Denoiser Properties; An analysis

Chege Simon[1], Dr. Suman Mishra[2]

[1]M.Tech, ECE, SVCET (JNTUA Affiliated), RVS Nagar, 517127, AP, India

[2]Professor and HoD, ECE, SVCET (JNTUA Affiliated), RVS Nagar, 517127, AP, India

Abstract-*The main role of a denoising algorithm is to remove noise, errors or perturbations from a signal. A lot of research has been achieved in this area and therefore today's denoisers can effectively remove large amounts of additive noise. A compressive sensing (CS) reconstruction algorithm scheme seeks to recover a structured signal acquired using a relatively small number of randomized measurements. Typical CS reconstruction algorithms schemes can be cast as iteratively estimating a signal from a perturbed observation. There is an ongoing research on how to effectively employ a generic Denoiser in a CS reconstruction algorithm. The AMP reconstruction technique has proven to integrate with most denoisers (D-AMP)and offers an enhanced CS recovery performance while operating tens of times faster than competing methods. This paper seeks to look into an explanation of the exceptional performance of D-AMP by analyzing some of its theoretical properties and features.*

Keywords—*Denoising, Compressive Sensing, Approximate Message Passing*

I. INTRODUCTION

A denoiser's role is designed to estimate a signal $x_0$that is in a class of signals $C \subset \mathbb{R}^n$ from noisy observations, $x_0 + \sigma\epsilon$ where $\epsilon \sim N(0, I)$, and $\sigma > 0$ denotes the standard deviation of the noise. Let D_σ denote a family of denoisers indexed with the standard deviation of the noise. At every value of σ ,D_σ takes $x_0 + \sigma\epsilon$ as the input and returns an estimate ofx_0. [1]

To give a proper analysis of the D-AMP, it will be required that the denoisers' family to be *(near) proper*, *monotone*, and *Lipschitz continuous*. Since most denoisers easily satisfy these first two properties, and can be modified to satisfy the third, the requirements do not overly restrict the analysis.

II. DENOISER PROPERTIES

A. *Definition 1*

$D_\sigma$will be called a proper family of denoisers of level k $\left(k \in (0,1)\right)$ for the class of signals C if

$$\sup_{x_0 \in C} \frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \le k\sigma^2 \qquad (1)$$

for every $\sigma > 0$. Note that the expectation is with respect to $\epsilon \sim N(0, I)$.

The above definition can be clarified as under, consider the following examples:

Example 1: Let C denote a k-dimensional subspace of $\mathbb{R}^n$ ($k < n$). Also, let $D_\sigma(y)$ be the projection of y onto subspace C denoted by $P_C(y)$. Then,

$$\frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \le \frac{k}{n}\sigma^2 \qquad (2)$$

for every $x_0 \in C$ and every σ^2. Hence, this family of denoisers is proper of level k/n

Proof: First note that since the projection onto a subspace is a linear operator and since$P_C(x_0) = x_0$, we have

$$\mathbb{E}\|P_C(x_0+\sigma\epsilon)-x_0\|_2^2 = \mathbb{E}\|x_0+\sigma P_C(\epsilon)-x_0\|_2^2 = \sigma^2\mathbb{E}\|P_C(\epsilon)\|_2^2 \quad (3)$$

Also note that since $P_C^2 = P_C$, all the eigenvalues of P_C are either zero or one. Furthermore, since the null space of P is n - k dimensional, the rank of $P_C$is k. Hence, P_C has k eigenvalues equal to 1 and the rest are zero. Hence $\|P_C(\epsilon)\|_2^2$ follows a χ^2 distribution with k degrees of freedom and $\mathbb{E}\|P_C(x_0+\sigma\epsilon)-x_0\|_2^2 = k\sigma^2$.

Secondly consider a slightly more complicated example that has been popular in signal processing for the last twenty-five years. Let Γ_k denote the set of k-sparse vectors.

Example 2: Let $\eta(y;\tau\sigma) = (|y| - \tau\sigma) +$ sign (y) denote the family of soft-thresholding denoisers, then

$$\sup_{x_0 \in \Gamma_k} \frac{\mathbb{E}\|\eta(x_0+\sigma\epsilon;\tau\sigma)-x_0\|_2^2}{n} = \left[\frac{(1+\tau^2)k}{n} + \frac{n-k}{n}\mathbb{E}\left(\eta(\epsilon_1;\tau)\right)^2\right]\sigma^2 \quad (4)$$

Proof: For notational simplicity we assume that the first k coordinates of x_0 are non-zero and the rest are equal to zero.

$$\frac{\mathbb{E}\|\eta(x_0+\sigma\epsilon;\tau\sigma)-x_0\|_2^2}{n\sigma^2} = \frac{\sum_{k=1}^{k}\mathbb{E}\left(\eta\left(x_{0,i}+\sigma\epsilon_i;\tau\sigma\right)-x_{0,i}\right)^2}{n\sigma^2} + \frac{n-k}{n\sigma^2}\mathbb{E}\left(\eta(\sigma\epsilon_n;\tau\sigma)\right)^2$$

$$= \frac{\sum_{k=1}^{k}\mathbb{E}\left(\eta\left(\frac{x_{0,i}}{\sigma}+\epsilon_i;\tau\right)-\frac{x_{0,i}}{\sigma}\right)^2}{n} + \frac{n-k}{n}\mathbb{E}\left(\eta(\epsilon_n;\tau)\right)^2$$
(5)

Note that $\mathbb{E}\left(\eta\left(\frac{x_{0,i}}{\sigma}+\epsilon_i;\tau\right)-\frac{x_{0,i}}{\sigma}\right)^2$ is an increasing function of $\frac{x_{0,i}}{\sigma}$[59]. Therefore, it is straightforward to see that

$$\mathbb{E}\left(\eta\left(\frac{x_{0,i}}{\sigma}+\epsilon_i;\tau\right)-\frac{x_{0,i}}{\sigma}\right)^2 \le \lim_{x_{0,i\to\infty}} \mathbb{E}\left(\eta\left(\frac{x_{0,i}}{\sigma}+\epsilon_i;\tau\right)-\frac{x_{0,i}}{\sigma}\right)^2 = 1+\tau^2$$
(6)

where the last step swaps the lim and E (by the dominated convergence theorem). We obtain the desired result by combining (5) and (6).

It is to be noted that the optimal threshold τ to use within soft-thresholding depends on the sparsity $k/n$of the signal being denoised. One can optimize the parameter τ for every value of k/n and obtain an optimized family of denoisers. Figure 1 displays the level k of the optimized soft-thresholding in terms of k/n. Note that for sparse signals (k/nsmall) soft-thresholding is an effective denoiser and thus k is small. The previous denoisers both utilized prior knowledge about the structure of the signal (its dimensionality and its sparsity) in order to denoisex_0. When nothing is known about x_0 a proper denoiser might be too much to ask for. For instance, consider the maximum likelihood estimator.

Example 3: If $D_\sigma(x_0+\sigma\epsilon)$ is the maximum likelihood estimate of $x_0$from$x_0+\sigma\epsilon$, then

$$\frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n\sigma^2} = 1 \qquad (7)$$

So, these families of denoisers are not proper of level k for any k < 1. The proof of this statement is straightforward and hence has been shown that ([2, Ch. 5]) for any denoiser$\widehat{D}_\sigma$ we have

$$\sup_{x_0\in\mathbb{R}^n}\frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n\sigma^2} = 1 \qquad (8)$$

In this example the class of signals which have been considered is generic and hence the denoiser cannot employ any specific structure in x_0.There are occasions when theresearcher wants to deal with denoisers that are not proper because of an error/bias term that is independent of the noise level. To deal with scenarios such as these, the definition near proper is introduced as below.

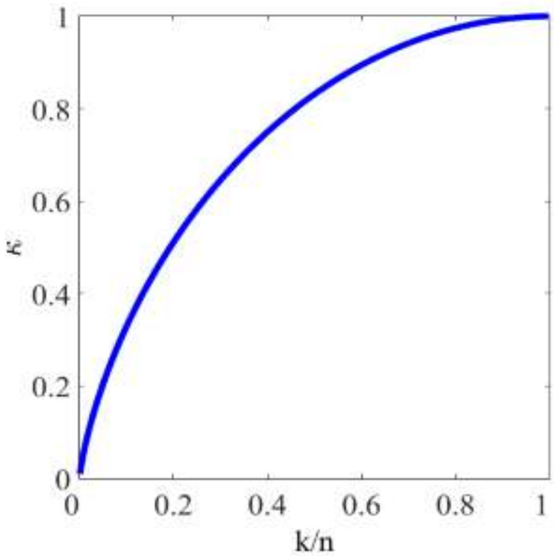

Fig. 1. The level k of optimal soft-thresholding method as a function of normalized sparsity For k/n sparse signals, soft-thresholding is a high performance denoiser.

B. *Definition 2*

$D_\sigma$is called a near proper family of denoisers of levels k $(k\in(0,1))$ and $B(B\in\mathbb{R}_+)$ for the class of signals C if

$$\sup_{x_0\in C}\frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \le k\sigma^2 + B \qquad (9)$$

for every $\sigma > 0$. Note that the expectation is with respect to ϵ ~ N(0, I).

As in Definition 1, the constants k and B determine the quality of the denoiser family. Better denoisers have smaller constants.

Example 4: Let $C_p = \{x \in \mathbb{R}^n : \|x\|_p \le 1\}$for some $0 \le p \le 1$. For a fixed k, let D_σ denote a denoiser that, through oracle information, knows the indices of the k largest elements of x and projects the noisy observation $x_0+\sigma\epsilon$ onto those coordinates. Then,

$$\sup_{x_0\in C}\frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \le \frac{k}{n}\sigma^2 + \frac{k^{1-2/p}}{n(2/p-1)}$$
(10)

for every $x_0\in C_p$ and every σ^2. Hence this family of denoisers is near proper with $k=\frac{k}{n}$ and $B=\frac{(k+1)^{1-2/p}}{n(2/p-1)}$

Proof: Let Λ denote the set of indices of the k-largest coefficients of x_0. For a vector x, define x_Λ in the following way, $x_{\Lambda,i} = x_i$ if $i\in\Lambda$ and otherwise $x_{\Lambda,i}=0$. Note that $x_{0,\Lambda}$is the best k-term approximation of x_0. We have

$$\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2 = \mathbb{E}\left\|\left(x_{0,\Lambda}+\sigma\epsilon_\Lambda\right)-x_0\right\|_2^2$$

$$= \left\|x_{0,\Lambda}-x_0\right\|_2^2 + \sigma^2\mathbb{E}\|\epsilon_\Lambda\|_2^2 \qquad (11)$$

Following the same logic as in example 1, it can be deduced that$\mathbb{E}\|\epsilon_\Lambda\|_2^2 = k$. The term $\left\|x_{0,\Lambda}-x_0\right\|_2^2$ above is simply the squared ℓ_2- norm of the smallest $n-k$ values ofx_0. Note that since $x_0 \in C_p$, we have

$$\sum_{i=1}^{n}\left|x_{0,i}\right|^p \le 1$$

In subsequent sections it is assumed that the signal in consideration belongs to a class C for which there is a proper or near proper family of denoisers D_σ . The class and denoiser can be very general. For instance, it can be assumed C to be the class of natural images and D_σ to denote the BM3D algorithm at different noise levels [3]

C. *Definition 3*

A denoiser is called Monotone if for everyx_0*, its risk function*

$$R(\sigma^2, x_0) = \frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \qquad (12)$$

is a non-decreasing function of σ^2.

Few remarks regarding monotone denoisers can be made as under;

Remark 1: Monotonicity is a natural property to expect from denoisers. Many standard denoisers such as soft-thresholding and group soft-thresholding are monotone if we optimize over the threshold parameter.

Remark 2: If a family of denoisers D_σ *is not monotone, then it is straightforward to construct a new denoiser that outperforms* D_σ*. Here is a simple proof. Suppose that for*$\sigma_1 < \sigma_2$ *we have*

$$R({\sigma_1}^2, x_0) > R({\sigma_2}^2, x_0)$$

Then construct a new denoiser for noise level σ_1 *in the following way:*

$$\widehat{D}_{\sigma_1}(y) = \mathbb{E}_{\hat{\epsilon}} D_{\sigma_2}\left(y+\sqrt{\sigma_2^2-\sigma_1^2}\,\hat{\epsilon}\right) \qquad (13)$$

Where $\hat{\epsilon} \sim N(0, I)$ *is independent of* y *and* $\mathbb{E}_{\hat{\epsilon}}\left(y+\sqrt{\sigma_2^2-\sigma_1^2}\,\hat{\epsilon}\right)$ *denotes the expected value with respect to* $\hat{\epsilon}$*. Let* $\hat{\sigma}_2 = \sqrt{\sigma_2^2-\sigma_1^2}$*. A simple application of Jensen's inequality shows that*

$$\frac{\mathbb{E}\left\|\widehat{D}_{\sigma_1}(x_0+\sigma_1\epsilon)-x_0\right\|_2^2}{n} = \frac{\mathbb{E}_\epsilon\left(\left\|\mathbb{E}_{\hat{\epsilon}} D_{\sigma_2}(x_0+\sigma_1\epsilon+\hat{\sigma}_2\hat{\epsilon})-x_0\right\|_2^2\right)}{n}$$

$$\le \frac{\mathbb{E}_{\epsilon,\hat{\epsilon}}\left(\left\|D_{\sigma_2}(x_0+\sigma_1\epsilon+\hat{\sigma}_2\hat{\epsilon})-x_0\right\|_2^2\right)}{n} \qquad (14)$$

Note that since $\hat{\epsilon}$ *and* ϵ *are independent,* $\frac{\mathbb{E}_{\epsilon,\hat{\epsilon}}\left(\left\|D_{\sigma_2}(x_0+\sigma_1\epsilon+\hat{\sigma}_2\hat{\epsilon})-x_0\right\|_2^2\right)}{n} = R({\sigma_2}^2, x_0)$*Therefore,*$\widehat{D}$ *improves* D *and does not violate the monotone property. Therefore, as is clear from this statement, non-monotone denoisers are not desirable in general since we can easily improve them..*

III. STATE EVOLUTION

A key ingredient in the analysis of D-AMP is the state evolution; a series of equations that predict the intermediate MSEof AMP algorithms at each iteration. Here an introduction of a new"deterministic" state-evolution to predict the performance ofD-AMP is done. Starting from $\theta^0 = \frac{\|x_0\|_2^2}{n}$the state evolution generatesa sequence of numbers through the following iterations:

$$\theta^{t+1}(x_0, \delta, \sigma_\omega^2) = \frac{1}{n}\mathbb{E}\left\|D_{\sigma^t}(x_0+\sigma^t\epsilon)-x_0\right\|_2^2 \qquad (15)$$

Where $(\sigma^t)^2 = \frac{\theta^t}{\delta}(x_0, \delta, \sigma_\omega^2) + \sigma_\omega^2$ and the expectation is with respect to $\epsilon \sim N(0, I)$. Note that the notation $\theta^{t+1}(x_0, \delta, \sigma_\omega^2)$ is set to emphasize that σ^t may depend on the signal x_0, the under-determinacy δ, and the measurement noise. Consider the iterations of D-AMP and let x^t denote its estimate at iteration t. The empirical findings show that the MSE of D-AMP is predicted accurately by the state evolution. The findings are therefore stated formally as below;

A. *Finding 1*

If the D-AMP algorithm starts from $x^0 = 0$*then for large values of* m *and* n*, state evolution predicts themean square error (MSE) of D-AMP, i.e.,*

$$\theta^t(x_0, \delta, \sigma_\omega^2) \approx \frac{1}{n}\|x^t - x_0\|_2^2 \qquad (16)$$

Based on extensive simulations, this findings are believed to be true if the following properties are satisfied:

i. The elements of the matrix **A** are i.i.d. Gaussian (or sub-Gaussian) with mean zero and standard deviation $1/m$.
ii. The noise ω is also i.i.d. Gaussian.
iii. The denoiser D is Lipchitz continuous. A denoiser is said to be L-Lipchitz

continuous if for every $x_1, x_2 \in C$ we have
$\|D(x_1) - D(x_2)\|_2^2 \le L\|x_1 - x_2\|_2^2$.

Most advanced image denoisers have no closed form expression, thus it is very hard to verify whether or not they are Lipschitz continuous. That said, every advanced denoisers tested was found to closely follow the state evolution equations (Finding 1), suggesting they are in fact Lipschitz.

In all the simulations the elements of A are i.i.d. Gaussian. The same is true for the elements of ω.

Figure 2 compares the state evolution predictions of D-AMP (based on the BM3D denoising algorithm [39]) with the empirical performance of D-AMP and D-IT. As is clear from this figure, the state evolution is accurate for D-AMP but not for D-IT. The validity of the above findings for the following denoising algorithms has been checked: (i) BM3D [3], (ii) BLS-GSM [4], (iii) Non-local means [35], (iv) AMP with soft-wavelet-thresholding [5], [6].

In this work, it's assumed that the state evolution is accurate for D-AMP and derive some of the main features of D-AMP based on this assumption.

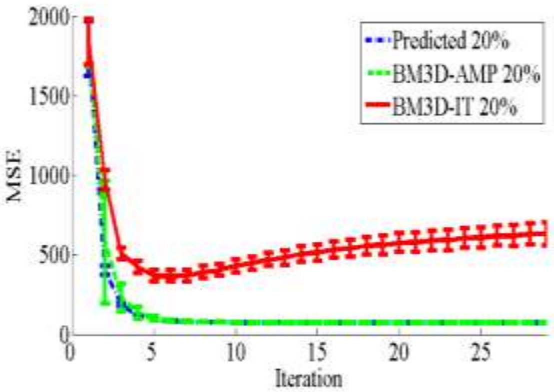

Fig.2. The MSE of the intermediate estimate versus the iteration count forBM3D-AMP and BM3D-IT alongside their predicted state evolution. Notice that BM3D-AMP is well predicted by the state evolution whereas BM3D-IT is not.

IV. ANALYSIS OF D-AMP IN THE ABSENCE OF MEASUREMENT NOISE

In this section we consider the noiseless setting $\sigma_\omega^2 = 0$ and characterize the number of measurements D-AMP requires (under the validity of the state evolution framework) to recover the signal x_0 exactly. We consider monotone denoisers, as defined in definition 3. Consider the state evolution equation under the noiseless setting $\sigma_\omega^2 = 0$:

$$\theta^{t+1}(x_0, \delta, \sigma_\omega^2) = \frac{1}{n}\mathbb{E}\left\|D_{\sigma^t}(x_0 + \sigma^t\epsilon) - x_0\right\|_2^2$$

(17)

where $(\sigma^t)^2 = \frac{\theta^t}{\delta}(x_0, \delta, 0)$. Starting with $\theta^0(x_0, \delta, 0) = \frac{\|x_0\|_2^2}{n}$ depending on the value of δ there are two conceivable scenarios for the state evolution equation:

i. $\theta^t(x_0, \delta, 0) \to 0$ as $t \to \infty$
ii. $\theta^t(x_0, \delta, 0) \nrightarrow 0$ as $t \to \infty$

(18)

$\theta^t(x_0, \delta, 0) \to 0$ implies the success of D-AMP algorithm, while $\theta^t(x_0, \delta, 0) \nrightarrow 0$ implies its failure in recovering x_0. The main goal of this section is to study the success and failure regions.

A. Lemma 1

For monotone denoisers, if for δ_0 , $\theta^t(x_0, \delta, 0) \to 0$ *then for any* $> \delta_0$, $\theta^t(x_0, \delta, 0) \to 0$ *as well.*

Proof: Define $(\sigma^t)^2 = \frac{\theta^t}{\delta}(x_0, \delta, \sigma_\omega^2)$. Clearly, since $\theta^t(x_0, \delta, \sigma_\omega^2) \to 0$, so does σ^t. The first claim is that for every $\sigma^2 < \frac{\|x_0\|_2^2}{n\delta_0} = (\sigma^0)^2$ (this is where D-AMP is initialized) we have

$$\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0 + \sigma^t\epsilon) - x_0\|_2^2 < \sigma^2, \quad \forall \sigma^2 > 0$$

(19)

Suppose that this is not true and define

$$\sigma_*^2 = \sup_{\sigma^2 \le \frac{\|x_0\|_2^2}{n\delta_0}} \left\{\sigma^2 : \frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0 + \sigma^t\epsilon) - x_0\|_2^2 \ge \sigma^2\right\}$$

(20)

Claim that if $\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0 + \sigma^0\epsilon) - x_0\|_2^2 < (\sigma^0)^2$, then $\sigma^t \to \sigma_*$ as $t \to \infty$. First, it can be concluded that $\frac{1}{n\delta_0}\mathbb{E}\left\|D_{\sigma_*}(x_0 + \sigma_*\epsilon) - x_0\right\|_2^2 = {\sigma_*}^2$. For $\sigma > \sigma_*$, its known that $\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0 + \sigma\epsilon) - x_0\|_2^2 < \sigma^2$.

By using the Monotonicity of the denoiser, then for every $\sigma \ge \sigma_*$,

$$\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0 + \sigma\epsilon) - x_0\|_2^2 \le \frac{1}{n\delta_0}\mathbb{E}\left\|D_{\sigma_*}(x_0 + \sigma_*\epsilon) - x_0\right\|_2^2 = {\sigma_*}^2$$

This (through simple induction) implies that for every t, $(\sigma^t)^2 \ge (\sigma_*)^2$. Furthermore according to the definition of ${\sigma_*}^2$ and the fact that $\sigma^t > \sigma_*$, we have

$$(\sigma^{t+1})^2 = \frac{1}{n\delta_0}\mathbb{E}\left\|D_{\sigma^t}(x_0 + \sigma^t\epsilon) - x_0\right\|_2^2 \le (\sigma^t)^2$$

Therefore, σ^{t+1} is a decreasing sequence with lower bound σ_*. Hence, σ^t converges as $\sigma^\infty \ge \sigma_*$. The last step is to show that $\sigma^\infty = \sigma_*$. If this is not the case, then $\sigma^\infty > \sigma_*$. But according to the definition of σ_* and the supposition that $\sigma^\infty > \sigma_*$ we have,

$$\frac{1}{n\delta_0}\mathbb{E}\|D_{\sigma^\infty}(x_0+\sigma^\infty\epsilon)-x_0)\|_2^2 \le (\sigma^\infty)^2$$

which is a contradiction to σ^∞ being a fixed point. Hence, $\sigma^\infty = \sigma_*$. Since$\sigma^\infty = 0$, then it can be concluded that $\sigma_* = 0$, Thus,

$$\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2 < \sigma^2, \quad \forall\, \sigma^2 > 0$$

Since $> \delta_0$, then

$$\frac{1}{n\delta}\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2 < \sigma^2, \quad \forall\, \sigma^2 > 0 \qquad (21)$$

Hence the only fixed point of this equation is also at zero and hence $\theta^t(x_0,\delta,0) \to 0$. Note that all the above argument is based on the assumption that $\frac{1}{n\delta_0}\mathbb{E}\|D_\sigma(x_0+\sigma^0\epsilon)-x_0\|_2^2 < (\sigma^0)^2$

Note that for very small values of δ , it is straightforward to see that $\theta^t(x_0,\delta,0) \nrightarrow 0$ as $t\to\infty$ If we combine this result with Lemma 1 then the following simple result can be made: For small values of δ D-AMP fails in recovering x_0. As δ increases, after a certain value of δ D-AMP will successfully recover x_0 from its under-sampled measurements. Define

$$\delta^*(x_0) = \inf_{\delta\in(0,1)}\{\delta : \ \theta^t(x_0,\delta,0)\to 0 \ \text{ as } t\to\infty\} \qquad (22)$$

$\delta^*(x_0)$denotes the minimum number of measurements required for the successful recovery of x_0. Our goal is to characterize $\delta^*(x_0)$ in terms of the performance of the denoising algorithm. However, since the number of measurements $\delta^*(x_0)$ depends on the signal x_0, a more natural question in the design of a system is the following: How many measurements does D-AMP require to recover every signal $x_0 \in C$. The following result addresses this question.

B. Proposition 1

Suppose that for signal class C the denoiser D_σ is proper at level κ. Then

$$\sup_{x_0\in C}\delta^*(x_0) \le \kappa \qquad (23)$$

Proof: The proof of this proposition is a simple application of the state evolution equation. Similar to the proof of Lemma 1 define

$$\left(\sigma^t(x_0,\delta,\sigma_\omega^2)\right)^2 = \frac{\theta^t(x_0,\delta,\sigma_\omega^2)}{\delta} \qquad (24)$$

Also for notational simplicity we use the notation σ^t instead of $\sigma^t(x_0,\delta,0)$ in the equation below. According to state evolution we have

$$(\sigma^{t+1})^2 = \frac{1}{n\delta}\mathbb{E}\left\|D_{\sigma^t}(x_0+\sigma^t\epsilon)-x_0\right\|_2^2$$
$$= \frac{(\sigma^t)^2}{n\delta(\sigma^t)^2}\mathbb{E}\left\|D_{\sigma^t}(x_0+\sigma^t\epsilon) - x_0\right\|_2^2$$
$$\le \frac{(\sigma^t)^2}{\delta}\sup_{x_0\in C}\frac{\mathbb{E}\left\|D_{\sigma^t}(x_0+\sigma^t\epsilon)-x_0\right\|_2^2}{n(\sigma^t)^2}$$
$$\le \frac{\kappa(\sigma^t)^2}{\delta} \qquad (25)$$

It is straightforward to see that

$$\left(\sigma^t(x_0,\delta,0)\right)^2 \le \left(\frac{\kappa}{\delta}\right)^t\left(\sigma^0(x_0,\delta,0)\right)^2 \qquad (26)$$

Here, if $\delta > \kappa$ then $\left(\sigma^t(x_0,\delta,0)\right)^2 \to 0$ as $t\to\infty$.

Proposition 1 can be applied to the examples of Denoiser properties and derive some well-known results, such as the phase transition of AMP with the soft-threshold denoiser. If the denoiser is only nearly proper, perfect recovery may not be possible. However, the same technique to bound the recovery error of D-AMP can be implemented.

C. *Lemma 1*

Let $D_\sigma$denote a near proper family of denoiserswith levels κ and B, as defined in definition 2. Then, if $\delta > \kappa$the error of D-AMP is upper bounded by

$$\lim_{t\to\infty}\left(\sigma^t(x_0,\delta,0)\right)^2 \le \frac{B}{\delta-\kappa} \qquad (27)$$

Proof: The proof of this result is much like the one used for proper denoisers. Again define $\left(\sigma^t(x_0,\delta,\sigma_\omega^2)\right)^2 = \frac{\theta^t(x_0,\delta,\sigma_\omega^2)}{\delta}$. Using the state evolution and the definition of near proper wehave

$$\left(\sigma^{t+1}(x_0,\delta,0)\right)^2 = \frac{1}{n\delta}\mathbb{E}\left\|D_{\sigma^t(x_0,\delta,0)}(x_0 + \sigma^t(x_0,\delta,0)\epsilon) - x_0\right\|_2^2$$
$$\le \kappa\frac{\left(\sigma_t(x_0,\delta,0)\right)^2 + B}{\delta} \qquad (28)$$

Hence,

$$\left(\sigma^t(x_0,\delta,0)\right)^2 \le \left(\frac{\kappa}{\delta}\right)^t\frac{\|x_0\|_2^2}{n} + \frac{1-(\kappa/\delta)^t}{1-\kappa/\delta}\frac{B}{\delta} \qquad (29)$$

For $\delta > \kappa$, the limit of the sequence is as follows,

$$\lim_{t\to\infty}\left(\sigma^t(x_0,\delta,0)\right)^2 \le \frac{B}{\delta-\kappa} \qquad (30)$$

Note that the proof techniques employed above was first developed in [5] and was later employed to establish the phase transition of AMP extensions [7]. There are some minor differences between this work's derivation and the derivations presented in the other works since this work has not adopted the mini-max setting.

V. NOISE SENSITIVITY OF D-AMP

In section IV, the performance of D-AMP in the noiseless setting where $\sigma_\omega^2 = 0$ was considered.

This section will be devoted to the analysis of D-AMP in the presence of the measurement noise. Here it is assumed that the denoiser is near proper at levels κ and B, i.e.,

$$\sup_{\sigma^2} \sup_{x_0 \in C} \frac{\mathbb{E}\|D_\sigma(x_0+\sigma\epsilon)-x_0\|_2^2}{n} \leq \kappa\sigma^2 + B \quad (31)$$

The following result shows that D-AMP is robust to the measurement noise. Let $\theta^\infty(x_0, \delta, \sigma_\omega^2)$ denote the fixed point of the state evolution equation. Since there is measurement noise, $\theta^\infty(x_0, \delta, \sigma_\omega^2) \neq 0$, i.e., D-AMP will not recover x_0 exactly. We define the noise sensitivity of D-AMP as

$$NS(\delta, \sigma_\omega^2) = \sup_{x_0 \in C} \theta^\infty(x_0, \delta, \sigma_\omega^2) \quad (32)$$

The following proposition provides an upper bound for the noise sensitivity as a function of the number of measurements and the variance of the measurement noise.

A. Proposition 2

Let$D_\sigma$denote a near proper family ofdenoisers at levels κ and B. Then, for $\delta > \kappa$ the noisesensitivity of D-AMP satisfies

$$NS(\delta, \sigma_\omega^2) \leq \frac{\kappa\sigma_\omega^2 + B}{1-\kappa/\delta} \quad (33)$$

Proof: Note that $\theta^\infty(x_0, \delta, \sigma_\omega^2)$ is a fixed point of the state evolution equation and hence it satisfies

$$\theta^\infty(x_0, \delta, \sigma_\omega^2) = \frac{1}{n\delta}\mathbb{E}\|D_{\sigma^\infty}(x_0 + \sigma^\infty(x_0, \delta, \sigma_\omega^2)\epsilon) - x_0\|_2^2 \quad (34)$$

Where $\theta^\infty(x_0, \delta, \sigma_\omega^2) = \sqrt{\theta^\infty(x_0, \delta, \sigma_\omega^2)/\delta + \sigma_\omega^2}$. Therefore,

$$NS(\delta, \sigma_\omega^2) = \sup_{x_0 \in C} \theta^\infty(x_0, \delta, \sigma_\omega^2)$$

$$= \sup_{x_0 \in C} \frac{1}{n}\mathbb{E}\|D_{\sigma^\infty}(x_0 + \sigma^\infty(x_0, \delta, \sigma_\omega^2)\epsilon) - x_0\|_2^2$$

$$\leq \sup_{x_0 \in C} \kappa\left(\sigma^\infty(x_0, \tau, \sigma_\omega^2)\right)^2 + B$$

$$= \sup_{x_0 \in C} \kappa\left(\frac{\sigma^\infty(x_0, \delta, \sigma_\omega^2)}{\delta} + \sigma_\omega^2\right) + B$$

$$= \frac{\kappa}{\delta} NS(\delta, \sigma_\omega^2) + \kappa\sigma_\omega^2 + B \quad (35)$$

Substituting in $B = 0$ into the above result gives the noise sensitivity for proper denoisers.

$$NS(\delta, \sigma_\omega^2) \leq \frac{\kappa\sigma_\omega^2}{1-\kappa/\delta} \quad (36)$$

There are several interesting features of this proposition that we would like to emphasize.

D. Remark 1

The bound presented in Proposition 2 is a worst case analysis. The bound may be achieved for certain signals in C and certain noise variances. However, for most signals in C and most noise variances D-AMP will perform better than what is predicted by the bound. Figure 3 shows the performance of BM3D-AMP in terms of the standard deviation of the noise.

The technique employed above was first developed in [8]. The result derived in Proposition 2 can be considered as a generalization of the result of [8] to much broader class of denoisers. As an aside, upper and lower bounds were recently derived for the mini-max noise sensitivity of anyrecovery algorithm when the measurement matrix is i.i.d. Gaussian and the compressively sampled signal is sparse [9]. Note that while the results can be applied to sparse signals, they have been derived under much more general setting.

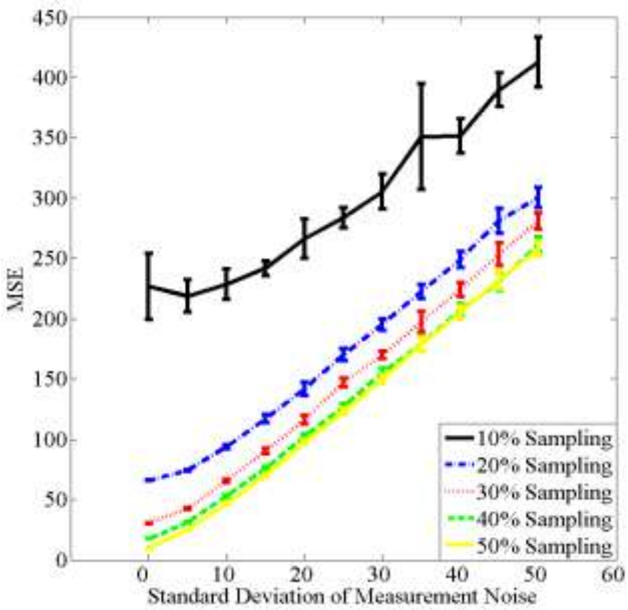

Fig. 3. The MSE of BM3D-AMP reconstructions of 128 × 128 Barbara test image with varying amounts of measurement noise at different sampling rates (δ).

VI. ADDITIONAL MISCELLANEOUS PROPERTIES OF D-AMP

A. Better Denoisers Lead to Better Recovery;

This intuitive result is a key feature of D-AMP. We formalize it below.

Theorem 1: Let a family of denoisers D_σ^1*be a betterdenoiser than a family* D_σ^2*for signal* x_0*in the following sense:*

$$\frac{\mathbb{E}\|D_\sigma^1(x_0+\sigma\epsilon)-x_0\|_2^2}{n\sigma^2} \le \frac{\mathbb{E}\|D_\sigma^2(x_0+\sigma\epsilon)-x_0\|_2^2}{n\sigma^2}, \quad \forall\sigma^2 > 0 \tag{37}$$

Also,$\sigma_{D_i}^{\infty}(x_0,\delta,\sigma_\omega^2)$ *let denote the fixed point of state evolutionfor denoiser* D_i*. then,*$\sigma_{D_1}^{\infty}(x_0,\delta,\sigma_\omega^2) \le \sigma_{D_2}^{\infty}(x_0,\delta,\sigma_\omega^2)$

Proof: The proof of this result is straightforward. Since, the state evolution of D^1 is uniformly lower thanD^2, its fixed point is lower as well.

B. D-AMP as a Regularization Technique

Explicit regularization is a popular technique to recover signals from an under-sampled set of linear measurements [10]. In these approaches a cost function, $\boldsymbol{J}(x)$, also known as a regularized, is considered on $\mathbb{R}^n$. This function returns large values for $x \notin C$and returns small values for $x \in C$. Regularized techniques recover $x_0$from measurements y by setting up and solving the following optimization problem:

$$\hat{x} = \text{argmin}_x \frac{1}{2}\|y - \boldsymbol{A}x\|_2^2 + \lambda\boldsymbol{J}(x) \tag{38}$$

Since in many cases $\boldsymbol{J}(x)$ is non-convex and non-differentiable, iterative heuristic methods have been proposed for solving the above optimization problem. D-AMP provides another heuristic approach for solving (38). It has two main advantages over the other heuristics:

a) D-AMP can be analyzed by the state evolution theoretically. Hence, we can theoretically predict the number of measurements required and the noise sensitivity of D-AMP.
b) The performances of most heuristic methods depend on their free parameters.

VII. CONCLUSIONS

Through extensive testing, it has been demonstrated that theapproximate message passing (AMP) compressed sensing recovery algorithm can be extended to use arbitrary denoisers to great effect. Variations of this denoising-based AMP algorithm (D-AMP) deliver state-of-the-art compressively sampled image recovery performance while maintaining a low computational footprint. The theoretical results and simulations showthat the performance of D-AMP can be predicted accurately by state evolution. Finally, can be shown that D-AMP is extremely robust to measurement noise. D-AMP represents a plug and play method to recover compressively sampled signals of arbitrary class; simply choose a denoiser well matched to the signal model and plug it in the AMP framework. Since designing denoising algorithms that employ complicated structures is usually much easier than designing recovery algorithms, D-AMP can benefit many different application areas.

ACKNOWLEDGMENT

Thanking the management of SVCET(JNTUA affiliated) for the support, especially Dr. Suman Mishra for his dedicated efforts to see this work complete. Thanking the scholarship provider (ICCR) for the chance and funding. Special thanks to my classmates for their encouragement throughout the research work and lastly, the anonymous reviewers for the constructive comments and suggestions.

REFERENCES

[1] C. A. Metzler, A. Maleki, and R. G. Baraniuk. (Apr. 2014). “From denoising to compressed sensing.” [Online].Available: http://arxiv.org/abs/1406.4175

[2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simon celli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[3] A. Buades, B. Coll, and J.-M. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[4] U. S. Kamilov, S. Rangan, M. Unser, and A. K. Fletcher, “Approximate message passing with consistent parameter estimation and applications to sparse learning,” in *Proc. Adv. Neural Process. Syst. (NIPS)*, 2012, pp. 2438–2446.

[5] C. Chen, E. W. Tramel, and J. E. Fowler, “Compressed-sensing recovery of images and video using multihypothesis predictions,” in *Proc. 45th Asilomar Conf. Signals Syst. Comput.*, Nov. 2011, pp. 1193–1198.

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

[7] S. Kudekar and H. D. Pfister, “The effect of spatial coupling on compressive sensing,” in *Proc. 48th Annu. Allerton Conf. Commun.,Control, Comput.*, Sep./Oct. 2010, pp. 347–353.

[8] K. Egiazarian, A. Foi, and V. Katkovnik, “Compressed sensing image reconstruction via recursive spatially adaptive filtering,” in *Proc. IEEEInt. Conf. Image Process. (ICIP)*, vol. 1. Sep./Oct. 2007, pp. I-549–I-552.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Trans.Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[10] S. Mun and J. E. Fowler, “Block compressed sensing of images using directional transforms,” in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 3021–3024.