An Empirical-Bayes Approach to Compressed Sensing via Approximate Message Passing

Jeremy Vila and Philip Schniter
Department of Electrical and Computer Engineering, The Ohio State University

Abstract

Traditional approaches to compressive sensing (e.g., Lasso, basis-pursuit denoising) take a worst-case approach to signal recovery, and are able to give guarantees for the recovery of any deterministic signal. Furthermore, recent work has connected these traditional approaches to minimax optimal estimation. Although the minimax property is admirable from a robustness perspective, estimation performance can be improved when the signal prior is known to differ from the least-favorable one. At the same time, care must be taken not to assume too much about the signal prior.

In this work, we propose an empirical Bayesian approach to compressive sensing, where we assume an i.i.d Gaussian-Mixture signal model and an i.i.d Gaussian noise model, but make no assumptions on the parameters of the signal distribution and the noise distribution. We then employ the EM algorithm to jointly learn the distributional parameters (in an ML sense) and recover the signal (in an MMSE sense), leveraging the recently proposed approximate message passing (AMP) algorithm for compressive sensing.

Compressive Sensing Problem Statement

Consider estimating a sparse signal $\mathbf{x} \in \mathbb{R}^N$ from undersampled measurements $\mathbf{y} \in \mathbb{R}^M$ in the presence of AWGN. The system is modeled as $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$.

- The signal $\mathbf{x}$ is sparse with otherwise unknown distribution.
- The mixing matrix $\mathbf{A}$ is known and elements are i.i.d with zero-mean and variance $1/M$.
- The dimension of the measurements is less than the signal (i.e. M<N).

Signal and Noise Models

Assume an i.i.d Gaussian-Mixture model for the signal and a Gaussian Model for the noise.

$$p(x_j) = \sum_{\ell=1}^{L} \lambda_\ell \mathcal{N}(x_j; \theta_\ell, \phi_\ell) \qquad p(w_k) = \mathcal{N}(w_k; 0, \psi)$$

where $\lambda_\ell, \theta_\ell, \phi_\ell, L, \psi$ are the **unknown** weighting components, means, signal variances, number of components, and noise variances respectively.

AMP Algorithm Highlights

- Exact belief propagation (BP) techniques infer the marginals $p(x_j|\mathbf{y})$. As the dimension of the signal grows, the number of calculations grow exponentially. This potentially high complexity motivates us to find a simpler scheme.
- The AMP algorithm [Donoho, Maleki, Montanari, '09] approximates these BP techniques.
  - By knowing the prior $p(\mathbf{x})$ and measurements $\mathbf{y}$, AMP provides estimates of $\mathbb{E}_{\mathbf{x}}[\mathbf{x}|\mathbf{y}]$ and var$[\mathbf{x}|\mathbf{y}]$.
  - Asymptotically, AMP follows a state evolution whose fixed points can be rigorously analyzed.
  - The AMP algorithm exhibits $\mathcal{O}(MN)$ complexity. When the $\mathbf{A}$ can be implemented as a fast transform (e.g., FFT), AMP requires $\mathcal{O}(N \log N)$ complexity.
- The GAMP algorithm [Rangan '10] generalizes the AMP algorithm for arbitrary signal and noise priors. It also provides estimates for individual component variances.

Estimation of $\lambda_\ell, \theta_\ell$, and ϕ_ℓ

The EM algorithm is an iterative process that computes the maximum likelihood (ML) estimate of the parameters $\lambda_\ell, \theta_\ell$, and ϕ_ℓ for which the observed data is most likely.

$$\lambda_{\ell_{\text{ML}}} = \arg\max_{\lambda_\ell} p(\mathbf{y}; \lambda_\ell) \qquad \theta_{\ell_{\text{ML}}} = \arg\max_{\theta_\ell} p(\mathbf{y}; \theta_\ell) \qquad \phi_{\ell_{\text{ML}}} = \arg\max_{\phi_\ell} p(\mathbf{y}; \phi_\ell)$$

Calculations show that the parameter estimates are dependent on a conditional expectation (hence expectation maximization).

$$\hat{\lambda}_\ell^{i+1} = \arg\max_{\lambda_\ell} \sum_{j=1}^{n} \mathbb{E}_{x_j}\left[\ln p(x_j; \lambda_\ell, \hat{\lambda}^i_{-\ell}, \hat{\theta}^i, \hat{\phi}^i)|\mathbf{y}, \hat{\lambda}^i\right]$$

$$\hat{\theta}_\ell^{i+1} = \arg\max_{\theta_\ell} \sum_{j=1}^{n} \mathbb{E}_{x_j}\left[\ln p(x_j; \theta_\ell, \hat{\theta}^i_{-\ell}, \hat{\lambda}^i, \hat{\phi}^i)|\mathbf{y}, \hat{\theta}^i\right]$$

$$\hat{\phi}_\ell^{i+1} = \arg\max_{\phi_\ell} \sum_{j=1}^{n} \mathbb{E}_{x_j}\left[\ln p(x_j; \phi_\ell, \hat{\phi}^i_{-\ell}, \hat{\lambda}^i, \hat{\theta}^i)|\mathbf{y}, \hat{\phi}^i\right]$$

- The ML estimates are guaranteed to converge to a *local* maxima as EM iterates.
- Current estimates of the parameters are based upon their previous estimates.
- The *closed-form* solution to the parameters are a function of $\hat{x}_j$, μ_j^x, the mean and variance of the element x_j provided by the GAMP algorithm.

Estimating Noise Variance ψ

The GAMP algorithm returns the estimates $(\hat{x}_j)$ for the signal $\mathbf{x}$ based upon the measurements and the signal prior. As a result, it also returns the estimate $(\hat{\mathbf{w}})$ for $\mathbf{w}$.

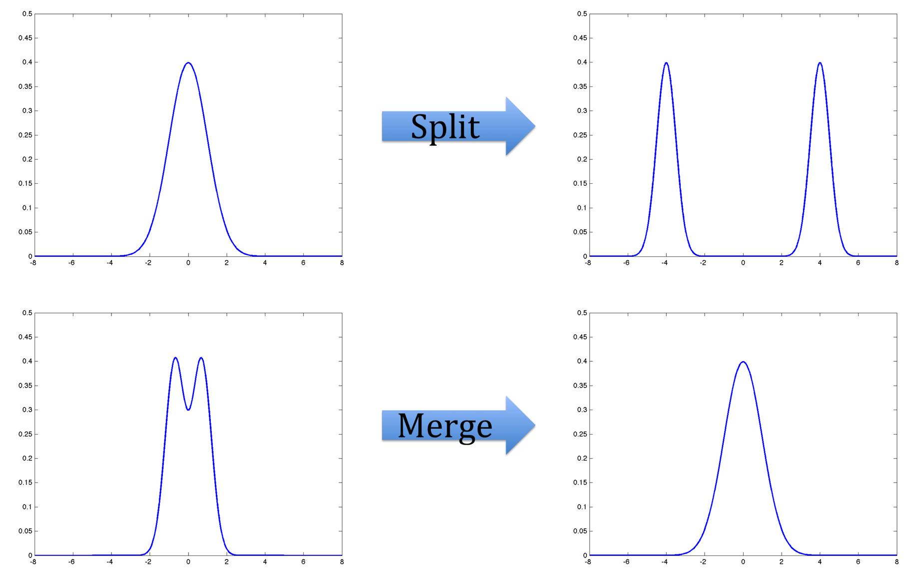
$$\hat{\mathbf{w}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$$

Then the estimate of the parameter ψ is simply the sample variance of $\hat{\mathbf{w}}$.

Estimating L

Two common problems with using standard EM for Gaussian Mixtures are that the number L components may be apriori unknown, and that it can converge to local maxima. The proposed SMEM algorithm [Ueda '00] proposes a solution to the problem:

- A component that poorly describes the data may need to be split into two components.
- Two similar components may need to be merged into one.



Greedy Expectation-Maximization AMP (GEM-AMP) Algorithm

The GEM-AMP algorithm attempts to approximate the true sparse signal prior $p(x_j)$ with a Gaussian Mixture. We use the EM algorithm to compute the ML estimates of those parameters and use them to form an estimated signal prior.

The GEM-AMP process is described by:

1. Initialize λ^0, θ^0, ϕ^0, and ψ^0
2. At iteration i, AMP calculates the mean and variance vectors of $\mathbf{x}$, $(\hat{x}^i_j$ and $\mu^{x^i}_j)$.
3. EM then returns $\lambda^{i+1}, \theta^{i+1}, \phi^{i+1}$, and ψ^{i+1}.
4. Reinitialize GEM-AMP with new parameters and repeat 2 and 3 until λ converges or a maximum number of iterations. Compute the log-likelihood Q.
5. Decide which component is best to split and which two components are best to merge. Perform EM for both cases. Then compute Q_{split} and Q_{merge}.
6. Find $\hat{Q} = \max\{Q, Q_{\text{split}}, Q_{\text{merge}}\}$. If $\hat{Q} = Q$ then stop. Otherwise, reinitialize with the parameters of $\hat{Q}$ and repeat 1-6 until 6 is satisfied or a maximum number of iterations.

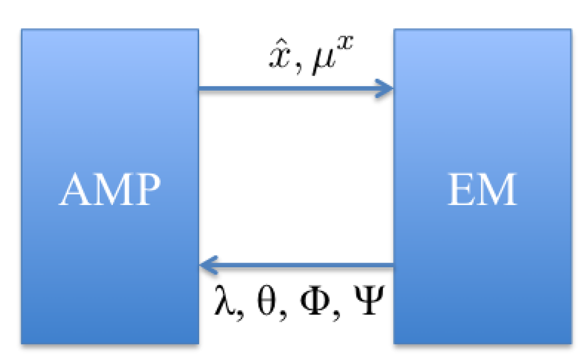


Figure: EM-AMP Process

The goal of 5 and 6 is to avoid local maxima and obtain the optimal L. Essentially choose either to merge or split depending on whichever maximizes the log-likelihood. If either decrease the log-likelihood then we assume the previous Gaussian-Mixture is the optimal one.

EM-AMP Initialization

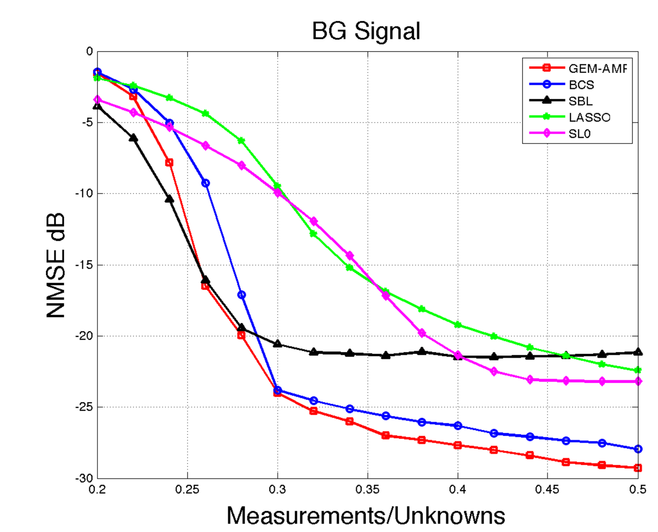
We initialize the Gaussian-Mixture with $L = 2$ terms. One term is modeled as Bernoulli as to capture the zero valued elements. The other Is modeled as Gaussian to approximate the active elements.

- Initialize $\psi \ll \min \phi$ since estimation in low SNR environments can be hard.
- Assume symmetry of the true prior so that $\theta = \{0, 0\}$.
- Use measurement/unknown (ρ) ratio to determine λ. $\lambda = \left\{\frac{2M}{5N}, 1 - \frac{2M}{5N}\right\}$ worked best for simulations.
- Use measurement energy to initialize the active component of ϕ. Set $\phi_{\text{inactive}} = 0.001$

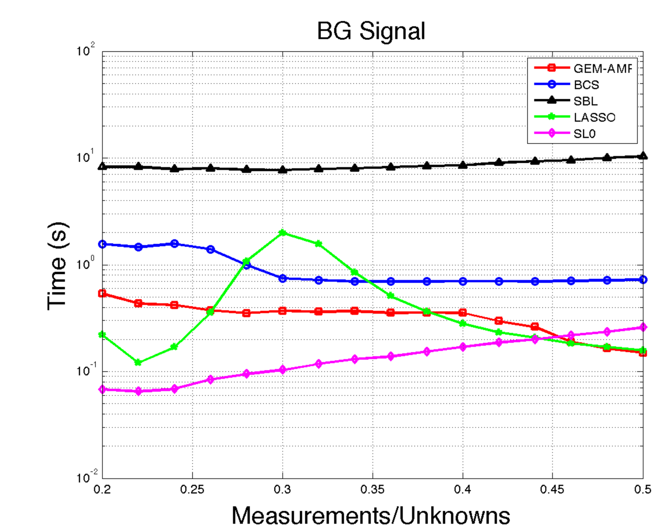
$$||\mathbf{y}||^2 = \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = ||\mathbf{x}_{\text{non-zero}}||^2 = \lambda \phi_{\text{active}} N$$

- This yields $\phi = \left\{\frac{5||\mathbf{y}||^2}{2M}, 0.001\right\}$.

Algorithm Comparisons

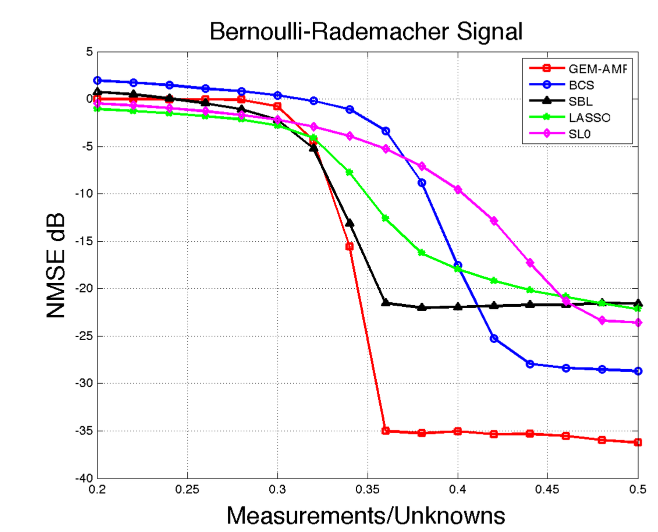


(a) NMSE of Signal Estimates (dB) vs M/N

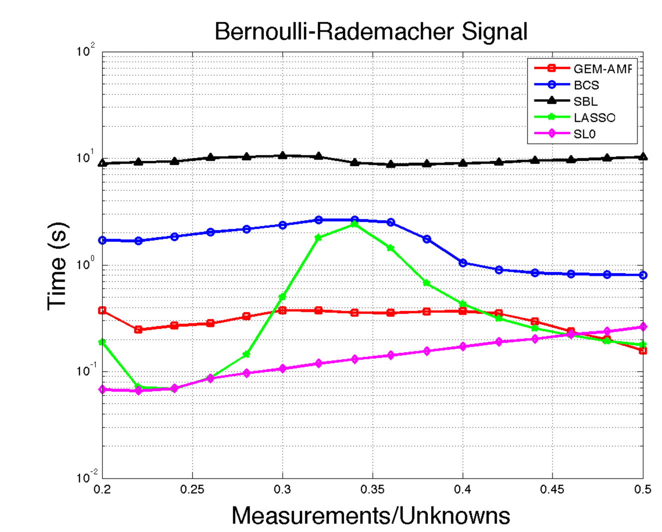


(b) Total computation time vs. M/N

Figure: Bernoulli-Gaussian prior with $N = 1000$, 100 active coefficients, and SNR =25 dB.



(a) NMSE of Signal Estimates (dB) vs M/N



(b) Total computation time vs. M/N

Figure: Bernoulli-Rademacher prior (± 1) with $N = 1000$, 100 active coefficients, and SNR =25 dB.

Noisy Phase Transition

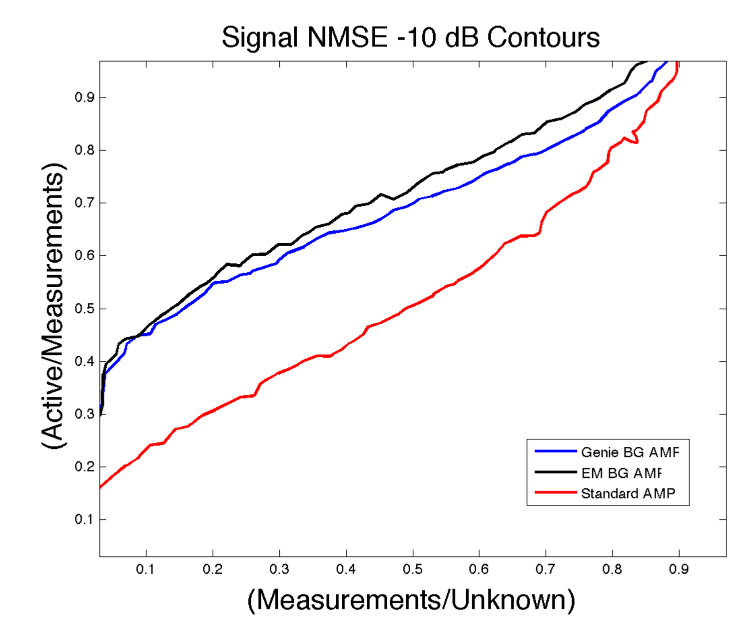


Figure: Phase Transition Contours for $\phi = 1$, $\psi = 10^{-4}$

- A sparse signal was generated as Bernoulli-Gaussian.
- -10 dB contours of phase transition curves are shown for three algorithms.
  1. BG-EM-AMP algorithm.
  2. AMP assuming Bernoulli Gaussian signal prior with known parameters.
  3. Standard AMP assuming Laplacian signal prior.
- Areas beneath curves represent scenarios for accurate signal estimation.

Related Works

- Sparse Bayesian Learning [Wipf, Rao '04]
- Bayesian CS [Ji, Dunson, Carin, '07]
- Bayesian CS using Laplace priors [Babacan, Molina, Katsaggelos '10]
- Solving Inverse Problems with PLE: From GMM to Structured Sparsity [Yu, Sapiro, Mallat '10]

Summary

- Presented a model with unknown parameters for describing sparse signals.
- Developed a GEM-AMP algorithm that simultaneously models and estimates a sparse signal.
  - Achieved linear complexity by only calculating $\mathcal{O}(MN)$ multiplications.
- Found GEM-AMP performed the best in a NMSE sense and comparable in computational complexity.
- Evaluated sparsity-undersampling conditions for accurate recovery of the signal.

Sponsors

This work has been supported in part by NSF-I/UCRC grant IIP-0968910, by NSF grant CCF-1018368, and by DARPA/ONR grant N66001-10-1-4090.

vilaj@ece.osu.edu