Structured Matrix Estimation via Approximate Message Passing

Phil Schniter and Jason Parker

THE OHIO STATE UNIVERSITY

With support from NSF CCF-1218754, NSF CCF-1527162, and an AFOSR Lab Task (under Dr. Arje Nachman).

ITA Workshop (San Diego) — Feb 5, 2016

The Generalized Bilinear Model

- Bilinear model:

  Infer $\boldsymbol{b} \in \mathbb{R}^{N_b}$ and $\boldsymbol{c} \in \mathbb{R}^{N_c}$ from bilinear measurements

  $$y_m = \boldsymbol{b}^\mathsf{T} \boldsymbol{\Phi}_m \boldsymbol{c} + w_m, \quad m = 1...M,$$

  where $\{\boldsymbol{\Phi}_m\}$ are known matrices and $\{w_m\}$ are independent noise samples.

- Generalized bilinear model:

  Infer $\boldsymbol{b}$ and $\boldsymbol{c}$ from

  $$y_m = f\big(\boldsymbol{b}^\mathsf{T} \boldsymbol{\Phi}_m \boldsymbol{c} + w_m\big), \quad m = 1...M,$$

  where $f(\cdot)$ is possibly non-linear (e.g., quantization, loss of phase).

Some Applications of the Generalized Bilinear Model

1. Self-Calibration

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

2. Blind Deconvolution

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

3. Joint channel-symbol estimation

   Observe $\boldsymbol{Y} = \text{Conv}(\boldsymbol{b})\boldsymbol{C} + \boldsymbol{W}$.
   Recover sparse channel coefficients $\boldsymbol{b}$ and coded finite-alphabet symbols $\boldsymbol{C}$.

4. Recovery of low-rank plus sparse matrix

   Observe $y_m = \text{tr}\{\boldsymbol{\Phi}_m^{\mathsf{T}}(\boldsymbol{L} + \boldsymbol{S})\} + w_m$ with known $\boldsymbol{\Phi}_m$ for $m = 1...M$.
   Recover low-rank matrix $\boldsymbol{L}$ and sparse matrix $\boldsymbol{S}$.

5. 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 $\{\boldsymbol{\Phi}_i\}$ and componentwise $f(\cdot)$.
   Recover sparse vector $\boldsymbol{c}$.

6. and many more ...

Extends Earlier "BiG-AMP" for Matrix Factorization

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 excellent phase transitions,
- BiG-AMP gives competitive runtimes.

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

Parametric Bilinear Generalized AMP (PBiG-AMP)

- Separable probabilistic model:

  Recall $y_m = f(\underbrace{\boldsymbol{b}^\mathsf{T}\boldsymbol{\Phi}_m\boldsymbol{c}}_{\triangleq\, z_m} + w_m), \quad m = 1...M.$

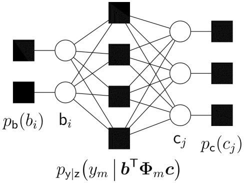
  Treat $\boldsymbol{b} \sim \prod_i p_\mathsf{b}(b_i)$, $\boldsymbol{c} \sim \prod_j p_\mathsf{c}(c_j)$, and $\boldsymbol{y} \sim \prod_m p_{\mathsf{y}|\mathsf{z}}(y_m|z_m)$.

- Treat $\boldsymbol{\Phi} \triangleq \{\boldsymbol{\Phi}_1, ..., \boldsymbol{\Phi}_M\}$ as i.i.d. Gaussian and consider large-system limit:

  $M, N_b, N_c \to \infty$ with N_b/M and N_c/M converging to fixed constants.

- Sum-product algorithm simplifies:
  - Messages become Gaussian
  - Number of messages reduces from $2M(N_b+N_c)$ to $2(M+N_b+N_c)$

  $\rightsquigarrow$ "Approximate Message Passing" [Donoho,Maleki,Montanari'09],[Rangan'10]



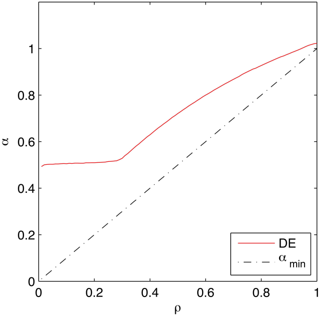
PBiG-AMP: Density Evolution

Collaborators Christophe Schülke and Lenka Zdeborova showed that...

- For i.i.d. Gaussian $\boldsymbol{\Phi}$ in the large-system limit, PBiG-AMP is characterized by a scalar density evolution.

- Example: Gaussian $p_{\mathsf{y|z}}$
  Bernoulli-Gaussian $p_{\mathsf{b}} = p_{\mathsf{c}}$
  $N_b = N_c \triangleq N$

  Density evolution predicts the phase transition:



  where

  ρ: sparsity ratio $\triangleq K/N$
  α: sampling rate $\triangleq M/(2N)$
  $-\cdot$: counting bound

PBiG-AMP: Implementation & Extensions

- Our implementation (`https://sourceforge.net/projects/gampmatlab/`) allows...
  - non-identical priors $\{p_{\mathsf{b}_i}\}, \{p_{\mathsf{c}_j}\}$ and likelihood $\{p_{\mathsf{y}_m|\mathsf{z}_m}\}$
  - complex-valued quantities
  - fast implementations of generic $\boldsymbol{\Phi}_m$ (e.g., FFTs, sparse, etc.).

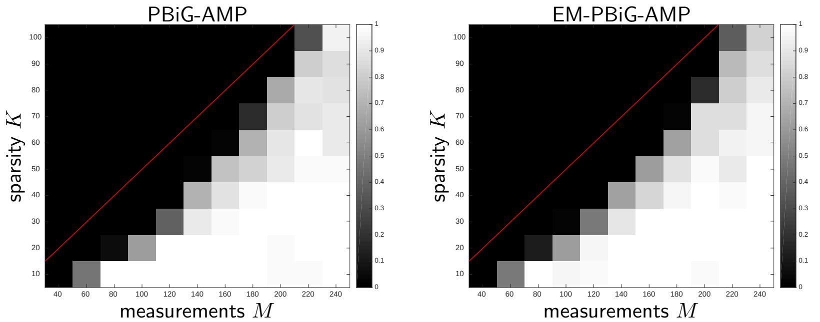
- Prior/likelihood parameters (e.g., sparsity, noise variance) can be tuned online using the same expectation maximization (EM) methods proposed for AMP [Schniter/Vila'11].

- Can relax the separability assumptions

$$\boldsymbol{b} \sim \prod_i p_{\mathsf{b}_i}(b_i), \quad \boldsymbol{c} \sim \prod_j p_{\mathsf{c}_j}(b_j), \quad \boldsymbol{y}|\boldsymbol{z} \sim \prod_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m)$$

by embedding PBiG-AMP in a larger factor graph (i.e., "turbo-AMP") [Schniter'10].

Example 1: Empirical Phase Transition under iid $\mathcal{N}$ $\mathbf{\Phi}$

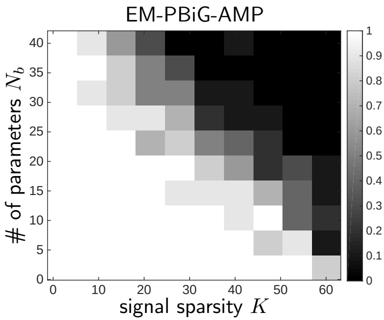


- Measure noiseless $y_m = \boldsymbol{b}^\mathsf{T}\mathbf{\Phi}_m\boldsymbol{c}$ for $m = 1...M$ with $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^{100}$

  Recover iid: $\mathsf{b}_i, \mathsf{c}_j \sim \mathcal{BG}(K/100, 0, 1)$

  Known iid Gaussian $\mathbf{\Phi}_m$
- Phase transition (NMSE$< 10^{-6}$) close to counting bound (red line).
- Due to finite-size effects, empirical performance beats the density evolution!

Example 2: Self-Calibration



- Measure noiseless $\boldsymbol{y} = \mathcal{D}(\boldsymbol{F}\boldsymbol{b})\boldsymbol{\Phi}\boldsymbol{c}$

  Recover unknown Gaussian $\boldsymbol{b} \in \mathbb{R}^{N_b}$, K-sparse $\boldsymbol{c} \in \mathbb{R}^{256}$.

  Known DFT $\boldsymbol{F} \in \mathbb{C}^{128\times N_b}$ and i.i.d. Gaussian $\boldsymbol{\Phi} \in \mathbb{R}^{128\times 256}$
- $M \gtrsim O(N_b+K)$ measurements suffice for EM-PBiG-AMP
- $M \gtrsim O(N_b K)$ are needed for SparseLift from [Ling/Strohmer'14].

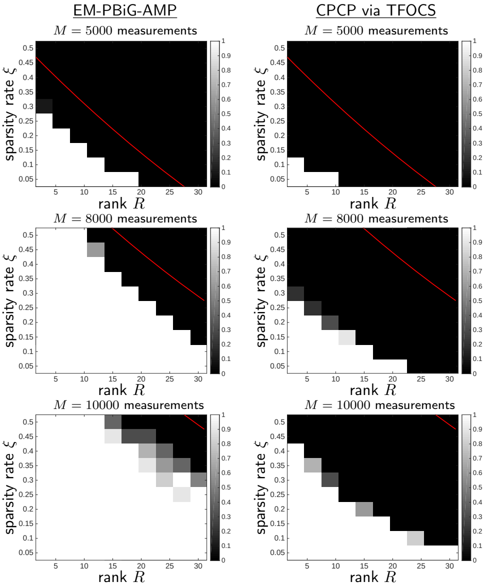
Example 3: Matrix Compressive Sensing

- For $m = 1...M$, measure $y_m = \mathrm{tr}\{\boldsymbol{\Phi}_m^{\mathsf{T}}(\boldsymbol{L}+\boldsymbol{S})\}$ with 50-sparse iid $\boldsymbol{\Phi}_m$.

  Recover sparse $\boldsymbol{S}$ and rank-R $\boldsymbol{L} \in \mathbb{R}^{100\times 100}$

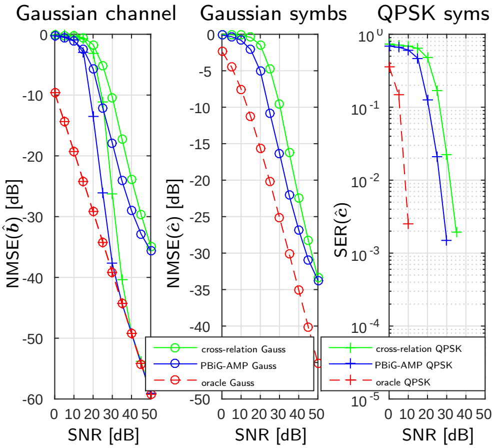
- EM-PBiG-AMP outperforms convex relaxation known as Compressive Principal Components Pursuit [Wright/Ganesh/Min/Ma'13]

- PBiG-AMP runs significantly faster than CPCP via TFOCS.

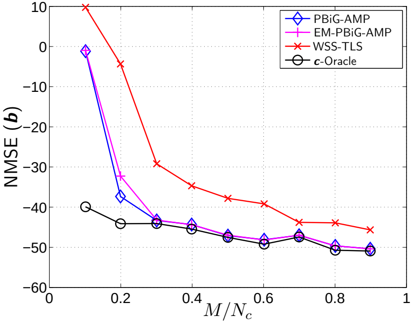


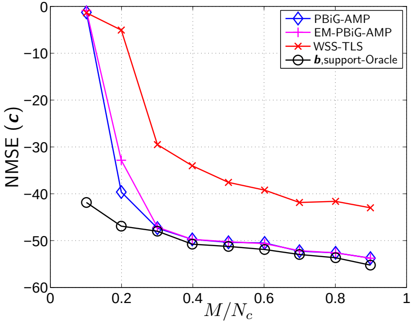
Example 4: Totally Blind Deconvolution

- Measure linear convolution outputs $y_m = b_m * c_m + w_m$.
  - Recover both b_m and c_m.
  - Zero-valued guards ensure identifiability [Manton'03].
- PBiG-AMP outperforms blind Cross Relation method [Hua'96]
- With QPSK at high SNR, PBiG-AMP achieves oracle performance.



Example 5: CS with Structured Matrix Uncertainty





- Measure: $\boldsymbol{y} = \left(\boldsymbol{\Phi}_0 + \sum_{i=1}^{10} b_i \boldsymbol{\Phi}_i\right)\boldsymbol{c} + \boldsymbol{w}$, $(N_c = 256, \text{ SNR} = 40\text{dB})$

  Recover iid: $w_m \sim \mathcal{N}(0, \nu^w)$, $b_i \sim \mathcal{N}(0,1)$, $c_j \sim \mathcal{BG}(0.04, 0, 1)$

  Known iid: $[\boldsymbol{\Phi}_0]_{mj} \sim \mathcal{N}(0, 10)$, $[\boldsymbol{\Phi}_i]_{mj} \sim \mathcal{N}(0,1)$
- EM-PBiG-AMP outperforms oracle-tuned WSS-TLS [Zhu/Leus/Giannakis'11]

Summary

- Proposed an AMP algorithm for the generalized bilinear model.
- Assumes unknown independent random vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ are related to observations $\{y_m\}$ through a conditionally independent likelihood of the form

$$p(y_m|\boldsymbol{b}^\mathsf{T}\boldsymbol{\Phi}_m\boldsymbol{c}) \text{ with large i.i.d. Gaussian } \boldsymbol{\Phi}_m.$$

- Behavior characterized by a scalar density evolution.
- Generalizes previous work on matrix-factorization AMP.
- Numerical experiments demonstrate performance near oracle bounds for various interesting $\boldsymbol{\Phi}$ that are *not* i.i.d. Gaussian.
- Full paper can be found at `http://arxiv.org/abs/1508.07575`.

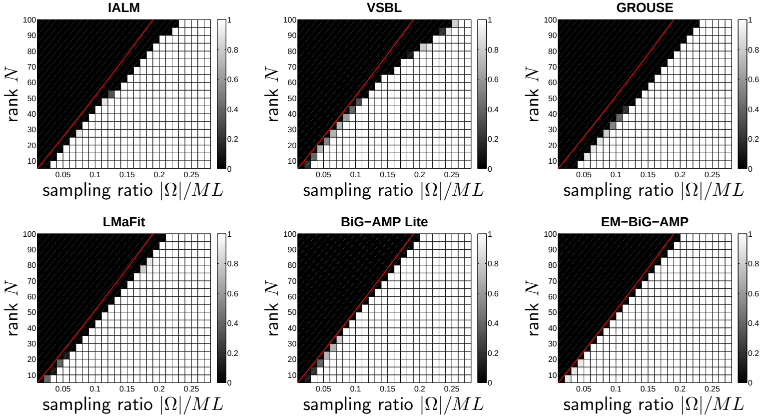
References

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. J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Message Passing—Part 1: Derivation," *IEEE Trans. Signal Process.*, 2014.
4. J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Message Passing—Part 2: Applications," *IEEE Trans. Signal Process.*, 2014.
5. J. T. Parker and P. Schniter, "Parametric Bilinear Generalized Approximate Message Passing," `http://arxiv.org/abs/1508.07575`, 2015.
6. J. P. Vila and P. Schniter, "Expectation-Maximization Gaussian-Mixture Approximate Message Passing," *IEEE Trans. Signal Process.*, 2013.
7. P. Schniter, "Turbo reconstruction of structured sparse signals," *Proc. Conf. Inform. Science & Syst.*, 2010.
8. S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, 2015.
9. H. Zhu, G. Leus, and G. Giannakis, "Sparsity-Cognizant Total Least-Squares for Perturbed Compressive Sampling," *IEEE Trans. Signal Process.*, 2011.
10. J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Inform. Inference*, 2013.
11. J. H. Manton and W. D. Neumann, "Totally blind channel identication by exploiting guard intervals," *Syst. Control Lett.*, 2003.
12. Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, 1996.

Thanks for listening!

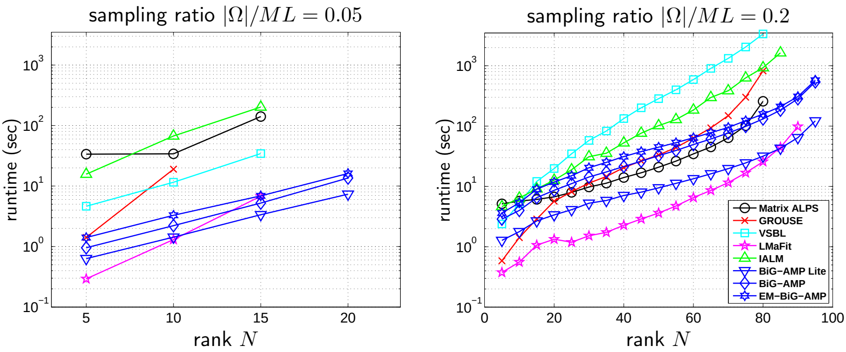
Matrix Completion: Phase Transitions

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.

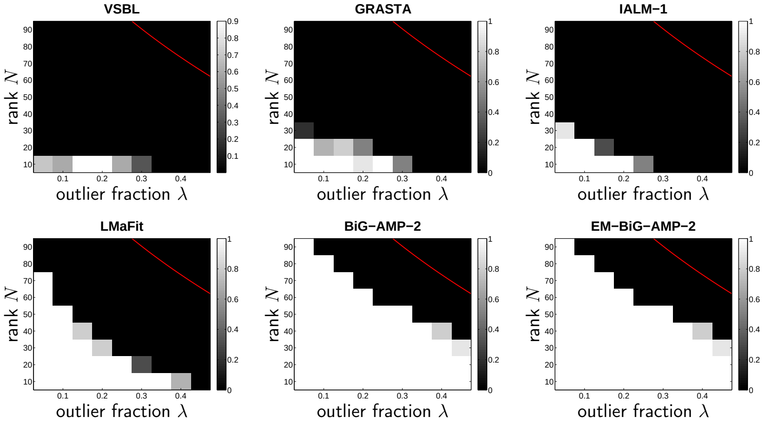
Matrix Completion: Runtime to NMSE=-100 dB



- 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

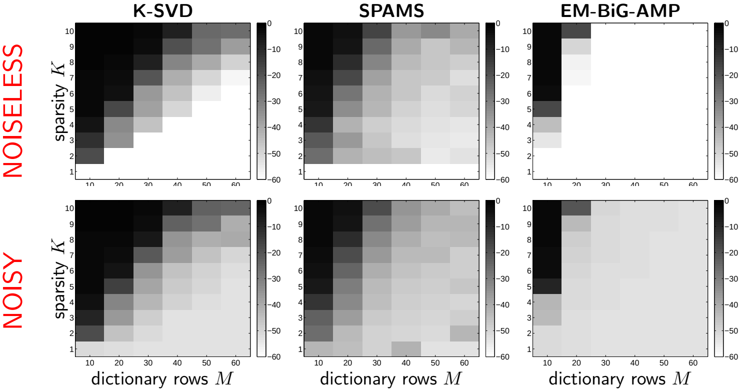
Empirical probability of NMSE < -80 dB over 10 realizations for noiseless recovery of the low-rank component of a 200×200 outlier-corrupted matrix.



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.