Sparse reconstruction algorithms aim to retrieve high-dimensional sparse signals from a limited amount of measurements under suitable conditions. These algorithms exhibit sharp algorithmic phase transition boundaries where the retrieval breaks down as the number of variables go to infinity. Sparse reconstruction algorithms are usually stated as optimization problems. These have been analyzed in the literature by defining associated statistical mechanical problems at a finite temperature, which are treated in the mean field approximation using the replica trick, and subsequently analyzed in the literature by defining associated statistical mechanical problems at a finite temperature. However, it is now well known that a computationally tractable convex relaxation of this optimization problem, termed Basis Pursuit (28), (4), (8), nevertheless yields the correct result for sufficiently small $K$ (sufficient sparsity). The key idea is to replace the $l_0$ norm by the $l_1$ norm. Namely, the estimated vector $\hat{x}$, defined by

$$\hat{x} = \lim_{\sigma \to 0+} x(\sigma^2),$$

is the solution of the resulting reduced matrix equation. Note that for the wrong choice of nonzero variables, the equations would typically be inconsistent, while, for the right choice, there would, typically, be a unique solution.

However, this procedure is not computationally efficient since $(\binom{N}{K}) \sim (\frac{N}{K})^K \binom{N}{K}^{N-K}$ combinations have to be tried out. An alternative formulation for the problem is to pose it as an optimization problem, defining $x(\lambda \sigma^2) = \arg \min_x \frac{1}{2\sigma^2} (y - Hx)^2 + \lambda ||x||_1$, where $||x||_p = \sqrt[p]{\sum |x_i|^p}$ is the $l_p$ norm and $||x||_0 = \lim_{p \to 0} ||x||_p$. Strictly speaking, $||x||_0$ is a norm only for $p \geq 1$, but in the context of sparse reconstruction, the term $l_0$ norm has been in use. The so-called $l_0$ norm, $||x||_0$, really counts the number of nonzero elements of $x$.

The point of adding $\lambda ||x||_0$, with $\lambda > 0$, to the least square term is to penalize the number of nonzero entries of $x$. Note that the solution of the optimization problem only depends on $\lambda \sigma^2$, which we will call $\vartheta$. The sparse retrieval solution is given by $\hat{x} = \lim_{\sigma \to 0+} x(\vartheta)$. One can argue that this retrieval would be exact, i.e. $\hat{x} = x_0$, for almost all random choices of $H$ and of $x_0$.

Unfortunately, this is a non-convex optimization problem and it also has exponential computational complexity. However, it is now well known that a computationally tractable convex relaxation of this optimization problem, termed Basis Pursuit (28), (4), (8), nevertheless yields the correct result for sufficiently small $K$ (sufficient sparsity). The key idea is to replace the $l_0$ penalty by the $l_1$ norm. Namely, the estimated vector $\hat{x}$, defined by

$$\hat{x}(\lambda \sigma^2) = \arg \min_x \frac{1}{2\sigma^2} (y - Hx)^2 + \lambda ||x||_1 \quad (1)$$

$$\hat{x} = \lim_{\vartheta \to 0+} \hat{x}(\vartheta), \quad (2)$$

The cavity method for phase transitions in sparse reconstruction algorithms

Mohammad Ramezanali, Partha P. Mitra, and Anirvan M. Sengupta

1 Dept. of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Rd, Piscataway, NJ 08854 USA

2 Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724 USA

3 BioMaPS Institute for Quantitative Biology, Rutgers University, 176 Frelinghuysen Rd, Piscataway, NJ 08854 USA

I. INTRODUCTION

In traditional statistics, the number of observations are typically much larger than the number of unknown parameters being estimated. In the last couple of decades, however, the opposite limit has been studied in a variety of fields, ranging from communication technology to business informatics to systems biology: the associated data vectors are large, but the underlying feature space can be of even larger dimensionality. A popular method for regularizing such ill-posed problems has been the imposition of a sparsity constraint on the feature space in a particular basis [3, 6]. These methods have found a broad variety of applications.

The usual setting of the sparse retrieval problem is an ill-posed linear equation (noise free case) is $y = Hx$, where $y$ is a $M$ dimensional measurement vector, $H$ is an $M \times N$ known measurement matrix, and $x$ is an $N$ dimensional unknown parameter vector ($M < N$). Let $y$ be generated by $Hx_0$, where $x_0$ is an unknown $N$ dimensional vector with $N - K$ components being zero. Our task is to reconstruct this unknown vector. The ill-posedness of the linear problem is removed by imposing back sparsity constraint. For typical $H$, as long as the number of unknowns, $K$, is less than the number of measurements $M$, this linear equation has a unique sparse solution with high probability. This can be seen by setting $N - K$ entries of $x$ to zero, and trying to solve
is an exact reconstruction of \( x_0 \), in a certain region of parameter space.

Remarkably, the boundary between the parameter region where the correct unique solution is found by this algorithm, and the parameter region where the algorithm fails, becomes sharp as \( N \to \infty \), with \( \frac{M}{N} \) and \( \frac{K}{N} \) being held fixed. This is an algorithmic phase transition or zero-one law, with the probability of correct reconstruction (suitably defined) jumping from zero to one at the transition boundary. Analytical formulae for the phase transition boundary exist in the literature, but the corresponding derivations are complex. In this paper we explore these transition boundaries using alternative analytical approaches for computing these boundaries, and obtain additional insights into the problem as well as insights about a more general set of optimization problems.

For measurement matrices that have independent and identically distributed (iid) Gaussian entries, Basis Pursuit requires at least \( M > O(K \log(N/K)) \) measurements for perfect reconstruction [5]. Rigorous results, based on the so called Restricted Isometry Property [2] of the matrix \( H \), provide parameter regions where the algorithm is guaranteed to work. However, these results do not give the exact threshold where the procedure breaks down. On the other hand, analyses based on the message-passing method, and on the replica formalism borrowed from statistical physics, indicate that the performance failure of the \( l_1 \) norm minimization method and other analogous algorithms with polynomial time complexity occurs at a sharp boundary analogous to a second-order phase transition [7, 9, 10]. The replica method enables computation of the phase transition boundary, but it is difficult to gain intuitive insight into the problem from this approach. These are both “mean field” approaches, which typically produce approximate answers, but for the current problems they produce exact results that have been verified by numerical simulations.

Here, we take an alternate approach in understanding the nature of the associated phase transitions using the cavity method [14, 15]. The cavity method is a powerful approach which was originally designed for understanding the nature of ground states of certain spin glass models. The method has since been applied to a wider class of problems including algorithmic phase transitions, some examples being the satisfiability problem [16, 19] and Hopfield neural networks [23]. The cavity method leads to the same results as obtained by replica trick [15] and is closely related to the message-passing algorithm in graphical models [1]. However, we find that for the problem at hand, the cavity method provides more insight and intuition in comparison to the replica formalism.

The paper is organized as follows. We first briefly review a finite noise/finite temperature formulation of the problem, and a recapitulation of replica approach to its solution, presented previously in the literature. We do this for a generalized regularization or penalty function \( V \). Then, we introduce a susceptibility matrix associated with this problem and explore its structure. In the next sections, we derive the self-consistent mean field equations for the mean square error of estimation via a two-step cavity method at zero temperature. This approach is not only different from the replica method but also from the method based on iterations in a message-passing algorithm [13]. We find a simple way to arrive at the two phases and the phase boundary. We treat the simple case of Ridge Regression [29], recover the known results for Basis Pursuit [4] and obtain new ones for the Elastic Net [30]. We end by pointing out potential arenas of application where our approach is more natural than conventional replica theory.

II. FORMULATION OF THE SPARSE RECONSTRUCTION PROBLEM

Here, we set up the general framework for investigating the performance of sparse reconstruction algorithms that could be stated as a penalized regression problem. We assume that the data \( y = Hx_0 + \zeta \) are generated by a probability distribution \( p(y|x_0, H) \), given an (unknown) sparse signal \( x_0 \) and a (known) matrix \( H \), and an (unknown) Gaussian noise vector \( \zeta \) whose components are i.i.d. samples from \( \mathcal{N}(0, \sigma^2) \). The vector \( x_0 \) is considered to be a random sample from a distribution \( P_0(x_0) = \prod a p_0(x_{a0}) \). The distribution \( p_0(x_{a0}) \) a distribution with a continuous part and a delta function at origin:

\[
p_0(x_a) = \rho \pi(x_a) + (1 - \rho) \delta(x_a).
\]

Here \( \rho \) is the sparsity parameter \( K/N \) and \( \pi(x_0) \) is the distribution of non-zero entries of the parameter vector.

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

\[
[H]_{ab} = 0
\]

\[
[H]_{ia} = \frac{1}{M} \delta_{ij} \delta_{ab}
\]

We study the performance of an estimator of \( x_0 \), namely the location \( \hat{x} \) of the minimum of a cost function

\[
E_0(x) = \frac{(y - Hx)^2}{2\sigma^2} + V(x).
\]

As we reformulate this exercise in estimation as a statistical mechanics problem, the cost function will play the role of energy. We assume the penalty/potential term \( V(x) \) is such that there is a unique minimum of \( E_0 \). Note that \( \hat{x} = \arg\min_x E_0(x) \) depends on \( y, H \), meaning that it can be written as a function \( \hat{x} = g(x_0, H, \zeta) \), using the fact that \( y = Hx + \zeta \). Since we set up an ensemble of problem instances by specifying the probability distribution of the variables \( x_0, H, \zeta \), we could study the performance of the estimator over this distribution. For example, we could
study the distribution of the estimation error \( \hat{x} - x_0 \) for this problem instance ensemble.

In order to make a connection between the optimization problem and statistical mechanics, one could choose a probability distribution of \( x \) parametrized by \( \beta \), playing the role of inverse temperature,

\[
p_\beta(x|y, H) = \frac{1}{Z} \exp\left( -\beta \mathcal{E}_0(x) \right) = \frac{1}{Z(\beta, y, H)} \exp\left\{ -\beta \left( \frac{(y - Hx)^2}{2\sigma^2} + V(x) \right) \right\}
\]

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

\[
\begin{align*}
Z(\beta, y, H) &= \int d^N x \frac{1}{Z} \exp(-\beta \mathcal{E}_0(x)) \\
&= \int d^N x \exp\left\{ -\beta \left( \frac{(y - Hx)^2}{2\sigma^2} + V(x) \right) \right\}.
\end{align*}
\]

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

We will consider averages of functions of the form \( f(x, x_0) \) containing both the original sparse signal and the variable related to the estimate. For example, we are interested in the estimation error, which can be quantified by a suitable norm placed on the difference between the original and the estimated signal. The average of the function \( f(x, x_0) \) is given by

\[
\langle f(x, x_0) \rangle = \frac{\int d^N x f(x, x_0) \exp\left\{ -\beta \left( \frac{(y - Hx)^2}{2\sigma^2} - \beta V(x) \right) \right\}}{\int d^N x \exp\left\{ -\beta \left( \frac{(y - Hx)^2}{2\sigma^2} - \beta V(x) \right) \right\}}.
\]

This ‘thermal’ average, represented by \( < \cdots > \), depends on the random variables \( x_0, H \) and \( \zeta \). Note that in the limit \( \beta \to \infty \), this average should become \( f(x, x_0) \), for continuous \( f \). Averaging the result of this calculation over the random instances of \( x_0, H \) and \( \zeta \) is a technical challenge related to quenched averages in disordered systems [15].

The function \( f(x, x_0) = \frac{1}{N}(x - x_0)^2 \) plays an important role in our analysis. Its average corresponds to the mean squared estimation error.

\[
\text{MSE} \equiv \frac{1}{N} \sum_{a=1}^{N} \left( (x_a - x_{0a})^2 \right)^{\text{av}}_{x_0, H, \zeta} = \frac{1}{N} \left[ \langle (x - x_0)^2 \rangle \right]^{\text{av}}_{x, x_0, H, \zeta}
\]

We will use \( \ldots \text{av} \) to denote quenched averages, with the relevant quenched variables indicated in the subscript, when necessary. We use the notation \( u = x - x_0 \) to indicate the estimation error vector. The size of the vector \( u \) provides a measure of the inaccuracy of the reconstruction.

In the context of penalized regression, the penalty function is often chosen to be a sum of potentials involving single variables, namely, \( V(x) = \sum_a U(x_a) \). We will focus on \( V(x) \) of this nature. An important special case for sparse reconstruction is regularizing potential \( U(x) = \lambda_1|x| + \lambda_2 \frac{x^2}{2} \) which will be treated in great detail. In this paper, we would also restrict ourselves to the noiseless case, \( \zeta = 0 \), although the same methods could be used to analyze the noisy case as well.

For \( \zeta = 0 \), we will be interested in the result of the constrained optimization problem of minimizing \( V(x) \) subject to the constraint \( y = Hx \). In the \( K, M, N \to \infty \) limit, this problem may exhibit a phase transition from a perfect reconstruction phase to an error-prone phase, with this problem may exhibit a phase transition from a per-

III. REPLICA APPROACH

In this section, we review the replica approach to the problem [9, 10], presenting the mean field equations in terms of a distribution of asymptotically independent single-variable problems with a set of self-consistency conditions. In order to calculate quantities like the MSE, we need to compute quenched averages of the form \( \left\langle (f(x, x_0)) \right\rangle^{\text{av}}_n \), which is complicated by the presence of the denominator in Eq. (9). Formally, the denominator is handled by introducing \( n \)-non-interacting replicas of the system and taking \( n \to 0 \), as shown below. In the noiseless case, \( \mathcal{E}_0(x) \) depends on \( x \) as well as on \( x_0, H \). To emphasize those additional dependences, we write \( \mathcal{E}_0(x) \) as \( \mathcal{E}_0(x, x_0, H) \) in the next few equations.

\[
\begin{align*}
\langle f(x, x_0) \rangle_x &= \frac{\int d^N x f(x, x_0) \exp\left( -\beta \mathcal{E}_0(x, x_0, H) \right)}{\int d^N x \exp\left( -\beta \mathcal{E}_0(x, x_0, H) \right)} \\
&= \lim_{n \to 0} \left( \int d^N x \exp\left( -\beta \mathcal{E}_0(x, x_0, H) \right) \right)^{n-1} \\
&= \lim_{n \to 0} \int d^N x f(x, x_0) \exp\left( -\beta \mathcal{E}_0(x, x_0, H) \right) \\
&= \lim_{n \to 0} \int f(x_1, x_0) \prod_{\mu=1}^{n} \{ d^N x_\mu \exp\left( -\beta \mathcal{E}_0(x_\mu, x_0, H) \right) \}.
\end{align*}
\]
Averaging over the quenched variables $x_0$ and $H$, we get
\[
\langle f(x, x_0) \rangle_{x_0, H} = \lim_{n \to 0} \left[ \prod_{\mu=1}^{n} \{ d^N x_\mu \} f(x_1, x_0) \exp \left\{ - \beta \sum_{\mu} E_0(x_\mu, x_0, H) \right\} \right]_{x_0, H}^{av}
\]
(12)

Using $y = Hx_0$ in the noiseless case, the energy function for the $n$-th replica would be
\[
E_0(x_\mu, x_0, H) = \frac{(y - Hx_\mu)^2}{2\sigma^2} + V(x_\mu)
\]
(13)
rewritten in terms of the error variables $u_\mu = x_\mu - x_0, \mu = 1, \ldots, n$. Thus, we are interested in average quantities in the replicated ensemble whose partition function is given by
\[
\left[ Z^{av} \right]_{x_0, H}^{av} = \left[ \prod_{\mu=1}^{n} \{ du_\mu \} \exp \left\{ - \beta \left( \sum_{\mu=1}^{n} \frac{(H u_\mu)^2}{2\sigma^2} + V(u_\mu + x_0) \right) \right\} \right]_{x_0, H}
\]
(14)

In order to average over $\mathcal{P}(H)$, the only quantity that needs to be computed is
\[
\left[ \exp \left\{ - \sum_{\mu=1}^{n} \frac{\beta}{2\sigma^2} (H u_\mu)^2 \right\} \right]_H^{av} = \frac{1}{Z_0} \int dH \exp \left\{ - \frac{M}{2} \text{Tr}(H^2) - \frac{\beta}{2\sigma^2} \sum_{\mu=1}^{n} u_\mu^2 H^T H u_\mu \right\}
\]
(15)

where $Z_0$ is the normalization term for the Gaussian distribution of $H$, and $\alpha = M/N$ is the sampling ratio. The elements of the $n \times n$ matrix $Q$ are defined by $Q_{\mu\nu} = \frac{1}{N} u^\dagger_\mu u_\nu$. Using the Fourier representation of the $\delta$ function
\[
\delta(u^\dagger_\mu u_\nu - N Q_{\mu\nu}) = \frac{1}{2\pi^2} \int dR_{\mu\nu} \exp(-i R_{\mu\nu}(u^\dagger_\mu u_\nu - N Q_{\mu\nu}))
\]
and inserting this delta function with an integral over $Q_{\mu\nu}$ in Eq. (14), we get
\[
\left[ Z^{av} \right]_{x_0, H}^{av} = \int \{ dQ_{\mu\nu} dR_{\mu\nu} \exp \left\{ - S(Q, R) \right\} \}
\]
(17)

\[
S[Q, R] = \frac{M}{2} \text{Tr} \log (I_n + \frac{\beta}{\alpha \sigma^2} Q) - i N \text{Tr} \text{Tr}(RQ)
\]
\[
- \log \left[ \prod_{\mu=1}^{n} \{ du_\mu \} \exp \left\{ - i \sum_{\mu, \nu} R_{\mu\nu} u^\dagger_\mu u_\nu + \sum_{\mu} V(u_\mu + x_0) \right\} \right]_{x_0}^{av}
\]
(18)

This integral over $Q, R$ can be evaluated using the saddle point method \cite{9, 10} when $M, N \to \infty$, holding $\alpha = \frac{M}{N}$ fixed. The saddle point $Q = \bar{Q}, R = -i \bar{R}$ satisfies the conditions:
\[
Q_{\mu\nu} = \frac{1}{N} \langle \{ u^\dagger_\mu u_\nu \} \rangle
\]
(19)
\[
R = \frac{\beta}{2\sigma^2} \left[ I_n + \frac{\beta}{\alpha \sigma^2} Q \right]^{-1}
\]
(20)
obtained by differentiating $S(Q, R)$ with respect to the elements of $Q, R$. The expectation $\langle \{ u^\dagger_\mu u_\nu \} \rangle$ depends on $\bar{R}$ via
\[
\langle \{ u^\dagger_\mu u_\nu \} \rangle = \frac{\partial F(\bar{R})}{\partial R_{\mu\nu}}
\]
(21)

\[
\langle \{ u^\dagger_\mu u_\nu \} \rangle = \left[ \int \{ d^N u_\mu \} \exp \left\{ - \sum_{\mu, \nu} \bar{R}_{\mu\nu} u^\dagger_\mu u_\nu - \beta \sum_{\mu} V(u_\mu + x_0) \right\} \right]_{x_0}^{av}
\]
(22)

If $U(x)$ is a convex function, we expect a unique state and a replica symmetric solution for $Q, R$. This implies $Q_{\mu\nu} = (Q - q) \delta_{\mu\nu}$ and $R_{\mu\nu} = (R - r) \delta_{\mu\nu}$. With that ansatz,
\[
\int \{ d^N u_\mu \} \exp \left\{ - \sum_{\mu, \nu} \bar{R}_{\mu\nu} u^\dagger_\mu u_\nu - \beta \sum_{\mu} V(u_\mu + x_0) \right\}
\]
\[
= \int \{ d^N u_\mu \} \exp \left\{ - (R - r) \sum_{\mu} u^2_\mu - r \sum_{\mu} V(u_\mu + x_0) \right\}
\]
\[
= \int \{ d^N \xi \} \exp(-\frac{\xi^2}{2\sigma^2}) \int \{ d^N u_\mu \} \exp \left\{ - \beta \frac{\xi^2}{2\sigma^2} \sum_{\mu} u^2_\mu + \frac{\beta}{\sigma^2} \xi^\dagger (\sum_{\mu} u_\mu) - \beta \sum_{\mu} V(u_\mu + x_0) \right\}
\]
(23)
identifying $R - r \equiv \frac{\beta}{\sigma^2}$ and $r \equiv \frac{\beta^2 \sigma^2}{2\sigma^2}$. We have used
\[
\int \{ d^N \xi \} \exp(-\frac{\xi^2}{2\sigma^2}) \exp\left( \frac{\beta}{\sigma^2} \xi^\dagger (\sum_{\mu} u_\mu) \right)
\]
\[
= \exp\left( \frac{\beta^2 \sigma^2}{2\sigma^2} (\sum_{\mu} u_\mu)^2 \right)
\]
(24)
to decouple the item replica coupling in the $(\sum_{\mu} u_\mu)^2$ term, at the cost of introducing another quenched variable $\xi$. Note that we require $R - r > 0$ and $r < 0$ for this approach to work. These inequalities follow from (20) and from $Q - q > 0$ and $q > 0$. The conditions on $Q$ and $q$ would be obvious once we look at interpretation of these quantities described below.

For $V(x) = \sum_a U(x_a)$ we can simplify further. Remembering that we also need to do the quenched average over
Thus, in the saddle point approximation, each of the $N$ components of $u$ become effectively independent and the saddle point conditions reduce to a self-consistent problem for each component $a = 1, \ldots, N$. Since this self-consistent problem is similar for each index, we suppress the subscript $a$ in $u_{a\mu}$ and in $x_{a\mu}$. For each $a$, we have the integral of the form

$$\left[ \int \prod_{\mu=1}^{n} du_{\mu} \exp \left[ -\beta \left( \frac{1}{2\sigma_{\text{eff}}^{2}} \sum_{\mu} (u_{\mu}^{2} - \xi u_{\mu}) + \sum_{\mu} U(u_{\mu} + x_{0\mu}) \right) \right] \right]_{\xi, x_{0\mu}}^{\text{av}}$$

This replica problem corresponds to a single variable $u$ follows the effective distribution

$$P_{\text{eff}}(u|x_{0}, \xi) = \frac{1}{Z(x_{0}, \xi)} e^{-\beta E_{\text{eff}}(u;x_{0}, \xi)},$$

with an effective mean-field Hamiltonian

$$E_{\text{eff}}(u;x_{0}, \xi) = \frac{1}{2\sigma_{\text{eff}}^{2}} (u^{2} - 2\xi u) + U(u + x_{0})$$

which depends on two quenched variables $x_{0}$ and $\xi$. The variable $x_{0}$ has the probability distribution $p_{0}(x_{0}) = \rho \pi(x_{0}) + (1 - \rho) \delta(x_{0})$, whereas $\xi$ is distributed according to a Gaussian distribution with mean zero and variance $\sigma_{\xi}^{2}$. The two parameters $\sigma_{\text{eff}}^{2}$ and $\sigma_{\xi}^{2}$ are given by the following set of self-consistency conditions.

$$q = \left[ \langle u^{2} \rangle_{x_{0}, \xi} \right]^{\text{av}}$$

$$\Delta Q = Q - q = \left[ \langle (u - \langle u \rangle)^{2} \rangle \right]_{x_{0}, \xi}^{\text{av}}$$

$$\sigma_{\text{eff}}^{2} = \sigma^{2} + \frac{\beta \Delta Q}{\alpha}$$

$$\sigma_{\xi}^{2} = \frac{q}{\alpha}$$

where the thermal averages $\langle \cdots \rangle$ over $u$ are performed in the $P_{\text{eff}}$ ensemble and the so-called quenched average $\langle \cdots \rangle_{x_{0}, \xi}$ is over variables $\xi$ and $x_{0}$.

In order to study the phase transition, we need to take the limits $\beta \to \infty$, and then $\sigma \to 0$. A nontrivial aspect of the zero temperature limit ($\beta \to \infty$) is that the quantity $\beta \Delta Q$ in Eq. (30) behaves differently in different phases. Using Eq. (29), this quantity is just $\beta$ times the thermal fluctuation in $u$. The fluctuation-dissipation relation [11] implies that this quantity may be interpreted as a local susceptibility. In our following considerations based on the zero temperature cavity method, we formally introduce a susceptibility and use its properties to give a more transparent derivation of the same equations.

### IV. SUSCEPTIBILITY AND CONJUGATE VARIABLES

The optimization problem associated with the regularized sparse reconstruction problem involves minimizing the energy function $E_{0}(x) = \langle \frac{1}{2\sigma^{2}} u^{T}H^{T}Hu + V(x) \rangle$. For the noise free case, using $y = Hx_{0}$, the energy to be minimized may be rewritten as

$$E(u) = \frac{1}{2\sigma^{2}} u^{T}H^{T}Hu + V(u + x_{0}).$$

where $u = x - x_{0}$. Note that, unlike the function $E_{0}(x)$, which is parametrized by known quantities (the data $y$ and the measurement matrix $H$) and can therefore be empirically optimized with respect to its argument, the closely related function $E(u) = E_{0}(u + x_{0})$ depends on the knowledge of the original signal $x_{0}$. The purpose of dealing with this function is not to provide an algorithm to estimate this signal given measured data, but to study the statistical behavior of this function and its minima over the distribution of problem instances, namely, input signals and the measurement matrices. For example, we can calculate the distribution of each component of the estimation error vector $u$, given the distributions of $x_{0}$ and $H$. We will be working with $E(u)$, although the susceptibility for a particular problem instance, to be defined below, could be defined completely in terms of $E_{0}(x)$.

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

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

$$E(u; f) = \frac{1}{2\sigma^{2}} u^{T}H^{T}Hu + V(u + x_{0}) - f \cdot u.$$
with the variables \( f \), which are conjugate to \( u \). Optimizing \( \mathcal{E}(u, f) \) will produce an \( f \) dependent answer \( u = \hat{u}(f) \).

For small \( f \) we expect

\[
\hat{u}(f) = \hat{u} + \chi f + \ldots
\]

defining the susceptibility matrix \( \chi \).

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

\[
f = \nabla_u \mathcal{E}(u)
\]

Where \( f = 0 \), \( u \) is at its optimal value \( \hat{u} \). If perturbation \( f \) is small and \( \mathcal{E} \) is differentiable to higher orders, we can expect \( \delta u \) to be small and, therefore, Taylor expand \( \mathcal{E}(u + \delta u) \) around \( u = \hat{u} \)

\[
\mathcal{E}(\hat{u} + \delta u) = \mathcal{E}(\hat{u}) + \frac{1}{2} \sum_{ab} \delta u_a \delta u_b \left. \frac{\partial^2 \mathcal{E}}{\partial u_a \partial u_b} \right|_{u = \hat{u}} + \ldots
\]

From (35) and (36), we can identify the inverse susceptibility \( (\chi^{-1})_{ab} = \frac{\partial^2 \mathcal{E}}{\partial u_a \partial u_b} \left|_{u = \hat{u}} \right. \)

\[
\text{and can show that}
\]

\[
\min_u \mathcal{E}(u, f) = \mathcal{E}(\hat{u}) - \hat{u}^T f - \frac{1}{2} f^T \chi f + \ldots
\]

This expansion will appear several times in our derivations.

We will now discuss the structure of the susceptibility matrix \( \chi \). It is simplest to study the properties of \( \chi \), when the potential \( U(x) \) has continuous second derivatives. Several observations from this case would carry over to singular examples like \( U(x) = \lambda |x| \), which we have to ultimately deal with. If we could Taylor expand around the solution \( u = \hat{u} \), we will have

\[
\mathcal{E}(\hat{u} + \delta u; f) = \mathcal{E}(\hat{u}; 0) + \frac{1}{2} \delta u^T \left[ \frac{H^T H}{\sigma^2} + W(x) \right] \delta u - f^T \chi f + \ldots
\]

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

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

When \( H \) is a large random matrix, we can make asymptotic estimates of the mean and the variance of different components of the susceptibility matrix \( \chi \). One way to approach this problem is to formally expand the R.H.S of Eq. (39) in powers of \( \frac{H^T H}{\sigma^2} \) (see Fig. 1) and compute moments by averaging over \( H_{ia} \) diagrammatically. Namely, we expand

\[
\chi = W^{-1} - \frac{1}{\sigma^2} W^{-1} H^T H W^{-1} + \frac{1}{\sigma^4} W^{-1} H^T H W^{-1} - M H^T H W^{-1} - \ldots
\]

and then compute moments of the form

\[
[\chi^{ab \cdots}_{a_1 b_1 \cdots c} H_{i_1 i_2 \cdots j}] \chi^{a b \cdots}_{a_1 b_1 \cdots c} H_{i_1 i_2 \cdots j}
\]

using Wick’s theorem, since \( H \) distribution is Gaussian with mean and covariance specified by Eq. (4) and Eq. (5), respectively.

Following earlier work on singular values of random matrices [21, 22], we use the fact that, in the large \( M, N \) limit, only the planar diagrams survive. \([\chi(x, H)]_{\text{planar}}\) could be written as \([W(x) - \Sigma(x) I_N]^{-1}\), where \( \Sigma(x) \) is a self-energy term. The planar contributions to the self-energy are shown in Fig. 2 and can be re-summed as

\[
\Sigma(x) = -\frac{1}{\sigma^2} \frac{1}{1 + \frac{M}{\sigma^2} \sum_a \chi^{aa}(x)}
\]

Hence, the mean susceptibility (holding \( x_a \)’s fixed but averaging over \( H \)) is given by

\[
\chi^{aa}(x) \equiv [\chi(x, H)]_{\text{planar}} = \left[ W(x) + \frac{M}{\sigma^2 + \text{Tr}[\chi^{aa}(x)]} I_N \right]^{-1}
\]

Moreover, one can show that the variance of each element of the \( \chi \) matrix is of the order of \( 1/M \) and vanishes in the \( N, M \to \infty \) limit (for \( a = M/N \) held fixed). Covariance of \( \chi \), \([\chi^{ab}(x, H) \chi^{cd}(x, H)]_{\text{planar}}\) could be computed using the diagrams in Fig. 3 and they are suppressed in the large \( M, N \) limit, since their contributions are \( O \left( \frac{1}{M^2 N^2} \right) \).

Note that only the diagonal terms \( \chi^{aa} \) have non-trivial means, whereas the off-diagonal terms average to zero. For diagonal terms, namely local susceptibilities, we get
the following equations:
\[ [\chi^{ab}(x)]^{av}_{\mathbf{H}} = \left[ W_{aa}(x_a) + \frac{1}{\sigma^2 + \chi(x)} \right]^{-1} \]
\[ \chi(x) = \frac{1}{N} \sum_{a} [\chi_{aa}(x)]^{av}_{\mathbf{H}}. \]
Thus, given the distribution of \( W_{aa} \), which themselves depend on the distribution of \( x_a \) (where \( x_a = u_a + x_{0a} \)), we can determine the distribution of \( [\chi^{aa}]^{av}_{\mathbf{H}} \), self-consistently, from these equations.

One should note, although \( [\chi_{ab}]^{av}_{\mathbf{H}} = 0 \) for \( a \neq b \), for a particular choice of \( \mathbf{H} \), \( \chi^{ab} \) is a \( \mathbf{H} \)-dependent number of the order \( 1/\sqrt{M} \). Even if these off-diagonal terms are small compared to the self-averaging diagonal terms, they have an important effect on the self-consistency equations via the so-called Onsager reaction term [20]. In particular, we will need the correlation of \( \chi^{ab} \) with the corresponding matrix elements of \( \mathbf{H}^T \mathbf{H} \).

Using the identity \( [\mathbf{W} + \mathbf{H}^T \mathbf{H}] \chi = \mathbf{I}_N \), we can prove a useful corollary of the result in Eq. (42).
\[ \frac{1}{\sigma^2} [\mathbf{H}^T \mathbf{H} \chi(x, \mathbf{H})]^{av}_{\mathbf{H}} = \mathbf{I}_N - \mathbf{W}(x) \chi^{av}(x) \]
\[ = \mathbf{I}_N - \mathbf{W}(x) \left[ \mathbf{W}(x) + \frac{M}{\sigma^2 + \chi^{av}(x)} \mathbf{I}_N \right]^{-1} \]
\[ = \frac{M}{\sigma^2 + \chi^{av}(x)} \left[ \mathbf{W} + \frac{M}{\sigma^2 + \chi^{av}(x)} \mathbf{I}_N \right]^{-1} \chi^{av}(x) \]
\[ = \frac{\alpha \chi^{av}(x)}{\alpha \sigma^2 + \chi(x)} \]
\[ = \frac{\alpha \chi^{av}(x)}{\alpha \sigma^2 + \chi(x)} \]
In particular, Eq. (45) implies
\[ [\mathbf{I}^{\mathbf{H}}]^{av}_{\mathbf{H}} = \frac{M \sigma^2 \chi^{av}(x)}{\alpha \sigma^2 + \chi(x)} \]
which will be a useful identity in the next section.

V. ZERO-TEMPERATURE CAVITY METHOD: REMOVING A VARIABLE NODE

As a motivation for introduction of the cavity method [14], notice that Eq. (43) could be written as
\[ [\chi^{aa}]^{av}_{\mathbf{H}} = 1/(W_{aa} + \sigma^2). \]
This expression for local susceptibility would make sense if we could break this problem into effectively independent optimization problems for individual variables in the following manner: For the variable \( u_a \), the function to be optimized is
\[ \mathcal{E}_{\text{eff}}(u_a; x_{0a}, \xi_a) = \frac{1}{2\sigma_{\text{eff}}^2} (u_a^2 - 2u_a u_a) + \frac{1}{2} \frac{\xi_a}{\alpha}, \]
which is also the leading planar diagrams in covariance computation are of the order \( O(\frac{1}{\sqrt{M}}, \frac{1}{N}) \), as can be seen from counting a factor of \( M \) or \( N \) for appropriate index loop, and counting a factor of \( \frac{1}{\sqrt{M}} \) for each double-line contraction coming from averaging over the matrix elements.

Before we leave the section, we should mention that many observations made here are independent of the assumption that \( U(x) \) has a continuous second derivative. For example, we use \( U(x) = |x| \) in the Sec. VIII. We could define a second-differentiable function \( U_x(x) \) such that \( \lim_{x \to 0} U_x(x) = U(x) \), for example, \( U_x(x) = \sqrt{x^2 + \epsilon x} \) or \( U_x(x) = \frac{1}{2} \log(2 \cosh(\epsilon)) \). If \( x_a = x_{0a} + u_a \) goes to zero as \( \epsilon \) vanishes, then the corresponding \( W_{aa} = U_x''(x_a) \) diverges. However, the corresponding local susceptibility, \( \chi^{aa} \), just becomes zero in this limit. Therefore, as \( \epsilon \to 0 \), the idea of using effective single variable optimization problems and determining the self-consistent distribution of \( x_a \) and \( \chi^{aa} \) becomes meaningful. We just need to separate out the set of variables \( x_a \) for which \( W_{aa} \) diverges and treat this set carefully. As a consequence of \( \chi \) remaining well-defined in the \( \epsilon \to 0 \) limit, many relations derived in this section, such as Eqs. (37), (45) and (46), remain valid even if the potential \( U(x) \) becomes singular at some values of \( x \).

FIG. 3. The leading planar diagrams in covariance computation are of the order \( O(\frac{1}{\sqrt{M}}, \frac{1}{N}) \), as can be seen from counting a factor of \( M \) or \( N \) for appropriate index loop, and counting a factor of \( \frac{1}{\sqrt{M}} \) for each double-line contraction coming from averaging over the matrix elements.
graph [27], with the circles being the variable nodes and the squares being the ‘check’ nodes. The system with $N$ variables (circles) and $M$ data constraints (squares) would be represented as the $(N, M)$ system. Our task is to relate properties of the $(N, M)$ system to $(N - 1, M)$ system and obtain self consistency conditions based on quantities that converge in the thermodynamic limit, $N, M \to \infty$.

We pick a particular node $a$ and partition the cost function

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{H}^T \mathbf{H} \mathbf{u} + V(\mathbf{u} + \mathbf{x}_0) \quad (47)$$

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

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

$$+ \frac{1}{\sigma^2} \left( \sum_{b \neq a} u_b^2 \right)^2 \sum_{b \neq a} U(u_b + x_{0b}) \quad (48)$$

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

$$[\mathbf{h}_a^2]_{\text{av}} = \sum_i \left( \mathbf{H}_{i\alpha}^{\text{av}} \right)^2 = \frac{1}{M} \sum_{i=1}^M \frac{1}{M} = 1, \quad (49)$$

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

The system without node ‘$a$’, i.e., the system with a ‘cavity’ (see Fig. 5), will have its own optimum values $u_b = \hat{u}_b$, for all $b \neq a$. The variable $u_a$ interacts with the rest of the system through the quantity $\mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b$. The program of cavity method is to characterize the distribution of this quantity in terms of some parameters relating to the $(N - 1, M)$ system, and then use the fact that node ‘$a$’ is statistically the same as every other node to relate these parameters to the distribution of $u_a$.

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

Unfortunately, such a ‘straightforward’ approach is wrong on two counts. First, the components of $u_a$ cannot be replaced by $\hat{u}_\alpha$. Second, the components of the optimum vector $\mathbf{u}_\alpha$ are correlated with the vectors $\mathbf{h}_b$, making the variance estimate wrong.

Correcting the first error requires introducing the Onsager reaction terms [20]. One can start by rewriting Eq. (48) as follows

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) + \mathcal{E}_\alpha(\mathbf{u}_\alpha) - \mathbf{u}_\alpha^T \mathbf{f}_\alpha. \quad (50)$$

The cost function of the $(N-1,M)$ system is $\mathcal{E}_\alpha(\mathbf{u}_\alpha)$. We identify $(\mathbf{f}_\alpha)_b = -\frac{1}{\alpha} \mathbf{h}_b \cdot \mathbf{h}_a u_a$ to be the local force exerting on each node $u_a$, due to presence of node $u_a$. Since we are looking for the ground state, we minimize the expression in Eq. (50)

$$\min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = \min_{u_a, u_{\alpha}} \left( \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) \right)$$

$$+ \mathcal{E}_\alpha(\mathbf{u}_\alpha) - \mathbf{u}_\alpha^T \mathbf{f}_\alpha. \quad (51)$$

Given that $\mathbf{h}_b \cdot \mathbf{h}_a$ is of the order $1/\sqrt{M}$. $\mathbf{f}_\alpha$ is small, and we can invoke the definition of susceptibility $\chi_{\alpha}$ for the $(N-1,M)$ system and use expansion of the minimized cost function (37).

$$\min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = \min_{u_a, u_{\alpha}} \left( \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) \right)$$

$$+ \mathcal{E}_\alpha(\mathbf{u}_\alpha) - \mathbf{u}_\alpha^T \mathbf{f}_\alpha - \frac{1}{\sigma^2} \chi_{\alpha} \mathbf{f}_\alpha \chi_{\alpha} \mathbf{f}_\alpha (52)$$

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

$$\min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = \min_{u_a, u_{\alpha}} \left( \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) \right)$$

$$+ \mathcal{E}_\alpha(\mathbf{u}_\alpha) - \mathbf{u}_\alpha^T \mathbf{\hat{u}}_\alpha - \frac{1}{\sigma^2} \chi_{\alpha} (\mathbf{\hat{u}}_\alpha - \mathbf{\hat{u}}_\alpha^T \mathbf{\hat{u}}_\alpha) \quad (53)$$
The quantity \( \sum_{b,c=a} (h_b)(h_c) \chi \gamma^{bc} \) is independent of \( h_a \). As a result, \( \sum_{b,c=a} (h_b)(h_c) \chi \gamma^{bc} \) can be replaced by \( \sum_{b,c=a} (h_b)(h_c) \chi \gamma_{ia}^{bc} / M \), thanks to the self-averaging of \( (h_b)(h_c) \). Using Eq. (46) for the \((N-1,M)\) system,

\[
\sum_{b,c=a} h_b h_c \chi \gamma_{ia}^{bc} = \frac{M}{\alpha \sigma^2} \chi \gamma_{ia}^{bc} = \alpha \sigma^2 \chi \gamma_{ia}^{bc} / \sigma^2 + \chi
\]

with the last step having to do with \( \chi \) becoming independent of \( N,M \) asymptotically. Using this last relation Eq. (55), in Eq. (54) we get

\[
\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\chi}{\alpha}
\]

Looking at Eq. (53), the node variable \( u_a \) is coupled to the rest of the system, via the vector \( v = \sum_{b \neq a} h_b \hat{u}_b \), which is independent of \( h_a \) and depends solely on the \((N-1,M)\) system. Since \( u_a \) is coupled to \( h_a \cdot v \), we need to know the moments of this dot product. Because \( v \) is independent of \( h_a \), for mean, we have

\[
[h_a \cdot v]_{\text{av}} = [h_a]_{\text{av}} \cdot [v]_{\text{av}} = 0
\]

and for variance,

\[
[(h_a \cdot v)^2]_{\text{av}} = \sum_{ij} [H_{ia}H_{ja}]_{\text{av}} [v_i v_j]_{\text{av}} = \frac{M}{\sum_i [v_i^2]_{\text{av}}}.
\]

The order \( k \) cumulants go as \( M^{1-k/2} \) and, for \( k > 2 \), they tend to zero as \( M \) goes to infinity. Therefore, we will stop with the variance and treat \( h_a \cdot v \) as a zero-mean Gaussian variable. We still need the variance, for which we need a condition to determine \( [v_i^2]_{\text{av}} \). This requires a second step of the cavity method.

**VI. ZERO TEMPERATURE CAVITY METHOD: REMOVING A CONSTRAINT NODE**

The subtlety in determining \( [v_i^2]_{\text{av}} \) involves accounting for correlation between matrix elements \( H_{ib} \) and the optimal values \( \hat{u}_b \) of the \((N-1,M)\) system. To do this, we need to set up an \((N-1,M-1)\) system with the constraint \( i' \) removed (see Fig. 6). Such a two-stage cavity method has been used in the context of Hopfield neural networks before [23].

To find \( v_i = \sum_{b \neq a} H_{ib} \hat{u}_b \), we break up the minimization over \( u_a \) into two steps:

\[
\min_{u_a} \mathcal{E}_{ia}(u_a) = \min_{v_i} \left\{ \min_{u_a, \sum_{b \neq a} H_{ib} \hat{u}_b = v_i} \{ \mathcal{E}_{ia}(u_a) \} + \frac{1}{2\sigma^2} v_i^2 \right\}
\]

(59)

FIG. 6. The \((N-1,M-1)\) cavity system. Node \( a \) and constraint \( i \) have been removed from the system by removing the links to them.

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

Our strategy would be to compute the effect of perturbation in terms of the system susceptibility.

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

\[
\min_{v_i} \mathcal{E}_{ia}(u_a) = \max_{\gamma_i} \min_{u_a,v_i} \{ \mathcal{E}_{ia}(u_a) \} + \frac{1}{2\sigma^2} v_i^2 - \gamma_i (v_i - \sum_{b \neq a} H_{ib} \hat{u}_b) \}.
\]

(60)

Minimizing Eq. (60) with respect to \( v_i \) we get \( v_i = \sigma^2 \gamma_i \), and making that substitution for \( v_i \) into the cost function we get

\[
\min_{u_a} \mathcal{E}_{ia}(u_a) = \max_{\gamma_i} \min_{\gamma_i} \{ \mathcal{E}_{ia}(u_a) - \frac{1}{2\sigma^2} \gamma_i^2 - u_a^T g \} = \max_{\gamma_i} \left\{ -\frac{1}{2\sigma^2} \gamma_i^2 + \mathcal{E}_{\gamma_i}^*(g) \right\}
\]

(61)

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

\[
\mathcal{E}_{\gamma_i}^*(g) = \min_{u_a} \{ \mathcal{E}_{ia}(u_a) - u_a^T g \}
\]

(62)

where the presence of \( g \) alters the optimal \( u_a \) from the unconstrained optimum \( u_a' \). Since each component of \( g \) is small \((O(1/\sqrt{M}))\), we can expand around \( u_a' \) using \( \chi_{ai} \), the susceptibility of the \((N-1,M-1)\) system, as in Eq. (37). Therefore, \( \mathcal{E}_{\gamma_i}^*(g) \) can be written as

\[
\mathcal{E}_{\gamma_i}^*(g) = \mathcal{E}_{\gamma_i}(u_a') - u_a'^T g - \frac{1}{2} g^T \chi_{ai} g + \cdots
\]

(64)

Now, Eq. (62) becomes

\[
\min_{u_a} \mathcal{E}_{ia}(u_a) = \min_{\gamma_i} \left\{ -\frac{1}{2} \gamma_i^2 + \mathcal{E}_{ia}(u_a') - u_a'^T g - \frac{1}{2} g^T \chi_{ai} g \right\}
\]

(65)
The quadratic term $g^T \chi_{ai} g = \gamma_i^2 \sum_{ij} H_{ib} H_{ic} \chi_{ai}$ can be simplified because of self-averaging. We have
\[
\sum_{ij} [H_{ib} H_{ic}]_{av} \chi_{ai} = \frac{1}{M} \sum_b \chi_{ab} = \frac{1}{\alpha} \chi_{ai},
\]
(66)
once more using the fact that the average local susceptibility $\chi$ is nearly the same for the $(N, M)$ system and the $(N-1, M-1)$ system.

Putting everything together
\[
\min_{u_a} \mathcal{E}_u(u_a) = \max \left\{ \frac{\sigma^2}{2} (1 + \frac{1}{\alpha \sigma^2}) \gamma_i^2 + \gamma_i \sum_{b=0}^N H_{ib} u_b' \right\},
\]
(67)
maximizing with respect to $\gamma_i$ and then using $\gamma_i = \sigma^2 / \gamma_i$ gives us
\[
v_i = \frac{1}{1 + \frac{1}{\alpha \sigma^2} \sum_{b=0}^N H_{ib} u_b'},
\]
(68)
The denominator $(1 + \frac{1}{\alpha \sigma^2})$ ‘scales down’ the unconstrained answer $\sum_{b=0}^N H_{ib} u_b'$. It is the same factor that relates $\sigma^2$ to $\sigma_{\alpha}^2$.

Given that this result is true for any $i$'s, (53) becomes
\[
\min_{u} \mathcal{E}(u) = \min_{u_a} \left\{ \frac{1}{2 \sigma_{\alpha}^2} u_a^2 - \frac{\xi}{\sigma^2 (1 + \frac{1}{\alpha \sigma^2})} u_a + U(u_a + x_{0a}) \right\}
\]
(69)
with
\[
\xi = - \sum_i H_{ia} \sum_{b=0}^N H_{ib} u_b'
\]
(70)
being a random Gaussian variable with mean zero and variance
\[
\sigma^2 \xi = \left\{ \xi^2 \right\}_{x_0, \xi} = \sum_{i,j} [H_{ia} H_{ja}]_{av} \sum_{b=0}^N [H_{ib} H_{ij}]_{av} \left\{ u_b' u_c' \right\}_{x_0, \xi}
\]
\[
= \sum_{i,j} \sum_{b,c} \sum_{b,c} \frac{\delta_{ij}}{M} \left\{ u_b' u_c' \right\}_{x_0, \xi}
\]
\[
= \frac{1}{M} \sum_{b=0}^N \left\{ u_b' \right\}_{x_0, \xi} = \frac{q}{\alpha}
\]
(71)
thanks to $H_{ja}$ and $H_{ib}$ being independent for $a \neq b$, as well as $u_b'$'s being independent of those matrix elements. The quantity $q \equiv \frac{1}{N-1} \sum_{b,c} \left\{ u_b^2 \right\}_{x_0, \xi}$ is the MSE for the $(N-1, M-1)$ system.

In summary, the zero temperature problem boils down to a collection of independent single variable optimization
\[
\min_{u_a} \left\{ \frac{1}{2 \sigma_{\alpha}^2} u_a^2 - 2 \xi a u_a + U(u_a + x_{0a}) \right\}
\]
(72)
which has the same effective cost function as Eq. (27) in Sec. III. The variables $\xi_a$ are chosen independently from $\mathcal{N}(0, \sigma_{\alpha}^2)$. With $x_{0a}, \xi_a$ chosen randomly, we can calculate a distribution of $u_a$ and $\chi_{\alpha a}$. The self-consistency require $u_a$ average to satisfy $\sigma_{\alpha}^2 = q/\alpha$ and $\chi_{\alpha a}$ average $\chi$ to be related to $\sigma_{\alpha}^2$ by $\sigma_{\alpha}^2 = \sigma^2 + \chi/\alpha$.

In this formulation, we do not need to invoke temperature. One can consider $\chi$ to be equivalent to the $\beta Q$ in the replica approach. As we will see in the next section, we could use $\chi$ to distinguish phases around the zero-temperature transition described by Donoho and Tanner [8].

\section{Ridge Regression}

Let us start with simplest form of regularization with $U(x) = \frac{1}{2} x^2$, a penalty function that does not impose sparsity on the solutions. This is just ridge regression with Tikhonov regularization [29]
\[
\hat{x} = \arg \min_{x} \left\{ \frac{1}{2 \sigma_{\alpha}^2} \left( H(x - x_0) \right)^2 + \frac{\lambda}{2} x^2 \right\}
\]
(73)
We could explicitly minimize $x$ and proceed with our analysis using random matrix theory, but, instead, we will apply first the self-consistent formalism we have developed.
\[
\min_{u} \left\{ \frac{1}{2 \sigma_{\alpha}^2} (u^2 - 2 \xi u) + \frac{\lambda}{2} (u + x_0)^2 \right\}
\]
(74)
Minimizing we get
\[
u = \frac{-\lambda x_0 + \frac{1}{\sigma_{\alpha}^2} \xi}{\lambda + \frac{1}{\sigma_{\alpha}^2}} \implies x = x_0 + \frac{\xi}{\lambda + \frac{1}{\sigma_{\alpha}^2}}
\]
(75)
Also, by adding a local field, one can see that local susceptibility is the same everywhere, and it is $\left( \lambda + \frac{1}{\sigma_{\alpha}^2} \right)^{-1}$. We will use $\chi$ to denote the same, for the sake of continuity of notation. Using Eq. (56), namely, $\sigma_{\alpha}^2 = \sigma^2 + \chi/\alpha$, we have a self-consistency condition for $\chi$:
\[
\chi = \chi \left[ \lambda + \frac{1}{\sigma^2 + \frac{2}{\alpha}} \right]^{-1},
\]
(76)
which, in turn, allows us to determine $\sigma_{\alpha}^2$. Also, $\sigma_{\alpha}^2$ is determined by
\[
\sigma_{\alpha}^2 = \frac{q}{\alpha} = \frac{1}{\alpha} \left[ u^2 \right]_{x_0, \xi} = \frac{\sigma^2 \chi}{\alpha (1 + \lambda \sigma_{\alpha}^2)^2}
\]
(77)
where $[\ldots]_{x_0, \xi}$ means average over $\pi(x_0)$.

It is particularly insightful to consider the $\sigma^2 \to 0$ limit, where we find the $x$ with the minimal $\ell_2$ norm, subject to linear constraints $Hx = Hx_0$. Rewriting Eq. (76) as
\[
1 - \frac{1}{\lambda \chi} = \frac{\sigma^2 (1 - \lambda \chi)}{\chi}
\]
(78)
we see that in this limit, $\lambda_\infty = 1 - \alpha$, and, therefore, $1 + \lambda \sigma^2 \rightarrow 1 + \lambda_\infty \alpha = \alpha^{-1}$. The Eqs. (75), (77) leads to

$$x = \alpha(x_0 + \xi) \quad \sigma^2 = \frac{(1 - \alpha)\rho_{x_0}^2}{\alpha}.$$

(79) (80)

In other words, $x$ is a Gaussian variable, with $x_0$ as its mean and $(1 - \alpha)\rho_{x_0}^2$ as its variance. Hence, $x_0$ is a projection operator, and we expect the fluctuation of $x$ around zero to be of the order of $(1 - \alpha)\rho_{x_0}^2 x_0$. We could have arrived at this conclusion by choosing a threshold from a different point of view. Elementary derivation could parametrize the result, of such a procedure. When $\rho < \alpha/(1 - \alpha)$, it is possible to choose a threshold $\theta$ such that $((1 - \alpha)\rho_{x_0}^2 x_0)^{1/2} \ll \theta \ll \alpha((x_0^2)^{1/2})$. With such a threshold, both error rates would be small.

Of course, we could have arrived at this conclusion from a different point of view. Elementary derivation leads us to an explicit expression:

$$x = \frac{H^T H}{\sigma^2} \left[ \frac{H^T H}{\sigma^2} + \lambda I_N \right]^{-1} x_0 = \sum_{i=1}^M \frac{s_i^2}{\sigma^2} \lambda \sigma^2 V_i V_i^T x_0.$$

(81)

where we use the singular vector basis of the matrix $H$, with $s_i$ being the non-zero singular values, and $V_i$ the corresponding right singular vectors. We take the limit of vanishing $\sigma^2$, we just have a projection of the $N$ dimensional vector $x_0$ to a $M$ dimensional projection spanned by $V_i$’s. In other words

$$x_a = \sum_{b=1}^N \sum_{i=1}^M V_{ia} V_{ib} x_{0a} = \sum_{a=1}^N P_{ab} x_{0a}.$$

(82)

$P$ being the projection matrix. For random $H$, $V_i$’s are just a random choice of $M$ orthonormal vectors. Thus, the properties of the estimate depends on the statistics of projection matrix to a random $M$ dimensional subspace.

$$[P_{ab}]_{H}^\text{av} = \frac{M}{N} \sum_{i=1}^M \delta_{ab} = \alpha \delta_{ab} \Rightarrow [\hat{x}_a]_{H}^\text{av} = \alpha x_{0a}.$$

(83)

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

$$\sum_{a} [P_{ab} P_{ac}]_{H}^\text{av} = \sum_{a} [P_{ba} P_{ac}]_{H}^\text{av} = [P_{bc}]_{H}^\text{av} = \alpha \delta_{bc}.$$

(84)

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

$$[P_{aa} P_{ac}]_{H}^\text{av} \approx (\sum_{a} [P_{aa}]_{H}^\text{av})^2 = \alpha^2.$$

(85)

Using the two constraints, represented by Eqs. (84) and (85), we can determine $A$ and $B$, in the large $M, N$

limit, leading to,

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

(86)

The variance is now given by,

$$[\hat{x}_a]_{H}^\text{av} - ([\hat{x}_a]_{H}^\text{av})^2 = \frac{\alpha(1 - \alpha)}{N} \delta_{bc} + \alpha^2 \delta_{ab} \delta_{bc} x_{0b} x_{0c} - (\alpha x_{0a})^2$$

(87)

recovering our earlier result.

VIII. PHASE TRANSITION IN BASIS PURSUIT: $\ell_1$-NORM MINIMIZATION

In this section, we reconsider the much-analyzed case where the penalty function is the $\ell_1$ norm of $x$, i.e., $U(x)$ is $\lambda |x|$ [2, 5, 8]. The reconstructed sparse solution is given by

$$x = \min_x \left\{ \frac{1}{2\sigma^2} (H(x - x_0))^2 + \lambda ||x||_1 \right\}.$$

(88)

Our task is to solve the self-consistency problem for the potential $U(x) = \lambda |x|$, equations (71) and (30) in the $\sigma^2 \rightarrow 0$ limit.

In this case $U''(x)$ is zero everywhere except at $x = 0$, where it is formally infinite. Consequently, $\chi^{aa} = 0$, if $x_a = 0$

$$\chi^{aa} = \sigma_{\text{eff}}^2,$$ otherwise.

(89)

The fact the $\chi^{aa}$ is the same for all non-zero values makes the analysis particularly simple. We define $\hat{\rho}$ to be fraction of $x_a$’s that are non-zero (think of this number as the estimated sparsity). Then $\chi = \hat{\rho} \sigma_{\text{eff}}^2$ and

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

(90)

implying

$$\sigma_{\text{eff}}^2 (1 - \hat{\rho}) = \sigma^2.$$

(91)

This equation is central to understanding the $\sigma^2 \rightarrow 0$ limit and the associated phase transition. When $\sigma$ goes to zero, we either have $\sigma_{\text{eff}}^2 = 0$ or $\hat{\rho} = \alpha$. These two conditions correspond to the two phases of the system, the first being the perfect reconstruction phase and the second, the non-zero error regime. In terms of average local susceptibility, the first phase has $\chi = \hat{\rho} \sigma_{\text{eff}}^2 = 0$, while the second one has $\chi \neq 0$.

Note the condition (91) involves $\sigma^2$ and $\sigma_{\text{eff}}^2$ explicitly, while the optimization result depends solely on $\hat{\rho} = \lambda \sigma^2$.
and not on $\sigma$ and $\lambda$ separately. However, just by multiplying the equation with $\lambda$, we could rewrite it as

$$\theta(1 - \frac{\hat{\xi}}{\alpha}) = \vartheta$$

(92)

with $\theta = \lambda\sigma_{\text{eff}}^2$. Further analysis of the self-consistency conditions depend crucially upon $\theta$. One could, in principle, describe the system solely in terms of $\theta$ and $\lambda$.

Now let us set up the notation for the single variable optimization problem. More precisely, by searching for the solutions of Eq. (28) and (29), we arrive at the following soft-thresholding function for the estimated value of $\hat{x}$ that we will denote by $\eta_{\text{soft}}(t; \theta)$, with the variable $t = x_0 + \xi$ and the parameter $\theta = \lambda\sigma_{\text{eff}}^2$.

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

(93)

## A. Perfect Reconstruction Regime

Now let us focus on the case where, as $\sigma^2$ becomes small, $\sigma_{\text{eff}}^2$ becomes small as well. According to Eq. (93), one source of error is when the $x_0$ was originally non-zero, but the estimated $\hat{x}$, thanks to the shift by $\xi$, has fallen into the $[-\lambda\sigma_{\text{eff}}^2, \lambda\sigma_{\text{eff}}^2]$ interval and then been truncated to zero (see Fig. 7). The probability of this event becomes small when $\sigma_{\text{eff}}^2$ gets to be small, and would be ignored for the moment being. Under this circumstance, if $\xi$ remains order one, the error is dominated by $\xi$ and $q = \sigma_{\text{eff}}^2$, which is not consistent with $\sigma_{\text{eff}}^2 = q/\alpha$, unless $\alpha^2 = 0$. Hence in this regime, we need to consider $\sigma_{\text{eff}}^2$ that is comparable to $\lambda\sigma_{\text{eff}}^2$. As $\sigma^2 \to 0$, we will have $\sigma_{\text{eff}}^2 \to 0$ and $q \to 0$ as well, making the reconstruction perfect. Let us now derive the self-consistency condition for this limit when $\sigma^2, \sigma_{\text{eff}}^2, \sigma_{\text{eff}}^2 \to 0$ with $\tau = \lambda\sigma_{\text{eff}}^2$ order one.

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

$$\rho[(\hat{x} - x_0)^2]_{x_0, \xi} = \rho[(\xi - \lambda\sigma_{\text{eff}}^2 \text{sgn}(\hat{x}))^2]_{x_0, \xi} = \rho(\sigma_{\text{eff}}^2 + (\lambda\sigma_{\text{eff}}^2)^2)$$

(94)

The other source of error is the event when the $x_0$ is zero but $\hat{x}$ has fallen outside the interval $[-\lambda\sigma_{\text{eff}}^2, \lambda\sigma_{\text{eff}}^2]$ and has been estimated to be non-zero. In this case, the contribution to MSE is

$$(1 - \rho)[\hat{x}^2]_{x_0, \xi} = 2(1 - \rho) \int_{\frac{\xi}{\lambda\sigma_{\text{eff}}^2}}^{\infty} d\xi \frac{1}{\sqrt{2\pi\sigma_{\text{eff}}^2}} e^{-\frac{\xi^2}{2\sigma_{\text{eff}}^2}} (\xi - \lambda\sigma_{\text{eff}}^2)^2$$

$$= 2\sigma_{\text{eff}}^2 (1 - \rho) \left(1 + \tau^2\right) \Phi(\tau) - \tau\phi(\tau).$$

(95)

Adding the contributions from Eq. (94) and (95), we get the total MSE, $q$ (i.e. $\alpha\sigma_{\text{eff}}^2$). Therefore Eq. (71), $\sigma_{\text{eff}}^2 = q/\alpha$, becomes

$$\alpha = 2(1 - \rho) \left(1 + \tau^2\right) \Phi(\tau) - \tau\phi(\tau) + \rho(1 + \tau^2)$$

(96)

where $\Phi(\tau) = \int_{-\infty}^{\infty} dz \phi(z)$, and $\phi(\tau) = \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2}$.

If $x_0 = 0$, we have to have $|\xi| > \theta$ to lead to a non-zero $x$. A non-zero $x_0$ remains non-zero with probability approaching one in this case, as argued in the beginning of the section. Counting all sources of the non-zero $\hat{x}$’s, we have in the

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

(97)

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

In the error-prone phase $q, \sigma_{\text{eff}}^2$ and therefore $\sigma_{\text{eff}}^2$ need to be non-zero, forcing $\hat{\rho} = \rho$. If the transition happens continuously, the condition for phase boundary is $\alpha = \hat{\rho} = 2(1 - \rho) \Phi(\tau) + \rho$. Hence the relation between $\alpha$ and $\rho$ at the phase boundary is obtained by solving and eliminating $\tau$ from

$$\alpha = 2(1 - \rho) \left(1 + \tau^2\right) \Phi(\tau) - \tau\phi(\tau) + \rho(1 + \tau^2)$$

(98)

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

(99)

Alternatively, Eq. (98) and (99) can be solved for $\alpha, \rho$ at the phase boundary and expressed parametrically as a function of $\tau$

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

(100)

$$\rho = \frac{1 - \tau\Phi(\tau)}{\phi(\tau)}$$

(101)

leading to the phase diagram depicted in Fig. 8.

In the extremely sparse limit, $\rho \ll 1$, one can obtain a more explicit asymptotic relation between $\alpha$ and $\rho$. In this limit $\tau$ is large, and Eq. (98) and (99) behave, respectively, as

$$\alpha \approx \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{\tau^2}{2}}}{\tau}$$

(102)

Therefore, in this extreme sparsity limit, we have, $\alpha(\rho) \sim 2\rho \log \frac{1}{\rho}$, a result that is known for some time and comes close to the bounds from restricted isometry property [2].
B. Into the Error-prone Regime

To get a better understanding of the nature of this phase transition and its behavior as one decreases $\alpha$ from above $\alpha_c(\rho)$ to below, we should search for solutions of Eqs. (103) and (104) in the error regime where both $\sigma^2_{\text{eff}}$ and $\alpha^2_{\xi}$ remain $O(1)$. In this case, we have to deal carefully with the possibility that $\bar{x}$ has been set to zero, because $x_0 + \xi$ fell within $\pm \lambda \sigma^2_{\text{eff}}$. The self-consistency equation for $\sigma^2_{\xi}$ could be rewritten as

$$
\alpha = 2(1 - \rho) \left\{ (1 + \tau^2) \Phi(\tau) - \tau \phi(\tau) \right\} + \rho \left[ \tau^2 \{ 1 - \Phi(\tau + \tau_0) - \Phi(\tau - \tau_0) \} \right]^{\text{av}}_{x_0} \nonumber
$$

(103)

where $\left[ \ldots \right]_{x_0}^{\text{av}}$ means average over $\pi(x_0)$ and $\tau_0 = \frac{x_0}{\xi}$. The quantity $\tau$ and functions $\Phi(\tau)$ and $\phi(\tau)$ are defined as before.

The self-consistency equation for $\sigma^2_{\text{eff}}$ becomes

$$
\alpha = \frac{\alpha \sigma^2_{\xi}}{\sigma^2_{\text{eff}}} + 2(1 - \rho) \Phi(\tau) + \rho \left[ \Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \right]^{\text{av}}_{x_0} \nonumber
$$

(104)

$$
\rightarrow 2(1 - \rho) \Phi(\tau) + \rho \left[ \Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \right]^{\text{av}}_{x_0} \nonumber
$$

(105)

in the error-prone phase, because, as $\sigma^2 \rightarrow 0$, $\sigma^2_{\text{eff}}$ remains non-zero, making the term $\frac{\alpha \sigma^2_{\xi}}{\sigma^2_{\text{eff}}}$ drop off from Eq. (104).

Note that, when $|\tau_0| = \frac{x_0}{\xi} \rightarrow \infty$, $\Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \rightarrow 1$ and $(\tau - \tau_0) \phi(\tau + \tau_0), (\tau + \tau_0) \phi(\tau - \tau_0) \rightarrow 0$. The $\tau_0$ dependent expression inside $\left[ \ldots \right]_{x_0}^{\text{av}}$ in Eq. (103) goes from $2 \left\{ (1 + \tau^2) \Phi(\tau) - \tau \phi(\tau) \right\}$ to $1 + \tau^2$ as $\tau_0$ goes from zero to infinity. We will write this expression as $1 + \tau^2 - \psi(\tau_0, \tau)$. The function $\psi(\tau_0, \tau)$ is an even function of $\tau_0$ that falls off quickly as $\tau_0$ becomes much larger than 1. In the same manner, we write $\Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) = 1 - \psi(\tau_0, \tau)$, $\psi(\tau_0, \tau)$ being another even function of $\tau_0$ that becomes small as $\tau_0$ becomes much larger than 1.

Now we are ready to understand the behavior of Eqs. (103) and (105) close to the transition, with $\sigma^2_{\text{eff}}$ and $\sigma^2_{\xi}$ small. For that purpose, it is better to rewrite these equations as

$$
\alpha = 2(1 - \rho) \left\{ (1 + \tau^2) \Phi(\tau) - \tau \phi(\tau) \right\} + \rho (1 + \tau^2) \nonumber
$$

$$
- \rho \left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0} \nonumber
$$

(106)

$$
\alpha = 2(1 - \rho) \Phi(\tau) + \rho \left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0}. \nonumber
$$

(107)

Compared to Eqs. (98) and (99) which give us the phase boundary, these have the extra terms involving averages $\left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0}$ and $\left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0}$. Since

$$
\left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0} = \int d\tau_0 \pi(x_0) \psi(\tau_0, \tau) = \sigma_\xi \int d\tau_0 \pi(\tau_0) \psi(\tau_0, \tau) \nonumber
$$

(108)

The small $\sigma^2_{\xi}$ behavior of this average depends on how $\pi(x)$ behaves at small $x$. If $\pi(x)$ has a non-zero limit $\pi(0)$, then $\left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0} \approx \sigma_\xi \pi(0) \int d\tau_0 \psi(\tau_0, \tau) \sim \sigma_\xi$. Similarly $\left[ \psi(\tau_0, \tau) \right]^{\text{av}}_{x_0} \sim \sigma_\xi$. Thus, the perturbations added to Eqs. (98) and (99) are of the order of $\sigma_\xi$.

Let us express the phase boundary as $\alpha = \alpha_c(\rho)$, $\tau = \tau_c(\rho)$ solving Eqs. (98) and (99). Close to the boundary, $\alpha = \alpha_c(\rho) - \delta \alpha$ and $\tau = \tau_c(\rho) - \delta \tau$. Since the perturbations are of the order $\sigma_\xi$, we get

$$
\delta \alpha \sim \sigma_\xi = \sqrt{\frac{n}{\alpha}}. \nonumber
$$

(109)

This equation shows that, for nonzero terms drawn from a distribution with nonzero density at the origin, Eq. (109) tells us that the mean square error rises as

$$
q \sim (\alpha_c - \alpha)^2 \nonumber
$$

(110)

which can be confirmed by numerical experiments (see Fig. 9). Therefore, although the phase boundary $\alpha_c(\rho)$ does not depend on the distributions of non-zeros, the rise of the error is continuous, i.e. it is a second order phase transition, and its critical exponent depends on the behavior of $\pi(x_0)$ near $x_0 = 0$.

IX. ELASTIC NET

As a quick application of our zero temperature cavity method, we consider how phase transition would be affected if we generalize our penalty function $V(x)$ by adding a quadratic term $|x|^2$ to the $l_1$ norm. This penalty
function is used in Elastic Net method of variable selection and regularization [30]. The optimization problem becomes

\[
x_{\text{EN}} = \min_x \frac{1}{2\sigma^2} (y - Hx)^2 + \lambda_1 |x| + \frac{\lambda_2}{2} |x|^2
\]

(111)

In the noiseless reconstruction problem, \( y = Hx_0 \). We take the limit \( \sigma^2 \to 0 \) and choose the distribution of \( H \) and \( x_0 \) to be the same as in the previous sections.

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

\[
\chi^{\text{aa}} = 0, \text{ if } x_a = 0
\]

\[
\chi^{\text{aa}} = \frac{\sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2}, \text{ otherwise.}
\]

(112)

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

\[
\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\hat{\rho} \sigma_{\text{eff}}^2}{\alpha (1 + \lambda_2 \sigma_{\text{eff}}^2)}
\]

(113)

implying

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

(114)

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

For the corresponding single variable optimization problem, we can still use the soft-thresholding function described in Eq. (93). The estimated value of \( \hat{x} \) is once more given by \( \eta_{\text{soft}}(t; \theta) \), but with \( t = \frac{\Phi_{\text{fin}}}{1 + \lambda_2 \sigma_{\text{eff}}^2} \) and \( \theta = \frac{\lambda_2 \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2} \).

As before, we start in the perfect reconstruction phase, where \( \sigma^2, \sigma_{\text{eff}}^2, \sigma^2_x \to 0 \) with \( \tau = \frac{\lambda_2 \sigma_{\text{eff}}^2}{\sigma^2} \) order one. In this phase, we ignore the case of non-zero \( x_0 \) leading to \( \hat{x} = 0 \). The contribution to MSE for the non-zero \( x_0 \) is slightly different

\[
\rho [(\hat{x} - x_0)^2]_{\alpha,\xi} = \rho \left( \frac{x_0 + \xi - \lambda_1 \sigma_{\text{eff}}^2 \text{sgn}(\hat{x}) - x_0}{1 + \lambda_2 \sigma_{\text{eff}}^2} \right)^2
\]

(115)

The key approximations is that \( [x_0 \text{sgn}(\hat{x})]_{\alpha,\xi} \approx [x_0]_{\alpha,\xi} \), since, in this limit, typically \( |\xi| \ll |x_0| \) implying \( \hat{x} \) and \( x_0 \) have the same sign. The other source of error is the event when the \( x_0 \) is zero but \( \hat{x} \) has fallen outside the interval \([-\lambda_2 \sigma_{\text{eff}}^2, \lambda_2 \sigma_{\text{eff}}^2]\) and has been estimated to be non-zero. In this case, the contribution to MSE is

\[
(1 - \rho) [x_0^2]_{\alpha,\xi} = 2(1 - \rho) \int_{\lambda_1 \sigma_{\text{eff}}^2}^{\infty} \frac{\rho}{2\sigma^2} \left( \frac{\xi - \lambda_1 \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2} \right)^2 (1 + \gamma^2) (\Theta - \tau \Phi (\tau)) \text{d}x
\]

(116)

Combining Eq. (115) and (116) in the self-consistency equation for \( \sigma_{\text{eff}}^2 \) and remembering that \( \sigma^2, \sigma_{\text{eff}}^2 \to 0 \) with \( \tau = \frac{\lambda_1 \sigma_{\text{eff}}^2}{\sigma^2} \) order one, we have

\[
\alpha = 2(1 - \rho) \left( 1 + \gamma^2 \right) \Phi(\tau) - \tau \Phi(\tau)
\]

\[
+ \rho \left( 1 + \gamma^2 \right) \left( 1 + \lambda_2^2 \lambda_1^2 \sigma_{\text{fin}}^2 + \lambda_2^2 \left[ x_0 \right]_{\alpha,\xi}^2 \right)
\]

(117)

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

\[
\hat{\rho} = 2(1 - \rho) \Phi(\tau) + \rho
\]

(118)

The condition for phase boundary is \( \alpha = \hat{\rho} (1 + \lambda_2 \sigma_{\text{eff}}^2) \to \hat{\rho} = 2(1 - \rho) \Phi(\tau) + \rho \). Thus, for Elastic Net method, the phase boundary is obtained by solving and eliminating \( \tau \) from

\[
\alpha = 2(1 - \rho) \left( 1 + \gamma^2 \right) \Phi(\tau) - \tau \Phi(\tau)
\]

\[
+ \rho \left( 1 + \gamma^2 \right) \left( 1 + \lambda_2^2 \lambda_1^2 \sigma_{\text{fin}}^2 + \lambda_2^2 \left[ x_0 \right]_{\alpha,\xi}^2 \right)
\]

\[
\alpha = 2(1 - \rho) \Phi(\tau) + \rho
\]

(119)

(120)

For the same, \( \rho \), the transition happens at higher \( \alpha \) (see Fig. 9). For Gaussian \( \pi(x_0) \) with variance \( \sigma^2_{\alpha,\xi} \), the key
dimensionless parameter is \( \frac{\lambda x_0}{\sigma} \), which determines the relative strength of the quadratic penalty term.

Naively, one might think that we expect the Elastic Net [30] to perform worse than Basis Pursuit [4]. However, one could argue that in presence of strong correlations between different covariates, Elastic Net may have less false negatives compared to Basis Pursuit, when it comes to variable selection. Treatment of these issues and the role of additive (and other kinds of) noise would be left to a future study.

X. AFTERWORD

In this study, we deal with the phase transition in compressed sensing. In the case of a full random measurement matrix, we directly treat the optimization problem and show how to adapt the cavity method for doing mean field theory in the context. The mean field theory leads to a self-consistency condition on average mean squared error (MSE), since error in estimating one variable affects error in others. Careful derivation of the self-consistency condition, without using replica trick, involve accounting for subtle correlations in the system. To take care of these correlations, we needed a two-step cavity approach: one step removing a variable and then, another, removing a data constraint.

We had not discussed the role of additive noise, but the derivation of the modified self-consistency equations in presence of noise is rather straightforward. Although, we have emphasized the zero-temperature treatment, the cavity method can be used for finite temperature results as well. For completeness, we have provided the corresponding derivation in Appendix A. The key connection with the zero-temperature treatment is via the fluctuation-dissipation theorem [11], which relates thermal fluctuation with susceptibility.

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

The cavity approach looks at the behavior of the system for a particular choice of quenched variables, \( \mathbf{H} \) and \( x_0 \) in this case. In contrast, the replica approach centers on immediately averaging those quenched variables away. In the context of compressed sensing, one can imagine many problems, where the matrix \( \mathbf{H} \) is non-random. Currently there is no obvious way to extend the replica mean field treatment for such sensing matrices. The cavity method could be a more versatile tool in this regard. Extensions of this tool to other classes of compressed sensing problems would be a goal of future studies.

Appendix A: Finite Temperature Cavity Method

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

\[
E(\mathbf{u}) = \frac{1}{2\sigma^2}(\mathbf{H}\mathbf{u})^2 + V(\mathbf{u} + x_0)
\]  

(A1)

we define the Boltzmann distribution \( P(\mathbf{u}|\mathbf{H},x_0) \):

\[
P(\mathbf{u}|\mathbf{H},x_0) = \frac{1}{Z(\beta|\mathbf{H},x_0)} e^{-\beta E}
\]

(A2)

with the normalization factor/partition function given by

\[
Z(\beta|\mathbf{H},x_0) = \int d\mathbf{u} e^{-\beta E}
\]

(A3)

We now apply the first step of the two-step cavity method. First, we rewrite \( E \) as an interaction between variable \( u_a \) and the rest of the variables

\[
E(\mathbf{u}) = \frac{1}{2\sigma^2}(u^2_a + \frac{1}{\sigma^2} u_a h_a \sum_{b \neq a} h_b u_b + U(u_a + x_{0a}) + \mathcal{E}_{\mathcal{V}}(u_a))
\]

(A4)

By defining

\[
\eta_a \equiv \frac{h_a \cdot \sum_{b \neq a} h_b u_b}{h^2_a}
\]

(A5)

and using \( h^2_a = 1 + \mathcal{O}(\frac{1}{\sqrt{N}}) \) we have

\[
E = \frac{1}{2\sigma^2}(u^2_a - 2u_a \eta_a) + U(u_a + x_{0a}) + \mathcal{E}_{\mathcal{V}}(u_a)
\]

(A6)

with \( \mathbf{u}_a, \mathcal{E}_{\mathcal{V}} \) etc defined as in Sec. V. Equation (A6) indicates that the variable \( u_a \) interacts with all the others only through \( \eta_a \). Therefore, we rewrite the marginal distribution \( P(u_a) \) as an integral over the joint distribution of \( \eta_a \) and \( u_a \), \( P(u_a,\eta_a) \).

\[
P(u_a) = \frac{1}{Z} \int d\mathbf{u}_a e^{-\beta E} = \int d\eta_a P(u_a,\eta_a)
\]

(A7)

where

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

(A8)

for all \( a = 1, \ldots, N \). Now we introduce a cavity ‘field’ distribution of \( \eta_a \) at the removed node \( a \) as

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

(A9)

By comparing (A8) and (A9), we get

\[
P(u_a) = \frac{\int d\eta_a \exp\left[-\beta \left\{ \frac{1}{2\sigma^2}(u^2_a - 2u_a \eta_a) + U(u_a + x_{0a}) \right\} \right] P_{\mathcal{V}}(\eta_a)}{\int d\mathbf{u}_a d\eta_a \exp\left[-\beta \left\{ \frac{1}{2\sigma^2}(u^2_a - 2u_a \eta_a) + U(u_a + x_{0a}) \right\} \right] P_{\mathcal{V}}(\eta_a)}
\]

(A10)
The assumption of continuity of the global ground state, even in the presence of the cavity after removing node \(a\), is equivalent to the replica symmetric (RS) hypothesis. This is a valid assumption when the penalty function \(V\) is convex. Therefore, in the limit of \(N \to \infty\), even if the nodes of \((N - 1, M)\) system are weakly correlated, \(\eta_a\) is still a sum of many variables and \(P(\eta_a)\) can well be approximated by a Gaussian distribution.

\[
P_a(\eta_a) \propto e^{-\frac{(\eta_a - \langle \eta_a \rangle_a)^2}{2\sigma_a^2}} \quad (A11)
\]

Then (A10) becomes

\[
P(u_a) = \frac{\exp(-\frac{\beta}{2\sigma^2}(1-\frac{\beta}{2\sigma^2}(\delta v_i^2)_{ia} + 2\beta M \delta u_i \eta_i))}{\int du_a \exp(-\frac{\beta}{2\sigma^2}(1-\frac{\beta}{2\sigma^2}(\delta v_i^2)_{ia} + 2\beta M \delta u_i \eta_i))}
\]

\[
(A12)
\]

Therefore, only the thermal averages \(\langle \eta_a \rangle_a\) and the thermal fluctuation strength \(\langle \delta v_i^2 \rangle_{ia} = \langle (\eta_a - \langle \eta_a \rangle_a)^2 \rangle\) of the field \(\eta_a\) for the distribution \(P_a(\eta_a)\) are left to be computed. In that process the effects of (weak) correlation between the \(u_a\)'s have to be accounted for. Define, as in Sec. V,

\[
v_i = \sum_{b \neq a} H_{ib} u_b \quad (A13)
\]

and utilize our definition,

\[
\eta_a = -\sum_i H_{ia} v_i \quad (A14)
\]

then we arrive at

\[
\langle \eta_a \rangle_a = -\sum_i H_{ia} \langle v_i \rangle \quad (A15)
\]

and

\[
\langle \delta v_i^2 \rangle_{ia} = \sum_{ij} H_{ia} H_{ja} \langle \delta v_i \delta v_j \rangle
\]

\[
\approx \frac{1}{M} \sum_{ij} \delta v_i \delta v_j = \frac{1}{M} \sum_{ij} \langle \delta v_i^2 \rangle \quad (A16)
\]

Having done that we need to compute \(\langle v_i \rangle\) and \(\langle \delta v_i^2 \rangle\). To do so, this time in addition to site \(a\) we exclude site \(i\). Hence from (A6) we get

\[
E_{ia}(u_{ia}) = \frac{1}{2\sigma^2} v_i^2 + E_{ia}(u_{ia}) \quad (A17)
\]

After carrying out the same computation as in (A7), (A8), and (A10) for the marginal distribution \(Q_{a}(v_i)\), we arrive at

\[
Q_{a}(v_i) = \frac{\exp\left\{-\frac{\beta}{2\sigma^2} v_i^2 - \frac{(v_i - \langle v_i \rangle_a)^2}{2\langle \delta v_i^2 \rangle_{ia}}\right\}}{\int dv_i \exp\left\{-\frac{\beta}{2\sigma^2} v_i^2 - \frac{(v_i - \langle v_i \rangle_a)^2}{2\langle \delta v_i^2 \rangle_{ia}}\right\}} \quad (A18)
\]

Therefore

\[
Q_{a}(v_i) = \frac{\exp\left\{-\frac{\beta}{2\sigma^2} (1 + \frac{\sigma^2}{\beta}) (v_i - \langle v_i \rangle_a)^2\right\}}{\int dv_i \exp\left\{-\frac{\beta}{2\sigma^2} (1 + \frac{\sigma^2}{\beta}) (v_i - \langle v_i \rangle_a)^2\right\}}
\]

and then

\[
\langle v_i \rangle_a = \frac{\langle v_i \rangle_{ia}}{1 + \frac{\beta \Delta Q}{\sigma^2}} \quad (A21)
\]

and

\[
\langle \delta v_i^2 \rangle_{ia} = \frac{1}{1 + \frac{\beta \Delta Q}{\sigma^2}} \quad (A22)
\]

Notice how both these moments for the (\(N - M\)) system are weakly correlated, as the second approximate equality becoming exact in the thermodynamic limit. Then, we have

\[
\langle \delta v_i^2 \rangle_{ia} = \Delta Q/\alpha. \quad (A23)
\]

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

\[
\langle \delta v_i^2 \rangle_{ia} = \frac{1}{1 + \frac{\beta \Delta Q}{\sigma^2}} \quad (A24)
\]

Therefore from (A16), (A20), and (A34)

\[
\frac{\beta}{\sigma^2} \langle \delta \eta_a^2 \rangle_a = \frac{1}{1 + \frac{\beta \Delta Q}{\sigma^2}} \quad (A25)
\]

and from (A15), (A21), and (A34)

\[
\langle \eta_a \rangle_a = \frac{\sum H_{ia} \sum_{b \neq a} H_{ib} \delta u_{ia}}{1 + \frac{\beta \Delta Q}{\sigma^2}} \quad (A26)
\]
Moreover, we define
\[\xi_a = \sum_i H_{ia} \sum_{b \neq a} H_{ab} \langle u_b \rangle_{a|i} \] (A27)
which has variance \(\sigma^2 = q/\alpha\) with \(q\)
\[q = \frac{1}{N} \sum_a (u_a)^2 = \frac{1}{N-1} \sum_b (u_b)^2\] (A28)
being the mean squared error. Therefore, by plugging (A26) and (A27) into Eq. (A12), the marginal distribution for single variable \(u_a\) becomes
\[P(u_a) = \frac{\exp\left(-\frac{\beta}{2\sigma^2}(u_a^2 - 2u_a \xi_a) - \beta U(x_{0a} + u_a)\right)}{\int \exp\left(-\frac{\beta}{2\sigma^2}(u_a^2 - 2u_a \xi_a) - \beta U(x_{0a} + u_a)\right) du_a} \] (A29)
with \(\sigma^2_{\text{eff}} = \sigma^2(1 + \frac{\beta Q}{\alpha \sigma^2})\), and the effective cost function for the individual node is
\[\mathcal{E}(u_a) = \frac{1}{2\sigma^2_{\text{eff}}} (u_a^2 - 2u_a \xi_a) + U(x_{0a} + u_a) \] (A30)
Therefore, with \(\mathcal{E}\) replaced by a set of effectively decoupled nodes, and the sum over index \(a\) replaced by a quenched average over \(\xi_a, x_{0a}\). As a result, the self-consistency conditions for the MSE
\[q = \frac{1}{N} \sum_{a=1}^{N} (u_a)^2\] (A31)
and for
\[\Delta Q = \frac{1}{N} \sum_{a=1}^{N} (\delta u_a^2)\] (A32)
reduce to
\[q = \langle(u)^2\rangle_{\xi,x_0}^{\text{av}}\] (A33)
and
\[\Delta Q = \langle(\delta u^2)\rangle_{\xi,x_0}^{\text{av}}\] (A34)
where the thermal average \(\langle \ldots \rangle_{\text{eff}}\) is performed with respect to the effective individual node distribution (26) and \(\langle \ldots \rangle_{\xi,x_0}^{\text{av}}\) is the quenched average over variables \(\xi, x_0\), with \(\xi\) drawn from \(\mathcal{N}(0, q/\alpha)\) and signal \(x_0\) drawn independently from a distribution \(P(x_0)\). These self-consistency equations are exactly the same those from the RS ansatz in Sec. III.

**Appendix B: Correlated Random Matrices**

In this appendix we present the generalization of our results for Gaussian \(\mathcal{P}(\mathbf{H})\) with the matrix elements correlated in a ‘factorized’ manner, a special case of correlation that appears in many practical problems [17, 18, 21, 22, 25, 26]. More precisely, we have
\[\langle H_{ia} \rangle^{\text{av}} = 0\] (B1)
and
\[\langle H_{ia} H_{jb} \rangle^{\text{av}} = \frac{1}{M} C_{ij} D_{ab}.\] (B2)

In this case, one could still follow that diagrammatics of Section IV and show that Eq. (42) and Eq. (45) generalize to
\[\mathbf{H}^{\text{av}} = \left[ \mathbf{W} + \frac{1}{M\sigma^2} \text{Tr}(\mathbf{C}(\mathbf{I}_M + \frac{\text{Tr}(\mathbf{D}[\mathbf{X}^{\text{av}}(\mathbf{H})]^{\text{av}})}{M\sigma^2})^{-1}) \right]^{-1}\] (B3)
\[\frac{1}{\sigma^2} \mathbf{H}^{T} \mathbf{H}^{\text{av}} = \text{Tr}(\mathbf{C}(M\sigma^2 \mathbf{I}_M + \text{Tr}(\mathbf{D}[\mathbf{X}^{\text{av}}(\mathbf{H})]^{\text{av}})C^{-1})) \] (B4)
These results are closely related to earlier works by two of the authors on singular values of correlated matrices [21, 22]. Once more, the variance of each element of the matrix \(\mathbf{H}\) is of the order of \(1/M\) and vanishes in the \(N, M \rightarrow \infty\) limit (for \(\alpha = M/N\) held fixed).

The finite temperature replica saddle point equations [24] generalize to The saddle point \(Q = \mathbf{Q}, R = -\alpha \mathbf{R}\) satisfies the conditions:
\[\bar{Q}_{\mu \nu} = \frac{1}{N} \langle(u)^T_{\mu} \mathbf{D}_{\nu} \rangle\] (B5)
\[\mathbf{R} = \frac{\beta}{2\sigma^2} \text{Tr}_M \left[ \mathbf{C} \otimes \mathbf{I}_n (\mathbf{I}_M \otimes \mathbf{I}_n + \frac{\beta}{\alpha \sigma^2} \mathbf{C} \otimes \mathbf{Q})^{-1} \right] \] (B6)
obtained by differentiating \(S(Q, R)\) with respect to the elements of \(Q, R\). The trace \(\text{Tr}_M\) is a partial trace only applying to the \(M\) dimensional space. The expectation \(\langle(u)^T_{\mu} \mathbf{D}_{\nu} \rangle\) depends on \(\mathbf{R}\) via
\[\langle(u)^T_{\mu} \mathbf{D}_{\nu}\rangle = \beta \frac{\partial F(Q)}{\partial R_{\mu \nu}}\] (B7)
with \(\exp(-\beta F(Q))\)
\[= \left[ \int \prod_{\mu=1}^{n} \langle dN_u \rangle \exp\left(- \sum_{\mu, \nu} \bar{R}_{\mu \nu} u_{\mu}^* u_{\nu} - \beta \sum_{\mu} V(u_{\mu} + x_0)\right) \right]^{\text{av}}_{x_0}\] (B8)
For \(U(x)\) convex, we have replace symmetric ansatz for \(Q, R\): \(\bar{Q}_{\mu \nu} = (Q - q) \delta_{\mu \nu} + q\) and \(\bar{R}_{\mu \nu} = (R - r) \delta_{\mu \nu} + r\). The zero-temperature limit turns out to be a self-consistent problem of coupled nodes in presence of quenched correlated noise. The optimization problem reduces to finding
\[\min_u \left\{ \frac{1}{2\sigma^2} \langle u^T \mathbf{D} u - 2\xi^T \mathbf{D} u \rangle + V(u + x_0) \right\}\] (B9)
with
\[\frac{1}{\sigma^2_{\text{eff}}} = \frac{1}{M} \text{Tr}\left[ \mathbf{C} \left( \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D}[\mathbf{X}^{\text{av}}(\mathbf{H})]^{\text{av}})}{M}\right)^{-1} \right]^{-1}\] (B10)
where the Gaussian quenched vector \(\xi\) has mean zero and the covariance matrix given by
\[\frac{q \text{Tr}[\mathbf{C} \left( \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D}[\mathbf{X}^{\text{av}}(\mathbf{H})]^{\text{av}})}{M}\right)^{-1}]}{\alpha \left( \text{Tr}[\mathbf{C} \left( \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D}[\mathbf{X}^{\text{av}}(\mathbf{H})]^{\text{av}})}{M}\right)^{-1} \right)^2} \] (B11)
where

\[ q = \frac{1}{N}[u^TDu]_{\text{av}}^{\text{av}}, \quad (B12) \]

\[ q = \frac{1}{N}[u^TDu]_{\text{av}}^{\text{av}}. \]

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation INSPIRE (track 1) award 1344069.

[1] Alfredo Braunstein, Marc Mézard, and Riccardo Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

[2] Emmanuel Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[3] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[4] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[5] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[6] David L Donoho. For most large underdetermined systems of linear equations the minimal. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[7] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[8] David L Donoho and Jared Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9446–9451, 2005.

[9] Surya Ganguli and Haim Sompolinsky. Statistical mechanics of compressed sensing. *Physical review letters*, 104(18):188701, 2010.

[10] Yoshiyuki Kabashima, Tadashi Wadayama, and Toshiyuki Tanaka. A typical reconstruction limit for compressed sensing based on \( l_p \)-norm minimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(09):L09003, 2009.

[11] Rep Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255, 1966.

[12] H.-A Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 21(1):28–41, 2004.

[13] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[14] Marc Mézard, Giorgio Parisi, and MA Virasoro. SK model: The replica solution without replicas. *Europhys. Lett.*, 1(2):77–82, 1986.

[15] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9. World scientific Singapore, 1987.

[16] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

[17] Aris L Moustakas, Harold U Baranger, Leon Balents, Anirvan M Sengupta, and Steven H Simon. Communication through a diffusive medium: Coherence and capacity. *Science*, 287(5451):287–290, 2000.

[18] Aris L Moustakas, Steven H Simon, and Anirvan 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*, 49(10):2545–2561, 2003.

[19] Roberto Mulet, Andrea Pagnani, Martin Weigt, and Riccardo Zecchina. Coloring random graphs. *Physical Review Letters*, 89(26):268701, 2002.

[20] Lars Onsager. Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, 1936.

[21] Anirvan M Sengupta and Partha P Mitra. Distributions of singular values for some random matrices. *Physical Review E*, 60(3):3389, 1999.

[22] Anirvan Mayukh Sengupta and Partha Pratim Mitra. Capacity of multivariate channels with multiplicative noise: Random matrix techniques and large-\( n \) expansions (2). *Journal of Statistical Physics*, 125(5-6):1223–1242, 2006.

[23] Maoz Shamir and Haim Sompolinsky. Thouless-anderson-palmer equations for neural networks. *Physical Review E*, 61(2):1839, 2000.

[24] Koujin Takeda and Yoshiyuki Kabashima. Statistical mechanical analysis of compressed sensing utilizing correlated compression matrix. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1538–1542. IEEE, 2010.

[25] Koujin Takeda and Yoshiyuki Kabashima. Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals. *EPL (Europhysics Letters)*, 95(1):18006, 2011.

[26] Koujin Takeda and Yoshiyuki Kabashima. Reconstruction algorithm in compressed sensing based on maximum a posteriori estimation. In *Journal of Physics: Conference Series*, volume 473, page 012003. IOP Publishing, 2013.

[27] Robert Michael Tanner. A recursive approach to low complexity codes. *Information Theory, IEEE Transactions on*, 27(5):533–547, 1981.

[28] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[29] Andrey Nikolayevich Tikhonov. On the stability of inverse problems. *Dokl. Akad. Nauk SSSR*, 39(5):195–198, 1943.

[30] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.