Approximate Message Passing-based Compressed Sensing Reconstruction with Generalized Elastic Net Prior

Xing Wang, Jie Liang[1],[*]

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

Abstract

In this paper, we study the compressed sensing reconstruction problem with generalized elastic net prior (GENP), where a sparse signal is sampled via a noisy underdetermined linear observation system, and an additional initial estimation of the signal (the GENP) is available during the reconstruction. We first incorporate the GENP into the LASSO and the approximate message passing (AMP) frameworks, denoted by GENP-LASSO and GENP-AMP respectively. We then focus on GENP-AMP and investigate its parameter selection, state evolution, and noise-sensitivity analysis. A practical parameterless version of the GENP-AMP is also developed, which does not need to know the sparsity of the unknown signal and the variance of the GENP. Simulation results with 1-D data and two different imaging applications are presented to demonstrate the efficiency of the proposed schemes.

Keywords:
Compressed sensing, approximate message passing, elastic net prior, state evolution, phase transition.

1. Introduction

The problem of reconstructing a sparse signal from its noisy linear measurement is crucial to many applications. In this case, the observation $y \in \mathbb{R}^m$ can be written as

$$y = Ax + w, \tag{1}$$

where $x \in \mathbb{R}^n$ is a k-sparse signal, *i.e.*, with k nonzero entries ($k \ll n$). $A \in \mathbb{R}^{m \times n}$ is a known linear measurement matrix, and $w \in \mathbb{R}^m$ is an addi-

tive white Gaussian noise with variance σ^2, *i.e.*, $w \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. In this paper, the following ratios are frequently used:

$$\delta = m/n, \ \ \varepsilon = k/n, \ \ \rho = \varepsilon/\delta = k/m. \tag{2}$$

When $m < n$, the problem is underdetermined and has been studied extensively recently via the compressed sensing (CS) theory. It is shown in [1] that when A satisfies certain condition and m is larger than some bound, ℓ_1-based algorithms can successfully recover the sparse signal. Many reconstruction algorithms have been developed to estimate the sparse signal x from y, including, *e.g.*, convex optimization [1], greedy method [2], and iterative thresholding algorithm [3]. However, precise performance analyses of these methods are not available.

Estimation theory can also be used to analyse the performance of CS. In [4], with the help of the replica method from statistical physics, a sharp prediction is derived for the performance of the LASSO or Basis Pursuit Denoising method (BPDN) [5, 6], which is an ℓ_1-regularized least-square optimization problem. However, the replica assumption is not rigorous and it cannot be checked for specific problems.

In [7, 8, 9, 10], an approximate message passing (AMP) algorithm is developed for Gaussian sampling matrices, which reduces the complexity of classic message passing [11]. More importantly, the AMP is rigorous and can predict the final reconstruction performance accurately. Some generalizations of AMP have been developed. For example, in [12], a generalized AMP (GAMP) is developed to handle arbitrary noise distributions and arbitrary prior distributions. In [13], the Gaussian mixture model and expectation-maximization (EM) algorithm are used to learn the distribution of the signal's nonzero coefficients. It is also shown empirically and theoretically that AMP-type solvers work well with various types of matrices, such as Rademacher matrices and Fourier matrices [13, 14, 15]. Moreover, the AMP offers a unified framework to exploit other prior knowledge or side information (SI) about the signal [9, 10], *e.g.*, non-negativity or positivity constraint [10, 16] and non-uniformly sparsity [17]. Other forms of SI can also be incorporated in the AMP. For example, in [18], the support of the signal is time-invariant and the signal amplitudes are slowly varying over time. In [19], the support of the signal is also allowed to change over time.

In this paper, we consider another kind of SI where there is an initial estimation of the sparse signal x. Intuitively, this initial estimation can help the reconstruction of x. For example, compared to the case without any side information, better reconstruction quality or faster convergence can be achieved with the same sampling rate. This kind of SI could exist in many applications. For example, in dense sensor networks, the sample of a sensor can be estimated from those of its neighboring sensors. This can help the encoding of the sample, as shown in the distributed source coding [20]. As another example, in hybrid multiview imaging systems (as demonstrated in Sec. 7), some cameras are traditional cameras and some are CS cameras [21, 22, 23, 24, 25]. Since neighboring cameras are very close to each other, strong correlations exist among their views. Without losing the generality, we assume that the left and right cameras are

traditional cameras while the middle camera is a CS camera. Therefore, by exploiting the geometric relationship between neighboring views, disparity estimation and depth-based image rendering techniques can be used to obtain a prediction of the middle view from its neighboring views. As another example, in dynamic systems, the current state can be estimated from the previous state through the state evolution equation [26, 27, 18, 19].

In this paper, we model the initial estimation or SI of the signal as a noisy version of the unknown sparse signal, and modify the LASSO and AMP frameworks to incorporate the initial estimation. In [28], an additional ℓ_2 penalty term is added to LASSO, and the scheme is called elastic net-regularized LASSO. In the optimization framework derived in Sec. 3 of this paper, there is also an additional ℓ_2 penalty term to LASSO. When the SI is zero, our scheme reduces to that in [28]. Therefore the SI in our framework can be viewed as a generalized elastic net prior (GENP), and we denote the GENP-aided LASSO and AMP as GENP-LASSO and GENP-AMP respectively. Although [29] is the first to study elastic net prior using AMP, it focuses on the binary classification problem and there is no theoretical performance analysis.

After developing the frameworks of GENP-LASSO and GENP-AMP, we focus on the GENP-AMP, and investigate its parameter selection, state evolution, asymptotic prediction performance, and noise-sensitivity analysis. However, these theoretical analyses require the knowledge of the sparsity of the unknown sparse signal and the variance of the generalized elastic net prior. In practices, these parameters have to be estimated. In [30], a parameterless AMP is developed using Stein's unbiased risk estimate (SURE). Inspired by [30], we apply the SURE theory to GENP-AMP and develop a parameterless version of GENP-AMP. Simulation results with 1-D data and two different imaging applications are presented to demonstrate the efficiency of the proposed schemes.

1.1. Related work

There have been some efforts on exploiting various prior knowledge in CS. In [31], the CS is formulated as a Bayesian problem, and the CS solution is shown to be the maximum a posteriori estimate with a Laplacian prior. In [32], the CS is related to the more general notion of Bayesian experimental design, which allows non-random sampling matrices. The sampling matrix can also be optimized using the Bregman matrix theory in [33]. In this paper, we focus on the widely used random matrix-based CS problem.

In [27], the problem of CS with partially known support is studied, which shows that by finding the signal that satisfies the measurement constraint and is the sparsest outside the partially known support, the CS reconstruction can be improved, and bounds on the reconstruction error are derived. However, the method is time-consuming. Another relevant approach is to recover the estimation error instead of the sparse signal [21], based on the assumption that the prediction error between the initial estimation and the sparse signal is sparser than the signal itself, and is thus easier to be recovered, but this method lacks theoretical analysis. It is also possible that the prediction error is denser than the original sparse signal, if the initial estimation has poor quality.

In [24], the belief-propagation-based CS framework (BPCS) in [34] is used to exploit the SI from neighboring cameras in multiview imaging systems, where the SI is used as the starting point for belief propagation. In [25], a squared-error-constrained penalty term is added to the CS of multiview images. It also considers a more general case, where the variances of the prediction errors are different at different entries. A fast solution is developed based on the Gradient Projection for Sparse Reconstruction (GPSR) algorithm [35].

The sparsity-constrained dynamic system estimation scheme proposed in [26] and the dynamic compressed sensing via approximate message passing (DCS-AMP) proposed in [18, 19] are closely related to our framework. In [26], a prediction of the signal is obtained from the state evolution model, and the norm of the prediction error is added as a penalty term in the objective function of LASSO or BPDN method. In [18, 19], the sparse signal is modeled as the Bernoulli-Gaussian distribution and the correlation between the active amplitudes in different time slots is assumed to be a stationary steady-state Gaussian-Markov process. The EM and AMP are applied to learn the hidden parameters and perform the inference. Although the model in [18, 19] is similar to ours, it relies on sequential data to learn the hidden parameters, and cannot be applied to solve the problem discussed here directly. In fact, it is not clear how to extend the method in [18] to solve the problem in this paper.

Several papers have also studied the theoretical contribution of the prior knowledge [27, 36]. In [27], the authors have provided some sharp bounds on the necessary number of CS measurements to successfully reconstruct the original sparse signal, based on null space property and geometry interpretations. However, it is mainly on the noiseless case. The performance of noisy case remains unknown. Kamilov *et al.* have taken the first step towards a theoretical understanding of EM-based algorithms [18, 19, 36], although the complete analysis is still not available. Our method does not involve any loose constant, and can accurately predict the performance.

On the other hand, the GENP considered in this paper can be incorporated into the GAMP [12]. However, even if the GENP is known to the GAMP, the GAMP still needs to know the exact prior distribution $p(x)$. Therefore in practice some learning-based methods such as the EM algorithm have to be used to learn $p(x)$ [13]. Our scheme does not need to know $p(x)$, and only assumes that x is sparse.

In Sec. 7 of this paper, we will present simulation results with 1-D data and two different imaging applications. We will show that the overall performance of our methods is better than the AMP, GAMP, the method in [21], the modified CS in [27], the linear minimum mean squared error (LMMSE) method, and direct denoising. Our parameterless method also works very well below the phase transition boundary of AMP, although its performance still needs to be improved above the boundary, because the estimated variance of the prior using the method in [30] is unstable in this case.

Some preliminary results of this paper have been reported in [37]. Due to the importance of the problem of side/prior information-based CS reconstruction, earlier versions of this paper have received attention from other researchers

4

[38, 39]. In [38], only the noiseless CS sampling scenario is considered, and an ℓ_1 or ℓ_2 constraint of the prior information is added to the ℓ_1 Basis Pursuit objective function of the unknown signal. However, only some loose bounds of different constraints are presented in it. In [40], similar ℓ_1 constraint is added and an adaptive-weighted approach is developed. Instead of uniformly random sampling, an adaptive sampling is exhibited in each step until a sufficient number of samples has been obtained. In [39], the classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information is discussed. The high-dimensional signals are assumed to follow Gaussian mixture model (GMM) that can be learned from training data. The fundamental limits are derived based on this assumption. In our paper, we do not make any assumption about the target signal except sparsity and there is no training data involved.

2. Background: Minimax MSE of Soft Thresholding Algorithm

In this section, we briefly review the minimax MSE of the soft thresholding algorithm [8, 41], which plays an important role in AMP. Suppose we need to recover a k-sparse n-vector $x^0 = (x^0(i) : i \in [n])$ (where $[n] \equiv \{1, \ldots, n\}$) contaminated by a Gaussian white noise, *i.e.*,

$$y(i) = x^0(i) + z^0(i), \ i \in [n],$$

where $z^0(i) \sim \mathcal{N}(0, \sigma^2)$ is independent and identically distributed. One way to estimate the signal is to solve the following LASSO or ℓ_1-regularized least-squares problem,

$$\hat{x}^\lambda = \arg\min_x \ \frac{1}{2}\|y - x\|_2^2 + \lambda\|x\|_1. \tag{3}$$

An important fact is that the solution of this problem is equivalent to that of the well-known soft thresholding algorithm in wavelet denoising [41],

$$\hat{x}^\lambda(i) = \eta(y(i); \lambda), \ i \in [n],$$

where the soft thresholding operation with threshold θ is

$$\eta(x; \theta) = \begin{cases} x - \theta & \text{if } x > \theta, \\ 0 & \text{if } -\theta \leqslant x \leqslant \theta, \\ x + \theta & \text{if } x < -\theta. \end{cases} \tag{4}$$

A reasonable choice of the threshold λ in (3) is a scaled version of the noise standard deviation, *i.e.*, $\lambda = \alpha\sigma$. The MSE of the soft thresholding algorithm can thus be written as

$$\text{mse}(\sigma^2; p, \alpha) \equiv E\{[\eta(X + \sigma Z; \alpha\sigma) - X]^2\}, \tag{5}$$

where the expectation is with respect to independent random variables $Z \sim$

$\mathcal{N}(0,1)$ and $X \sim p$.

The soft thresholding method is scale-invariant [8], *i.e.*,

$$\mathrm{mse}(\sigma^2; p, \alpha) = \sigma^2 \mathrm{mse}(1; p_{1/\sigma}, \alpha), \tag{6}$$

where p_s is a scaled version of p, $p_s(S) = p(\{x : \ sx \in S\})$. Therefore we only need to focus on $\sigma = 1$, and the notation $\mathrm{mse}(1; p, \alpha)$ can be simplified into $\mathrm{mse}(p, \alpha)$.

Since x^0 is k-sparse, we can define the following set of probability measures with small non-zero probability,

$$\mathcal{F}_\varepsilon \equiv \{p : \ p \text{ is a probability measure with } p(\{0\}) \geqslant 1 - \varepsilon\}, \tag{7}$$

where $\varepsilon = k/n$ is defined in (2).

The minimax threshold MSE is thus defined as [8]

$$M^{\pm}(\varepsilon) = \inf_{\alpha > 0} \sup_{p \in \mathcal{F}_\varepsilon} \mathrm{mse}(p, \alpha), \tag{8}$$

which is the minimal MSE of the worst distribution in $\mathcal{F}_\varepsilon$, where $\pm$ means a nonzero estimand can take either sign.

For a given α, the worst case MSE in (8) is given by [8]

$$\sup_{p \in F_\varepsilon} \mathrm{mse}(p, \alpha) = \varepsilon(1 + \alpha^2) + (1 - \varepsilon)[2(1 + \alpha^2)\Phi(-\alpha) - 2\alpha\phi(\alpha)], \tag{9}$$

with $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ being the standard normal density, and $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$ the Gaussian cumulative distribution function. Moreover, the supremum can be achieved by the following three-point probability distribution on the extended real line $\mathbb{R} \cup \{-\infty, \infty\}$

$$p_\varepsilon^* = (1 - \varepsilon)\delta_0 + \frac{\varepsilon}{2}\delta_\infty + \frac{\varepsilon}{2}\delta_{-\infty},$$

where δ_t is a Dirac delta function at t. In practice, we are more interested in the near-worse-case signals with finite values. It is known that the following c-least-favorable distribution can achieve a MSE that is a fraction of $(1 - c)$ of the worst case,

$$p_{\varepsilon,c} = (1 - \varepsilon)\delta_0 + \frac{\varepsilon}{2}\delta_{h^{\pm}(\varepsilon,c)} + \frac{\varepsilon}{2}\delta_{-h^{\pm}(\varepsilon,c)}, \tag{10}$$

where $h^{\pm}(\varepsilon, c) \sim \sqrt{2\log(\varepsilon^{-1})}$ as $\varepsilon \to 0$.

3. GENP-aided LASSO

In this paper, we study the generalized elastic net prior (GENP)-aided CS reconstruction, where in addition to the CS sampling as in (1), an initial estimation of x, denoted by $\tilde{x}$, is available during reconstruction, which can be seen

as a noisy version of x. The error of this estimation, $e = \tilde{x} - x$, is assumed to be i.i.d. additive white Gaussian with variance σ_s^2, i.e., $e \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I})$. This Gaussian noise model can be used in applications such as image acquisition with poor illumination, high temperature, or transmission error, and has been widely used in image denoising [42]. The ratio between the noise variance of $\tilde{x}$ and that of the compressed sampling noise in Eq. (1) will be used later for noise sensitivity analysis.

$$\gamma_s^2 = \sigma_s^2 / \sigma^2. \tag{11}$$

To exploit the $\tilde{x}$ in the CS reconstruction, we propose the following optimization framework,

$$\hat{x}(\lambda, \tau_s) = \arg\min_{z \in \mathbb{R}^n} \left(\frac{1}{2} \|y - Az\|_2^2 + \lambda \|z\|_1 + \frac{\tau_s}{2} \|\tilde{x} - z\|_2^2 \right), \tag{12}$$

which is a generalized version of the LASSO in (3) with an additional ℓ_2 penalty term caused by the initial estimation $\tilde{x}$ to ensure the solution close to this initial estimation. When $\tilde{x} = 0$, the problem reduces to the elastic net-regularized LASSO in [28]. Therefore we call $\tilde{x}$ generalized elastic net prior (GENP), and the problem in Eq. (12) generalized elastic net prior-aided LASSO (GENP-LASSO). A is assumed to be Gaussian measurement matrix in this paper.

A special case of our framework is that when $p(x)$ follows the Laplacian distribution, the result of Eq. (12) is equivalent to the maximum a posteriori (MAP) solution. However, our framework in Eq. (12) is more general than this special case because we do not rely on any assumption about $p(x)$, except that x should be sparse as defined in Eq. (7). In the following theoretical analysis, we will apply the minimax estimator introduced in Sec. 2 to study the parameter selection, state evolution and MSE performance of the optimization problem in Eq. (12).

Similarly, although the GENP in our framework can also be incorporated into the GAMP scheme in [12], it should be noted that GAMP also needs to know the exact prior distribution $p(x)$. Therefore learning algorithms such as the EM have to be used to learn the prior distribution [13]. In Sec. 7, we will compare our method to the EMGMAMP in [13] and a modified EMGMAMP that incorporates the GENP, and show that our method has better overall performance.

In LASSO, the ratio ρ in Eq. (2) cannot be larger than 1, i.e., the number of selected atoms is bounded by the number of samples, whereas it is shown in [28] that in the elastic net-regularized LASSO, the quadratic penalty term removes this limitation. Our noise sensitivity analysis in Sec. 5 will show that $\rho < 1$ is also not necessary in the GENP-LASSO.

The parameters λ and τ_s in Eq. (12) are closely related to σ_s^2, the noise variance of the GENP. How to tune the two parameters λ and τ_s will be addressed later in the paper.

The proposed GENP-LASSO in (12) is a convex optimization problem and can be solved by, e.g., the interior point methods (as used in the CVX package

[43]) and the gradient methods. For example, to incorporate the GENP into the Orthant-Wise Limited-memory Quasi-Newton (OWLQN) algorithm [44], which is a popular gradient-based method for large-scale LASSO problems, we can replace the ℓ_2 regularization term $\|z\|_2^2$ in it by the quadratic penalty term $\|\tilde{x} - z\|_2^2$. However, both interior point and gradient methods are quite slow for large-scale problems.

In this paper, we will solve the GENP-LASSO problem by modifying the fast AMP algorithm, which enjoys several advantages, *e.g.*, low complexity and the capability of predicting the final performance accurately.

Note that we can also combine y and $\tilde{x}$ as follows.

$$
\begin{bmatrix} y \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} A \\ I \end{bmatrix} x + \begin{bmatrix} w \\ e \end{bmatrix}. \tag{13}
$$

This is an overdetermined system of x with $(m + n)$ equations. Therefore x can be solved directly using the least-squares (LS) or the linear minimum mean squared error (LMMSE) method. However, we will show in Sec. 7 that the performance of the LMMSE method is not as good as the proposed method (the LS solution is even worse than that of the LMMSE, and is not included due to space limitation). Note that the LMMSE solution also requires the knowledge of $p(x)$.

Finally, we point out that the framework in Eq. (12) can still be applied when z itself is not sparse, but is sparse in an orthogonal basis Ψ, *i.e.*, $\alpha = \Psi z$ is sparse, for example, when z is a natural image. In this case, Eq. (12) can be written as

$$
\hat{x}(\lambda, \tau_s) = \arg\min_{z \in \mathbb{R}^n} \left(\frac{1}{2} \|y - Az\|_2^2 + \lambda \|\Psi z\|_1 + \frac{\tau_s}{2} \|\tilde{x} - z\|_2^2 \right). \tag{14}
$$

However, since Ψ is orthogonal, this is equivalent to

$$
\hat{\alpha}(\lambda, \tau_s) = \arg\min_{\alpha \in \mathbb{R}^n} \left(\frac{1}{2} \|y - A\Psi^{-1}\alpha\|_2^2 + \lambda \|\alpha\|_1 + \frac{\tau_s}{2} \|\tilde{\alpha} - \alpha\|_2^2 \right), \tag{15}
$$

where $\tilde{\alpha} = \Psi\tilde{x}$. This has the same format as Eq. (12), except that the sensing matrix becomes $A\Psi^{-1}$. However, given Ψ and Gaussian A, $A\Psi^{-1}$ is still Gaussian. Therefore the AMP theory can still be applied. Without loss of generality, we focus on Eq. (12) in the following derivations.

4. GENP-aided Approximate Message Passing

In this section, we present the formulae of GENP-AMP. We then study its connections with the GENP-LASSO, and derive its corresponding parameter selections and state evolution.

4.1. The Formula of GENP-AMP

In [9], the following iterative formulas of AMP are obtained after simplifying the traditional min-sum-based message passing algorithm using the quadratic approximation.

$$\hat{x}_0^t = x^t + A^T r^t,$$
$$x^{t+1} = \eta(\hat{x}_0^t; \theta_t), \tag{16}$$

$$b_t = \frac{1}{m} \|x^t\|_0, \tag{17}$$

$$r^t = y - Ax^t + b_t r^{t-1}. \tag{18}$$

Each iteration of AMP only needs to update the estimate x^t in (16) and the residual r^t in (18), which have only $m + n$ entries. The complexity is thus much lower than traditional message passing methods that need $2mn$ updates. Note that the AMP is parameterized by two sequences of scalar parameters: the thresholds $\{\theta_t\}_{t \geq 0}$ and the factors $\{b_t\}_{t \geq 0}$.

To incorporate the GENP into AMP, we modify the local message of each AMP variable node from $\lambda\|z\|_1$ to $\lambda\|z\|_1 + \frac{\tau_s}{2}\|\tilde{x} - z\|_2^2$. By the same simplifications and derivations in [9], we can get the following iterative estimate of the n-vector signal x. The details are skipped due to space limitation.

$$\hat{x}_0^t = \frac{u_t}{1 + u_t}\tilde{x} + \frac{1}{1 + u_t}(x^t + A^T r^t), \tag{19}$$

$$x^{t+1} = \eta(\hat{x}_0^t; \theta_t), \tag{20}$$

$$b_t = \frac{1}{1 + u_{t-1}} \frac{\|x^t\|_0}{m}, \tag{21}$$

$$r^t = y - Ax^t + b_t r^{t-1}. \tag{22}$$

Compared to AMP, $\hat{x}_0^t$ in our scheme is a linear combination of $x^t + A^T r^t$ and the GENP, adaptively controlled by a new sequence of scalar parameters, $\{u_t\}_{t \geq 0}$. The factor b_t is also affected by u_{t-1}. When $u_t = 0$, $\tilde{x}$ has no contribution, and the proposed framework reduces to the standard AMP in [7, 8, 9, 10]. The iteration is applied to each entry. Hence, if the variances of different $\tilde{x}_i$ are different, the method can still be applied by changing the scalar u_t to vector $\mathbf{u_t} = [u_{t,1}, u_{t,2}, ..., u_{t,n}]$ and the scalar θ_t to its vector case.

4.2. Connections to GENP-LASSO

As shown in [9], the parameters $\{\theta_t\}_{t \geq 0}$ and $\{b_t\}_{t \geq 0}$ are constrained by its connection with the min-sum algorithm. This is also true for the new parameter $\{u_t\}_{t \geq 0}$. However, the following proposition shows that GENP-AMP provides a very general solution for the GENP-LASSO problem in Eq. (12). When there is no GENP ($u_t = 0$), the proposition reduces to Prop. 5.1 in [9] for LASSO.

Proposition 4.1. *Let (x^*, r^*) be the fixed point of the GENP-AMP algorithm given by (19) and (22) for fixed $\theta_t = \theta$, $u_t = u$, and $b_t = b$. Then x^* is also a minimum of the GENP-LASSO problem in (12) with*

$$\lambda = (1+u)\theta(1-b), \tag{23}$$

$$\tau_s = u(1-b). \tag{24}$$

PROOF. The fixed-point condition of Eq. (19) is

$$x^* = \frac{u}{1+u}\tilde{x} + \frac{1}{1+u}(x^* + A^T r^*) - \theta v^*, \tag{25}$$

where $v_i^* = sign(x_i^*)$ if $x_i^* \neq 0$ and $v_i^* \in [-1, +1]$ otherwise. Similarly, from (22), we get $(1-b)r^* = y - Ax^*$, or $r^* = (y - Ax^*)/(1-b)$. Plugging into the equation above, we get

$$(1+u)\theta(1-b)v^* + u(1-b)(x^* - \tilde{x}) = A^T(y - Ax^*).$$

On the other hand, in Eq. (12), by setting the derivative of the GENP-LASSO objective function with respect to z to zero, we get the stationary condition

$$\lambda v^* + \tau_s(x^* - \tilde{x}) = A^T(y - Ax^*). \tag{26}$$

Comparing the two equations above leads to the conclusion.

4.3. GENP-AMP State Evolution and Parameter Selection

In this part, we derive the state evolution of GENP-AMP and investigate its parameter selection. The state evolution was first developed to describe the asymptotic limit of the AMP estimates as $m, n \to \infty$ for any fixed t, but with the same sample ratio $\delta = m/n$, as defined in (2) [9]. It enables the accurate prediction of the MSE of AMP by solving a fixed-point equation. This part is based on Sec. IV of [8].

First, we define the MSE map Ψ as

$$\Psi(q^2, u, \delta, \sigma, \sigma_s, \alpha, p) \equiv \text{mse}(\text{npi}(q^2, u; \delta, \sigma, \sigma_s); p, \alpha),$$

which is the MSE of the soft thresholding as defined in (5) with npi (noise-plus interference) as the noise variance, where q^2 is the variance of the thresholded estimator, and npi is the variance of the unthresholded estimator in (19), which can be written as (see Appendix A for the derivation)

$$\text{npi}(q^2, u; \delta, \sigma, \sigma_s) = (\frac{u}{1+u})^2\sigma_s^2 + (\frac{1}{1+u})^2(\sigma^2 + \frac{q^2}{\delta}). \tag{27}$$

As pointed out in [9], the choice of the AMP parameter θ_t can be quite flexible. A good option is $\theta_t = \alpha\xi_t$, where $\alpha > 0$, and ξ_t is the root MSE of the unthresholded estimation $\hat{x}_0^t$ in (19). From this, based on the i.i.d. normalized

10

distribution of A and the large system limit [8], it can be shown that

$$\xi_t^2 = \mathrm{npi}(q_t^2, u_t^2; \delta, \sigma, \sigma_s) \approx \left(\frac{u_t}{1+u_t}\right)^2 \sigma_s^2 + \left(\frac{1}{1+u_t}\right)^2 \frac{\|r^t\|_2^2}{m}. \qquad (28)$$

Besides, we have $\|x^t\|_0/n \approx \mathrm{E}\{\eta'(x_0 + \sigma^t Z; \alpha\sigma^t)\}$. According to Eq. (21, 23, 24), Prop. 4.1 can be rewritten as

$$
\begin{aligned}
\lambda &= (1+u_*)\alpha\xi_* \left[1 - \frac{1}{1+u_*}\frac{\mathrm{E}\{\eta'(x_0+\xi_* Z; \alpha\xi_*)\}}{\delta}\right], \\
\tau_s &= u_* \left[1 - \frac{1}{1+u_*}\frac{\mathrm{E}\{\eta'(x_0+\xi_* Z; \alpha\xi_*)\}}{\delta}\right],
\end{aligned}
\qquad (29)
$$

where $\xi_* = \lim_{t\to\infty}\xi_t$. Since the computation of q^2 is nontrivial, Eq. (28) is useful for practical algorithm design, whereas Eq. (27) is mainly for theoretical analysis.

The state of GENP-AMP is defined as a 7-tuple $(q^2, u; \delta, \sigma, \sigma_s, \alpha, p)$. The state evolution follows the rule

$$
\begin{aligned}
(q_t^2, u_t; \delta, \sigma, \sigma_s, \alpha, p) &\mapsto (\Psi(q_t^2, u_t), \Upsilon(q_t^2, u_t); \delta, \sigma, \sigma_s, \alpha, p), \\
t &\mapsto t+1,
\end{aligned}
$$

where q_t^2 and u_t are the MSE and the weighting parameter in the t-th iteration, and Ψ and Υ are the evolution functions of q_t^2 and u_t, respectively. As $(\delta, \sigma, \sigma_s, \alpha, \upsilon)$ are fixed during the evolution, we only need the following state evolutions of q_t^2 and u_t (See Appendix Appendix A for the derivation).

$$
\begin{aligned}
q_t^2 &\mapsto q_{t+1}^2 \equiv \Psi(q_t^2, \frac{\sigma^2 + q_t^2/\delta}{\sigma_s^2}), \\
u_t &\mapsto u_{t+1} = \Upsilon(q_t^2, u_t) = \frac{\sigma^2 + \Psi(q_t^2, (\sigma^2 + q_t^2/\delta)/\sigma_s^2)/\delta}{\sigma_s^2},
\end{aligned}
\qquad (30)
$$

where the formula for u_t is the result of the following proposition.

Proposition 4.2. *The optimal weighting parameter u_t that combines the GENP $\tilde{x}$ and the previous iteration result in the GENP-AMP is given by*

$$u_t = \frac{\sigma^2 + q_t^2/\delta}{\sigma_s^2}. \qquad (31)$$

PROOF. The optimal u_t should minimize the MSE between the original sparse signal and the unthresholded estimation $\hat{x}_0^t$ in (19), which can be obtained by minimizing $(\frac{u_t}{1+u_t})^2\sigma_s^2 + (\frac{1}{1+u_t})^2(\sigma^2 + \frac{q_t^2}{\delta})$ over u_t.

Replacing u in Eq. (27) by Eq. (31), $\mathrm{npi}(q^2, u; \delta, \sigma, \sigma_s)$ can be simplified into

$$\mathrm{npi}(q^2) = \frac{\sigma_s^2(\sigma^2 + q^2/\delta)}{\sigma_s^2 + \sigma^2 + q^2/\delta}. \tag{32}$$

The fixed point condition of the state evolution is

$$q_*^2 = \Psi(q_*^2, \frac{\sigma^2 + q_*^2/\delta}{\sigma_s^2}) = \mathrm{mse}(\mathrm{npi}(q_*^2); p, \alpha). \tag{33}$$

If we treat $\xi^2 = \mathrm{npi}(q_*^2)$ as an unknown variable, plugging (33) into (32) yields a fixed-point equation for ξ^2,

$$\xi^2 = \frac{\sigma_s^2(\sigma^2 + \mathrm{mse}(\xi^2; p, \alpha)/\delta)}{\sigma_s^2 + \sigma^2 + \mathrm{mse}(\xi^2; p, \alpha)/\delta} \equiv F(\xi^2, \alpha). \tag{34}$$

The following result shows that with an appropriate choice of α, the fixed-point equation has a unique solution, from which we can predict the final MSE performance of the GENP-AMP algorithm.

Proposition 4.3. *Let $\alpha_{\min} = \alpha_{\min}(\delta, \gamma_s)$ be the unique non-negative solution of the equation*

$$(1 + \alpha^2)\Phi(-\alpha) - \alpha\phi(\alpha) = \frac{\delta}{2}\frac{(\gamma_s^2 + 1)^2}{\gamma_s^4}, \tag{35}$$

where $\phi(z)$ and $\Phi(z)$ are defined after Eq. (9), and γ_s^2 is defined in Eq. (11). Then for any $\alpha > \alpha_{\min}(\delta, \gamma_s)$, the fixed-point equation $\xi^2 = F(\xi^2, \alpha)$ in (34) admits a unique solution $\xi_* = \xi_*(\alpha)$, and $\lim_{t \to \infty} \xi_t = \xi_*(\alpha)$.

PROOF. This proof is an extension of Case $\chi = \pm$ in Appendix C of [7]. It is easy to find that if γ_s^2 goes to ∞, the whole equation is exactly the one in [9].

Since we want to have $F < \xi^2$, following the same setup as the one in Case $\chi = \pm$ in Appendix C of [7], we need to consider the boundary point, which can be found by solving the boundary condition $\frac{dF}{d\xi^2}|_{\xi^2=0} = 1$. This leads to $\frac{\sigma_s^4 d(\Psi/\delta)/d\xi^2}{(\sigma_s^2 + \sigma^2 + \Psi/\delta)^2}|_{\xi^2=0} = 1$. If $\xi^2 \to 0$, we know that $q^2/\delta = 0$, and the expression of $\frac{d(q^2/\delta)}{d\xi^2}$ can be obtained as in [7]. Then the problem is transformed into

$$\frac{d(q^2/\delta)}{d\xi^2}|_{\xi^2=0} = \frac{(1 + \gamma_s^2)^2}{\gamma_s^4}. \tag{36}$$

The numerator of Eq. (36) becomes $\frac{(1+\gamma_s^2)^2}{\gamma_s^4}(1 - \frac{\gamma_s^4}{(1+\gamma_s^2)^2}\frac{2}{\delta}[(1+\alpha^2)\Phi(-\alpha) - \alpha\phi(\alpha)])$ instead of $1 - \frac{2}{\delta}[(1 + \alpha^2)\Phi(-\alpha) - \alpha\phi(\alpha)]$ as in the classical case in Eq. (6.6) of [9]. Comparing these two expressions, from Proposition 6.2 in [9], we can reach the conclusion.

If the threshold α and the distribution p_0 of X_0 are given, we can obtain

the fixed point ξ_* by solving Eq. (34). Therefore, the MSE performance of the GENP-AMP algorithm can be predicted.

Based on Prop. 4.1, λ and τ_s can be determined if the necessary parameters are known. Conversely, if either λ or τ_s is given, combining Eq. (35) with Eq. (29), we can get the corresponding α and ξ_*. Thus the other parameter can be uniquely determined.

5. Noise Sensitivity Analysis of GENP-AMP

The noise sensitivity phase transition is a curve in the (δ, ρ) plane [8], where $\rho = k/m$ and $\delta = m/n$, as defined in (2). For many classical compressed sensing algorithms, the MSE is bounded below the phase transition curve, and unbounded above the curve. It is known that the optimal phase transition can be achieved by methods such as the AMP [8]. ℓ_1-based methods (such as the CVX package [43]) can also have good phase transition performance. For large-scale problems, the OWLQN algorithm in [44] has similar empirical phase transition boundary to ℓ_1 methods, but its complexity is higher.

In this section, we show that there is no phase transition boundary for GENP-AMP, *i.e.*, its MSE is bounded in the entire plane, thanks to the GENP. We also prove that $\rho < 1$ is no longer needed, which agrees with Lemma 1 in [28] for the elastic net-regularized LASSO.

First, for the GENP-LASSO problem in (12), we define the MSE per entry when the empirical distribution of the signal converges to p_0:

$$\text{MSE}(\sigma^2; \sigma_s^2, p_0, \lambda, \tau_s) = \lim_{n \to \infty} \frac{1}{n} E\{\|\hat{x}(\lambda, \tau_s) - x_0\|_2^2\}, \qquad (37)$$

where the limit is taken along a converging sequence. Since the class $\mathcal{F}_\epsilon$ in (7) is scale-invariant, where $\varepsilon = k/n = \rho\delta$ according to (2), the minimax risk of the GENP-LASSO can be written as

$$\inf_{\lambda, \tau_s} \sup_{p_0 \in F_{\rho\delta}} \text{MSE}(\sigma^2; \sigma_s^2, p_0, \lambda, \tau_s) = M^*(\delta, \rho, \gamma_s^2)\sigma^2, \qquad (38)$$

which indicates the sensitivity of the GENP-LASSO to the noise variance in the measurements, where γ_s^2 is defined in Eq. (11), and the expression of noise sensitivity $M^*(\delta, \rho, \gamma_s^2)$ is given by the following proposition. We also give closed-form expressions of the tuning parameters λ and τ_s that achieve the minimax risk bound.

Before presenting the proposition, we first define the formal mean square error (fMSE) and formal noise-plus interference level (fNPI), following Definitions 3.1−3.4 in [8]. fMSE is defined as the MSE of an observable in a large system framework $\text{LSF}(\delta, \rho, \sigma, \gamma_s, p)$, where $\text{LSF}(\delta, \rho, \sigma, \gamma_s, p)$ denotes a sequence of problem instances $(y; A, x)_{m,n}$ as per Eq. (1) indexed by the problem sizes,

13

and m and n grow proportionally such that $m/n = \delta$. fNPI is expressed as

$$\text{fNPI} = (\frac{u^*}{1+u^*})^2 \sigma_s^2 + (\frac{1}{1+u^*})^2(\sigma^2 + \text{fMSE}/\delta),$$

$$u^* = \frac{\sigma^2 + \text{fMSE}/\delta}{\sigma_s^2}.$$

Its minimax value is $\text{NPI}^*(\delta, \rho, \gamma_s^2) \equiv \frac{\gamma_s^2 \sigma^2(1+M^*(\delta,\rho,\gamma_s^2)/\delta)}{\gamma_s^2+1+M^*(\delta,\rho,\gamma_s^2)/\delta}$ by replacing fMSE in the equation above with its minimax risk $M^*(\delta, \rho, \gamma_s^2)$.

Proposition 5.1. *(1) For any point in the surface, i.e., $\rho \leqslant 1/\delta$ (since $\delta\rho = \varepsilon \leqslant 1$), the minimax risk of GENP-LASSO is bounded, and $M^*(\delta, \rho, \gamma_s^2)$ is given by*

$$M^*(\delta, \rho, \gamma_s^2) = \frac{-G(\delta, \rho, \gamma_s^2) + \sqrt{G(\delta, \rho, \gamma_s^2)^2 + 4\delta\gamma_s^2 M^{\pm}(\delta\rho)}}{2}, \tag{39}$$

where $G(\delta, \rho, \gamma_s^2) = \delta\gamma_s^2 + \delta - \gamma_s^2 M^{\pm}(\delta\rho)$.
  (2)For $c > 0$, define

$$h^*(\delta, \rho, \gamma_s^2; c) \equiv h^{\pm}(\delta\rho, c) \cdot \sqrt{NPI^*}.$$

Then similar to Eq. (10), the distribution $p \in \mathcal{F}_{\delta\rho}$ with a fraction $(1 - \delta\rho)$ of its mass at zero and the remaining mass equally at $\pm h^*(\delta, \rho, \gamma_s^2; c)$ is c-nearly-least-favorable, i.e., the formal noise sensitivity of $\hat{x}(\lambda, \tau_s)$ is

$$\frac{-G(\delta, \rho, \gamma_s^2; c) + \sqrt{G(\delta, \rho, \gamma_s^2)^2 + 4(1-c)\delta\gamma_s^2 M^{\pm}(\delta\rho)}}{2}, \tag{40}$$

where $G(\delta, \rho, \gamma_s; c) = \delta\gamma_s^2 + \delta - (1-c)M^{\pm}(\delta\rho)\gamma_s^2$.
  (3) The formal minimax parameters are given by

$$\lambda(\upsilon; \delta, \rho, \sigma, \sigma_s) \equiv (1 + u^*) \cdot \alpha^{\pm}(\delta\rho) \cdot \sqrt{fNPI(\alpha^{\pm}; \delta, \rho, \sigma, \sigma_s, \upsilon)}$$

$$\times (1 - \frac{1}{1+u^*}EqDR(\upsilon; \alpha^{\pm}(\delta\rho))/\delta), \tag{41}$$

$$\tau_s(\upsilon; \delta, \rho, \sigma, \sigma_s) \equiv u^*(1 - \frac{1}{1+u^*}EqDR(\upsilon; \alpha^{\pm}(\delta\rho))/\delta),$$

where EqDR is the equilibrium detection rate, i.e., the asymptotic fraction of coordinates that are estimated to be nonzero, i.e., $EqDR = P\{\eta(x_\infty; \theta_\infty) \neq 0\}$, as in Eq. (4.5) in [8].

PROOF. The proof is given in Appendix Appendix B.

To show that the noise sensitivity analysis presented here is indeed a generalized result, we next discuss three special cases and show that the result here degrades to the existing known conclusions. First, let $\gamma_s^2 = \infty$. In this case, Eq. (39) degrades to the formulae of the bounded MSE below the phase transition

boundary of AMP, *i.e.*, Eq. (4.8) in [8] . The phase transition boundary only exists in this extreme case for GENP-AMP. Second, if $\gamma_s^2 = 0$, *i.e.*, $\tilde{x} = x$, we do not need to run the AMP; hence the MSE is 0, which coincides with Eq. (39) when $\gamma_s^2 = 0$. Last, if $\delta = 0$, which means there is no compressed measurement, solving the minimization problem in Eq. (12) is equivalent to scalar denoising, and the minimax MSE is $M^\pm(\rho\delta)\sigma_s^2$, which also agrees with the denoising of scalars introduced in Sec. 2.

When there is no initial estimation $\tilde{x}$, the formal MSE noise sensitivity above the phase transition is infinite. However, this is no longer the case in the presence of the GENP, as we can at least assign τ_s to ∞ while keeping λ to be finite, and the formal MSE noise sensitivity is thus bounded by γ_s^2. We can do even better by exploiting the measurement and the sparsity of the original signal, as shown below.

It is easy to verify that $\partial M^*(\delta, \rho, \gamma_s^2)/\partial\gamma_s^2$ is positive, so $M^*(\delta, \rho, \gamma_s^2)$ is a monotonically increasing function of γ_s^2. Since GENP-AMP reduces to AMP when $\gamma_s^2 = \infty$, this means that the minimax bound of GENP-LASSO is no greater than that of LASSO, *i.e.*,

$$M^*(\delta, \rho, \gamma_s^2) \leqslant M^b(\delta, \rho), \tag{42}$$

where $M^b(\delta, \rho) = \frac{M^\pm(\delta\rho)}{1 - M^\pm(\delta\rho)/\delta}$ is the bound of LASSO minimax risk.

Besides, we can also verify that for a fixed sparsity, i.e., $\varepsilon = \delta\rho$ is a constant, $\partial M^*(\delta, \rho, \gamma_s^2)/\partial\delta$ is non-positive (only equal to 0 when $\delta = 0$), and $M^*(\delta, \rho, \gamma_s^2)$ is a monotonically decreasing function of δ. Since GENP-AMP reduces to denoising via soft-thresholding described in Sec. 2 when $\delta = 0$, we conclude that the minimax bound of GENP-LASSO is no greater than that of scalar denoising,

$$M^*(\delta, \rho, \gamma_s^2) \leqslant M^\pm(\delta\rho)\gamma_s^2. \tag{43}$$

In fact, Eq. (42) and (43) have proved that GENP-AMP outperforms AMP and the scalar denoising via soft-thresholding. More importantly, Eq. (42) measures the benefit brought by the generalized elastic net prior while Eq. (43) measures the benefit brought by the linear CS measurements.

We can find more properties of this minimax risk bound. For a fixed δ, the only function of ρ is $M^\pm(\delta\rho)$. From [8], we know that $M^\pm(\delta\rho)$ is monotonically increasing with respect to ρ, and $M^\pm(0) \to 0$, $M^\pm(1) \to 1$. Besides, we can find that $M^*(\delta, \rho, \gamma_s^2)$ is monotonically increasing with respect to $M^\pm(\delta\rho)$. The maximum value of $M^\pm(\delta\rho)$ is 1. The maximum value of $M^*(\delta, \rho, \gamma_s^2)$ is thus

$$\max_{M^\pm(\delta\rho)} M^*(\delta, \rho, \gamma_s^2)$$
$$= \frac{\sqrt{(\delta\gamma_s^2 - \gamma_s^2 + \delta)^2 + 4\delta\gamma_s^2} - (\delta\gamma_s^2 - \gamma_s^2 + \delta)}{2}, \tag{44}$$

where the maximum is achieved at $\rho = 1/\delta$.

6. Parameterless GENP-AMP

In the GENP-AMP proposed above, two parameters need to be known in advance: (1) the sparsity of the signal, $\varepsilon = k/n$, in order to select the appropriate thresholding parameter in soft thresholding function in Sec. 2; (2) the variance of the prior $\tilde{x}$, σ_s^2, in order to determine the weighting parameter u_t as in Prop. 4.2. This makes the algorithm impractical.

The original AMP also needs to know the sparsity. However, recently two types of parameterless AMP algorithms have been developed in [30] and [13, 16]. In [30], Stein's unbiased risk estimate (SURE) framework is used to automatically determine the optimal thresholding parameter in AMP using the gradient descent method. The methods in [13, 16] are both based on the GAMP [12], and try to approximate the MMSE result by learning the prior distribution of the sparse signal through Expectation Maximization (EM) method.

In this part, we follow the approach in [30] due to its theoretical guarantee, since the complete analysis of the EM algorithm used in [13, 16] is still not available. However, the method in [30] cannot be applied in this paper directly since it does not consider the GENP. In the following proposition, using the SURE theory, we develop a practical parameterless version of the GENP-AMP (P-GENP-AMP) that can simultaneously select the thresholding parameter and estimate the variance of the GENP.

Proposition 6.1. *The variance of the GENP $\tilde{x}$ can be approximated by*

$$\sigma_s^2 \approx \frac{\|\tilde{x} - x_{AMP}\|_2^2 - \lim_{t\to\infty} \hat{r}(\theta^t)}{n}, \tag{45}$$

where x_{AMP} is the sparse signal estimated by the AMP with the same setup (fixed A, δ, and ρ), $\lim_{t\to\infty} \hat{r}(\theta^t)/n$ is the MSE of AMP predicted by the SURE method in [30], and

$$\frac{\hat{r}(\theta_t)}{n} = \frac{1}{n} \left\| \eta(\hat{x}_0^t; \theta_t) - \hat{x}_0^t \right\|_2^2 + \sigma_t^2 + \frac{1}{n}\sigma_t^2 [1^T(\eta'(\hat{x}_0^t; \theta_t) - 1)] \tag{46}$$

is Eq. (13) in [30], in which σ_t^2 is the noise-plus interference level in the tth iteration of the standard AMP.

PROOF. The proof is given in Appendix Appendix C.

In fact, thanks to the state evolution analysis, the choice of x_{AMP} can be quite flexible. Another good choice is $\hat{x}_0^*$, the unthresholded estimator in the last iteration of AMP, whose variance is σ_*^2, mentioned in Eq. (16). Then, σ_s^2 can also be approximated by

$$\sigma_s^2 \approx \frac{\|\tilde{x} - \hat{x}_0^*\|_2^2 - \sigma_*^2}{n}. \tag{47}$$
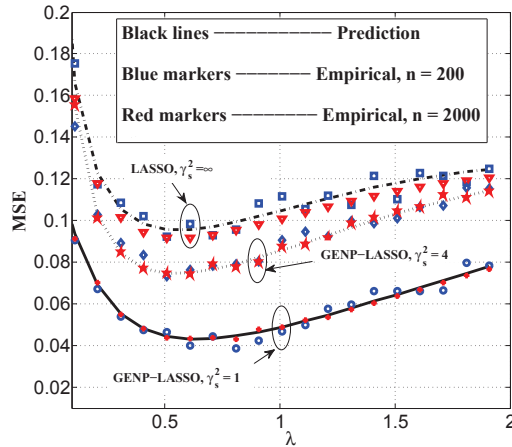
Figure 1: The predicted and actual MSEs of LASSO and GENP-LASSO with different regularization parameter λ. The sample rate is $\delta = 0.64$.

Note that as shown in Prop. 6.1 and its proof in Appendix Appendix C, the approximation of σ_s^2 relies on the approximation of the standard AMP. Therefore, above the phase transition boundary of AMP, the AMP approximation is unstable since the MSE is unbounded, making the approximation $\lim_{t\to\infty} \hat{r}(\theta^t)/n$ unbounded. A tiny mismatch between $\lim_{t\to\infty} \hat{r}(\theta^t)/n$ and MSE of AMP will cause large error when estimating σ_s^2. On the other hand, below the phase transition boundary, the MSE of AMP is bounded. The approximation is very stable.

Once σ_s^2 is estimated, the remaining problem is to determine the thresholding parameter in Eq. (19). Since the iteration formulae and the state evolutions of GENP-AMP are similar to those of AMP, we only need to replace the explicit expressions of σ_t^2 in Eq. (46) with npi(q_t^2) in Eq. (32). The subsequent steps are exactly the same as those in [30], *i.e.*, determining the thresholding parameter θ_t using gradient descent, and updating the estimator and the residual according to Eq. (19) and (22).

7. Numerical Experiments

In this section, we present simulation results with 1-D data and two different imaging applications to demonstrate the performances of the proposed GENP-LASSO and GENP-AMP. Comparisons with some other methods are also included.

7.1. Performance of GENP-LASSO

We first compare the predicted and empirical MSEs of GENP-LASSO and LASSO. Note that GENP-LASSO reduces to LASSO when $\gamma_s^2 = \infty$. We generate the signal vector x_0 by randomly choosing each entry from $\{+1, 0, -1\}$ with

δ	ρ	h^*	λ^*	τ^*	fMSE GENP-AMP	eMSE GENP-OWLQN	eMSE GENP-AMP	fMSE AMP	eMSE OWLQN	eMSE AMP	fMSE DN	eMSE DN	eMSE LMMSE
0.1	0.10	2.83	2.59	1.00	0.03	0.03	0.03	0.14	0.12	0.13	0.06	0.06	0.07
0.1	0.14	2.81	2.36	0.99	0.05	0.04	0.05	0.38	0.39	0.43	0.08	0.08	0.10
0.1	0.17	2.80	2.26	0.99	0.06	0.06	0.06	1.05	1.20	1.09	0.09	0.09	0.11
0.1	0.18	2.80	2.22	0.99	0.06	0.06	0.06	2.06	1.96	3.16	0.09	0.10	0.12
0.1	1.90	2.66	0.92	0.95	0.41	0.41	0.41	UB	UB	UB	0.49	0.48	0.53
0.25	0.13	2.58	2.03	1.00	0.09	0.09	0.09	0.37	0.37	0.37	0.15	0.15	0.17
0.25	0.20	2.55	1.80	0.99	0.12	0.12	0.12	1.03	1.21	1.14	0.20	0.20	0.21
0.25	0.24	2.53	1.69	0.99	0.14	0.14	0.14	2.83	2.71	2.91	0.23	0.23	0.24
0.25	0.25	2.53	1.66	0.99	0.15	0.15	0.15	5.58	6.67	5.68	0.24	0.24	0.25
0.25	1.90	2.28	0.51	0.97	0.62	0.63	0.63	UB	UB	UB	0.80	0.79	0.59
0.5	0.19	2.36	1.51	1.00	0.18	0.18	0.18	0.85	0.85	0.86	0.32	0.32	0.29
0.5	0.29	2.31	1.28	0.99	0.25	0.25	0.25	2.33	2.34	2.41	0.41	0.42	0.35
0.5	0.35	2.29	1.17	0.99	0.28	0.28	0.28	6.37	7.23	6.31	0.46	0.47	0.37
0.5	0.37	2.29	1.14	0.99	0.29	0.30	0.29	12.43	15.67	12.17	0.48	0.48	0.39
0.5	1.90	1.25	0.05	0.99	0.69	0.69	0.70	UB	UB	UB	0.98	0.97	0.46

Table 1: Empirical and predicted MSEs of different methods for different points in the sampling space.

probabilities $P(x_{0,i} = +1) = P(x_{0,i} = -1) = 0.064$. The entries of the measurement matrix A are drawn from the i.i.d. Gaussian distribution $\mathcal{N}(0, 1/m)$. The sampling noise w are drawn from $\mathcal{N}(0, 0.2)$, and the noise e of the GENP $\tilde{x}$ are drawn from $\mathcal{N}(0, 0.2\gamma_s^2)$. The simulation setup is the same as that in [9], except for the GENP.

As shown in Sec. 4, the MSE of GENP-LASSO is controlled by two regularization parameters λ and τ_s, but they are connected by the hidden parameter u. If one of them is given, using Prop. 4.1, Prop. 4.2, and Prop. 4.3, the other parameters can be uniquely determined.

Fig. 1 shows the predicted and the empirical MSEs of LASSO and GENP-LASSO with different λ. Three γ_s^2 are tested, each with two different values of n. In this example, the predicted MSEs of GENP-LASSO are given by the state evolution of GENP-AMP. The empirical results of LASSO and GENP-LASSO for $n = 200$ are obtained by the Matlab-based CVX package [43]. The empirical results of LASSO for $n = 2000$ are obtained by the OWLQN algorithm [44], which is written in C++. The empirical results of GENP-LASSO for $n = 2000$ are obtained by modifying the OWLQN to incorporate the GENP, as described in Sec. 3. We denote this as GENP-OWLQN.

It can be seen from Fig. 1 that the predicted MSE is quite accurate in both LASSO and GENP-LASSO. The result of LASSO (with $\gamma_s^2 = \infty$) is the same as Fig. 9 in [9]. When $\gamma_s^2 = 4$ or $\gamma_s^2 = 1$, the minimal MSE of GENP-LASSO can be reduced by about 20% and 50%, respectively, compared to the standard LASSO without any prior.

7.2. Comparison of AMP, GENP-AMP, Denoising and Least Squares

We now compared the performances of AMP, GENP-AMP, the LMMSE solution for Eq. (13), and scalar denoising via soft thresholding of the initial estimation when they are operated at different points of the sampling plane, including points below and above the phase transition boundary of the standard AMP. We will compare the predicted and empirical MSEs of GENP-AMP and AMP using the nearly-least-favorable signal generated by Eq. (10). We also use OWLQN and GENP-OWLQN to find the LASSO solution $\hat{x}(\lambda)$ and the GENP-LASSO solution $\hat{x}(\lambda, \tau_s)$ for Eq. (12), but OWLQN-based methods could not

γ_s^2	δ	ρ	h^*	λ^*	τ^*	fMSE (GENP-AMP)	eMSE (GENP-OWLQN)	eMSE (GENP-AMP)	fMSE (DN)	eMSE (DN)	eMSE (LMMSE)
	0.1	0.10	3.47	2.11	0.50	0.05	0.05	0.05	0.12	0.11	0.11
	0.1	0.14	3.51	1.88	0.50	0.07	0.08	0.08	0.16	0.13	0.15
2	0.1	0.17	3.54	1.78	0.50	0.09	0.09	0.09	0.18	0.16	0.17
	0.1	0.18	3.55	1.75	0.49	0.09	0.09	0.09	0.19	0.17	0.19
	0.1	1.90	3.72	0.63	0.45	0.79	0.81	0.81	0.97	0.87	1.03
	0.1	0.10	4.09	1.79	0.25	0.07	0.07	0.07	0.23	0.15	0.14
	0.1	0.14	4.27	1.54	0.25	0.11	0.11	0.11	0.32	0.21	0.23
4	0.1	0.17	4.38	1.43	0.25	0.13	0.13	0.13	0.36	0.24	0.29
	0.1	0.18	4.41	1.40	0.25	0.14	0.15	0.15	0.38	0.25	0.29
	0.1	1.90	5.22	0.40	0.20	1.57	1.57	1.57	1.94	1.46	2.05

Table 2: Empirical and predicted MSEs of different methods with different γ_s^2.

predict the MSE, and the regularized parameters need to be chosen manually. The number of iterations of GENP-AMP and AMP for empirical results is fixed as 60.

We first generate in each case 20 random realizations of size $n = 2000$, with parameters , $\gamma_s^2 = 1$, $\sigma^2 = 1$, $\delta \in \{0.10, 0.25, 0.50\}$, $\rho \in \{\frac{1}{2}\rho(\delta), \frac{3}{4}\rho(\delta), \frac{9}{10}\rho(\delta), \frac{19}{20}\rho(\delta), 1.9\}$, where $\rho(\delta)$ represents the phase transition boundary of the standard AMP. The results are summarized in Table 1, where eMSE and fMSE denote the empirical MSE and predicted formal MSE respectively. DN denotes the denoising method, and UB represents unbounded MSE. More results with different γ_s^2 are shown in Table 2.

Some observations can be drawn from Tables 1 and 2. First, the MSE of GENP-AMP is much lower than those of AMP and denoising. Secondly, the fMSE and eMSE of GENP-AMP match very well, even when the number of measurements is smaller than the sparsity. For example, for $\rho = 1.9$, the fMSE of GENP-AMP is still very close to eMSE. For AMP, this ρ is much higher than its phase transition boundary. Its MSE is thus unbounded. Thirdly, since the denoising method is equivalent to GENP-AMP with $\delta = 0$, the performance difference between GENP-AMP and denoising shows the contribution of the CS measurements. Moreover, the LMMSE solution is comparable to DN solution. The exceptions happen when $\delta = 0.25$, $\rho = 1.9$ and $\delta = 0.5$, $\rho = 1.9$. This can be expected since LMMSE can be interpreted as assuming the target signal x follows Gaussian distribution. When $\varepsilon = \delta\rho$ is sufficiently large, the distribution of x is close to Gaussian distribution, according to central limit theorem. In this case, the LMMSE result is near-optimal. Especially, when $\delta = 0.5$, $\rho = 1.9$, i.e., $\varepsilon = 0.95$, almost all entries of x are nonzero, LMMSE outperforms other methods. However, in all other cases, LMMSE is worse than our proposed algorithm.

Finally, although the empirical MSE of GENP-OWLQN is very similar to that of GENP-AMP, GENP-OWLQN is much slower, since it needs to calculate the gradients in each iteration. For example, on a computer with Intel Core i7 3.07GHz CPU and 6.00 GB memory, our Matlab implementation of GENP-AMP is about 10 times faster than the C++ implementation of GENP-OWLQN.

7.3. Performance of the Parameterless GENP-AMP

In the previous two simulations, the sparsity ε and the variance σ_s^2 of the prior $\tilde{x}$ are assumed to be known. In this subsection, we show the performance
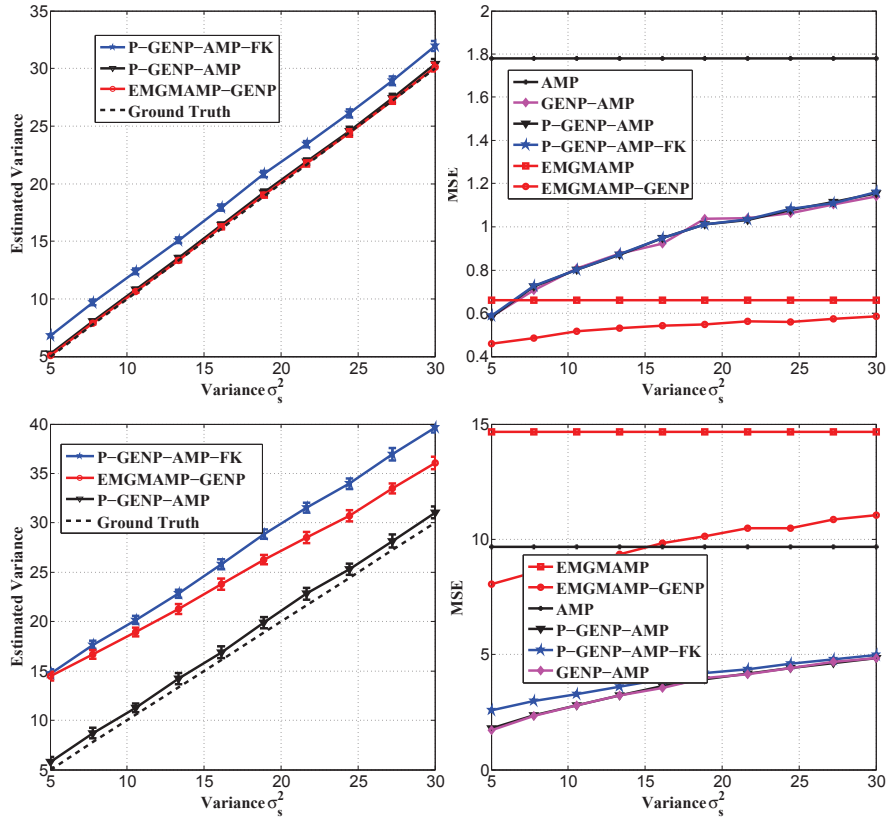
Figure 2: Performances of parameterless algorithms with $\delta = 0.5$ and $\varepsilon = 0.2$. First row (from left to right): (a) Estimated σ_s^2 with SNR=20 dB. The confidence level of the error bar is 0.95. (b) MSEs with SNR=20 dB. Second row: (c) Estimated σ_s^2 with SNR=5 dB. The confidence level of the error bar is 0.95. (d) MSEs with SNR=5 dB.

of the parameterless GENP-AMP (P-GENP-AMP), which can estimate σ_s^2. A similar setup to the previous experiments is used, except for the following. The non-zero coefficients of the sparse signal x follow i.i.d. $\mathcal{N}(0, 100)$. The sampling noise w are drawn from $\mathcal{N}(0, \sigma^2)$ where the variance σ^2 is set according to signal-to-noise ratio (SNR) defined as SNR $= 10\log_{10}(\frac{1}{m} \|Ax\|_2^2 / \sigma^2)$, and the noise e of the GENP $\tilde{x}$ are drawn from $\mathcal{N}(0, \sigma_s^2)$. The number of Monte-Carlo simulations is 100.

For comparison purpose, we also estimate σ_s^2 using the following method

$$\sigma_s^2 \approx \frac{1}{n} \|\tilde{x} - x_{\text{AMP}}\|_2^2, \tag{48}$$

i.e., we first reconstruct the sparse signal using standard CS reconstruction methods such as AMP, and then use the reconstructed signal and $\tilde{x}$ to estimate σ_s^2. And we name such kind of algorithm as Parameterless GENP-AMP with

20

σ^2, σ_s^2	δ	Alg1	Alg2	Alg3	Alg4	Alg5	Alg6	Alg7	Alg8	Alg9
1e3, 1e3	1/5	24.73	**26.34**	**26.40**	24.48	23.24	17.79	24.95	17.83	23.30
	1/2	26.15	**26.87**	**26.86**	20.58	20.04	18.14	24.95	26.49	23.78
1e3, 2.5e3	1/5	24.73	**25.91**	**25.89**	24.48	23.97	14.00	24.15	6.73	20.72
	1/2	26.15	**26.57**	**26.61**	20.58	19.75	14.70	24.15	26.03	21.52

Table 3: PSNRs of different methods for the reconstruction of "Lena". For $\sigma_s^2 = 1e3$, the PSNR of the corrupted upsampled version are all 18.13 dB, whereas when $\sigma_s^2 = 2.5e3$, the PSNR is 14.00 dB.

faked variance (P-GENP-AMP-FK). In fact, the only difference between Eq. (45) and Eq. (48) is the term $\lim_{t\to\infty} \hat{r}(\theta^t)/n$, the estimated MSE by the SURE framework proposed in [30].

We also compare with the method in [13], denoted as EMGMAMP, using its source code from [45]. We modify its source code to incorporate the GENP, and treat the variance of GENP as an additional hidden parameter, which can also be updated by the Expectation-Maximization algorithm in [13]. This algorithm is denoted as EMGMAMP-GENP in the following figures. The updating rule follows

$$\sigma_s^2(t) = \frac{1}{n} \sum_{i=1}^{n} [(\tilde{x}_i - \hat{x}_i(t))^2 + \mu_i^x(t)^2], \tag{49}$$

where $\hat{x}_i(t)$ and $\mu_i^x(t)$ is the approximate MMSE result, and its standard deviation in the t-th iteration, respectively.

In the first experiment, we consider a high SNR of 20 dB. From Fig. 2(a), we can see that P-GENP-AMP, and P-GENP-AMP-FK can both provide good approximations of the variance σ_s^2 while the gap between the ones estimated by P-GENP-AMP and GENP-AMP is exactly the MSE of AMP shown in Fig. 2 (b). It can also be seen from Fig. 2 (b) that all GENP-based algorithms achieve better performances. EMGMAMP-GENP outperforms the others, since it can learn the prior distribution of the sparse signal through EM and thus achieves near MMSE result. Although the full understanding of EM algorithm is still not available, its efficiency can be proven empirically in this high SNR example. On the other hand, both P-GENP-AMP and P-GENP-AMP-FK perform almost the same as GENP-AMP with known GENP variance. The reason is that at high SNR, the MSE of AMP is very small. Therefore Eq. (45) and Eq. (48) are very similar.

Fig. 2 (c) and (d) show the results with a low SNR of 5 dB. In this case, EMGMAMP-GENP no longer achieves an accurate estimate of σ_s^2, whereas the proposed P-GEMP-AMP still performs well. Moreover, P-GENP-AMP and GENP-AMP are still very close and are much better than other algorithms. The failure of EMGMAMP-GENP is because there are many approximations in EMGMAMP, *e.g.*, using the GAMP approximated posterior as the true one and learning the hidden parameters through EM. At low SNRs, these approximations are not accurate, and the method cannot achieve near MMSE result. Its performance can be even worse than the AMP.

Test sequence	σ^2, σ_s^2	δ	Alg1	Alg2	Alg3	Alg4	Alg5	Alg6	Alg7	Alg8	Alg9
Balloons	1e2, 1e2	1/5	31.27	**33.72**	**33.72**	32.65	**34.50**	27.25	32.04	32.31	32.18
		1/2	34.71	**35.63**	**35.79**	30.41	30.65	28.04	32.04	35.62	33.05
	1e2, 1e3	1/5	31.27	**32.71**	32.61	32.65	**33.20**	18.02	28.69	14.28	25.53
		1/2	34.71	**35.07**	**35.10**	30.43	30.20	19.45	28.69	32.91	27.59
	1e3, 1e3	1/5	27.83	**30.36**	**30.42**	27.08	25.70	18.01	28.69	15.38	24.48
		1/2	29.06	**30.87**	**30.94**	21.17	20.60	18.52	28.69	29.81	24.92
Kendo	1e2, 1e2	1/5	33.08	**35.88**	**35.82**	34.37	35.56	27.57	33.51	34.77	32.95
		1/2	36.22	**37.05**	37.04	30.79	30.89	28.28	33.51	**37.33**	33.47
	1e2, 1e3	1/5	33.08	34.73	**34.76**	34.37	**35.20**	18.07	30.20	16.77	25.85
		1/2	36.22	**36.63**	**36.64**	30.77	30.59	19.50	30.20	35.11	27.79
	1e3, 1e3	1/5	28.15	**31.86**	**32.00**	28.07	25.98	18.04	30.20	22.30	24.73
		1/2	30.26	**32.20**	**32.31**	21.32	20.64	18.57	30.20	31.04	25.14
Pantomime	1e2, 1e2	1/5	31.65	**34.41**	34.20	33.42	33.51	27.43	31.93	24.79	32.39
		1/2	**36.46**	36.24	36.36	30.89	30.29	28.20	31.93	**37.62**	33.16
	1e2, 1e3	1/5	31.65	33.73	**33.77**	33.42	**34.40**	18.06	29.77	24.58	25.66
		1/2	36.46	**36.62**	**36.66**	30.88	30.57	19.48	29.77	34.41	27.63
	1e3, 1e3	1/5	28.50	**31.39**	**31.49**	28.01	25.74	17.63	29.77	26.38	24.63
		1/2	30.32	**31.86**	**32.01**	21.34	20.66	18.56	29.77	31.11	25.11

Table 4: PSNRs of different methods for multiview images. For $\sigma_s^2 = 1e3$, the PSNRs of the corrupted virtual middle views are all 18.03 dB, whereas when $\sigma_s^2 = 1e2$, the PSNRs are 26.96 dB for "Balloons", 27.35 dB for "Kendo", and 27.20 dB for "Pantomime".
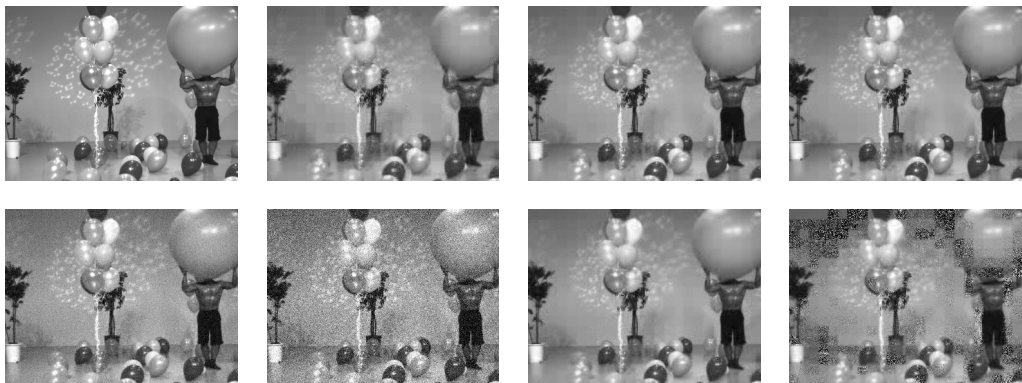

Figure 3: The reconstructed "Balloons" with $\sigma^2 = 1e3, \sigma_s^2 = 1e3, \delta = 1/5$. First row (from left to right): original, AMP (PSNR: 27.83dB) , P-GENP-AMP (30.36dB), GENP-AMP (30.42dB). Second row: EMGMAMP-GENP (25.70dB), Residual AMP (18.02dB), Denoising (28.69dB), and Modified CS (14.28dB).

7.4. Application in Compressive Image Sampling

We next consider a compressive image sensing example. The target image is the image "Lena" with resolution 512×512. We assume that the receiver has access to a 128×128 low-resolution version of the image, which is then upsampled to 512×512 and corrupted by Gaussian noises with different variances, to simulate the noises in poor illumination, high temperature, or transmission error. This is used as the GENP prior of our method.

The full size image is partitioned into overlapped blocks of size 48×48 pixels, with an overlap of 6 pixels to reduce the blocking artifacts. The DCT is used as the sparsifying transform. The same i.i.d. Gaussian sensing matrix is applied to each block to obtain the CS measurements. Eight algorithms are compared: AMP (denoted as Alg1), P-GENP-AMP (Alg2), GENP-AMP (Alg3), EMG-MAMP (Alg4), EMGMAMP-GENP (Alg5), the residual AMP similar to [21] (Alg6), the direct denoising of the prior image via soft-thresholding (Alg7), the

modified CS [27] (Alg8), which finds the sparsest signal outside the support set detected from the prior $\tilde{x}$, and the ℓ_1/ℓ_1 algorithm (Alg9) in [38] that adds the ℓ_1 constraint of the prior information to the ℓ_1 constraint in basis pursuit with $\beta = 1$. For the denoising algorithm, the parameterless SURE framework in [30] is applied to automatically choose the tuning parameter, and σ_s^2 is assumed to be known.

The results are summarized in Table 3. The top-two best results in each case are highlighted in bold. We can see that our proposed P-GENP-AMP and GENP-AMP always outperform other algorithms. Besides, at low SNRs ($\sigma^2 = 1e3$), the performance of EMGMAMP-GENP is quite poor. Note that the performance of Algorithms 4 and 5 degrade when given more samples, due to the instability of EM-based algorithms.

7.5. Application in Hybrid Multi-View Imaging System

We next apply the GENP-AMP to the hybrid multi-view imaging system [21, 24, 25], where a group of cameras capture the scene from different locations. Some cameras are traditional cameras, and others are CS cameras such as the single pixel cameras [22]. For each CS camera, we assume its left and right neighbouring cameras are traditional cameras. To help the reconstruction from CS sampling, the left and right views are used to generate a virtual view, which is corrupted by Gaussian noise and serves as the initial estimate or the GENP of the middle view.

We test the multiview image sequences "Balloons", "Kendo", and "Pantomime" under various channel noise levels. The setup is similar to Sec. 7.4. The virtual middle image is generated by Version 3.5 of the MPEG view synthesis reference software (VSRS) [46], and the test sequences are downloaded from [47].

Table 4 reports the PSNRs (dB) of the reconstructions given by the eight methods under different σ^2, σ_s^2, and δ. The following can be observed. First, almost all the top-two results are P-GENP-AMP and GENP-AMP, and there is no noticeable gap between them, verifying the efficiency of the proposed algorithms. In particular, when $\sigma^2 = 1e3$ and $\sigma_s^2 = 1e3$, *i.e.*, both the CS samples and GENP have low quality, our algorithms always perform the best. Second, when the channel noise level is low and sampling rate is high, *i.e.*, $\sigma^2 = 1e2$, $\sigma_s^2 = 1e2$, and $\delta = 1/2$, the modified CS (Alg6) is comparable to or even better than the proposed methods Alg2 and Alg3. This is as expected, since detecting the support of the virtual view $\tilde{x}$ is easier under low noise levels. However, as the noise level increases, the performance of the modified CS degrades quickly. It also requires the knowledge of σ^2, which is not needed in AMP-based algorithms. Third, at high SNR ($\sigma^2 = 1e2$), EMGMAMP-GENP outperforms the proposed P-GENP-AMP, but our method is better at low SNRs. Finally, Our methods are also about 20 times faster than the CVX-based modified CS and CVX-based ℓ_1/ℓ_1 method, and comparable to EMGMAMP and EMGMAMP-GENP.

Some examples of the reconstructed images are shown in Fig. 3. Our P-GENP-AMP and GENP-AMP provide the best visual quality. All other methods have some limitations. For example, some artifacts exist in the AMP

23

and EMGMAMP. Blurs happen when thresholding-based denoising is used, and Gaussian noises cannot be removed by the residual AMP. Although some parts can be well recovered by the modified CS, it also introduces severe artifacts in certain areas, due to its poor detection rate of the support set in high noise levels.

8. Conclusions and Future Work

This paper studies the generalized elastic net prior (GENP)-aided compressed sensing problem, where an additional noisy version of the original signal is available for CS reconstruction. We develop a GENP-aided approximate message passing algorithm (GENP-AMP), and study its parameter selection, state evolution, and noise sensitivity. The contribution of the GENP is also examined. We also develop a parameterless GENP-AMP that does not need to know the sparsity of the unknown signal and the variance of the GENP. Simulation results with 1-D data and two imaging applications demonstrate the performances of the proposed methods.

For the future work, a parameterless GENP-AMP algorithm that can accurately work in the whole plane needs to be developed. According to the noise sensitivity analysis in Sec. 5, there is no phase transition boundary, and the MSE is bounded in the whole plane. However, the parameterless GENP-AMP proposed in Sec. 6 only works well below the phase transition boundary of the standard AMP, due to the unbounded MSE above the phase transition boundary of the standard AMP and the approximation accuracy of SURE.

The original AMP is based on the simple soft thresholding in each iteration. Recently, it is found in [48, 49] that other denoising methods can be employed in AMP to further improve the reconstruction. For example, using the BM3D denoising algorithm [42], state-of-the-art CS reconstructions can be achieved in imaging applications. This approach can also be adopted into the GENP-AMP framework in this paper.

Applying the proposed schemes to multiview videos instead of multiview images is another attractive topic, where the approaches in [18, 19] could be useful. It is also worthwhile to find other applications of the proposed GENP-AMP method.

Appendix A. A heuristic derivation of the state evolution of GENP-AMP

In this section, we derive the state evolution of GENP-AMP in Eq. (30) of Sec. 4.3. The derivation is generalized from that in [9] for AMP. We start from the GENP-AMP iteration in (19) and (22), but introduce the following three modifications: (i) The random matrix A is replaced by a new i.i.d. $A(t)$ at each iteration t, where $A_{ij}(t) \sim N(0, 1/m)$; (ii) The corresponding observation becomes $y^t = A(t)x + w$; (iii) The last term in the update equation for r^t is

24

eliminated. We thus get the following dynamics:

$$x^{t+1} = \eta\left(\frac{u_t}{1+u_t}\tilde{x} + \frac{1}{1+u_t}(x^t + A(t)^T r^t); \theta_t\right), \tag{A.1}$$

$$r^t = y^t - A(t)x^t. \tag{A.2}$$

Eliminating r^t, the first equation becomes:

$$\begin{aligned}
x^{t+1} &= \eta\left(\frac{u_t}{1+u_t}\tilde{x} + \frac{1}{1+u_t}(A(t)^T y^t + (I - A(t)^T A(t))x^t; \theta_t\right) \\
&= \eta\left(x + \frac{u_t}{1+u_t}(\tilde{x} - x) + \frac{1}{1+u_t}(A(t)^T w + B(t)(x^t - x)); \theta_t\right),
\end{aligned} \tag{A.3}$$

where $B(t) = I - A(t)^T A(t)$.

Since the large system limit is assumed here, similar to [8], q_t^2 in Sec. 4.3 can be approximated by $\lim_{n\to\infty} \|x^t - x\|_2^2 / n$. It can be shown using the central limit theorem that $B(t)(x^t - x)$ converges to a vector with i.i.d. normal entries, and each entry has zero mean and variance q_t^2/δ. In addition, the entries of $A(t)^T w$ have zero mean and variance of σ^2, and they are independent of $B(t)(x^t - x)$. Therefore, each entry of the vectors in the argument of η in Eq. (A.3) converges to $X_0 + \xi_t Z$ with $Z \sim N(0,1)$ independent of X_0, and

$$\xi_t^2 = \left(\frac{u_t}{1+u_t}\right)^2 \sigma_s^2 + \left(\frac{1}{1+u_t}\right)^2 (\sigma^2 + \frac{1}{\delta}q_t^2). \tag{A.4}$$

On the other hand, by Eq. (A.3), each entry of $x^{t+1} - x$ converges to $\eta(X_0 + \xi_t Z; \theta_t) - X_0$. Therefore

$$q_{t+1}^2 = \lim_{n\to\infty} \frac{1}{n}\|x^{t+1} - x\|_2^2 = E\{[\eta(X_0 + \xi_t Z; \theta_t) - X_0]^2\}. \tag{A.5}$$

From Eq. (A.4) and Eq. (A.5), we can obtain the state evolution in Eq. (30).

This is a heuristic proof, more rigorous proof can be achieved following the proof in [50].

Appendix B. Proof of Proposition 5.1

In this part, we prove Prop. 5.1, which studies the bound of the MSE of the GENP-AMP in the (ρ, δ) plane.

Proof Consider $p_0 \in \mathcal{F}_{\delta\rho}$, $\sigma^2 = 1$ and let $\alpha^*(\delta, \rho) = \alpha^\pm(\delta\rho)$ minimax the MSE. To simplify the notation, we define

$$\begin{aligned}
\Psi(q^2, u; p) &= \Psi(q^2, u, \delta, \sigma = 1, \sigma_s, \alpha^*, p) \\
&= mse(npi(q^2, u, 1, \sigma_s, \delta); p, \alpha^*).
\end{aligned} \tag{B.1}$$

Then, by the definition of fixed point, we get

$$q_*^2 = \Psi(q_*^2, u^*; p),$$

$$u^* = \frac{1 + \frac{q_*^2}{\delta}}{\gamma_s^2}.$$

Using the scale invariance, we have $mse(\sigma^2; p, \alpha^*) = \sigma^2 mse(1; \tilde{p}, \alpha^*)$, where $\tilde{p}$ is a rescaled probability measure, $\tilde{p}\{x \cdot \sigma \in B\} = p\{x \in B\}$. For $p \in F_{\delta\rho}$, we have $\tilde{p} \in F_{\delta\rho}$ as well. Therefore,

$$\begin{aligned}
q_*^2 &= mse(npi(q_*^2, u^*, 1, \sigma_s, \delta); p, \alpha^*) \\
&= mse(1; \tilde{p}, \alpha^*) \cdot npi(q_*^2, u^*, 1, \sigma_s, \delta) \\
&\leqslant M^{\pm}(\delta\rho) \cdot npi(q_*^2, u^*, 1, \sigma_s, \delta)
\end{aligned}$$

Hence,

$$\frac{q_*^2}{npi(q_*^2, u^*; 1, \sigma_s, \delta)} \leqslant M^{\pm}(\delta\rho),$$

where we use the fact that $\sigma = 1$ and $\gamma_s = \sigma_s$.

By the definition of npi in Eq. (27), we have

$$\frac{q_*^2}{(\frac{u^*}{1+u^*})^2 \gamma_s^2 + (\frac{1}{1+u^*})^2(1 + \frac{q_*^2}{\delta})} \leqslant M^{\pm}(\delta\rho).$$

Replacing u^* by (31), we get

$$q_*^2 \leqslant \frac{-G(\delta, \rho, \gamma_s^2) + \sqrt{G(\delta, \rho, \gamma_s^2)^2 + 4\delta\gamma_s^2 M^{\pm}(\delta\rho)}}{2} \tag{B.2}$$

where $G(\delta, \rho, \gamma_s^2) = \delta\gamma_s^2 + \delta - \gamma_s^2 M^{\pm}(\delta\rho)$.

It is easy to verify that the phase transition boundary only exists when $\gamma_s^2 = \infty$ from the inequality above. If we let $(\gamma_s^2 + 1)\delta < \gamma_s^2 M^{\pm}(\delta\rho)$, $G(\delta, \rho, \gamma_s^2)$ in the right hand side of Eq. (B.2) is positive. In such case, if γ_s^2 goes to ∞, then $\delta < M^{\pm}(\delta\rho)$, we can get $q_*^2 \leqslant \infty$, i.e., the mean square error is unbounded, corresponding to the classical AMP phase transition boundary.

To prove the second part of Prop. 5.1, we make a specific choice $\bar{p}$ of p, and fix a small constant $c > 0$.

Now for $\varepsilon = \delta\rho$, define $h = h^{\pm}(\varepsilon, c) \cdot \sqrt{\mathrm{NPI}^*}$. Let $\bar{p} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{-h} + (\varepsilon/2)\delta_h$, similar to (10). Denote $q_*^2 = q_*^2(\bar{p})$ the highest fixed point corresponding to the signal distribution. Again, by the scale invariance, we have

$$\begin{aligned}
q_*^2 &= mse(npi(q_*^2, u^*, 1, \gamma_s, \delta); \bar{p}, \alpha^*) \\
&= mse(1; \tilde{p}, \alpha^*) \cdot npi(q_*^2, 1, \gamma_s, \delta),
\end{aligned}$$

where $\tilde{p}$ is a scaled probability measure, and $\tilde{p}\{x \cdot \sqrt{npi(q_*^2, 1, \gamma_s, \delta)} \in B\} =$

$\bar{p}\{x \in B\}$. Since $q_*^2 \leqslant M^*$, we have $\text{npi}(q_*^2, 1, \gamma_s, \delta) \leqslant \text{NPI}^*$ and hence

$$\frac{h}{\sqrt{\text{npi}(q_*^2, 1, \gamma_s, \delta)}} = h^\pm(\varepsilon, c) \cdot \sqrt{\frac{\text{NPI}^*}{\text{npi}(q_*^2, 1, \gamma_s, \delta)}} > h^\pm(\varepsilon, c).$$

Note that $\text{mse}(q; (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{-x} + (\varepsilon/2)\delta_x, \alpha)$ increases monotonically in $|x|$. Recall that $p_{\varepsilon,c} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{-h^\pm(\varepsilon,c)} + (\varepsilon/2)\delta_{h^\pm(\varepsilon,c)}$ is nearly-least-favorable for the minimax problem. Consequently,

$$\text{mse}(1; \tilde{p}, \alpha^*) \geqslant \text{mse}(1; p_{\delta\rho,c}, \alpha^*) = (1 - c) \cdot M^\pm(\delta, \rho).$$

By the scale-invariant property, we conclude that

$$\frac{q_*^2}{\text{npi}(q_*^2, 1, \gamma_s, \delta)} \geqslant (1 - c) \cdot M^\pm(\delta\rho).$$

Then, we can get the inequality

$$(q_*^2)^2 + [\delta(\gamma_s^2 + 1) - (1 - c)M^\pm(\delta, \rho)\gamma_s^2]q_*^2$$
$$- (1 - c)M^\pm(\delta\rho)\gamma_s^2\delta \geqslant 0.$$

Therefore,

$$\text{fMSE}(\alpha^*; \delta, \rho, 1, \gamma_s^2, \bar{p}) \geqslant \frac{-[\delta(\gamma_s^2 + 1) - (1 - c)M^\pm(\delta, \rho)\gamma_s^2]}{2}$$
$$+ \frac{\sqrt{[\delta(\gamma_s^2 + 1) - (1 - c)M^\pm(\delta, \rho)\gamma_s^2]^2 + 4(1 - c)M^\pm(\delta\rho)\gamma_s^2\delta}}{2},$$

where $\text{fMSE}(\alpha; \delta, \rho, \sigma, \gamma_s^2, p)$ is the equilibrium formal MSE for GENP-AMP (λ, τ_s) for the large system framework [8].

As $c > 0$ is arbitrary, we conclude

$$\sup_{p \in F_{\delta\rho}} \text{fMSE}(\alpha^*; \delta, \rho, 1, \gamma_s^2, p) \geqslant \frac{-[\delta(\gamma_s^2 + 1) - M^\pm(\delta, \rho)\gamma_s^2]}{2}$$
$$+ \frac{\sqrt{[\delta(\gamma_s^2 + 1) - M^\pm(\delta, \rho)\gamma_s^2]^2 + 4M^\pm(\delta\rho)\gamma_s^2\delta}}{2}.$$

Also, following the same procedure as Prop. 4.2 in [8], it can be shown that $M^* = \inf_\alpha \sup_{p \in F_{\delta\rho}} \text{fMSE}(\alpha; \delta, \rho, \sigma = 1, \gamma_s^2, p)$.

The last part of Prop. 5.1 can be proven by simply substituting the fixed point results in the second part of Prop. 5.1 for the ones in Eq. (29).

Appendix C. Proof of Proposition 6.1

In this part, we prove Prop. 6.1, which provides an accurate estimation of the variance of the prior $\tilde{x}$, $i.e.$, σ_s^2. This is an important step of the parameterless

GENP-AMP.

Proof From the definition of the GENP $\tilde{x}$, we get

$$
\begin{aligned}
\sigma_s^2 &= E[(\tilde{X} - X_0)^2] \\
&= E[(\tilde{X} - X_{\text{pos}} - X_0 + X_{\text{pos}})^2] \\
&= \underbrace{E[(\tilde{X} - X_{\text{pos}})^2]}_{(a)} + \underbrace{E[(X_0 - X_{\text{pos}})^2]}_{(b)} \\
&\quad - 2\underbrace{E[(\tilde{X} - X_{\text{pos}})(X_0 - X_{\text{pos}})]}_{(c)}
\end{aligned}
\tag{C.1}
$$

where X_{pos} is the estimated sparse signal by GENP-AMP based on an postulated variance $\sigma_{\text{s-pos}}^2$. $\tilde{X}$ and X_{pos} can be explicitly expressed as follows.

$$
\begin{aligned}
\tilde{X} &= X_0 + e, \ e \sim N(0, \sigma_s^2) \\
X_{\text{pos}} &= \eta(X_0 + \sigma_* Z; \theta), \ Z \sim N(0, 1),
\end{aligned}
\tag{C.2}
$$

where σ_*^2 is the variance of the unthresholded estimator in the last iteration of GENP-AMP.

Next, we look at each part of Eq. (C.1). Part (c) can be rewritten as

$$
E[(\tilde{X} - X_{\text{pos}})(X_0 - X_{\text{pos}})] = E[(X_0 - X_{\text{pos}})^2] + E[e(X_0 - X_{\text{pos}})].
\tag{C.3}
$$

Thus Eq. (C.1) becomes

$$
\sigma_s^2 = E[(\tilde{X} - X_{\text{pos}})^2] - E[(X_0 - X_{\text{pos}})^2] - 2E[e(X_0 - X_{\text{pos}})].
\tag{C.4}
$$

If $\sigma_{\text{s-pos}}^2$ is set to ∞, GENP-AMP degrades to AMP, which does not use $\tilde{x}$. This implies that a perfect candidate of X_{pos} is the signal recovered by AMP, X_{AMP}. Therefore, the two Gaussian noises $\sigma_* Z$ and e are uncorrelated. As a result, $E[e(X_0 - X_{\text{AMP}})] = 0$, and σ_s^2 can be further represented as

$$
\sigma_s^2 = E[(\tilde{X} - X_{\text{AMP}})^2] - E[(X_0 - X_{\text{AMP}})^2].
\tag{C.5}
$$

Part (a) can be rewritten as $E[(\widehat{X} - \eta(\widehat{X} + \sigma_* Z - e; \theta))^2]$ This term can exactly be seen as a denoising operator. According to the large system limit [8], when n is sufficiently large,

$$
E[(\tilde{X} - x_{\text{AMP}})^2] \approx \frac{\|\tilde{x} - x_{\text{AMP}}\|_2^2}{n}.
\tag{C.6}
$$

Next, $E[(X_0 - X_{\text{AMP}})^2]$ can be estimated by the method proposed in [30], inspired by the SURE theory. According to Theorem 4.3 and Theorem 4.7 in [30], it can be predicted by $\lim_{N \to \infty} \frac{\widehat{r}^t(\tau^t)}{N}$ when $t \to \infty$, where t is the inner iteration index of AMP. Usually it will converge in a few iterations.

Summarizing the analyses above, we can prove Prop. 6.1.

Acknowledgement

References

[1] E. J. Candès, T. Tao, Decoding by linear programming, IEEE Transaction on Information Theory 51 (12) (2005) 4203–4215.

[2] J. A. Tropp, A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transaction on Information Theory 53 (12) (2007) 4655–4666.

[3] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal in Imgace Sciences 2 (1) (2009) 183–202.

[4] S. Rangan, A. K. Fletcher, V. K. Goyal, Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing, IEEE Transaction on Information Theory 58 (3) (2012) 1902–1923.

[5] R. Tibshirani, Regression shrinkage and selection with the lasso, J. Royal. Statist. Soc. B 58 (1996) 267–288.

[6] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61.

[7] D. Donoho, A. Maleki, A. Montanari, Message passing algorithms for compressed sensing, Proceedings of the National Academy of Sciences 106 (45) (2009) 18914–18919.

[8] D. Donoho, A. Maleki, A. Montanari, The noise-sensitivity phase transition in compressed sensing, IEEE Transaction on Information Theory 57 (10) (2011) 6920–6941.

[9] A. Montanari, Graphical models concepts in compressed sensing, in: Compressed Sensing Theory and Applications, Cambrige University Press, 2012, pp. 394–438.

[10] A. Maleki, Approximate message passing algorithms for compressed sensing, Ph.D. thesis, Stanford University (2010).

[11] F. R. Kschischang, B. J. Frey, H. A. Loeliger, Factor graphs and the sum-product algorithm, IEEE Transaction on Information Theory 47 (2) (2001) 498–519.

[12] S. Rangan, Generalized approximate message passing for estimation with random linear mixing, arXiv: 1010.5141, Aug. 2012.

[13] J. P. Vila, P. Schniter, Expectation-Maximization Gaussian-mixture approximate message passing, IEEE Transaction on Signal Processing 61 (19) (2013) 4658–4672.

[14] M. Bayati, M. Lelarge, A. Montanari, Universality in polytope phase transitions and message passing algorithms, arXiv: 1207.7321, 2012.

[15] S. Rangan, P. Schniter, A. Fletcher, On the convergence of approximate message passing with arbitrary matrices, arXiv: 1402.3210, 2014.

[16] J. P. Vila, P. Schniter, An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals, arXiv: 1310.2806.

[17] S. Som, L. C. Potter, P. Schniter, On approximate message passing for reconstruction of non-uniformly sparse signals, in: IEEE National Aerospace and Electronics Conference, 2010, pp. 223–229.

[18] J. Ziniel, P. Schniter, Efficient high-dimensional inference in the multiple measurement vector problem, IEEE Transactions on Signal Processing 61 (2) (2013) 340–354.

[19] J. Ziniel, P. Schniter, Dynamic compressive sensing of time-varying signals via approximate message passing, IEEE Transactions on Signal Processing 61 (21) (2013) 5270–5284.

[20] Z. Xiong, A. D. Liveris, S. Cheng, Distributed source coding for sensor networks, IEEE Signal Process. Mag. (2004) 80–94.

[21] M. Trocan, T. Maugey, J. E. Fowler, B. Pesquet-Popescu, Disparity-compensation compressed-sensing reconstruction for multiview images, in: IEEE International Conference on Multimedia and Expo, 2010, pp. 1225–1228.

[22] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, R. G. Baraniuk, Single-pixel imaging via compressive sampling, IEEE Signal Processing Magazine 25 (2) (2008) 83–91.

[23] L. W. Kang, C. S. Lu, Distributed compressive video sensing, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 1169–1172.

[24] P. Beigi, X. Xiu, J. Liang, Compressive sensing based multiview image coding with belief propagation, in: Proc. Asilomar Conference on Signals, Systems, and Computers, 2010, pp. 430–433.

[25] X. Wang, J. Liang, View interpolation confidence-aided compressed sensing of multiview images, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2013, pp. 1651–1655.

[26] A. S. Charles, M. S. Asif, J. Romberg, C. J. Rozell, Sparsity penalties in dynamic system estimation, in: Conference on Information Science and Systems, 2011, pp. 1–6.

[27] N. Vaswani, W. Lu, Modified-CS: Modifying compressive sensing for problems with partially known support, IEEE Transaction on Signal Processing 58 (9) (2010) 4595–4607.

[28] H. Zou, T. Hastie, Regularization and variable selection via elastric net, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2) (2005) 301–320.

[29] J. Ziniel, P. Schniter, Binary linear classification and feature selection via generalized approximate message passing, arXiv: 1401.0872, Jan. 2014.

[30] A. Mousavi, A. Maleki, R. G. Baraniuk, Parameterlss optimal approximate message passing, arXiv: 1311.0035, Oct. 2013.

[31] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Proc. 56 (6) (2008) 2346–2356.

[32] M. W. Seeger, H. Nickisch, Compressed sensing and Bayesian experimental design, in: Proc. 25th Inter. Conf. Machine Learning, 2008, pp. 1–8.

[33] L. Wang, D. E. Carlson, M. R. D. Rodrigues, E. Calderbank, L. Carin, A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels, IEEE Trans. Inf. Theory 60 (5) (2014) 2611–2629.

[34] R. Baron, S. Sarvoham, R. G. Baraniuk, Bayesian compressive sensing via belief propagation, IEEE Trans. Signal Proc. 58 (1) (2010) 269–280.

[35] M. A. T. Figueiredo, R. D. Nowak, S. J. Wright, Gradient projection for sparse reconstruction, IEEE Journal of Selected Topics in Signal Processing 1 (4) (2007) 586–597.

[36] U. S. Kamilov, S. Rangan, A. K. Fletcher, M. Unser, Approximate message passing with consistent parameter estimation and applications to sparse learning, IEEE Trans. Inf. Theory 60 (5) (2014) 2969–2985.

[37] X. Wang, J. Liang, Side information-aided compressed sensing reconstruction via approximate message passing, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2014, pp. 3354–3358.

[38] J. Mota, N. Deligiannis, M. Rodrigues, Compressed sensing with prior information: Optimal strategies, geometry, and bounds, arXiv: 1408.5250, Aug. 2014.

[39] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, M. R. D. Rodrigues, Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information, arXiv: 1412.0614, 2014.

[40] L. Weizman, Y. Eldar, D. B. Bashat, Compressed sensing for longitudinal mri: An adaptive-weighted approach, arXiv: 1407.2602, Jul. 2014.

[41] D. L. Donoho, I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biomefrika 81 (3) (1994) 425–455.

[42] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, IEEE Trans. Image Proc. 16 (8) (2007) 2080–2095.

[43] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.0 beta, `http://cvxr.com/cvx` (Sep. 2013).

[44] G. Andrew, J. Gao, Scalable training of ℓ_1-regularized log-linear models, in: Proc. of International Conference on Machine Learning, 2007, pp. 33–40.

[45] S. Rangan, A. Fletcher, V. Goyal, U. Kamilov, J. Parker, P. Schniter, J. Vila, J. Ziniel, M. Borgerding, gampmatlab: Generalized approximate message passing, `http://sourceforge.net/projects/gampmatlab/files/` (May. 2014).

[46] M. Tanimoto, T. Fujii, K. Suzuki, View synthesis algorithm in view synthesis reference software 3.5 document m16090, ISO/IEC JTC1/SC29/WG11 (MPEG).

[47] Fujii lab's multi-view sequences download lists, `http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/`.

[48] J. Tan, Y. Ma, D. Baron, Compressive imaging via approximate message passing with image denoising, arXiv: 1405.4429, May 2014.

[49] C. A. Metzler, A. Maleki, R. G. Baraniuk, From denoising to compressed sensing, arXiv: 1406.4175, Jun. 2014.

[50] M. Bayati, A. Montanari, The dynamics of message passing on dense graphs, with applications to compressed sening, IEEE Trans. Inf. Theory 57 (2) (2011) 1462–1474.