Near Optimal Compressed Sensing without Priors: Parametric SURE Approximate Message Passing

Chunli Guo, University College London
Mike E. Davies, University of Edinburgh

Talk Outline

- Motivation for Parametric SURE-AMP
  - What is approximate message passing (AMP) algorithm ?
  - Iterative Gaussian denoising nature of AMP

- Parametric SURE-AMP Algorithm
  - SURE based denoiser design
  - Parameterization & optimization of denoisers

- Numerical Reconstruction Examples

What is AMP ?

- The CS reconstruction problem $y = \Phi x_0$ with $\Phi \in \mathbb{R}^{m\times n}$, $m < n$
- The Generic AMP algorithm for i.i.d Gaussian Φ [Donoho 09]
- Initialized with $\hat{x}^0 = 0,\ z^0 = y$

For t = 0, 1….

$$r^t = \hat{x}^t + \Phi^T z^t$$

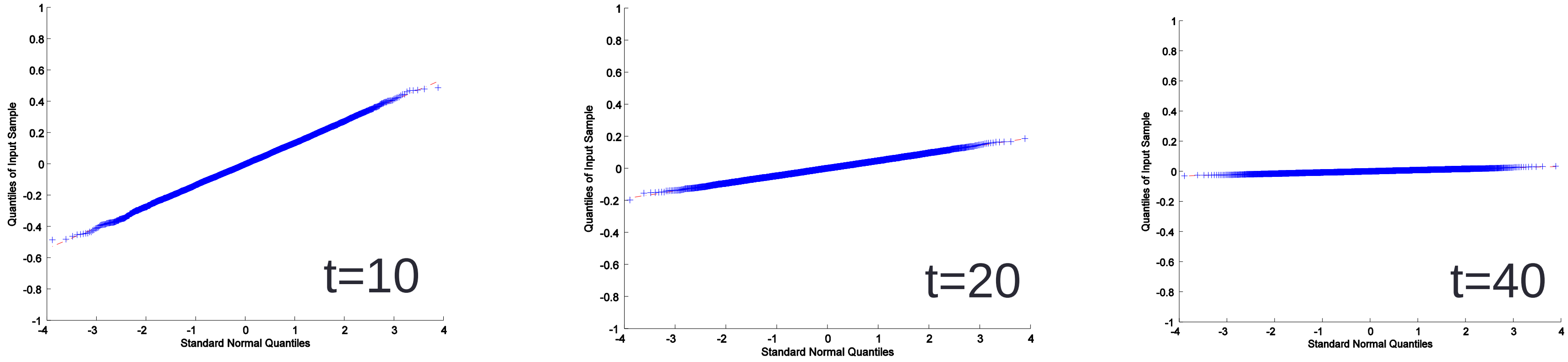
$$\hat{x}^{t+1} = \eta_t(r^t)$$

$$z^{t+1} = y - \Phi\hat{x}^{t+1} + \frac{n}{m} z^t < \eta_t^{'}(r^t) >$$ ⟵ Onsager reaction term

Where $\eta_t(\bullet)$ is the non-linear function applied element-wise to the vector r^t

Iterative Gaussian denoising nature of AMP

Quantile-Quantile Plot for $r^t - x_0$ against Gaussian distribution



$$r^t = x_0 + w\sqrt{c^t} \quad \text{Where} \quad w \sim N(0,1)$$

c^t is the effective noise variance at each AMP iteration

AMP variants:

- L1-AMP: $\eta_t(\bullet)$ being the soft-thresholding function
- Bayesian optimal AMP: $\eta_t(\bullet)$ being the MMSE estimator

Motivation for parametric SURE-AMP

- L1-AMP treats the signal denoising as a 1-d problem while the true signal pdf is visible in the noisy estimate r^t in the large system limit.

- Reconstruction goal: achieve recovery with minimum MSE (BAMP reconstruction) without the prior $p(x_0)$

- Solution:
- Fitting the prior with finite number of Gaussians iteratively

  EM-GAMP algorithm [Vila et al. 2013] – indirect way to minimize MSE
- Optimize the parametric denoiser iteratively

  Parametric SURE-AMP – direct way to minimize MSE

Parametric SURE-AMP algorithm

Initialized with $\hat{x}^0 = 0,\ z^0 = y,\ c^0 = <\| z^0 \|^2>$

For t = 0,1,….

$$r^t = \hat{x}^t + \Phi^T z^t$$

$$\theta^t = H_t(r^t, c^t)$$ ← **parameter selection function**

$$\hat{x}^{t+1} = f_t(r^t, c^t \mid \theta^t)$$ ← **parametric denoiser**

$$\upsilon^{t+1} = \left\langle f_t^{'}(r^t, c^t \mid \theta^t) \right\rangle$$

$$z^{t+1} = y - \Phi \hat{x}^{t+1} + \frac{n}{m} \upsilon^{t+1} z^t$$

$$c^{t+1} = \left\langle \left\| z^{t+1} \right\|^2 \right\rangle$$

SURE: Unbiased estimate of MSE

- Ideally we would like a denoiser with the mimimum MSE. Calculating MSE requires x_0 , thus we need to find a surrogate for MSE
- Let $r = x_0 + w\sqrt{c}$ be the noisy observation of x_0 with $w \sim N(0,1)$

  The denoised signal is obtained via

$$\hat{x} = f(r, c \mid \theta) = r + g(r, c \mid \theta)$$

Theorem [Stein 1981]

SURE is defined as the expected value over the noisy data alone and is the unbiased estimate of the MSE

$$\mathrm{E}_{\hat{x},x_0}\left\{\left(\hat{x} - x_0\right)^2\right\} = \mathrm{E}_{x_0,\gamma}\left\{\left[f(r,c \mid \theta) - x_0\right]^2\right\}$$

$$= c + \mathrm{E}_r\left\{g^2(r,c \mid \theta) + 2cg'(r,c \mid \theta)\right\}$$

Parameter Selection Function

The denoiser parameters are iteratively selected according to

$$\theta^t = H_t(r^t, c^t)$$
$$= \underset{\theta}{argmin} \left\langle g^2\left(r^t, c^t \mid \theta\right) + 2c^t g^{'}(r^t, c^t \mid \theta) \right\rangle$$

- The parameters optimization relies purely on the noisy data and the effective noise variance.

- If all MMSE estimators are included in the parametric family, the parametric SURE-AMP achieves the BAMP performance without prior.

Practical Parametric Denoiser

- The denoiser is parameterized as the weighted sum of kernel functions

$$f(\gamma, c \mid \theta) = r + g(\gamma, c \mid \theta) = \sum_{i=1}^{k} \lambda_i f_i(r \mid \Delta_i(c))$$

- The non-linear parameters of the kernels are tied up with the effective noise variance

$$\Delta_i(c) = \omega_i c$$

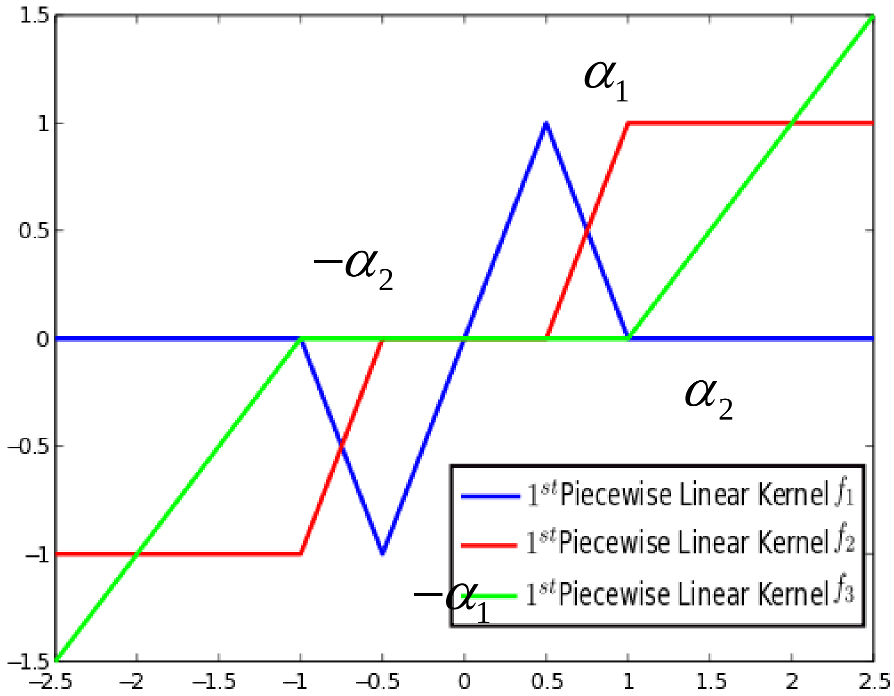
where ω_i is fixed for all iterations.

- The linear weight for the kernels are optimized by solving

$$\varepsilon = c + \left\langle g^2(r, c \mid \theta) + 2cg^{'}(r, c \mid \theta) \right\rangle$$

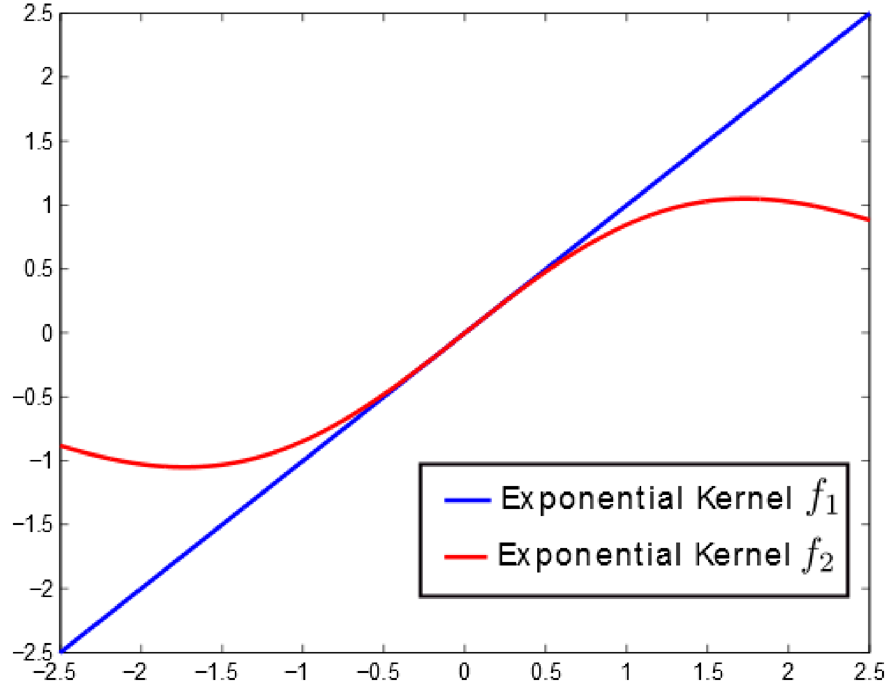
$$\frac{d\varepsilon}{d\lambda_i} = \left\langle 2g(r, c \mid \theta) \frac{d}{d\lambda_i} g(r, c \mid \theta) + c \frac{d}{d\lambda_i} g^{'}(r, c \mid \theta) \right\rangle = 0$$

Kernel Function Examples



Piecewise Linear Kernel [Donoho et al. 2012]

$$\alpha_1 = 2\sqrt{c}, \quad \alpha_2 = 4\sqrt{c}$$

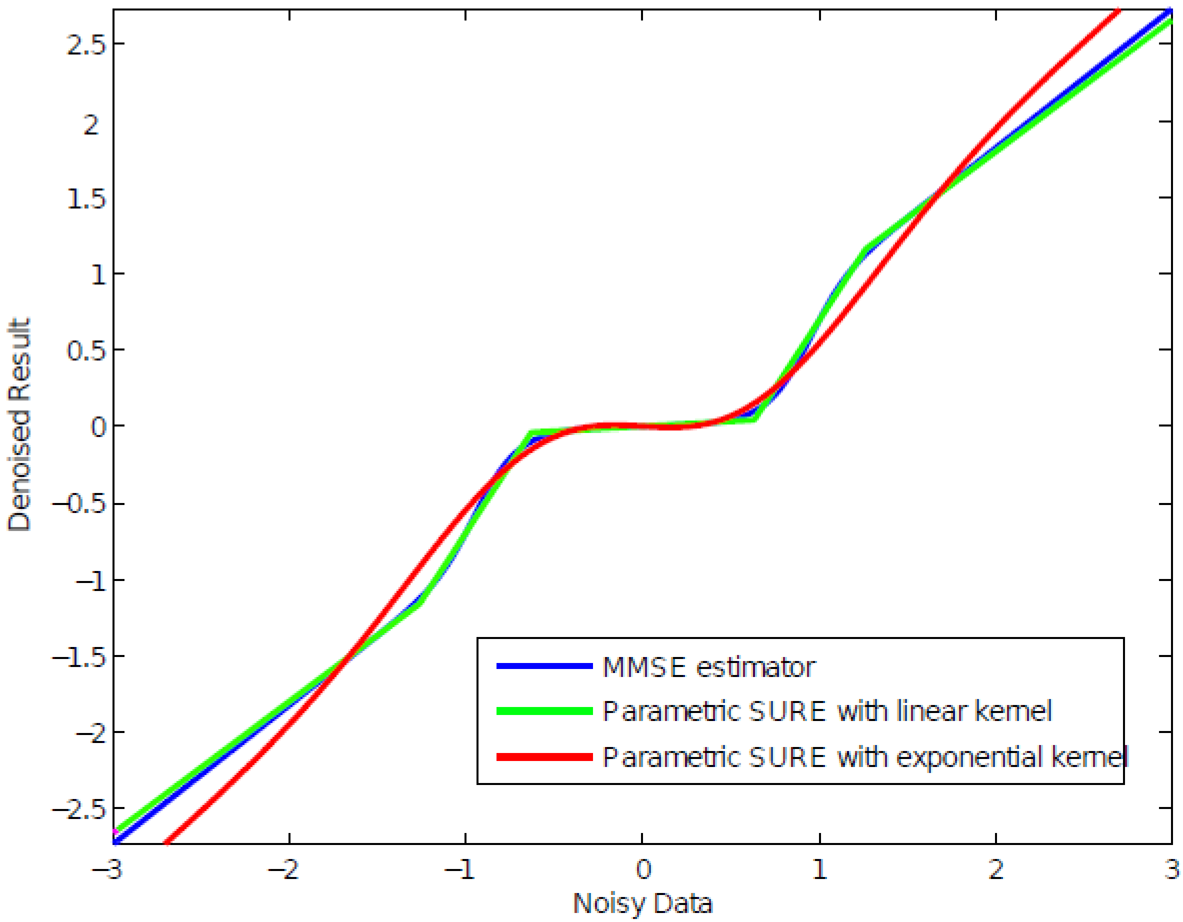


Exponential Kernel [Luisier et al. 2007]

$$f_1(\gamma) = \gamma,\ f_2(\gamma \mid \mathrm{T}) = \gamma e^{-\frac{\gamma^2}{2T^2}}$$

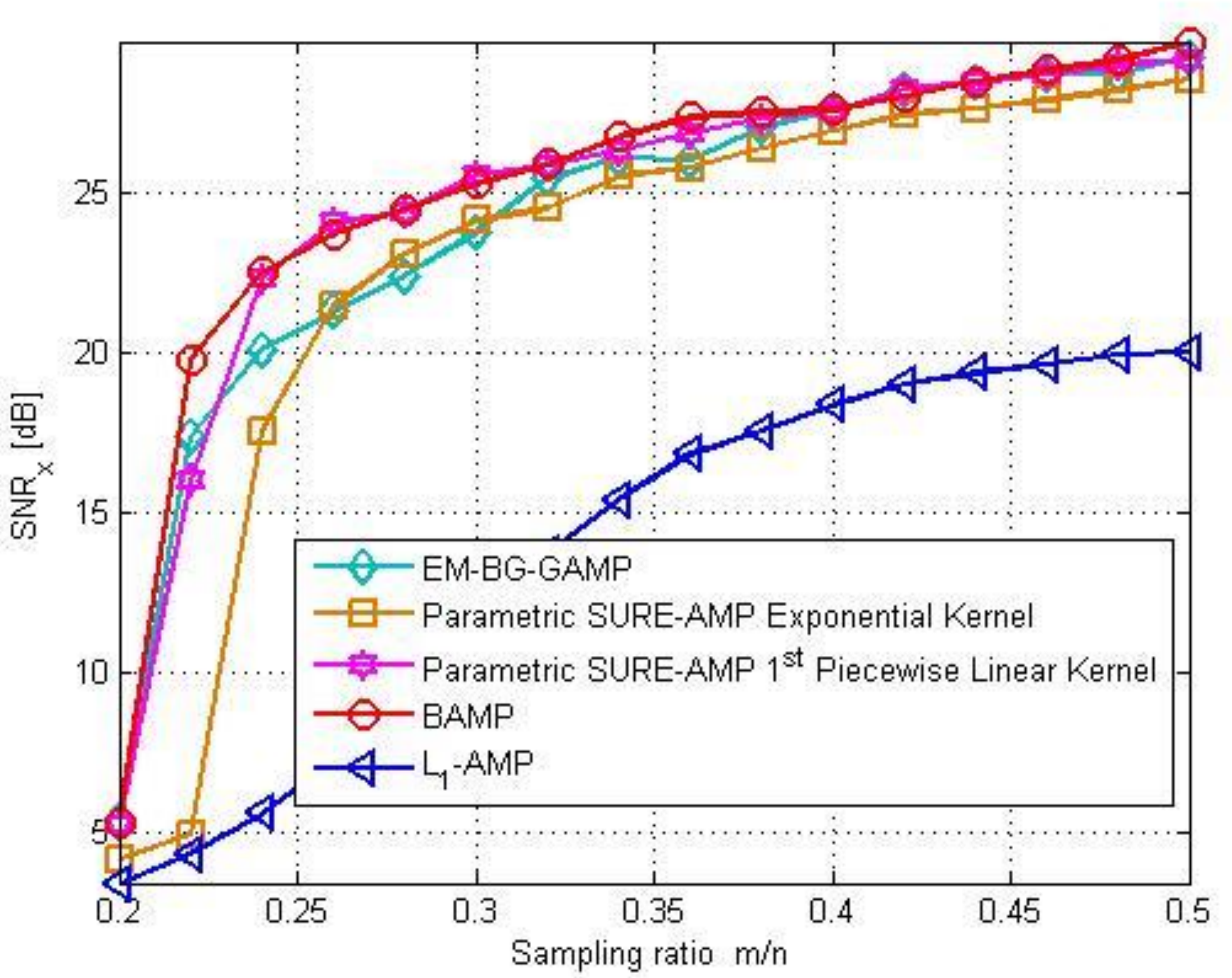
$$T = 6\sqrt{c}$$

MMSE estimator V.S. Kernel Based Denoiser



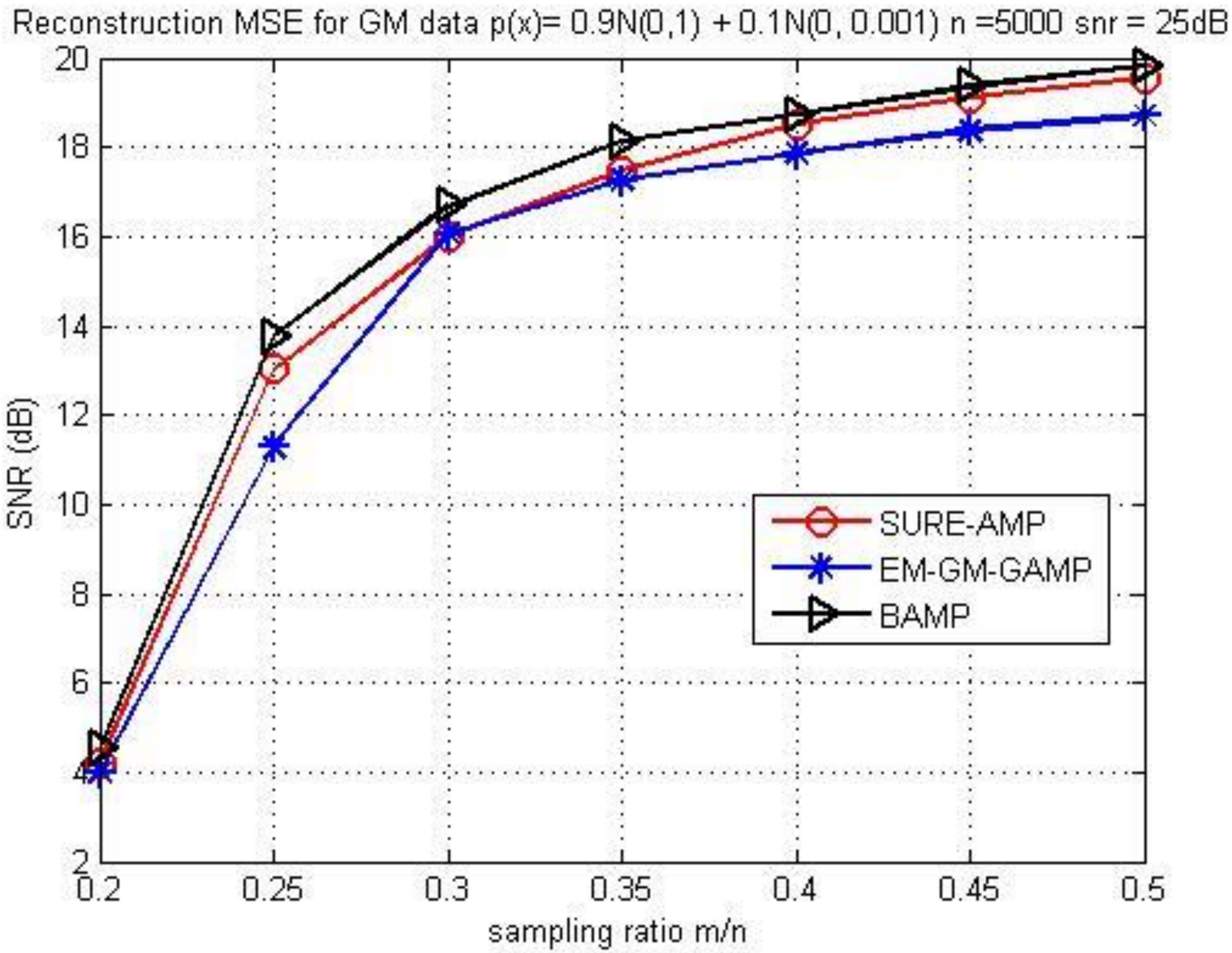
$$p(\mathrm{x}) = 0.1\mathrm{N}(0,1) + 0.9\delta(\mathrm{x})$$

Reconstruction Comparison



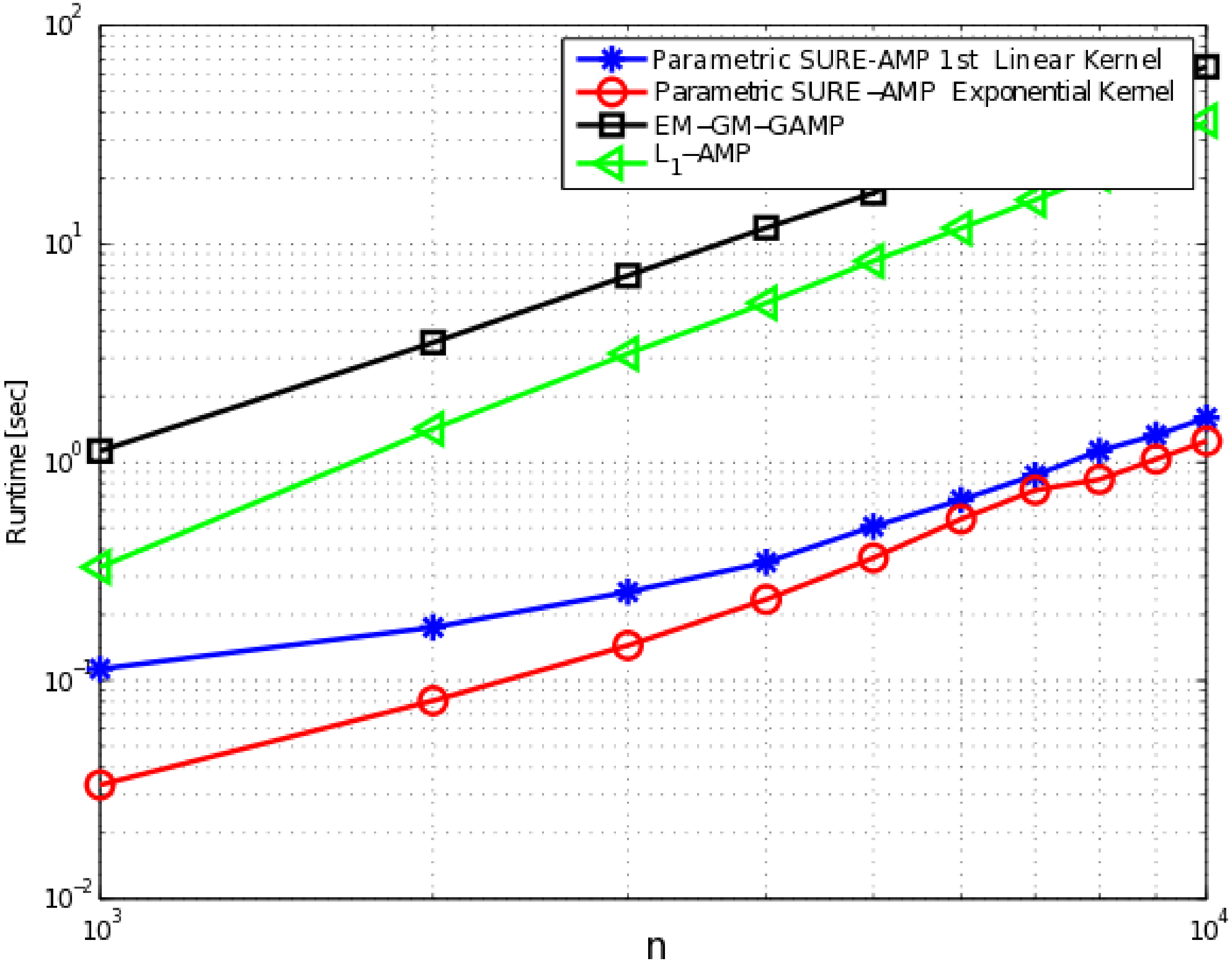
$$p(\mathrm{x}) = 0.1\,\mathrm{N}(0,1) + 0.9\,\delta(\mathrm{x})$$

Reconstruction Comparison



$$p(\mathrm{x}) = 0.1\,\mathrm{N}(0,1) + 0.9\,N(0, 0.01)$$

Runtime Comparison



20 times faster than the EM-GM-GAMP algorithm for Bernoulli-Gaussian

Natural Images Reconstruction

Natural Images Reconstruction

Boat	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
ℓ_1-AMP	13.065	13.55	14.77	14.87	15.90	17.93
EM-GM-GAMP	14.04	14.32	15.90	16.14	16.90	18.32
SURE-AMP	**14.44**	**15.00**	**16.20**	**16.26**	**17.27**	**18.94**
House	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
ℓ_1-AMP	14.33	14.87	15.33	15.53	17.14	17.64
EM-GM-GAMP	15.29	15.78	15.85	**17.09**	18.91	19.56
SURE-AMP	**15.63**	**16.29**	**16.59**	17.02	**18.95**	**19.82**
Lena	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
ℓ_1-AMP	12.04	13.18	14.12	14.62	15.56	16.40
EM-GM-GAMP	13.56	13.92	14.80	15.71	16.80	17.55
SURE-AMP	**13.82**	**13.97**	**15.08**	**16.34**	**17.19**	**19.22**
Cameraman	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
ℓ_1-AMP	12.20	12.65	13.55	14.00	14.56	16.78
EM-GM-GAMP	**13.10**	13.87	14.55	15.63	16.39	**18.55**
SURE-AMP	12.78	**14.12**	**14.69**	**16.33**	**16.65**	18.30
Bridge	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
ℓ_1-AMP	13.62	14.21	14.41	15.68	16.21	17.42
EM-GM-GAMP	14.25	15.15	15.70	16.64	**18.20**	**18.83**
SURE-AMP	**14.29**	**15.31**	**16.13**	**16.81**	18.19	18.78

Conclusion

- The parametric SURE-AMP ***directly minimizes the MSE*** of the reconstructed signal at each iteration.
- With proper design of the parametric family, the parametric SURE-AMP algorithm ***achieves the BAMP performance without the signal prior***.
- The parametric SURE is ***cheap in terms of the computational cost.***
- Further research involves considering more sophisticated kernel families and the rigorous proof for the state evolution dynamics.