Optimal Recovery from Compressive Measurements via Denoising-based Approximate Message Passing

Christopher A. Metzler[1], Arian Maleki[2], and Richard G. Baraniuk[1]

[1]Department of Electrical and Computer Engineering at Rice University
[2]Department of Statistics at Columbia University

***Abstract*—Recently progress has been made in compressive sensing by replacing simplistic sparsity models with more powerful denoisers. In this paper, we develop a framework to predict the performance of denoising-based signal recovery algorithms based on a new deterministic state evolution formalism for approximate message passing. We compare our deterministic state evolution against its more classical Bayesian counterpart. We demonstrate that, while the two state evolutions are very similar, the deterministic framework is far more flexible. We apply the deterministic state evolution to explore the optimality of denoising-based approximate message passing (D-AMP). We prove that, while D-AMP is suboptimal for certain classes of signals, no algorithm can uniformly outperform it.**

***Index Terms*—Denoising, Approximate Message Passing, State Evolution, Compressed Sensing**

I. Introduction

Compressive sensing (CS) is a new sampling paradigm that uses prior knowledge, beyond just bandlimitedness, to sample a signal more efficiently. A CS system takes a series of m linear measurements of a length n signal x_o. This process can be written as the system of equations $y = \mathbf{A}x_o + w$, where y is a length m vector containing the measurements, $\mathbf{A}$ is an $m \times n$ measurement matrix that models the measurement process, x_o is a length n vector representing the signal of interest, and w is a vector representing measurement noise. A CS recovery algorithm then uses prior knowledge about x_o to recover $\hat{x}$ from this underdetermined system [1], [2].

Most existing work in CS has assumed that the signal x_o has transform-domain sparsity. Recent work has made a fundamentally different assumption [3]–[9]. Rather than assuming that x_o is bandlimited or sparse, we assume that we know how to denoise it. That is, we assume that x_o belongs to class C and that we have a denoiser designed to denoise signals of class C. This line of reasoning has enabled the development of the denoising-based approximate message passing (D-AMP) algorithm [3]–[5], which exhibits state-of-the-art recovery of compressively sampled images.

D-AMP's exceptional performance raises questions: Can CS recovery algorithms be further improved? Is CS recovery a solved problem?

Answering these questions requires a method to predict recovery performance on an arbitrary problem. Potentially, one could predict D-AMP's performance using the approximate message passing (AMP) algorithm's Bayesian state evolution (SE): a framework for predicting an AMP algorithm's expected mean squared error (MSE) at each iteration [10], [11]. Unfortunately, the Bayesian SE is inadequate for this purpose: It requires explicitly knowing the probability density function (pdf) of the sampled signal. For most problems of interest, like imaging, we do not have an accurate model for the sampled signal's distribution.

To circumvent this obstacle, we develop a new deterministic SE. Like the traditional Bayesian SE, our deterministic SE predicts the expected MSE of D-AMP. However, unlike the Bayesian SE, the deterministic SE predicts the recovery error of a particular signal; no signal pdfs are involved. This distinction makes the deterministic SE applicable to a far wider set of problems.

Armed with this flexible deterministic SE, we analyze the optimality of D-AMP. We first demonstrate that D-AMP cannot be uniformly improved; for some problems, no algorithm can recover a signal from fewer measurements than the amount required by D-AMP. We then demonstrate that there exist problems for which, even with an optimal denoiser, D-AMP is suboptimal; D-AMP is not a panacea for CS recovery.

II. D-AMP

D-AMP extends AMP by replacing its thresholding operator with a more general denoiser [3]–[5].[1] D-AMP can be written as follows:

$$\begin{aligned} x^{t+1} &= D_{\hat{\sigma}^t}(x^t + \mathbf{A}^* z^t), \\ z^t &= y - \mathbf{A}x^t + z^{t-1} \mathrm{div} D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{A}^* z^{t-1})/m, \\ (\hat{\sigma}^t)^2 &= \frac{\|z^t\|_2^2}{m}, \end{aligned} \tag{1}$$

where, x^t and z^t are the estimates of x_o and the residue $y - Ax_o$ at iteration t, respectively. The term $\mathbf{A}^*$ denotes the conjugate transpose of $\mathbf{A}$, and $\mathrm{div} D_{\hat{\sigma}^{t-1}}$ denotes the

[1]D-AMP was not the first algorithm to use denoisers in AMP, see [12] and [6]. However, it uses far more general denoisers than previous efforts.

The work of C. Metzler supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. The work of C. Metzler and R. Baraniuk was supported in part and by the grants NSF CCF-0926127, NSF CCF-1117939, ONR N00014-12-1-0579, ONR N00014-11-1-0714, and ARO MURI W911NF-09-1-0383. The work of A. Maleki was supproted by the grant NSF CCF-1420328.

978-1-4673-7353-1/15/$31.00 ©2015 IEEE

divergence of the denoiser.[2] The term $z^{t-1}\text{div}D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{A}^* z^{t-1})/m$ is known as the *Onsager correction term.* As in the original AMP algorithm, $x^t + \mathbf{A}^* z^t$ can be written as $x_o + v^t$, where v^t is known as the effective noise. We can estimate the variance of v^t with $(\hat{\sigma}^t)^2 = \frac{\|z^t\|_2^2}{m}$ [13]. D-AMP works well because the Onsager correction term Guassianizes v^t, and because typical denoisers are designed to handle additive white Gaussian noise.

For more information on the D-AMP algorithm, including information on the Gaussianity of the effective noise and how to approximate the Onsager correction term, see [4].

III. State Evolution

The state evolution (SE) refers to a series of equations that predict the intermediate MSE of AMP algorithms at each iteration. Here we introduce D-AMP's deterministic SE framework and compare it with AMP's Bayesian framework, which was first introduced in [10], [11].

A. Deterministic state evolution

The "deterministic" SE assumes that x_o is an arbitrary but fixed vector in C. Starting from $\theta^0 = \frac{\|x_o\|_2^2}{n}$, the deterministic SE generates a sequence of numbers through the following iterations:

$$\theta^{t+1}(x_o, \delta, \sigma_w^2) = \frac{1}{n}\mathbb{E}_\epsilon \|D_{\sigma^t}(x_o + \sigma^t \epsilon) - x_o\|_2^2, \qquad (2)$$

where $(\sigma^t)^2 = \frac{1}{\delta}\theta^t(x_o, \delta, \sigma_w^2) + \sigma_w^2$, the scalar σ_w represents the standard deviation of the measurement noise w, and $\epsilon \sim N(0, I)$. Note that our notation $\theta^{t+1}(x_o, \delta, \sigma_w^2)$ emphasizes that θ^t may depend on x_o, the under-determinacy δ, and the measurement noise. Consider the iterations of D-AMP and let x^t denote its estimate at iteration t. Our empirical findings, presented in [4], show that, when the below assumptions are satisfied, the MSE of D-AMP is predicted accurately by the SE (2). We formally state our finding.

Finding 1. *Assume the following: (i) The elements of the matrix* $\mathbf{A}$ *are i.i.d. Gaussian with mean zero and standard deviation* $1/m$. *(ii) The noise* w *is also i.i.d. Gaussian. (iii) The denoiser* D *is Lipschitz continuous.*[3] *Under these conditions, if the D-AMP algorithm starts from* $x^0 = 0$*, then for large values of* m *and* n*, the SE predicts the mean square error of D-AMP, i.e.,*

$$\theta^t(x_o, \delta, \sigma_w^2) \approx \frac{1}{n}\|x^t - x_o\|_2^2.$$

In testing we found the above deterministic SE (2) predictions were accurate to within 1% for AMP based on soft-wavelet-thresholding [11], [14] and for D-AMP based on the NLM [15], BLS-GSM [16], and BM3D [17] denoisers. Because hard-thresholding is not Lipschitz, the SE predictions were inaccurate for D-AMP based on hard-wavelet-thresholding. However, we found the SE prediction held if we smoothed the hard-thresholding operator (see Section VI of [4]).

Finding 1 is important, because it enables us to easily analyze the performance of D-AMP for an arbitrary signal x_o. For additional details on the simulations that lead to this finding, see Sections III.C and VII.C of [4]. We have posted our code online[4] to let other researchers check the validity of our findings in more general settings and explore the validity of this conjecture on a wider range of denoisers.

B. Bayesian state evolution

The Bayesian SE assumes that x_o is a vector drawn from a pdf p_x, where the support of p_x is a subset of C. Starting from $\bar{\theta}^0 = \frac{\|x_o\|_2^2}{n}$, the Bayesian SE generates a sequence of numbers through the following iterations:

$$\bar{\theta}^{t+1}(p_x, \delta, \sigma_w^2) = \frac{1}{n}\mathbb{E}_{x_o,\epsilon}\|D_{\bar{\sigma}^t}(x_o + \bar{\sigma}^t\epsilon) - x_o\|_2^2, \qquad (3)$$

where $(\bar{\sigma}^t)^2 = \frac{1}{\delta}\bar{\theta}^t(p_x, \delta, \sigma_w^2) + \sigma_w^2$. Note the important difference between this SE equation and the one we discussed previously: in (3), the expected value is with respect to $\epsilon \sim N(0, I)$ and $x_o \sim p_x$, while in (2), x_o was considered an arbitrary but fixed vector in C. We have used the notation $\bar{\theta}$ to distinguish the Bayesian SE from its deterministic counterpart.

C. Connection between the two state evolutions

While the deterministic and Bayesian SEs are different, we can establish a connection between them by employing standard results in theoretical statistics regarding the connection between the minimax risk and the Bayesian risk. Let $\bar{\theta}^\infty(p_x, \delta, \sigma_w^2)$ denote the fixed point of the Bayesian SE (3) for the distribution p_x. Also, let $\theta^\infty(x_o, \delta, \sigma_w^2)$ denote the fixed point of deterministic SE (2) for the family of minimax denoisers. (See Definition 3 in Section IV-C for a formal definition of a minimax denoiser.)

Theorem 1. *Let* $\mathcal{P}$ *denote the set of all distributions whose support is a subset of* C*. Then,*

$$\sup_{p_x \in \mathcal{P}} \bar{\theta}^\infty(p_x, \delta, \sigma_w^2) \leqslant \sup_{x_o \in C} \theta^\infty(x_o, \delta, \sigma_w^2).$$

Proof. For an arbitrary family of denoisers D_σ we have

$$\mathbb{E}_{x_o,\epsilon}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2 \leqslant \sup_{x_o \in C} \mathbb{E}_\epsilon \|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2. \quad (4)$$

If we take the minimum with respect to D_σ on both sides of (4), we obtain the following inequality

$$\mathbb{E}_{x_o,\epsilon}\|\tilde{\eta}_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2 \leqslant \sup_{x_o \in C} \mathbb{E}_\epsilon\|\eta_{MM}(x_o + \sigma\epsilon) - x_o\|_2^2,$$

where η_{MM} denotes the minimax denoiser and $\tilde{\eta}_\sigma(\ x_o + \sigma\epsilon)$ denotes $\mathbb{E}(x_o \mid x_o + \sigma\epsilon)$. Let $(\bar{\sigma}^\infty)^2 = \frac{\bar{\theta}^\infty(x_o, \delta, \sigma_w^2)}{\delta} + \sigma_w^2$ and $(\sigma_{mm}^\infty)^2 = \frac{\theta^\infty(x_o, \delta, \sigma_w^2)}{\delta} + \sigma_w^2$. Also, for notational simplicity

[2] In the context of this work the divergence $\text{div}D(x)$ is simply the sum of the partial derivatives with respect to each element of x, i.e., $\text{div}D(x) = \sum_{i=1}^{n} \frac{\partial D(x)}{\partial x_i}$, where x_i is the i^{th} element of x.

[3] Many advanced image denoisers have no closed form expression, thus it is very hard to verify whether they are Lipschitz continuous. That said, every advanced denoisers we tested was found to closely follow the SE equations (Finding 1), suggesting they are in fact Lipschitz.

[4] http://dsp.rice.edu/software/DAMP-toolbox

assume that $\sup_{x_o \in C} \mathbb{E}_\epsilon \|\eta_{MM}(x_o + \sigma\epsilon) - x_o\|_2^2$ is achieved at a certain value x_{mm}. We then have

$$\begin{aligned} \bar{\theta}^\infty(p_x, \delta, \sigma_w^2) &= \frac{\mathbb{E}_{x_o,\epsilon}\|\tilde{\eta}_{\bar{\sigma}^\infty}(x_o + \bar{\sigma}^\infty\epsilon) - x_o\|_2^2}{n} \\ &\leqslant \frac{\mathbb{E}_\epsilon\|\eta^{mm}_{\sigma^\infty}(x_{mm} + \sigma^\infty\epsilon) - x_{mm}\|_2^2}{n}. \end{aligned}$$

This inequality implies that $\bar{\theta}^\infty(p_x, \delta, \sigma_w^2)$ is below the fixed point of the deterministic SE using η^{mm} at x_{mm}. Therefore, because $\sup_{x_o \in C} \theta^\infty(x_o, \delta, \sigma_w^2)$ will be equal to or above the fixed point of η^{mm} at x_{mm}, it will satisfy $\sup_{p_x \in \mathcal{P}} \bar{\theta}^\infty(p_x, \delta, \sigma_w^2) \leqslant \sup_{x_o \in C} \theta^\infty(x_o, \delta, \sigma_w^2)$. □

Under some general conditions it is possible to prove that

$$\begin{aligned} &\sup_{\pi \in \mathcal{P}} \inf_{D_\sigma} \mathbb{E}_{x_o,\epsilon}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2 \\ &\quad = \inf_{D_\sigma} \sup_{x_o \in C} \mathbb{E}_\epsilon\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2. \end{aligned} \tag{5}$$

For instance, if we have

$$\begin{aligned} &\sup_{\pi \in \mathcal{P}} \inf_{D_\sigma} \mathbb{E}_{x_o,\epsilon}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2 \\ &\quad = \inf_{D_\sigma} \sup_{\pi \in \mathcal{P}} \mathbb{E}_{x_o,\epsilon}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2, \end{aligned}$$

then (5) holds as well. Since we work with square loss in the SE, swapping the infimum and supremum is permitted under quite general conditions on $\mathcal{P}$. For more information, see Appendix A of [18]. If (5) holds, then we can follow similar steps as in the proof of Theorem 1 to prove that under the same set of conditions we can have

$$\sup_{x_o \in C} \theta^\infty(x_o, \delta, \sigma_w^2) = \sup_{p_x \in \mathcal{P}} \bar{\theta}^\infty(p_x, \delta, \sigma_w^2).$$

In words, *the supremums of the fixed points of the deterministic and Bayesian SEs are equivalent.*

D. Why bother?

Considering that the deterministic and Bayesian SEs look so similar, and under certain conditions have the same supremums, it is natural to ask why we developed the deterministic SE at all. That is, what is gained by using SE (2) rather than (3)?

The deterministic SE is useful, because it enables us to deal with signals with poorly understood distributions. Take, for instance, natural images. To use the Bayesian SE on imaging problems, we would first need to characterize all images according to some generalized, almost assuredly inaccurate, pdf. In contrast, the deterministic SE deals with specific signals, not distributions. Thus, even without knowledge of the underlying distribution, so long as we can come up with representative test signals, we can use the deterministic SE. Because the SE shows up in the parameter tuning, noise sensitivity, and performance guarantees of AMP algorithms, being able to deal with arbitrary signals is invaluable [4].

To further demonstrate its utility, we next use the deterministic SE of (2) to investigate the optimality of D-AMP.

IV. Optimality of D-AMP

Is D-AMP optimal? In other words, given a family of denoisers, D_σ, for a set C, can we come up with an algorithm for recovering x_o from $y = \mathbf{A}x_o + w$ that outperforms D-AMP? Note that this problem is ill-posed in the following sense: the denoising algorithm might not capture all the structure that is present in the signal class C. Hence, a recovery algorithm that captures additional structure not used by the denoiser (and thus not used by D-AMP) might outperform D-AMP. In the following sections we consider two different approaches to analyze the optimality of D-AMP. Before we proceed, we first need to formalize our notion of denoiser.

A. Denoiser properties

The role of a denoiser is to estimate a signal x_o belonging to a class of signals $C \subset \mathbb{R}^n$ from noisy observations, $f = x_o + \sigma\epsilon$, 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 by the standard deviation of the noise. At every value of σ, D_σ takes $x_o + \sigma\epsilon$ as the input and returns an estimate of x_o.

To analyze D-AMP, we require the denoiser family to be *proper*, *monotone*, and *Lipschitz continuous* (proper and monotone are defined below). Because most denoisers easily satisfy these first two properties, and can be modified to satisfy the third (see Section VI of [4]), the requirements do not overly restrict our analysis.

Definition 1. *D_σ is called a proper family of denoisers of level κ ($\kappa \in (0,1)$) for the class of signals C if*

$$\sup_{x_o \in C} \frac{\mathbb{E}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2}{n} \leqslant \kappa\sigma^2 \tag{6}$$

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

To clarify the above definition, we consider the following examples.

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

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

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

Proof. Note that since the projection onto a subspace is a linear operator and since $P_C(x_o) = x_o$, we have

$$\mathbb{E}\|P_C(x_o+\sigma\epsilon)-x_o\|_2^2 = \mathbb{E}\|x_o+\sigma P_C(\epsilon)-x_o\|_2^2 = \sigma^2\mathbb{E}\|P_C(\epsilon)\|_2^2.$$

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

Below we 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 $\eta(f;\tau\sigma) = (|f| - \tau\sigma)_+\text{sign}(f)$ denote the family of soft-thresholding denoisers. Then*

$$\sup_{x_o \in \Gamma_k} \frac{\mathbb{E}\|\eta(x_o + \sigma\epsilon; \tau\sigma) - x_o\|_2^2}{n\sigma^2} = \tfrac{(1+\tau^2)k}{n} + \tfrac{n-k}{n}\mathbb{E}(\eta(\epsilon_1;\tau))^2,$$

where ϵ_1 denotes the first element of the noise vector ϵ.

Similar results can be found in the literature, including [19]. See Section III.B of [4] for a short proof.

Definition 2. *A denoiser is called monotone if for every x_o its risk function*

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

is a non-decreasing function of σ^2.

Remark 1. *Monotonicity is a natural property to expect from denoisers; increasing the variance of the noise should not improve the quality of a denoiser's estimate. Many standard denoisers, including soft-thresholding and group soft-thresholding, are monotone if we optimize over the threshold parameter. See Lemma 4.4 in [20] for more information.*

In subsequent sections we assume our signal belongs to a class C for which we have a proper family of monotone and Lipschitz denoisers D_σ. The class and denoiser can be very general. For instance, C can represent the class of natural images and D_σ can denote the BM3D algorithm [17] at different noise levels.

B. Uniform optimality

Let $\mathcal{E}_\kappa$ denote the set of all classes of signals C for which there exists a family of denoisers D_σ^C that satisfies

$$\sup_{\sigma^2} \sup_{x_o \in C} \frac{\mathbb{E}\|D_\sigma^C(x_o + \sigma\epsilon) - x_o\|_2^2}{n\sigma^2} \leqslant \kappa. \quad (7)$$

We know from Proposition 1 of [4] that for any $C \in \mathcal{E}_k$, D-AMP recovers all the signals in C from $\delta > \kappa$ measurements.

We now ask our uniform optimality question: *Does there exist any other signal recovery algorithm that can recover all the signals in all these classes with fewer measurements than D-AMP?* If the answer is affirmative, then D-AMP is suboptimal in the uniform sense, meaning that there exists an approach that outperforms D-AMP uniformly over all classes in $\mathcal{E}_\kappa$. The following proposition shows that any recovery algorithm requires at least $m = \kappa n$ measurements for accurate recovery, i.e., D-AMP is optimal in this sense.

Proposition 1. *If m^* denotes the minimum number of measurements required (by any recovery algorithm) for a set $C \in \mathcal{E}_\kappa$, then*

$$\sup_{C \in \mathcal{E}_\kappa} \frac{m^*(C)}{n} \geqslant \kappa.$$

Proof. According to Example 1, any κn dimensional subspace of $\mathbb{R}^n$ belongs to $\mathcal{E}_\kappa$ (assuming that κn is an integer). From the fundamental theorem of linear algebra we know that to recover the vectors in a k dimensional subspace we require at least k measurements. Hence

$$\sup_{C \in \mathcal{E}_\kappa} \frac{m^*(C)}{n} \geqslant \frac{\kappa n}{n} = \kappa.$$

□

According to this simple result, D-AMP is optimal for at least certain classes of signals and certain denoisers. Hence, it cannot be uniformly improved.

C. Single class optimality

The uniform optimality framework we introduced above considers a set of signal classes and measures the performance of an algorithm on every class in this set. However, in many applications, such as imaging, we are interested in the performance of D-AMP on a specific class of signals, such as natural images. Therefore, in this section we evaluate the optimality of D-AMP using a *single class optimality* framework.

Let C denote a class of signals. Instead of assuming that we are given a family of denoisers for the signals in class C, we assume that we can find the denoiser that brings out the best performance from D-AMP. This ensures that D-AMP employs as much information as it can about C. Let $\theta_D^\infty(x_o, \delta, \sigma_w^2)$ denote the fixed point of the SE equation given in (2). Note that we have added a subscript D to our notation for θ to indicate the dependence of this quantity on the choice of the denoiser. The best denoiser for D-AMP is a denoiser that minimizes $\theta_D^\infty(x_o, \delta, \sigma_w^2)$. According to Finding 1, $\theta_D^\infty(x_o, \delta, \sigma_w^2)$ corresponds to the mean square error of the final estimate that D-AMP returns.

Definition 3. *A family of denoisers D_σ^* is called minimax optimal for D-AMP at noise level σ_w^2, if it achieves*

$$\inf_{D_\sigma} \sup_{x_o \in C} \theta_D^\infty(x_o, \delta, \sigma_w^2).$$

Note that, according to our definition, the optimal denoiser may depend on both σ_w^2 and δ and is not necessarily unique. We denote the versions of D-AMP that employ D_σ^* by D*-AMP.

Armed with this definition, we formally ask the single class optimality question: *Can we provide a new algorithm that can recover signals in class C with fewer measurements than* D*-*AMP?* The following definition and proposition help answer the question.

Definition 4. *The minimax risk of a set of signals C at the noise level σ^2 is defined as*

$$R_{MM}(C, \sigma^2) = \inf_D \sup_{x_o \in C} \mathbb{E}\|D(x_o + \sigma\epsilon) - x_o\|_2^2,$$

where the expected value is with respect to $\epsilon \sim N(0, I)$. If D_σ^M achieves $R_{MM}(C, \sigma^2)$, then it will be called the family of minimax denoisers for the set C under the square loss.

Proposition 2. *The family of minimax denoisers for C is a family of optimal denoisers for D-AMP. Furthermore, in order to recover every $x_o \in C$,* D*-*AMP requires at least $n\kappa_{MM}$ measurements:*

$$\kappa_{MM} = \sup_{\sigma^2>0} \frac{R_{MM}(\sigma^2)}{n\sigma^2}.$$

The proof of this result can be found in Appendix A of [4].

Based on this result, we can simplify the single class optimality question: *Does there exist any recovery algorithm that can recover every $x_o \in C$ from fewer observations than $n\kappa_{MM}$?* Unfortunately, the answer is affirmative.

Consider the following extreme example. Let B_k^n denote the class of signals that consist of k ones and $n-k$ zeros. Define $\rho = k/n$ and let $\phi(z)$ denote the density function of a standard normal random variable.

Proposition 3. *For very high dimensional problems, there are recovery algorithms that can recover signals in B_k accurately from 1 measurement. On the other hand, the number of measurements that D*-*AMP requires to recover signals from this class is given by $n(\kappa_{MM} - o(1))$, where*

$$\kappa_{MM} = \sup_{\sigma^2>0} \frac{1}{\sigma^2}\mathbb{E}_{z_1\sim\phi}\left(\frac{\rho\phi_\sigma(z_1)}{\rho\phi_\sigma(z_1)+(1-\rho)\phi_\sigma(z_1+1)} - 1\right)^2 \rho + \mathbb{E}_{z_1\sim\phi}\left(\frac{\rho\phi_\sigma(z_1-1)}{\rho\phi_\sigma(z_1-1)+(1-\rho)\phi_\sigma(z_1)}\right)^2 (1-\rho),$$

where $\phi_\sigma(z) = \phi(z/\sigma)$.

The proof of this result can be found in Appendix B of [4]. According to this proposition, since κ_{MM} is non-zero, the number of measurements D*-AMP requires is proportional to the ambient dimension n, while the actual number of measurements required for recovery is equal to 1. Hence, in such cases D*-AMP is suboptimal.

V. Conclusions

In this paper, we have explored the performance of denoising-based approximate message passing (D-AMP) using a new deterministic state evolution (SE) framework. The new SE could be of independent interest because, unlike the traditional Bayesian SE, it requires no knowledge of a signal's pdf. As a result, it can be applied to problems where the traditional SE framework is intractable. Using this new SE, we found that, while D-AMP is suboptimal for certain classes of signals, no algorithm can uniformly outperform it.

There are several avenues for continued research in this arena. We would like to develop a better understanding of the signal classes and denoisers for which D-AMP is optimal. In particular, we would like to determine whether D-AMP is optimal for images, or certain classes of images. Additionally, the SE framework currently applies only to Gaussian measurement matrices. We would like to extend the state-evolution to more general measurement matrices, especially sub-sampled Fourier measurements.

References

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4203 – 4215, Dec. 2005.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[3] C. Metzler, A. Maleki, and R. Baraniuk, "From denoising to compressed sensing," *ELEC599 Project Report*, Apr. 2014.

[4] C. Metzler, A. Maleki, and R. Baraniuk, "From denoising to compressed sensing," *arXiv preprint:1406.4175*, 2014.

[5] C. Metzler, A. Maleki, and R. Baraniuk, "BM3D-AMP: A new image recovery algorithm based on BM3D denoising," *submitted to Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2015.

[6] J. Tan, Y. Ma, and D. Baron, "Compressive imaging via approximate message passing with image denoising," *arXiv preprint:1405.4429*, 2014.

[7] Y. Ma, J. Zhu, and D. Baron, "Compressed sensing via universal denoising and approximate message passing," *Proc. Allerton Conf. Communication, Control, and Computing*, 2014.

[8] C. Guo and M. E. Davies, "Near optimal compressed sensing without priors: Parametric sure approximate message passing," *European Signal Processing Conference*, 2014.

[9] S. Hamzehei and M. F. Duarte, "Compressive parameter estimation via approximate message passing," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2015.

[10] A. Maleki and D. L. Donoho, "Optimally tuned iterative thresholding algorithm for compressed sensing," *IEEE J. Select. Top. Signal Processing*, Apr. 2010.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[12] D. L. Donoho, I. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE Trans. Inform. Theory*, vol. 59, no. 6, pp. 3396–3433, 2013.

[13] A. Maleki, "Approximate message passing algorithm for compressed sensing," *Stanford University PhD Thesis*, Nov. 2010.

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

[15] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Simul*, vol. 4, pp. 490–530, 2005.

[16] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, pp. 2080–2095, Aug. 2007.

[18] I. M. Johnstone, "Function estimation and gaussian sequence models," *Unpublished manuscript*, 2013.

[19] D. L. Donoho, A. Maleki, and A. Montanari, "Noise sensitivity phase transition," *IEEE Trans. Inform. Theory*, vol. 57, Oct. 2011.

[20] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Parameterless optimal approximate message passing," *arXiv preprint:1311.0035*, 2013.