Expectation-Maximization Bernoulli-Gaussian Approximate Message Passing

Asilomar 2011

Jeremy Vila and Philip Schniter

The Ohio State University
Department of Electrical and Computer Engineering
vilaj@ece.osu.edu and schniter@ece.osu.edu

November 8th, 2011

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.

CS Problem Statement

- Recover a signal from undersampled measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \quad \mathbf{x} \in \mathbb{R}^N \quad \mathbf{y}, \mathbf{w} \in \mathbb{R}^M \quad M < N$$

where $\mathbf{x}$ is K-sparse (or compressible) with $K < M$.

- With sufficient sparsity and appropriate conditions on the mixing matrix $\mathbf{A}$ (e.g. RIP, nullspace), signal recovery is possible.

- Common approach (LASSO) is to solve

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_1.$$

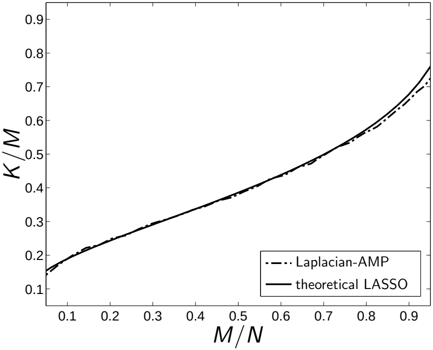
where α must be tuned in accordance with sparsity and SNR.

LASSO Phase Transition

- Region beneath the curve shows (M, N, K) combinations where LASSO can perfectly recover a noiseless signal.

- If the true pdf of $\mathbf{x}$ is i.i.d. $p(x_n) = \lambda f(x_n) + (1-\lambda)\delta(x_n)$, and $\lambda \triangleq \frac{K}{N}$, then the LASSO PTC is unaffected by $f(\cdot)$.

- This implies LASSO is robust to signal distribution, but it cannot benefit when $\mathbf{x}$ belongs to an "easier" class.



Empirical noiseless Bernoulli-Gaussian PTCs

Bayesian Interpretation

- The sparse signal recovery problem can be interpreted through a Bayesian framework.
- Minimizing the LASSO criterion $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1$ is equivalent to finding the MAP estimate from $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ when $\mathbf{w}$ is i.i.d. Gaussian and $\mathbf{x}$ is i.i.d. Laplacian.
- Alternative Bayesian approaches to the CS problem follow from different assumptions on the signal and noise priors, and/or from seeking the MMSE rather than MAP estimate of $\mathbf{x}$.
- MAP estimation using assumed i.i.d. signal/noise priors has the form

$$\max_{\mathbf{x}} \sum_{m=1}^{M} \ln p(y_m|\mathbf{a_m}^{\mathsf{T}}\mathbf{x}) + \sum_{n=1}^{N} \ln p(x_n).$$

Approximate Message Passing (AMP)

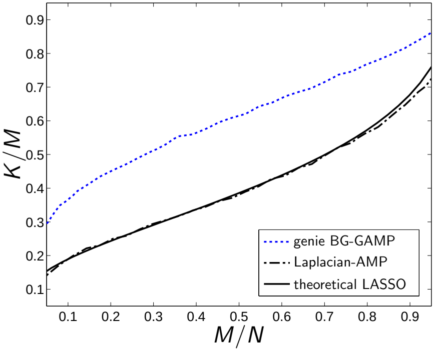
- Efficient algorithms for Bayesian CS can be constructed using loopy belief propagation using carefully constructed message approximations:
  - The "original" AMP [Donoho, Maleki, Montanari '09] solves the LASSO problem (i.e., Laplacian MAP) under i.i.d. matrices $\mathbf{A}$.
  - The "Bayesian" AMP [Donoho, Maleki, Montanari '10] framework tackles MMSE inference under generic signal priors.
  - The "generalized AMP" [Rangan '10] framework tackles MAP or MMSE inference under generic signal and noise priors and generic matrices $\mathbf{A}$.

- All of these AMP algs are sophisticated iterative thresholding algs, thus complexity is dominated by two applications of $\mathbf{A}$ per iteration and ≈ 15 iterations (for any M and N).

Bernoulli-Gaussian GAMP

- Suppose the signal is known to be i.i.d Bernoulli Gaussian. That is, $p(x_n) = \lambda\mathcal{N}(x_n; \theta, \phi) + (1 - \lambda)\delta(x_n)$, where a genie supplies us with the true parameters (λ, θ, ϕ)

- For such signals, the PT improves:



Empirical noiseless Bernoulli-Gaussian PTCs

Expectation-Maximization BG-GAMP (EM-BG-GAMP)

- In practice, the pdf parameter values $\mathbf{q} = (\lambda, \theta, \phi, \psi)$ are unknown. Thus, we propose to learn them via the EM algorithm while simultaneously recovering $\mathbf{x}$.

- In our EM algorithm, we treat both $\mathbf{x}$ and $\mathbf{w}$ as missing data, and perform element-wise incremental updates.

- The update of λ equates to solving the E and M steps

$$\text{(E-step)} \quad Q(\lambda|\lambda^i) = \sum_{n=1}^{N} \mathrm{E}\left\{ \ln p(x_n; \lambda, \theta^i, \phi^i) \middle| \mathbf{y}; \mathbf{q}^i \right\}$$

$$\text{(M-step)} \quad \lambda^{i+1} = \arg\max_{\lambda \in (0,1)} Q(\lambda|\lambda^i).$$

  Updates of (θ, ϕ, ψ) have a similar form.

- All quantities required to compute the EM conditional expectation are provided by GAMP!

Parameter Initialization

Smart initialization is critical since the EM algorithm can converge to local maxima of the likelihood function.

- Set the sparsity $\lambda^0 = \frac{M}{N}\rho_{\mathsf{SE}}(\frac{M}{N})$, where $\rho_{\mathsf{SE}}(\frac{M}{N})$ is the theoretical LASSO PTC.
- Assume signal prior is symmetric and initialize the active mean $\theta^0 = 0$.
- Given a hypothesis SNR^0 we find that the active variance ϕ and noise variance ψ can be initialized based on the energy of the measurements $\|\mathbf{y}\|_2^2$.

$$\psi^0 = \frac{\|\mathbf{y}\|_2^2}{(\mathsf{SNR}^0 + 1)M}, \quad \phi^0 = \frac{\|\mathbf{y}\|_2^2 - M\psi^0}{\mathrm{tr}(\mathbf{A}^\mathsf{T}\mathbf{A})\lambda^0}$$

EM-BG-GAMP Algorithm

Initialize EM parameters $(\lambda^0, \theta^0, \phi^0 \psi^0)$ and GAMP mean/variance $(\mathbf{\hat{x}^0}, \boldsymbol{\nu}^\mathbf{0})$

for $i = 1, 2, \ldots,$ max EM iters

for $t = 1, 2, \ldots,$ max GAMP iters

Update soft signal estimates $(\mathbf{\hat{x}^t}, \boldsymbol{\nu}^\mathbf{t})$ assuming prior params q^{i-1}

Break if early convergence

end;

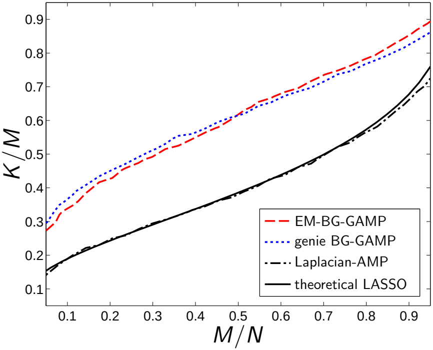
Update prior parameters $(\lambda^i, \theta^i, \phi^i, \psi^i)$ using GAMP outputs.

Break if early convergence

end;

EM-BG-GAMP Phase Transition Curve

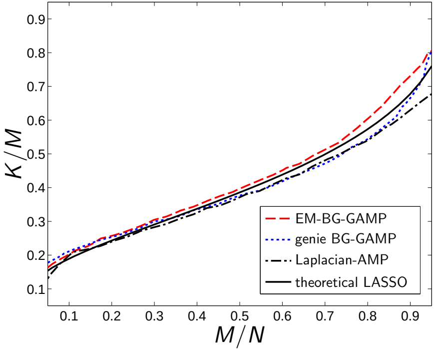
- We now demonstrate EM-BG-GAMP performance for noiseless BG signals.
- As shown, EM-BG-GAMP learns the signal prior parameters well enough to perform as good as genie BG-AMP!
- EM-BG-GAMP performs significantly better than LASSO for this signal class.



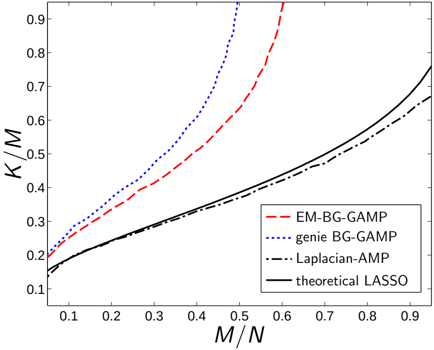
Empirical noiseless Bernoulli-Gaussian PTCs

EM-BG-GAMP PTC (cont.)

- The good performance of EM-BG-AMP is not limited to BG signals.



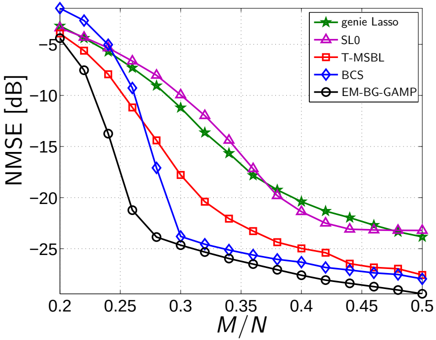
Empirical noiseless Bernoulli-Rademacher PTCs



Empirical noiseless Bernoulli PTCs

- For Bernoulli distributions, EM-BG-GAMP was able to recover nearly all signal realizations (99.8%) when $M/N > 0.65$!

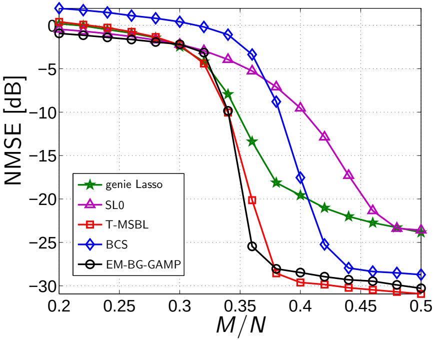
- We now compare EM-BG-GAMP to state-of-the-art CS algorithms for noisy signal recovery using normalized MSE.
- For BG signals, fix $N = 1000, K = 100, \mathrm{SNR} = 25\mathrm{dB}$ and vary M.
- EM-BG-GAMP outperforms the other algorithms for all meaningful M/N.
- The other "Bayesian" approaches, BCS and SBL, exhibit the next best performance.



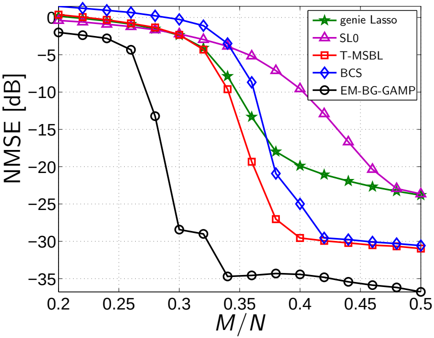
Noisy Bernoulli-Gaussian recovery NMSE.

Noisy Signal Recovery (cont.)

- We also see excellent NMSE for other K-sparse distributions:



Noisy Bernoulli-Rademacher recovery NMSE.

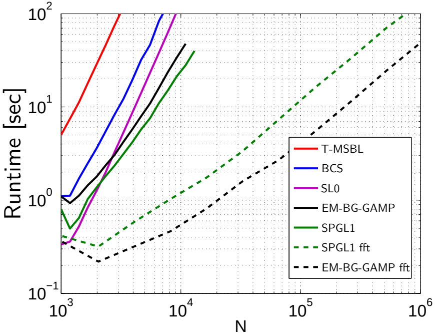


Noisy Bernoulli recovery NMSE.

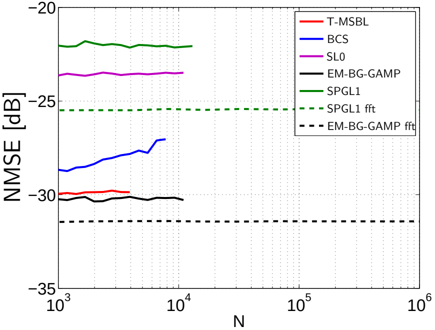
- For Bernoulli signals especially, EM-BG-GAMP exhibits a huge improvement over the other algorithms.

Algorithm Complexity

- We now compare algorithm complexity. Fix $M = 0.5N$, $K = 0.1N$, SNR $= 25$dB, and vary N. Results averaged over 50 iterations.



Noisy Bernoulli-Rademacher recovery time.

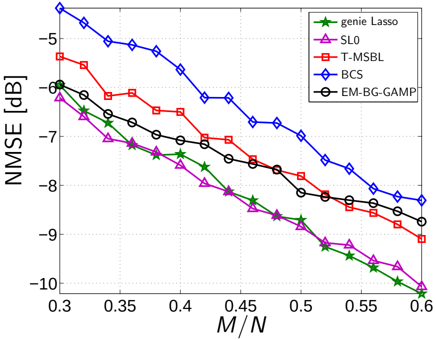


Noisy Bernoulli-Rademacher recovery NMSE.

- For large N, EM-BG-AMP has state-of-the-art complexity.

EM-BG-GAMP Limitations

- EM-BG-GAMP is outperformed by genie-LASSO and SL0 with a *non-compressible* Student's-t signal.



Noisy Student's-t recovery NMSE.

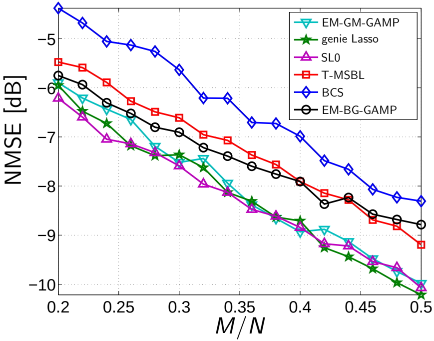
- Interestingly, the algorithms that performed best for sparse signals performed the worse for the Student's-t.

Conclusions

- We proposed an extension of BG-AMP wherein the signal and noise distributional parameters were automatically learned via the EM algorithm.
- Advantages of EM-BG-AMP
  - State-of-the-art NMSE performance for a wide class of signal/matrix types.
  - State-of-the-art complexity scaling as problem dimensions get large.
  - No tuning parameters.
- Limitations of EM-BG-AMP
  - If the true signal/noise pdfs cannot be well matched by BG/Gaussian priors, then performance may suffer.
- To address this limitation, we are working on a Gaussian-Mixture version (EM-GM-AMP) with automatic selection of the mixture order.

EM-GM-GAMP Teaser

- Our new EM-GM-GAMP algorithm may alleviate the shortcomings seen in recovering a *non-compressible* Student's-t signal.



Noisy Student's-t recovery NMSE.

- Details coming soon.

Matlab code is publicly available at
`http://ece.osu.edu/~vilaj/EMBGAMP/EMBGAMP.html`

Thanks!

Explicit Results

- GAMP outputs:

$$\begin{aligned}\hat{x} &= \pi(\hat{r},\nu^r;\mathbf{q})\,\gamma(\hat{r},\nu^r;\mathbf{q})\\ \nu^x &= \pi(\hat{r},\nu^r;\mathbf{q})\left(\beta(\hat{r},\nu^r;\mathbf{q})+|\gamma(\hat{r},\nu^r;\mathbf{q})|^2\right)-\left(\pi(\hat{r},\nu^r;\mathbf{q})\right)^2|\gamma(\hat{r},\nu^r;\mathbf{q})|^2,\end{aligned}$$

where

$$\begin{aligned}p(s=1|y) \triangleq \pi(\hat{r},\nu^r;\mathbf{q}) &\triangleq \frac{1}{1+\left(\frac{\lambda}{1-\lambda}\frac{\mathcal{N}(\hat{r};\theta,\phi+\nu^r)}{\mathcal{N}(\hat{r};0,\nu^r)}\right)^{-1}}\\ \mathbb{E}\left[x|y,s=1\right] \triangleq \gamma(\hat{r},\nu^r;\mathbf{q}) &\triangleq \frac{\hat{r}/\nu^r+\theta/\phi}{1/\nu^r+1/\phi}\\ \mathrm{var}(x|y,s=1) \triangleq \beta(\hat{r},\nu^r;\mathbf{q}) &\triangleq \frac{1}{1/\nu^r+1/\phi}.\end{aligned}$$

- EM updates:

$$\lambda^{i+1}=\frac{1}{N}\sum_{n=1}^{N}\pi(\hat{r}_n,\nu_n^r;\mathbf{q}^i). \qquad \theta^{i+1}=\frac{1}{\lambda^{i+1}N}\sum_{n=1}^{N}\pi(\hat{r}_n,\nu_n^r;\mathbf{q}^i)\gamma(\hat{r}_n,\nu_n^r;\mathbf{q}^i)$$

$$\phi^{i+1}=\frac{1}{\lambda^{i+1}N}\sum_{n=1}^{N}\pi(\hat{r}_n,\nu_n^r,\mathbf{q}^i)\Big(\left|\theta^i-\gamma(\hat{r}_n,\nu_n^r;\mathbf{q}^i)\right|^2+\beta(\hat{r}_n,\nu_n^r;\mathbf{q}^i)\Big)$$