S-AMP: Approximate Message Passing for General Matrix Ensembles
Cakmak, Burak; Winther, Ole; Fleury, Bernard Henri

Published in:
Information Theory Workshop (ITW), 2014 IEEE

DOI (link to publication from Publisher):
10.1109/ITW.2014.6970819

Publication date:
2014

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
S-AMP: Approximate Message Passing for General Matrix Ensembles

Burak Çakmak (AAU)
Ole Winther (DTU) and Bernard H. Fleury (AAU)

Technical University of Denmark (DTU) and Aalborg University (AAU)

02 November 2014, Hobart, Tasmania
Motivation

• **Low complexity** and near optimal inference algorithms for linear observation models

\[
y = A x + \epsilon, \quad \epsilon \sim \mathcal{N}(\epsilon|0, \sigma^2 I), \quad x \sim \prod_{k=1}^{K} p_k(x_k)
\]

• \(A: N \times K\), known and **drawn from known matrix ensemble**

• \(N, K \gg 1\)
Loopy BP

\[ p(y_1 | (Ax)_1) \quad p(y_2 | (Ax)_2) \quad p(y_3 | (Ax)_3) \]

\[ m_{n \rightarrow k}(x_k) = \int p(y_n | (Ax)_n) \prod_{l \neq k} m_{l \rightarrow n}(x_l) \, dx_l \]

\[ m_{l \rightarrow n}(x_l) \approx p(x_l) \prod_{m \neq n} m_{m \rightarrow l}(x_l) \]
**Local Cavity Argument:**

\[
h_{n,k} = \sum_{l \neq k} A_{ nl } x_l, \quad x_l \sim m_{l\rightarrow n}(x_l)
\]

- Due to CLT, \( h_{n,k} \) is approximated by Gaussian.
- This leads Loop BP to Loopy EP

\[
m_{n\rightarrow k}(x_k) = \int p(y_n|(Ax)_n) \prod_{l \neq k} m_{l\rightarrow n}(x_l) dx_l
\]

\[
m_{l\rightarrow n}(x_l) \cong p(x_l) \prod_{m \neq n} m_{m\rightarrow l}(x_l)
\]
Define $q_k(x_k) \sim p_k(x_k) \prod_{n \in N} m_n \rightarrow k(x_k)$

Let $\tilde{q}_k(x_k) \equiv N(x_k | \mu_k, \sigma_k^2)$ such that

$$
\mu_k = \mathbb{E}[x_k | q(x_k)] \\
\sigma_k^2 = \text{Var}[x_k | q(x_k)]
$$

Then loopy EP update rule is

$$
m_{n \rightarrow k}(x_k) = \int p(y_n | Ax)_n \prod_{l \in K \setminus k} m_{l \rightarrow n}(x_l) dx_l
$$

$$
m_{l \rightarrow n}(x_l) = \exp \left(-\frac{\lambda_{l \rightarrow n}}{2} x_l^2 + \gamma_{l \rightarrow n} x_l \right)
$$
Loopy EP

Define

\[ q_k(x_k) \approx p_k(x_k) \prod_{n \in N} m_{n \to k}(x_k) \]

Let \( \tilde{q}_k(x_k) \triangleq N(x_k | \mu_k, \sigma_k^2) \) such that

\[ \mu_k = \mathbb{E}[x_k | q(x_k)] \]
\[ \sigma_k^2 = \text{Var}[x_k | q(x_k)] \]

Then loopy EP update rule is

\[ m_{n \to k}(x_k) = \int p(y_n|(Ax)_n) \prod_{l \in K \setminus k} m_{l \to n}(x_l) \, dx_l \]

\[ m_{l \to n}(x_l) = \exp \left( -\frac{\lambda_{l \to n}}{2} x_l^2 + \gamma_{l \to n} x_l \right) \]

\[ m_{k \to n}(x_k) = \frac{\tilde{q}_k(x_k)}{\prod_{n \in N} m_{n \to k}(x_k)} \]
Approximate message passing (AMP) [Donoho et al 2009]

- Assume $A_{nk}$ zero mean-iid, $\overline{A_{nk}^2} = 1/N$, $N, K \to \infty$, $\alpha \equiv N/K$ finite
- Reduces the number of messages to $N + K$ means.

\[
\mu^{t+1} = \eta_t \left( A^T z^t + \mu^t \right)
\]
\[
z^t = y - A\mu^t + \frac{1}{\alpha} \left\langle \eta_{t-1}' \left( A^T z^{t-1} + \mu^{t-1} \right) \right\rangle z^{t-1}
\]

with $\langle u \rangle \triangleq \sum_{k=1}^{K} u_k / K$.

- $\eta_t(\kappa_k)$ and $\eta_t'(\kappa_k)/\tau$ are (in some cases) the mean and variance of (Krzakala et al 2012)

\[
q_k(x_k) \cong p_k(x_k)N(x_k|\kappa_k, 1/\tau)
\]

- Return to $\tau$ when discussing EP and S-AMP
EP [Minka 2001], [Opper and Winther 2000]

Define

\[ q_k(x_k) \approx p_k(x_k) m_{N \to k}(x_k) \]

Let \( \tilde{q}_k(x_k) = N(x_k | \mu_k, \sigma_k^2) \) such that

\[ \mu_k = \mathbb{E}[x_k | q_k(x_k)] \]
\[ \sigma_k^2 = \mathbb{V}ar[x_k | q(x_k)] \]

Then EP update rule is

\[ m_{N \to k}(x_k) = \int p(y | Ax) \prod_{l \in \mathcal{K} \setminus k} m_{l \to N}(x_l) \, dx_l \]

\[ m_{k \to N}(x_k) = \exp \left( -\frac{\Lambda_{kk}}{2} x_k^2 + \gamma_k x_k \right) \]

\[ m_{k \to N}(x_k) = \frac{\tilde{q}_k(x_k)}{m_{N \to k}(x_k)} \]
S-AMP

• generalizes AMP for arbitrary (orthogonally invariant) matrix ensembles.

\[ \mu^{t+1} = \eta_t (A^\dagger z^t + \mu^t) \]

\[ z^t = y - A\mu^t + \left(1 - \frac{1}{s_{A}^{t-1}}\right) z^{t-1} \]

\[ s_{A}^{t-1} \triangleq S_A \left(-\langle \eta_{t-1}'(A^\dagger z^{t-1} + \mu^{t-1}) \rangle\right) \]

\( S_A \) denotes the S-transform (in free probability theory) of the limiting eigenvalue distribution (LED) of \( A^\dagger A \).

• Indeed when the entries of \( A \) be iid with zero mean variance \( 1/N \):

\[ S_A(\omega) = \frac{1}{1 + \omega/\alpha} \]

which yields AMP iteration steps.
EP→S-AMP: Start with EP Update Rule

\[ p(y|Ax) \]

\[ N \]

\[ m_{N \rightarrow k}(x_k) = \int p(y|Ax) \prod_{l \in K \setminus k} m_{l \rightarrow N}(x_l) \, dx_l \]

\[ m_{k \rightarrow N}(x_k) = \exp \left( -\frac{\Lambda_{kk}}{2} x_k^2 + \gamma_k x_k \right) \]

\[ q_k(x_k) \sim p_k(x_k) m_{N \rightarrow k}(x_k) \]

\[ m_{N \rightarrow k}(x_k) = \tilde{q}_k(x_k) / m_{k \rightarrow N}(x_k) \]

\[ \bullet \text{ Let } J = A^\dagger A / \sigma^2 \text{ and } \theta = A^\dagger y / \sigma^2. \text{ Define} \]

\[ \Sigma = (\Lambda + J)^{-1} \quad \mu = \Sigma(\theta + \gamma) \]

\[ \bullet \text{ Then we have} \]

\[ m_{N \rightarrow k}(x_k) = \exp \left\{ -\frac{1}{2} \left( \frac{1}{\Sigma_{kk}} - \Lambda_{kk} \right) x_k^2 + \left( \frac{\mu_k}{\Sigma_{kk}} - \gamma_k \right) x_k \right\} \]
Let $\tau_k = \frac{1}{\Sigma_{kk}} - \Lambda_{kk}$ and $\kappa_k = (\frac{\mu_k}{\Sigma_{kk}} - \gamma_k)/\tau_k$.

Hence we can write

$$m_{\mathcal{N} \rightarrow k}(x_k) \approx \mathcal{N}(x_k | \kappa_k, 1/\tau_k)$$

Write $q_k(x_k)$ in the form of

$$q_k(x_k) = \frac{p_k(x_k)\mathcal{N}(x_k | \kappa_k, 1/\tau_k)}{Z(\kappa_k, \tau_k)}$$

Define

$$\eta(\kappa_k; \tau_k) \triangleq \kappa_k + \frac{1}{\tau_k} \frac{\partial \log Z(\kappa_k, \tau_k)}{\partial \kappa_k}$$

and

$$\eta'(\kappa_k; \tau_k) \triangleq \frac{\partial \eta(\kappa_k; \tau_k)}{\partial \kappa_k}$$

where $\eta(\kappa_k; \tau_k)$ and $\eta'(\kappa_k; \tau_k)/\tau_k$ are respectively the mean and the variance of $q_k(x_k)$ [Krzakala et al. 2012].
EP $\rightarrow$ S-AMP: Move to ADATAP [Opper and Winther 2001]

• Note that

$\mu = (\Lambda + J)^{-1}(\gamma + \theta) \iff (\Lambda + J)\mu = \gamma + \theta$
• Note that

\[ \mu = (\Lambda + J)^{-1}(\gamma + \theta) \iff (\Lambda + J)\mu = \gamma + \theta \]

• Putting everything together leads EP to

\[ \mu_k = \eta(\kappa_k; \tau_k) \]

\[ \kappa_k = \frac{1}{\tau_k \sigma^2} \sum_{n \in \mathcal{N}} A_{nk} \left( y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l \right) + \mu_k \]

\[ \tau_k = \frac{1}{\sum_{kk} - \Lambda_{kk}}, \quad \Lambda_{kk} = \frac{\tau_k}{\eta'(\kappa_k; \tau_k)} - \tau_k \]

exactly coincides ADATAP for the linear observation models.
Define

$$z_{n,k} \triangleq \frac{1}{\tau_k \sigma^2} \left( y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l \right).$$

Using this definition we "devise" the following identity:

$$z_{n,k} = y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l + (1 - \sigma^2 \tau_k) z_{n,k}.$$
EP $\rightarrow$ S-AMP: Apply Adaptive Damping

Define
\[ z_{n,k} \triangleq \frac{1}{\tau_k \sigma^2} \left( y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l \right). \]

Using this definition we “devise” the following identity:
\[ z_{n,k} = y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l + (1 - \sigma^2 \tau_k) z_{n,k}. \]

Doing so leads to
\[ \mu_k = \eta \left( \sum_{n \in \mathcal{N}} A_{nk} z_{n,k} + \mu_k ; \tau_k \right) \]
\[ z_{n,k} = y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l + (1 - \sigma^2 \tau_k) z_{n,k} \]
\[ \tau_k = \frac{1}{\sum_{kk} - \Lambda_{kk}}, \quad \Lambda_{kk} = \frac{\tau_k}{\eta' (\kappa_k ; \tau_k)} - \tau_k \]

This equations can be thought as a finite size interpretation of AMP.

• We can recover self-averaging matrix ensembles \( \tau_k \rightarrow \tau \):

\[
\Sigma_{kk} = \left[ (\Lambda + J)^{-1} \right]_{kk} = \frac{\partial}{\partial \Lambda_{kk}} \ln \det(\Lambda + J)
\]

• by using

\[
\frac{1}{K} \ln \det(\Lambda + J) \rightarrow \frac{1}{K} \mathbb{E}_J [\ln \det(\Lambda + J)] \quad \text{for} \quad K \rightarrow \infty
\]

- We can recover self-averaging matrix ensembles $\tau_k \rightarrow \tau$:

$$\Sigma_{kk} = [(\Lambda + J)^{-1}]_{kk} = \frac{\partial}{\partial \Lambda_{kk}} \ln \det (\Lambda + J)$$

- by using

$$\frac{1}{K} \ln \det (\Lambda + J) \rightarrow \frac{1}{K} \mathbb{E}_J [\ln \det (\Lambda + J)] \quad \text{for} \quad K \rightarrow \infty$$

- Doing so leads $\tau_k$ to $\tau$ that is the solution of

$$\sigma^2 \tau = \frac{1}{\sigma^2} R_A \left( - \frac{\langle \eta'(A^\dagger z + \mu; \tau) \rangle}{\sigma^2 \tau} \right)$$

$R_A$ is the R-transform (in free probability theory) of the LED of $A^\dagger A$ and

$$z = y - A\mu + (1 - \sigma^2 \tau)z$$
EP → S-AMP: Move to S-transform

• Recall that
  \[ \sigma^2 \tau = R_A \left( \frac{- < \eta' (A^\dagger z + \mu; \tau)}{\sigma^2 \tau} \right) \]

• By invoking the fact [Haagerup and Larsen 2001]
  \[ S_A(\omega) = \frac{1}{R_A(\omega S_A(\omega))} \]

• we have
  \[ \sigma^2 \tau = \frac{1}{S_A(- < \eta' (A^\dagger z + \mu; \tau) >)} \]
EP $\rightarrow$ S-AMP: Move to $S$-transform

- Recall that
  \[
  \sigma^2 \tau = R_A \left( \frac{-< \eta' (A^\dagger z + \mu; \tau)}{\sigma^2 \tau} \right)
  \]

- By invoking the fact [Haagerup and Larsen 2001]
  \[
  S_A(\omega) = \frac{1}{R_A(\omega S_A(\omega))}
  \]

- we have
  \[
  \sigma^2 \tau = \frac{1}{S_A(-< \eta' (A^\dagger z + \mu; \tau)>)}
  \]

- this completes the mapping at "fixed points":
  \[
  \mu = \eta (A^\dagger z + \mu; \tau)
  \]
  \[
  z = y - A \mu + \left(1 - \frac{1}{S_A}\right)z
  \]
  \[
  s_A = S_A \left( -< \eta' (A^\dagger z + \mu; \tau) > \right)
  \]
What is S-AMP?

• In summary

\[
\begin{align*}
\mu^{t+1} &= \eta_t (A^\dagger z^t + \mu^t) \\
z^t &= y - A\mu^t + \left(1 - \frac{1}{S_{A}^{t-1}}\right) z^{t-1} \\
S_{A}^{t-1} &= S_{A} (\langle \eta'_t (A^\dagger z^{t-1} + \mu^{t-1}) \rangle)
\end{align*}
\]

where \( \eta_t(x^t) = \eta(x^t; \tau^t) \) and

\[
\tau^t = \frac{1}{\sigma^2 S_A \langle - \langle \eta' (A^\dagger z^t + \mu^t; \tau^t) \rangle \rangle}
\]

• Oops, S-AMP includes a hard fixed point equation.
• As a matter of fact we don’t know what is the best update rule for \( \tau^t \)
A Variant of S-AMP

- By making analogy with the state evolution formula [Bayati and Montari 2011]

\[
\begin{align*}
\mu^{t+1} &= \eta \left( A^\dagger z^t + \mu^t; \tilde{\tau}^t \right) \\
z^t &= y - A \mu^t + \left( 1 - \frac{1}{s_{A}^{t-1}} \right) z^{t-1} \\
s_{A}^{t-1} &\triangleq S_A \left(- \left\langle \eta' \left( A^\dagger z^{t-1} + \mu^{t-1}; \tilde{\tau}^{t-1} \right) \right\rangle \right)
\end{align*}
\]

where \( \tilde{\tau}^t \) is updated by using the solution

\[
\tilde{\tau}^t = \frac{1}{\sigma^2 S_A \left( - \frac{\tilde{\tau}^t}{\tilde{\tau}^{t-1}} \left\langle \eta' \left( A^\dagger z^{t-1} + \mu^{t-1}; \tilde{\tau}^{t-1} \right) \right\rangle \right)}
\]
A Variant of S-AMP

- By making analogy with the state evolution formula [Bayati and Montari 2011]

\[
\mu^{t+1} = \eta \left( A^\dagger z^t + \mu^t; \tilde{\tau}^t \right)
\]

\[
z^t = y - A \mu^t + \left( 1 - \frac{1}{s_{A}^{t-1}} \right) z^{t-1}
\]

\[
s_{A}^{t-1} \triangleq S_A \left( -\left< \eta' \left( A^\dagger z^{t-1} + \mu^{t-1}; \tilde{\tau}^{t-1} \right) \right> \right)
\]

where \( \tilde{\tau}^t \) is updated by using the solution

\[
\tilde{\tau}^t = \frac{1}{\sigma^2 S_A \left( -\frac{\tilde{\tau}^t}{\tilde{\tau}^{t-1}} \left< \eta' \left( A^\dagger z^{t-1} + \mu^{t-1}; \tilde{\tau}^{t-1} \right) \right> \right)}
\]

- i.e.

\[
\tilde{\tau}^t = \frac{1}{\sigma^2 R_A} \left( -\frac{\left< \eta' \left( A^\dagger z^{t-1} + \mu^{t-1}; \tilde{\tau}^{t-1} \right) \right>}{\sigma^2 \tilde{\tau}^{t-1}} \right)
\]
Application: Row Orthogonal Ensembles in Compressed Sensing

• A random row orthogonal ensemble defined as

\[ A = \alpha^{-\frac{1}{2}} P_{\alpha} O, \quad \alpha \leq 1 \]

where \( P_{\alpha} \) is the \( N \times K \) matrix with entries \( (P_{\alpha})_{ij} = \delta_{ij}, \forall ij \).

• In this case we have

\[ S_{A}(z) = \frac{1 + z}{1 + z/\alpha} \]

\[ R_{A}(z) = \frac{z - \alpha + \sqrt{(\alpha - z)^2 + 4\alpha^2z}}{2\alpha z} \]
Simulation Results

- Let \( p_k(x_k) = (1 - \rho)\delta(x_k) + \rho N(x_k | 0, 1) \), with \( \rho \in (0, 1) \).
- For the closed-forms of \( \eta_t(\cdot) \) and \( \eta'_t(\cdot) \), see [Krzakala et.al. 2012].

- S-AMP for the row orthogonal matrix ensemble (solid curves) and the iid zero-mean ensemble (dashed curves).
- Confidence intervals (CIs) are also shown for \( \alpha = 1/3 \).
- We set \( \sigma^2 = -20 \text{ dB} \) and \( \rho = 0.1 \), and \( K = 1200 \).
- The numbers in the plot are the predictions of replica theory [Kabashima and Vekapera 2014].