

The Minimax Noise Sensitivity in Compressed Sensing

Galen Reeves and David Donoho

Department of Statistics
Stanford University

Abstract—Consider the compressed sensing problem of estimating an unknown k -sparse n -vector from a set of m noisy linear equations. Recent work focused on the noise sensitivity of particular algorithms – the scaling of the reconstruction error with added noise. In this paper, we study the *minimax* noise sensitivity – the minimum is over all possible recovery algorithms and the maximum is over all vectors obeying a sparsity constraint. This fundamental quantity characterizes the difficulty of recovery when nothing is known about the vector other than the fact that it has at most k nonzero entries. Assuming random sensing matrices (i.i.d. Gaussian), we obtain non-asymptotic bounds which show that the minimax noise sensitivity is finite if $m \geq k + 3$ and infinite if $m \leq k + 1$.

We also study the large system behavior where $\delta = m/n \in (0, 1)$ denotes the undersampling fraction and $k/n = \varepsilon \in (0, 1)$ denotes the sparsity fraction. There is a phase transition separating successful and unsuccessful recovery: the minimax noise sensitivity is bounded for any $\delta > \varepsilon$ and is unbounded for any $\delta < \varepsilon$. One consequence of our results is that the Bayes optimal phase transitions of Wu and Verdú can be obtained uniformly over the class of all sparse vectors.

I. INTRODUCTION

The following estimation problem has been studied extensively in the compressed sensing literature. Consider the noisy underdetermined system of linear equations:

$$\mathbf{y} = A\mathbf{x} + \mathbf{w} \quad (1)$$

where matrix A is $m \times n$, $m < n$, the n -vector \mathbf{x} is k -sparse (i.e. has at most k nonzero entries) and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 I)$ is Gaussian white noise. Both \mathbf{y} and A are known, both \mathbf{x} and \mathbf{w} are unknown, and we seek an approximation to \mathbf{x} .

Recovery is often said to be *stable* [1], [2] with respect to noise if the mean-squared error (MSE) of an estimate $\hat{\mathbf{x}}$ is proportional to the noise power, i.e.

$$\mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|^2] \leq c \cdot \sigma^2.$$

The best possible constant in this inequality – the *noise sensitivity* [3] – is important in both theory and practice since it characterizes the degree to which noise degrades the quality of estimation.

In recent work, the noise sensitivity has been studied using two different approaches: one line of research has focused on the worst-case behavior over fixed (and suboptimal) classes of estimators [3]–[6], while another line of research analyzed the expected behavior for optimal Bayes estimators when \mathbf{x} is a random vector with a known distribution [7]–[9].

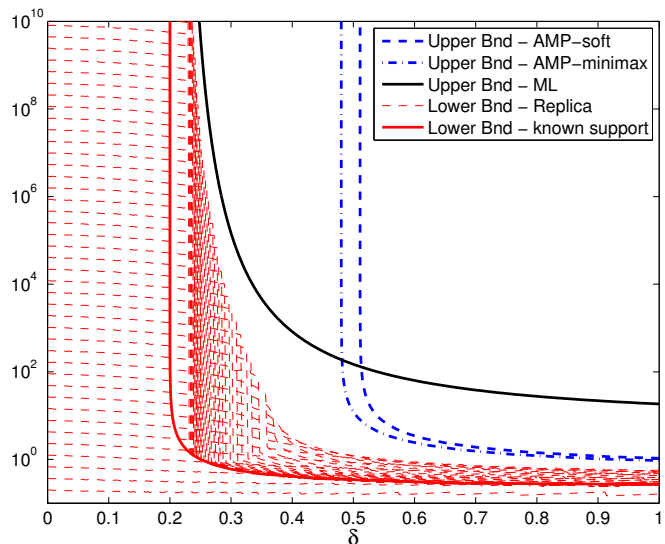


Fig. 1. Bounds on the asymptotic minimax noise sensitivity $M^*(\delta, \varepsilon)$ as a function of δ for $\varepsilon = 0.2$. The ML upper bound is given in Theorem 3; the lower bound for known support is given in Corollary 5; the replica lower bounds are given by Proposition 6 for various distribution; the AMP-soft upper bound corresponds to AMP with soft thresholding [3]; and the AMP-minimax upper bound corresponds to AMP with the minimax scalar denoiser [6].

By contrast this paper takes an approach inspired by statistical decision theory [10] in which we seek to understand the noise sensitivity of the optimal estimator when: *nothing is known about \mathbf{x} other than the fact that it is k -sparse*. We analyze the problem from a minimax perspective where the minimum is over all possible recovery algorithms and the maximum is over all possible vectors of a given sparsity.

We obtain non-asymptotic bounds on minimax approximation error as a function of the tuple (k, m, n) in the setting where A is a random matrix with i.i.d. Gaussian entries. Our bounds can be computed numerically and show that the expected minimax noise sensitivity is finite if $m \geq k + 3$ and infinite if $m \leq k + 1$.

We also obtain information about the behavior of the minimax noise sensitivity by large system analysis; let k, m, n each tend to infinity so that $m/n \rightarrow \delta$ and $k/n \rightarrow \varepsilon$; thus we have a phase space $0 \leq \delta, \varepsilon \leq 1$, expressing different combinations of undersampling δ and sparsity ε . For each choice of asymptotic sparsity/undersampling, there is an asymptotic minimax noise sensitivity $M^*(\delta, \varepsilon)$ (defined below); we develop bounds on the quantity, illustrated in Figure 1 above.

A. Problem Formulation

We focus on the compressed sensing model (1) under the additional assumption that A is a random matrix with i.i.d. Gaussian entries $A_{i,j} \sim \mathcal{N}(0, 1)$. The quality of an estimator $g : \mathbb{R}^{m \times n} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ at noise level σ^2 is measured by the mean-squared error:

$$\text{mse}(\mathbf{x}, g, \sigma^2) = \mathbb{E} \left[\frac{1}{n} \|g(A, A\mathbf{x} + \mathbf{w}) - \mathbf{x}\|^2 \right],$$

where expectation is taken with respect to the matrix A and the noise \mathbf{w} .

The noise sensitivity is given by the ratio of the MSE to the noise power and the *minimax noise sensitivity* is defined formally by:

$$M_n^*(m, k) = \inf_g \sup_{\mathbf{x} \in B^n(k)} \sup_{\sigma^2 > 0} \frac{\text{mse}(\mathbf{x}, g, \sigma^2)}{\sigma^2}, \quad (2)$$

where the infimum is over all measurable functions $g(\cdot)$ and $B^n(k)$ denotes the set of all n -dimensional vectors with at most k nonzero entries.

For any sampling rate $\delta \in (0, 1)$ and sparsity rate $\varepsilon \in (0, 1)$, the asymptotic minimax mean squared error is given by

$$M^*(\delta, \varepsilon) = \limsup_{n \rightarrow \infty} M_n^*([\delta \cdot n], [\varepsilon \cdot n]). \quad (3)$$

B. Relation to Previous Work

There is now a massive literature in compressed sensing providing a wide variety of performance guarantees for various algorithms and model assumptions. In the following, we discuss recent results which study the precise behavior of the noise sensitivity in the compressed sensing model considered in this paper.

In [3]–[5] the authors study a greedy algorithm known as approximate message passing (AMP) using a soft thresholding scalar denoiser. The behavior of AMP using the optimal (i.e. minimax) denoiser is characterized in [6]. In both cases, it is shown that there is a phase transition curve $\delta = \delta^*(\varepsilon)$: the noise sensitivity of AMP is bounded for any $\delta > \delta^*(\varepsilon)$ and is unbounded for any $\delta < \delta^*(\varepsilon)$.

In [9], Wu and Verdu consider the performance of optimal estimators in the setting where \mathbf{x} is a random variable with i.i.d. entries. They show that the fundamental phase transition for these algorithms is given by the curve $\delta = \varepsilon$ (which lies well below the phase transition $\delta^*(\varepsilon)$ for AMP). Moreover, they give an upper bound on the noise sensitivity for all $\delta > \varepsilon$. Their proof relies crucially on Kirschbraun's extension theorem for Lipschitz functions and does not describe an explicit algorithm.

In comparison to the previous work, this paper shows that the optimal phase transition $\delta = \varepsilon$ of Wu and Verdu is achieved uniformly over the class of all sparse vectors by the maximum likelihood estimator, and we give an explicit upper bound on its noise sensitivity.

II. UPPER BOUNDS

This section studies the behavior of the maximum likelihood (ML) estimator and derives upper bounds on the minimax noise sensitivity.

Under the assumed white Gaussian noise, the ML estimate $\hat{\mathbf{x}}^{\text{ML}}$ is the solution of the minimization problem

$$\min_{\mathbf{u} \in \mathbb{R}^n} \|\mathbf{y} - A\mathbf{u}\|^2 \quad \text{subject to} \quad \mathbf{u} \in B^n(k). \quad (4)$$

Whenever $m > k$ our assumptions make the minimizer of (4) unique almost surely.

Our first result bounds the probability that the ML reconstruction error is large relative to the noise; the proof is given in Section IV.

Theorem 1. *For any vector $\mathbf{x} \in B^n(k)$, noise power $\sigma^2 > 0$, and $m > k$, the maximum likelihood estimator obeys*

$$\mathbb{P} \left[\frac{\|\hat{\mathbf{x}}^{\text{ML}} - \mathbf{x}\|^2}{n \sigma^2} > \mathcal{T}_{k,m,n} \cdot (1 + u) - \frac{m}{n} \right] \leq 2e^{-(m-k)\Delta(u)}, \quad (5)$$

for all $u \geq 0$, with

$$\mathcal{T}_{k,m,n} = \frac{m(m+1)}{n(m-k+1)} \left[\mathcal{L}^{-1} \left(\frac{n}{m-k} H \left(\frac{k}{n} \right) \right) \right]^2, \quad (6)$$

and

$$\Delta(u) = \max_{0 \leq s \leq u} \min \left\{ \mathcal{L} \left(\frac{\sqrt{1+u}}{\sqrt{1+s}} \right), \frac{(\sqrt{1+2s}-1)^2}{4} \right\}, \quad (7)$$

where $\mathcal{L}(x) = \log(x) + 1/x - 1$, $\mathcal{L}^{-1}(\cdot)$ denotes the functional inverse of $\mathcal{L}(\cdot)$ restricted to the domain $[1, \infty)$, and $H(p) = -p \log_2(p) - (1-p) \log_2(1-p)$ is the binary entropy function.

A key contribution of Theorem 1 is the that the upper bound depends only on the integers (k, m, n) and is independent of the vector \mathbf{x} and the noise power σ^2 . The next result uses this fact to derive an upper bound on the minimax noise sensitivity.

Theorem 2. *For any $m > k + 2$, the minimax noise sensitivity $M_n^*(m, k)$ is finite and obeys*

$$\begin{aligned} M_n^*(m, k) &\leq \sup_{\mathbf{x} \in B^n(k)} \sup_{\sigma^2 > 0} \frac{\text{mse}(\mathbf{x}, g^{\text{ML}}, \sigma^2)}{\sigma^2} \\ &\leq \mathcal{T}_{k,m,n} \cdot \left(1 + \int_0^\infty e^{-(m-k)\Delta(t)} dt \right) - \frac{m}{n}. \end{aligned} \quad (8)$$

Proof: The first inequality follows directly from the definition of $M_n^*(m, k)$. For the second inequality, observe that the normalized mean-squared error of the ML estimator can be expressed using the relationship:

$$\frac{\mathbb{E} \left[\frac{\|\hat{\mathbf{x}}^{\text{ML}} - \mathbf{x}\|^2}{n \sigma^2} \right]}{n \sigma^2} = \int_0^\infty \mathbb{P} \left[\frac{\|\hat{\mathbf{x}}^{\text{ML}} - \mathbf{x}\|^2}{n \sigma^2} \geq t \right] dt. \quad (9)$$

Replacing the integrand on the right-hand side of (9) with the upper bound (5) and taking the supremum over \mathbf{x} and σ^2 leads to (8). Convergence of the integral for all $m > k + 2$ follows from the fact that $\Delta(u) \sim (1/2) \log(u)$ as $u \rightarrow \infty$. ■

In the proportional growth setting with $m/n \rightarrow \delta$ and $k/n \rightarrow \varepsilon$, Theorem 2 provides the following upper bound on the asymptotic minimax noise sensitivity.

Theorem 3. *For any pair $(\delta, \varepsilon) \in (0, 1)^2$ with $\delta > \varepsilon$, the asymptotic minimax noise sensitivity $M^*(\delta, \varepsilon)$ obeys:*

$$M^*(\delta, \varepsilon) \leq \delta \left(\frac{\delta}{\delta - \varepsilon} \left[\mathcal{L}^{-1} \left(\frac{H(\varepsilon)}{\delta - \varepsilon} \right) \right]^2 - 1 \right). \quad (10)$$

Proof: This result follows from taking the limit of (8) and noting that $C_{m-k} \rightarrow 0$ as $m - k \rightarrow \infty$. ■

III. LOWER BOUNDS

We now consider lower bounds on the minimax noise sensitivity by analyzing the worst-case behavior in the Bayesian setting where \mathbf{x} is a random vector with a known distribution. Specifically, we define the worst-case Bayes risk to be

$$M_n^B(m, k) = \sup_{\sigma^2 > 0} \sup_{P \in \mathcal{P}^n(k)} \inf_g \frac{\mathbb{E}[\text{mse}(\mathbf{x}, g, \sigma^2)]}{\sigma^2} \quad (11)$$

where $\mathcal{P}^n(k)$ denotes the class of all probability distributions P supported on $B^n(k)$ and the expectation is taken with respect to the a random vector $\mathbf{x} \sim P$.

Since $M_n^B(m, k)$ corresponds to the minimax noise sensitivity defined in (2) with the order of the minimization and maximization swapped, it provides a lower bound on the minimax noise sensitivity:

$$M_n^B(m, k) \leq M_n^*(m, k).$$

A. Bounds when sparsity pattern is known

The following lower bound on $M_n^B(m, k)$ corresponds to the setting where the sparsity pattern (i.e. the locations of the nonzero entries) is known to the decoder.

Theorem 4. *The worst-case Bayes noise sensitivity $M_n^B(m, k)$ is infinite for $m \leq k + 1$ and obeys:*

$$M_n^B(m, k) \geq \frac{mk}{n} \mathbb{E} \left[\frac{1}{W} \right] \quad (12)$$

for $m > k + 1$ where W is a chi-square random variable with $m - k + 1$ degrees of freedom.

Proof: Let \mathbf{x} be a random vector whose first k entries are i.i.d. $\mathcal{N}(0, \gamma)$ and whose remaining entries are equal to zero. It is well known (see e.g. [11]) that the minimum mean squared error at noise power $\sigma^2 = 1$ is given explicitly by

$$\inf_g \mathbb{E}[\text{mse}(\mathbf{x}, g, 1)] = \frac{1}{n} \mathbb{E} \left[\text{tr} \left\{ (\gamma^{-1} I + H^* H)^{-1} \right\} \right]$$

where $H \in \mathbb{R}^{m \times k}$ corresponds to the first k columns of A . In the limit $\gamma \rightarrow \infty$, this expression converges to

$$\frac{1}{n} \mathbb{E} \left[\text{tr} \left\{ (H^* H)^{-1} \right\} \right] = \frac{k}{n} \mathbb{E} \left[\left[(H^* H)^{-1} \right]_{1,1} \right] = \frac{mk}{n} \mathbb{E} \left[\frac{1}{W} \right]$$

where the last equality follows from results in [12].

Finally, note that if $m = k + 2$, then W has an exponential distribution and $\mathbb{E}[1/W] = \infty$. Since the minimax noise

sensitivity is nonincreasing in m , this proves that it is infinite for all $m \leq k + 1$. ■

In the large system limit, Theorem 4 gives the following upper bound on the asymptotic worst-case Bayes noise sensitivity, defined by

$$M^B(\delta, \varepsilon) = \limsup_{n \rightarrow \infty} M_n^B(\lceil \delta \cdot n \rceil, \lceil \varepsilon \cdot n \rceil).$$

Corollary 5. *For any pair $(\delta, \varepsilon) \in (0, 1)^2$ with $\delta > \varepsilon$, the asymptotic worst-case Bayes noise sensitivity $M^B(\delta, \varepsilon)$ obeys:*

$$M^B(\delta, \varepsilon) \geq \frac{\delta \varepsilon}{\delta - \varepsilon}. \quad (13)$$

We note that the bound (13) is related to the signal-to-interference ratio of the decorrelator in multiuser detection (see e.g. [12, Theorem 3.1]).

B. Asymptotic bounds using the replica heuristics

Next, we study the asymptotic behavior of the minimum mean-squared error using the replica method from statistical physics; this approach was adapted to vector estimation in [7], [8] and has been used for the compressed sensing model in a number of recent papers (see e.g. [9], [13], [14]).

We note that the analysis based on the replica method is heuristic since it depends on certain key assumptions – notably replica symmetry – currently lacking rigorous justification in the context of our model.

Proposition 6. *Assume that the replica assumptions hold. Let X be a random variable obeying $\Pr[X \neq 0] \leq \varepsilon$ and*

$$\Pr[X^2 > t] \leq \exp(-c \cdot t)$$

for some positive constant c . Then, the asymptotic worst-case Bayes noise sensitivity $M^B(\delta, \varepsilon)$ obeys

$$M^B(\delta, \varepsilon) \geq \delta(\tau^* - 1) \quad (14)$$

where τ^* is the largest minimizer τ of the free energy

$$G(\tau) = \delta[\log(\tau) + \tau^{-1}] + 2I(X; X + \sqrt{\tau}Z), \quad (15)$$

with $Z \sim \mathcal{N}(0, 1)$ independent of X .

Proof Sketch: The right-hand side of (14) is the asymptotic MMSE, as predicted by the replica method, when the entries of \mathbf{x} are i.i.d. copies of X and $\sigma^2 = 1$. The fact that this expression gives a lower bound on $M^B(\delta, \varepsilon)$ follows straightforwardly from the assumptions on X ; the details are omitted due to space constraints. ■

Observe that each choice of X in Proposition 6 gives a lower bound on the asymptotic minimax noise sensitivity $M^*(\delta, \varepsilon)$. For each pair (δ, ε) , it can be verified that the greatest lower bound is given by the supremum over the class of all probability distributions placing all but ε of their mass at the origin.

IV. PROOF OF THEOREM 1

The ML estimator is scale invariant in the sense that

$$g^{\text{ML}}(A, A\mathbf{x} + \sigma\mathbf{w}) = g^{\text{ML}}(A, A\mathbf{x}/\sigma + \mathbf{w})$$

for all $\sigma > 0$. Since the set $B^n(k)$ is also scale invariant, it is sufficient to consider the case $\sigma^2 = 1$.

Next, observe that any $\mathbf{u} \in B^n(k)$ lies in the range space $\mathcal{R}(D)$ of an $n \times k$ matrix D whose columns are drawn from the $n \times n$ identity matrix. Accordingly, the minimization in the right-hand side of (4) can be expressed equivalently as

$$\min_{D \in \mathbb{R}^{n \times k}} \left\{ \min_{\mathbf{u} \in \mathcal{R}(D)} \|\mathbf{y} - A\mathbf{u}\|^2 \right\}, \quad (16)$$

where the outer minimization is over all $\binom{n}{k}$ unique choices of the matrix D (up to permutations of the columns).

By basic linear algebra, the inner minimization in (16) has a unique solution given explicitly by

$$\mathbf{u}_D = D(AD)^\dagger \mathbf{y}, \quad (17)$$

where $(AD)^\dagger$ denotes the Moore–Penrose pseudoinverse.

To proceed, it is useful to define the random events

$$\begin{aligned} \mathcal{A}_s(D) &= \left\{ \frac{\|\mathbf{y} - A\mathbf{u}_D\|^2}{m-k} \leq 1+s \right\} \\ \mathcal{B}_t(D) &= \left\{ \frac{\|\mathbf{x} - \mathbf{u}_D\|^2}{n} \geq \frac{m}{n} \left(\frac{m+1}{m-k+1} (1+t) - 1 \right) \right\}. \end{aligned}$$

If we let D_0 denote the minimizer of the outer minimization in (16), then the left-hand side of (5) corresponds to $\mathbb{P}[\mathcal{B}_t(D_0)]$ (with an appropriate substitution of t). By conditioning on the event $\mathcal{A}_s(D_0)$, we can upper bound this probability as follows:

$$\begin{aligned} \mathbb{P}[\mathcal{B}_t(D_0)] &= \mathbb{P}[\mathcal{B}_t(D_0) \cap \mathcal{A}_s(D_0)] + \mathbb{P}[\mathcal{B}_t(D_0) \cap \mathcal{A}_s^c(D_0)] \\ &\leq \inf_s \left[\mathbb{P}[\mathcal{B}_t(D_0) \cap \mathcal{A}_s(D_0)] + \mathbb{P}[\mathcal{A}_s^c(D_0)] \right] \end{aligned} \quad (18)$$

where the superscript c denotes the complement of an event.

By the union bound, we have

$$\mathbb{P}[\mathcal{B}_t(D_0) \cap \mathcal{A}_s(D_0)] \leq \binom{n}{k} \max_D \mathbb{P}[\mathcal{B}_t(D) \cap \mathcal{A}_s(D)] \quad (19)$$

Furthermore, by the definition of D_0 , we have

$$\mathbb{P}[\mathcal{A}_s^c(D_0)] \leq \min_D \mathbb{P}[\mathcal{A}_s^c(D)]. \quad (20)$$

The following technical result is proved in Section IV-A.

Lemma 7. *For any vector $\mathbf{x} \in \mathbb{R}^n$ and matrix $D \in \mathbb{R}^{n \times k}$ with orthonormal columns, the events $\mathcal{A}_s(D)$ and $\mathcal{B}_t(D)$ are independent and their probabilities are given by*

$$\mathbb{P}[\mathcal{A}_s(D)] = \mathbb{P} \left[\frac{V}{\mathbb{E}[V]} \leq \frac{1+s}{1+\mu_D} \right] \quad (21)$$

$$\mathbb{P}[\mathcal{B}_t(D)] = \mathbb{P} \left[\frac{B}{\mathbb{E}[B]} \leq \frac{1+\mu_D}{1+t} \right] \quad (22)$$

where $\mu_D = \frac{1}{n} \|(I - DD^*)\mathbf{x}\|^2$, V is chi-square with $m-k$ degrees of freedom, and B is Beta($\frac{m-k+1}{2}, \frac{k}{2}$).

Using standard Chernoff bounds, it can be shown that

$$\begin{aligned} \mathbb{P}[V \leq (1/x)\mathbb{E}[V]] &\leq \exp\left(-\frac{(m-k)}{2} \mathcal{L}(x)\right) \\ \mathbb{P}[B \leq (1/x)\mathbb{E}[B]] &\leq \exp\left(-\frac{(m-k)}{2} \mathcal{L}(x)\right), \end{aligned}$$

for all x where $\mathcal{L}(x) = [\log(x) + 1/x - 1]$ for all $x \geq 1$ and is zero otherwise. Combining these bounds with Lemma 7 gives, for $0 \leq s \leq t$,

$$\begin{aligned} \max_D \mathbb{P}[\mathcal{A}_s(D) \cap \mathcal{B}_t(D)] &= \max_D \mathbb{P}[\mathcal{A}_s(D)] \mathbb{P}[\mathcal{B}_t(D)] \\ &\leq \max_D \exp\left(-\frac{(m-k)}{2} \left[\mathcal{L}\left(\frac{1+\mu_D}{1+s}\right) + \mathcal{L}\left(\frac{1+t}{1+\mu_D}\right) \right]\right) \\ &\leq \sup_{\mu \geq 0} \exp\left(-\frac{(m-k)}{2} \left[\mathcal{L}\left(\frac{1+\mu}{1+s}\right) + \mathcal{L}\left(\frac{1+t}{1+\mu}\right) \right]\right) \\ &= \exp\left(-(m-k) \mathcal{L}\left(\frac{\sqrt{1+t}}{\sqrt{1+s}}\right)\right). \end{aligned} \quad (23)$$

Also, by the definition of D_0 and the fact that $\mu_D = 0$ for the matrix corresponding to the support of \mathbf{x} , we have

$$\begin{aligned} \min_D \mathbb{P}[\mathcal{A}_s^c(D)] &= \min_D \mathbb{P} \left[\frac{V}{\mathbb{E}[V]} > \frac{1+s}{1+\mu_D} \right] \\ &\leq \exp\left(-\frac{(m-k)}{4} (\sqrt{1+2s} - 1)^2\right), \end{aligned} \quad (24)$$

where (24) follows from a Chernoff bound [15, pg. 1325].

At this point, putting everything together gives an upper bound in terms of the parameter t ,

$$\begin{aligned} \mathbb{P}[\mathcal{B}_t(D_0)] &\leq \inf_{0 \leq s \leq t} \left[\binom{n}{k} \exp\left(-(m-k) \mathcal{L}\left(\frac{\sqrt{1+t}}{\sqrt{1+s}}\right)\right) \right. \\ &\quad \left. + \exp\left(-\frac{(m-k)}{4} (\sqrt{1+2s} - 1)^2\right) \right]. \end{aligned}$$

To simplify the expression, we make the substitution

$$1+t = \left[\mathcal{L}^{-1}\left(\frac{n}{m-k} H\left(\frac{k}{n}\right)\right) \right]^2 (1+u).$$

Using the that $\mathcal{L}(ab) \geq \mathcal{L}(a) + \mathcal{L}(b)$ for all $a, b \geq 1$, we have, for $0 \leq s \leq u$,

$$\mathcal{L}\left(\frac{\sqrt{1+t}}{\sqrt{1+s}}\right) \geq \frac{n}{m-k} H\left(\frac{k}{n}\right) + \mathcal{L}\left(\frac{\sqrt{1+u}}{\sqrt{1+s}}\right) \quad (25)$$

Combining (25) with the inequality $\binom{n}{k} \leq \exp(nH(\frac{k}{n}))$ (see e.g. [16, pg. 353]), and taking the supremum over $s \in [0, u]$ leads to (5). This concludes the proof of Theorem 1.

A. Proof of Lemma 7

Let $H = AD$ and $\mathbf{z} = A(I - DD^*)\mathbf{x} + \mathbf{w}$. Then, we have

$$\mathbf{y} = HD^*\mathbf{x} + \mathbf{z} \quad \text{and} \quad \mathbf{u}_D = DH^\dagger \mathbf{y}.$$

where the entries in \mathbf{z} are i.i.d. $\mathcal{N}(0, 1 + \mu_D)$; the entries in H are i.i.d. $\mathcal{N}(0, 1/m)$; and H and \mathbf{z} are independent.

The residual sum of squares can be expressed as

$$\|\mathbf{y} - A\mathbf{u}_D\|^2 = \|(I - HH^\dagger)\mathbf{y}\|^2 = \|(I - HH^\dagger)\mathbf{z}\|^2.$$

With probability one, $(I - HH^\dagger)$ has $m - k$ eigenvalues equal to one and the rest equal to zero, and thus

$$V = \frac{\|(I - HH^\dagger)\mathbf{z}\|^2}{1 + \mu_D}$$

has a chi-square distribution with $m - k$ degrees of freedom. This completes the proof of (21).

Next, if we assume that the columns of H are linearly independent (an event which occurs with probability one) then $H^\dagger H = I$, and so

$$\begin{aligned} \|\mathbf{u}_D - \mathbf{x}\|^2 &= \|(DH^\dagger HD^* - I)\mathbf{x} + DH^\dagger \mathbf{z}\|^2 \\ &= m \cdot \mu_D + \|DH^\dagger \mathbf{z}\|^2. \end{aligned}$$

At this point, the independence of $\|\mathbf{y} - A\mathbf{u}_D\|^2$ and $\|\mathbf{u}_D - \mathbf{x}\|^2$ follows from the rotational invariance of the Gaussian distribution and the fact that $(I - HH^\dagger)$ and HH^\dagger are orthogonal projections onto complementary linear subspaces.

To characterize the distribution of $\|\mathbf{u}_D - \mathbf{x}\|^2$, observe that by the rotational invariance of the Gaussian distribution,

$$\|DH^\dagger \mathbf{z}\|^2 = \|H^\dagger \mathbf{z}\|^2 \stackrel{d}{=} \|(H^* H)^{-1/2} \tilde{\mathbf{z}}\|^2 \quad (26)$$

where $\stackrel{d}{=}$ denotes equality in distribution and $\tilde{\mathbf{z}} \in \mathbb{R}^k$ corresponds to the first k entries in \mathbf{z} .

Next, we use that well known fact that $\tilde{\mathbf{z}}$ can be decomposed into two independent quantities:

$$T = \|\tilde{\mathbf{z}}\|^2 / (1 + \mu_D) \quad \text{and} \quad \mathbf{s} = \tilde{\mathbf{z}} / \|\tilde{\mathbf{z}}\|$$

where T has a chi-square distribution with k degrees of freedom and \mathbf{s} is distributed uniformly on the unit sphere. plugging these expressions back into (26), we have

$$\begin{aligned} \|(H^* H)^{-1/2} \tilde{\mathbf{z}}\|^2 &= \|(H^* H)^{-1/2} \mathbf{s}\|^2 T (1 + \mu_D) \\ &\stackrel{d}{=} [(H^* H)^{-1}]_{1,1} T (1 + \mu_D) \end{aligned}$$

where the second step follows from the rotational invariance of the Gaussian distribution and the fact that H and \mathbf{s} are independent. Finally, it is shown in [12] that

$$W = \frac{m}{[(H^* H)^{-1}]_{1,1}}$$

has a chi-square distribution with $m - k + 1$ degrees of freedom.

Putting everything back together we have

$$\frac{\|\mathbf{u}_D - \mathbf{x}\|^2}{m} = \mu_D + (1 + \mu_D) \frac{T}{W} = \frac{1 + \mu_D}{B} - 1$$

where $B = W/(W + T)$ has a Beta($\frac{m-k+1}{2}, \frac{k}{2}$) distribution. This completes the proof of (22).

V. DISCUSSION

This paper considers the minimax behavior of the noise sensitivity in compressed sensing over the class of sparse vectors; it presents non-asymptotic bounds on the minimax noise sensitivity $M_n^*(m, k)$ and studies the large system limit $M^*(\delta, \varepsilon)$.

Our bounds on $M^*(\delta, \varepsilon)$ are plotted in Figure 1 as a function of δ for $\varepsilon = 0.2$. The ML upper bound corresponds to

Theorem 1 and is finite for all $\delta > \varepsilon$. The replica lower bounds are given by evaluating Proposition 6 for the Bernoulli-Gaussian distribution with various noise powers.

For comparison, Figure 1 also shows upper bounds on $M^*(\delta, \varepsilon)$ given by exact formulas for the noise sensitivity of AMP soft thresholding [3] and AMP with the optimal denoiser [6]. While those formulas yield improved upper bounds for large δ , they blow up as δ approaches their respective AMP phase transitions, given by $\delta^*(0.2|\text{AMP-soft}|) \approx 0.5110$ for soft-threshold and $\delta^*(0.2|\text{AMP-minimax}|) \approx 0.4803$ for the minimax denoiser.

One problem to attack in future work is to evaluate the minimax Replica MSE, i.e. to characterize the real-valued probability distribution maximizing the Replica MSE given by the right hand side of (14). Assuming the validity of the replica assumptions, such a distribution would constitute a least-favorable prior for the case of i.i.d. vectors with sparsity constraint ε . Our calculations indicate that this distribution is different than the related problem of minimaxing the AMP MSE.

REFERENCES

- [1] 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.
- [2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, Feb 2006.
- [3] D. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6920–6941, Oct 2011.
- [4] D. L. Donoho, I. M. Johnstone, A. Maleki, and A. Montanari, "Compressed sensing over ℓ_p -balls: Minimax mean square error," in *Proceedings of the IEEE International Symposium on Information Theory*, St. Petersburg, Russia, Aug 2011.
- [5] M. Bayati and A. Montanari, "The LASSO risk for gaussian matrices," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 1997–2017, Apr 2012.
- [6] D. L. Donoho, I. M. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," 2011, <http://arxiv.org/abs/1111.1041>.
- [7] 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.
- [8] D. Guo, S. Shamai, and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun 2005.
- [9] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241 – 6263, Oct 2012.
- [10] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New York: Wiley, 2009.
- [11] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA: now Publisher Inc., 2004.
- [12] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, Jan 2000.
- [13] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proceedings of the 47-th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct 2009.
- [14] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.
- [15] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1302 – 1338, 2000.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.