Plug-And-Play Learned Gaussian-mixture Approximate Message Passing

Osman Musa, Peter Jung, and Giuseppe Caire

Communications and Information Theory Group, Technische Universität Berlin

Email:{osman.musa,peter.jung,caire}@tu-berlin.de.

Introduction

Compressed Sensing

- In Compressed Sensing (CS) we take M possibly noisy linear measurements $\{y_i\}_{i=1}^{m}$ of an N-dimensional K-sparse (has at most K nonzero components) vector $\mathbf{x}$, according to

$$\mathbf{y} = \mathbf{A}\mathbf{x} \; (+\mathbf{w}),$$

where w is independent and identically distributed (i.i.d.) additive noise.

- Recovery is possible if $\mathbf{A} \in \mathbb{R}^{m\times N}$ satisfies the Restricted Isometry Property (RIP).
- Matrices whose elements are randomly drawn from a (sub-)Gaussian distribution satisfy RIP with high probability.

Why LAMP?

- AMP is very appealing for its efficiency and accurate recovery.
- It approximates the computationally intractable high-dimensional integration involved with calculating

$$\hat{\mathbf{x}} \approx \mathbb{E}\{\mathbf{x} \mid \mathbf{y}\}.$$

- Its parameters can be learned from training data $\Rightarrow$ LAMP.
- LAMP significantly improves upon both LISTA and AMP.
- Is the chosen parametric family of denoisers good for the source prior? - We don't know!

Main Idea

- We model the source signal prior as an independent and identically distributed (i.i.d.) Gaussian-mixture (GM) distribution.
- We adopt the optimal denoiser function that would minimize MSE had the assumed GM prior match the true unknown prior.
- The parameters of the GM are learned from training data.
- We, therefore, examine a general-purpose denoiser within the LAMP algorithm.

Why GM?

- The resulting denoiser function $\eta(\cdot)$, and it's derivative $\frac{d}{dr}\eta(\cdot)$ can be calculated analytically.
- If the overall objective is to minimize MSE, a good approximation of a discrete component in the source prior is a Gaussian distribution with matching mean and a very small variance.
- A Gaussian mixture can model a variety of continuous distributions.

Learned Gaussian-mixture AMP

The Steps of LAMP

- The LAMP algorithm is initialized (at $t = 0$) according to

$$\hat{\mathbf{x}}^0 = 0_{N\times 1}, \quad \mathbf{z}^{(0)} = 0_{m\times 1}, \quad \mathbf{B}.$$

- At every iteration $t = 1, 2, \ldots, T_{\max}$, the algorithm computes

$$\mathbf{r}^{(t)} = \mathbf{x}^{(t-1)} + \mathbf{B}\mathbf{z}^{(t)},$$
$$\mathbf{x}^{(t+1)} = \eta(\mathbf{r}^{(t)};\, \sigma^{2(t)}, \Theta^{(t)}),$$
$$b^{(t)} = \frac{1}{m}\sum_{i=1}^{N} \frac{\partial[\eta(\mathbf{r}^{(t)}, \sigma^{2(t)}, \Theta^{(t)})]_i}{\partial r_i},$$
$$\mathbf{z}^{(t)} = \mathbf{y} - \mathbf{A}\mathbf{x}^{(t-1)} + b^{(t-1)}\mathbf{z}^{(t-1)}.$$

- $\mathbf{B}$ is the learned weight (filter) matrix.
- $\sigma^{2(t)}$ is the effective noise variance at the t-th layer, which can be estimated with $\sigma^{(t)} = \|\mathbf{z}^{(t)}\|_2/\sqrt{m}$.
- The denoiser function $\eta(\cdot,\cdot,\cdot)$ that minimizes MSE is given by

$$\eta(r_n^{(t)};\, \sigma^{(t)}, \Theta^{(t)}) = \mathbb{E}[x_n \mid r_n^{(t)} = r_n^{(t)};\, \sigma^{(t)}, \Theta^{(t)}].$$

- The Onsager term allows for the decoupled measurement model, i.e., $\mathbf{r}^{(t)} \sim \mathbf{x} + \mathbf{v}^{(t)}$, where $\mathbf{v}^{(t)} \sim \mathcal{N}(\mathbf{0}_N, \sigma^{2(t)}\mathbf{I}_N)$.

GM prior distribution

- We assume GM prior distribution:

$$p(x_n;\, \Theta_{\mathrm{GM}}) = \sum_{l=1}^{L} \omega_l \mathcal{N}(x_n; \mu_l, \sigma_l^2),$$

where $\sum_{l=1}^{L} \omega_l = 1$, $0 \le \omega_l \le 1$, $\forall l \in [L]$.

- $\Rightarrow$ the conditional pdf of x_n given $r_n^{(t)}$ can be written as

$$p(x_n \mid r_n;\, \sigma^2, \Theta_{\mathrm{GM}}) = \sum_{l=1}^{L} \bar{\beta}_{n,l}\mathcal{N}(x_n; \gamma_{n,l}, \nu_{n,l}).$$

Learning the L-GM-AMP Parameters

- A network with $T_{\max}$ layers which has $Nm + 3LT_{\max}$ tunable parameters $\mathbf{B} \cup_{l=1}^{L} \{\omega_l, \mu_l, \sigma_l^2\}$.

Algorithm 1: Tied LAMP-GMP parameter learning

$\mathbf{B} = \mathbf{A}^T, \Theta_{\mathrm{GM}}^{(0)} = \cup_{l=1}^{L}\{\omega_{l,0}, \mu_{l,0}, \sigma_{l,0}^2\}$;
while $t \le T_{max}$ **do**
  Initialize $\Theta_{\mathrm{GM}}^{(t)} = \Theta_{\mathrm{GM}}^{(t-1)}$;
  Learn $\Theta_{\mathrm{GM}}^{(t)}$ with fixed $\Theta_{\mathrm{tied}}^{(t-1)}$;
  Refine $\Theta_{\mathrm{tied}}^{(t)} = \{\mathbf{B}, \cup_{t'=1}^{t}\Theta_{\mathrm{GM}}^{(t')}\}$;
end

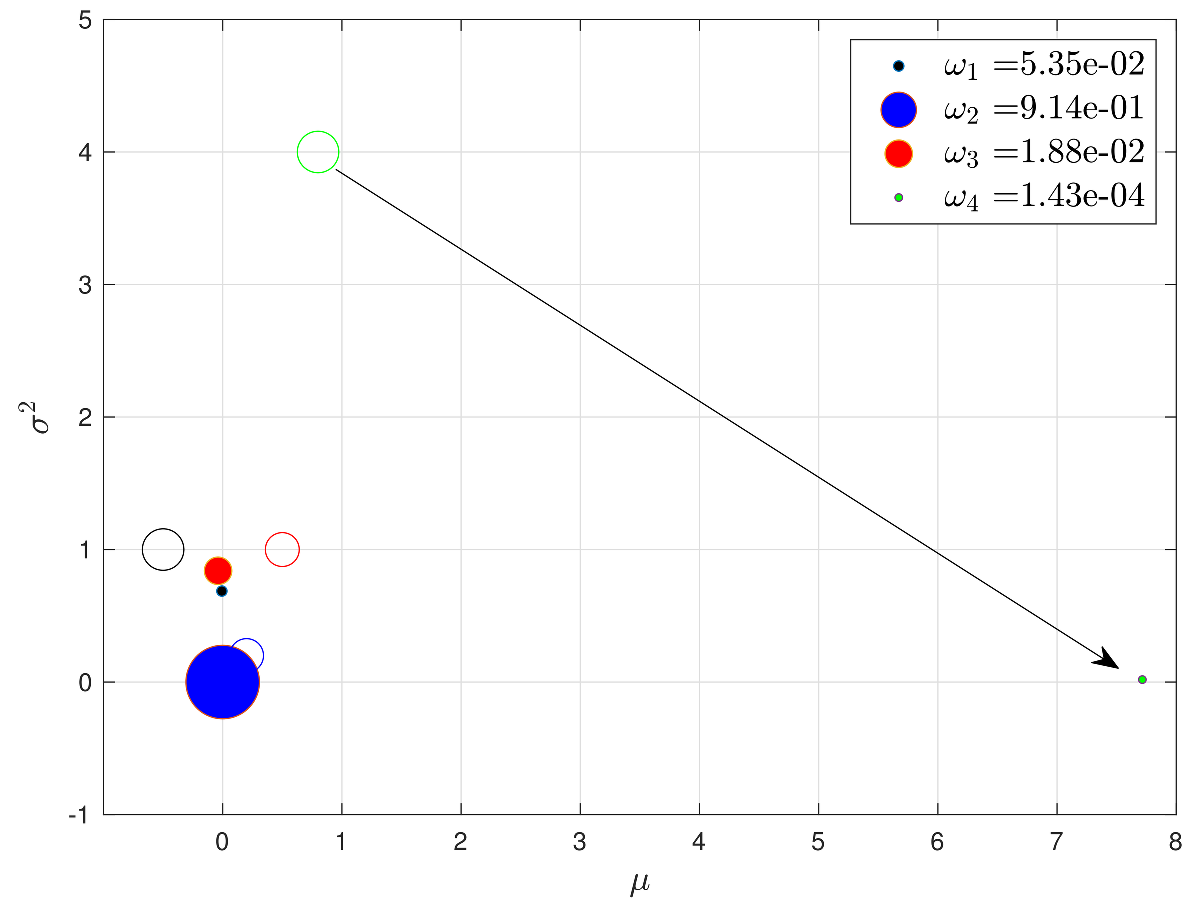
Numerical Results

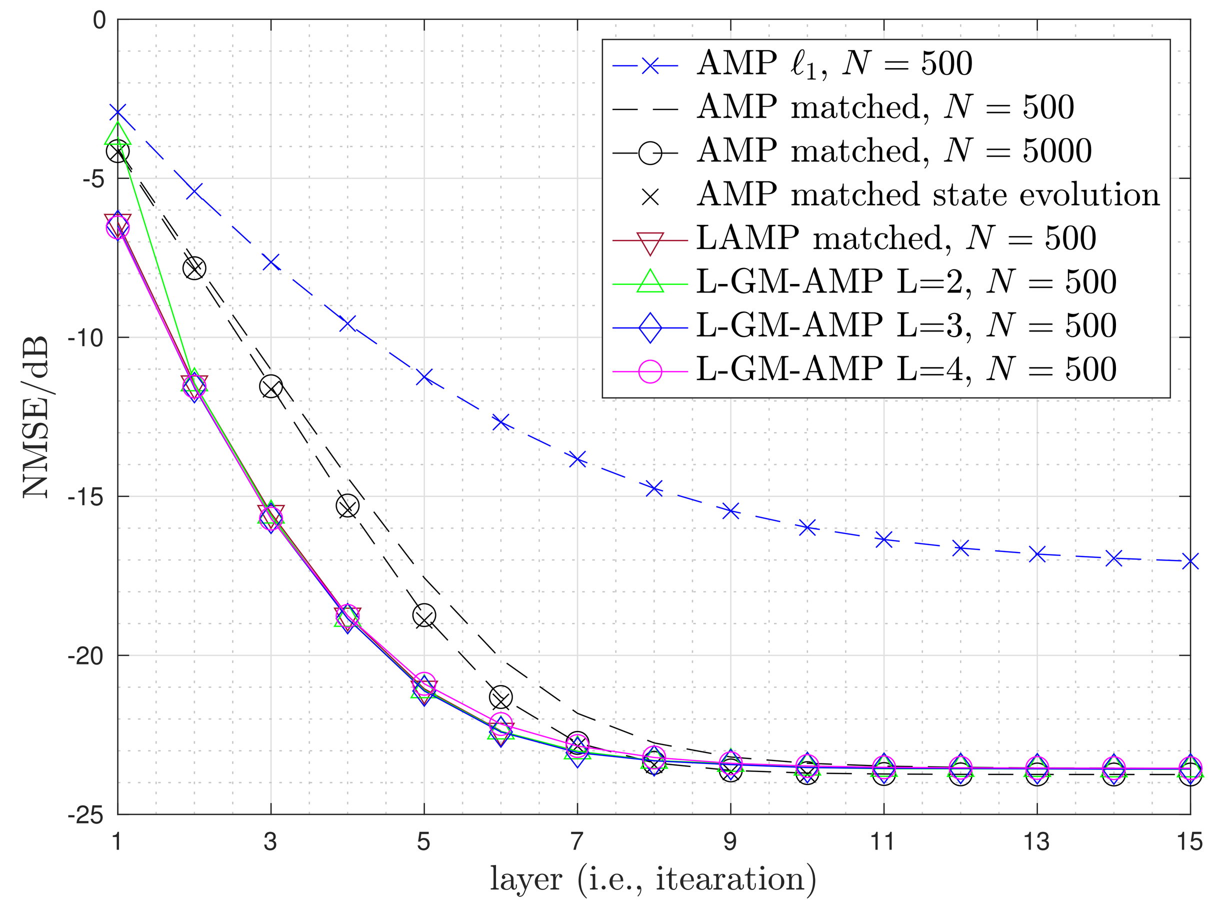
Simulation Setup

- The entries of the sensing matrix A are drawn once independently from a zero-mean Gaussian distribution with variance 1/m, and kept fixed.
- The per-iteration NMSE is used as the performance metric

$$\mathrm{NMSE}^{(t)} = \mathbb{E}\big[\|\hat{\mathbf{x}}^{(t)} - \mathbf{x}\|_2^2 \,\big|\, \mathbf{A}\big] / \mathbb{E}\big[\|\mathbf{x}\|_2^2\big].$$

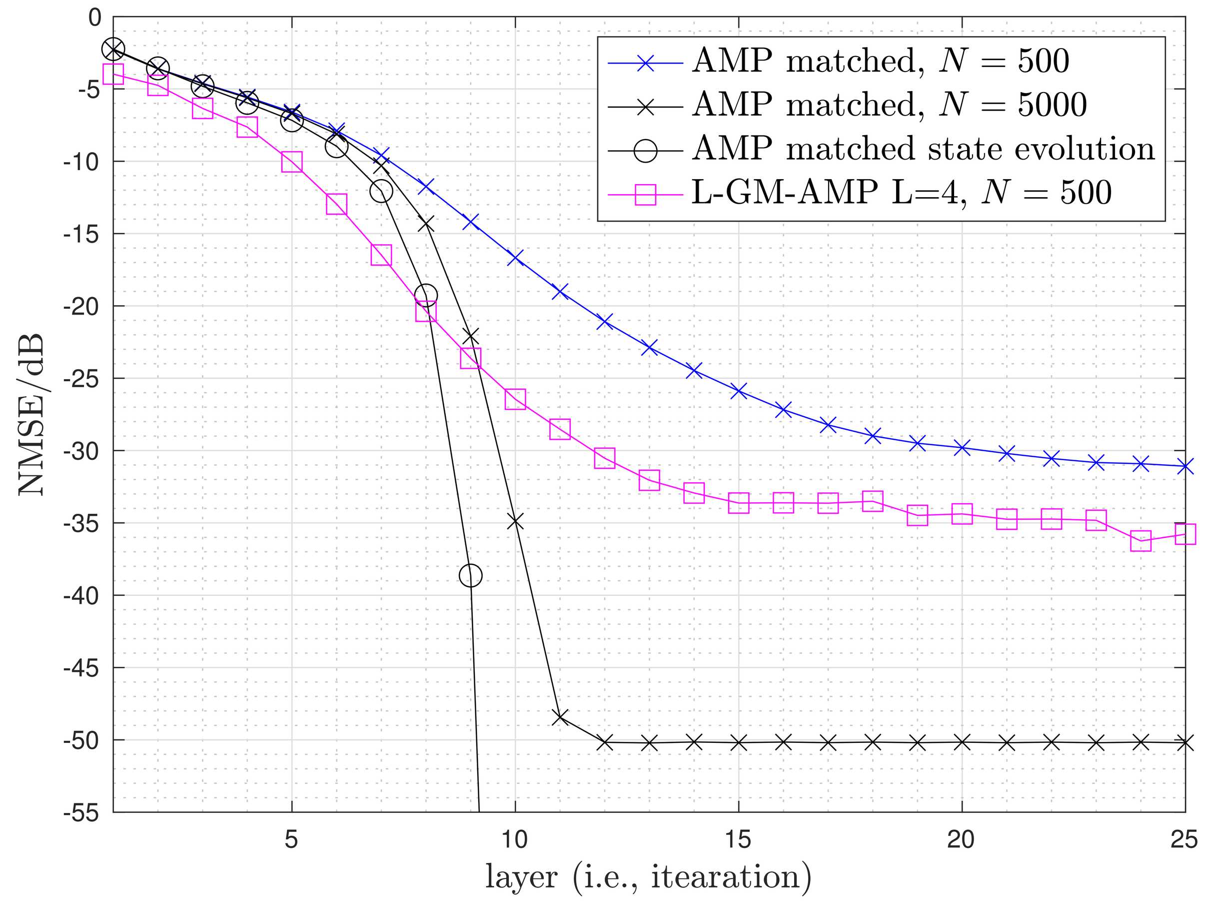
- Given δ and ϵ, we take $m = \delta N$ measurements of a 500-long ϵN-sparse source vector.
- Noise power is calculated as $\mathrm{SNR} = \mathbb{E}\big[\|\mathbf{y}\|^2\big]/\mathbb{E}\big[\|\mathbf{w}\|^2\big] = \frac{\epsilon\,\sigma_x^2}{\delta\,\sigma_w^2}$.
- The network is trained using Adam optimizer.

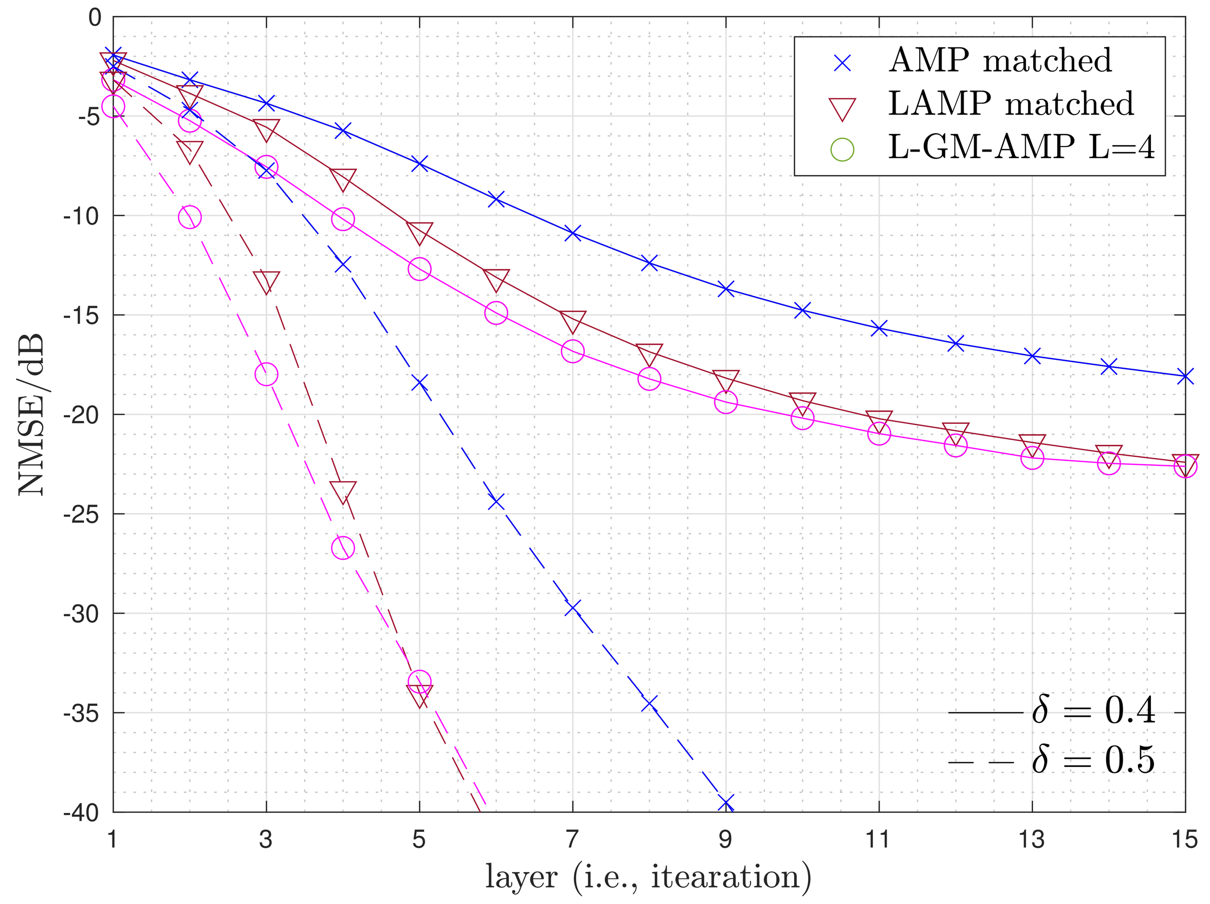




Results

- L-GM-AMP matches the performance of matched LAMP.
- L-GM-AMP does not suffer from over-parametrization.
- L-GM-AMP is also capable of learning discrete priors.





Conclusion

- Although reminiscent of Borgerdings LAMP [1], it differs in the adoption of a universal plug and play denoising function.
- L-GM-AMP algorithm achieves state-of-the-art performance offered by (L)AMP with perfect knowledge of the source prior.

[1] M. Borgerding, P. Schniter, and S. Rangan, AMP-inspired deep networks for sparse linear inverse problems, IEEE Transactions on Signal Processing, vol. 65, no. 16, pp. 42934308, Aug 2017.