Mean Field Analysis of Sparse Reconstruction with Correlated Variables

Mohammad Ramezanali
Department of Physics and Astronomy,
Rutgers University, 136 Frelinghuysen Rd,
Piscataway, NJ 08854 USA
E-mail: mrr@physics.rutgers.edu

Partha P. Mitra
Cold Spring Harbor Laboratory,
1 Bungtown Road,
Cold Spring Harbor, NY 11734 USA
E-mail: mitra@cshl.edu

Anirvan M. Sengupta
Department of Physics and Astronomy,
Rutgers University, 136 Frelinghuysen Rd,
Piscataway, NJ 08854 USA
E-mail: anirvans@physics.rutgers.edu

***Abstract*—Sparse reconstruction algorithms aim to retrieve high-dimensional sparse signals from a limited number of measurements. A common example is LASSO or Basis Pursuit where sparsity is enforced using an ℓ_1-penalty together with a cost function $||\mathbf{y} - \mathbf{Hx}||_2^2$. For random design matrices $\mathbf{H}$, a sharp phase transition boundary separates the 'good' parameter region where error-free recovery of a sufficiently sparse signal is possible and a 'bad' regime where the recovery fails. However, theoretical analysis of phase transition boundary of the correlated variables case lags behind that of uncorrelated variables. Here we use replica trick from statistical physics to show that when an N-dimensional signal $\mathbf{x}$ is K-sparse and $\mathbf{H}$ is $M \times N$ dimensional with the covariance $E[H_{ia}H_{jb}] = \frac{1}{M}C_{ij}D_{ab}$, with all $D_{aa} = 1$, the perfect recovery occurs at $M \sim \psi_K(\mathbf{D})K\log(N/M)$ in the very sparse limit, where $\psi_K(\mathbf{D}) \geq 1$, indicating need for more observations for the same degree of sparsity.**

***Index Terms*—Compressed sensing, structured matrices, replica method, Basis Pursuit.**

I. Introduction

Compressed sensing [1], [2] plays an important role in modern signal processing, and is based on the idea that the observed signal is sparse in a suitable basis. Recovering the signal from measurements in a compressed sensing framework requires employing a sparse retrieval algorithm. The signal is modeled as $\mathbf{y} = \mathbf{Hx}$, where $\mathbf{y}$ is generated by $\mathbf{Hx}_0$, is an M-dimensional measurement vector, $\mathbf{H}$ is an $M \times N$ design matrix, and $\mathbf{x}_0$ is the N dimensional vector to be retrieved from the knowledge of $\mathbf{y}$ and $\mathbf{H}$. It is a priori known that $\mathbf{x}_0$ has at most K nonzero components and $M \ll N$. Algorithms such as Least absolute shrinkage and selection operator (LASSO) [3], the Elastic Net [4] as well as greedy algorithms [5] guarantee such recovery with high probability for sufficiently sparse signals.

Much of the theory concerning guaranteed performance bounds for CS is based on random design matrices corresponding to uncorrelated dependent variables. In particular, for design matrices that have independent and identically distributed Gaussian entries, the CS systems can robustly recover K-sparse signal from just $M = O(K\log(N/K))$ measurements [1], [2]. Previous analyses show that the performance failure of the ℓ_1 norm minimization method and other analogous algorithms with polynomial time complexity occurs at a sharp boundary as $N \to \infty$, with $\frac{M}{N}$ and $\frac{K}{N}$ being held fixed, analogous to a second-order (continuous) phase transition [6]–[9].

In practice, the typical design matrices are not I.I.D Gaussian, but can have very specific structure. In this case, the treatment of correlated design matrices is limited [10]–[13], and the reconstruction algorithm may not recover the original signal even in the noise-free case [14]. So far, most studies have focused on the constructions of design matrices with the goal of establishing performance bounds that are comparable to those of random matrices, for instance by satisfying the restricted isometry property (RIP) [15]. However, fewer analytical results on the behavior of the sparse recovery phase transition in the presence of correlations is available.

The *replica method* has been previously used in other contexts to study the behavior of design matrices with a factorized/Kronecker correlation structure [16]–[18]. In the case of compressed sensing, the authors of [12] employed this analytical technique to achieve perfect recovery close to the theoretical bound in linear time. However, this paper treat quasi-one dimensional correlations. To our knowledge, there are as of yet no analytical formulae for the phase boundary for general correlated design matrices. We provide a derivation of such formulae in this article, for the extremely sparse limit.

Outline of the Paper: We begin in Section II with a formalization of the problem setup and we propose the self-consistency mean field equations using replica method when the correlated design matrices are present. The derivation of these mean field equations is outlined in Appendix A. As a check, we solve these equations for the well studied example of Basis Pursuit in Section III. In the Section IV, we generalize the calculations to the case of correlated random design matrices in the very sparse limit. We then present numerical results for the special case of a symmetric Toeplitz matrix for the correlations.

II. Problem Formulation

Sparse retrieval can be accomplished by penalized regression with suitable penalties. This leads to estimates of the form $\hat{\mathbf{x}}(\vartheta = \lambda\sigma^2) = \arg\min_{\mathbf{x}} \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{Hx})^2 + \lambda \mathrm{V}(\mathbf{x})$. ϑ is a non-negative parameter giving relative weight between the first and second term and V is the penalty function. We focus on $V(\mathbf{x}) = \sum_a U(x_a)$ that is convex and separable and also

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

on the noise-free case where $\vartheta \to 0^+$. This is equivalent to the constrained optimization problem of minimizing $V(\mathbf{x})$ subject to the constraint $\mathbf{H}\mathbf{x}_0 = \mathbf{H}\mathbf{x}$. To understand the properties of the estimation error, we study the optimization of the function $\mathcal{E}(\mathbf{u})$ in terms of the error variable $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$, where

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

We consider the Gaussian distributed design matrices $\mathcal{P}(\mathbf{H})$ with the matrix elements correlated in a "factorized" manner, a special case that appears in many practical problems [16]–[18]. More precisely, the distribution $\mathcal{P}(\mathbf{H})$ has mean zero and covariance $\left[H_{ia}H_{jb}\right]^{\rm av} = \frac{1}{M}C_{ij}D_{ab}$. We use $\left[\cdots\right]^{\rm av}_{\rm vars}$ for quenched averages, with the relevant quenched variables indicated in the subscript, when necessary. Here $\mathbf{C}$,$\mathbf{D}$ are positive semidefinite symmetric matrices. The vector $\mathbf{x}_0$ is drawn from a distribution $P_0(\mathbf{x}_0) = \prod_a p_0(x_{a0})$. We choose the sparsity promoting distribution $p_0(x_{a0})$ which has a continuous part and a delta function at origin: $p_0(x_a) = \rho\pi(x_a) + (1-\rho)\,\delta(x_a)$.

Using replica mean field theory we show that the optimization of Eq. (1) reduces to the minimization of the following self-consistency equations (see Appendix A):

Proposition 1 (Effective Optimization for the Correlated Matrices)**.**

$$\hat{\mathbf{u}} = \min_{\mathbf{u}} \big\{ \frac{1}{2\sigma^2_{\rm eff}}(\mathbf{u}^\top\mathbf{D}\mathbf{u} - 2\boldsymbol{\xi}^\top\mathbf{D}\mathbf{u}) + \lambda V(\mathbf{u}+\mathbf{x}_0)\big\}, \tag{2}$$

We define

$$q = \frac{1}{N}[\hat{\mathbf{u}}^\top\mathbf{D}\hat{\mathbf{u}}]^{\rm av}_{\mathbf{x}_0,\boldsymbol{\xi}}. \tag{3}$$

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

$$\mathrm{Cov}(\xi,\xi) = (\frac{q}{\alpha})\frac{\mathbf{D}^{-1}\mathrm{tr}_M\big(\big[\mathbf{C}\Big\{\mathbf{I}_M\sigma^2 + \frac{\mathrm{tr}(\mathbf{D}\overline{\boldsymbol{\chi}})}{M}\mathbf{C}\Big\}^{-1}\big]^2\big)/M}{\Big(\mathrm{tr}_M\big(\mathbf{C}\Big\{\mathbf{I}_M\sigma^2 + \frac{\mathrm{tr}(\mathbf{D}\overline{\boldsymbol{\chi}})}{M}\mathbf{C}\Big\}^{-1}\big)/M\Big)^2} \tag{4}$$

and $\sigma^2_{\rm eff}$ can be obtained from

$$\frac{1}{\sigma^2_{\rm eff}} = \frac{1}{M}\mathrm{tr}_M\big(\mathbf{C}\Big\{\mathbf{I}_M\sigma^2 + \frac{\mathrm{tr}(\mathbf{D}\overline{\boldsymbol{\chi}})}{M}\mathbf{C}\Big\}^{-1}\big) \tag{5}$$

The symbol tr_M is a trace applying to the M dimensional space. The local susceptibility $\boldsymbol{\chi}$ is an $N\times N$ matrix obtained from the relation $\delta\mathbf{u} = \boldsymbol{\chi}\mathbf{f}$ with $f_a \to 0$ in which, $\hat{u}_a(f)$ is obtained by minimizing $\min_{\mathbf{u}}\big\{\frac{1}{2\sigma^2_{\rm eff}}(\mathbf{u}^\top\mathbf{D}\mathbf{u} - 2\boldsymbol{\xi}^\top\mathbf{D}\mathbf{u}) + \lambda V(\mathbf{u}+\mathbf{x}_0) - \mathbf{f}^\top\mathbf{u}\big\}$. Averaging over all the instances of the design matrix is expected to be self averaging in the large M, N limit yielding the average susceptibility matrix, $\overline{\boldsymbol{\chi}}$ [19]. We summarize the symbols used in the next section in the Table I.

III. Basis Pursuit with Uncorrelated design Matrix

In this section, we consider the well-studied case where the penalty function is the ℓ_1-norm of $\mathbf{x}$ and each element of the matrix $\mathbf{H}$ is I.I.D. normally distributed as $\mathcal{N}(0, 1/M)$ [1], [2], [20]. In this limit, optimization of Eq. (1) gives rise to a set of uncoupled univariate optimization problems:

TABLE I: Symbols that are being used in this article.

Symbol	Description
u_a	Measure of residual error $x_a - x_{0a}$
q	$\frac{1}{N}[\hat{\mathbf{u}}^\top\mathbf{D}\hat{\mathbf{u}}]^{\rm av}_{\mathbf{x}_0,\boldsymbol{\xi}}$ (MSE for Proposition 2)
α	Measure for the number of constraints, $\frac{M}{N}$
ρ	Measure for the sparsity, $\frac{K}{N}$
λ	ℓ_1-norm regression coefficient
σ^2	Error variance on the constraint $\mathbf{y} = \mathbf{H}\mathbf{x}$
ϑ	$\lambda\sigma^2$
$\sigma^2_{\rm eff}$	Effective σ^2 given in the large M, N limit
θ	$\lambda\sigma^2_{\rm eff}$
σ^2_ξ	$\frac{q}{\alpha}$ (In the uncorrelated case)
τ^2	$\frac{\alpha\theta^2}{q}$

Proposition 2 (Effective Individual Optimization)**.**

$$\hat{u}_a = \min_{u_a}\{\frac{1}{2\sigma^2_{\rm eff}}\left(u_a^2 - 2\xi_a u_a\right) + \lambda|u_a + x_{0a}|\} \tag{6}$$

$$q \equiv \sum_a [\hat{u}_a^2]^{\rm av}_{x_0,\xi} \tag{7}$$

$$\xi_a \in \mathcal{N}(0,\sigma^2_\xi) \;\text{ with }\; \sigma^2_\xi \equiv \frac{q}{\alpha} \tag{8}$$

$$\sigma^2_{\rm eff} \equiv \sigma^2 + \frac{\overline{\chi}}{\alpha} \quad \& \quad \overline{\chi} \equiv \frac{1}{N}\sum_a \chi^{aa} \tag{9}$$

where the asymptotic estimates of the local susceptibilities are given by

$$[\chi^{aa}(\mathbf{x})]^{\rm av} = \left[\left((\lambda|u_a + x_{0a}|)''\delta_{ab} + \frac{1}{\sigma^2_{\rm eff}}\right)^{-1}\right]^{\rm av}. \tag{10}$$

To determine $\sigma^2_{\rm eff}$, we look at the local susceptibilities in Proposition 2. In this case $U''(x)$ is zero everywhere except at $x = 0$, where it is formally infinite. Consequently,

$$\begin{aligned} &\chi^{aa} = 0, \text{ if } x_a = 0 \\ &\chi^{aa} = \sigma^2_{\rm eff}, \text{ otherwise.} \end{aligned} \tag{11}$$

The fact that χ^{aa} is the same for all non-zero values makes the analysis particularly simple. We define $\hat{\rho}$ to be the detection rate or estimated sparsity, i.e. the fraction of estimated x_a's that are non-zero. Therefore $\overline{\chi} = \hat{\rho}\sigma^2_{\rm eff}$ ($\lambda\overline{\chi} = \hat{\rho}\theta$) and $\sigma^2_{\rm eff} = \sigma^2 + \frac{\overline{\chi}}{\alpha}$ implying

$$\theta(1 - \frac{\hat{\rho}}{\alpha}) = \vartheta. \tag{12}$$

Thus when ϑ goes to zero, we either have $\theta = 0$ ($\hat{\rho} \neq \alpha$) or $\hat{\rho} = \alpha$ ($\theta \neq 0$). These two conditions correspond to the two phases of the system, the first being the perfect reconstruction phase and the second, the error phase. In terms of average local susceptibility, the first phase has $\overline{\chi} = \hat{\rho}\theta = 0$, while the second one has $\overline{\chi} \neq 0$.

Computation of the mean squared error (MSE), q (i.e. $\alpha\sigma^2_\xi$) using the soft-threshholding properties of the ℓ_1-norm, and

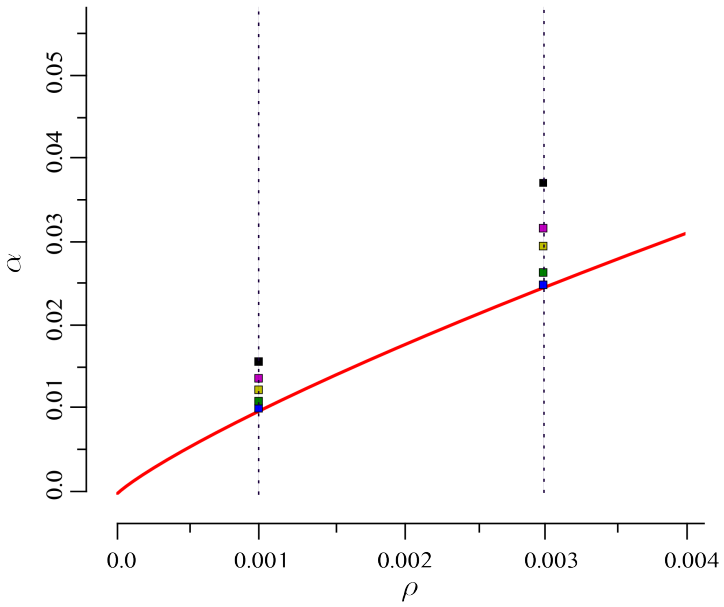

Fig. 1: The red curve is the theoretical phase boundary in the very sparse limit for the uncorrelated case obtained by solving Eq. (14) and (15). Numerical data for 'transition points' $c = 0, 0.2, 0.3, 0.4, 0.5$ corresponds to blue, green, yellow, magenta, and black markers for the Toeplitz matrix with entries $c^{|a-b|}$.

using Eq. (8), $\sigma_\xi^2 = q/\alpha$, leads to a parametric expression in the perfect reconstruction phase:

$$\alpha = 2(1-\rho)\big\{(1+\tau^2)\Phi(\tau) - \tau\phi(\tau)\big\} + \rho(1+\tau^2), \quad (13)$$

where $\Phi(\tau) = \int_\tau^\infty dz\,\phi(z)$, and $\phi(\tau) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\tau^2/2}$. To determine $\hat{\rho}$, one can notice that if $x_0 = 0$, we have to have $|\xi| > \theta$ to obtain a false positive. On the other hand, as θ goes to zero, a non-zero x_0 remains non-zero with the false negative probability approaching zero. Counting all sources of positive detections we have $\hat{\rho} = 2(1-\rho)\Phi(\tau) + \rho$.[1] Recall that in the error phase $\hat{\rho} = \alpha$. Equating these expressions at the phase transition line we obtain the parametric form for the boundary:

$$\alpha = 2(1-\rho)\big\{(1+\tau^2)\Phi(\tau) - \tau\phi(\tau)\big\} + \rho(1+\tau^2) \quad (14)$$

$$\alpha = 2(1-\rho)\Phi(\tau) + \rho. \quad (15)$$

Thus, Eq. (14) and (15) provide a parametric representation of α and ρ at the phase boundary, leading to the red curve depicted in Fig. 1.

IV. Extreme Sparse Limit

In this section, we reconsider the case of the extremely sparse limit, in which $\rho, \alpha << 1$. In this limit τ is large and the dominant contributions are the first term, $2(1-\rho)\Phi(\tau)$, from Eq. (15) which yields

$$\alpha \approx \sqrt{\frac{2}{\pi}}\frac{e^{-\frac{\tau^2}{2}}}{\tau}, \quad (16)$$

and the second term, $\rho(1+\tau^2)$, from Eq. (14) giving

$$\rho \approx \frac{\alpha}{\tau^2}. \quad (17)$$

Eq. (16) implies that $\tau^2 \approx 2\log(1/\alpha)$. Plugging this result into Eq. (17) we obtain $\rho \sim \alpha/(2\log(1/\alpha))$. This relation is identical to the bound that is found in [21] in the limit $\alpha \to 0$. Moreover, apart from a coefficient, it has a similar form to the RIP bounds [22].

From the above approximation, it is clear that in the very sparse limit near the transition we can make two observations,

1) The dominant contribution to the error comes from the shrinkage of the non-zero variables, namely from the term $\rho\tau^2$, with the false negative rate being negligible.
2) The false positives contribute to $\hat{\rho}$ and influence the phase boundary which is $\hat{\rho} = \alpha$.

In the case of the full-rank matrices $\mathbf{C}, \mathbf{D}$ in Eqs. (4) and (5), as we send $\sigma^2 \to 0^+$ limit, only the terms with coefficient $\mathrm{tr}(\mathbf{D}\bar{\boldsymbol{\chi}})/M$ stay relevant. Therefore, noise-free limit results in a significant simplification as the $\mathbf{C}$ terms cancel each other out:

$$\sigma^2_{\text{eff}} = \frac{1}{M}\mathrm{tr}(\mathbf{D}\bar{\boldsymbol{\chi}}), \quad (18)$$

$$\mathrm{Cov}(\xi,\xi) = \frac{q\mathbf{D}^{-1}}{\alpha}. \quad (19)$$

We assume that if we choose $\mathbf{D}$ close to $\mathbf{I}_N$, the nature of the solution does not change drastically. From the observation 1, since the major contribution to the error q comes from the shrinkage of the non-zero variables, we minimize the expression in Eq. (2), with $V(\mathbf{x}) = ||\mathbf{u} + \mathbf{x}_0||_1$, ignoring the term involving with $\boldsymbol{\xi}$. Considering the contributions of only true positives components, we get

$$\hat{\mathbf{u}}_A = -\lambda\sigma^2_{\text{eff}}(\mathbf{D}_A)^{-1}\mathbf{sgn}(\mathbf{x}_{0A}) \quad (20)$$

where A is the set of indices of the non-zero variables and the subscript A in $\hat{\mathbf{u}}_A, \mathbf{D}_A, \mathbf{x}_{0A}$ refer to $\hat{\mathbf{u}}, \mathbf{D}, \mathbf{x}_0$ restricted to the indices in A. Therefore

$$\begin{aligned} q &= \frac{1}{N}[\hat{\mathbf{u}}^\top\mathbf{D}\hat{\mathbf{u}}]^{\text{av}}_{\mathbf{x}_0,\boldsymbol{\xi}} = \frac{1}{N}[\hat{\mathbf{u}}_A^\top\mathbf{D}_A\hat{\mathbf{u}}_A]^{\text{av}}_{\mathbf{x}_0,\boldsymbol{\xi}} \\ &= \frac{\theta^2}{N}\Big[\sum_{a,a'\in A}\mathbf{sgn}(\mathbf{x}_0)_a\mathbf{sgn}(\mathbf{x}_0)_{a'}(\mathbf{D}_A)^{-1}_{aa'}\Big]^{\text{av}}_{\mathbf{x}_0} \\ &= \frac{\theta^2}{N}\big[\mathrm{tr}(\mathbf{D}_A^{-1})\big]^{\text{av}}_A \end{aligned} \quad (21)$$

We define $\psi_K(\mathbf{D}) = \big[\frac{1}{K}\mathrm{tr}(\mathbf{D}_A^{-1})\big]^{\text{av}}_A$ and rewrite

$$q = \rho\theta^2\psi_K(\mathbf{D}) \quad (22)$$

implying

$$\frac{1}{\tau^2} = \frac{q}{\alpha\theta^2} = \frac{\rho}{\alpha}\psi_K(\mathbf{D}). \quad (23)$$

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

which is the generalization of Eq. (17) for correlated design matrices. From observation 2, to determine the position of the phase boundary we should minimize Eq. (2). Thus we get $(\mathbf{Du})_a = (\mathbf{D}\boldsymbol{\xi})_a \mp \theta$ implying that if $x_0 = 0$, we should have $|(\mathbf{D}\boldsymbol{\xi})_a| > \theta$ to lead to a non-zero x. It is straightforward to show that the variance of $(\mathbf{D}\boldsymbol{\xi})_a$ equals $\frac{q}{\alpha}\mathbf{D}_{aa}$. Therefore, counting the average number of false positives gives

$$\alpha = \frac{2}{N}\sum_a \Phi(\tau/\sqrt{\mathbf{D}_{aa}}). \tag{24}$$

In the extremely sparse limit, τ is large and Eq. (24) becomes:

$$\alpha = \sqrt{\frac{2}{\pi}}\frac{1}{N}\sum_a \frac{e^{-\frac{\tau^2}{2\mathbf{D}_{aa}}}}{\tau/\sqrt{\mathbf{D}_{aa}}}. \tag{25}$$

which generalizes Eq. (16) to the correlated case. Therefore, by solving Eqs. (23) and (25) self-consistency, one can obtain the asymptotic relation for the correlated design matrix in the very sparse limit.

For the specific case which all the D_{aa}'s are 1, Eq. (25) suggests $\tau^2 \sim 2\ln(\frac{1}{\alpha})$. Combining this result and Eq. (23), we get

$$\alpha(\rho) \sim 2\psi_K(\mathbf{D})\rho\ln(\frac{1}{\alpha}). \tag{26}$$

We can show, that in this case, $\psi_K(\mathbf{D}) \geq 1$ (using the fact that $[\frac{1}{K}\mathrm{tr}(\mathbf{D}_A)]_A^{\mathrm{av}}[\frac{1}{K}\mathrm{tr}(\mathbf{D}_A^{-1})]_A^{\mathrm{av}} \geq 1$). Thus, in the presence of correlations, the number of measurements required for perfect reconstruction usually increases, as one would intuitively expect.

As an example we consider the case when the matrix $\mathbf{D}$ is the symmetric Toeplitz matrix with $D_{ab} = c^{|a-b|}$, $(c < 1)$, and $a, b = 1, 2, \ldots, N$. This result is relevant when correlations between sites decay as a function of the difference between their distance, e.g. spatial frequency or spatial location. For small ρ, $\psi_K(\mathbf{D}) = 1 + 2\rho c^2/(1-c^2) + \cdots$. Note that unless c is very close to one, so that $\rho/(1-c)$ is not insignificant, we do not get a big correction. This condition is related to the average spacing between indices in the active set A being comparable to the correlation length for the Toeplitz matrix. We could alternatively consider block-correlated matrices. The simplest example would be $D_{ab} = \delta_{ab} + \epsilon(1-\delta_{ab})$. In that problem, $\psi_K(\mathbf{D}) = 1/(1-\epsilon)$, showing once more an increase in α_c.

To investigate numerically the effect on the reconstruction limit α_c, we use the homotopy method [23] to solve the ℓ_1 norm optimization of (1). We obtain the $M \times N$ matrix $\mathbf{H}$ is filled with correlated entries $E[H_{ia}H_{jb}] = \frac{1}{M}C_{ij}D_{ab}$ where $\mathbf{D}$ is the symmetric Toeplitz matrix and $\mathbf{C}$ is the identity matrix. In the example shown, the size of the vector $\mathbf{x}$ is $N = 2 \times 10^4$, and are chosen to have two different values, $K = 20$, and $K = 60$ randomly placed elements driven from a standard Gaussian distribution. The failure is decided when MSE $> 10^{-4}$. For a fixed ρ, α_c increases with increasing correlation (c), as illustrated in the Fig. 1. Unfortunately, because the relatively small value of K, we have to deal, we do not expect to see quantitative agreement in this limit.

V. Conclusion

In this paper, we developed scheme to evaluate the typical reconstruction limit of LASSO with the correlated design matrix. When the dependent variables have strong correlations, as is often the case in real applications, LASSO can arbitrarily pick one out of a group of strongly correlated variables, rather than identifying the whole group. This is considered undesirable behavior and has led to the proposal of alternative algorithms. Here we employed the replica trick from statistical mechanics and provided analytical results of an effective optimization problem emerged by taking the average over the ensemble of the design matrix. In particular, we showed that when the N-dimensional signal $\mathbf{x}$ is K-sparse and the random design matrix $\mathbf{H}$ is $M \times N$ dimensional with the covariance $E[H_{ia}H_{jb}] = \frac{1}{M}C_{ij}D_{ab}$, the perfect recovery phase transition occurs at $M \sim \psi_K(\mathbf{D})K\log(N/M)$ in the very sparse limit $\frac{K}{N} \to 0$. Numerical experiments show qualitatively the same behavior.

Acknowledgment

This work was supported by the National Science Foundation INSPIRE (track 1) award 1344069. The final version of this paper was written while two of the authors (AMS and MR) were visiting Simons Center for Data Analysis at Simons Foundation. We are grateful for their hospitality.

Appendix

In order to make a connection between the optimizations problem and statistical mechanics, one could choose a probability distribution of $\mathbf{x}$ parametrized by β, i.e. $p_\beta(\mathbf{x}|\mathbf{y},\mathbf{H}) = \frac{1}{Z}\exp\big(-\beta\mathcal{E}(\mathbf{x})\big)$. The cost function $\mathcal{E}(\mathbf{x})$ is given by Eq. (1) and the normalization factor $Z = Z(\beta,\mathbf{y},\mathbf{H})$, known as the partition function. If we send β to ∞, the probability gets concentrated at the minimum of the cost function. Here we define β to be dimensionless. We will consider averages of the function $\mathcal{O}(\mathbf{x},\mathbf{x}_0)$ containing both the original sparse signal and the variable related to the estimate. The 'thermal' average of the function $\mathcal{O}(\mathbf{x},\mathbf{x}_0)$ over the distribution $p_\beta(\mathbf{x}|\mathbf{y},\mathbf{H})$ is represented by $< \mathcal{O}(\mathbf{x},\mathbf{x}_0) >$, depends on the random variables $\mathbf{x}_0$ and $\mathbf{H}$. For certain self averaged quantities we compute a further average over $\mathbf{x}_0$ and $\mathbf{H}$, denoted by $\big[\langle\mathcal{O}(\mathbf{x},\mathbf{x}_0)\rangle\big]^{\mathrm{av}}_{\mathbf{x}_0,\mathbf{H}}$. Computation of the quenched averages is complicated by the presence of the partition Z in the denominator of $\langle\mathcal{O}(\mathbf{x},\mathbf{x}_0)\rangle$. Formally, the denominator is handled by introducing n non-interacting replicas of the system and taking $n \to 0$, as shown below. $\mathcal{E}(\mathbf{x})$ depends on $\mathbf{x}$ as well as on $\mathbf{x}_0, H$. To emphasize those additional dependences, we write $\mathcal{E}(\mathbf{x})$ as $\mathcal{E}(\mathbf{x}_\mu,\mathbf{x}_0,\mathbf{H})$ in the next few equations.

$$\begin{aligned}\langle\mathcal{O}(\mathbf{x},\mathbf{x}_0)\rangle_{\mathbf{x}} &= \frac{\int d^N\mathbf{x}\mathcal{O}(\mathbf{x},\mathbf{x}_0)\exp\big(-\beta\mathcal{E}(\mathbf{x},\mathbf{x}_0,\mathbf{H})\big)}{\int d^N\mathbf{x}\exp\big(-\beta\mathcal{E}(\mathbf{x},\mathbf{x}_0,\mathbf{H})\big)} \\ &= \lim_{n\to 0}\int \mathcal{O}(\mathbf{x}_1,\mathbf{x}_0)\prod_{\mu=1}^{n}\big\{d^N\mathbf{x}_\mu\exp\big(-\beta\mathcal{E}(\mathbf{x}_\mu,\mathbf{x}_0,\mathbf{H})\big)\big\}.\end{aligned} \tag{27}$$

Thus, we need the replicated ensemble with partition function

$$\begin{aligned}&\big[Z^n\big]^{\rm av}_{\mathbf{x}_0,\mathbf{H}}\\ &=\Big[\int\prod_{\mu=1}^{n}d\mathbf{u}_\mu\exp\big[-\beta\big\{\sum_{\mu=1}^{n}\frac{(\mathbf{H}\mathbf{u}_\mu)^2}{2\sigma^2}+\lambda V(\mathbf{u}_\mu+\mathbf{x}_0)\big\}\big]\Big]^{\rm av}_{\mathbf{x}_0,\mathbf{H}}.\end{aligned}\tag{28}$$

rewritten in terms of the error variables $\mathbf{u}_\mu=\mathbf{x}_\mu-\mathbf{x}_0,\mu=1,..,n$. After averaging over $\mathcal{P}(\mathbf{H})$, introducing auxilliary variables $\mathbf{Q}$ and $\mathbf{R}$, and evaluating the saddle point $\bar{\mathbf{Q}},\bar{\mathbf{R}}$ in the limit $M,N\to\infty$, holding $\alpha=\frac{M}{N}$ fixed, we arrive at

$$\big[Z^n\big]^{\rm av}_{\mathbf{x}_0,\mathbf{H}}\propto\exp(-\beta F_n(\bar{\mathbf{R}})+\sum_{\mu,\nu}\bar{R}_{\mu\nu}\bar{Q}_{\mu\nu})\tag{29}$$

$$\begin{aligned}\exp(-\beta F_n(\bar{\mathbf{R}}))=\Big[\int\prod_{\mu=1}^{n}d^N\mathbf{u}_\mu\exp\big[-\sum_{\mu,\nu}\bar{R}_{\mu\nu}\mathbf{u}_\mu^\top\mathbf{D}\mathbf{u}_\nu\\ -\beta\sum_\mu\lambda V(\mathbf{u}_\mu+\mathbf{x}_0)\big]\Big]^{\rm av}_{\mathbf{x}_0}.\end{aligned}\tag{30}$$

$$\bar{Q}_{\mu\nu}=\frac{1}{N}\langle\langle\mathbf{u}_\mu^\top\mathbf{D}\mathbf{u}_\nu\rangle\rangle,\tag{31}$$

$$\bar{\mathbf{R}}=\frac{\beta}{2\sigma^2}\mathrm{tr}_M\big[\mathbf{C}\otimes\mathbf{I}_n(\mathbf{I}_M\otimes\mathbf{I}_n+\frac{\beta}{\alpha\sigma^2}\mathbf{C}\otimes\mathbf{Q})^{-1}\big].\tag{32}$$

The trace tr_M is a trace only applying to the M dimensional space. The expectation $\langle\langle\mathbf{u}_\mu^\top\mathbf{D}\mathbf{u}_\nu\rangle\rangle$ depends on $\bar{\mathbf{R}}$ via

$$\langle\langle\mathbf{u}_\mu^\top\mathbf{D}\mathbf{u}_\nu\rangle\rangle=\beta\frac{\partial F_n(\bar{\mathbf{R}})}{\partial\bar{R}_{\mu\nu}}.\tag{33}$$

If $V(x)$ is a convex function, we expect a unique minimum and a replica symmetric solution [8], [24] for $\mathbf{Q},\mathbf{R}$. This implies $\bar{Q}_{\mu\nu}=(Q-q)\delta_{\mu\nu}+q$ and $\bar{R}_{\mu\nu}=(R-r)\delta_{\mu\nu}+r$. Using this ansatz in Eqs. (31) and (32) and eliminating $\mathbf{R}$ by introducing another quenched variable $\boldsymbol{\xi}$, one can rewrite the right hand side of Eq. (30) as

$$\begin{aligned}\Big[\int\prod_{\mu=1}^{n}\{d^N\mathbf{u}_\mu\}\exp\Big[-\beta\Big\{\frac{1}{2\sigma_{\rm eff}^2}\sum_\mu(\mathbf{u}_\mu^\top\mathbf{D}\mathbf{u}_\mu-2\boldsymbol{\xi}^\top\mathbf{D}\mathbf{u}_\mu)\\ +\lambda\sum_\mu V(\mathbf{u}_\mu+\mathbf{x}_0)\Big\}\Big]\Big]^{\rm av}_{\boldsymbol{\xi},\mathbf{x}_0}.\end{aligned}\tag{34}$$

In order to study the regularized least-squares reconstruction, one should take the limits $\beta\to\infty$, and then $\sigma\to0$. However, $\sigma^2_{\rm eff}$ and the second moment of the Gaussian quenched vector $\boldsymbol{\xi}$ depend on $\beta\Delta Q$. This makes the computation of these quantities unnecessarily nontrivial. As it is shown in [19], one can resolve this issue by identifying $\beta\Delta Q$ as $[\boldsymbol{\chi}]^{\rm av}_{\mathbf{H}}$ where $\boldsymbol{\chi}$ is the local susceptibility matrix. Therefore, the optimization of $\mathcal{E}(\mathbf{x})$ corresponds to the optimization in Proposition 1.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[4] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[5] 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, 2007.

[6] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1902–1923, 2012.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[8] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on ℓ_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 09, p. L09003, 2009.

[9] S. Ganguli and H. Sompolinsky, "Statistical mechanics of compressed sensing," *Physical review letters*, vol. 104, no. 18, p. 188701, 2010.

[10] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[11] K. Takeda and Y. Kabashima, "Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals," *EPL (Europhysics Letters)*, vol. 95, no. 1, p. 18006, 2011.

[12] M. Chiara Angelini, F. Ricci-Tersenghi, and Y. Kabashima, "Compressed sensing with sparse, structured matrices," 2012.

[13] P. Bühlmann, P. Rütimann, S. van de Geer, and C.-H. Zhang, "Correlated variables in regression: clustering and sparse estimation," *Journal of Statistical Planning and Inference*, vol. 143, no. 11, pp. 1835–1858, 2013.

[14] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4053–4085, 2011.

[15] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[16] A. Sengupta and P. P. Mitra, "Distributions of singular values for some random matrices," *Physical Review E*, vol. 60, no. 3, p. 3389, 1999.

[17] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: Random matrix techniques and large-n expansions (2)," *Journal of Statistical Physics*, vol. 125, no. 5-6, pp. 1223–1242, 2006.

[18] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "Mimo capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large n analysis," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2545–2561, 2003.

[19] M. Ramezanali, P. P. Mitra, and A. M. Sengupta, *arXiv preprint arXiv:1501.03194*.

[20] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9446–9451, 2005.

[21] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[22] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[23] D. L. Donoho and Y. Tsaig, "Fast solution of-norm minimization problems when the solution may be sparse," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4789–4812, 2008.

[24] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond*. World scientific Singapore, 1987, vol. 9.