# Compressed Sensing Phase Transitions: Rigorous Bounds versus Replica Predictions

Galen Reeves
Department of Statistics
Stanford University

Michael Gastpar[†]
Department of EECS
University of California, Berkeley

***Abstract*—In recent work, two different methods have been used to characterize the fundamental limits of compressed sensing. On the one hand are rigorous bounds based on information-theoretic arguments or the analysis of specific algorithms. On the other hand are exact but heuristic predictions made using the replica method from statistical physics. In this paper, it is shown that, for certain problem settings, these bounds are in agreement, and thus provide a rigorous and accurate characterization of the compressed sensing problem. This characterization shows that the limits of sparse recovery can be quantified succinctly in terms of an effective signal-to-interference-plus-noise ratio, that depends on the number of measurements and the behavior of the sparse components themselves. Connections with the MMSE dimension by Wu and Verdu and minimax behavior of approximate message passing by Donoho et al. are discussed.**

## I. Introduction

The following sparse recovery problem is studied extensively in the developing literature of compressed sensing. Suppose that a vector $\mathbf{x}$ of length $n$ is known to have a small number $k$ of nonzero entries, but the values and locations of the nonzero entries are unknown and must be estimated from a set of $m$ noisy linear projections (or samples) given by the vector

$$\mathbf{y} = A\mathbf{x} + \frac{1}{\sqrt{\mathsf{snr}}}\mathbf{w} \tag{1}$$

where $A$ is a known $m \times n$ measurement matrix and $\mathbf{w} \sim \mathcal{N}(0, I_{m\times m})$ is white Gaussian noise.

Initial results in compressed sensing showed that under certain conditions recovery can be stable (e.g. the mean-squared-error or the fraction of detection errors can be made arbitrarily small by increasing $\mathsf{snr}$) even if the number of measurements $m$ is less than the vector length $n$ [1], [2]. The focus of a great deal of current research is to find a sharp characterization of what can and cannot be recovered as a function of the dimensions $(k, n, m)$, the signal-to-noise ratio $\mathsf{snr}$, and key properties of the nonzero entries in $\mathbf{x}$.

One line of analysis is to derive matching necessary and sufficient conditions for a given recovery task (see e.g. [3]–[7]). Typically, the necessary conditions follow from information theoretic arguments (e.g, Fano's inequality) whereas the sufficient follow from the analysis of specific algorithms (e.g. ML, LASSO, Basis Pursuit). Unfortunately, due in part to the difficulty of analyzing optimal recovery algorithms, there exist gaps between the best known necessary and sufficient conditions for most recovery tasks of interest.

An alternative line of analysis is based on the replica method from statistical physics, which provides a precise characterization in terms of an effective signal-to-interference-plus-noise ratio (SINR) of the problem when $m$ and $n$ are large and the measurement matrix $A$ has i.i.d. entries [6]–[12]. The caveat here, however, is that these results depend on the currently unproven assumption of replica symmetry, and are therefore heuristic.

The purpose of this paper is to use the strengths of both approaches outlined above to improve the understanding of the compressed sensing problem. We do so in the context of sparsity pattern recovery, which is the detection problem of determining which entries in $\mathbf{x}$ are nonzero. We compare rigorous bounds on the *detection error rate* with the precise predictions made using the replica analysis, and show that these results are in close agreement in the high SNR setting.

Our results have a further interpretation that extends beyond sparsity pattern recovery: if it is indeed true that the difficulty of the the compressed sensing problem can be characterized by an effective SINR, then our bounds on the detection error rate imply corresponding bounds on this SINR. These bounds on the SINR, in turn, imply necessary and sufficient conditions for a wide variety of recovery tasks.

### A. Problem Formulation

We assume throughout that the entries of $\mathbf{x}$ are i.i.d. with a distribution $p_X$ that has second moment equal to one and point-mass of weight $(1-\kappa)$ at the origin, i.e.

$$p_X = (1-\kappa)\,\delta_0 + \kappa\, p_{\tilde{X}}.$$

The parameter $\kappa$ corresponds to the fraction of nonzero entries and is referred to as the *sparsity rate*.

We further assume that the entries of $A$ are i.i.d. Gaussian $\mathcal{N}(0, 1/n)$, and we consider the large system limit where the vector length $n$ and the number of measurements $m$ tend to infinity with a constant ratio $\rho = m/n$. We refer to $\rho$ as the *sampling rate*.

The sparsity pattern $S^*$ corresponds to the locations of the nonzero entries in $\mathbf{x}$, i.e.

$$S^* = \{i \in [n] : x_i \neq 0\}.$$

[†]Also with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland.

For any estimate $\hat{S}$, the *missed detection rate* (MDR) is the fraction of entries in $S^*$ omitted from $\hat{S}$ and the *false alarm rate* (FAR) is the fraction of entries in $\hat{S}$ that are not present in $S^*$. A *detection error rate* $D \in [0,1]$ is said to be achievable if there exists a recovery algorithm such that

$$\Pr[\max(\text{MDR}, \text{FAR}) > D] \to 0, \quad \text{as } n \to \infty,$$

where probability is taken with respect to the vector $\mathbf{x}$, the matrix $A$, the noise $\mathbf{w}$, and any additional randomness used by the recovery algorithm. The *distortion function* $D(\rho, \mathsf{snr})$ is the infimum over achievable distortions:

$$D(\rho, \mathsf{snr}) = \inf \left\{ D \geq 0 \,:\, D \text{ is achievable} \right\}.$$

We note that the case $D = 0$ corresponds to exact recovery of the sparsity pattern.

**Notation:** For functions $f(x)$ and $g(x)$ with parameter $x$ tending to $y$, we use the asymptotic notation:

$$\begin{aligned} f(x) \sim g(x) \quad &\Leftrightarrow \quad \lim_{x \to y} f(x)/g(x) = 1 \\ f(x) \gtrsim g(x) \quad &\Leftrightarrow \quad \liminf_{x \to y} f(x)/g(x) \geq 1 \\ f(x) \lesssim g(x) \quad &\Leftrightarrow \quad \limsup_{x \to y} f(x)/g(x) \leq 1. \end{aligned}$$

### B. Overview of Main Results

The central object of interest in the paper is the distortion function $D(\rho, \mathsf{snr})$. As the sampling rate $\rho$ increases (and eventually becomes greater than one), its behavior can be analyzed straightforwardly using classical techniques. The main challenge, therefore, is to understand its behavior in the setting relevant to compressed sensing, where $\rho$ is small and fixed. In this regime, we study the behavior of $D(\rho, \mathsf{snr})$ as $\mathsf{snr}$ becomes large.

Our results are based on the analysis of four different sparsity pattern recovery algorithms, outlined in Table I. Section III, uses the replica analysis to derive an exact, but heuristic, expression for the detection error rate achieved by thresholding the minimum mean-squared-error (MMSE) estimate; Section IV-A uses a recent result of Bayati and Montanari [13] to derive an exact and rigorous result for an iterative algorithm known as approximate message passing algorithm (AMP); Section IV-B analyzes a recent upper bound corresponding to the maximum likelihood (ML) estimate [6]; and Section IV-C analyzes an information theoretic lower bound developed in [7]. The corresponding bounds on the distortion function $D(\rho, \mathsf{snr})$ are illustrated in Figure 1 as a function of the SNR for two different values of the sampling rate.

TABLE I
COMPARISON OF SPARSITY PATTERN RECOVERY RESULTS

| Algorithm | Comp. Efficientt | Characterization | Type |
|---|---|---|---|
| MMSE | No | Exact | Heuristic |
| AMP | Yes | Exact | Rigorous |
| ML | No | Upper Bound | Rigorous |
| Optimal | No | Lower Bound | Rigorous |

The results of this paper rely heavily on an analogy between the vector recovery problem outlined in (1) and a related scalar recovery problem of the form:

$$Y = X + \frac{1}{\sqrt{s}} W \tag{2}$$

where $X \sim p_X$ and $W \sim \mathcal{N}(0,1)$ are independent and $s$ is an effective signal-to-noise ratio. The analogue of sparsity pattern detection in the scalar setting is to determine whether $X$ is nonzero.

Accordingly, we let $S^* = \mathbf{1}(X \neq 0)$ and define the scalar distortion function

$$D_{\text{scalar}}(s) = \min_T \max \left( \frac{\Pr[\text{MD}]}{\kappa}, \frac{\Pr[\text{FA}]}{1-\kappa} \right) \tag{3}$$

where $\Pr[\text{MD}]$ and $\Pr[\text{FA}]$ denote the missed detection and false alarm probabilities and the minimization is over all hypothesis tests of the form $\hat{S} = \mathbf{1}(|Y| \geq T)$.

The following result consolidates the main findings of this paper. For simplicity, this result is stated in terms the Bernoulli-Gaussian distribution (i.e. a distribution $p_X$ whose nonzero part is $\mathcal{N}(0, 1/\kappa)$); generalizations to other distributions follow directly from results given in the following sections.

**Theorem 1.** *Suppose that $p_X$ is Bernoulli-Gaussian with sparsity rate $\kappa$ and fix any sampling rate $\rho > \kappa$. As $\mathsf{snr} \to \infty$, the scaling behavior of the distortion function $D(\rho, \mathsf{snr})$ can be characterized in terms of the scalar distortion function $D_{scalar}(\cdot)$ as follows:*

*(a)* **MMSE Replica Prediction:** *Assume that the replica analysis assumptions outlined in Section III hold. Then,*

$$D(\rho, \mathsf{snr}) \lesssim D_{scalar}\big((\rho - \kappa) \cdot \mathsf{snr}\big). \tag{4}$$

*(b)* **AMP-MMSE Bound:** *There exists a* stability coefficient $\varrho^{(AMP\text{-}MMSE)} \in (\kappa, 1)$ *such that if $\rho > \varrho^{(AMP\text{-}MMSE)}$, then*

$$D(\rho, \mathsf{snr}) \lesssim D_{scalar}\big((\rho - \kappa) \cdot \mathsf{snr}\big). \tag{5}$$

*(c)* **ML Upper Bound:**

$$D(\rho, \mathsf{snr}) \lesssim D_{scalar}\big(0.028 \cdot (\rho - \kappa) \cdot \mathsf{snr}\big). \tag{6}$$

*(d)* **Lower Bound:**

$$D(\rho, \mathsf{snr}) \gtrsim D_{scalar}\big(2.25 \cdot \rho \cdot \mathsf{snr}\big). \tag{7}$$

Theorem 1 corresponds to the high SNR behavior of the bounds in Figure 1. In the left panel, the sampling rate $\rho = 5 \times 10^{-4}$ is greater than the sparsity rate $\kappa = 10^{-4}$, but less than the stability coefficient $\varrho^{(\text{AMP-MMSE})} \approx 5.11 \times 10^{-4}$ needed for the AMP-MMSE bound. In the right panel, the sampling rate $\rho = 6 \times 10^{-4}$ is larger than $\varrho^{(\text{AMP-MMSE})}$ and the AMP-MMSE and MMSE bounds are in close agreement.

Observe that the scaling factors preceding the term $(\rho - \kappa) \cdot \mathsf{snr}$ in Theorem 1 correspond to the relative shifts (along the $x$-axis) between the bounds in Figure 1. For example, the factor $0.028$ in part (c) corresponds to a shift of approximately 15.4 dB, whereas the factor $2.25 \cdot \rho / (\rho - \kappa)$ in part (d) corresponds to a shift of approximately $4.5$ dB with $\rho/\kappa = 5$ and 4.3 dB with $\rho/\kappa = 6$.

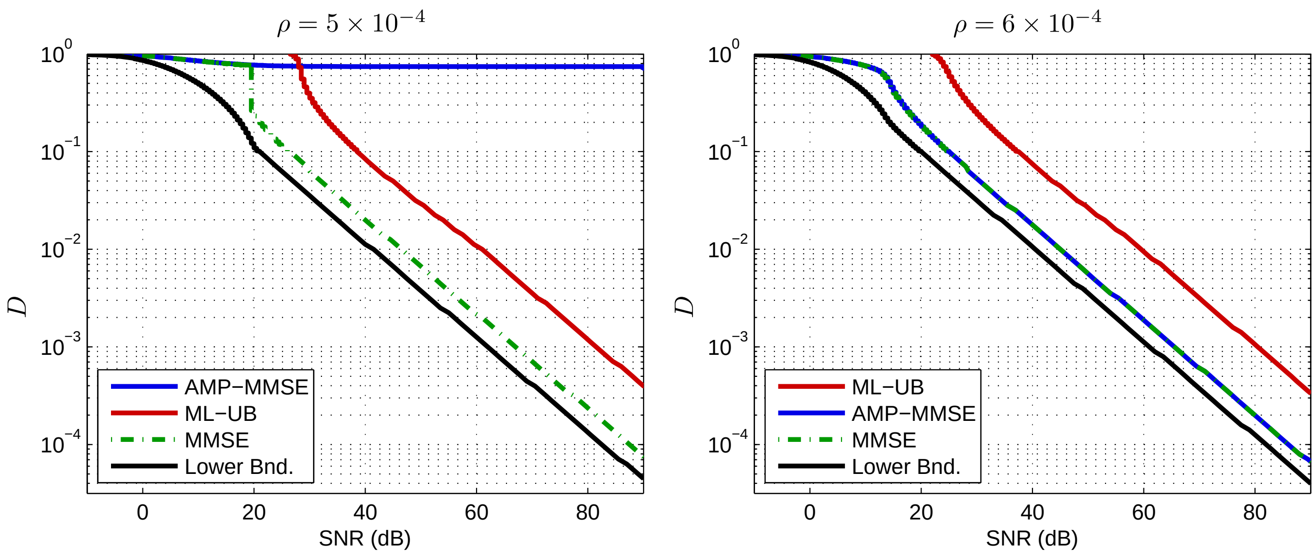


Fig. 1. Characterization of the detection error rate $D$ as a function of the SNR for two different sampling rates $\rho = m/n$ when the entries in $\mathbf{x}$ are Bernoulli-Gaussian with sparsity rate $\kappa = 10^{-4}$. In this setting, the MMSE stability coefficient is equal to the sparsity rate (i.e. $\varrho^{(\text{MMSE})} = \kappa$) and the AMP-MMSE stability coefficient $\varrho^{(\text{AMP-MMSE})}$ is approximately $5.11 \times 10^{-4}$.

## II. The Scalar Setting

This section studies the behavior of the scalar detection problem.

A probability distribution $p_X$ is said to have *polynomial decay rate* $L \in (0, \infty)$ if

$$\Pr[|X| \le x | X \neq 0] \sim \tau\, x^L, \quad x \to 0 \tag{8}$$

for some limiting constant $\tau \in (0, \infty)$. The decay rate measures the relative probability that $X$ is close to the origin. Note that this probability decreases with $L$.

The following is a new result which shows that $L$ and $\tau$ provide an explicit characterization of the high SNR behavior of the scalar detection error rate. Its proof follows from the analysis of the inverse function $s(D)$ studied in [6].

**Proposition 1.** *If $p_X$ has polynomial decay rate $L$ and limiting constant $\tau$, then*

$$D_{\textit{scalar}}(s) \sim \tau \sqrt{\frac{L \log(s)}{s^L}}, \quad s \to \infty. \tag{9}$$

The Bernoulli-Gaussian distribution is an example of a distribution with polynomial decay rate $L = 1$ and limiting constant $\tau = \sqrt{2\kappa/\pi}$. By Proposition 1, the limiting behavior of its scalar distortion function $D_{\text{scalar}}(s)$ is given by

$$D_{\text{scalar}}(s) \sim \sqrt{\frac{2\kappa \log(s)}{\pi\, s}}, \quad s \to \infty. \tag{10}$$

## III. Replica Predictions

This section analyzes the behavior of a sparsity pattern recovery algorithm based on thresholding the MMSE estimator. The analysis is based on the powerful but heuristic *replica method* from statistical physics which has been applied to the vector estimation problem studied in this paper by a series of recent papers [8]–[12].

A detailed explanation of the replica analysis is beyond the scope of this paper. The assumptions needed for our results are summarized below.

**Replica Analysis Assumptions:** The key assumptions underlying the replica analysis are stated explicitly by Guo and Verdú in [9]. A concise summary can also be found in [11, Appendix A]. Two assumptions that are used—and generally accepted throughout the literature—are the validity the "replica trick" and the self averaging property of a certain function defined on the random matrix $\mathbf{A}$. A further assumption that is also required is that of *replica symmetry*. This last assumption is problematic, however, since it is known that there are cases where it does not hold, and there is currently no test to determine whether or not it holds in the setting of this paper.

The following result is due to Guo and Verdu [9].

**Proposition 2** (MMSE [9])**.** *Assume that the replica analysis assumptions hold. For each problem of size $n$, let $\nu_n$ be the (random) empirical probability measure on $\mathbb{R}^2$ that places mass $1/n$ at each point $\{(X_i, \mathbb{E}[X_i|\mathbf{Y}])\}_{i\in[n]}$. Then, $\nu_n$ converges weakly, in probability, to the distribution of the pair $(X, \mathbb{E}[X|X + s^{-1/2}W])$ where $s$ is given by*

$$s = \arg\min_{u \ge 0} \Gamma(u) \tag{11}$$

*with*

$$\Gamma(u) = u/\mathsf{snr} + 2I(X; X + u^{-1/2}W) - \rho \log(u). \tag{12}$$

Roughly speaking, Proposition 2 says that the vector problem (1) is equivalent to $n$ copies of the scalar problem (2) where the effective signal-to-noise ratio $s$ is expressed precisely as a function of the tuple $(\rho, \mathsf{snr}, p_X)$.

An important implication of this result is that the impact of the parameters $\rho$ and $\mathsf{snr}$ are quantified entirely by a single parameter $s$. In other words, all pairs $(\rho, \mathsf{snr})$ with the same value of $s$ yield exactly the same performance. We refer to $s$ as the effective signal-to-interference-plus-noise ratio (SINR), to emphasize the fact that it is due to the original noise in the problem as well as the interference caused by the nonzero entries.

### A. MMSE + Thresholding

We now consider implications of Proposition 2 for the problem of sparsity pattern recovery. Since the scalar function $\varphi(y) = \mathbb{E}[X | X + s^{-1/2} W = y]$ is one-to-one, we can apply the functional inverse $\varphi^{-1}(\cdot)$ componentwise to the vector estimate $\mathbb{E}[\mathbf{x}|\mathbf{y}]$ to obtain an $n$-dimensional vector $\mathbf{z}$. By Proposition 2, it then follows that the empirical distribution of the entries $\{(X_i, Z_i)\}_{i \in [n]}$ converges weakly, in probability, to the distribution on $(X, X + s^{-1/2} W)$ where $s$ is given by (11).

At this point, one way to estimate the sparsity pattern $S^*$ is to threshold the magnitudes of $\mathbf{z}$ at a level $T$, i.e.

$$\hat{S} = \{ i \in [n] \, : \, |Z_i| \geq T \}.$$

We will assume that the threshold $T$ is chosen to minimize the detection error rate. Note that this can be achieved in practice by choosing a value of $T$ yielding an estimate of size $k$.

The following result is a direct consequence of Proposition 2 and the definition of the scalar distortion function.

**Proposition 3.** *Assume that the replica analysis assumptions hold. Then, the detection error rate of MMSE + Thresholding is given by*

$$D^{(MMSE)}(\rho, \mathsf{snr}) = D_{scalar}(s) \tag{13}$$

*where* $D_{scalar}(\cdot)$ *is given in* (3) *and* $s$ *is given in* (11).

Proposition 3 shows that sparsity pattern recovery can be characterized precisely by first finding the equivalent scalar SINR, and then plugging it into the scalar distortion function.

### B. High SNR Predictions

We now let $\rho$ be fixed and study the behavior of $s$ (and hence also $D(\rho, \mathsf{snr})$ as $\mathsf{snr}$ becomes large.

Since the function $\Gamma(u)$ is differentiable and grows without bound as either $u \to \infty$ or $u \to 0$, any possible value of $s$ belongs to the set $\{u : \Gamma'(u) = 0\}$. Using the fact [14]:

$$\frac{d}{ds} 2 I(X; X + u^{-1/2} W) = \mathsf{mmse}(u), \tag{14}$$

where $\mathsf{mmse}(s)$ denotes the minimum means-squared error:

$$\mathsf{mmse}(u) = \mathbb{E}[(X - \mathbb{E}[X | X + u^{-1/2} W])^2], \tag{15}$$

it follows that $\Gamma'(u) = 0$ is equivalent to

$$u = (\rho - u \, \mathsf{mmse}(u)) \, \mathsf{snr}. \tag{16}$$

By the optimality of the MMSE estimate (with respect to mean squared error), $s$ is nondecreasing in $\mathsf{snr}$. Therefore, as $\mathsf{snr}$ becomes large, there are only two possibilities: the first is that $s \to 0$ (and hence the distortion becomes arbitrarily small) and the second is that $s$ is bounded away from zero. By analyzing the behavior of these fixed points (see [6, Appendix C-B]), it can be shown that the first possibility occurs if and only if $\rho > \varrho^{(\mathrm{MMSE})}$ where

$$\varrho^{(\mathrm{MMSE})} = \lim_{u \to 0} u \, \mathsf{mmse}(u). \tag{17}$$

Since $\varrho^{(\mathrm{MMSE})}$ is the infimum over all sampling rates $\rho$ for which recovery can be made arbitrarily good by increasing the SNR we refer to it as the MMSE *stability coefficient*.

We remark that $\varrho^{(\mathrm{MMSE})}$ corresponds to the *MMSE Dimension* of $p_X$ [15] and its role in the context sparse recovery has been investigated in recent work [16].

Returning to the expression in (16), it follows that if $\rho > \varrho^{(\mathrm{MMSE})}$, then

$$s \sim (\rho - \varrho^{(\mathrm{MMSE})}) \, \mathsf{snr}, \quad \mathsf{snr} \to \infty. \tag{18}$$

Combining Proposition 3 and (18) leads to the following result.

**Proposition 4.** *Assume the replica analysis assumptions hold. If* $\rho > \varrho^{(MMSE)}$*, then*

$$D^{(MMSE)}(\rho, \mathsf{snr}) \sim D_{scalar}\big((\rho - \varrho^{(MMSE)}) \cdot \mathsf{snr}\big), \tag{19}$$

*as* $\mathsf{snr} \to \infty$.

We remark that the dependence of the SINR on the distribution $p_X$ is encapsulated entirely by the stability coefficient $\varrho^{(\mathrm{MMSE})}$. By contrast, the behavior of $D_{\mathrm{scalar}}$ depends critically on the behavior of $p_X$ around the origin.

## IV. Rigorous Bounds

The analysis in the previous section yielded a sharp, but heuristic, characterization of the sparsity pattern recovery problem. This section provides the other side of the story and presents rigorous bounds.

### A. Approximate Message Passing + Thresholding

We begin with an algorithm known as *approximate message passing* (AMP) [17]. The general AMP algorithm is characterized by a sequence of scalar de-noising functions $\{\eta_t(\cdot)\}_{t \geq 0}$. For the purposes of this paper, we consider the special case where this sequence is given by

$$\eta_t(y) = \mathbb{E}[X | X + u_t^{-1/2} W = y], \tag{20}$$

with $u_0 = \rho / (1 + 1/\mathsf{snr})$ and

$$u_t = \frac{\rho}{\mathsf{snr}^{-1} + \mathsf{mmse}(u_{t-1})}. \tag{21}$$

We refer to this version of AMP as AMP-MMSE. The output of the algorithm is an estimate $\hat{\mathbf{x}}^{(\text{AMP-MMSE})}$ of the unknown vector $\mathbf{x}$.

The recent work of Donoho et al. [17] and Bayati and Montanari [13] provides a precise characterization of the behavior of AMP, similar in nature to the characterization of the MMSE based on the replica analysis.

**Proposition 5** (AMP-MMSE [13])**.** *For each problem of size* $n$*, let* $\nu_n$ *be the (random) empirical probability measure on* $\mathbb{R}^2$

*that places mass $1/n$ at each point $\{(X_i, X_i^{(AMP\text{-}MMSE)})\}_{i\in[n]}$. Then, $\nu_n$ converges weakly, almost surely, to the distribution of the pair $(X, \mathbb{E}[X|X + s^{-1/2}W])$ where $s$ is given by*

$$s = \arg\min_{u\geq 0}\{u : \Gamma'(u) = 0\} \tag{22}$$

*and $\Gamma(u)$ is given by (12).*

There are two important differences between Propositions 2 and 5. First, the analysis of the AMP algorithm does not require any unproven assumptions and thus provides a rigorous upper bound on the performance of the optimal recovery algorithm.

Second, the SINR $s$ achieved by AMP-MMSE corresponds to the worst-case (i.e. smallest) solution $u$ to the fixed point equation (16). By contrast, the MMSE scalar SNR corresponds to the fixed point minimizing the functional $\Gamma(s)$. Under the replica analysis assumptions, this implies that, for each distribution $p_X$, the $(\rho, \mathsf{snr})$ plane can be divided into two regions: in one region, AMP-MMSE is equivalent to MMSE and in the other region AMP-MMSE is strictly worse than MMSE.

Using similar arguments as in Section III, we may consider sparsity pattern recovery by first applying the inverse mapping $\varphi^{-1}(\cdot)$ to the output $\mathbf{x}^{(\text{AMP-MMSE})}$, and then thresholding based on the magnitudes of the entries. This approach leads to the following result (see [6] for full details).

**Proposition 6.** *The detection error rate of AMP-MMSE + Thresholding is given by*

$$D^{(AMP\text{-}MMSE)}(\rho, \mathsf{snr}) = D_{scalar}(s) \tag{23}$$

*where $D_{scalar}(\cdot)$ is given in (3) and $s$ is given in (22).*

With these results in hand, we may now consider the high SNR behavior of $D^{(\text{MMSE})}(\rho, \mathsf{snr})$. Using the analysis given in [6, Section IV], it can be shown that the behavior of $s$ for AMP-MMSE is similar as for MMSE except that the stability coefficient is given by

$$\varrho^{(\text{MMSE-AMP})} = \sup_{u>0} u\, \mathsf{mmse}(u). \tag{24}$$

This behavior is stated explicitly as follows.

**Proposition 7.** *If $\rho > \varrho^{(MMSE\text{-}AMP)}$, then*

$$D^{(AMP\text{-}MMSE)}(\rho, \mathsf{snr}) \sim D_{scalar}\big((\rho - \varrho^{(MMSE)}) \cdot \mathsf{snr}\big) \tag{25}$$

*as $\mathsf{snr} \to \infty$.*

### *B. Maximum Likelihood Upper Bound*

We next consider the performance of the maximum likelihood estimate which is given by

$$\hat{\mathbf{x}}^{(\text{ML})} = \arg\min_{\tilde{\mathbf{x}}\,:\,\tilde{x}\text{ is }k\text{-sparse}} \|\mathbf{y} - A\tilde{\mathbf{x}}\|. \tag{26}$$

In [6, Theorem 1], it is shown that a detection error rate $D$ is achievable using the ML estimate if

$$\rho > \kappa + \max_{\delta\in[D,1]} \Lambda^{(\text{ML-UB})}(\delta; \mathsf{snr}, p_X) \tag{27}$$

where the function $\Lambda^{(\text{ML-UB})}(\delta; \mathsf{snr}, p_X)$ is finite for all $\delta > 0$ and tends to infinity as $\delta \to 0$. Here, it is interesting to note that the maximization over $\delta$ is reminiscent of the minimization of the functional $\Gamma(u)$ in Proposition 2.

Combining the ML upper bound (27) with the analysis in [6, Section IV], leads to the following result.

**Proposition 8.** *If $p_X$ has polynomial decay rate $L$ then, for any $\rho > \kappa$,*

$$D^{(ML)}(\rho, \mathsf{snr}) \lesssim D_{scalar}\big(C_L \cdot (\rho - \kappa)\, \mathsf{snr}\big), \tag{28}$$

*as $\mathsf{snr} \to \infty$, where $C_L \in (0, 1)$ is given by*

$$C_L = \frac{3 - \sqrt{8}}{2(1 + 2/L)}. \tag{29}$$

The main advantage of Proposition 8 is that it applies generally for all sampling rates $\rho > \kappa$. The main drawback, however, is the constant $C_L$ which makes the bound weaker than the AMP-MMSE bound in cases where $\rho$ is much larger than $\varrho^{(\text{AMP-MMSE})}$. We note that the constant $C_1$ corresponds to the 15.43 dB shift (along the $x$-axis) between the ML-UB and MMSE bounds in Figure 1.

### *C. Information-Theoretic Lower Bound*

Lastly, we consider an information-theoretic lower bound which applies generally to any possible recovery algorithm.

**Proposition 9** ( Lower Bound [7, Corollary 3.2]). *A distortion $D$ is not achievable for any recovery algorithm if*

$$\rho < \max_{\delta\in[D,1]} \frac{2(1 - \kappa + \kappa\delta) R\big(D/\delta, \frac{\kappa\delta}{1-\kappa+\kappa\delta}\big)}{\log(1 + P(\delta)\, \mathsf{snr})} \tag{30}$$

*where $R(D, p) = H_b(p) - pH_b(D) - (1-p)H_b\big(\frac{p}{1-p}D\big)$, $H(p)$ is the binary entropy function, and*

$$P(\delta) = \int_0^\infty \max\Big(\Pr[X^2 > u] - (1 - \delta)\kappa, 0\Big) du. \tag{31}$$

Analyzing the behavior of the right hand side of (30) evaluated with $\delta = (1 + 2/L)D$ leads to the following lower bound.

**Proposition 10.** *If $p_X$ has polynomial decay rate $L$ then*

$$D(\rho, \mathsf{snr}) \gtrsim D_{scalar}(\tilde{C}_L \cdot \rho \cdot \mathsf{snr})$$

*for any recovery algorithm where $\tilde{C}_L \in (1, e)$ is given by*

$$\tilde{C}_L = \big(1 + L/2\big)^{2/L}. \tag{32}$$

Note that if $L$ is large and $\rho$ is large relative to $\varrho^{(\text{MMSE})}$, then this result is in close agreement with the MMSE behavior in Proposition 4.

The main weakness of the bound, however, is that $\mathsf{snr}$ is proportional to $\rho$ instead of $\rho - \varrho^{(\text{MMSE})}$. Note that in Figure 1, the constant $\tilde{C}_1 = 9/4$ corresponds to a 3.5 dB shift (along the $x$-axis) between the LB and MMSE bounds. The remainder of the gap is due to the ratio $\rho/(\rho - \kappa)$.

## V. Discussion and Future Work

This paper considers the fundamental limits of sparsity pattern recovery in compressed sensing. In particular, we provide a simplified characterization of the high SNR behavior of the detection error rate. Our bounds are presented in terms of a related scalar detection problem with an effective SINR that depends on the sampling rate (i.e. the ratio of measurements to the length of the vector), the SNR, and the distribution of the nonzero entries.

One contribution of our analysis is that the detection error rate can be assessed in two steps: first compute the SINR of the compressed sensing problem, and then analyze the corresponding scalar detection problem, evaluated with this SINR. This perspective is particularly useful since the SINR and the performance of the scalar detection problem depend on different aspects of the underlying distribution.

A further contribution of our analysis is that bounds on the detection error rate imply bounds on the SINR. According to the heuristic replica analysis, the high SNR behavior of the SINR has the simple expression:

$$\text{SINR} \sim \left(\rho - \varrho^{(\text{MMSE})}\right) \cdot \mathsf{snr}.$$

The rigorous upper and lower bounds given in parts (b)-(d) of Theorem 1 (or more generally in Propositions 7, 8, and 10) correspond to lower and upper bounds on this SINR, respectively.

The significance of the SINR is characterized by the convergence in distribution described in Proposition 2. It is important to emphasize that this convergence in distribution is proved rigorously for the special case of AMP in Proposition 5, but is currently unproven for the MMSE and ML estimates. Our bounds on the SINR for these algorithms are derived, instead, from our rigorous analysis of sparsity pattern recovery.

One question for further research is whether the MMSE + Thresholding algorithm discussed in Section III achieves the optimal distortion, i.e. if $D^{(\text{MMSE})}(\rho, \mathsf{snr}) = D(\rho, \mathsf{snr})$. Another question of interest is if it is possible to improve the lower bound given in Proposition 10 so that the corresponding upper bound on the SINR is proportional to $\rho - \rho^{(\text{MMSE})}$.

## Acknowledgment

This work was supported in part by ARO MURI No. W911NF-06-1-0076, by EPFL fund F-1218 (where the first author was a postdoctoral scholar in 2011), and an NSF VIGRE Postdoctoral Fellowship.

## References

[1] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. on Pure and Applied Math.*, vol. 59, pp. 1207–1223, Feb. 2006.

[2] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory.*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[3] M. J. Wainwright, "Information-theoretic limitations on sparsity recovery in the high-dimensional and noisy setting." *IEEE Trans. Inf. Theory*, vol. 55, pp. 5728–5741, Dec. 2009.

[4] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[5] M. Akcakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[6] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity tbe samplig rate-distortion tradeoff for spasity pattern recovery in compressed sensing," *To appear in IEEE Trans. Inf. Theory*.

[7] G. Reeves, "Sparsity pattern recovery in compressed sensing," Ph.D. dissertation, University of California, Berkeley, 2011.

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

[9] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[10] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit of compressed sensing based on lp norm minimization," *J. Stat. Mech.*, 2009.

[11] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," in *Proc. Neural Information Processing Systems Conf.*, vol. 22, Vancouver, CA, Dec. 2009, pp. 1545–1553.

[12] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sep. 2009.

[13] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

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

[15] Y. Wu and S. Verdú, "Renyi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.

[16] Y. Wu and S. Verdu, "Optimal phase transitions in compressed sensing," Nov. 2011, arXiv:1111.6822v1.

[17] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," in *Proc. Nat. Acad. Sci.*, vol. 106, 2009, pp. 18 914–18 919.