The Replica-Symmetric Prediction for Compressed Sensing with Gaussian Matrices is Exact

Galen Reeves*† and Henry D. Pfister*
*Department of Electrical Engineering, Duke University
†Department of Statistical Science, Duke University

***Abstract*—This paper considers the fundamental limit of compressed sensing for i.i.d. signal distributions and i.i.d. Gaussian measurement matrices. Its main contribution is a rigorous characterization of the asymptotic mutual information (MI) and minimum mean-square error (MMSE) in this setting. Under mild technical conditions, our results show that the limiting MI and MMSE are equal to the values predicted by the replica method from statistical physics. This resolves a well-known problem that has remained open for over a decade.**

I. Introduction

The canonical compressed sensing problem can be formulated as follows. The signal is a random n-dimensional vector $X^n = (X_1, \dots, X_n)$ whose entries are drawn independently from a common distribution P_X with finite variance. The signal is observed using noisy linear measurements of the form

$$Y_k = \langle A_k, X^n \rangle + W_k,$$

where $\{A_k\}$ is a sequence of n-dimensional measurement vectors, $\{W_k\}$ is a sequence of standard Gaussian random variables, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product between vectors. The primary goal is to reconstruct X^n from the set of m measurements $\{(Y_k, A_k)\}_{k=1}^m$. Since the reconstruction problem is symmetric under simultaneous scaling of X^n and $\{W_k\}$, the unit-variance assumption on $\{W_k\}$ incurs no loss of generality. In matrix form, the relationship between the signal and a set of m measurements is given by

$$Y^m = A^m X^n + W^m$$

where A^m is an $m \times n$ measurement matrix whose k-th row is A_k.

This paper analyzes of the minimum mean-square error (MMSE) reconstruction in the asymptotic setting where the number of measurements m and the signal length n increase to infinity. The focus is on scaling regimes in which the measurement ratio $\delta_n = m/n$ converges to a number $\delta \in (0, \infty)$. The objective is to show that the normalized mutual information (MI) and MMSE converge to limits,

$$\mathcal{I}_n(\delta_n) \triangleq \frac{1}{n} I(X^n; Y^m \mid A^m) \to \mathcal{I}(\delta)$$
$$\mathcal{M}_n(\delta_n) \triangleq \frac{1}{n} \mathsf{mmse}(X^n \mid Y^m, A^m) \to \mathcal{M}(\delta),$$

almost everywhere and to characterize these limits in terms of the measurement ratio δ and the signal distribution P_X.

Using the replica method from statistical physics, Guo and Verdú [1] provide an elegant characterization of these limits in the setting of i.i.d. measurement matrices. Their result was stated originally as a generalization of Tanaka's replica analysis of code-division multiple-access (CDMA) with binary signaling [2]. The replica method was also applied specifically to compressed sensing in [3]–[8]. The main issue, however, is that the replica method is not rigorous. It requires an exchange of limits that is unjustified, and it requires the assumption of replica symmetry, which is unproven in the context of the compressed sensing problem.

The main result of this paper is that replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X, has bounded fourth moment and satisfies a certain 'single-crossing' property. The proof differs from previous approaches in that we first establish some properties of the finite-length MMSE and MI sequences, and then use these properties to uniquely characterize their limits.

A. The Replica-Symmetric Prediction

We now describe the results predicted by the replica method. For a signal distribution P_X the function $R : \mathbb{R}_+^2 \to \mathbb{R}_+$ is defined as

$$R(\delta, z) = I_X\left(\frac{\delta}{1+z}\right) + \frac{\delta}{2}\left[\log(1+z) - \frac{z}{1+z}\right],$$

where $I_X(s) = I(X; \sqrt{s}X + N)$ is the scalar mutual information function (in nats) of $X \sim P_X$ under independent Gaussian noise $N \sim \mathcal{N}(0,1)$ with signal-to-noise ratio $s \in \mathbb{R}_+$ [1], [4].

Definition 1. The replica-MI function $\mathcal{I}_{\mathrm{RS}} \colon \mathbb{R}_+ \to \mathbb{R}_+$ and the replica-MMSE function $\mathcal{M}_{\mathrm{RS}} \colon \mathbb{R}_+ \to \mathbb{R}_+$ are defined as

$$\mathcal{I}_{\mathrm{RS}}(\delta) = \min_{z \ge 0} R(\delta, z)$$
$$\mathcal{M}_{\mathrm{RS}}(\delta) \in \arg\min_{z \ge 0} R(\delta, z).$$

The function $\mathcal{I}_{\mathrm{RS}}(\delta)$ is increasing, concave, and thus differentiable almost everywhere. The function $\mathcal{M}_{\mathrm{RS}}(\delta)$ is decreasing and, thus, continuous almost everywhere. If the minimizer is not unique, then $\mathcal{M}_{\mathrm{RS}}(\delta)$ may have jump discontinuities and may not be uniquely defined at those points; see Figure 1.

B. Statement of Main Result

In order to state our results, we need some further definitions. Let $R_z(\delta, z) = \frac{\partial}{\partial z} R(\delta, z)$ denote the partial derivative of $R(\delta, z)$ with respect to z. The *fixed-point curve* FP is the set of (δ, z) pairs where z is a stationary point of $R(\delta, z)$, i.e.

$$\mathrm{FP} = \left\{(\delta, z) \in \mathbb{R}_+^2 \,:\, R_z(\delta, z) = 0\right\}.$$

To emphasize the connection with mutual information, we often plot this curve using the change of variables $z \mapsto \frac{1}{2}\log(1+z)$.

The work of G. Reeves was supported in part by funding from the Laboratory for Analytic Sciences (LAS). The work of H. Pfister was supported part by the NSF under Grant No. 1545143. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

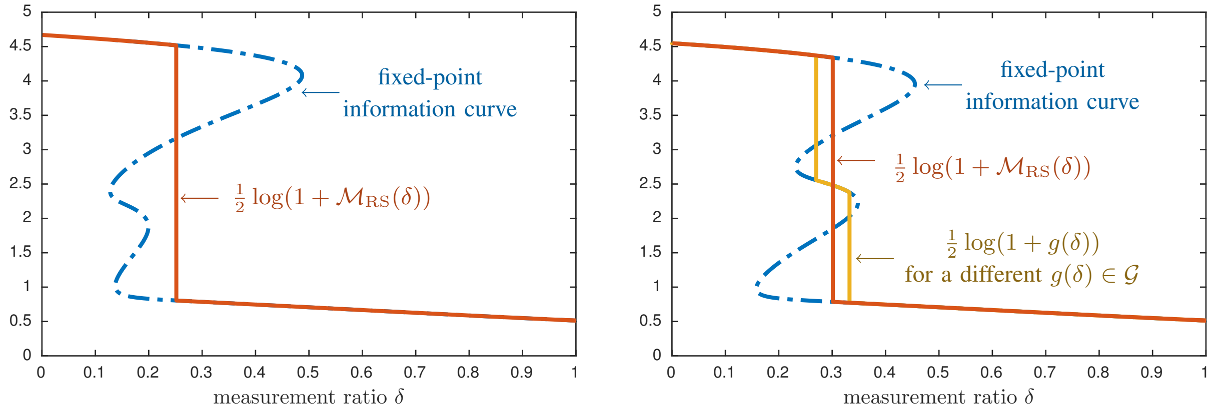

Fig. 1. Plot of the replica-MMSE as a function of the measurement ratio δ. The signal distribution is given by a three-component Gaussian mixture of the form $P_X = 0.4\mathcal{N}(0,5) + \alpha\mathcal{N}(40,5) + (0.6-\alpha)\mathcal{N}(220,5)$. In the left panel, $\alpha = 0.1$ and the distribution satisfies the single-crossing property. In the right panel, $\alpha = 0.3$ and the distribution does not satisfy the single-crossing property. The fixed-point information curve (dashed blue line) is given by $\frac{1}{2}\log(1+z)$ where z satisfies the fixed-point equation $R_z(\delta, z) = 0$.

The resulting curve, $(\delta, \frac{1}{2}\log(1+z))$, is called the *fixed-point information curve*; see Figure 1.

Definition 2 (Single-Crossing Property)**.** A signal distribution P_X has the single-crossing property if the replica-MMSE crosses the fixed-point curve FP at most once.

Assumption 1 (IID Gaussian Measurements)**.** The rows of the measurement matrix $\{A_k\}$ are independent Gaussian vectors with mean zero and covariance $n^{-1}I_n$. Furthermore, the noise $\{W_k\}$ is i.i.d. Gaussian with mean zero and variance one.

Assumption 2 (IID Signal Entries)**.** The signal entries $\{X_i\}$ are independent copies of a random variable $X \sim P_X$ with bounded fourth moment $\mathbb{E}[X^4] \leq B$.

Assumption 3 (Single-Crossing Property)**.** The signal distribution P_X satisfies the single-crossing property.

Theorem 1. *Under Assumptions 1-3, we have*

(i) The sequence of MI functions $\mathcal{I}_n(\delta)$ converges to the replica prediction. In other words, for all $\delta \in \mathbb{R}_+$,

$$\lim_{n\to\infty} \mathcal{I}_n(\delta) = \mathcal{I}_{\text{RS}}(\delta).$$

(ii) The sequence of MMSE functions $\mathcal{M}_n(\delta)$ converges almost everywhere to the replica prediction. In other words, for all continuity points of $\mathcal{M}_{\text{RS}}(\delta)$,

$$\lim_{n\to\infty} \mathcal{M}_n(\delta) = \mathcal{M}_{\text{RS}}(\delta).$$

Remark 1. The primary contribution of Theorem 1 is for the case where $\mathcal{M}_{\text{RS}}(\delta)$ has a discontinuity. This occurs, for example, in applications such as compressed sensing with sparse priors and CDMA with finite alphabet signaling. For the special case where $\mathcal{M}_{\text{RS}}(\delta)$ is continuous, the validity of the replica prediction can also be established by combining the AMP analysis with the I-MMSE relationship [9]–[13].

Remark 2. For a given signal distribution P_X the single-crossing property can be verified numerically by evaluating the replica-MI and replica-MMSE functions.

C. Related Work

The replica method was developed originally to study mean-field approximations in spin glasses [14], [15]. It was first applied to linear estimation problems in the context of CDMA wireless communication [1], [2], [16], with subsequent work focusing on the compressed sensing directly [3]–[8].

Within the context of compressed sensing, the results of the replica method have been proven rigorously in a number of settings. One example is given by message passing on matrices with special structure, such as sparsity [9], [17], [18] or spatial coupling [19]–[21]. However, in the case of i.i.d. matrices, the results are limited to signal distributions with a unique fixed point [10], [12] (e.g. Gaussian inputs [22], [23]). For the special case of i.i.d. matrices with binary inputs, it has also been shown that the replica prediction provides an upper bound for the asymptotic mutual information [24]. Bounds on the locations of discontinuities in the MMSE with sparse priors have also been obtain by analyzing the problem of approximate support recovery [6]–[8].

Recent work by Huleihel and Merhav [25] addresses the validity of the replica MMSE directly in the case of Gaussian mixture models, using tools from statistical physics and random matrix theory [26], [27].

II. Overview of Proof

A. Properties of the Replica Prediction

We begin with an alternate characterization of the replica limits. Using the I-MMSE relationship [28], it can be shown that the $R_z(\delta, z) = 0$ condition implies that the replica MMSE must obey the fixed-point equation

$$\mathcal{M}_{\text{RS}}(\delta) = \mathsf{mmse}_X\left(\frac{\delta}{1+\mathcal{M}_{\text{RS}}(\delta)}\right), \tag{1}$$

where $\mathsf{mmse}_X(s) = \mathsf{mmse}(X \mid \sqrt{s}X + N)$.

Furthermore, for all δ where $\mathcal{I}'_{\text{RS}}(\delta)$ exists, one can combine the envelope theorem [29] with the fixed-point condition (1) to show that the derivative satisfies

$$\mathcal{I}'_{\text{RS}}(\delta) = \frac{1}{2}\log(1+\mathcal{M}_{\text{RS}}(\delta)),$$

almost everywhere. Additionally, one can show that $\mathcal{I}_{\text{RS}}(\delta) = I_X(\delta) + o(\delta)$ and this implies that

$$\lim_{\delta\to\infty}\left| I_X(\delta) - \int_0^\delta \frac{1}{2}\log(1+\mathcal{M}_{\text{RS}}(\gamma))\mathrm{d}\gamma \right| = 0. \tag{2}$$

Definition 3. Let $\mathcal{G}$ be the subset of non-increasing functions from $\mathbb{R}_+ \to \mathbb{R}_+$ such that all $g \in \mathcal{G}$ satisfy both

$$g(\delta) = \mathsf{mmse}_X\left(\frac{\delta}{1+g(\delta)}\right) \tag{3}$$

almost everywhere and

$$\lim_{\delta\to\infty}\left|I_X(\delta) - \int_0^\delta \frac{1}{2}\log(1+g(\gamma))\mathrm{d}\gamma\right| = 0. \tag{4}$$

By combining (1) and (2), one can show that the replica-MMSE function satisfies $\mathcal{M}_{\mathrm{RS}}(\delta) \in \mathcal{G}$. At this point, it is natural to ask whether these conditions essentially define $\mathcal{M}_{\mathrm{RS}}(\delta) \in \mathcal{G}$ uniquely. The answer to this question is yes if the signal distribution P_X satisfies the single-crossing property. This is illustrated graphically in Figure 1.

Lemma 2. *If P_X has the single-crossing property, then $\mathcal{M}_{\mathrm{RS}}(\delta)$ has at most one discontinuity and all $g \in \mathcal{G}$ satisfy $g(\delta) = \mathcal{M}_{\mathrm{RS}}(\delta)$ almost everywhere.*

B. Properties of the MMSE and MI Sequences

For a fixed problem of size n, the MMSE sequence M_m and mutual information sequence I_m are defined as

$$M_m \triangleq \frac{1}{n}\mathbb{E}[\mathrm{tr}(\mathsf{Cov}(X^n \mid Y^m, A^m))]$$
$$I_m \triangleq I(X^n; Y^m, A^m).$$

The MMSE sequence is bounded and non-increasing. The MI sequence is non-decreasing with

$$\lim_{m\to\infty} I_m = \begin{cases} H(X^n), & \text{if } P_X \text{ has finite entropy} \\ +\infty, & \text{otherwise.} \end{cases}$$

The first and second order MI difference sequences are defiend as $I'_m = I_{m+1} - I_m$ and $I''_m = I'_{m+1} - I'_m$. The next result follows from the chain rule for mutual information.

Lemma 3. *Under Assumption 1, the first order and second order mutual information difference sequences obey*

$$I'_m = I(X^n; Y_{m+1} \mid Y^m, A^{m+1}) \tag{5}$$
$$I''_m = -I(Y_{m+1}; Y_{m+2} \mid Y^m, A^{m+2}).$$

The finite-length sequences are extended to functions of a continuous parameter $\delta \in \mathbb{R}_+$ using

$$\mathcal{M}_n(\delta) \triangleq M_{\lfloor \delta n \rfloor}, \qquad \mathcal{I}'_n(\delta) \triangleq I'_{\lfloor \delta n \rfloor}$$
$$\mathcal{I}_n(\delta) \triangleq \int_0^\delta \mathcal{I}'_n(\gamma)\mathrm{d}\gamma,$$

By construction, $\mathcal{I}_n(\delta)$ corresponds to the normalized mutual information sequence and obeys $\mathcal{I}_n(m/n) = \frac{1}{n} I_m$.

The following results provide the foundations of our proof and show that the MMSE and MI functions asymptotically satisfy the same constraints as the replica prediction.

Theorem 4. *Under Assumptions 1 and 2, the sequence of MMSE functions $\mathcal{M}_n(\delta)$ obeys*

$$\lim_{n\to\infty}\int_0^T\left|\mathcal{M}_n(\delta) - \mathsf{mmse}_X\left(\frac{\delta}{1+\mathcal{M}_n(\delta)}\right)\right|\mathrm{d}\delta = 0$$

for all $T \in \mathbb{R}_+$.

Theorem 5. *Under Assumptions 1 and 2, the sequences $\mathcal{M}_n(\delta)$ and $\mathcal{I}'_n(\delta)$ obey*

$$\lim_{n\to\infty}\int_0^T\left|\mathcal{I}'_n(\delta) - \frac{1}{2}\log(1+\mathcal{M}_n(\delta))\right|\mathrm{d}\delta = 0$$

for all $T \in \mathbb{R}_+$.

Theorem 6. *Under Assumptions 1 and 2, the sequence of mutual information functions $\mathcal{I}_n(\delta)$ satisfies, for all $\delta > 2$,*

$$|\mathcal{I}_n(\delta) - I_X(\delta)| \le \frac{1}{\delta - 2}.$$

III. Key Steps in the Proof

Due to space constraints, we are only able to provide an overview of some of the main ideas. An extended version of the paper will contain the full details [30]. In the following, we use C to denote an absolute constant and C_B to denote a number that depends on B but is independent of the all other problem parameters. The Euclidean norm is denoted by $\|\cdot\|$.

A. Proof Sketch of Theorem 4

We use a reparameterization argument based on applying an orthogonal transformation to the measurements that zeros out all but one entry in the first column of the measurement matrix; the details of this argument can be found in [31]. To establish the asymptotic convergence, we use an approach that is similar to the one outlined below for Theorem 5.

B. Proof Sketch of Theorem 5

The centered measurement $\bar{Y}_{m+1}$ is defined to be the difference between a new measurement and its conditional expectation given the previous data:

$$\bar{Y}_{m+1} \triangleq Y_{m+1} - \mathbb{E}\big[Y_{m+1} \mid Y^m, A^{m+1}\big].$$

By the linearity of expectation, the centered measurement can also be viewed as a noisy linear projection of the error:

$$\bar{Y}_{m+1} = \langle A_{m+1}, X^n - \mathbb{E}[X^n \mid Y^m, A^m]\rangle + W_{m+1}. \tag{6}$$

In this expression, the measurement vector A_{m+1} and noise W_{m+1} are Gaussian and independent of everything else. Consequently, the distribution of the centered measurement can be expressed as a Gaussian scale mixture with

$$P_{\bar{Y}_{m+1}|X^n,Y^m,A^m} = \mathcal{N}(0, 1+\mathcal{E}_m),$$

where $\mathcal{E}_m = \frac{1}{n}\|X^n - \mathbb{E}[X^n \mid Y^m, A^m]\|^2$ is the squared error. Moreover, the variance satisfies the identity

$$\mathsf{Var}(\bar{Y}_{m+1}) = 1 + M_n. \tag{7}$$

At this point, the key question for our analysis is the extent to which the (random) conditional distribution of $\bar{Y}_{m+1}$ given (Y^m, A^{m+1}) can be approximated by a Gaussian distribution with same mean and variance as $\bar{Y}_{m+1}$. The non-Gaussianness is measured by the expected KL divergence and is defined as

$$\Delta_m \triangleq \mathbb{E}\Big[D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m,A^{m+1}}\,\Big\|\,\mathcal{N}(0, \mathsf{Var}(\bar{Y}_{m+1}))\Big)\Big].$$

Combining (5) and (7) with the fact that the KL divergence with respect to a Gaussian distribution is equal to the difference in differential entropies [32], the non-Gaussianness can be expressed as a function of the MMSE and the MI difference:

$$\Delta_m = \frac{1}{2}\log(1+M_m) - I'_m.$$

This identity shows that the integral relationship between mutual information and MMSE in Theorem 5 can be stated equivalently in terms of the non-Gaussianness of the centered measurements. Furthermore, it also shows that a phase transition occurs if and only if the conditional distribution of new measurements is highly non-Gaussian.

One of the key steps in our proof is the following result, which shows that the non-Gaussianness converges to zero almost everywhere in the limit.

Lemma 7. *Under Assumptions 1 and 2, the non-Gaussianness of the centered measurements satisfies, for all* $\delta \in (0,\infty)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{\lceil \delta n \rceil} \Delta_m = 0$$

The first step in our proof of Lemma 7 is to condition on the *posterior variance* $V_m \triangleq \mathbb{E}[\mathcal{E}_m \mid Y^m, A^m]$, which is the conditional expectation of the squared error. This allows us to decompose the non-Gaussianness as follows:

$$\Delta_m = \mathbb{E}\Big[D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m,A^{m+1}} \,\Big\|\, \mathcal{N}(0, 1+V_m)\Big)\Big] + \frac{1}{2}\mathbb{E}\left[\log\left(\frac{1+M_m}{1+V_m}\right)\right], \tag{8}$$

where the first term on the right measures the non-Gaussianness with respect to a Gaussian approximation that depends on (Y^m, A^m) and the second term measures the deviation of the posterior variance. Our approach to bounding these terms is described in the following subsections.

C. Conditional CLT for new measurements

A necessary condition for first term on the right-hand side of (8) to be small is that the conditional variance of the centered measurement given (Y^m, A^{m+1}) is close to the posterior variance V_m, and thus does not depend strongly on the measurement vector A_{m+1}. In light of (6), this is possible only if the error corresponding to the conditional expectation is nearly isotropic.

In order to bound this term, we use the following two results. The first result shows that certain properties of the error distribution can be bounded in terms of the second order MI difference. The second result shows that these properties are sufficient to bound the non-Gaussianness.

Lemma 8. *Under Assumptions 1 and 2, the posterior variance and the posterior covariance matrix satisfy*

$$\mathbb{E}[|\mathcal{E}_m - V_m|] \le C_B \cdot |I''_m|^{\frac{1}{4}}$$
$$\frac{1}{n}\mathbb{E}[\|\mathsf{Cov}(X^n \mid Y^m, A^m)\|_F] \le C_B \cdot |I''_m|^{\frac{1}{4}}.$$

Lemma 9 (Conditional CLT for Random Projections [33])**.** *Let U be an n-dimensional random vector with mean zero and $\mathbb{E}\big[\|U\|^4\big] \le n^2 B$. Consider the noisy random projection*

$$Y = \langle A, U\rangle + W,$$

where A is an n-dimensional Gaussian vector with mean zero and covariance $n^{-1} I_n$ and $W \sim \mathcal{N}(0,1)$. Then, the expected KL divergence between the random conditional distribution $P_{Y|A}$ and the Gaussian distribution with the same mean and variance as P_Y satisfies

$$\mathbb{E}_A\big[D_{\mathrm{KL}}\big(P_{Y|A} \,\big\|\, \mathcal{N}(0, \mathsf{Var}(Y))\big)\big] \le C_B \cdot \Big|\tfrac{1}{n}\|\mathsf{Cov}(U)\|_F + \tfrac{1}{n}\mathbb{E}\big[\big|\|U\|^2 - \mathbb{E}[\|U\|^2]\big|\big]\Big|^{\frac{4}{9}}.$$

Using Lemmas 8 and 9, it can be shown that

$$\mathbb{E}\Big[D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m,A^{m+1}} \,\Big\|\, \mathcal{N}(0,1+V_m)\Big)\Big] \le C_B \cdot |I''_m|^{\frac{1}{9}}. \tag{9}$$

This inequality shows that the non-Gaussianness of the centered measurement with respect to $\mathcal{N}(0, 1+V_m)$ converges to zero everywhere that the MI sequence is smooth, i.e. everywhere except phase transitions.

D. Concentration of Posterior Variance

The second term on the right-hand side of (8) is nonnegative and measures the deviation of the posterior variance from its expectation. In order to bound this term, we focus on the variance of the *instantaneous mutual information*, which is the random variable defined by

$$\imath(X^n; Y^m \mid A^m) \triangleq \log\left(\frac{f_{Y^m|X^n,Y^m,A^m}(Y^m \mid X^n, Y^m, A^m)}{f_{Y^m|A^m}(Y^m \mid A^m)}\right).$$

Note that the expected value of $\imath(X^n; Y^m \mid A^m)$ is equal to the mutual information I_m.

Lemma 10. *Under Assumptions 1 and 2, the variance of the instantaneous mutual information satisfies*

$$\mathsf{Var}(\imath(X^n; Y^m \mid A^m)) \le C_B \cdot \Big(1 + \frac{m}{n}\Big)^2 n.$$

Using Lemmas 8, 9 and 10 as well as further smoothness properties of the posterior variance, it can eventually be shown that, for all $\delta \in (0,\infty)$,

$$\frac{1}{n}\sum_{m=1}^{\lfloor \delta n \rfloor} \mathbb{E}\left[\log\left(\frac{1+M_m}{1+V_m}\right)\right] \le C_B \cdot (1+\delta^3) n^{-\frac{1}{27}}. \tag{10}$$

At a high level, the proof of this inequality requires relating the expected increase in the instantaneous mutual information associated with a new measurement to a function of the posterior variance. Combining (8), (9) and (10) leads to Lemma 7.

E. Proof Sketch of Theorem 1

The setting for this proof is the set of uniformly-bounded non-increasing functions from $\mathbb{R}_+$ to $\mathbb{R}_+$. It is well-known that these functions are continuous almost everywhere (e.g., except for a countable set of jump discontinuities) [34]. Two such functions are called equivalent if they are equal at all points of continuity. Let $\mathcal{D}$ be the set of equivalence classes induced by this equivalence relation. For negative arguments, we extend $f \in \mathcal{D}$ to $\tilde{f} \colon \mathbb{R} \to \mathbb{R}_+$ using the rule $\tilde{f}(x) = f(0)$ for $x < 0$. The following metric turns $\mathcal{D}$ into a metric space.

Definition 4. Adapting the Lévy metric [35] to $\mathcal{D}$ gives

$$d_L(f,g) \triangleq \inf\Big\{\epsilon > 0 \,\Big|\, g(x+\epsilon) - \epsilon \le f(x) \le g(x-\epsilon) + \epsilon, \forall x \in \mathbb{R}\Big\}.$$

Let $\mathcal{D}_0 \subset \mathcal{D}$ be the subset that also satisfies the upper bound $f(x) \le B/(x-2)$ for $x \ge 3$. One can analyze $\mathcal{D}_0$ based on its close connection to tight families of cumulative distribution functions [35].

Lemma 11. *Consider the sequences $\mathcal{M}_n(\delta)$ and $\mathcal{I}'_n(\delta)$ in the compact metric space $(\mathcal{D}_0, d_L)$. For any limit point of these sequences, there is a $g \in \mathcal{G}$ and a subsequence $n(i)$ such that*

$$\mathcal{M}_{n(i)}(\delta) \overset{d_L}{\to} g(\delta)$$
$$\mathcal{I}'_{n(i)}(\delta) \overset{d_L}{\to} \frac{1}{2}\log(1 + g(\delta)).$$

Sketch of Proof: First, we use Theorem 4 to show that any limit point of any subsequence of $\mathcal{M}_n(\delta)$ must satisfy (3) in Definition 3. Next, we use Theorems 4 and 6 to show that any limit point of any subsequence of $\mathcal{M}_n(\delta)$ must satisfy (4) in Definition 3. Since $\mathcal{G}$ is defined by (3) and (4), it follows that any limit point of any subsequence must lie in $\mathcal{G}$. ■

The proof of Theorem 1 is completed by first applying Lemma 11. Then, the single-crossing property is combined with Lemma 2 to show that the limiting $g \in \mathcal{G}$ must equal the replica prediction almost everywhere. Finally, for a sequence in a compact metric space, if all subsequences have the same limit point, then the sequence must converge to that limit point.

IV. Conclusion

In this paper, we present a rigorous derivation of the fundamental limits of compressed sensing for i.i.d. signal distributions and i.i.d. Gaussian measurement matrices. We show that the limiting MI and MMSE are equal to the values predicted by the replica method from statistical physics. This resolves a well-known open problem.

References

[1] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1983–2010, June 2005.

[2] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[3] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on l_p-norm minimization," *J. Stat. Mech.*, 2009.

[4] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, (Monticello, IL), Oct. 2009.

[5] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, pp. 1902 – 1923, March 2012.

[6] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, pp. 3065–3092, May 2012.

[7] G. Reeves and M. Gastpar, "Compressed sensing phase transitions: Rigorous bounds versus replica predictions," in *Proc. Conf. Inform. Sciences and Systems*, (Princeton, NJ), March 2012.

[8] A. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inform. Theory*, vol. 59, pp. 4243–4271, July 2013.

[9] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, (Chengdu, China), pp. 194–198, Oct. 2006.

[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 18914–18919, Nov. 2009.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, pp. 6920–6941, Oct. 2011.

[12] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, pp. 764–785, February 2011.

[13] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *Int. Symp. Inform. Theory and its Appl.*, (Boston, MA), July 2012.

[14] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, pp. 965–974, 1975.

[15] M. Mézard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[16] R. Muller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Trans. Signal Process.*, vol. 51, pp. 2821–2828, Nov. 2003.

[17] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inform. Theory Workshop*, (Punta del Este, Uruguay), pp. 160–164, 2006.

[18] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, 2010.

[19] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, (Monticello, IL), 2010.

[20] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, May 2012.

[21] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inform. Theory*, vol. 59, pp. 7434–7464, July 2013.

[22] S. Verdú and S. Shamai, "Spectral efficiency of cdma with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, March 1999.

[23] D. N. C. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwith and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, March 1999.

[24] S. B. Korada and N. Macris, "Tight bounds on the capicty of binary input random CDMA systems," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5590–5613, Nov. 2010.

[25] W. Huleihel and N. Merhav, "Asymptotic MMSE analysis under sparse representation modeling," 2015. Available at http://arxiv.org/abs/1312.3417v2.

[26] N. Merhav, D. Guo, and S. Shamai, "Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1400–1416, 2010.

[27] N. Merhav, "Optimum estimation via gradients of partition functions and information measures: A statistical-mechanical perspective," *IEEE Trans. Inform. Theory*, vol. 57, pp. 3887–3898, June 2011.

[28] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1261–1282, April 2005.

[29] P. Milgrom and I. Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, vol. 70, pp. 583–601, March 2002.

[30] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact." to be posted to arXiv, 2016.

[31] W. van den Boom, D. B. Dunson, and G. Reeves, "Quantifying uncertainty in variable selection with arbitrary matrices," in *CAMSAP*, 2015.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, second ed., 2006.

[33] G. Reeves, "Conditional central limit theorems for random projections." to be posted to arXiv, 2016.

[34] W. Rudin, *Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)*. New York, NY: McGraw-Hill, 1976.

[35] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.