

Analysis of Sparse Recovery Algorithms via the Replica Method

A Comprehensive Introduction to the Applications of the Replica Method in Analysis of Large Inference Problems

This is the initial version of the contribution to the book “Compressed Sensing in Information Processing”. The peer-reviewed version has been published in [12].

Ali Bereyhi, Ralf R. Müller and Hermann Schulz-Baldes

Abstract This manuscript goes through the fundamental connections between statistical mechanics and estimation theory by focusing on the particular problem of *compressive sensing*. We first show that the asymptotic analysis of a sparse recovery algorithm is mathematically equivalent to the problem of calculating the free energy of a *spin glass* in the thermodynamic limit. We then use the *replica method* from statistical mechanics to evaluate the performance in the asymptotic regime. The asymptotic results have several applications in communications and signal processing. We briefly go through two instances of these applications: Characterization of joint sparse recovery algorithms used in distributed compressive sensing, and tuning of receivers employed for detection of spatially modulated signals.

1 Introduction

Statistical mechanics deals with the analysis of very large many-particle systems and seeks the following ultimate goal: Starting from the *microscopic* behavior of individual particles, it tries to find out the *macroscopic* properties of the system. The system size is, however, so large that it is not possible to solve the microscopic equations of motion. Statistical mechanics follows an alternative approach: It describes

Ali Bereyhi

Institute for Digital Communications, Friedrich-Alexander Universität Erlangen-Nürnberg, Cauerstrasse 7, 91058, Erlangen e-mail: ali.bereyhi@fau.de

Ralf R. Müller

Institute for Digital Communications, Friedrich-Alexander Universität Erlangen-Nürnberg, Cauerstrasse 7, 91058, Erlangen e-mail: ralf.r.mueller@fau.de

Hermann Schulz-Baldes

Department of Mathematics, Friedrich-Alexander Universität Erlangen-Nürnberg, Cauerstrasse 11, 91058, Erlangen e-mail: schuba@mi.uni-erlangen.de

the microscopic behavior of the system particles via a stochastic model and extracts the desired *deterministic* properties via statistical analysis.

The goal and techniques of statistical mechanics are in various aspects similar to those of information theory. This connection has been widely investigated in the literature [29]. In addition to all interesting theoretical aspects of this connection, the links between the two theories lead to a key achievement: The analytical tools of statistical mechanics can be used to address asymptotic analysis in information theory and its applications.

In this manuscript, we use one particular statistical mechanical tool, namely the replica method, to investigate the asymptotic performance of a large class of sparse recovery algorithms. The interest in characterizing the asymptotic performance has several origins: The most natural one is to have an analytic bound on the performance of a given recovery scheme. This is however not the only application. Sparse recovery is used in several other applications in which the asymptotic performance characterization is useful for system design. In Section 7, we give two particular instances; namely, the example of deriving error bounds in distributed compressive sensing, and tuning algorithms used for detection of spatially modulated signals.

The focus of this manuscript is on the asymptotic analysis of a generic compressive sensing setting via the replica method. As a result, the contents give a comprehensive overview on the replica method and its applications to asymptotic analyses in communications and signal processing. The discussions in this manuscript can be followed in greater detail in [2].

2 A Multi-Terminal Setting for Compressive Sensing

We consider a generic multi-terminal setting. This setting includes the classical *single-terminal* compressive sensing setting, as well as other scenarios of sparse recovery.

Consider a distributed sensing network (DSN) with J *correlated sparse* source signals, namely $x_j(t)$ for $j \in [J]$. Here, the notation $[J]$ is defined as

$$[J] := \{1, \dots, J\}, \quad (1)$$

and is used through the manuscript to shorten the presentation. The source signals are sampled at the time instances, $t = t_n$ for $n \in [N]$. Let $\mathbf{x}_j \in \mathbb{X}^{N \times 1}$ for $\mathbb{X} \subset \mathbb{R}$ denote the vector of samples collected from the j -th source signal. We assume that the sampling is performed, such that the *temporal* correlation among different time samples is negligible¹. As the result, the sample vectors are statistically modeled as follows: $\mathbf{x}_1, \dots, \mathbf{x}_J$ are independent and identically distributed (i.i.d.) such that the n -th time sample of the source signals are *spatially* correlated.

The spatial correlation of the time samples at $t = t_n$ is modeled via the joint probability distribution $p_{\mathbb{X}}(\mathbf{x}_n^J)$, where we define

¹ This is typically the case in classic signal sampling techniques.

$$\mathbf{x}_n^J := (x_{1n}, \dots, x_{Jn}). \quad (2)$$

Hence, the joint distribution of the all signal samples is given by

$$p(\mathbf{x}^J) = \prod_{n=1}^N p_X(x_n^J), \quad (3)$$

where the notation \mathbf{x}^J is defined as

$$\mathbf{x}^J := (\mathbf{x}_1, \dots, \mathbf{x}_J). \quad (4)$$

Considering source signal j , an individual sensing unit collect M_j linear (and potentially noisy) observations of the samples. Denoting the vector of observations by $\mathbf{y}_j \in \mathbb{R}^{M_j \times 1}$, we can write

$$\mathbf{y}_j = \mathbf{A}_j \mathbf{x}_j + \mathbf{z}_j. \quad (5)$$

Here, $\mathbf{A}_j \in \mathbb{R}^{M_j \times N}$ denotes the *sensing matrix* of unit j which describe the linear transform from the signal samples to the observations, and $\mathbf{z}_j \in \mathbb{R}^{M_j \times 1}$ is measurement noise at terminal j .

The observations, as well as the sensing matrices, are given to a *single* data-fusion center. The data fusion center recovers the signal samples using a *joint* recovery algorithm, i.e., it finds the estimates $\hat{\mathbf{x}}^J = \hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_J$ as

$$\hat{\mathbf{x}}^J = \mathbf{g}(\mathbf{y}_1, \dots, \mathbf{y}_J | \mathbf{A}_1, \dots, \mathbf{A}_J), \quad (6)$$

via some recovery algorithm $\mathbf{g}(\cdot | \mathbf{A}_1, \dots, \mathbf{A}_J)$. At this point, we consider a generic form for the recovery algorithms. We will later focus on a specific (but broad) class of sparse recovery algorithms which use the method of least-squares.

2.1 Characterization of the Recovery Performance

Before illustrating the details of the system model, let us clarify the ultimate goal of this manuscript, i.e., the asymptotic analysis of sparse recovery algorithms. To this end, we first need to define a metric which characterizes the performance of a recovery algorithm $\mathbf{g}(\cdot | \mathbf{A}_1, \dots, \mathbf{A}_J)$. This metric is defined in the following definition:

Definition 1 (Average distortion) Consider a distortion function

$$\Delta(\cdot; \cdot) : \mathbb{R}^J \times \mathbb{R}^J \mapsto \mathbb{R}. \quad (7)$$

Using this function, the distortion between the source samples \mathbf{x}^J and their corresponding estimates $\hat{\mathbf{x}}^J$ is determined as

$$\Delta(\hat{\mathbf{x}}^J; \mathbf{x}^J) = \sum_{n=1}^N \Delta(\hat{x}_n^J; x_n^J). \quad (8)$$

The average distortion is then given by

$$D_N = \frac{1}{N} \mathbb{E} \left\{ \Delta(\hat{\mathbf{x}}^J; \mathbf{x}^J) \right\} \quad (9)$$

where $\mathbb{E} \{ \cdot \}$ indicates mathematical expectation.

The average distortion describes the quality of the recovery algorithm. Depending on the choice of the distortion function, the average distortion determines different forms of estimation errors. For example, by setting

$$\Delta(\hat{x}_n^J; x_n^J) = \sum_{j=1}^J |\hat{x}_{jn} - x_{jn}|^2, \quad (10)$$

the average distortion reduces to the well-known mean squared error (MSE).

It is important to keep in mind that the average distortion explicitly depends on the recovery algorithm. In fact, depending on the choice of $\mathbf{g}(\cdot | \mathbf{A}_1, \dots, \mathbf{A}_J)$, the estimated samples in $\hat{\mathbf{x}}^J$ change, and consequently, D_N varies.

Our ultimate goal is to find the average distortion for a large class of sparse recovery algorithms when the number of signal samples per each terminal, i.e., N , is very large. For a most known sparse recovery algorithms, this is a hard task to do, due to reasons which we explain in Section 4.

In order to address this goal, we need to specify a model for every component of the setting. We do this in the following section.

2.2 Stochastic Model of System Components

A typical model for the noise processes is the additive white Gaussian noise (AWGN) model. This follows from the fact that noise in sensing devices is physically caused by several random independent processes whose spectral density in the bandwidth of interest is well approximated by that of AWGN. Considering the AWGN model, the vector \mathbf{z}_j is considered to be an i.i.d. Gaussian random vector whose elements are zero-mean with variance σ_j^2 .

We also model the sensing matrices as they are generated by a random process. Although stochastic modeling of noise is widely accepted, considering such model for sensing matrices requires a bit of illustration. A stochastic model for sensing matrices assumes that each sensing matrix \mathbf{A}_j is taken at random from a predefined

ensemble. The logic behind considering such a model is as follows: From the compressive sensing literature, we know that sensing matrices require to satisfy some specific properties, such that a certain recovery performance is guaranteed [20]. Many random ensembles are shown to satisfy these properties. This means that by generating a sensing matrix from these ensembles at random, the anticipated recovery performance is achieved with a high probability. To incorporate this fact into the analysis, the classic approach is to assume that the sensing matrices are given by a random ensemble. From the mathematical viewpoint, such an assumption does not harm the generality of the analysis, as structured sensing matrices can be described by an ensemble as well.

In this manuscript, we assume that the sensing matrices are *right rotationally invariant* random matrices. This is generic assumption, since it includes most well-known random ensembles, e.g., the class of i.i.d. sensing matrices. We introduce this random ensemble in the sequel. However, before defining that let us first define the density of states for a given matrix.

Definition 2 (Density of states) Let $\mathbf{S} \in \mathbb{R}^{N \times N}$ be a self-adjoint square matrix whose eigenvalues are given by $\lambda_1, \dots, \lambda_N \in \mathbb{R}$. The density of states for this matrix is defined as the empirical cumulative distribution function (CDF) of its eigenvalues, i.e., for $\lambda \in \mathbb{R}$

$$F_{\mathbf{S}}^N(\lambda) = \frac{1}{N} \sum_{n=1}^N \mathbf{1}\{\lambda_n < \lambda\}. \quad (11)$$

We are now ready to define right rotationally invariant random matrices.

Definition 3 (Right rotationally invariant matrices) $\mathbf{A}_j \in \mathbb{R}^{M_j \times N}$ is right rotationally invariant if the Gramian $\mathbf{J}_j = \mathbf{A}_j^T \mathbf{A}_j$ has the decomposition

$$\mathbf{J}_j = \mathbf{U}_j \mathbf{D}_j \mathbf{U}_j^T \quad (12)$$

where \mathbf{U}_j and \mathbf{D}_j have the following properties:

1. $\mathbf{U}_j \in \mathbb{R}^{N \times N}$ is a Haar distributed matrix².
2. $\mathbf{D}_j \in \mathbb{R}^{N \times N}$ denotes the diagonal matrix of eigenvalues whose density of states is equal to the density of states of \mathbf{J}_j , i.e., $F_{\mathbf{D}_j}^N(\lambda) = F_{\mathbf{J}_j}^N(\lambda) = F_{\mathbf{A}_j}^N(\lambda)$.

The class of right rotationally invariant matrices includes most well-known random ensembles in compressive sensing. Examples of such ensembles are i.i.d. and row-orthogonal random sensing matrices. Note that different forms of random matrices will have different densities of states. We clarify this point further in Section 4 when we formally formulate the asymptotic analysis of a sparse recovery algorithm.

² A Haar distributed matrix is a random matrix generated uniformly on the set of all orthogonal matrices.

2.3 Stochastic Model for Jointly Sparse Signals

As indicated, we assume that there exists a spatial correlation among the signals for various sources. Noting that the signals are sparse, we interpret this spatial coupling as *joint sparsity*. From statistical viewpoint, this means that the joint probability density function (PDF) $p_X(x_n^J)$ contains one or multiple Dirac impulses. These impulses occur on subspaces of \mathbb{X}^J , where at least one of the vector entries is set to zero.

To give an intuition on joint sparsity, we focus in the sequel on a special joint sparsity model in which the sample n of signal j is written as

$$x_{jn} = c_n w_{0n} + s_{0n} w_{jn} + s_{jn} u_{jn}. \quad (13)$$

Here, s_{0n} , w_{0n} , c_n , w_{jn} , s_{jn} and u_{jn} are independent i.i.d. sequences whose distributions are as follows:

1. w_{0n} , w_{jn} and u_{jn} are random variables on \mathbb{X} whose probabilities of being zero are equal to zero.
2. s_{0n} , c_n and s_{jn} are Bernoulli random variables with

$$\Pr\{c_n = 1\} = 1 - \Pr\{c_n = 0\} = \mu_c \quad (14a)$$

$$\Pr\{s_{0n} = 1\} = 1 - \Pr\{s_{0n} = 0\} = \mu_0 \quad (14b)$$

$$\Pr\{s_{jn} = 1\} = 1 - \Pr\{s_{jn} = 0\} = \mu_j. \quad (14c)$$

In this model, the samples of a terminal are given as the superposition of three sparse components. The first component, whose n -th entry is given by $c_n w_{0n}$, is a sparse vector which is common among the all terminals. The second sparse component, represented by $s_{0n} w_{jn}$ for $n \in [N]$, has a common support³ across all the terminals; however, the value of the non-zero entries are drawn from independent processes. The last component contains a sparse signal whose support and non-zero entries are independently generated for each terminal.

Although the given model for joint sparsity is not the most general one, it includes most well-known sparse recovery settings in the literature. In the sequel, we address the main settings for sparse recovery. These settings are derived from our system model as special cases.

2.4 Special Cases

The three main settings for sparse recovery are classical compressive sensing, the problem of multiple measurement vectors (MMV), and distributed compressive sensing (DCS). In the sequel, we briefly go through these settings and illustrate how they are derived from our generic multi-terminal setting.

³ By support, we refer to the index of non-zero entries in a vector.

2.4.1 Classical Compressive Sensing

In classical compressive sensing, also called the single measurement vector problem, a sparse signal is observed linearly via a *single* terminal and is to be recovered from the underdetermined set of observations. This settings is simply derived by setting $J = 1$ in our model. As spatial correlation among terminals has no meaning in this case, one can further set $c_n = s_{0n} = 0$ for $n \in [N]$ in the sparsity model.

2.4.2 Multiple Measurement Vectors

In MMV, multiple sparse signals are observed with a common sensing matrix and recovered at a single data fusion center. This setting is straightforwardly derived from our model by letting

$$\mathbf{A}_1 = \dots = \mathbf{A}_J. \quad (15)$$

In general, the joint sparsity model given in Section 2.3 is a valid model in MMV. Nevertheless, in many applications of MMV, it is common to assume a the *common support* model for the spatial correlation. This model assumes that the samples of different terminals have common support; however, the non-zero entries are drawn from independent processes. The common support model is derived from the joint sparsity model in Section 2.3 by setting $s_{jn} = c_n = 0$ for $n \in [N]$ and $j \in [J]$.

2.4.3 Distributed Compressive Sensing

DCS describes the most generic setting which fits to our model. In this problem, the jointly sparse signals of different terminals are observed with different matrices and recovered at a common fusion center. Similar to MMV, the joint sparsity model in Section 2.3 is generally valid for DCS. A common model is however the *common-innovation* model in which the signal of each terminal is given as a common sparse component superposed by an independent sparse innovation term. This model is derived from the one given in Section 2.3 by setting $s_{jn} = 0$ for $n \in [N]$ and $j \in [J]$.

3 RLS-Based Recovery Algorithms

We focus on the class of regularized least squares (RLS)-based recovery algorithms. These algorithms recover the signal samples by minimizing a penalized residual sum of squares. In general, an RLS-based algorithm is of the following form:

$$\mathbf{g}(\mathbf{y}_1, \dots, \mathbf{y}_J | \mathbf{A}_1, \dots, \mathbf{A}_J) = \underset{\mathbf{v}_1, \dots, \mathbf{v}_J \in \mathbb{X}^N}{\operatorname{argmin}} \sum_{j=1}^J \frac{1}{2\lambda_j} \|\mathbf{y}_j - \mathbf{A}_j \mathbf{v}_j\|^2 + u_v(\mathbf{v}^J). \quad (16)$$

Here, $\mathbf{v}^J := (\mathbf{v}_1, \dots, \mathbf{v}_J)$, and $u_{\mathbf{v}}(\cdot) : \mathbb{R}^{JN \times 1} \mapsto \mathbb{R}^+$ is the *regularization function* which penalizes the residual sum of squares, and $\|\cdot\|$ denotes the Euclidean norm. In the sequel, we assume that $u_{\mathbf{v}}(\cdot)$ decouples, i.e., there exist $u(\cdot) : \mathbb{R}^{1 \times J} \mapsto \mathbb{R}^+$, such that

$$u_{\mathbf{v}}(\mathbf{v}^J) = \sum_{n=1}^N u(v_n^J). \quad (17)$$

Furthermore, $\lambda_1, \dots, \lambda_J$ are tunable factors, referred to as *regularizers*.

The interest in this class of recovery schemes comes from its generality. In fact, the recovery scheme in (16) includes a diverse set of sparse recovery algorithms, e.g., ℓ_p -norm minimization techniques in classical compressive sensing and $\ell_{p,q}$ -norm minimization techniques in DCS; we will discuss these examples with more details later on in this section. But before that, let us give a Bayesian interpretation of an RLS-based recovery scheme.

3.1 Bayesian Interpretation

In the Bayesian view point, an RLS-based recovery scheme is a maximum-a-posteriori (MAP) estimator in which

1. The *prior* joint distribution of samples at t_n is *postulated* to be proportional to $\exp\{-u(\cdot)\}$. This means that $p_X(x_n^J)$ is *assumed* to be

$$p_X(x_n^J) = \frac{\exp\{-u(\cdot)\}}{Z} \quad (18)$$

for some normalization factor Z .

2. The noise processes are postulated to be Gaussian.
3. The variance of noise at terminal j is *postulated* to be λ_j .

It is worth mentioning that the postulated parameters are not necessarily matched to the true ones. Hence, the MAP estimator is in general *mismatched*.

3.2 Special Cases

As mentioned, the RLS-based scheme includes several sparse recovery algorithms. In the sequel, we discuss two well-known examples, namely ℓ_p -norm minimization for classical compressive sensing and $\ell_{p,q}$ -norm minimization for DCS.

3.2.1 ℓ_p -Norm Minimization

Most algorithms in compressive sensing with a single terminal recover the sparse signal by finding a vector of samples whose residual sum of squared is bounded, i.e., finding \mathbf{v} , such that⁴

$$\|\mathbf{y} - \mathbf{A}\mathbf{v}\|^2 \leq \epsilon \quad (19)$$

for some ϵ , and whose ℓ_p -norm for some p is minimum. The most common choice of p is $p = 1$ which results in the least absolute shrinkage and selection operator (LASSO) or basis pursuit algorithm.

Using the method of Lagrange multipliers, it is shown that there exists a regularizer λ , for which the RLS-based algorithm with decoupled regularization function $u(v_n) = |v_n|^p$ performs identical to this algorithm.

3.2.2 $\ell_{p,q}$ -Norm Minimization

For multi-terminal settings, the classic ℓ_p -norm minimization techniques are often extended to $\ell_{p,q}$ -norm minimization techniques. The feasible set in this case is constructed with the same approach, i.e., finding $\mathbf{v}_1, \dots, \mathbf{v}_J$ for which

$$\|\mathbf{y}_j - \mathbf{A}_j\mathbf{v}_j\|^2 \leq \epsilon_j \quad (20)$$

with some ϵ_j for $j \in [J]$. The recovered samples are then found by searching the feasible for vectors whose $\ell_{p,q}$ -norm is minimum. For a collection of J vectors $\mathbf{v}_1, \dots, \mathbf{v}_J$, the $\ell_{p,q}$ -norm is defined as

$$\|\mathbf{v}_1, \dots, \mathbf{v}_J\|_{p,q} = \sum_{n=1}^N \left[\left(\sum_{j=1}^J |v_{jn}|^p \right)^{1/p} \right]^q. \quad (21)$$

The most well known $\ell_{p,q}$ -norm minimization technique is the group LASSO technique in which $p = 2$ and $q = 1$.

Similar to ℓ_p -norm minimization, one can invoke the method of Lagrange multipliers and show that there exist regularizers $\lambda_1, \dots, \lambda_J$, for which the RLS-based algorithm with decoupled regularization function $u(v_n^j) = (\|v_n^j\|_p)^q$ performs identical to $\ell_{p,q}$ -norm minimization.

4 Asymptotic Characterization

Now that the system model and recovery algorithm are presented, we are ready to formally formulate the asymptotic performance of a sparse recovery algorithm.

⁴ Note that index j is dropped, as we consider a single terminal setting.

For asymptotic analysis, we consider a sequence of settings. The number of signal samples N and the number of measurements at terminal j , i.e., M_j , in this sequence grow large, such that M_j is a deterministic function of N . We assume that N grows unboundedly large, and M_j grows with N linearly. This means that there exists a fixed ρ_j (typically $\rho_j \leq 1$) for each $j \in [J]$, such that

$$\rho_j := \lim_{N \uparrow \infty} \frac{M_j}{N} < \infty. \quad (22)$$

We refer to ρ_j as the j -th terminal compression ratio.

For every DSN in the sequence, we use an RLS-based algorithm to recover the signal samples. Let D_N denote the average distortion between the true signal samples and their estimates in the DSN whose index is N . The asymptotic analysis intends to find the asymptotic limit of this sequence distortions, i.e.,

$$D := \lim_{N \uparrow \infty} D_N. \quad (23)$$

The derivation of D and its dependence on the various model parameters are not straightforward from both analytical and computational points of view. In fact, depending on the regularization function, the derivation of D deals with one or two of the following issues:

- For some regularization functions, the RLS-based algorithm solves a *convex* optimization problem. This means that the recovery is posed as a linear programming. Hence, it is performed in polynomial time. Although RLS-based recovery in this case is computationally tractable, there is no guarantee that the problem is also *analytically* tractable. For asymptotic analysis, one needs to determine the sequence of average distortions for any integer index N and take the limit when N goes to ∞ . For some particular RLS-based algorithms, this task can be done via basic analytic tools; nevertheless, there are several forms whose limit is not known analytically via the basic tools.
- Several RLS-based algorithms are not only analytically, but also computationally intractable. An example is the ℓ_0 -norm minimization algorithm in which the regularization function is proportional to the ℓ_0 -norm. In such cases, the recovery algorithm deals with a nondeterministic polynomial time (NP)-hard problem [20]. Clearly, for these forms, asymptotic characterization is not tractable.

The above analytical and computational issues can be addressed via the replica method. As it becomes clear later, the replica method invokes several non-rigorous tricks to bypass the analytical obstacles of the problem. The term *non-rigorous tricks* will be clarified in the next sections of this manuscript while we illustrate how the replica method exactly does that.

Now that the asymptotic analysis is formulated in principle, we can state explicitly our main purpose as follows:

The main purpose of this manuscript is to illustrate how the asymptotic distortion D is derived for an RLS-based algorithm via the replica method.

4.1 Stieltjes and R-Transforms

Before we start with the illustration of the replica method, we give some basic definitions which are used throughout the derivations via the replica method. These definitions enables us to represent compactly the statistics of the sensing matrices.

To start with the definitions, consider the sequence of DSNs indexed by N . For each terminal, there exists a corresponding sequence of densities $F_j^N(\lambda)$ which for a given index N describes the density of states of $\mathbf{J}_j = \mathbf{A}_j^\top \mathbf{A}_j$. We assume that this sequence converges as $N \rightarrow \infty$ to a deterministic density of state $F_j(\lambda)$ for each $j \in [J]$. For these asymptotic densities, the Stieltjes and R-transforms are defined as follows [42]:

Definition 4 (Stieltjes Transform) For the asymptotic density $F_j(\lambda)$, the Stieltjes transform is given by

$$G_j(s) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - s} dF_j(\lambda) \quad (24)$$

for some complex s with $\text{Im } s \geq 0$, where $\text{Im } s$ is the imaginary part of s .

Definition 5 (R-Transform) For the asymptotic density $F_j(\lambda)$, the R-transform is defined as

$$R_j(\omega) = G_j^{-1}(-\omega) - \frac{1}{\omega} \quad (25)$$

such that

$$\lim_{\omega \downarrow 0} R_j(\omega) = \int_{-\infty}^{+\infty} \lambda dF_j(\lambda). \quad (26)$$

$G_j^{-1}(\cdot)$ denotes the inverse of the Stieltjes transform with respect to composition.

This definition of the R-transform is further extended to matrix arguments: Consider a matrix $\mathbf{S}_{N \times N}$ with the eigendecomposition

$$\mathbf{S} = \mathbf{\Sigma} \mathbf{\Lambda} \mathbf{\Sigma}^{-1}. \quad (27)$$

For this matrix, we use the notation $R_j(\mathbf{S})$ to refer to

$$R_j(\mathbf{S}) := \mathbf{\Sigma} \text{Diag} \{R_j(\lambda_1), \dots, R_j(\lambda_n)\} \mathbf{\Sigma}^{-1} \quad (28)$$

where $\text{Diag} \{a_1, \dots, a_N\}$ denotes an $N \times N$ diagonal matrix whose diagonal entries are a_1, \dots, a_N .

5 Building a Bridge to Statistical Mechanics

As mentioned before, the replica method was initially developed in statistical mechanics for analysis of spin glasses. Nevertheless, it found its way to several other fields, such as coding, information theory and signal processing. The key point in employing the replica method for asymptotic analysis is to make a connection between the problem in hand and the theory of spin glasses. In this section, we illustrate how this connections is made. To this end, we need first to give a quick overview on basic definitions in statistical mechanics.

5.1 Introduction to Statistical Mechanics

A thermodynamic system consists of N particles with each having a microscopic parameter $v_n \in \mathbb{V}$ for $n \in [N]$ and some set \mathbb{V} . This parameter describes a macroscopic property of the corresponding particle, e.g., the velocity. In general, a microscopic parameter could be a vector of continuous or discrete entries. For sake of brevity, we assume that v_n is a *continuous* scalar. The extension to cases with discrete \mathbb{V} can be followed in [2, Chapter 3]. For this system, the *microstate* is defined as a vector in \mathbb{V}^N which collects microscopic parameters of all the particles, i.e.

$$\mathbf{v} = [v_1, \dots, v_N]^T. \quad (29)$$

Corresponding to this system, a *Hamiltonian* $\mathcal{E}(\cdot)$ is defined which describes the physical properties of the system. The Hamiltonian is a function which assigns to microstate \mathbf{v} a non-negative energy level $\mathcal{E}(\mathbf{v})$.

A Side Note

Here, we have defined the Hamiltonian in an abstract form. For a physical system, the explicit form of the Hamiltonian is derived from the physical theories which describe the interactions of microscopic parameters in the system.

For a thermodynamic system, the explicit calculation of macroscopic parameters is intractable⁵. To address this issue, statistical mechanics follows a stochastic approach. In this approach, the microstate is considered to be a random vector whose distribution depends on the *temperature*. We denote this distribution by $p_\beta(\mathbf{v})$ where β is the *inverse temperature*, i.e., $\beta = 1/T$ with T being the temperature.

Using stochastic analysis, statistical mechanics derives physical features of the thermodynamic system from this stochastic model. These physical features are known as *macroscopic* parameters of the system. Mathematically, a thermodynamic system

⁵ This follows the same reasons given in Section 4 for the asymptotic analysis of RLS-based algorithms.

can be described by two main macroscopic parameters, namely the *entropy* and *free energy*⁶. These parameters are defined as follows:

Definition 6 (Normalized entropy) For a given thermodynamic system with N particles, the normalized entropy at inverse temperature β is defined as

$$\mathcal{H}_N(\beta) := -\frac{1}{N} \int_{\mathbb{V}^N} p_\beta(\mathbf{v}) \log p_\beta(\mathbf{v}) \, d\mathbf{v} \quad (30)$$

Definition 7 (Normalized free energy) Consider a thermodynamic system with N particles and Hamiltonian $\mathcal{E}(\cdot)$. At inverse temperature β , the normalized free energy is defined as

$$\mathcal{F}_N(\beta) := \frac{1}{N} \mathbb{E} \{ \mathcal{E}(\mathbf{v}) \} - \frac{1}{\beta} \mathcal{H}_N(\beta). \quad (31)$$

5.1.1 Second Law of Thermodynamics

The fundamental rule in stochastic analysis of thermodynamic systems is the second law of thermodynamic. This law indicates that the microstate in thermal equilibrium⁷ is distributed such that the free energy is minimized. Since $\mathcal{F}_N(\beta)$ is convex with respect to $p_\beta(\mathbf{v})$, it is concluded that the microstate is distributed with the *Boltzmann-Gibbs* distribution. This means that at thermal equilibrium

$$p_\beta(\mathbf{v}) = \frac{\exp\{-\beta\mathcal{E}(\mathbf{v})\}}{\mathcal{Z}_N(\beta)} \quad (32)$$

at inverse temperature β . $\mathcal{Z}_N(\beta)$ in the denominator is a normalization factor, i.e.,

$$\mathcal{Z}_N(\beta) = \int_{\mathbb{V}^N} \exp\{-\beta\mathcal{E}(\mathbf{v})\} \, d\mathbf{v}, \quad (33)$$

and is called the *partition function*. $p_\beta(\mathbf{v})$ reduces to some well-known distributions for several choices of the Hamiltonian, e.g., it reduces to the Gaussian distribution when $\mathcal{E}(\mathbf{v}) \propto \|\mathbf{v}\|^2$.

⁶ In fact, the main two macroscopic parameters of a thermodynamic system are entropy and energy. The free energy is derived by applying the second law of thermodynamic as the Lagrange dual function. We however use directly the free energy in our formulation, for sake of brevity.

⁷ This means that there is no energy flow.

A Side Note

The stated form of the second law of thermodynamics is a simplified interpretation of the original form. In fact, the law states that the entropy in an isolated system grows constantly. This is interpreted as a constrained optimization problem in which the normalized entropy is maximized subject to an energy constraint. Using the method of Lagrange multipliers, the free energy is derived as the objective function of the dual unconstrained optimization. It is then shown that the Lagrange multiplier is in fact the temperature.

Substituting the Boltzmann-Gibbs distribution in the definition of the free energy, it is concluded that

$$\mathcal{F}_N(\beta) = -\frac{1}{\beta N} \log \mathcal{Z}_N(\beta). \quad (34)$$

This is a fundamental identity indicating that the free energy of a system in thermal equilibrium is calculated explicitly from the partition function. Starting from this equation, it is shown that all other macroscopic parameters of the system are directly derived from $\mathcal{F}_N(\beta)$. For instance,

$$\mathcal{H}_N(\beta) = \beta^2 \frac{d}{d\beta} \mathcal{F}_N(\beta). \quad (35)$$

Therefore, the partition function completely describes the macroscopic features of the system in thermal equilibrium.

5.1.2 Spin Glasses

Spin glasses are thermodynamic systems whose particles choose to interact randomly. This means that the Hamiltonian of a spin glass is not only a function of the microstate, but also a randomizer. This randomizer is realized once from a random ensemble and remains fixed as the system is in thermal equilibrium⁸.

Similar to thermodynamic systems, the stochastic analysis of spin glasses follows the second law of thermodynamic. Let Ω denote the randomizer of a spin glass. The Hamiltonian of this spin glass is given by

$$\mathcal{E}(\cdot|\Omega) : \mathbb{V}^N \mapsto \mathbb{R}^+. \quad (36)$$

In other words, for every realization of Ω , we have a specific Hamiltonian function. By the same lines of derivations explained in Section 5.1.1, one can show that *conditioned* to the randomizer, the microstate of the spin glass in thermal equilibrium

⁸ In statistical mechanics, this randomizer is known to have *quenched* randomness. This is different from the type of randomness considered for the microstate which is called *annealed* randomness.

is distributed with the Boltzmann-Gibbs distribution. This means that

$$p_{\beta}(\mathbf{v}|\mathbf{\Omega}) = \frac{\exp\{-\beta\mathcal{E}(\mathbf{v}|\mathbf{\Omega})\}}{\mathcal{Z}_N(\beta|\mathbf{\Omega})} \quad (37)$$

with random partition function

$$\mathcal{Z}_N(\beta|\mathbf{\Omega}) = \int_{\mathbb{V}^N} \exp\{-\beta\mathcal{E}(\mathbf{v}|\mathbf{\Omega})\} d\mathbf{v} \quad (38)$$

The normalized free energy in thermal equilibrium is hence written as

$$\mathcal{F}_N(\beta|\mathbf{\Omega}) = -\frac{1}{\beta N} \log \mathcal{Z}_N(\beta|\mathbf{\Omega}), \quad (39)$$

and the *conditional* entropy is determined from the free energy by

$$\mathcal{H}_N(\beta|\mathbf{\Omega}) = \beta^2 \frac{d}{d\beta} \mathcal{F}_N(\beta|\mathbf{\Omega}). \quad (40)$$

In the remaining, we focus on spin glasses. This is due to the fact that our problem is formulated in terms of a spin glass.

5.1.3 Thermodynamic Limit

Spin glasses are studied in the *thermodynamic limit*. This means that the macroscopic parameters are derived for the case, in which the number of particles tends to infinity, i.e., the asymptotic limit $N \uparrow \infty$. Suggested by physical intuition, in the thermodynamic limit, a spin glass has deterministic macroscopic parameters. This means that in the asymptotic limit, the free energy $\mathcal{F}_N(\beta|\mathbf{\Omega})$ converges to its expected value⁹. This property of spin glasses is known as *self averaging*; more discussions in this respect can be followed in [33, 22, 21].

Following the self-averaging property, the free energy of a spin glasses in the thermodynamic limit is calculated as follows:

1. Determining sequence of *expected* free energies $\bar{\mathcal{F}}_N(\beta|\mathbf{\Omega})$ indexed by N as

$$\bar{\mathcal{F}}_N(\beta) = \mathbb{E}\{\mathcal{F}_N(\beta|\mathbf{\Omega})\}, \quad (41)$$

where the expectation is taken with respect to $\mathbf{\Omega}$.

2. Taking the asymptotic limit of the expected sequence, i.e., calculating

$$\bar{\mathcal{F}}(\beta) = \lim_{N \uparrow \infty} \bar{\mathcal{F}}_N(\beta). \quad (42)$$

⁹ Here, the expectation is taken over the randomizer $\mathbf{\Omega}$.

5.1.4 Averaging Trick

Before we start with the derivations, let us illustrate the key *averaging trick* in statistical mechanics. Consider a function $\psi_N(\cdot)$ which for each microstate $\mathbf{v} \in \mathbb{V}^N$ determines an scalar parameter. The macroscopic parameter corresponding to this function is defined as

$$\bar{\psi}_N = \frac{1}{N} \mathbb{E} \{ \psi_N(\mathbf{v}) \} \quad (43)$$

where the expectation is taken first with respect to the conditional Boltzmann-Gibbs distribution, i.e., $p_\beta(\mathbf{v}|\mathbf{\Omega})$, and then with respect to $\mathbf{\Omega}$.

The classic approach for determining ψ_N in statistical mechanics is to use the averaging trick. This trick modifies the partition function with a dummy factor h as follows

$$\mathcal{Z}_N(\beta, h|\mathbf{\Omega}) = \int_{\mathbf{v} \in \mathbb{V}^N} \exp \{ -\beta \mathcal{E}(\mathbf{v}|\mathbf{\Omega}) + h \psi_N(\mathbf{v}) \} d\mathbf{v} \quad (44)$$

For this modified partition function, the normalized free energy, conditioned to a realization of the randomizer, is

$$\mathcal{F}_N(\beta, h|\mathbf{\Omega}) = -\frac{1}{\beta N} \log \mathcal{Z}_N(\beta, h|\mathbf{\Omega}), \quad (45)$$

and its expected value $\bar{\mathcal{F}}_N(\beta, h)$ is determined by calculating the expectation over $\mathbf{\Omega}$, i.e., as in (41).

By standard derivations, it is readily shown that

$$\bar{\psi}_N = -\beta \frac{\partial}{\partial h} \bar{\mathcal{F}}_N(\beta, h) |_{h=0} \quad (46)$$

Exchanging limiting procedures, one has in the thermodynamic limit

$$\bar{\psi} := \lim_{N \uparrow \infty} \bar{\psi}_N \quad (47a)$$

$$= -\beta \frac{\partial}{\partial h} \bar{\mathcal{F}}(\beta, h) |_{h=0}. \quad (47b)$$

5.2 Corresponding Spin Glass

The connection between the sparse recovery problem and statistical mechanics is illustrated by introducing the concept of *corresponding spin glass*. In fact, for an RLS-based recovery algorithm, we can define an *imaginary* spin glass whose macroscopic parameters are the asymptotic performance metric of the recovery algorithm. We clarify this connection in the sequel.

Remember the system model in Section 2 with sensing matrices $\mathbf{A}_1, \dots, \mathbf{A}_J$ and observation vectors $\mathbf{y}_1, \dots, \mathbf{y}_J$. We define corresponding spin glass as follows:

Corresponding Spin Glass

The corresponding spin glass is a spin glass whose microstate is described by $\mathbf{v}^J = (\mathbf{v}_1, \dots, \mathbf{v}_J)$ where $\mathbf{v}_j \in \mathbb{X}^N$ for $j \in [J]$. The randomizer of this spin glass is

$$\mathbf{\Omega} = \{\mathbf{A}_1, \dots, \mathbf{A}_J, \mathbf{y}_1, \dots, \mathbf{y}_J\}, \quad (48)$$

and its Hamiltonian is

$$\mathcal{E}(\mathbf{v}^J | \mathbf{\Omega}) = \sum_{j=1}^J \frac{1}{2\lambda_j} \|\mathbf{y}_j - \mathbf{A}_j \mathbf{v}_j\|^2 + u_{\mathbf{v}}(\mathbf{v}^J). \quad (49)$$

From our earlier discussions, we know that at inverse temperature β , the microstate in thermal equilibrium is conditionally distributed with

$$p_{\beta}(\mathbf{v}^J | \mathbf{\Omega}) = \frac{\exp\{-\beta \mathcal{E}(\mathbf{v}^J | \mathbf{\Omega})\}}{\mathcal{Z}_N(\beta | \mathbf{\Omega})} \quad (50)$$

where partition function $\mathcal{Z}_N(\beta | \mathbf{y}, \mathbf{A})$ reads

$$\mathcal{Z}_N(\beta | \mathbf{\Omega}) = \int_{\mathbf{v}_j \in \mathbb{X}^N} \exp\{-\beta \mathcal{E}(\mathbf{v}^J | \mathbf{\Omega})\} d\mathbf{v}^J. \quad (51)$$

The key property of this spin glass which connects it to our sparse recovery problem is its *ground state property*.

Ground State Property

For a given realization of $\mathbf{\Omega}$, assume that the Hamiltonian has a unique minimizer denoted by $\mathbf{v}_{\star}^J(\mathbf{\Omega})$. Then, as the temperature goes to zero, i.e., $\beta \uparrow \infty$, the microstate of the corresponding spin glass converges in distribution to the deterministic vector $\mathbf{v}_{\star}^J(\mathbf{\Omega})$. This means that for every realization of $\mathbf{\Omega}$

$$\lim_{\beta \uparrow \infty} p_{\beta}(\mathbf{v}^J | \mathbf{\Omega}) = \begin{cases} 1 & \mathbf{v}^J = \mathbf{v}_{\star}^J(\mathbf{\Omega}) \\ 0 & \mathbf{v}^J \neq \mathbf{v}_{\star}^J(\mathbf{\Omega}) \end{cases}. \quad (52)$$

In fact, this is a well-known property in statistical mechanics: At zero temperature, the microstate converges in distribution to a realization whose energy level is minimized. The appellation follows the fact that this realization, i.e., $\mathbf{v}_{\star}^J(\mathbf{\Omega})$, is called the *ground state* of the system.

The ground state property clarifies the connection between our problem and this spin glass: In fact, the ground state is what the RLS-based algorithm recovers, i.e.,

$$\mathbf{v}_{\star}^J(\boldsymbol{\Omega}) = \hat{\mathbf{x}}^J. \quad (53)$$

In other words, as the temperature goes to zero the microstate of the corresponding spin glass converges to the signal samples which are recovered via the algorithm. Hence, the performance metrics of this sparse recovery algorithm, e.g., the asymptotic distortion, are derived as the macroscopic parameters of this spin glass at zero temperature.

The corresponding spin glass shows several other interesting properties. Interested readers are suggested to see [2, Chapter 3].

5.2.1 Asymptotic Distortion as a Macroscopic Parameter

The main purpose of this manuscript is to determine the asymptotic distortion. As indicated, this metric can be defined as a macroscopic parameter of the corresponding spin glass. To show that, consider the following macroscopic parameter

$$\psi_N(\mathbf{v}^J) = \Delta(\mathbf{v}^J; \mathbf{x}^J) \quad (54)$$

where \mathbf{x}^J refer to the true signal samples. The macroscopic parameter defined by this function is

$$\bar{\psi} = \lim_{N \uparrow \infty} \frac{1}{N} \mathbb{E} \{ \psi_N(\mathbf{v}) \} \quad (55a)$$

$$= \lim_{N \uparrow \infty} \frac{1}{N} \mathbb{E} \{ \Delta(\mathbf{v}^J; \mathbf{x}^J) \} \quad (55b)$$

As the temperature goes to zero, $\beta \uparrow$, the microstate \mathbf{v}^J converges to $\hat{\mathbf{x}}^J$. Hence, at zero-temperature we have

$$\bar{\psi} \rightarrow \lim_{N \uparrow \infty} \frac{1}{N} \mathbb{E} \{ \Delta(\hat{\mathbf{x}}^J; \mathbf{x}^J) \} \quad (56a)$$

$$= \lim_{N \uparrow \infty} D_N \quad (56b)$$

$$= D. \quad (56c)$$

The last equation clarifies how the asymptotic distortion is derived from the corresponding spin glass.

Using the averaging trick, we can find D from the following expected modified free energy in the thermodynamic limit

$$\bar{\mathcal{F}}(\beta, h) = - \lim_{N \uparrow \infty} \frac{1}{\beta N} \mathbb{E} \{ \log \mathcal{Z}_N(\beta, h | \Omega) \}, \quad (57)$$

at zero temperature as

$$D = - \lim_{\beta \uparrow \infty} \beta \frac{\partial}{\partial h} \bar{\mathcal{F}}(\beta, h) |_{h=0}, \quad (58)$$

where the partition function is given by

$$\mathcal{Z}_N(\beta, h | \Omega) = \int_{\mathbf{v}_j \in \mathbb{X}^N} \exp \left\{ -\beta \mathcal{E}(\mathbf{v}^J | \Omega) + h \Delta(\mathbf{v}^J; \mathbf{x}^J) \right\} d\mathbf{v}^J. \quad (59)$$

5.3 The Replica Method

The variational problem derived in terms of the corresponding spin glass suffers from the same intractability issue we observed in the original problem. In the original problem, the intractability issue was due to the optimization problem. This is now transformed to a *logarithmic expectation* in (57). This is not a trivial task, and hence keeps the problem still intractable.

The replica method tries to calculate this logarithmic expectation with a series of tricks. The first step is to use the Riesz identity [35]:

Riesz Identity

For a non-negative random variable X , we have

$$\mathbb{E} \{ \log X \} = \lim_{\theta \downarrow 0} \frac{\log \mathbb{E} \{ X^\theta \}}{\theta} \quad (60)$$

Using this identity, one can rewrite the logarithmic expectation of (57) as

$$\mathbb{E} \{ \log \mathcal{Z}_N(\beta, h | \Omega) \} = \lim_{\theta \downarrow 0} \frac{\log \mathbb{E} \{ \mathcal{Z}_N^\theta(\beta, h | \Omega) \}}{\theta}. \quad (61)$$

The right hand side now deals with logarithm of an expectation. The problem is however still intractable, since θ in the right hand side of the identity is a *real* scalar: The intractability of logarithmic expectation is now transformed to the challenge of *calculating real moments*. Here, the second step is applied:

Replica Continuity

We assume that the moment function is real analytic on the real axis. This means that the function

$$f_M(\theta) = \mathbb{E} \left\{ \mathcal{Z}_N^\theta(\beta, h | \mathbf{\Omega}) \right\} \quad (62)$$

can be analytically continued from integer values to all real θ .

This second step is *not mathematically rigorous*. This is why the replica method is often called the *replica trick*. The available results suggest that this is a valid assumption; however, the proof is still an open problem.

Assuming θ to be an integer finally resolves the intractability issue at the expense of losing mathematical rigor. We now can write the moment function as¹⁰

$$f_M(\theta) = \mathbb{E} \left\{ \mathcal{Z}_N^\theta(\beta, h | \mathbf{\Omega}) \right\} \quad (63a)$$

$$= \mathbb{E} \left\{ \prod_{a=1}^{\theta} \int \exp \left\{ -\beta \mathcal{E}(\mathbf{v}_a^J | \mathbf{\Omega}) + h \Delta(\mathbf{v}_a^J; \mathbf{x}^J) \right\} d\mathbf{v}_a^J \right\} \quad (63b)$$

$$= \int \mathbb{E} \left\{ \exp \left\{ -\beta \sum_{a=1}^{\theta} \mathcal{E}(\mathbf{v}_a^J | \mathbf{\Omega}) + h \Delta(\mathbf{v}_a^J; \mathbf{x}^J) \right\} \right\} d\mathbf{v}_1^J \dots d\mathbf{v}_\theta^J. \quad (63c)$$

The latter integral is complicated but tractable. The main remained task is to calculate this integral and find it as an analytic function in θ . We then replace it into the Riesz identity, and take the limits. In the sequel, we give a quick overview on the derivations.

6 The Replica Analyses

The derivation of $f_M(\theta)$ from (63c) takes a long amount of time, and is out of the scope for this manuscript. We hence present the derivations steps and skip the details. Interested readers are referred to [2, Appendices A-D].

We start the derivation by taking expectation with respect to noise. This task is done via basic properties of Gaussian integrals. We then use the results in [18, 26, 23] on the asymptotic limit of *spherical integrals* to calculate the expectation with respect to the sensing matrices. Some short notes on spherical integrals and their asymptotic limits are found in [2, Appendix E]. Finally, we use the law of large numbers to take the expectation with respect to the true signal samples \mathbf{x}^J .

After taking the expectations, we finally conclude that

¹⁰ In the notation, we drop the integration set for sake of compactness.

$$f_M(\theta) = \int \exp \left\{ -N E_M(\mathbf{Q}^J, \mathbf{S}^J) + \epsilon_N \right\} d\mathbf{Q}^J d\mathbf{S}^J \quad (64)$$

where the exponent function $E_M(\mathbf{Q}^J, \mathbf{S}^J)$ is defined as

$$E_M(\mathbf{Q}^J, \mathbf{S}^J) = \sum_{j=1}^J [\mathcal{G}_j(\mathbf{T}_j \mathbf{Q}_j) + \text{tr}\{\mathbf{S}_j \mathbf{Q}_j\}] - \mathcal{M}(\mathbf{S}^J). \quad (65)$$

\mathbf{Q}^J and \mathbf{S}^J are further defined as

$$\mathbf{Q}^J = (\mathbf{Q}_1, \dots, \mathbf{Q}_J) \quad (66a)$$

$$\mathbf{S}^J = (\mathbf{S}_1, \dots, \mathbf{S}_J) \quad (66b)$$

with \mathbf{Q}_j and \mathbf{S}_j being symmetric $\theta \times \theta$ matrices for $j \in [J]$. The exact definition of integral measures $d\mathbf{Q}_j$ and $d\mathbf{S}_j$ is given in [2, Appendix A]. ϵ_N is a bounded sequence in N which converges to zero as N grows large. \mathbf{T}_j is defined as

$$\mathbf{T}_j = \frac{1}{2\lambda_j} \left(\mathbf{I}_\theta - \frac{\beta\sigma_j^2}{\lambda_j + \theta\beta\sigma_j^2} \mathbf{1}_\theta \right) \quad (67)$$

where \mathbf{I}_θ and $\mathbf{1}_\theta$ denote $\theta \times \theta$ identity and all-one matrices, respectively. The components of the exponent function $E_M(\mathbf{Q}^J, \mathbf{S}^J)$ are further defined as follows:

- The function $\mathcal{G}_j(\cdot)$ is given by

$$\mathcal{G}_i(\mathbf{M}) := \int_0^\beta \text{tr}\{\mathbf{M} \mathbf{R}_j(-2\mathbf{M}\omega)\} d\omega \quad (68)$$

for a $\theta \times \theta$ matrix \mathbf{M} .

- $\mathcal{M}(\mathbf{S}^J)$ is defined as

$$\mathcal{M}(\mathbf{S}^J) = \mathbb{E} \left\{ \log \int_{\mathbf{v}_j \in \mathbb{X}^\theta} \exp \left\{ \Xi(\mathbf{v}^J, \mathbf{x}^J | \mathbf{S}^J) + h\Delta(\mathbf{v}^J; \mathbf{x}^J) \right\} d\mathbf{v}^J \right\} \quad (69)$$

where the function $\Xi(\mathbf{v}^J, \mathbf{x}^J | \mathbf{S}^J)$ is given by

$$\Xi(\mathbf{v}^J, \mathbf{x}^J | \mathbf{S}^J) = \sum_{j=1}^J (\mathbf{x}_j - \mathbf{v}_j)^\top \mathbf{S}_j (\mathbf{x}_j - \mathbf{v}_j) - \beta u_v(\mathbf{v}^J). \quad (70)$$

In these equations, $\mathbf{v}^J = (\mathbf{v}_1, \dots, \mathbf{v}_J)$ and $\mathbf{x}^J = (\mathbf{x}_1, \dots, \mathbf{x}_J)$, where $\mathbf{v}_j \in \mathbb{X}^\theta$ and $\mathbf{x}_j = x_j \mathbf{1}_{\theta \times 1}$ for $j \in [J]$. $\mathbf{1}_{\theta \times 1}$ denotes a $\theta \times 1$ all-one vector and x_1, \dots, x_J are correlated random variables distributed jointly with $p_X(x_1, \dots, x_J)$. It is worth mentioning that the term $u_v(\mathbf{v}^J)$ is decomposed as

$$u_{\mathbf{v}}(\mathbf{v}^J) = \sum_{a=1}^{\theta} u(v_a^J) \quad (71)$$

using the decoupling property of the regularization function $u_{\mathbf{v}}(\cdot)$. Here, $\mathbf{v}_a^J = [v_{1a}, \dots, v_{Ja}]$ with v_{ja} denoting the a -th entry of \mathbf{v}_j .

Remark

The definition of $f_M(\theta)$ contain integrals over N -dimensional vectors. These integrals are transformed to integrals over θ -dimensional vectors in the final expression. This transform follows several steps and assumptions. The detailed derivations can be followed in [2, Appendix A].

6.1 General Form of the Solution

The final form of the moment function in (64) enables us to apply the saddle-point method to derive the free energy in the thermodynamic limit. After some lines of derivation, we conclude that the asymptotic distortion is given by

$$D = \lim_{u \downarrow 0} \lim_{\beta \uparrow \infty} \int_{\mathbf{v}_j \in \mathbb{X}^{\theta}} \mathbb{E} \left\{ \Delta(\mathbf{v}^J; \mathbf{x}^J) q_{\beta}(\mathbf{v}^J | \mathbf{x}^J) \right\} d\mathbf{v}^J. \quad (72)$$

The conditional distribution $q_{\beta}(\mathbf{v}^J | \mathbf{x}^J)$ in this equation is a Boltzmann-Gibbs distribution over the reduced dimension and is defined as

$$q_{\beta}(\mathbf{v}^J | \mathbf{x}^J) = \frac{\exp\{-\beta E_0(\mathbf{v}^J, \mathbf{x}^J)\}}{\int_{\mathbf{v}_j \in \mathbb{X}^{\theta}} \exp\{-\beta E_0(\mathbf{v}^J, \mathbf{x}^J)\} d\mathbf{v}^J} \quad (73)$$

where the exponent function is defined as

$$E_0(\mathbf{v}^J, \mathbf{x}^J) = \sum_{j=1}^J (\mathbf{x}_j - \mathbf{v}_j)^{\top} \mathbf{R}_j (\mathbf{x}_j - \mathbf{v}_j) + u_{\mathbf{v}}(\mathbf{v}^J), \quad (74)$$

and the expectation is taken with respect to \mathbf{x}^J . \mathbf{R}_j in the exponent function is further defined as

$$\mathbf{R}_j := \mathbf{T}_j \mathbf{R}_j \left(-2\beta \mathbf{T}_j \mathbf{Q}_j^* \right) \quad (75)$$

where the symmetric $\theta \times \theta$ matrix \mathbf{Q}_j^* for $j \in [J]$ is calculated from the following fixed-point equation:

$$\mathbf{Q}_j^* = \int_{\mathbf{v}_j \in \mathbb{X}^\theta} \mathbb{E} \left\{ (\mathbf{x}_j - \mathbf{v}_j) (\mathbf{x}_j - \mathbf{v}_j)^\top q_\beta(\mathbf{v}^J | \mathbf{x}^J) \right\} d\mathbf{v}^J. \quad (76)$$

Remark

To see how (76) describes a fixed-point equation, note that the conditional distribution $q_\beta(\mathbf{v}^J | \mathbf{x}^J)$ depends on \mathbf{Q}_j^* . As the result, the right hand side of this identity is calculated as a function of \mathbf{Q}_j^* , and (76) describes a fixed-point equation in \mathbf{Q}_j^* .

6.2 Constructing Parameterized \mathbf{Q}_j^*

The general solution of the replica method is given in terms of $\theta \times \theta$ matrices \mathbf{Q}_j^* . The reason for having such a solution is simply the *replica continuity* assumption. In this assumption, we postulated that θ is an *integer*. For an integer θ , having a $\theta \times \theta$ matrix is completely reasonable. Nevertheless, we aim to find the final solution as an analytic function in θ , so that we could use it also for *real* choices of θ .

To find an analytic solution, there exists a classic trick: *Assuming an structure on \mathbf{Q}_j^** . In this trick, we limit the search to a set of parameterized matrices. The parameterization is considered such that the solution of the fixed-point equation leads to an analytic moment function in θ .

In order to clarify this trick, consider the following illustration: We assume that \mathbf{Q}_j^* is a $\theta \times \theta$ matrix which is parameterized by L parameters $q^{(1)}, \dots, q^{(L)}$. This means that

$$\mathbf{Q}_j^* = W_j(q^{(1)}, \dots, q^{(L)}) \quad (77)$$

where $W_j(\cdot)$ is a deterministic function which determines a $\theta \times \theta$ matrix for given scalar arguments $q^{(1)}, \dots, q^{(L)}$. Note that L is an integer whose value is fixed and does not vary by changing θ . By inserting this matrix into the fixed-point equation, a system of L coupled equations in terms of $q^{(1)}, \dots, q^{(L)}$ is derived. We insert the solution of this equation system into the replica solution and calculate the limits analytically.

With respect to this trick the following question arises: *What is a meaningful structure for \mathbf{Q}_j^* ?* The answer to this question is based on physical intuitions and mathematical investigations of an energy model. These discussions are out of the scope of this overview; however, their results can be directly applied to our study. The investigations in theory of spin glasses suggests a set of recursively extendable structures drawn from the assumption of *replica symmetry (RS)*. These structures

start by a simple symmetric parameterization, known as RS, and then extends to more advanced structures by iteratively perturbing the RS.

6.2.1 Replica Symmetric Solution

RS considers the most basic structure on \mathbf{Q}_j^* which depends only on two parameters q_j and χ_j , and is given by

$$\mathbf{Q}_j^* = \frac{\chi_j}{\beta} \mathbf{I}_\theta + q_j \mathbf{1}_\theta. \quad (78)$$

Using this structure, the asymptotic distortion is determined as follows:

Replica Symmetric Solution

Consider random variables $\mathbf{x}^J = (x_1, \dots, x_J)$ distributed with $p_X(\mathbf{x}^J)$. Let y_j for $j \in [J]$ be calculated from x_j as

$$y_j = x_j + z_j \quad (79)$$

with z_j being independent Gaussian random variables with zero mean and variance ξ_j^2 . The variances ξ_j^2 are determined in terms of parameters χ_j and q_j as

$$\xi_j^2 = \left[\mathbf{R}_j \left(-\frac{\chi_j}{\lambda_j} \right) \right]^{-2} \frac{\partial}{\partial \chi_j} \left[\left(\sigma_j^2 \chi_j - \lambda_j q_j \right) \mathbf{R}_j \left(-\frac{\chi_j}{\lambda_j} \right) \right]. \quad (80)$$

Define $\hat{\mathbf{x}}^J$ as

$$\hat{\mathbf{x}}^J = \operatorname{argmin}_{\mathbf{v}_j \in \mathbf{X}} \sum_{j=1}^J \frac{1}{2\tau_j} (y_j - v_j)^2 + u(\mathbf{v}^J), \quad (81)$$

where τ_j is given by

$$\tau_j = \frac{\lambda_j}{\mathbf{R}_j \left(-\frac{\chi_j}{\lambda_j} \right)}. \quad (82)$$

The RS solution for asymptotic distortion in this case is given by

$$D = \mathbb{E} \left\{ \Delta \left(\hat{\mathbf{x}}^J; \mathbf{x}^J \right) \right\} \quad (83)$$

when χ_j and q_j are calculated from the following fixed point equations:

$$q_j = \mathbb{E} \left\{ (\hat{x}_j - x_j)^2 \right\} \quad (84a)$$

$$\theta_j^2 \chi_j = \tau_j \mathbb{E} \left\{ (\hat{x}_j - x_j) z_j \right\}. \quad (84b)$$

The expectation is taken over all random variables, i.e., \mathbf{x}^J and \mathbf{z}^J .

The RS solution is calculated readily. In fact, (81) is a J -dimensional optimization which can be solved analytically in various cases. For most well-known RLS-based recovery algorithms, the RS solution gives a valid prediction of the asymptotic distortion. Nevertheless, there are few particular cases in which the RS solution is invalid¹¹. This inconsistency is due to the simplicity of the RS structure. For those cases, one needs to break RS.

6.2.2 Replica Symmetry Breaking

For scenarios in which the RS solution is not valid, the search for \mathbf{Q}_j^* is extended to a wider set of parameterized matrices via the replica symmetry breaking (RSB) scheme. This scheme was introduced by Parisi in [32]. The scheme perturbs the RS gradually via a recursive technique. This perturbation is called *breaking*.

Replica Symmetry Breaking

Let θ be an integer multiple of an integer ζ , and \mathbf{Q}_ℓ represent a $\zeta \times \zeta$ matrix. RSB finds the new $\theta \times \theta$ matrix $\mathbf{Q}_{\ell+1}$ as

$$\mathbf{Q}_{\ell+1} = \mathbf{I}_{\frac{\theta}{\zeta}} \otimes \mathbf{Q}_\ell + q_{\ell+1} \mathbf{1}_\theta \quad (85)$$

for some real scalar $q_{\ell+1}$. Here, \otimes denotes the Kronecker product.

By letting \mathbf{Q}_0 to be an RS matrix, the RSB structures are recursively generated. The RSB solutions are of more complicated form. We hence skip them and refer interested readers to [2, Chapter 4].

7 Applications and Numerical Results

The asymptotic characterization of RLS-based recovery algorithms enables us to address several tasks which rise in various applications of sparse recovery. In this section, we briefly go through a few of them. The scope of these applications however is not limited to these instances. We have given more discussions in this respect in [2, 8, 14, 10, 11, 13, 16, 6, 37, 38, 17, 5, 36, 3, 15, 7, 1, 4].

¹¹ In fact, the predicted distortion deviates the known rigorous bounds.

7.1 Performance Analysis of Sparse Recovery

The most relevant application of the results is to employ them for asymptotic investigation of sparse recovery algorithms. A long discussion in this respect is found in [2, Chapter 6], as well as [10, 11, 6]. As a particular instance, we employ the asymptotic results to study the impact of spatial correlation in multi-terminal compressive sensing.

For sake of visualization, we consider a simple setting with two terminals. These terminals observe signals $x_1(t)$ and $x_2(t)$ which are jointly sparse. We assume that the joint sparsity follows the *common-innovation* model; see Section 2.4.3.

The fusion center can recover the sparse signal via two approaches:

1. Since each signal is sparse, the fusion center can use two separate sparse recovery algorithms to recover each sparse signal *individually*.
2. A *joint* recovery algorithm can be used to take into account the spatial correlation among the terminals.

From Slepian-Wolf theorem, it is known that the joint recovery can outperform an individual scheme [19]. To investigate this issue, we consider a sample RLS-based recovery algorithm for each approach and compare their performance using the asymptotic characterization. For the individual approach, we consider the well-known LASSO algorithm. This algorithm is realized by setting the regularization function to

$$u_v(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1\|_1 + \|\mathbf{v}_2\|_1. \quad (86)$$

As a comparable joint recovery, one can use an RLS-based joint recovery scheme with convex utility, e.g., the group LASSO algorithm in which

$$u_v(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1, \mathbf{v}_2\|_{2,1}. \quad (87)$$

In the sequel, we use the two-dimensional LASSO technique proposed initially in [6]. This algorithm extends the individual LASSO recovery approach by modifying the regularization function as

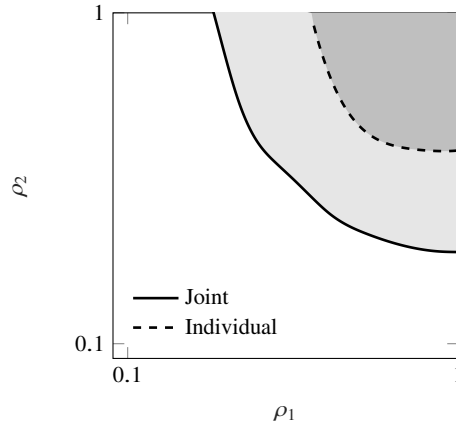
$$u_v(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1\|_1 + \|\mathbf{v}_2\|_1 + \phi\|\mathbf{v}_1 + \alpha\mathbf{v}_2\|_1 \quad (88)$$

for some scalars ϕ and α . The intuition behind this algorithm is that any linear combination of jointly sparse signals is also sparse and its sparsity level depends on the spatial correlation. The study in [6] has shown that this approach outperforms the classic group LASSO technique for the common-innovation joint sparsity model.

Using the RS solution, we can calculate the asymptotic MSE for both approaches. The asymptotic MSE is determined from the RS solution by setting the distortion function to the squared Euclidean distance between the true and recovered pairs. Using the asymptotic MSE, we plot the *rate-distortion* region for both schemes. It is found by fixing a threshold MSE and finding all pairs of compression rates, i.e., (ρ_1, ρ_2) for which the achievable MSE is smaller than the threshold. This region is shown in Fig. 1 for a particular example in which the common part is 30% sparse

and each terminal has a 10% sparse innovation component. The tunable factors in both algorithms are optimized to achieve minimal MSE. As the figure shows, using a spatially coupled regularization improves the recovery performance significantly. The Bayesian viewpoint illustrates this observation as follows: The postulated prior distribution of an RLS-based algorithm with spatially coupled regularization takes into account the spatial correlation, and hence outperforms the individual approach.

Fig. 1 Rate-distortion region for both joint and individual LASSO schemes.



The asymptotic characterization can be further used to investigate those recovery algorithms whose performances are unknown. An example is a non-convex $\ell_{p,q}$ -norm minimization, e.g.,

$$u_v(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1, \mathbf{v}_2\|_{2,q} \quad (89)$$

with $0 \leq q \leq 1$. These algorithms are not computationally tractable, and hence their performance is left unknown. The asymptotic characterization via the replica method can give a prediction on the performance of these algorithm. More discussions can be followed in [2, Chapter 6].

7.2 Tuning RLS-Based Algorithms

Compressive sensing is not the only application of sparse recovery. In fact, sparse recovery is used in various applications, such as communications, networking and machine learning; see some instances in [2, 14, 13, 16, 37, 38, 17, 5, 36, 3, 15, 7, 1, 4, 30]. In these applications, there is often a tuning task: Find the *regularizers* of an RLS-based algorithm, such that the performance is optimized. This task is readily addressed via the asymptotic characterization of the RLS-based recovery algorithms.

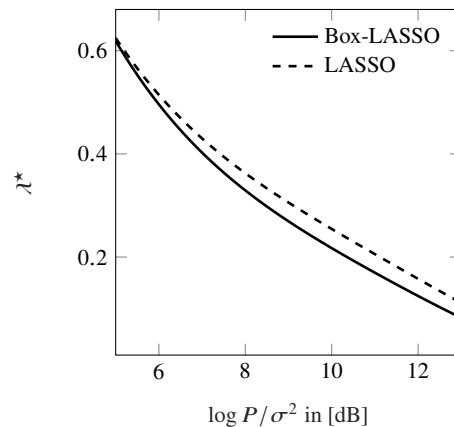
We can illustrate this application by considering a simple example of *spatial modulation*. The details on this example can be followed in [3, 4]. In spatial modulation,

the information is encoded in the support of the transmit signal: In each symbol interval, based on the data bits, a subset of available transmit antennas is set on and the remaining are turned off. As a result, the transmit signal is sparse, and hence an effective detection scheme at the receiver is to use a sparse recovery algorithms¹².

The common sparse recovery algorithms used in *spatial modulation* are formulated as RLS-based recovery schemes. Examples are the classic LASSO and *box-LASSO* techniques. We already know the classic LASSO scheme from the previous section. The *box-LASSO* technique is moreover an extension of LASSO in which the set \mathbb{X} is restricted to a box, e.g., $\mathbb{X} = [-B, B]$ for some real B . This box restriction is shown to enhance the performance, when we detect discrete-valued signals.

One of the challenges in these techniques is to find the optimal regularizers, which results in minimum bit error rate. Such a task is usually addressed via iterative tuning techniques. Nevertheless, in high data rates, the tuning techniques impose extra processing load on the system. The asymptotic characterization enables us to address this task analytically, and hence avoid the extra load. An instance of tuning via the asymptotic characterization is shown in Fig. 2. In this figure, a multiuser uplink scenario is considered in which the LASSO and box-LASSO techniques are used for detection. Here, P denotes the transmit power and σ^2 is the noise variance at the receiver. The sparsity of the transmit signal is assumed to be 12.5%. The figure shows the optimal regularizer, denoted by λ^* against $\log P/\sigma^2$. Although these results are derived via the asymptotic characterization, the study in [4] shows that they closely track the simulation results. Further discussions regarding the tuning of RLS-based algorithms via asymptotic results can be followed in [2, Chapters 6 and 7], as well as [17].

Fig. 2 Optimal regularizer for LASSO and box-LASSO.



¹² For sake of brevity, we skip the detailed system model. Interested readers are referred to [3, 4] and the references therein.

8 Summary and Final Discussions

The replica method is a powerful tool for large system analyses, as seen this in this manuscript. Following the prescription suggested by the replica method, we have found an analytic expression for the asymptotic distortion. The result could not be derived via basic analytical tools. This demonstrates the power of the *replica method*. To keep the contents of this manuscript straightforward, we have dropped the detailed derivations and only presented the major steps. The details can be followed in [2].

The presented analysis is extendable in various respects, and results in various further interesting conclusions. Going through all of these extensions and conclusions is not possible within a short manuscript. We hence skip them here and refer interested readers to [2] and the references therein. Nevertheless, to give you a flavor, we conclude this manuscript by giving a few highlights.

8.1 Decoupling Principle

Although this manuscript focused on the derivation of *asymptotic distortion*, the result can be further used to prove the so-called *decoupling principle*. This principle indicates that in the asymptotic regime the joint distribution of x_n^J and \hat{x}_n^J converges to the one described via an *equivalent* scalar system, often called the *decoupled system*. This decoupled setting is shown to consist of an equivalent additive noise term and a decoupled recovery scheme; see [2, Chapter 5]. The interesting point is that the decoupled recovery scheme remains the same for all solutions, i.e., the RS and RSB solutions, and it is only the distribution of the equivalent noise term which changes. A comprehensive illustration of the decoupling principle and its detailed derivations are given in [2, Chapter 5].

8.2 Nonuniform Sparsity Patterns

In various applications, the sparsity of signals varies over time. This form of sparsity is often called *nonuniform*, whereas the normal form is considered *uniform*. For nonuniform sparse signals, the stochastic model of samples is not i.i.d. anymore. They are still independent¹³; however, the joint distribution changes through time. The analysis in this manuscript extends to nonuniform patterns by some modifications. Some results on this direction can be followed in [9, 15].

¹³ Since temporal correlation is usually avoided by classic sampling approaches.

8.3 Extensions to Bayesian Estimation

In the Bayesian framework, the considered RLS-based algorithms are seen as MAP estimators. This is however not the only approach for Bayesian inference. In many other applications, e.g., signal processing and machine learning, other forms of Bayesian inference are used, e.g., the minimum MSE estimator or more generally estimators with minimal posterior distortion; see for example [30].

The replica based analysis in this manuscript is readily extended to these estimators as well. The derivations follow the same steps as illustrated in this manuscript, i.e., finding a corresponding spin glass and interpreting the desired metrics as its macroscopic parameters. The key difference here is that for other estimators, the desired metrics might be a macroscopic parameter at a *non-zero* temperature.

9 Bibliographical Notes

Primary studies on asymptotic analysis were limited to linear recovery techniques, e.g., studies in [24, 39]. Müller and Gerstaecker conjectured later that similar behavior extends to most nonlinear schemes, as well [31]. This conjecture was originated from the analytic results reported in a series of studies which employed the replica method to derive the asymptotic performance of multiuser detectors. This line of work started with the study by Tanaka in [40]. A key milestone in this direction was achieved in [25], where the authors determined the asymptotic performance metrics of a mismatched minimum MSE recovery scheme. This result was later extended to MAP estimators in [34] using standard large deviations techniques.

With respect to the problem of sparse recovery, most initial asymptotic analyses relied on the replica analyses presented in [40] and [25]. This led to results which enclosed restricted system models, e.g., single terminal and i.i.d. sensing matrix. The later lines of work deviated from this approach and used the replica method explicitly to derive the asymptotic characteristics; for instance [41, 43, 28, 27, 44]. These analyses were however limited to RS investigations. The complete replica analysis of RLS-based algorithms was given in a series of work in [11, 10, 8, 2] providing both RS and RSB solutions.

References

1. Asaad, S., Bereyhi, A., Müller, R.R., Schaefer, R.F.: Joint user selection and precoding in multiuser MIMO systems via group LASSO. Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (2019, Istanbul, Turkey)
2. Bereyhi, A.: Statistical Mechanics of Regularized Least Squares. Ph.D. Dissertation, Friedrich-Alexander University (2020)

3. Beryhi, A., Asaad, S., Gäde, B., Müller, R.R.: RLS-based detection for massive spatial modulation MIMO. *Proc. IEEE International Symposium on Information Theory (ISIT)* pp. 1167–1171 (2019, Paris, France)
4. Beryhi, A., Asaad, S., Gäde, B., Müller, R.R., Poor, H.V.: Detection of spatially modulated signals via RLS: Theoretical bounds and applications. *arXiv preprint arXiv:2011.06890* (2020)
5. Beryhi, A., Asaad, S., Müller, R.R., Chatzinotas, S.: RLS precoding for massive MIMO systems with nonlinear front-end. *Proc. IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (2019, Cannes, France)
6. Beryhi, A., Haghghatshoar, S., Müller, R.R.: Theoretical bounds on MAP estimation in distributed sensing networks. *Proc. IEEE International Symposium on Information Theory (ISIT)* (2018, Vail, CO, USA)
7. Beryhi, A., Jamali, V., Müller, R.R., Schober, G.F.R., Tulino, A.M.: PAPR-limited precoding in massive MIMO systems with reflect- and transmit-array antennas. *Proc. Asilomar Conference on Signals, Systems, and Computers (to be appeared)* (2019, Pacific Grove, CA, USA)
8. Beryhi, A., Müller, R., Schulz-Baldes, H.: RSB decoupling property of MAP estimators. *Proc. IEEE Information Theory Workshop (ITW)* pp. 379–383 (2016, Cambridge, UK)
9. Beryhi, A., Müller, R.R.: Maximum-a-posteriori signal recovery with prior information: Applications to compressive sensing. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* pp. 4494–4498 (2018, Calgary, AB, Canada)
10. Beryhi, A., Müller, R.R., Schulz-Baldes, H.: Replica symmetry breaking in compressive sensing. *Proc. IEEE Information Theory and Applications Workshop (ITA)* pp. 1–7 (2017, San Diego, CA, USA)
11. Beryhi, A., Müller, R.R., Schulz-Baldes, H.: Statistical mechanics of MAP estimation: General replica ansatz. *IEEE Transactions on Information Theory* **65**(12), 7896–7934 (2019)
12. Beryhi, A., Müller, R.R., Schulz-Baldes, H.: Analysis of Sparse Recovery Algorithms via the Replica Method, pp. 145–179. Springer International Publishing, Cham (2022)
13. Beryhi, A., Sedaghat, M.A., Asaad, S., Müller, R.: Nonlinear precoders for massive MIMO systems with general constraints. *Proc. VDE 21st International ITG Workshop on Smart Antennas (WSA)* pp. 1–8 (2017, Berlin, Germany)
14. Beryhi, A., Sedaghat, M.A., Müller, R.R.: Asymptotics of nonlinear LSE precoders with applications to transmit antenna selection. *Proc. IEEE International Symposium on Information Theory (ISIT)* pp. 81–85 (2017, Aachen, Germany)
15. Beryhi, A., Sedaghat, M.A., Müller, R.R.: RLS recovery with asymmetric penalty: Fundamental limits and algorithmic approaches. *Proc. 2nd International Balkan Conference on Communications and Networking* (2018, Podgorica, Montenegro)
16. Beryhi, A., Sedaghat, M.A., Müller, R.R.: Precoding via approximate message passing with instantaneous signal constraints. *Proc. International Zurich Seminar on Information and Communication (IZS)* pp. 128–132 (2018, Zürich, Switzerland)
17. Beryhi, A., Sedaghat, M.A., Müller, R.R., Fischer, G.: GLSE precoders for massive MIMO systems: Analysis and applications. *IEEE Transactions on Wireless Communications* **18**(9), 4450–4465 (2019)
18. Chandra, H.: Differential operators on a semisimple Lie algebra. *American Journal of Mathematics* pp. 87–120 (1957)
19. El Gamal, A., Kim, Y.H.: *Network information theory*. Cambridge University Press, Cambridge, UK (2011)
20. Foucart, S., Rauhut, H.: *A Mathematical Introduction to Compressive Sensing*. Springer, New York (2013). doi: 10.1007/978-0-8176-4948-7
21. Guerra, F., Toninelli, F.L.: The infinite volume limit in generalized mean field disordered models. *arXiv preprint cond-mat/0208579* (2002)
22. Guerra, F., Toninelli, F.L.: The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics* **230**(1), 71–79 (2002)
23. Guionnet, A., Zeitouni, O.: Large deviations asymptotics for spherical integrals. *Journal of functional analysis* **188**(2), 461–515 (2002)

24. Guo, D., Rasmussen, L.K., Lim, T.J.: Linear parallel interference cancellation in long-code CDMA multiuser detection. *IEEE Journal on Selected Areas in Communications* **17**(12), 2074–2081 (1999)
25. Guo, D., Verdú, S.: Randomly spread CDMA: Asymptotics via statistical physics. *IEEE Transactions on Information Theory* **51**(6), 1983–2010 (2005)
26. Itzykson, C., Zuber, J.B.: The planar approximation. II. *Journal of Mathematical Physics* **21**(3), 411–421 (1980)
27. Kabashima, Y., Wadayama, T., Tanaka, T.: A typical reconstruction limit for compressed sensing based on ℓ_p -norm minimization. *Journal of Statistical Mechanics: Theory and Experiment* **2009**(09), L09003 (2009)
28. Kabashima, Y., Wadayama, T., Tanaka, T.: Statistical mechanical analysis of a typical reconstruction limit of compressed sensing. *Proceedings of IEEE International Symposium on Information Theory (ISIT)* pp. 1533–1537 (2010)
29. Merhav, N.: Statistical physics and information theory. *Foundations and Trends in Communications and Information Theory* **6**(1-2), 1–212 (2010)
30. Müller, R.R., Bereyhi, A., Mecklenbräuker, C.F.: Oversampled adaptive sensing with random projections: Analysis and algorithmic approaches. *Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* pp. 336–341 (2018, Louisville, KY, USA)
31. Müller, R.R., Gerstacker, W.H.: On the capacity loss due to separation of detection and decoding. *IEEE Transactions on Information Theory* **50**(8), 1769–1778 (2004)
32. Parisi, G.: A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General* **13**(4), L115 (1980)
33. Pastur, L., Shcherbina, M.: Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model. *Journal of Statistical Physics* **62**(1-2), 1–19 (1991)
34. Rangan, S., Fletcher, A.K., Goyal, V.: Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing. *IEEE Transactions on Information Theory* **58**(3), 1902–1923 (2012)
35. Riesz, F.: Sur les valeurs moyennes des fonctions. *Journal of the London Mathematical Society* **1**(2), 120–121 (1930)
36. Schram, V., Bereyhi, A., Zaech, J.N., Müller, R.R., Gerstacker, W.H.: Approximate message passing for indoor THz channel estimation. *Proc. 3rd International Balkan Conference on Communications and Networking (2019, Skopje, North Macedonia)*
37. Sedaghat, M.A., Bereyhi, A., Müller, R.: A new class of nonlinear precoders for hardware efficient massive MIMO systems. *Proc. IEEE International Conference on Communications (ICC) (2017, Paris, France)*
38. Sedaghat, M.A., Bereyhi, A., Müller, R.R.: Least square error precoders for massive MIMO with signal constraints: Fundamental limits. *IEEE Transactions on Wireless Communications* **17**(1), 667–679 (2018)
39. Shamai, S., Verdú, S.: The impact of frequency-flat fading on the spectral efficiency of CDMA. *IEEE Transactions on Information Theory* **47**(4), 1302–1327 (2001)
40. Tanaka, T.: A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors. *IEEE Transactions on Information Theory* **48**(11), 2888–2910 (2002)
41. Tulino, A.M., Caire, G., Verdu, S., Shamai, S.: Support recovery with sparsely sampled free random matrices. *IEEE Transactions on Information Theory* **59**(7), 4243–4271 (2013)
42. Tulino, A.M., Verdú, S., et al.: Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory* **1**(1), 1–182 (2004)
43. Vehkaperä, M., Kabashima, Y., Chatterjee, S.: Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing. *IEEE Transactions on Information Theory* **62**(4), 2100–2124 (2016)
44. Wen, C.K., Zhang, J., Wong, K.K., Chen, J.C., Yuen, C.: On sparse vector recovery performance in structurally orthogonal matrices via lasso. *IEEE Transactions on Signal Processing* **64**(17), 4519–4533 (2016)