State Evolution Analysis of Approximate Message Passing with Side Information

Hangjin Liu, Cynthia Rush, and Dror Baron
NC State University and Columbia University

IEEE International Symposium on Information Theory
Paris, France
July 2019
Linear Inverse Problem Applications

- Medical Imaging
- Seismic Imaging
- Wireless Communication
Problem Setting

• High-dimensional linear regression model:
  \[ y = Ax + w \]

  • \( x \in \mathbb{R}^N \): unknown input
  • \( y \in \mathbb{R}^M \): noisy measurements
  • \( A \in \mathbb{R}^{M \times N} \): measurement matrix \((M < N), \delta = \frac{M}{N}\)
  • \( w \in \mathbb{R}^M \): noise

Goal: Reconstruct \( x \), given \( A, y \)
Methods for Signal Reconstruction

• Matching pursuit [Mallat et al. 1993]
• $l_1$ optimization [Chen et al. 2001]

• Statistical info on $x, w \rightarrow$ approximate message passing (AMP) [Donoho et al. 2009] and Generalized AMP (GAMP) [Rangan 2012]
• Bayes-optimal in large system limit ($M, N \rightarrow \infty, \delta = \frac{M}{N}$ fixed) [Bayati et al. 2011]
• Better reconstruction from fewer measurements
Approximate Message Passing (AMP)

[Donoho et al. 2009]

Iterate:

• Residual: $r^t = y - Ax^t + \text{correction term}$

• Pseudo-data: $v^t = x^t + A^T r^t$

• Denoising: $x^{t+1} = \eta_t(v^t)$
AMP

**Iterate:**

- Residual: $r^t = y - Ax^t + \frac{r^{t-1}}{\delta} \sum_{i=1}^{N} [\eta_{t-1}(v_{i}^{t-1})]$
Iterate:

- Pseudo-data: $v^t = x^t + A^T r^t \approx x + \mathcal{N}(0, \lambda^2_t I)$

Independent and identically distributed (i.i.d.) Gaussian matrix $A$, i.i.d. input $x$
AMP

Iterate:

- Denoising: $x^{t+1} = \eta_t(v^t)$

- $\eta_t(a) = \mathbb{E}[X|a = X + \mathcal{N}(0, \lambda_t^2)]$

Denoiser provides minimum mean squared error (MSE) signal estimate
AMP Iteration 1

Pseudo-data

denoising

Denoised image
AMP Iteration 3

Pseudo-data

denoising

Denoised image
AMP Iteration 10

Pseudo-data

Denoising

Denoised image
AMP State Evolution (SE)

Scalar recursion tracks AMP performance

- **State Evolution (SE) equation**

\[
\lambda_t^2 = \sigma_W^2 + \frac{1}{\delta} \mathbb{E}[(\eta_{t-1}(X + \mathcal{N}(0, \lambda_{t-1}^2)) - X)^2]
\]

- \(\lambda_t^2\): pseudo-data noise variance
- \(W\): i.i.d., \(\sigma_W^2 = \mathbb{E}[W^2]\)
- \(X\): independent \(Z \sim \mathcal{N}(0,1)\)
Side Information (SI)
SI: Application

• Hyperspectral Imaging

• Images at nearby frequencies (colors) contain SI about current image
AMP with SI (AMP-SI) [2017]
AMP-SI [Baron et al., 2017]

• SI: $\tilde{x} \in \mathbb{R}^N$, $(x, \tilde{x}) \sim f(X, \tilde{X})$

• AMP-SI uses modified denoising stage:
  • Residual: $r^t = y - Ax^t + \frac{r^{t-1}}{\delta} \sum_{i=1}^{N} [\eta_{t-1}'([v^{t-1}]_i, \tilde{x}_i)]$
  • Denoising: $x_i^{t+1} = \eta_t([v^t]_i, \tilde{x}_i)$

$$\eta_t : \mathbb{R}^2 \rightarrow \mathbb{R}, \, \eta_t(a, b) = \mathbb{E}[X|X + \mathcal{N}(0, \lambda_t^2) = a, \tilde{X} = b]$$

Modified denoiser provides minimum MSE estimate of signal
AMP-SI with SE

[New]
AMP-SI SE

• SE equation: \[ \lambda_t^2 = \sigma_w^2 + \frac{1}{\delta} \mathbb{E}[(\eta_{t-1}(X + \mathcal{N}(0, \lambda_{t-1}^2), \tilde{X}) - X)^2] \]

• Recall denoiser: \[ \eta_t(a, b) = \mathbb{E}[X|X + \mathcal{N}(0, \lambda_t^2) = a, \tilde{X} = b] \]

How do SE and AMP-SI relate?
AMP-SI SE

• General non-separable SE results [Berthier et al., 2017]

• Applicable to AMP-SI algorithm

Paris already has a rich tradition of Berthiers (Versailles Palace, July 2019) - many more to come!
Assumptions

• Measurement matrix $A$: i.i.d. Gaussian, mean 0, variance $1/M$

• $w \sim f(W), (x, \tilde{x}) \sim f(X, \tilde{X})$ all i.i.d., finite moments

• Denoisers $\eta_t(\cdot, \cdot)$: Lipschitz continuous for scalars $a_1, a_2, b_1, b_2$; constant $L > 0$,

\[
|\eta_t(a_1, b_1) - \eta_t(a_2, b_2)| \leq L \| (a_1, b_1) - (a_2, b_2) \|_2
\]
Results: SE for AMP-SI

- Under assumptions
  \[ \lim_{N \to \infty} \frac{1}{N} ||\nu^t - \mathbf{x}||^2 \overset{p}{=} \lambda_t^2 \]
  \[ \ell_2 \text{ loss function} \]

- Similarly
  \[ \lim_{N \to \infty} \frac{1}{N} \left\| \mathbf{x}^{t+1} - \mathbf{x} \right\|^2 \overset{p}{=} \delta (\lambda_{t+1}^2 - \sigma_w^2) \]
  \[ \mathbf{x}^{t+1} = \eta_t(\nu^t, \mathbf{x}) \]
Main Theorem

• Following assumptions, for \( PL(2) \) functions \( \phi: \mathbb{R}^2 \to \mathbb{R} \) and \( \psi: \mathbb{R}^3 \to \mathbb{R} \),

\[
\lim_{M} \frac{1}{M} \sum_{i=1}^{M} \phi(r_i^t, w_i) \overset{p}{=} \mathbb{E} \left[ \phi \left( W + \sqrt{\lambda_t^2 - \sigma_W^2} Z_1, W \right) \right]
\]

\[
\lim_{N} \frac{1}{N} \sum_{i=1}^{N} \psi(v_i^t, x_i, \tilde{x}_i) \overset{p}{=} \mathbb{E}[\psi(X + \lambda_t Z_2, X, \tilde{X})]
\]

• \( Z_1, Z_2 \): standard Gaussians, independent of \( W \sim f(W), (X, \tilde{X}) \sim f(X, \tilde{X}) \)
Proof steps

• Justify assumptions (C1)-(C5) [Berthier et al., 2017]

• The interesting assumptions are

(C2) $\tilde{\eta}_N^t(x) = \eta_t(x, \bar{x})$, uniformly Lipschitz, $\eta_t$ applied entry-wise

(C5) $\lim_{N \to \infty} \frac{1}{N} \mathbb{E}_Z \left[ x^T \tilde{\eta}_N^t(x + Z) \right] < \infty$, $\lim_{N \to \infty} \frac{1}{N} \mathbb{E}_{Z, Z'} \left[ \tilde{\eta}_N^t(x + Z) \tilde{\eta}_N^s(x + Z') \right] < \infty,$

where $(Z, Z') \sim \mathcal{N}(0, \Sigma \otimes I_N)$
Theorem 14 [Berthier et al., 2017]

• Following assumptions (C1)-(C5), for uniformly PL functions

\( \rho_M : \mathbb{R}^{2M} \to \mathbb{R} \), \( \gamma_n : \mathbb{R}^{2N} \to \mathbb{R} \),

\begin{align*}
\lim_{M} \rho_M(r^t, w) & \overset{p}{=} \lim_{M} \mathbb{E}_{z_1} \left[ \rho_M \left( w + \sqrt{\frac{\lambda_t^2 - \sigma_w^2}{\lambda_t}} z_1, w \right) \right] \\
\lim_{N} \gamma_N(x^t + A^T r^t, x) & \overset{p}{=} \lim_{N} \mathbb{E}_{z_2} \left[ \gamma_N (x + \lambda_t z_2, x) \right] \\
Z_1 & \sim \mathcal{N}(0, I_M), \quad Z_2 \sim \mathcal{N}(0, I_N)
\end{align*}
Proof steps

• Justify assumptions (C1)-(C5) [Berthier et al., 2017]

• Show that \( \frac{1}{M} \sum_{i=1}^{M} \phi(r_i^t, w_i), \frac{1}{N} \sum_{i=1}^{N} \psi(v_i^t, x_i, \bar{x}_i) \) are uniformly \( PL(2) \)

• Apply Berthier et al. Theorem 14

• Apply Strong Law of Large Numbers
Examples
Simple Example - GG Model

- i.i.d. Gaussian signal $X$; Gaussian SI: $\tilde{X} = X + \mathcal{N}(0, \sigma^2 I)$

- AMP-SI denoiser: $\eta_{t-1}(a, b) = f \cdot a + g \cdot b$; $f, g$: functions of $\sigma_x^2, \sigma^2, \lambda_{t-1}^2$

- SE: $\lambda_t^2 = \sigma_w^2 + \frac{1}{\delta} \text{function}(\sigma_x^2, \sigma^2, \lambda_{t-1}^2)$
AMP-SI (SE Prediction vs Empirical)

Empirical MSE performance of AMP-SI and SE prediction
(GG model, $\delta = 0.3, \sigma_x = 1, \sigma_\omega = 0.1$, and $\sigma = 0.2$)
BG Model

- Bernoulli-Gaussian signal: $x_i \sim \epsilon \mathcal{N}(0, 1) + (1 - \epsilon) \delta_0$
  - Zero w.p. $1 - \epsilon$, else $\mathcal{N}(0, 1)$

- Gaussian SI: $\tilde{X} = X + \mathcal{N}(0, \sigma^2 I)$
BG Model

- AMP-SI denoiser: $\eta_{t-1}(a, b) = \Pr(X \neq 0|a, b) \mathbb{E}[X|a, b, X \neq 0]$
  - $\mathbb{E}[X|a, b, X \neq 0] = f \cdot a + g \cdot b$; $f, g$: functions of $\sigma^2, \lambda^2_{t-1}$

- $\eta_{t-1}(a, b)$ has bounded partial derivative $\rightarrow$ Lipschitz continuous
AMP-SI (SE Prediction vs Empirical)

Empirical MSE performance of AMP-SI and SE prediction
(BG model, \(N = 10000\), \(M = 3000\), \(\epsilon = 0.2\), \(\sigma_x = 1\), \(\sigma_\omega = 0.1\))
Future Work

• AMP-SI State Evolution for
  • time-varying signals
  • non-i.i.d. signal and side information

• Apply AMP-SI to hyperspectral, video, ...
Thanks!

Supported by the National Science Foundation under the Grant EECS 1611112
Reference:


• The hyper spectral Lego scene was not endorsed by the trademark owners and is used here as fair use to illustrate qualitative aspects of hyper spectral image data. LEGO is a trademark of the LEGO Group, which does not sponsor, authorize or endorse the images in this presentation. The LEGO Group. All Rights Reserved. http://aboutus.lego.com/enus/legal-notice/fair-play/