Aalborg Universitet

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):
Cakmak, B., Winther, O., & Fleury, B. H. (2014). S-AMP: Approximate Message Passing for General Matrix Ensembles. In *Information Theory Workshop (ITW), 2014 IEEE* (pp. 192 - 196). IEEE Press. https://doi.org/10.1109/ITW.2014.6970819

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 20, 2024

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

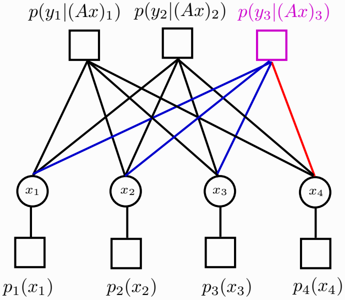
Motivation

- Low complexity and near optimal inference algorithms for
- linear observation models

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

- $\boldsymbol{A}$: $N \times K$, known and drawn from known matrix ensemble
- $N, K \gg 1$

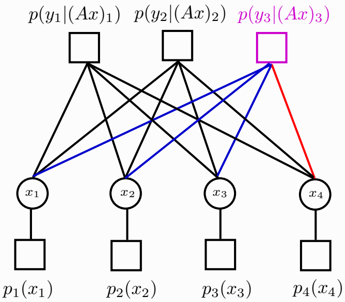
Loopy BP



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

$$m_{l \to n}(x_l) \cong p(x_l) \prod_{m \neq n} m_{m \to l}(x_l)$$

Loopy BP



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

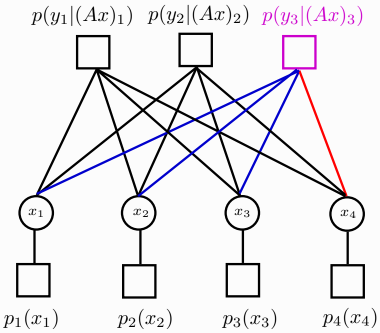
$$m_{l\to n}(x_l) \cong p(x_l) \prod_{m\neq n} m_{m\to l}(x_l)$$

- Local Cavity Argument:

$$h_{n,\setminus k} = \sum_{l\neq k} A_{nl} x_l, \quad x_l \sim m_{l\to n}(x_l)$$

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

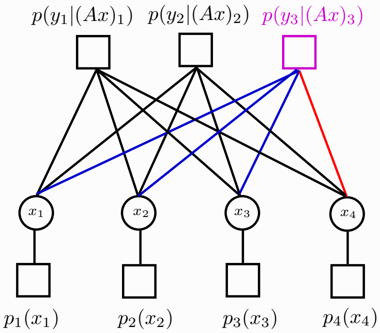
Loopy EP



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

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

Loopy EP



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

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

Define

$$q_k(x_k) \cong p_k(x_k) \prod_{n\in\mathcal{N}} m_{n\to k}(x_k)$$

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

$$\mu_k = \mathbb{E}[x_k|q(x_k)]$$

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

Then loopy EP update rule is

$$m_{k\to n}(x_k) = \frac{\tilde{q}_k(x_k)}{\prod_{n\in\mathcal{N}} m_{n\to k}(x_k)}$$

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.

$$\boldsymbol{\mu}^{t+1} = \eta_t \left(\boldsymbol{A}^T \boldsymbol{z}^t + \boldsymbol{\mu}^t \right)$$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}^t + \frac{1}{\alpha} \left\langle \eta'_{t-1}(\boldsymbol{A}^T \boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}) \right\rangle \boldsymbol{z}^{t-1}$$

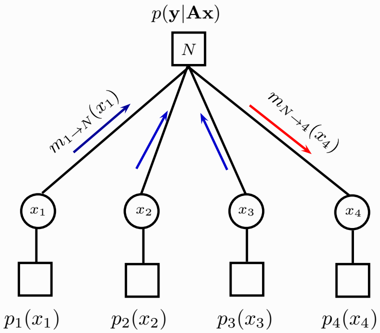
with $\langle \boldsymbol{u} \rangle \triangleq \sum_{k=1}^K u_k / K$.

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

$$q_k(x_k) \cong p_k(x_k) \mathcal{N}(x_k | \kappa_k, 1/\tau)$$

- Return to τ when discussing EP and S-AMP

EP [Minka 2001], [Opper and Winther 2000]



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

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

Define

$$q_k(x_k) \cong p_k(x_k) m_{N\to k}(x_k)$$

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

$$m_{k\to N}(x_k) = \frac{\tilde{q}_k(x_k)}{m_{N\to k}(x_k)}$$

S-AMP

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

$$\boldsymbol{\mu}^{t+1} = \eta_t \left(\boldsymbol{A}^\dagger \boldsymbol{z}^t + \boldsymbol{\mu}^t \right)$$
$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}^t + \left(1 - \frac{1}{\mathrm{s}_{\boldsymbol{A}}^{t-1}} \right) \boldsymbol{z}^{t-1}$$
$$\mathrm{s}_{\boldsymbol{A}}^{t-1} \triangleq \mathrm{S}_{\boldsymbol{A}} \left(- \left\langle \eta'_{t-1}(\boldsymbol{A}^\dagger \boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}) \right\rangle \right)$$

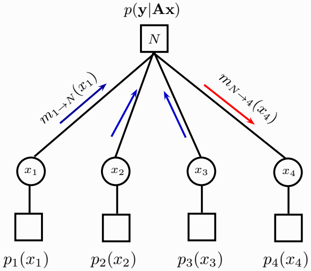
  $\mathrm{S}_{\boldsymbol{A}}$ denotes the S-transform (in free probability theory) of the limiting eigenvalue distribution (LED) of $\boldsymbol{A}^\dagger \boldsymbol{A}$.

- Indeed when the entries of $\boldsymbol{A}$ be iid with zero mean variance $1/N$:

$$\mathrm{S}_{\boldsymbol{A}}(\omega) = \frac{1}{1 + \omega/\alpha}$$

  which yields AMP iteration steps.

EP→S-AMP: Start with EP Update Rule



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

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

$$q_k(x_k) \cong p_k(x_k)m_{N\to k}(x_k)$$

$$m_{N\to k}(x_k) = \tilde{q}_k(x_k)/m_{k\to N}(x_k)$$

- Let $\boldsymbol{J} = \boldsymbol{A}^\dagger\boldsymbol{A}/\sigma^2$ and $\boldsymbol{\theta} = \boldsymbol{A}^\dagger\boldsymbol{y}/\sigma^2$. Define

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

- Then we have

$$m_{N\to k}(x_k) = \exp\left\{-\frac{1}{2}\left(\frac{1}{\Sigma_{kk}} - \Lambda_{kk}\right)x_k^2 + \left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right)x_k\right\}$$

EP $\rightarrow$ S-AMP: Use AMP Notations

- 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

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

- Write $q_k(x_k)$ in the form of

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

- Define

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

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

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 → S-AMP: Move to ADATAP [Opper and Winther 2001]

- Note that

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

EP $\rightarrow$ S-AMP: Move to ADATAP [Opper and Winther 2001]

- Note that

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

- Putting everything together leads EP to

$$\mu_k = \eta(\kappa_k; \tau_k)$$
$$\kappa_k = \frac{1}{\tau_k \sigma^2} \sum_{n \in \mathcal{N}} A_{nk} \left(y_n - \sum_{l \in \mathcal{K}} A_{nl} \mu_l \right) + \mu_k$$
$$\tau_k = \frac{1}{\Sigma_{kk}} - \Lambda_{kk}, \quad \Lambda_{kk} = \frac{\tau_k}{\eta'(\kappa_k; \tau_k)} - \tau_k$$

exactly coincides ADATAP for the linear observation models.

EP→ 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:

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

EP→ 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

$$\mu_k = \eta\left(\sum_{n \in \mathcal{N}} A_{nk}z_{n,k} + \mu_k; \tau_k\right)$$

$$z_{n,k} = y_n - \sum_{l \in \mathcal{K}} A_{nl}\mu_l + (1 - \sigma^2\tau_k)z_{n,k}$$

$$\tau_k = \frac{1}{\Sigma_{kk}} - \Lambda_{kk}, \quad \Lambda_{kk} = \frac{\tau_k}{\eta'(\kappa_k; \tau_k)} - \tau_k$$

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

EP$\rightarrow$ S-AMP: Invoke Self Averaging Ansatz [Opper and Winther 2001]

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

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

- by using

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

EP→ S-AMP: Invoke Self Averaging Ansatz [Opper and Winther 2001]

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

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

- by using

$$\frac{1}{K} \ln \det(\boldsymbol{\Lambda} + \boldsymbol{J}) \to \frac{1}{K} \mathbb{E}_{\boldsymbol{J}} \left[\ln \det(\boldsymbol{\Lambda} + \boldsymbol{J})\right] \quad \text{for} \quad K \to \infty$$

- Doing so leads τ_k to τ that is the solution of

$$\sigma^2 \tau = \frac{1}{\sigma^2} \mathrm{R}_{\boldsymbol{A}} \left(- \frac{< \eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu}; \tau) >}{\sigma^2 \tau} \right)$$

$\mathrm{R}_{\boldsymbol{A}}$ is the R-transform (in free probability theory) of the LED of $\boldsymbol{A}^\dagger \boldsymbol{A}$ and

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} + (1 - \sigma^2 \tau)\boldsymbol{z}$$

EP$\to$ S-AMP: Move to S-transform

- Recall that

$$\sigma^2\tau = \mathrm{R}_{\boldsymbol{A}}\left(\frac{- < \eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu}; \tau) >}{\sigma^2\tau}\right)$$

- By invoking the fact [Haagerup and Larsen 2001]

$$\mathrm{S}_{\boldsymbol{A}}(\omega) = \frac{1}{\mathrm{R}_{\boldsymbol{A}}(\omega \mathrm{S}_{\boldsymbol{A}}(\omega))}$$

- we have

$$\sigma^2\tau = \frac{1}{\mathrm{S}_{\boldsymbol{A}}(- < \eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu}; \tau) >)}$$

EP→ S-AMP: Move to S-transform

- Recall that

$$\sigma^2\tau = \mathrm{R}_{\boldsymbol{A}}\left(\frac{-<\eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu};\tau)>}{\sigma^2\tau}\right)$$

- By invoking the fact [Haagerup and Larsen 2001]

$$\mathrm{S}_{\boldsymbol{A}}(\omega) = \frac{1}{\mathrm{R}_{\boldsymbol{A}}(\omega \mathrm{S}_{\boldsymbol{A}}(\omega))}$$

- we have

$$\sigma^2\tau = \frac{1}{\mathrm{S}_{\boldsymbol{A}}(-<\eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu};\tau)>)}$$

- this completes the mapping at "fixed points":

$$\boldsymbol{\mu} = \eta\left(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu};\tau\right)$$

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} + (1 - \frac{1}{\mathrm{s}_{\boldsymbol{A}}})\boldsymbol{z}$$

$$\mathrm{s}_{\boldsymbol{A}} = \mathrm{S}_{\boldsymbol{A}}\left(-\left\langle \eta'(\boldsymbol{A}^\dagger \boldsymbol{z} + \boldsymbol{\mu};\tau)\right\rangle\right)$$

What is S-AMP?

- In summary

$$\boldsymbol{\mu}^{t+1} = \eta_t\left(\boldsymbol{A}^\dagger \boldsymbol{z}^t + \boldsymbol{\mu}^t\right)$$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}^t + \left(1 - \frac{1}{\mathrm{s}_{\boldsymbol{A}}^{t-1}}\right)\boldsymbol{z}^{t-1}$$

$$\mathrm{s}_{\boldsymbol{A}}^{t-1} \triangleq \mathrm{S}_{\boldsymbol{A}^\dagger \boldsymbol{A}}\left(-\left\langle \eta'_{t-1}(\boldsymbol{A}^\dagger \boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1})\right\rangle\right)$$

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

$$\tau^t = \frac{1}{\sigma^2 \mathrm{S}_{\boldsymbol{A}}\left(-\left\langle \eta'(\boldsymbol{A}^\dagger \boldsymbol{z}^t + \boldsymbol{\mu}^t; \tau^t)\right\rangle\right)}$$

- 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 τ^t

A Variant of S-AMP

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

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

$$\boldsymbol{z}^{t} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}^{t} + \left(1 - \frac{1}{\mathrm{s}_{\boldsymbol{A}}^{t-1}}\right)\boldsymbol{z}^{t-1}$$

$$\mathrm{s}_{\boldsymbol{A}}^{t-1} \triangleq \mathrm{S}_{\boldsymbol{A}}\left(-\left\langle \eta'(\boldsymbol{A}^{\dagger}\boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}; \tilde{\tau}^{t-1})\right\rangle\right)$$

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

$$\tilde{\tau}^{t} = \frac{1}{\sigma^2 \mathrm{S}_{\boldsymbol{A}}\left(-\frac{\tilde{\tau}^{t}}{\tilde{\tau}^{t-1}}\left\langle \eta'(\boldsymbol{A}^{\dagger}\boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}; \tilde{\tau}^{t-1}\right\rangle\right)}$$

A Variant of S-AMP

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

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

$$\boldsymbol{z}^{t} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}^{t} + \left(1 - \frac{1}{\mathrm{s}_{\boldsymbol{A}}^{t-1}}\right)\boldsymbol{z}^{t-1}$$

$$\mathrm{s}_{\boldsymbol{A}}^{t-1} \triangleq \mathrm{S}_{\boldsymbol{A}}\left(-\left\langle \eta'(\boldsymbol{A}^{\dagger}\boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}; \tilde{\tau}^{t-1})\right\rangle\right)$$

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

$$\tilde{\tau}^{t} = \frac{1}{\sigma^{2}\mathrm{S}_{\boldsymbol{A}}\left(-\frac{\tilde{\tau}^{t}}{\tilde{\tau}^{t-1}}\left\langle \eta'(\boldsymbol{A}^{\dagger}\boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}; \tilde{\tau}^{t-1})\right\rangle\right)}$$

- i.e.

$$\tilde{\tau}^{t} = \frac{1}{\sigma^{2}}\mathrm{R}_{\boldsymbol{A}}\left(-\frac{\left\langle \eta'(\boldsymbol{A}^{\dagger}\boldsymbol{z}^{t-1} + \boldsymbol{\mu}^{t-1}; \tilde{\tau}^{t-1})\right\rangle}{\sigma^{2}\tilde{\tau}^{t-1}}\right)$$

Application: Row Orthogonal Ensembles in Compressed Sensing

- A random row orthogonal ensemble defined as

$$\boldsymbol{A} = \alpha^{-\frac{1}{2}} \boldsymbol{P}_\alpha \boldsymbol{O}, \quad \alpha \leq 1$$

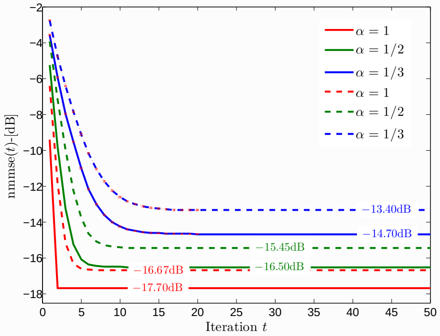
  where $\boldsymbol{P}_\alpha$ is the $N \times K$ matrix with entries $[\boldsymbol{P}_\alpha]_{ij} = \delta_{ij}, \forall ij$.
- In this case we have

$$\mathrm{S}_{\boldsymbol{A}}(z) = \frac{1+z}{1+z/\alpha}$$

$$\mathrm{R}_{\boldsymbol{A}}(z) = \frac{z - \alpha + \sqrt{(\alpha - z)^2 + 4\alpha^2 z}}{2\alpha z}$$

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 $\sigma^2 = -20$ dB and $\rho = 0.1$, and $K = 1200$.
- The numbers in the plot are the predictions of replica theory [Kabashima and Vekapera 2014]