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 $\Phi$ [Donoho 09]

- Initialized with $\hat{x}^0 = 0, \ z^0 = y$

  For $t = 0, 1...,$

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

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

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

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 $\hat{x}^0 = 0$, $z^0 = y$, $c^0 = \langle \| z^0 \| ^2 \rangle$

For $t = 0, 1, \ldots$

\[
\begin{align*}
    r^t &= \hat{x}^t + \Phi^T z^t \\
    \theta^t &= H_t (r^t, c^t) \quad \text{parameter selection function} \\
    \hat{x}^{t+1} &= f_t (r^t, c^t | \theta^t) \quad \text{parametric denoiser} \\
    \nu^{t+1} &= \langle f_t' (r^t, c^t | \theta^t) \rangle \\
    z^{t+1} &= y - \Phi \hat{x}^{t+1} + n \frac{n}{m} \nu^{t+1} z^t \\
    c^{t+1} &= \langle \| z^{t+1} \| ^2 \rangle
\end{align*}
\]
SURE: Unbiased estimate of MSE

- Ideally we would like a denoiser with the minimum 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

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

$$= c + E_r \left\{ g^2(r, c \mid \theta) + 2cg'(r, c \mid \theta) \right\}$$
Parameter Selection Function

The denoiser parameters are iteratively selected according to

\[ \theta^t = H_t(r^t, c^t) \]

\[ = \arg\min_{\theta} \left\langle g^2 \left( r^t, c^t \mid \theta \right) + 2c^t g' \left( r^t, c^t \mid \theta \right) \right\rangle \]

- 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 \( \omega_i \) is fixed for all iterations.

• The linear weight for the kernels are optimized by solving

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

\[ f_1(\gamma) = \gamma, \quad f_2(\gamma \mid T) = \gamma e^{-\frac{\gamma^2}{2T^2}} \]

\[ T = 6 \sqrt{c} \]
$p(x) = 0.1N(0,1) + 0.9\delta(x)$

MMSE estimator V.S. Kernel Based Denoiser
Reconstruction Comparison

\[ p(x) = 0.1N(0,1) + 0.9\delta(x) \]
Reconstruction Comparison

Reconstruction MSE for GM data $p(x) = 0.9N(0,1) + 0.1N(0,0.001)$ $n = 5000$ $\text{snr} = 25\text{dB}$

$p(x) = 0.1N(0,1) + 0.9N(0,0.01)$
Runtime Comparison

20 times faster than the EM-GM-GAMP algorithm for Bernoulli-Gaussian
Natural Images Reconstruction
## Natural Images Reconstruction

<table>
<thead>
<tr>
<th></th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boat</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ell_1$-AMP</td>
<td>13.065</td>
<td>13.55</td>
<td>14.77</td>
<td>14.87</td>
<td>15.90</td>
<td>17.93</td>
</tr>
<tr>
<td>SURE-AMP</td>
<td>14.44</td>
<td>15.00</td>
<td>16.20</td>
<td>16.26</td>
<td>17.27</td>
<td>18.94</td>
</tr>
<tr>
<td><strong>House</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ell_1$-AMP</td>
<td>14.33</td>
<td>14.87</td>
<td>15.33</td>
<td>15.53</td>
<td>17.14</td>
<td>17.64</td>
</tr>
<tr>
<td>EM-GM-GAMP</td>
<td>15.29</td>
<td>15.78</td>
<td>15.85</td>
<td>17.09</td>
<td>18.91</td>
<td>19.56</td>
</tr>
<tr>
<td>SURE-AMP</td>
<td>15.63</td>
<td>16.29</td>
<td>16.59</td>
<td>17.02</td>
<td>18.95</td>
<td>19.82</td>
</tr>
<tr>
<td><strong>Lena</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ell_1$-AMP</td>
<td>12.04</td>
<td>13.18</td>
<td>14.12</td>
<td>14.62</td>
<td>15.56</td>
<td>16.40</td>
</tr>
<tr>
<td>SURE-AMP</td>
<td>13.82</td>
<td>13.97</td>
<td>15.08</td>
<td>16.34</td>
<td>17.19</td>
<td>19.22</td>
</tr>
<tr>
<td><strong>Cameraman</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ell_1$-AMP</td>
<td>12.20</td>
<td>12.65</td>
<td>13.55</td>
<td>14.00</td>
<td>14.56</td>
<td>16.78</td>
</tr>
<tr>
<td><strong>Bridge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ell_1$-AMP</td>
<td>13.62</td>
<td>14.21</td>
<td>14.41</td>
<td>15.68</td>
<td>16.21</td>
<td>17.42</td>
</tr>
<tr>
<td>EM-GM-GAMP</td>
<td>14.25</td>
<td>15.15</td>
<td>15.70</td>
<td>16.64</td>
<td>18.20</td>
<td>18.83</td>
</tr>
<tr>
<td>SURE-AMP</td>
<td>14.29</td>
<td>15.31</td>
<td>16.13</td>
<td>16.81</td>
<td>18.19</td>
<td>18.78</td>
</tr>
</tbody>
</table>
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.