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 \( x \sim \prod_n p_x(x_n) \) from \( y = \Phi x + w \)
  
  with AWGN \( w \) and known \( \Phi \).

  [Donoho/Maleki/Montanari’09]

- for the **generalized linear model**:
  
  Infer \( x \sim \prod_n p_x(x_n) \) from \( y \sim \prod_m p_y|z(y_m|z_m) \)
  
  with hidden \( z = \Phi x \) and known \( \Phi \).

  [Rangan’10]

- and for the **generalized bilinear model**:
  
  Infer \( A \sim \prod_{m,n} p_a(a_{mn}) \) and \( X \sim \prod_{n,l} p_x(x_{nl}) \)
  
  from \( Y \sim \prod_{m,l} p_y|z(y_{ml}|z_{ml}) \)
  
  with hidden \( Z = AX \).

  [Schniter/Cevher/Parker’11]

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

- to the **parametric generalized bilinear model**:
  
  Infer \( b \sim \prod_i p_b(b_i) \) and \( c \sim \prod_j p_c(c_j) \)
  
  from \( Y \sim \prod_{m,l} p_y|z(y_{ml}|z_{ml}) \) with hidden

  \( Z = A(b)X(c) \) and known matrix-valued linear \( A(\cdot), X(\cdot) \).

  [Parker/Schniter’14]
Example Applications of BiG-AMP

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


4. **Non-negative Matrix Factorization**: Recover non-negative matrices $A$ and $X$ from noise-corrupted observations $Y = AX + W$.

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
1. **Nonlinear Compressed Sensing with Structured Matrix Uncertainty**
   Observe $y = f\left((\sum_i b_i \Phi_i)c + w\right)$ with known $\Phi_i$.
   Recover sparse vector $c$.

2. **Generalized Matrix Recovery**:
   Observe $Y = f(\Phi BC + W)$ with known $\Phi$ and separable nonlinearity $f(\cdot)$.
   Recover low-rank matrix $BC$.

3. **Array Calibration**:
   Observe $Y = \text{Diag}(b \otimes 1)\Phi C + W$ with known $\Phi$.
   Recover calibration parameters $b$ and signal matrix $C$.

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

5. **Data Fusion**:
   Observe $Y_i = \Phi_i BC \Omega_i + W_i$ for $i = 1, 2, \ldots, T$, with known $\Phi_i$ and $\Omega_i$.
   Estimate tall $B$ and wide/sparse $C$.

6. and many more . . .
The functions $A(\cdot)$ and $X(\cdot)$ are treated as random affine transformations.

In particular, if $b \in \mathbb{R}^{N_b}$ and $c \in \mathbb{R}^{N_c}$, then

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

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

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, but is more involved/tedious.

In practice, we also consider smooth, non-linear $A(\cdot)$ and $X(\cdot)$ with partial derivatives $a_{mn}^{(i)}(b)$ and $x_{nl}^{(j)}(c)$, 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 $A(\cdot)$ and $X(\cdot)$ (e.g., FFT-based).

- Although P-BiG-AMP requires knowledge of the priors on $b$ and $c$ and the likelihood function $p_{Y|Z}(y|\cdot)$, 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: $\mathbf{y} = \left( \mathbf{A}_0 + \sum_{i=1}^{N_b} b_i \mathbf{A}_i \right) \mathbf{c} + \mathbf{w}$, ($N = 256, N_b = 10, \text{SNR} = 40\text{dB}$)
- Unknown (all 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 (drawn iid): $[\mathbf{A}_0]_{mn} \sim \mathcal{N}(0, N_b)$, $[\mathbf{A}_i]_{mn} \sim \mathcal{N}(0, 1)$
- 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 (40 dB SNR) 2D Fourier measurements of a 128 × 128 image with 30 non-zero pixels
- An unknown random phase (uniformly distributed on [−90°, +90°]) is added to all the measurements from each row of the observations
- P-BiG-AMP jointly estimates phase errors and sparse image to −50 dB NMSE.
- Surrogate for simultaneous sparse imaging and autofocus [Önhon/Çetin’12]
Presented preliminary work on an algorithm for parametric, bilinear, generalized inference based on AMP principles.

Assumes unknown independent random vectors $b$ and $c$ are related to observations $Y$ through a conditionally independent likelihood of the form

$$p(Y | A(b)X(c))$$

with known affine $A(\cdot)$ and $X(\cdot)$.

Builds on previous Bilinear Generalized AMP work.

Can be combined with EM and turbo AMP methods.

Numerical experiments demonstrate performance near oracle bounds.


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.
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 $200 \times 200$ outlier-corrupted matrix.

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