Parametric Bilinear Generalized Approximate Message Passing

Phil Schniter and Jason Parker

The Ohio State University

With support from NSF CCF-1218754 and an AFOSR Lab Task (under Dr. Arje Nachman).

ITA — Feb 6, 2015

Approximate Message Passing (AMP) & Generalizations

Previously, AMP algorithms have been proposed ...

- for the linear model:

  Infer $\boldsymbol{x} \sim \prod_n p_{\mathsf{x}}(x_n)$ from $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w}$
  with AWGN $\boldsymbol{w}$ and known $\boldsymbol{\Phi}$. [Donoho/Maleki/Montanari'09]

- for the *generalized* linear model:

  Infer $\boldsymbol{x} \sim \prod_n p_{\mathsf{x}}(x_n)$ from $\boldsymbol{y} \sim \prod_m p_{\mathsf{y|z}}(y_m|z_m)$
  with hidden $\boldsymbol{z} = \boldsymbol{\Phi}\boldsymbol{x}$ and known $\boldsymbol{\Phi}$. [Rangan'10]

- and for the generalized *bilinear* model:

  Infer $\boldsymbol{A} \sim \prod_{m,n} p_{\mathsf{a}}(a_{mn})$ and $\boldsymbol{X} \sim \prod_{n,l} p_{\mathsf{x}}(x_{nl})$ from $\boldsymbol{Y} \sim \prod_{m,l} p_{\mathsf{y|z}}(y_{ml}|z_{ml})$
  with hidden $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{X}$. [Schniter/Cevher/Parker'11]

In this talk, we describe recent work extending AMP ...

- to the *parametric* generalized bilinear model:

  Infer $\boldsymbol{b} \sim \prod_i p_{\mathsf{b}}(b_i)$ and $\boldsymbol{c} \sim \prod_j p_{\mathsf{c}}(c_j)$ from $\boldsymbol{Y} \sim \prod_{m,l} p_{\mathsf{y|z}}(y_{ml}|z_{ml})$ with hidden
  $\boldsymbol{Z} = \boldsymbol{A}(\boldsymbol{b})\boldsymbol{X}(\boldsymbol{c})$ and known matrix-valued linear $\boldsymbol{A}(\cdot)$, $\boldsymbol{X}(\cdot)$. [Parker/Schniter'14]

Example Applications of BiG-AMP

1. Matrix Completion:

   Recover low-rank matrix $\boldsymbol{AX}$
   from noise-corrupted incomplete observations $\boldsymbol{Y} = \mathcal{P}_\Omega(\boldsymbol{AX} + \boldsymbol{W})$.

2. Robust PCA:

   Recover low-rank matrix $\boldsymbol{AX}$ and sparse matrix $\boldsymbol{S}$
   from noise-corrupted observations $\boldsymbol{Y} = \boldsymbol{AX} + (\boldsymbol{S} + \boldsymbol{W}) = [\boldsymbol{A}\ \boldsymbol{I}] \left[\begin{smallmatrix} \boldsymbol{X} \\ \boldsymbol{S} \end{smallmatrix}\right] + \boldsymbol{W}$.

3. Dictionary Learning:

   Recover dictionary $\boldsymbol{A}$ and sparse matrix $\boldsymbol{X}$
   from noise-corrupted observations $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{W}$.

4. Non-negative Matrix Factorization:

   Recover non-negative matrices $\boldsymbol{A}$ and $\boldsymbol{X}$
   from noise-corrupted observations $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{W}$.

A detailed numerical comparison[1] against state-of-the-art algorithms suggests

- BiG-AMP gives best-in-class phase transitions,
- BiG-AMP gives competitive runtimes.

[1] Parker,Schniter,Cevher, IEEE-TSP'14

Example Applications of *Parametric* BiG-AMP

1. Nonlinear Compressed Sensing with Structured Matrix Uncertainty

   Observe $\boldsymbol{y} = f\big((\sum_i b_i \boldsymbol{\Phi}_i)\boldsymbol{c} + \boldsymbol{w}\big)$ with known $\boldsymbol{\Phi}_i$.
   Recover sparse vector $\boldsymbol{c}$.

2. Generalized Matrix Recovery:

   Observe $\boldsymbol{Y} = f(\boldsymbol{\Phi}\boldsymbol{B}\boldsymbol{C} + \boldsymbol{W})$ with known $\boldsymbol{\Phi}$ and separable nonlinearity $f(\cdot)$.
   Recover low-rank matrix $\boldsymbol{B}\boldsymbol{C}$.

3. Array Calibration:

   Observe $\boldsymbol{Y} = \text{Diag}(\boldsymbol{b} \otimes \mathbf{1})\boldsymbol{\Phi}\boldsymbol{C} + \boldsymbol{W}$ with known $\boldsymbol{\Phi}$.
   Recover calibration parameters $\boldsymbol{b}$ and signal matrix $\boldsymbol{C}$.

4. Blind Deconvolution:

   Observe $\boldsymbol{Y} = \boldsymbol{\Phi}\text{Conv}(\boldsymbol{b})\boldsymbol{\Psi}\boldsymbol{C} + \boldsymbol{W}$ with known $\boldsymbol{\Phi}$ and dictionary $\boldsymbol{\Psi}$.
   Recover filter $\boldsymbol{b}$ and sparse signal coefficients $\boldsymbol{C}$.

5. Data Fusion:

   Observe $\boldsymbol{Y}_i = \boldsymbol{\Phi}_i\boldsymbol{B}\boldsymbol{C}\boldsymbol{\Omega}_i + \boldsymbol{W}_i$ for $i = 1, 2, \ldots, T$, with known $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Omega}_i$.
   Estimate tall $\boldsymbol{B}$ and wide/sparse $\boldsymbol{C}$

6. and many more . . .

Parametric BiG-AMP: Derivation

- The functions $\boldsymbol{A}(\cdot)$ and $\boldsymbol{X}(\cdot)$ are treated as random affine transformations.
- In particular, if $\boldsymbol{b} \in \mathbb{R}^{N_b}$ and $\boldsymbol{c} \in \mathbb{R}^{N_c}$, then

$$a_{mn}(\boldsymbol{b}) = \frac{1}{\sqrt{N_b}} a_{mn}^{(0)} + \sum_{i=1}^{N_b} b_i a_{mn}^{(i)} = \sum_{i=0}^{N_b} b_i a_{mn}^{(i)}$$

$$x_{nl}(\boldsymbol{c}) = \frac{1}{\sqrt{N_c}} x_{nl}^{(0)} + \sum_{j=1}^{N_c} c_j x_{nl}^{(j)} = \sum_{j=0}^{N_c} c_j x_{nl}^{(j)},$$

  where $a_{mn}^{(i)}$ and $x_{nl}^{(j)}$ are realizations of independent zero-mean r.v.s.
- We then consider the large-system limit where $N, M, L, N_b, N_c \to \infty$ such that M/N, L/N, N_b/N^2, and N_c/N^2 converge to fixed positive constants.
- The remainder of the derivation follows along the lines of BiG-AMP,[2] but is more involved/tedious.
- In practice, we also consider smooth, non-linear $\boldsymbol{A}(\cdot)$ and $\boldsymbol{X}(\cdot)$ with partial derivatives $a_{mn}^{(i)}(\boldsymbol{b})$ and $x_{nl}^{(j)}(\boldsymbol{c})$, although without rigorous justification.

[2] Parker, Schniter, Cevher, IEEE-TSP'14

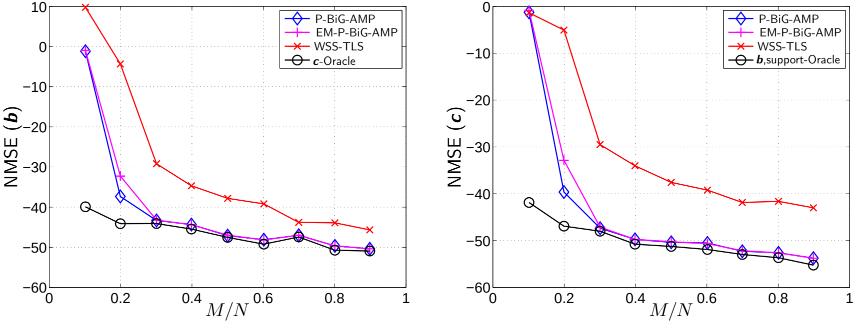
Parametric BiG-AMP: Features & Extensions

- The P-BiG-AMP algorithm exploits fast implementations of $\boldsymbol{A}(\cdot)$ and $\boldsymbol{X}(\cdot)$ (e.g., FFT-based).

- Although P-BiG-AMP requires knowledge of the priors on $\boldsymbol{b}$ and $\boldsymbol{c}$ and the likelihood function $p_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\cdot)$, the hyper-parameters can be learned from the data using the expectation maximization approach proposed for AMP in [Schniter/Vila'11].

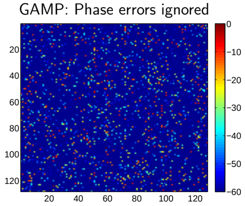
- Although P-BiG-AMP assumes independent $\{b_i\}$, independent $\{c_j\}$, and conditionally independent $\{y_{m,n}|z_{m,n}\}$, more general models can be handled using the turbo-AMP approach proposed in [Schniter'10].

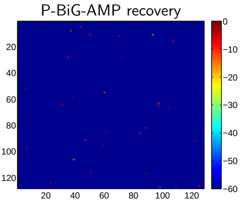
Example 1: CS with Structured Matrix Uncertainty

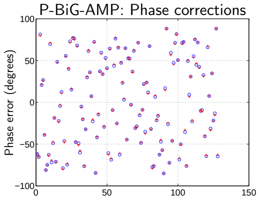


- Measure: $\boldsymbol{y} = \left(\boldsymbol{A}_0 + \sum_{i=1}^{N_b} b_i \boldsymbol{A}_i\right)\boldsymbol{c} + \boldsymbol{w}$, $(N = 256,\ N_b = 10,\ \text{SNR} = 40\text{dB})$
- Unknown (all iid): $\mathsf{w}_m \sim \mathcal{N}(0, \nu^w)$, $\mathsf{b}_i \sim \mathcal{N}(0,1)$, $\mathsf{c}_j \sim \mathcal{BG}(0.04, 0, 1)$
- Known (drawn iid): $[\boldsymbol{A}_0]_{mn} \sim \mathcal{N}(0, N_b)$, $[\boldsymbol{A}_i]_{mn} \sim \mathcal{N}(0,1)$
- EM-P-BiG-AMP outperforms oracle-tuned WSS-TLS [Zhu/Leus/Giannakis'11]

Example 2: Random 2D Fourier Measurements of a Sparse Image with Row-Wise Phase Errors







- Randomly sample 10% of the AWGN-corrupted (@40dB SNR) 2D Fourier measurements of a 128×128 image with 30 non-zero pixels
- An unknown random phase (uniformly distributed on $[-90^\circ, +90^\circ]$) is added to all the measurements from each row of the observations
- P-BiG-AMP jointly estimates phase errors and sparse image to -50dB NMSE.
- Surrogate for simultaneous sparse imaging and autofocus [Önhon/Çetin'12]

Summary

- Presented preliminary work on an algorithm for parametric, bilinear, generalized inference based on AMP principles.
- Assumes unknown independent random vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ are related to observations $\boldsymbol{Y}$ through a conditionally independent likelihood of the form

$$p(\boldsymbol{Y}|\boldsymbol{A}(\boldsymbol{b})\boldsymbol{X}(\boldsymbol{c})) \text{ with known affine } \boldsymbol{A}(\cdot) \text{ and } \boldsymbol{X}(\cdot).$$

- Builds on previous Bilinear Generalized AMP work.
- Can be combined with EM and turbo AMP methods.
- Numerical experiments demonstrate performance near oracle bounds.

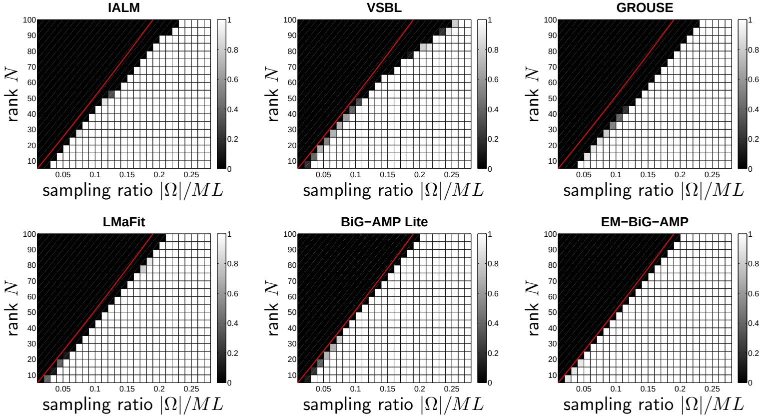
References

1. D.L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *PNAS*, 2009.
2. S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *ISIT*, 2011. (See also *arXiv:1010.5141*).
3. P. Schniter and V. Cevher, "Approximate message passing for bilinear models," *SPARS*, 2011.
4. J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Message Passing—Part 1: Derivation," *IEEE Trans. Signal Process.*, 2014.
5. J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Message Passing—Part 2: Applications," *IEEE Trans. Signal Process.*, 2014.
6. J. T. Parker, *Approximate Message Passing Algorithms for Generalized Bilinear Inference*, Ph.D. Dissertation, Dept. ECE, The Ohio State University, Columbus OH, 2014.
7. J. P. Vila and P. Schniter, "Expectation-Maximization Gaussian-Mixture Approximate Message Passing," *IEEE Trans. Signal Process.*, 2013.
8. P. Schniter, "Turbo reconstruction of structured sparse signals," *Proc. Conf. Inform. Science & Syst.*, 2010.
9. H. Zhu, G. Leus, and G. Giannakis, "Sparsity-Cognizant Total Least-Squares for Perturbed Compressive Sampling," *IEEE Trans. Signal Process.*, 2011.
10. N. Önhon and M. Çetin, "A Sparsity-Driven Approach for Joint SAR Imaging and Phase Error Correction," *IEEE Trans. Image Process.*, 2012.

Thanks for listening!

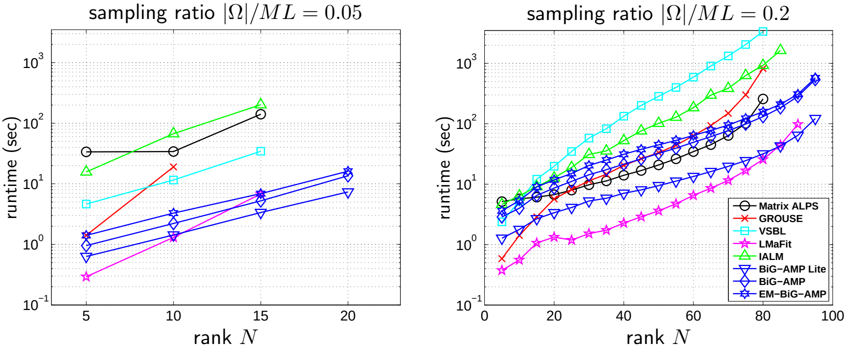
Matrix Completion: Phase Transitions

The following plots show empirical probability that NMSE < -100 dB (over 10 realizations) for noiseless completion of an $M \times L$ matrix with $M = L = 1000$.



Note that BiG-AMP-Lite and EM-BiG-AMP have the best phase transitions.

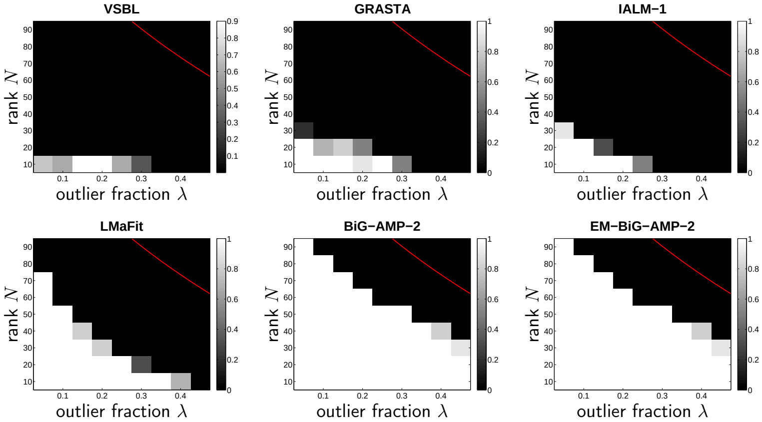
Matrix Completion: Runtime to NMSE=-100 dB



- Although LMaFit is the fastest algorithm at small rank N, BiG-AMP-Lite's superior complexity-scaling-with-N eventually wins out.
- BiG-AMP runs 1 to 2 orders-of-magnitude faster than IALM and VSBL.

Robust PCA: Phase Transitions

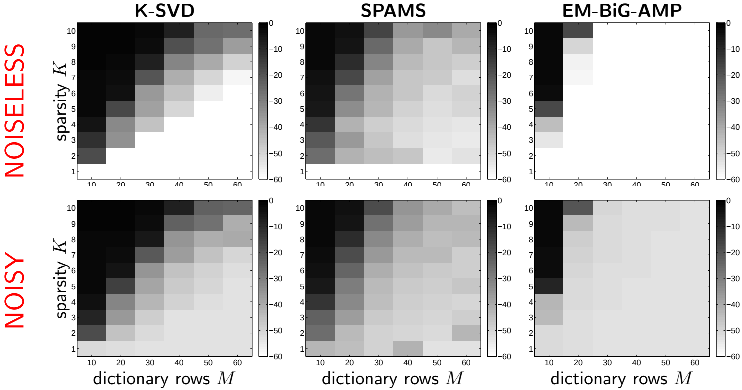
Empirical probability of NMSE < -80 dB over 10 realizations for noiseless recovery of the low-rank component of a 200×200 outlier-corrupted matrix.



As before, the BiG-AMP methods yield the best phase transitions.

Overcomplete Dictionary Recovery: Phase Transitions

Mean NMSE over 50 realizations for recovery of an $M \times (2M)$ dictionary from $L = 10M \log(2M)$ examples with sparsity K:



As before, the BiG-AMP methods yield the best phase transitions.