BAYESIAN OPTIMAL COMPRESSED SENSING WITHOUT PRIORS: PARAMETRIC SURE APPROXIMATE MESSAGE PASSING

Chunli Guo, Mike E. Davies

School of Engineering and Electronics, The University of Edinburgh
{c.guo, mike.davies}@ed.ac.uk

ABSTRACT

It has been shown that the Bayesian optimal approximate message passing (AMP) technique achieves the minimum mean-squared error (MMSE) optimal compressed sensing (CS) recovery. However, the prerequisite of the signal prior makes it often impractical. To address this dilemma, we propose the parametric SURE-AMP algorithm. The key feature is it uses the Stein's unbiased risk estimate (SURE) based parametric family of MMSE estimator for the CS denoising. Given that the optimization of the estimator and the calculation of its mean squared error purely depend on the noisy data, there is no need of the signal prior. The weighted sum of piecewise kernel functions is used to form the parametric estimator. Numerical experiments on both Bernoulli-Gaussian and k-dense signal justify our proposal.

Index Terms— Compressed sensing, approximate message passing, SURE estimator, denoising

1. INTRODUCTION

The analysis of the approximate message passing (AMP) technique suggests that the compressed sensing (CS) reconstruction can be interpreted as a recursive denoising problem: for each iteration, we observe a data set which is the original signal corrupted by some white noise [1]. The ℓ_1-AMP algorithm, behaving like LASSO, is proved to have a robust reconstruction that is invariant to the signal prior $p_{\mathbf{x}}(x)$ [1,2]. It iteratively solves a 1-d denoising that is minimax with respect to the pdf from which $\mathbf{x}$ is drawn. However, in CS we do not have only one but many parallel 1-d denoising problems. In contrast, Bayesian optimal AMP (BAMP) achieves a minimum mean squared error (MMSE) optimal performance by incorporating $p_{\mathbf{x}}(x)$ with the denoising procedure [3,4]. However, the requirement of $p_{\mathbf{x}}(x)$ to be known in advance can be restrictive. The benefit of using the density information and the fact that all noisy data naturally form a pdf motivate us to find an alternative reconstruction approach that performs as well as BAMP yet without the need of the signal prior.

Chunli Guo gratefully acknowledge the Maxwell Advanced Technology Fund at the University of Edinburgh in funding her PhD studies.

Towards this end, we propose the parametric SURE-AMP algorithm. The fundamental idea is to replace the MMSE estimator in BAMP with a Stein's unbiased risk estimate (SURE) based [5] parametric least square estimator. The virtue of the parametric estimator is that its optimization and MSE could be reformulated purely as a sample average of the corrupted data. In our work, we choose a parametric family of MMSE estimators to be the piecewise linear kernel functions. Most importantly, the kernel structures are not specifically designed to fit the signal prior but are inspired from its general sparsity pattern. The numerical experiments with both Bernoulli-Gaussian and k-dense data show that with a limited number of kernels, we are able to capture the evolving shape of the MMSE estimator in the AMP iteration.

The fact that signal distributions are rarely known a priori has been noticed and a number of AMP-based algorithms have been proposed to tackle this problem. In [4, 6–8], the mixture of Gaussians is used as the parametric representation for $p_{\mathbf{x}}(x)$ and the expectation-maximization (EM) approach is deployed to jointly learn the prior along with the estimation $\mathbf{x}$. Later in [9], the authors introduced an adaptive AMP framework which is not limited to the sum of Gaussians model. The key difference between the EM-GAMP and our proposal is that we resort to a parametric family of MMSE estimators instead of fitting the prior. The experiments suggest when the signal prior is easy to imitate, our parametric SURE-AMP serves as an alternative with an equivalent performance as EM-GAMP. When the signal prior is difficult to be approximated as the finite sum of Gaussians, parametric SURE-AMP provides a better solution.

Notation: For the rest of the paper, we use boldface capital letters e.g. $\mathbf{A}$, to represent matrices, and $\mathbf{A}^T$ to denote the transpose. We use boldface small letters like $\mathbf{x}$ to denote vectors and x_i to represent its i^{th} element. For a vector $\mathbf{x} \in \mathbb{R}^n$, we use $< \mathbf{x} >= \frac{1}{n}\sum_i x_i$ to represent its average.

2. PARAMETRIC SURE-AMP FRAMEWORK

We consider the standard CS setting: $\mathbf{y} = \mathbf{\Phi x} + \mathbf{w}$, with $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$, $\gamma = \frac{m}{n} < 1$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{w}; \mathbf{0}, \sigma_{\mathbf{w}}^{\mathbf{2}})$. In this section, we start with a simplified formulation of the Bayesian optimal AMP algorithm. Then we introduce the parametric

SURE estimator as a suitable surrogate for the MMSE estimator in the BAMP framework. Finally we present the parametric SURE-AMP algorithm.

2.1. Bayesian optimal AMP revisit

The intuitive idea behind the AMP algorithm is that for each iteration, we observe a white noise corrupted signal $\boldsymbol{\mu} = \mathbf{x} + \mathbf{s}$, $\mathbf{s} \sim \mathcal{N}(s; 0, \sigma_s^2)$. The reconstruction procedure is a recursive denoising problem until the noise variance σ_s^2 is decreased to a satisfactory level. Given the accurate prior $p_{\mathbf{x}}(x)$, the denoising is performed point-wise with the MMSE estimator in BAMP:

$$F(\boldsymbol{\mu}_i; \sigma_s^2) = \mathbb{E}_{x|\mu}(x|\mu = \boldsymbol{\mu}_i, \sigma_s^2) \tag{1}$$

and the conditional variance of x given that $\mu = \boldsymbol{\mu}_i$ is defined as:

$$G(\boldsymbol{\mu}_i; \sigma_s^2) = \mathrm{Var}_{x|\mu}(x|\mu = \boldsymbol{\mu}_i, \sigma_s^2) \tag{2}$$

It has been shown in [10] that for additive Gaussian noise, the optimal least square estimator $F(;)$ can be expressed entirely as a function of the noisy data.

$$F(\boldsymbol{\mu}_i; \sigma_s^2) = \boldsymbol{\mu}_i + \sigma_s^2 \frac{p'_{\boldsymbol{\mu}}(\boldsymbol{\mu}_i)}{p_{\boldsymbol{\mu}}(\boldsymbol{\mu}_i)} \tag{3}$$

By leveraging the property of the MMSE estimator (3) and with some straightforward calculations, we have the following relationship between $F(;)$ and $G(;)$ for any Gaussian noise corrupted data.

$$F'(\mu; \sigma_s^2) = \frac{1}{\sigma_s^2} G(\mu; \sigma_s^2) \tag{4}$$

Putting (4) into the generic BAMP algorithm [3,4,11], we have the simplified BAMP algorithm, as summarized below.

Algorithm 1 : Simplified Bayesian optimal AMP

1: **initialization:** $\mathbf{x}^0 \leftarrow \mathbf{0}$, $\mathbf{z}^0 \leftarrow \mathbf{y}$, $c^0 > \sigma_x^2$
2: **for** $t = 1, 2, \cdots$ **do**
3: $\quad \boldsymbol{\mu}^t = \boldsymbol{\Phi}^T \mathbf{z}^t + \mathbf{x}^t$
4: $\quad \mathbf{x}^{t+1} = F(\boldsymbol{\mu}^t; c^t)$
5: $\quad \nu^{t+1} = < G(\boldsymbol{\mu}^t; c^t) >$
6: $\quad \mathbf{z}^{t+1} = \mathbf{y} - \boldsymbol{\Phi}\mathbf{x}^{t+1} + \frac{\nu^{t+1}}{\gamma c^t}\mathbf{z}^t$
7: $\quad c^{t+1} = \sigma_w^2 + \frac{\nu^{t+1}}{\gamma}$
8: **end for**

The key yet impractical component of BAMP is the requirement of the signal prior in order to define $F(;)$ and $G(;)$. However, the AMP theory works in the large scale limit with $\mathbf{x}$ tending to some empirical distribution. In this case, the noisy estimates also follow an (observed) empirical distribution $p_{\boldsymbol{\mu}}(\mu)$ which we have access to. Thus the blindness of $p_{\mathbf{x}}(x)$ should not prevent us from achieving performance on par with the MMSE recovery. This motivates us to find a proper substitution of $F(;)$.

2.2. Parametric SURE estimator

We resort to a SURE-based parametric approximation of the MMSE estimator in the BAMP algorithm. SURE is a statistically unbiased estimator of the mean squared error (MSE) of an arbitrary estimator [5]. Since the MMSE estimator can be reformed as (3), we consider a family of estimators f_θ, parameterized by the vector $\boldsymbol{\theta}$:

$$f_\theta(\mu; \sigma_s^2) = \mu + g_\theta(\mu; \sigma_s^2) \tag{5}$$

It has been shown in [5] that for the Gaussian noise corrupted data, the SURE estimate of $f_\theta(;)$ can be expressed as an expectation over the noisy observation:

$$\begin{aligned} \nu =& \mathbb{E}_{\mathbf{x},\boldsymbol{\mu}}[(\mathbf{x} - f_\theta(\boldsymbol{\mu}; \sigma_s^2))^2] & (6) \\ =& \sigma_s^2 + \mathbb{E}_{\boldsymbol{\mu}}[g_\theta^2(\boldsymbol{\mu}; \sigma_s^2) + 2\sigma_s^2 g'_\theta(\boldsymbol{\mu}; \sigma_s^2)] & (7) \end{aligned}$$

In practice, the expectation can be approximated with the average over the observation $\boldsymbol{\mu}$. We can therefore use (7) to optimize $\boldsymbol{\theta}$ as in [10, 12].

$$\hat{\boldsymbol{\theta}} = \arg\min_\theta < g_\theta^2(\boldsymbol{\mu}; \sigma_s^2) + 2\sigma_s^2 g'_\theta(\boldsymbol{\mu}; \sigma_s^2) > \tag{8}$$

2.3. Parametric SURE-AMP

With $f_\theta(;)$ and ν solely depending on the noisy data, they can serve as a natural alternative of the MMSE estimator in the BAMP framework. Thus we present the parametric SURE-AMP algorithm below.

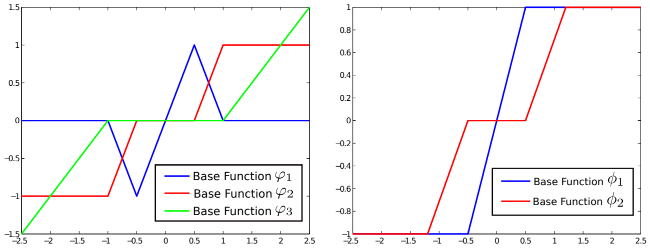
Algorithm 2 : Parametric SURE-AMP

1: **initialization:** $\mathbf{x}^0 \leftarrow \mathbf{0}$, $\mathbf{z}^0 \leftarrow \mathbf{y}$, $c^0 > \sigma_x^2$
2: **for** $t = 1, 2, \cdots$ **do**
3: $\quad \boldsymbol{\mu}^t = \boldsymbol{\Phi}^T \mathbf{z}^t + \mathbf{x}^t$
4: $\quad \mathbf{x}^{t+1} = f_\theta(\boldsymbol{\mu}^t; c^t)$
5: $\quad \nu^t = c^t + < g_\theta^2(\boldsymbol{\mu}^t; c^t) + 2c^t g'_\theta(\boldsymbol{\mu}^t; c^t) >$
6: $\quad \mathbf{z}^{t+1} = \mathbf{y} - \boldsymbol{\Phi}\mathbf{x}^{t+1} + \frac{\nu^{t+1}}{\gamma c^t}\mathbf{z}^t$
7: $\quad c^{t+1} = \sigma_w^2 + \frac{\nu^{t+1}}{\gamma}$
8: **end for**

The critical part is to find the proper parameter family to capture the evolving shape of the MMSE estimator. In our work, we choose the piecewise linear estimator.

3. COMPRESSED SENSING EXAMPLES

The reconstruction quality of the parametric SURE-AMP primarily counts on how accurately $f_\theta(;)$ can approximate the MMSE estimator $F(;)$. The common practice is to form the SURE estimator as a weighted sum of some kernel functions: In [10], authors used the "bump" functions as the kernels. In [12], exponential kernels are exploited. In this work, we choose the piecewise linear functions and show their effectiveness through the denoising of the Bernoulli-Gaussian and k-dense data.



(a) (b)

Fig. 1. Piecewise linear kernel functions, as used for linear parameterization of SURE: (a) Bernoulli-Gaussian signal (b) k-dense signal.

3.1. Bernoulli-Gaussian signal

For Bernoulli-Gaussian data, $\mathbf{x}$ is drawn i.i.d from the pdf

$$p_{\text{BG}}(x) = \lambda\mathcal{N}(x;0,1) + (1-\lambda)\delta(x) \tag{9}$$

We build a linearly parameterized estimator of the form:

$$f_{\text{BG}}(\mu) = a_1\varphi_1(\mu) + a_2\varphi_2(\mu) + a_3\varphi_3(\mu) \tag{10}$$

with the kernel functions being

$$\varphi_1(\mu) = \begin{cases} 0 & \mu \leq -2\alpha_1, \mu \geq 2\alpha_1 \\ -\frac{\mu}{\alpha_1} - 2 & -2\alpha_1 \leq \mu \leq -\alpha_1 \\ \frac{\mu}{\alpha_1} & -\alpha_1 \leq \mu \leq \alpha_1 \\ -\frac{\mu}{\alpha_1} + 2 & \alpha_1 \leq \mu \leq 2\alpha_1 \end{cases} \tag{11}$$

$$\varphi_2(\mu) = \begin{cases} -1 & \mu \leq -\alpha_2 \\ \frac{\mu+\alpha_1}{\alpha_2-\alpha_1} & -\alpha_2 < \mu < -\alpha_1 \\ 0 & -\alpha_1 \leq \mu \leq \alpha_1 \\ \frac{\mu-\alpha_1}{\alpha_2-\alpha_1} & \alpha_1 < \mu < \alpha_2 \\ 1 & \mu \geq \alpha_2 \end{cases} \tag{12}$$

$$\varphi_3(\mu) = \begin{cases} \mu + \alpha_2 & \mu \leq -\alpha_2 \\ 0 & -\alpha_2 < \mu < \alpha_2 \\ \mu - \alpha_2 & \mu \geq \alpha_2 \end{cases} \tag{13}$$

The hinge points α_1 and α_2 very much characterize the shape of the kernel function $\varphi_i(\mu)$. We adopt the recommendation in [13] and fix the hinge points to be proportional to the standard deviation of the noise

$$\alpha_1 = 2\sigma_s \qquad \alpha_2 = 4\sigma_s \tag{14}$$

Thus the denoising function $f_{\text{BG}}(\cdot)$ only depends linearly on the parameter set $\boldsymbol{\theta}_{\text{BG}} = [a_1, a_2, a_3]$. The exact optimization of $\boldsymbol{\theta}_{\text{BG}}$ is straightforward because the MSE estimate (7) has a quadratic form. Taking the derivative of the MSE with respect to a_i and setting to zero we have

$$\frac{d\nu}{da_i} = 2 < g_\theta(\boldsymbol{\mu};\sigma_s^2)\varphi_i(\boldsymbol{\mu};\sigma_s^2) > +2\sigma_s^2 < \varphi_i'(\boldsymbol{\mu};\sigma_s^2) >= 0 \tag{15}$$

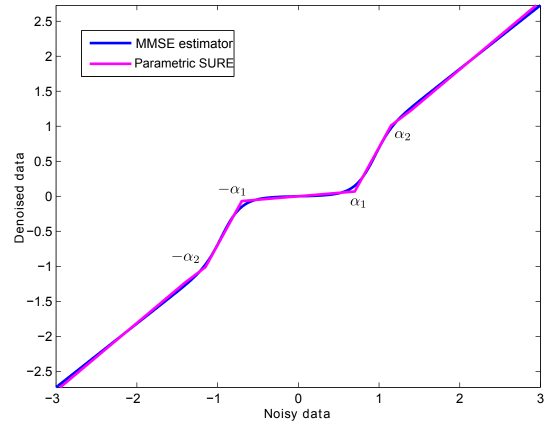


Fig. 2. MMSE estimator and parametric SURE for the noisy Bernoulli-Gaussian data. The noise variance σ_s^2 is 0.1. Hinge points are 0.64 and 1.26. The MSE for the MMSE estimator, the parametric SURE and the estimation using (7) are 0.0204, 0.0216 and 0.0222.

which leads to

$$\sum_k < \varphi_k(\boldsymbol{\mu};\sigma_s^2)\varphi_i(\boldsymbol{\mu};\sigma_s^2) > a_k = -\sigma_s^2 < \varphi_i'(\boldsymbol{\mu};\sigma_s^2) > \tag{16}$$

The summary of these equations form a linear system which can be solved directly via a simple matrix inversion.

Fig. 1.(a) is a plot of the kernel functions for the Bernoulli-Gaussian data. Fig. 2 illustrates both the MMSE estimator and its optimized-SURE approximation. We can see that the simple form of piecewise linear function is good enough to capture the key structure of the MMSE estimator.

3.2. K-dense signal

In this section we consider the k-dense signal which is a mixture of continuous and discrete elements.

$$p_{\text{KD}}(x) = \frac{(1-\lambda)}{2}\delta(x+1) + \frac{(1-\lambda)}{2}\delta(x-1) + \lambda\mathcal{U}(-1,1) \tag{17}$$

where $\mathcal{U}(a,b)$ denotes uniform distribution in the interior (a,b). The recovery of the noisy under-determined linear observation for such model is investigated in [1, 14]. It is proved that we need $\gamma > 0.5$ to make the recovery possible with linear programming. We observed a similar behavior for BAMP with Gaussian $\mathbf{\Phi}$. However the reconstruction quality is significantly improved.

The parametric estimator for the k-dense signal are formed as

$$f_{\text{KD}}(\mu) = b_1\phi_1(\mu) + b_2\phi_2(\mu) \tag{18}$$

where ϕ_1 and ϕ_2 is defined as:

$$\phi_1(\mu) = \begin{cases} -1 & \mu \leq -\beta_1 \\ \frac{\mu}{\beta_1} & -\beta_1 < \mu < \beta_1 \\ 1 & \mu \geq \beta_1 \end{cases} \tag{19}$$

$$\phi_2(\mu) = \begin{cases} -1 & \mu \le -\beta_2 \\ \frac{\mu+\beta_1}{\beta_2-\beta_1} & -\beta_2 < \mu < -\beta_1 \\ 0 & -\beta_1 \le \mu \le \beta_1 \\ \frac{\mu-\beta_1}{\beta_2-\beta_1} & \beta_1 < \mu < \beta_2 \\ 1 & \mu \ge \beta_2 \end{cases} \tag{20}$$

For the k-dense signal, the parameter vector $\boldsymbol{\theta}_{\text{KD}}$ has both linear (kernel weight) and nonlinear (hinge point) elements to be optimized. We resort to the gradient descend method to solve this problem. The optimization is not easy though, since the objective function $\nu(b_1, b_2, \beta_1, \beta_2)$ is not convex and possesses a lot of local minimas. To tackle this problem, we set up a searching grid for the starting position of the hinge points β_1, β_2 and choose the one delivers the least MSE.

1: **for** each starting point β_1^0, β_2^0 **do**
2: **for** $t = 1, 2, \cdots$ **do**
3: $b_1^t, b_2^t \in \arg\min_{b_1,b_2} \nu(\beta_1^{t-1}, \beta_2^{t-1}, b_1, b_2)$
4: $\beta_1^t \in \arg\min_{\beta_1} \nu(\beta_1, \beta_2^{t-1}, b_1^t, b_2^t)$
5: $\beta_2^t \in \arg\min_{\beta_2} \nu(\beta_1^t, \beta_2, b_1^t, b_2^2)$
6: **end for**
7: **end for**

The kernel functions for constructing the k-dense parametric SURE estimator are presented in Fig.1.(b). The MMSE estimator and SURE approximation for the k-dense signal are shown in Fig. 3. When implementing the parametric SURE-AMP, the region of the searching grid can be reduced over the AMP iteration, with the optimal hinge position from the previous iteration being the maximum searching point in the grid.

One thing worth noting is that the kernel structures are not specially designed to fit the Dirac and uniform combination. They are constructed based on the sparsity pattern (the position of the Dirac function). We believe these kernel structures are suitable for other k-dense models with the continuous components follow any smooth pdf.

4. SIMULATION

In this section, the reconstruction performance of the parametric SURE-AMP is reported for the CS noisy Bernoulli-Gaussian and k-dense data. For all experiments, we fixed $n = 10000$, $\lambda = 0.1$ and each numerical point is an average of 100 Monte Carlo iterations. The noise level in the measurement domain is known and quantified as $\text{SNR}_y = 10\log_{10} \| \mathbf{\Phi x} \|_2^2 \,/\, \| \mathbf{w} \|_2^2$. The reconstruction quality is evaluated in terms of the signal to noise ratio in the signal domain, defined as $\text{SNR}_x = 10\log_{10} \| \mathbf{x} \|_2^2 \,/\, \| \hat{\mathbf{x}} - \mathbf{x} \|_2^2$.

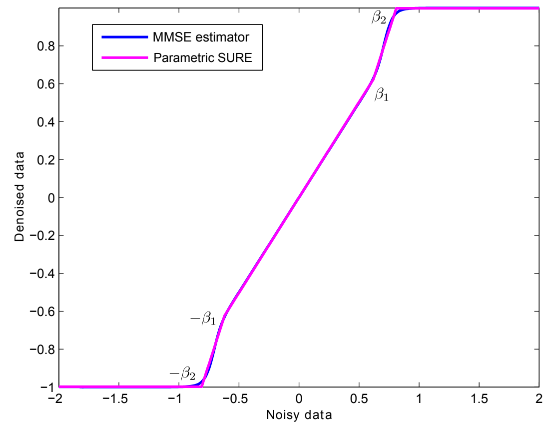


Fig. 3. MMSE estimator and parametric SURE for the noisy k-dense data. The noise variance σ_s^2 is 0.1. Hinge points are 0.62 and 0.81. The MSE for the MMSE estimator, the parametric SURE and the estimation using (7) are 1.810×10^{-3}, 1.825×10^{-3} and 1.737×10^{-3}.

4.1. Bernoulli-Gaussian

In [6], the authors have demonstrated that the EM-GM-AMP algorithm achieves the state-of-art performance compared to most of the existing CS algorithms that are blind to the signal prior. For comparison, we show the performance of the proposed parametric SURE-AMP, EM-GAMP, LASSO (via SPGL1 [15]) and BAMP (with true $p_{\text{BG}}(x)$) in Fig. 4. For LASSO, we used the same setting as described in [6]. The noise level is $\text{SNR}_y = 26$ dB.

As shown in Fig. 4, parametric SURE-AMP exhibits significant improvement over ℓ_1 minimization: the more than 7 dB increase of SNR_x demonstrates that exploring the density information indeed helps with CS reconstruction. Compared to EM-GM-AMP, our proposal delivers very close performance for γ between 0.36 and 0.5 and is roughly 0.5 dB better for γ between 0.24 and 0.36. Parametric SURE-AMP brings the sampling ratio breakpoint down to 0.24 while EM-GM-AMP pushes it down to 0.22. When compared with the BAMP result, parametric SURE-AMP exhibits nearly identical performance for γ larger than 0.36. In the small sampling ratio regime, there is about 2 dB gap to fill. However, given the hinge points are fixed for the parametric SURE-AMP in this simulation, we believe it has the potential to improve SNR_x even further with proper scheme to optimize the hinge points.

4.2. k-dense signal

Fig. 5 shows SNR_x for noisy recovery of the k-dense signal with $\text{SNR}_y = 46$ dB. Again we compare the parametric SURE-AMP with EM-GM-AMP, BAMP and convex optimization. The inverse incomplete linear system of the k-dense signal forms a convex problem [14]: $\hat{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \| \mathbf{y} - \mathbf{\Phi}\tilde{\mathbf{x}} \|_2$, s.t. $\| \tilde{\mathbf{x}} \|_\infty \le 1$, which can be solved with the

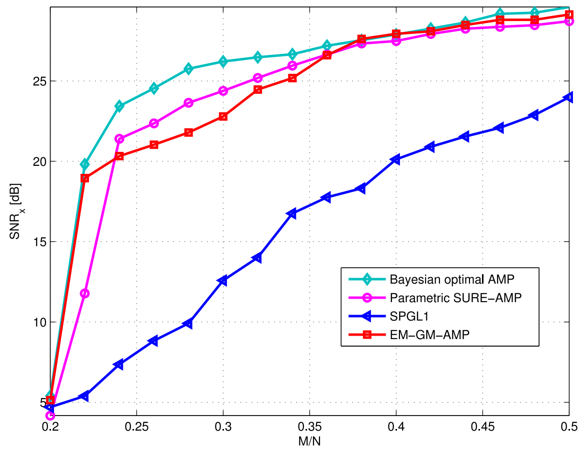


Fig. 4. Noisy Bernoulli-Gaussian CS recovery for different schemes under different sampling ratio.

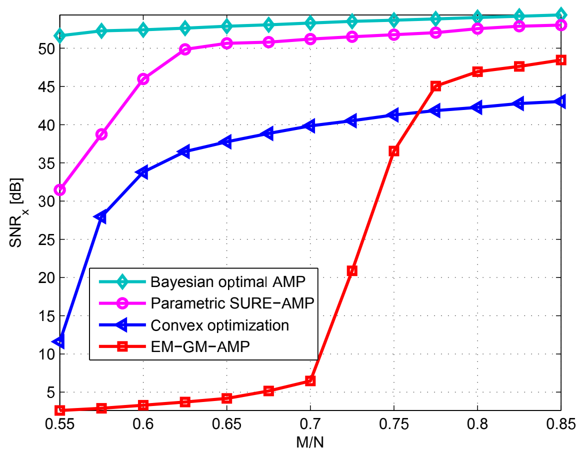


Fig. 5. Noisy k-dense CS recovery for different schemes under different sampling ratio.

gradient projection. For the EM-GM-AMP, we set he number of Gaussian to be the maximum 20 to approximate $p_{\text{KD}}(x)$.

With all kernel weights and hinge points optimized, parametric SURE-AMP achieves a nearly optimal performance for $\gamma > 0.625$: it is roughly 1.5 dB worse than BAMP. For $\gamma < 0.625$, its performance diverges from the Bayesian optimal curve due to the difficulty of locating the optimal hinge points. Again we observe an overall improvement over the convex minimization. The EM-GM-AMP algorithm experiences a failure for $\gamma < 0.725$ and is around 5 dB worse than the parametric SURE-AMP in the large sampling ratio regime. It demonstrates that when approximating the prior with Gaussian mixture models turns out to be a difficult task, resorting to its MMSE estimator approximation can be an ideal option.

5. CONCLUSION

In the paper, we present the parametric SURE-AMP algorithm, a complementary scheme for the EM-GM-AMP algorithm with the recovery quality approaching the Bayesian optimal. The selection of kernel functions plays a crucial role, similar to the parametric prior family in the EM-GM-AMP. Direction for further research would be: consider other parametric families, design better hinge selection scheme, and apply the parametric SURE-AMP to compressible signals.

REFERENCES

[1] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. of the Nat. Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[2] A. Maleki and A. Montanari, “Analysis of approximate message passing algorithm,” in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, March 2010, pp. 1–7.

[3] D. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction,” in *IEEE Inf. Theory Workshop (ITW)*. Dublin, Ireland, 2010, pp. 1–5.

[4] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices,” *J. Stat. Mech.*, vol. P08009, Aug. 2012.

[5] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *The annals of Statistics*, pp. 1135–1151, 1981.

[6] J. P. Vila and P. Schniter, “Expectation-maximization gaussian-mixture approximate message passing,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4658–4672, Oct 2013.

[7] J. P. Vila and P. Schniter, “Expectation-maximization bernoulli-gaussian approximate message passing,” in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, Nov 2011, pp. 799–803.

[8] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Statistical-physics-based reconstruction in compressed sensing,” *Phys. Rev. X*, vol. 2, pp. 021005(1–18), May 2012.

[9] U. Kamilov, S. Rangan, A. K. Fletcher, and M. Unser, “Approximate message passing with consistent parameter estimation and applications to sparse learning,” *IEEE Trans on Information Theory*, pp. 2969–2985, 2014.

[10] M. Raphan and E. P. Simoncelli, “Least squares estimation without priors or supervision,” *Neural computation*, vol. 23, no. 2, pp. 374–420, 2011.

[11] S. Som and P. Schniter, “Compressive imaging using approximate message passing and a markov-tree prior,” *IEEE Trans. on Signal Process.*, vol. 60, no. 7, pp. 3439–3448, July 2012.

[12] F. Luisier, T. Blu, and M. Unser, “A new sure approach to image denoising: Interscale orthonormal wavelet thresholding,” *Image Processing, IEEE Transactions on*, vol. 16, no. 3, pp. 593–606, March 2007.

[13] D.L. Donoho, Y. Tsaig, I. Drori, and J-L Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, Feb 2012.

[14] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[15] E. van den Berg and M. P. Friedlander, “Probing the pareto frontier for basis pursuit solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.