

Self-Averaging Expectation Propagation

Cakmak, Burak; Oppel, Manfred ; Fleury, Bernard Henri; Winther, Ole

Publication date:
2016

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Cakmak, B., Oppel, M., Fleury, B. H., & Winther, O. (2016). *Self-Averaging Expectation Propagation*. Poster presented at Advances in Approximate Bayesian Inference .
<http://approximateinference.org/accepted/CakmakEtAl2016.pdf>

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.

Self-Averaging Expectation Propagation

Burak Çakmak, Manfred Oppel, Bernard H. Fleury, and Ole Winther

Joint work of Aalborg Universitet, Technische Universität Berlin and Tekniske Universitet Denmark

Problem and Objective

Recover signal \mathbf{x} from the observation \mathbf{y} where

$$\mathbf{x} \rightarrow \mathbf{A}\mathbf{x} \rightarrow \mathbf{y}$$

For example:

- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$
- $\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x})$

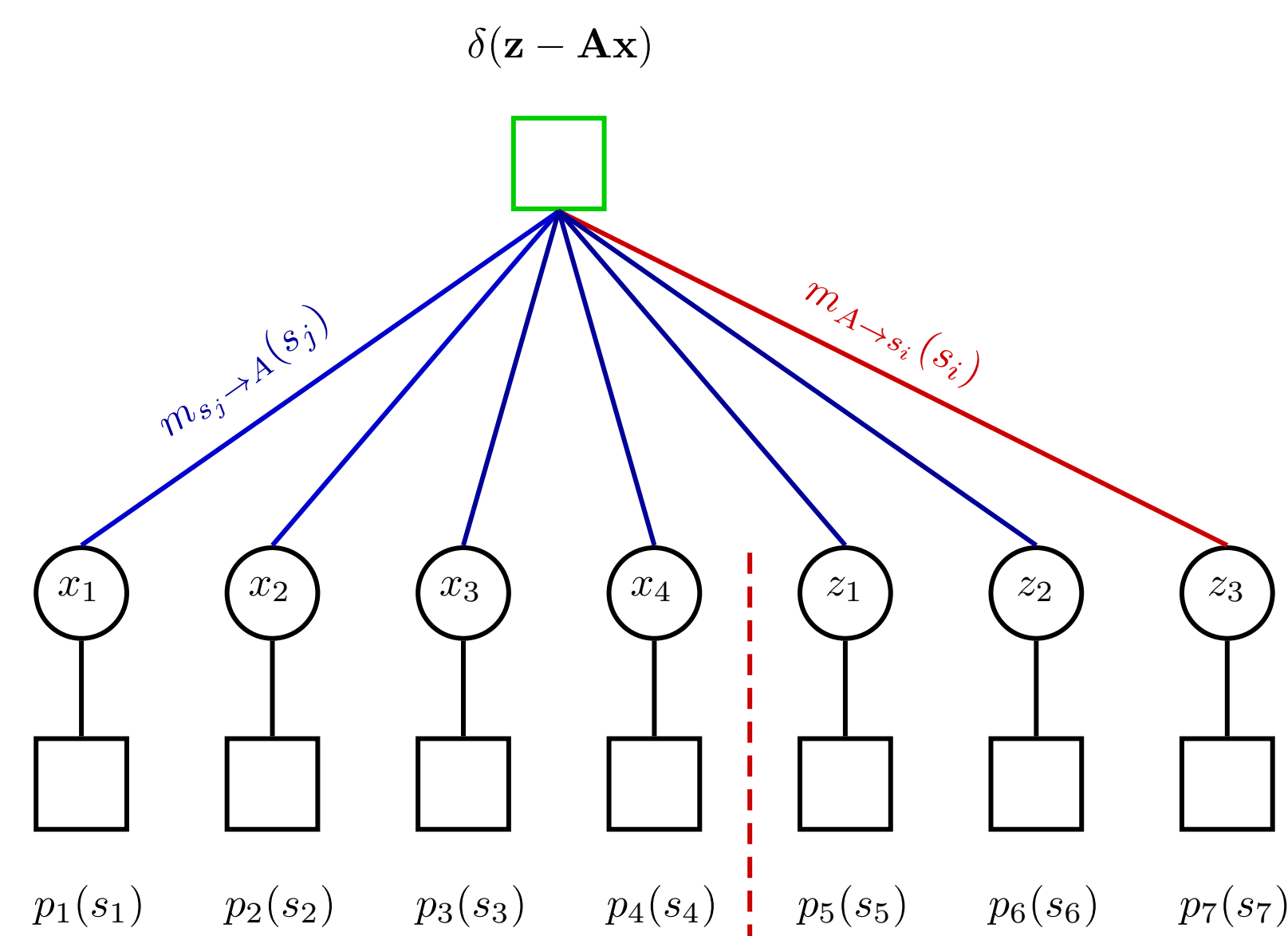
Assume that

- \mathbf{A} is drawn from a known ensemble
- The dimensions of \mathbf{A} are LARGE!

Obtain iterative estimation algorithms with

- Low computational complexity
- Good accuracy

Expectation Propagation (EP)



$$p(\mathbf{s}) \triangleq p(\mathbf{x})p(\mathbf{y}|\mathbf{z}) \quad \text{with} \quad \mathbf{s} \triangleq (\mathbf{x}, \mathbf{z})$$

$$m_{s_j \rightarrow A}(s_j) = \exp\left(-\frac{1}{2}\Lambda_{jj}s_j^2 + \gamma_j s_j\right)$$

$$m_{A \rightarrow s_i}(s_i) = \int \delta(\mathbf{z} - \mathbf{A}\mathbf{x}) \prod_{j \neq i} m_{s_j \rightarrow A}(s_j) d\mathbf{s}_j$$

$$\propto \exp\left(-\frac{1}{2}\mathbf{V}_{ii}s_i^2 + \rho_i s_i\right)$$

The two pdfs

$$q_i(s_i) \propto p_i(s_i) m_{A \rightarrow s_i}(s_i)$$

$$\tilde{q}_i(s_i) \propto m_{s_i \rightarrow A}(s_i) m_{A \rightarrow s_i}(s_i)$$

are consistent in the first- and second-moment:

$$\langle (s_i, s_i^2) \rangle_{q_i(s_i)} = \langle (s_i, s_i^2) \rangle_{\tilde{q}_i(s_i)}.$$

This work was partially supported by the research project VIRTU-OSO (funded by Intel Mobile Communications, Anite, Telenor, Aalborg University, and the Danish National Advanced Technology Foundation).

The Essence of the Issue: "Cavity Variances"

The update of the so-called cavity variances require matrix inversions.

- The exact posterior pdf of $\mathbf{s} = (\mathbf{x}, \mathbf{z})$ is given by

$$p(\mathbf{s}|\mathbf{y}, \mathbf{A}) \propto p(\mathbf{s})\delta(\mathbf{z} - \mathbf{A}\mathbf{x}) \quad \text{with} \quad p(\mathbf{s}) \triangleq p(\mathbf{x})p(\mathbf{y}|\mathbf{z}).$$

- EP approximates the exact posterior pdf in the form of

$$q(\mathbf{s}) \propto p(\mathbf{s}) \exp\left(-\frac{1}{2}\mathbf{s}^\dagger \mathbf{V} \mathbf{s} + \boldsymbol{\rho}^\dagger \mathbf{s}\right) \quad \text{with} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_z \end{pmatrix}$$

where $\{\mathbf{V}_{ii}\}$ are called cavity variances.

- The equations of $\boldsymbol{\rho} = (\boldsymbol{\rho}_x, \boldsymbol{\rho}_z)$ can be expressed by

$$\boldsymbol{\rho}_x = \mathbf{A}(\mathbf{V}_z \boldsymbol{\eta}_z - \boldsymbol{\rho}_z) + \mathbf{V}_x \boldsymbol{\eta}_x \quad \text{with} \quad (\boldsymbol{\eta}_x, \boldsymbol{\eta}_z) = \langle (\mathbf{x}, \mathbf{z}) \rangle_{q(\mathbf{x}, \mathbf{z})} = (\boldsymbol{\eta}_x, \mathbf{A}\boldsymbol{\eta}_x).$$

- The equations of cavity variances $\{\mathbf{V}_{ii}\}$ are

$$\chi_i = \frac{1}{\Lambda_{ii} + \mathbf{V}_{ii}} = \begin{cases} [(\Lambda_x + \mathbf{A}^\dagger \Lambda_z \mathbf{A})^{-1}]_{ii} & \Lambda_{ii} = [\Lambda_x]_{ii} \\ [\mathbf{A}(\Lambda_x + \mathbf{A}^\dagger \Lambda_z \mathbf{A})^{-1} \mathbf{A}^\dagger]_{jj} & \Lambda_{ii} = [\Lambda_z]_{jj} \end{cases}$$

where $\boldsymbol{\chi} \triangleq (\boldsymbol{\chi}_x, \boldsymbol{\chi}_z)$ is the variance of $q(\mathbf{x}, \mathbf{z})$.

- EP is accurate but has $O(K^3)$ computational complexity (per iteration) due to the update of cavity variances.

Self-Averaging Cavity Variances

Asymptotic freeness transforms the large-system challenges into opportunities.

- We use the concept of asymptotic freeness from random matrix theory to show that EP cavity variances are self-averaging.
- Specifically, $\mathbf{V}_x \simeq v_x \mathbf{I}$ and $\mathbf{V}_z \simeq v_z \mathbf{I}$ where

$$v_x = \frac{\alpha(1 - v_z \langle \boldsymbol{\chi}_z \rangle)}{\langle \boldsymbol{\chi}_x \rangle} \quad \& \quad v_z = \lambda_x S_{\mathbf{A}}(-\lambda_z \langle \boldsymbol{\chi}_z \rangle) \quad \text{with} \quad \lambda_a = \frac{1}{\langle \boldsymbol{\chi}_a \rangle} - v_a, \quad a \in \{x, z\}$$

$S_{\mathbf{A}}$ denotes the S-transform (in free probability) of the limiting spectrum of Gramian $\mathbf{A}\mathbf{A}^\dagger$.

- This self-averaging property reduces the complexity of EP from $O(K^3)$ to $O(K^2)$.
- E.g. let $\{A_{ij}\}$ be iid with zero mean and variance $1/K$, then $S_{\mathbf{A}}(z) = 1/(1 + \alpha z)$ with $\alpha = \dim(\mathbf{y})/\dim(\mathbf{x})$.

Illustrations via 1-bit Compressed Sensing

$$\text{Signal Model:} \quad \mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x}) \quad \text{with} \quad \mathbf{x} \sim (1 - \rho)\delta(\mathbf{x}) + \rho N(\mathbf{x}|\mathbf{0}, \tau \mathbf{I}).$$

- Signals are typically sparse in the discrete cosine transform (DCT) domain.
- Hence, we can consider that the rows of \mathbf{A} are pseudo-randomly drawn from the $K \times K$ DCT matrix.
- In this case, we have $S_{\mathbf{A}}(z) = 1$, i.e. $v_z = \frac{1}{\langle \boldsymbol{\chi}_x \rangle} - v_x$.

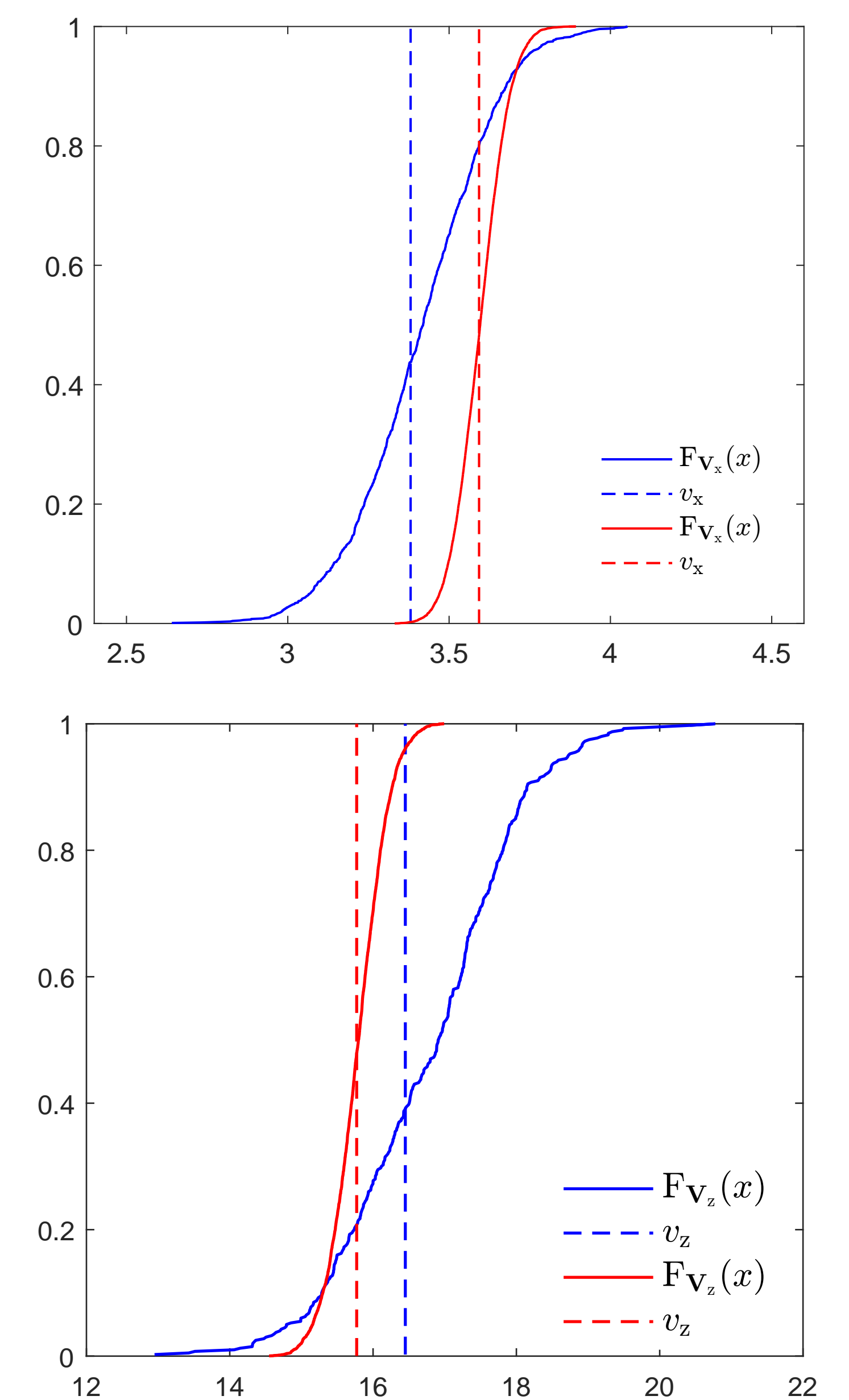


Figure 1: Empirical cumulative distribution function of the cavity variances. The dimensions of \mathbf{A} are $K/3 \times K$, $\rho = 0.1$ and $\tau = 1$. Blue curves are for $K = 1200$ and red curves are for $K = 9600$. The quantities v_x and v_z are obtained from the stable solutions of self-averaging EP.

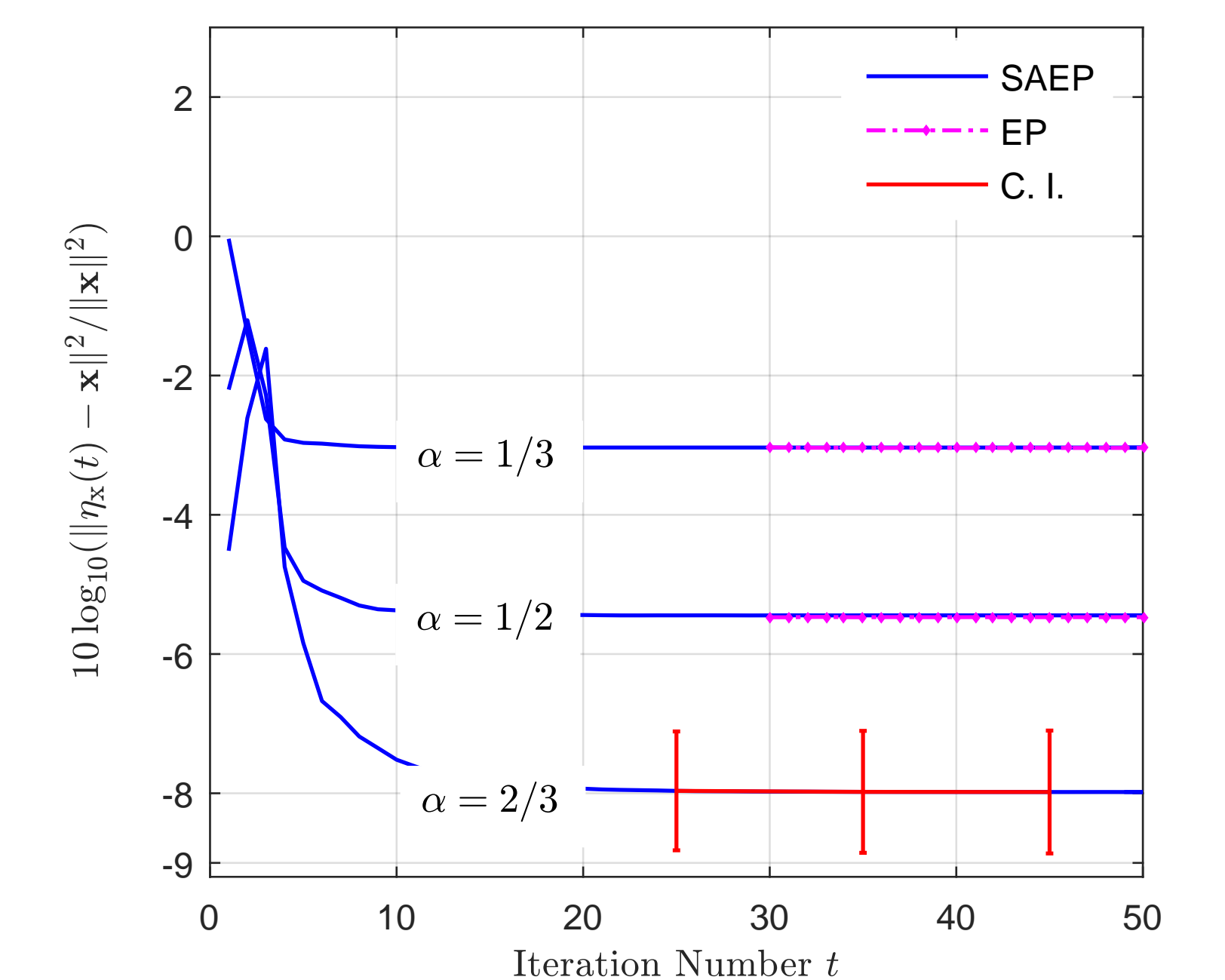


Figure 2: Mean-square-error of EP and self-averaging EP (SAEP) versus number of iterations: $\boldsymbol{\eta}_x(t)$ denotes the estimate of \mathbf{x} computed by an algorithm at iteration number t , the size of \mathbf{A} is $\alpha 1200 \times 1200$, $\rho = 0.1$ and $\tau = 1$. The reported figures are empirical averages over 100 and 1000 trials for $\alpha \in \{1/3, 1/2\}$ and $\alpha = 2/3$, respectively. C.I. denotes the confidence interval in dB.