On Approximate Message Passing for Reconstruction of Non-Uniformly Sparse Signals

Subhojit Som, Lee C. Potter, and Philip Schniter
Department of Electrical and Computer Engineering,
The Ohio State University, Columbus, OH 43210.
Email: som.4@osu.edu, potter@ece.osu.edu, schniter@ece.osu.edu

***Abstract*—This paper considers the reconstruction of non-uniformly sparse signals from noisy linear observations. By non-uniformly sparse, we mean that the signal coefficients can be partitioned into subsets that differ in the rate at which the coefficients tend to be active (i.e., nonzero). Inspired by recent work of Donoho, Maleki, and Montanari, we design a minimax-optimal approximate message passing (AMP) algorithm and we analyze it using a state evolution (SE) formalism that applies in the limit of very large problem dimensions. For the noiseless case, the SE formalism implies a phase transition curve (PTC) that bisects the admissible region of the sparsity-undersampling plane into two sub-regions: one where perfect recovery is very likely, and one where it is very unlikely. The PTC depends on the ratios of the activity rates and the relative sizes of the coefficient subsets. For the noisy case, we show that the same PTC also bisects the admissible region of the sparsity-undersampling plane into two sub-regions: one where the noise sensitivity remains finite and characterizable, and one where it becomes infinite (as the problem dimensions increase). Furthermore, we derive the formal mean-squared error (MSE) for (sparsity,undersampling) pairs in the region below the PTC. Numerical results suggest that the MSE predicted by the SE formalism closely matches the empirical MSE throughout the admissible region of the sparsity-undersampling plane, so long as the dimensions of the problem are adequately large.**[1]

I. Introduction

In classical sparse reconstruction, a key component of compressive sensing, the objective is to estimate a sparse vector $x^{\mathsf{true}} \in \mathbb{R}^N$ from the observations

$$y \;=\; Ax^{\mathsf{true}} + w, \tag{1}$$

where $A \in \mathbb{R}^{n\times N}$ is a known mixing matrix and $w \in \mathbb{R}^n$ is additive noise, which we take to be white Gaussian, i.e., $w \sim \mathcal{N}(0,\sigma^2 I)$. By "sparse", we mean that only $k \ll N$ entries of x^{true} are nonzero, though of course the locations of these nonzero elements are unknown. In many applications of interest, the dimensions n, N, and k are very large, motivating theoretical analysis of the so-called *large system limit*: $n, N, k \to \infty$ with $n/N \to \delta$ and $k/n \to \rho$ for $\delta, \rho \in [0,1]$. We will refer to ρ as the *sparsity rate* and to δ as the *undersampling ratio*. We will refer to $\epsilon \triangleq \rho\delta$ (i.e., the limiting value of k/N) as the *coefficient activity rate*.

In this paper, we consider a variant of the classical sparse reconstruction problem wherein there exists prior knowledge that certain subsets of signal coefficients have higher coefficient activity rates than others. We refer to this problem as *reconstruction of a non-uniformly sparse signal*. To make this problem more concrete, we assume that the coefficient index set can be partitioned into S disjoint subsets i.e., $\{1,\dots,N\} = \bigcup_{s=1}^{S} I_s$, where, for the s^{th} subset, $N_s \triangleq |I_s|$ denotes the size and k_s denotes the number of nonzero elements. Furthermore, we augment the previously described large system limit with the additional properties that $k_s/N_s \to \epsilon_s$ and $N_s/N \to \xi_s$ for $\epsilon_s, \xi_s \in [0,1]$. Thus, we can relate the local activity rates $\underline{\epsilon} \triangleq (\epsilon_1,\dots,\epsilon_S)$ to the global activity rate ϵ via $\epsilon = \sum_{s=1}^{S} \epsilon_s\xi_s$. In this work, we consider the case where the activity vector $\underline{\epsilon}$ is known apriori. While this may not always be a good assumption, we are motivated by so-called *turbo* sparse-reconstruction schemes whereby an external "pattern decoder" block supplies the sparse reconstruction algorithm with the activity vector [1].

Related work on the reconstruction of non-uniformly sparse signals includes [2] and [3], both of which considered weighted-ℓ_1 criteria.

II. Background

To set the stage for our work on the reconstruction of non-uniformly sparse signals, we first provide some background on the reconstruction of uniformly sparse signals (i.e., the classical sparse reconstruction problem).

A. Popular Algorithms

One of the most popular approaches to the classical sparse reconstruction problem is known as basis pursuit (BP) [4] or Lasso [5]. There, x^{true} is estimated by solving the convex optimization problem

$$\arg\min_x \left\{ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 \right\}, \tag{2}$$

for some carefully chosen regularization parameter $\lambda > 0$. We note that, in the noiseless case, (2) reduces to

$$\arg\min_x \|x\|_1 \text{ s.t. } y = Ax. \tag{3}$$

Another popular approach, known as *iterative thresholding* (IT), generates a sequence of estimates $x^1, x^2, x^3, \dots$ using the recursion

$$z^t \;=\; y - Ax^t \tag{4}$$

$$x^{t+1} \;=\; \eta_t(x^t + A^* z^t), \tag{5}$$

[1]This work was supported in part by NSF IUCRC Grant No. IIP-0968910.

initialized at $x^0 = 0$. In (4)-(5), $\eta_t(\cdot)$ is a component-wise denoising function that suppresses "small" values and keeps "large" ones, z^t is the iteration-t residual, and $(\cdot)^*$ denotes matrix transpose. While IT has advantages over BP in the realm of computational complexity, an extensive numerical study [6] of the noiseless case recently found that, under optimal tunings, the sparsity-undersampling tradeoff achieved by IT is inferior to that of BP. However, a small but important modification to IT recently proposed by Donoho, Maleki, and Montanari, and referred to as *approximate message passing* (AMP), was shown to yield empirical performance matching BP while maintaining complexity akin to IT [7]. Moreover, AMP's excellent performance has been rigorously justified in the large system limit when A has i.i.d Gaussian elements [8].

AMP [7] takes the form

$$z^t = y - Ax^t + \delta^{-1} z^{t-1} \langle \eta_t'(A^* z^{t-1} + x^{t-1}) \rangle \quad (6)$$

$$x^{t+1} = \eta_t(x^t + A^* z^t), \quad (7)$$

where $\eta_t'(a) = \frac{\partial}{\partial a}\eta_t(a)$ and $\langle \cdot \rangle$ denotes arithmetic average. AMP differs from IT in its inclusion of the term $\delta^{-1} z^{t-1} \langle \eta_t'(A^* z^{t-1} + x^{t-1}) \rangle$, whose role can be understood from the theory of belief propagation in graphical models [9], and which is analogous to the Onsager term from statistical physics.

B. State Evolution

While rigorous performance analysis of AMP under general A remains elusive, a great deal of intuition can be gained from a formalism known as *state evolution* (SE) [7], [10]–[13]. The motivation for SE starts with the IT algorithm, where it can be recognized that the input to the denoiser can be rewritten as

$$x^t + A^* z^t = x^t + A^* y - A^* A x^t \quad (8)$$

$$= x^t + A^* A x^{\mathsf{true}} + A^* w - A^* A x^t \quad (9)$$

$$= x^{\mathsf{true}} + \underbrace{(A^* A - I)}_{\triangleq H}(x^{\mathsf{true}} - x^t) + A^* w \quad (10)$$

Thus, the IT algorithm can be rephrased as

$$e^t \triangleq H(x^{\mathsf{true}} - x^t) + A^* w \quad (11)$$

$$x^{t+1} - x^{\mathsf{true}} = \eta_t(x^{\mathsf{true}} + e^t) - x^{\mathsf{true}}, \quad (12)$$

where e^t describes the time-t error in the denoiser's input. The essential properties of e^t can be deduced from the following lemma, which is a generalization of a similar lemma in [12].

Lemma 1. *Suppose, for some fixed x, that $u = Hx$, where $H = A^*A - I$ is generated from a random matrix A whose elements are drawn independently from a distribution that guarantees $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = \frac{1}{n}$, and $\mathbb{E}\{A_{ij}^4\} = \mathcal{O}(\frac{1}{n^2})$. Then $\mathbb{E}\{u_i\} = 0$, $\mathbb{E}\{u_i u_j\}\big|_{j \neq i} = \mathcal{O}(\frac{1}{n})$, and $\mathbb{E}\{u_i^2\} = \mathcal{O}(\frac{1}{n}) + \frac{1}{N\delta}\|x\|_2^2$, where the expectation is taken over A with the aforementioned properties.*

Under the conditions of Lemma 1, as $n \to \infty$, we have

- $\mathbb{E}\{e_i^t\} = 0$ for all i,
- $\mathbb{E}\{e_i^t e_j^t\} = 0$ for all i and $j \neq i$,
- $\mathbb{E}\{|e_i^t|^2\} = \frac{1}{N\delta}\|x^{\mathsf{true}} - x^t\|_2^2 + \sigma^2$ for all i.

where the expectations were taken over A and $w \sim \mathcal{N}(0, \sigma^2 I)$.

For very large N, the central limit theorem motivates the treatment of e^t as Gaussian. If we could furthermore justify the treatment of e^t as independent across iteration t, then we could rewrite the denoising step (12) as

$$x^{t+1} - x^{\mathsf{true}} = \eta_t(x^{\mathsf{true}} + \sigma_t v^t) - x^{\mathsf{true}}, \quad (13)$$

for independent noise $v^t \sim \mathcal{N}(0, I)$ and variance parameter

$$\sigma_t^2 \triangleq \frac{1}{N\delta}\mathbb{E}\{\|x^t - x^{\mathsf{true}}\|_2^2\} + \sigma^2. \quad (14)$$

Together, (13)-(14) yield the following SE for σ_t:

$$\sigma_{t+1}^2 = \frac{1}{N\delta}\mathbb{E}\left\{\|\eta_t(x^{\mathsf{true}} + \sigma_t v^t) - x^{\mathsf{true}}\|_2^2\right\} + \sigma^2. \quad (15)$$

From this SE, we are then fundamentally interested in knowing whether $\sigma_{t+1}^2 < \sigma_t^2$, since this property will determine whether or not the underlying iterative algorithm converges.

A critical assumption made in the development of the SE (15) was that e^t could be treated as independent Gaussian noise. The non-linear dynamics of the IT algorithm make this assumption difficult to justify and, indeed, empirical evidence suggests that the SE (15) does not accurately characterize the behavior of IT. However, a great deal of empirical evidence [7], [10]–[13] suggests that the SE (15) does accurately characterize both the transient and steady-state behaviors of AMP (6)-(7). Thus, the SE formalism appears to be a useful, if not rigorous, tool for analyzing the AMP algorithm.

Remarkably, the SE formalism predicts not only the performance limits of AMP, but also that of BP (3). In particular, for the noiseless case, [7] showed that the SE formalism predicts a *phase transition curve* (PTC) in the (δ, ρ) plane below which AMP perfectly reconstructs x^{true}, and above which it does not. Remarkably, the noiseless PTC of AMP coincides exactly with the PTC of BP, which was analyzed using a (much more complicated) combinatorial geometric approach in [14], [15]. Though these formal properties hold only in the large system limit, empirical performance at finite-N, n, k shows similar behavior. In the noisy case, [13] showed that the SE formalism predicts a PTC in the (δ, ρ) plane below which the *noise sensitivity* $\|x^{\mathsf{true}} - x^t\|_2^2/(N\sigma^2)$ converges to a finite and characterizable value, and above which it is possible to construct signals for which the noise sensitivity grows unboundedly (in the large system limit). Interestingly, this noisy PTC coincides with the noiseless PTC, and characterizes BP as well as AMP.

C. Soft Thresholding, Minimax MSE, and State Evolution

We now review some state evolution details that were presented in [13] for the uniform-sparse case, in preparation for our treatment of the non-uniform case in Section III.

For the remainder of the manuscript, we focus on a particular choice of AMP denoising function known as the *soft thresholding* function. In particular, we take $\eta_t(x) = \eta(x; \tau\hat{\sigma}_t)$,

where $\tau > 0$ is a tuning parameter, $\hat{\sigma}_t$ is an empirical estimate of the scale of the residuals, and

$$\eta(x;\theta) \triangleq \begin{cases} x-\theta & \text{if } x > \theta \\ 0 & \text{if } x \in [-\theta,\theta] \\ x+\theta & \text{if } x < -\theta \end{cases}. \quad (16)$$

With this denoiser, AMP (6)-(7) simplifies to [12]

$$z^t = y - Ax^t + \frac{\|x^t\|_0}{n} z^{t-1} \quad (17)$$

$$x^{t+1} = \eta(x^t + A^* z^t; \tau\hat{\sigma}_t), \quad (18)$$

where $\|x^t\|_0$ counts the number of nonzero elements in x^t and $\eta(\cdot;\cdot)$ in (18) performs component-wise denoising.

For the state evolution formalism in [13], it was assumed that the elements of x^{true} are i.i.d and drawn from some probability measure ν such that, for $X \sim \nu$, $\Pr\{X \neq 0\} \leq \epsilon$. The collection of all such measures will be denoted by $\mathcal{F}_\epsilon$.

The *soft thresholding MSE* is defined as

$$\mathsf{mse}(\sigma_t^2;\nu,\tau) \triangleq \mathbb{E}\left\{[\eta(X+\sigma_t V;\tau\sigma_t) - X]^2\right\}, \quad (19)$$

where the expectation is with respect to the independent variables $X \sim \nu$ and $V \sim \mathcal{N}(0,1)$. Notice that

$$\mathsf{mse}(\sigma_t^2;\nu,\tau) = \sigma_t^2 \mathsf{mse}(1;\nu^{1/\sigma_t},\tau), \quad (20)$$

where ν^{1/σ_t} is an appropriate dilation of the probability measure ν. Note that if $\nu \in \mathcal{F}_\epsilon$ then $\nu^{1/\sigma_t} \in \mathcal{F}_\epsilon$. Thus, it suffices to focus on the case $\sigma_t = 1$ and scale the MSE as needed, motivating the shorthand notation $\mathsf{mse}(\nu,\tau) \triangleq \mathsf{mse}(1;\nu,\tau)$.

The *minimax thresholding MSE* [16] is defined as

$$M^{\pm}(\epsilon) = \inf_{\tau>0} \sup_{\nu\in\mathcal{F}_\epsilon} \mathsf{mse}(\nu,\tau), \quad (21)$$

where the τ achieving the infimum will be denoted by $\tau^{\pm}(\epsilon)$. Figure 1 shows $M^{\pm}(\epsilon)$ as a function of ϵ. The threshold $\tau^{\pm}(\epsilon)$ is explicitly computable at $\epsilon > 0$ (see Fig. 2) and obeys

$$\lim_{\epsilon\to 0} \frac{\tau^{\pm}(\epsilon)}{\sqrt{2\log(\epsilon^{-1})}} = 1. \quad (22)$$

For any fixed τ, it holds that [13]

$$\sup_{\nu\in\mathcal{F}_\epsilon} \mathsf{mse}(\nu,\tau) = \epsilon(1+\tau^2) + (1-\epsilon)[2(1+\tau^2)\Phi(-\tau) - 2\tau\phi(\tau)], \quad (23)$$

where $\phi(a) \triangleq \exp(-a^2/2)/\sqrt{2\pi}$ denotes the normal pdf and $\Phi(a) \triangleq \int_{-\infty}^{a} \phi(t)\mathrm{dt}$ denotes the normal cdf.

The supremum in (23) is reached by the three-point mixture

$$\nu^*(\epsilon) = (1-\epsilon)\delta_0 + \frac{\epsilon}{2}\delta_\infty + \frac{\epsilon}{2}\delta_{-\infty} \quad (24)$$

on the extended real line $\mathbb{R} \cup \{-\infty,\infty\}$. [In (24), δ_a denotes the Dirac delta centered at a.] Although the least-favorable prior $\nu^*(\epsilon)$ is not practical, the authors of [13] demonstrate that there exist *almost*-least-favorable priors that are three-point mixtures with bounded variance. To do this, they define a so-called α-*least-favorable* prior as the distribution $\nu(\epsilon,\alpha) \in \mathcal{F}_\epsilon$ with the smallest second moment that satisfies

$$\mathsf{mse}\big(\nu(\epsilon,\alpha),\tau^{\pm}(\epsilon)\big) = (1-\alpha)\cdot \sup_{\nu\in\mathcal{F}_\epsilon} \mathsf{mse}(\nu,\tau^{\pm}(\epsilon)), \quad (25)$$

where $\alpha \in (0,1)$ is usually chosen so that $\alpha \ll 1$ (e.g., $\alpha = 0.02$ is often used in [13]). Indeed, the α-least-favorable prior is a three-point mixture of the form

$$\nu^*(\epsilon,\alpha) = (1-\epsilon)\delta_0 + \frac{\epsilon}{2}\delta_{\mu^{\pm}(\epsilon,\alpha)} + \frac{\epsilon}{2}\delta_{-\mu^{\pm}(\epsilon,\alpha)}, \quad (26)$$

where the value $\mu^{\pm}(\epsilon,\alpha)$ is explicitly computable when $\epsilon > 0$ and obeys the limit

$$\lim_{\epsilon\to 0} \frac{\mu^{\pm}(\epsilon,\alpha)}{\sqrt{2\log(\epsilon^{-1})}} = 1. \quad (27)$$

The *MSE map* is defined as

$$\Psi(m,\delta,\sigma,\tau,\nu) \triangleq \mathsf{mse}(\sigma^2 + m/\delta;\nu,\tau), \quad (28)$$

and *state evolution* (SE) is defined as the iterated application of Ψ to the state m_t (for fixed values of δ,σ,τ,ν):

$$\begin{cases} m_t \mapsto m_{t+1} \triangleq \Psi(m_t,\delta,\sigma,\tau,\nu) \\ t \mapsto t+1 \end{cases}, \quad (29)$$

initialized at $m_0 = \mu_2(\nu)$, where $\mu_2(\nu)$ denotes the second moment of the distribution ν. In [13], it was shown that the SE (29) converges to the *highest fixed point* of the continuous function $\Psi(\cdot,\delta,\sigma,\tau,\nu)$. In other words,

$$\lim_{t\to\infty} m_t = \mathrm{HFP}(\Psi) \quad (30)$$

for

$$\mathrm{HFP}(\Psi) \triangleq \sup\{s : \Psi(m) \geq m\}. \quad (31)$$

The authors of [13] defined the *formal MSE* (fMSE) as the value of MSE $\|x^{\mathsf{true}} - x^t\|_2^2/N$ predicted by the SE formalism. Below the PTC, it was shown that

$$\inf_{\tau} \sup_{\nu\in\mathcal{F}_\epsilon} \mathrm{fMSE}(\tau;\delta,\sigma=1,\nu) = \frac{M^{\pm}(\epsilon)}{1 - M^{\pm}(\epsilon)/\delta} \triangleq M^*(\epsilon,\delta). \quad (32)$$

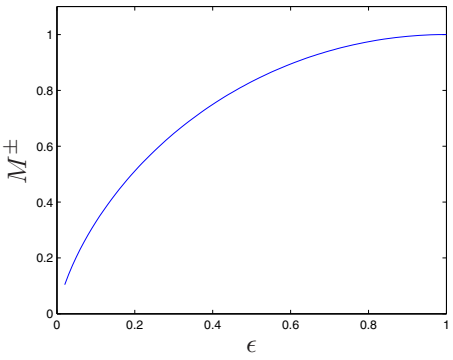

Fig. 1. The minimax thresholding MSE $M^{\pm}(\epsilon)$ versus sparsity ϵ.

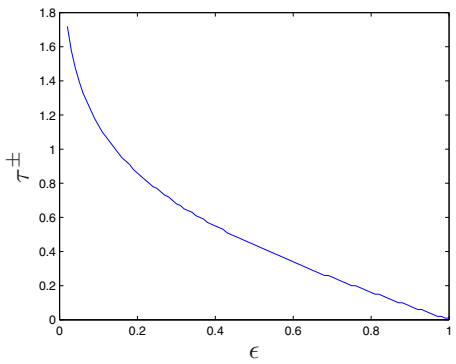

Fig. 2. The minimax-optimal threshold $\tau^{\pm}(\epsilon)$ versus sparsity ϵ.

Also, it was shown that the threshold policy $\tau^* = \tau^{\pm}(\epsilon)$ minimaxes the formal MSE, i.e.,

$$\sup_{\nu \in \mathcal{F}_\epsilon} \text{fMSE}(\tau^*; \delta, \sigma = 1, \nu) = \inf_{\tau} \sup_{\nu \in \mathcal{F}_\epsilon} \text{fMSE}(\tau; \delta, \sigma = 1, \nu). \tag{33}$$

III. Non-uniform Sparsity

We now generalize the concepts presented in Section II to the case of non-uniform sparsity, as defined in Section I.

A. The AMP-NU Algorithm

For AMP to handle non-uniform sparsity, we only need to modify the soft threshold function so that it chooses an appropriate threshold value for each coefficient subset I_s. In particular, for coefficient index $i \in I_s$, we choose the threshold value

$$\tau_s = \tau^{\pm}(\epsilon_s), \tag{34}$$

where ϵ_s was defined as the activity rate of the s^{th} subset, which is assumed to be known. The *AMP algorithm for non-uniform sparsity* (AMP-NU) then becomes

$$z^t = y - Ax^t + \frac{\|x^t\|_0}{n} z^{t-1} \tag{35}$$

$$x^{t+1} = \eta(x^t + A^* z^t; \hat{\sigma}_t \bar{\tau}), \tag{36}$$

where $\bar{\tau} \in \{\mathbb{R}^+\}^N$ is defined such that its i^{th} entry equals $\tau_s = \tau^{\pm}(\epsilon_s)$ when $i \in I_s$ [recalling (34)]. Here, we have extended the definition of η so that $\eta(r; \bar{\theta})$ performs component-wise denoising on $r \in \mathbb{R}^N$ according to the *vector* of thresholds $\bar{\theta} \in \{\mathbb{R}^+\}^N$.

B. State Evolution

For the SE formalism of AMP-NU, we treat the coefficients in x^{true} as independent and, furthermore, as identically distributed within each subset I_s but not across subsets. In particular, for $i \in I_s$, we treat $x_i^{\text{true}} \sim \nu_s$ for some (possibly worst-case) $\nu_s \in \mathcal{F}_{\epsilon_s}$.

We now generalize some of the concepts from [13] that were reviewed in Section II-C. For convenience, we define $\underline{\epsilon} \triangleq (\epsilon_1, \dots, \epsilon_S)$, $\underline{\tau} \triangleq (\tau_1, \dots, \tau_S)$, $\underline{\xi} \triangleq (\xi_1, \dots, \xi_S)$, $\underline{\nu} \triangleq (\nu_1, \dots, \nu_S)$, and $\mathcal{F}_{\underline{\epsilon}} \triangleq (\mathcal{F}_{\epsilon_1}, \dots, \mathcal{F}_{\epsilon_S})$. In the sequel, $\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}$ is taken to mean that $\nu_s \in \mathcal{F}_{\epsilon_s}$ for all $s \in \{1, \dots, S\}$.

First, we generalize the soft thresholding MSE (19) to

$$\begin{aligned} &\mathsf{mse}(\sigma_t^2; \underline{\nu}, \underline{\tau}, \underline{\xi}) \\ &\quad \triangleq \sum_{s=1}^{S} \xi_s \, \mathbb{E}\left\{ \left[\eta(X_s + \sigma_t V; \tau_s \sigma_t) - X_s \right]^2 \right\}, \end{aligned} \tag{37}$$

where the expectations are taken with respect to $X_s \sim \nu_s$ and $V \sim \mathcal{N}(0,1)$. We then apply the MSE scaling concept (20) to (37), giving

$$\mathsf{mse}(\sigma_t^2; \underline{\nu}, \underline{\tau}, \underline{\xi}) = \sigma_t^2 \mathsf{mse}(1; \underline{\nu}^{1/\sigma_t}, \underline{\tau}, \underline{\xi}), \tag{38}$$

and declare the abbreviation

$$\mathsf{mse}(\underline{\nu}, \underline{\tau}, \underline{\xi}) \triangleq \mathsf{mse}(1; \underline{\nu}, \underline{\tau}, \underline{\xi}). \tag{39}$$

Next, we generalize the minimax thresholding MSE in (21) to

$$M^{\pm}(\underline{\epsilon}, \underline{\xi}) = \inf_{\underline{\tau} > 0} \sup_{\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}} \mathsf{mse}(\underline{\nu}, \underline{\tau}, \underline{\xi}) \tag{40}$$

$$= \sum_{s=1}^{S} \xi_s M^{\pm}(\epsilon_s). \tag{41}$$

Finally, we generalize the MSE map (28) to

$$\Psi(m, \delta, \sigma, \underline{\tau}, \underline{\nu}, \underline{\xi}) \triangleq \mathsf{mse}\Big(\sigma^2 + \frac{m}{\delta}; \underline{\nu}, \underline{\tau}, \underline{\xi}\Big), \tag{42}$$

$$= \sum_{s=1}^{S} \xi_s \, \mathbb{E}\left\{ \left[\eta\big(X_s + V\sqrt{\sigma^2 + \tfrac{m}{\delta}}; \tau_s \sqrt{\sigma^2 + \tfrac{m}{\delta}}\big) - X_s \right]^2 \right\}. \tag{43}$$

C. PTCs and the Admissible Sub-region of (δ, ρ)

For the case of uniform sparsity, we saw that the phase transition curve (PTC) bisects the region $(\delta, \rho) \in [0,1]^2$ into sub-regions of good performance (lying below the curve) and bad performance (lying above the curve). There, the definitions of "good" and "bad" depend on whether the observations are noisy [13] or noiseless [7].

In the case of non-uniform sparsity, however, it is important to realize that—depending on $\underline{\xi}$ and the ratios between the elements in $\underline{\epsilon}$—some of the points $(\delta, \rho) \in [0,1]^2$ may be inadmissible, forcing us to define the *admissible subregion in the sparsity-undersampling plane*, and to consider PTCs over only that subregion.

We now define our notion of admissibility. First, we define $r_{s',s} \triangleq \epsilon_{s'}/\epsilon_s$ as the *activity ratio* between coefficient subset s' and s, and we collect $\{r_{s',s}\}_{\forall s',s}$ into the vector $\underline{r}$. Then, recalling that $\epsilon = \rho\delta$ and $\epsilon = \sum_{s=1}^{S} \xi_s \epsilon_s$, we can write

$$\epsilon_s = \frac{\epsilon_s \epsilon}{\epsilon} = \frac{\epsilon_s \rho \delta}{\sum_{s'=1}^{S} \xi_{s'} \epsilon_{s'}} = \frac{\rho\delta}{\sum_{s'=1}^{S} \xi_{s'} r_{s',s}} \tag{44}$$

where, by definition, we must have $\epsilon_s \in [0,1]$. We then declare that a given pair $(\delta, \rho) \in [0,1]^2$ is *admissible for* $(\underline{\xi}, \underline{r})$ if

$$\frac{\rho\delta}{\sum_{s'=1}^{S} \xi_{s'} r_{s',s}} \le 1 \quad \text{for all } s \in \{1, \dots, S\}. \tag{45}$$

Finally, we denote the set of all admissible pairs (δ, ρ) as the *admissible region of the sparsity-undersampling plane*, and note that it is clearly dependent on $(\underline{\xi}, \underline{r})$.

D. The Case of Noiseless Observations

For the noiseless (i.e., $\sigma = 0$) case, we have

$$\begin{aligned}
&\Psi(m, \delta, \sigma = 0, \underline{\tau}, \underline{\nu}, \underline{\xi}) \\
&= \mathsf{mse}\Big(\frac{m}{\delta}; \underline{\nu}, \underline{\tau}, \underline{\xi}\Big) && (46)\\
&= \frac{1}{\delta}\mathsf{mse}\big(m; \underline{\nu}^{\sqrt{\delta}}, \underline{\tau}, \underline{\xi}\big) && (47)\\
&= \frac{1}{\delta}\sum_{s=1}^{S} \xi_s \, \mathbb{E}\left\{\left[\eta(X_s + mV; \tau_s m) - X_s\right]^2\right\} && (48)
\end{aligned}$$

for $X_s \sim \nu_s^{\sqrt{\delta}}$. Since we have made no assumptions on ν_s other than $\nu_s \in \mathcal{F}_{\epsilon_s}$, and since $\nu_s \in \mathcal{F}_{\epsilon_s} \Rightarrow \nu_s^{\sqrt{\delta}} \in \mathcal{F}_{\epsilon_s}$, we can take $X_s \sim \nu_s$ w.l.o.g in (48).

The noiseless non-uniform SE

$$\begin{cases} m_t \mapsto m_{t+1} = \Psi(m_t, \delta, \sigma = 0, \underline{\tau}, \underline{\nu}, \underline{\xi}) \\ t \mapsto t+1 \end{cases} \tag{49}$$

obeys $\lim_{t\to\infty} m_t = 0$ (i.e., yields perfect recovery) iff $m_{t+1} < m_t \ \forall t$. Using (37)-(39), (48), and (49), we can see that this condition is equivalent to

$$\begin{aligned}
m_t &> \frac{1}{\delta}\mathsf{mse}(m_t; \underline{\nu}, \underline{\tau}, \underline{\xi}) = \frac{m_t}{\delta}\mathsf{mse}(\underline{\nu}^{1/\sqrt{m_t}}, \underline{\tau}, \underline{\xi}) && (50)\\
\Leftrightarrow \quad \delta &> \mathsf{mse}(\underline{\nu}^{1/\sqrt{m_t}}, \underline{\tau}, \underline{\xi}). && (51)
\end{aligned}$$

Since we assumed that $\underline{\tau}$ uses the minimax thresholds $\{\tau^{\pm}(\epsilon_s)\}_{s=1}^S$, (40) implies that the convergence condition (51) will hold for *any* $\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}$ whenever $\delta > M^{\pm}(\underline{\epsilon}, \underline{\xi})$. The following theorem then results from (41).

Theorem 1. *When the observations y are noiseless (i.e., $\sigma = 0$) and the thresholds $\underline{\tau}$ are minimax optimal, the SE formalism modeling AMP-NU converges to perfect signal recovery (i.e., $\lim_{t\to\infty} m_t = 0$) iff*

$$\sum_{s=1}^{S} \xi_s M^{\pm}(\epsilon_s) < \delta. \tag{52}$$

Theorem 1 implies a $(\underline{\xi}, \underline{r})$-dependent phase transition curve $\rho_{\text{SE}}(\delta)$ that bisects the $(\underline{\xi}, \underline{r})$-dependent admissible subregion of the sparsity-undersampling plane. In particular, for each $\delta \in [0, 1]$, the value $\rho_{\text{SE}}(\delta)$, if it exists, renders the pair $\big(\delta, \rho_{\text{SE}}(\delta)\big)$ admissible and satisfies

$$\delta = \sum_{s=1}^{S} \xi_s M^{\pm}\Big(\frac{\rho\delta}{\sum_{s'=1}^{S} \xi_{s'} r_{s',s}}\Big). \tag{53}$$

E. The Case of Noisy Observations

For the noisy case, the PTC is identical to the noiseless case. By explicit construction, the noise sensitivity can be made arbitrarily large above the PTC. Note that, in saying "above" the PTC, we restrict the domain to the admissible region of the sparsity-undersampling plane.

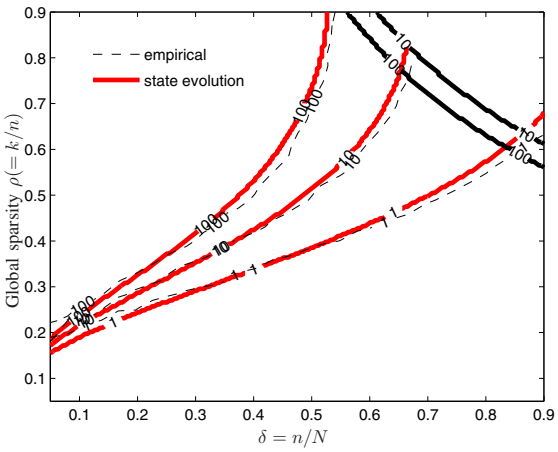

Fig. 3. Phase transition curves from state evolution (solid red) and empirical estimates (dashed black) for $S = 2$, $\xi_1 = \xi_2 = 1/2$, and $N = 1024$. Upper boundaries of the admissible region plane is shown in black. All curves are labeled according to the ratio $r = \epsilon_1/\epsilon_2$.

Proposition 1. *For any (δ, ρ) below the PTC,*

$$\begin{aligned}
&\inf_{\underline{\tau}} \sup_{\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}} \text{fMSE}(\underline{\tau}; \delta, \sigma = 1, \underline{\nu}) \\
&\qquad = \frac{M^{\pm}(\underline{\epsilon}, \underline{\xi})}{1 - M^{\pm}(\underline{\epsilon}, \underline{\xi})/\delta} \triangleq M^{\flat}(\underline{\epsilon}, \underline{\xi}, \delta). && (54)
\end{aligned}$$

Also, the threshold policy $\underline{\tau}^* = \underline{\tau}^{\pm}(\underline{\epsilon})$ minimaxes the formal MSE:

$$\sup_{\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}} \text{fMSE}(\underline{\tau}^*; \delta, \sigma = 1, \underline{\nu}) = \inf_{\underline{\tau}} \sup_{\underline{\nu} \in \mathcal{F}_{\underline{\epsilon}}} \text{fMSE}(\underline{\tau}; \delta, \sigma = 1, \underline{\nu}). \tag{55}$$

This proposition can be proved in a similar way to the proof [13] of (32)-(33) but by taking the corresponding vector quantities. In our case, the α-least favorable priors are parameterized by a vector $\underline{\mu}^{\pm}(\underline{\epsilon}, \alpha)$ rather than the scalar $\mu^{\pm}(\epsilon, \alpha)$.

IV. Numerical Results

A. The Case of Noiseless Observations

In case of non-uniform sparsity priors, the non-uniform minimax MSE $M^{\pm}(\underline{\epsilon}, \underline{\xi})$ completely defines the phase transition behavior of the state evolution at undersampling ratio δ. This state evolution phase transition agrees with empirically observed phase transition from AMP-NU algorithm. In Fig. 3, we considered the case $S = 2$ and $N_1 = N_2 = N/2$ with $N = 1024$. The global activity rate is $\epsilon = (\epsilon_1 + \epsilon_2)/2$. We chose different ratios $r = \epsilon_1/\epsilon_2 = 1, 10, 100$ for 30×30 grid on (δ, ρ) plane where $\epsilon = \rho\delta$. The dashed black lines represent empirically observed PTC and thick solid red lines represent the state evolution PTC $M^{\pm}(\underline{\epsilon}, \underline{\xi}) = \delta$. The PTC curve for $r = 10$ terminates at $\delta \approx 0.65$ because for larger values of δ, entire admissible region falls in the good region. The admissible region lies below the thick solid black lines.

We now study the non-uniform minimax MSE of the minimum risk thresholding policy $M^{\pm}(\underline{\epsilon}, \underline{\xi})$ in more detail.

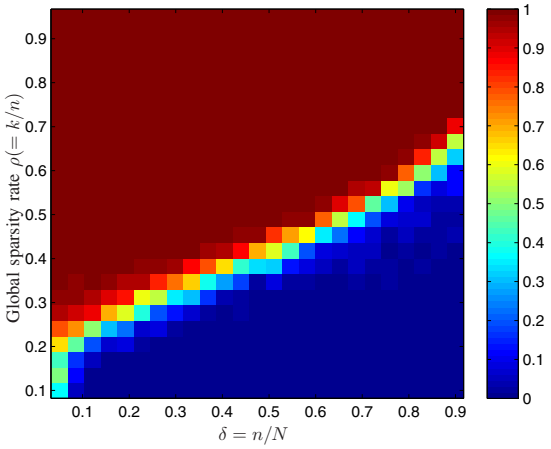

Fig. 4. Empirical phase diagram for $S = 1$, i.e., for uniform sparsity. The colormap shows fraction-of-failure from 100 trials at every point on a 30×30 grid.

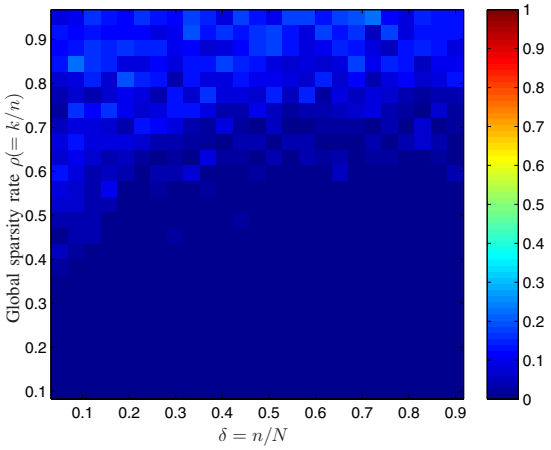

Fig. 5. Empirical phase diagram for $S = 2$, $\epsilon_1 = 1, \epsilon_2 = 0$, and $\xi_1 = \delta$, the extreme case of non-uniform sparsity. The colormap shows fraction-of-failure from 100 trials at every point on a 30×30 grid.

When $S = 1$, i.e., all coefficients have the same local activity rate ϵ, this reduces to the scalar minimax MSE $M^{\pm}(\epsilon)$. The PTC is the same as given in [7], [13]. This PTC can be seen in Fig. 3 corresponding to the ratio $r = 1$. Fig. 4 shows the corresponding phase diagram. We consider another extreme scenario when there are $S = 2$ subsets with $\epsilon_1 = 1$, and $\epsilon_2 = 0$. Then $M^{\pm}(\underline{\epsilon}, \underline{\xi}) = \xi_1$. From Theorem 1, perfect recovery happens *iff* $\xi_1 < \delta$. This is equivalent to the condition that the number of active coefficients is smaller than the number of observations. Fig. 5 shows the corresponding phase diagram.

B. The Case of Noisy Observations

Fig. 6 and 7 give the simulation results using the 0.02-least favorable signal for sparsity ratios $r = \epsilon_1/\epsilon_2 = 1$ and 100 respectively with $\xi_1 = \xi_2 = 1/2$ and $N = 1024$. The empirical

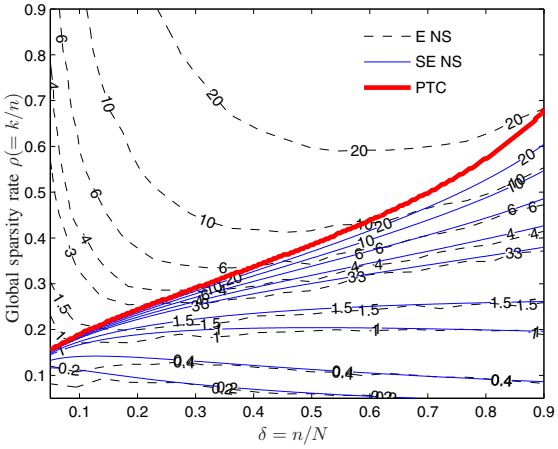

Fig. 6. For $\xi_1 = \xi_2 = 1/2$, $N = 1024$, and $r = \epsilon_1/\epsilon_2 = 1$, the empirical noise sensitivity contours, for AMP-NU using the 0.02-least favorable prior, are shown in dashed black. The PTC is shown as the thick red line, and $M^{\flat}(\underline{\epsilon}, \underline{\xi}, \delta)$ contours are shown as solid blue lines.

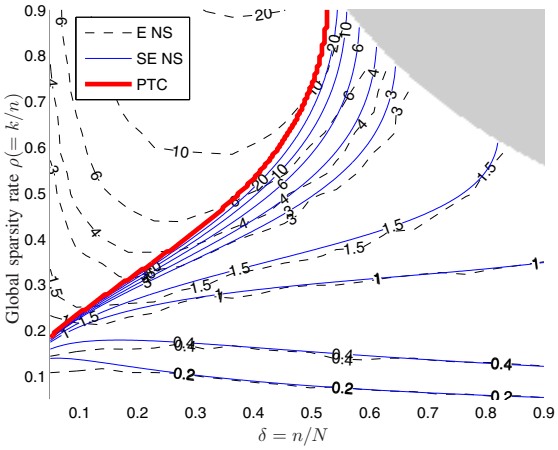

Fig. 7. For $\xi_1 = \xi_2 = 1/2$, $N = 1024$, and $r = \epsilon_1/\epsilon_2 = 100$, the empirical noise sensitivity contours, for AMP-NU using the 0.02-least favorable prior, are shown in dashed black. The PTC is shown as the thick red line, and $M^{\flat}(\underline{\epsilon}, \underline{\xi}, \delta)$ contours are shown as solid blue lines.

noise sensitivity contour plots from 100 trials are shown in solid blue and the corresponding $M^{\flat}(\underline{\epsilon}, \underline{\xi}, \delta)$ contour plots are shown in dashed black. The PTC is shown in thick solid red. The gray shaded region denotes the portion of the (δ, ρ) plane which is inadmissible for this choice of $\underline{\xi}$ and r. Fig. 8 shows the same phase behavior as Fig. 7 but for $N = 4096$. The empirical plots get closer to the state evolution predicted plots as the signal dimension increases.

V. Conclusion

In this paper, we considered the reconstruction of non-uniformly sparse signals from both noiseless and noisy observations, assuming that the activity rates are apriori known. First, we derived the minimax AMP algorithm for non-

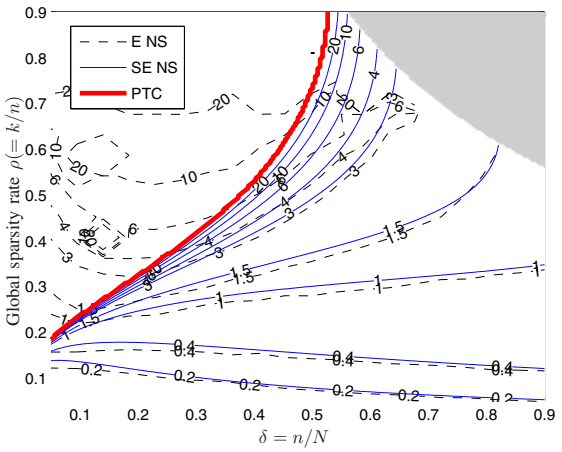

Fig. 8. For $\xi_1 = \xi_2 = 1/2$, $N = 4096$, and $r = \epsilon_1/\epsilon_2 = 100$, the empirical noise sensitivity contours, for AMP-NU using the 0.02-least favorable prior, are shown in dashed black. The PTC is shown as the thick red line, and $M^b(\underline{\epsilon}, \underline{\xi}, \delta)$ contours are shown as solid blue lines.

uniformly sparse signals (i.e., AMP-NU) and showed that it selects an appropriate threshold for each coefficient subset, depending on the activity of that subset. Next, we analyzed the behavior of AMP-NU usingthe state evolution (SE) formalism developed in [7], [13], and derived a noiseless phase transition curve (PTC) that predicts perfect recovery for (δ, ρ) values within the admissible region of the sparsity-undersampling plane. Both the PTC and the admissible region depend onthe relative sizes of the coefficient subsets and on the ratios of the coefficient activities. For the noisy case, we showed the existence of a sparse PTC that predicts whether the noise sensitivity will be finite for (δ, ρ) values within the admissible region of the sparsity-undersampling plane, and we showed that this noisy PTC coincides with the noiseless PTC. Then, for the region below the phase transition curve, the formal noise sensitivity was derived. Finally, some numerical results were presented to show that the empirical performance closely matches that of state evolution for large enough problem sizes.

REFERENCES

[1] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. 44th Annual Conf. Information Sciences and Systems*, March 2010.

[2] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior informa tion," Seoul, Korea, Jun. 2009, pp. 483–487.

[3] T. Tanaka and J. Raymond, "Optimal incorporation of sparsity information by weighted ℓ_1 optimization," Apr. 2010, arXiv:1001.1873v2.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998. [Online]. Available: http://link.aip.org/link/?SCE/20/33/1

[5] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal. Statist. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.

[6] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics in Signal Processing*, vol. 4, no. 2, pp. 330–341, 2010.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressive sensing," *Proc. National Academy of Sciences*, vol. 106, pp. 18 914–18 919, Nov 2009.

[8] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," arXiv:1001.3448, Jan 2010.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann, 1988.

[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. Information Theory Workshop*, Jan 2010.

[11] ——, "Message passing algorithms for compressed sensing: II. analysis and validation," in *Proc. Information Theory Workshop*, Jan 2010.

[12] A. Maleki and A. Montanari, "Analysis of approximate message passing algorithm," in *Proc. 44th Annual Conf. Information Sciences and Systems*, March 2010.

[13] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," arXiv:1004.1218, Apr 2010.

[14] D. L. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. National Academy of Sciences*, vol. 102, no. 27, pp. 9452–9457, 2005.

[15] ——, "Observed universality of phase transitions in high-dimensional ge ometry, with implications for modern data analysis and signal processing," *Phil. Trans. Royal Soc. A*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[16] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.