Self-Averaging Expectation Propagation

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

Publication date:
2016

Citation for published version (APA):
Cakmak, B., Opper, 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

Self-Averaging Expectation Propagation

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

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

Problem and Objective

Recover signal $\boldsymbol{x}$ from the observation $\boldsymbol{y}$ where

$$\boldsymbol{x} \to \boldsymbol{A}\boldsymbol{x} \to \boldsymbol{y}$$

For example:

- $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$
- $\boldsymbol{y} = \mathrm{sign}(\boldsymbol{A}\boldsymbol{x})$

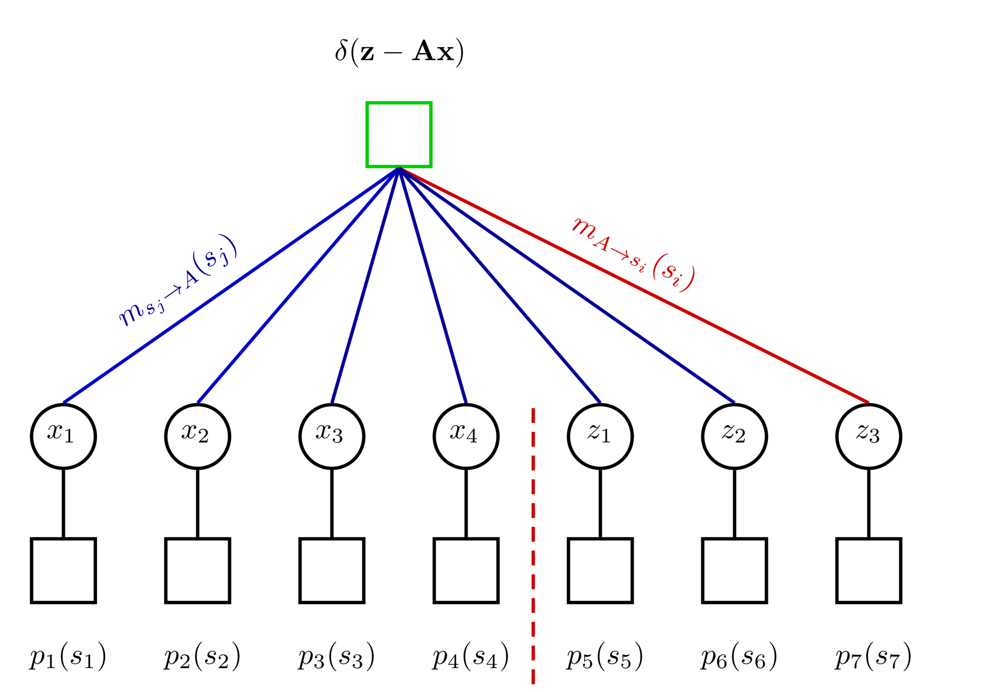
Assume that

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

Obtain iterative estimation algorithms with

- Low computational complexity
- Good accuracy

Expectation Propagation (EP)



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

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

$$m_{A\to s_i}(s_i) = \int \delta(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}) \prod_{j\neq i} m_{s_j\to A}(s_j)\mathrm{d}s_j \propto \exp\left(-\frac{1}{2}\mathrm{V}_{ii}s_i^2 + \rho_i s_i\right)$$

The two pdfs

$$q_i(s_i) \propto p_i(s_i) m_{A\to s_i}(s_i)$$
$$\tilde{q}_i(s_i) \propto m_{s_i\to A}(s_i) m_{A\to s_i}(s_i)$$

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 VIRTUOSO (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 $\boldsymbol{s} = (\boldsymbol{x}, \boldsymbol{z})$ is given by

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

- EP approximates the exact posterior pdf in the form of

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

  where $\{\mathrm{V}_{ii}\}$ are called cavity variances.
- The equations of $\boldsymbol{\rho} = (\boldsymbol{\rho}_\mathrm{x}, \boldsymbol{\rho}_\mathrm{z})$ can be expressed by

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

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

$$\chi_i = \frac{1}{\Lambda_{ii} + \mathrm{V}_{ii}} = \begin{cases} [(\boldsymbol{\Lambda}_\mathrm{x} + \boldsymbol{A}^\dagger \boldsymbol{\Lambda}_\mathrm{z}\boldsymbol{A})^{-1}]_{ii} & \Lambda_{ii} = [\boldsymbol{\Lambda}_\mathrm{x}]_{ii} \\ [\boldsymbol{A}(\boldsymbol{\Lambda}_\mathrm{x} + \boldsymbol{A}^\dagger \boldsymbol{\Lambda}_\mathrm{z}\boldsymbol{A})^{-1}\boldsymbol{A}^\dagger]_{jj} & \Lambda_{ii} = [\boldsymbol{\Lambda}_\mathrm{z}]_{jj} \end{cases}$$

  where $\boldsymbol{\chi} \triangleq (\boldsymbol{\chi}_\mathrm{x}, \boldsymbol{\chi}_\mathrm{z})$ is the variance of $q(\boldsymbol{x}, \boldsymbol{z})$.
- 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, $\mathbf{V}_\mathrm{x} \simeq v_\mathrm{x}\mathbf{I}$ and $\mathbf{V}_\mathrm{z} \simeq v_\mathrm{z}\mathbf{I}$ where

$$v_\mathrm{x} = \frac{\alpha(1 - v_\mathrm{z}\langle\boldsymbol{\chi}_\mathrm{z}\rangle)}{\langle\boldsymbol{\chi}_\mathrm{x}\rangle} \quad \& \quad v_\mathrm{z} = \lambda_\mathrm{x}\mathrm{S}_{\boldsymbol{A}}(-\lambda_\mathrm{z}\langle\boldsymbol{\chi}_\mathrm{z}\rangle) \quad \text{with} \quad \lambda_\mathrm{a} = \frac{1}{\langle\boldsymbol{\chi}_\mathrm{a}\rangle} - v_\mathrm{a}, \quad \mathrm{a} \in \{\mathrm{x}, \mathrm{z}\}$$

  $\mathrm{S}_{\boldsymbol{A}}$ denotes the S-transform (in free probability) of the limiting spectrum of Gramian $\boldsymbol{A}\boldsymbol{A}^\dagger$.
- 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 $\mathrm{S}_{\boldsymbol{A}}(z) = 1/(1+\alpha z)$ with $\alpha = \dim(\boldsymbol{y})/\dim(\boldsymbol{x})$.

Illustrations via 1-bit Compressed Sensing

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

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

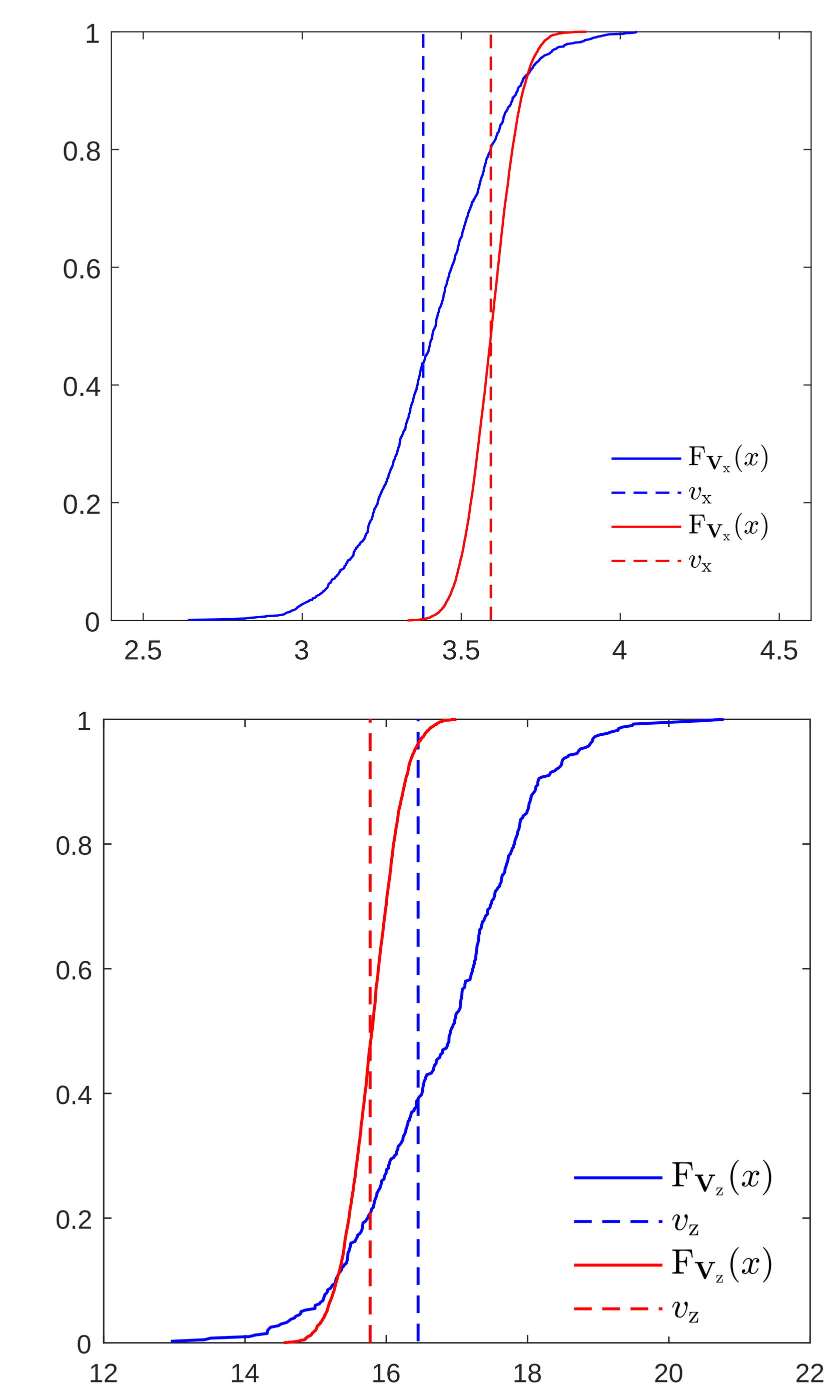


Figure 1: Empirical cumulative distribution function of the cavity variances. The dimensions of $\boldsymbol{A}$ are $K/3 \times K$, $\rho = 0.1$ and $\tau = 1$. Blue curves are for $K = 1200$ and red curves are for $K = 9600$. The quantities v_x and v_z are obtained from the stable solutions of self-averaging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

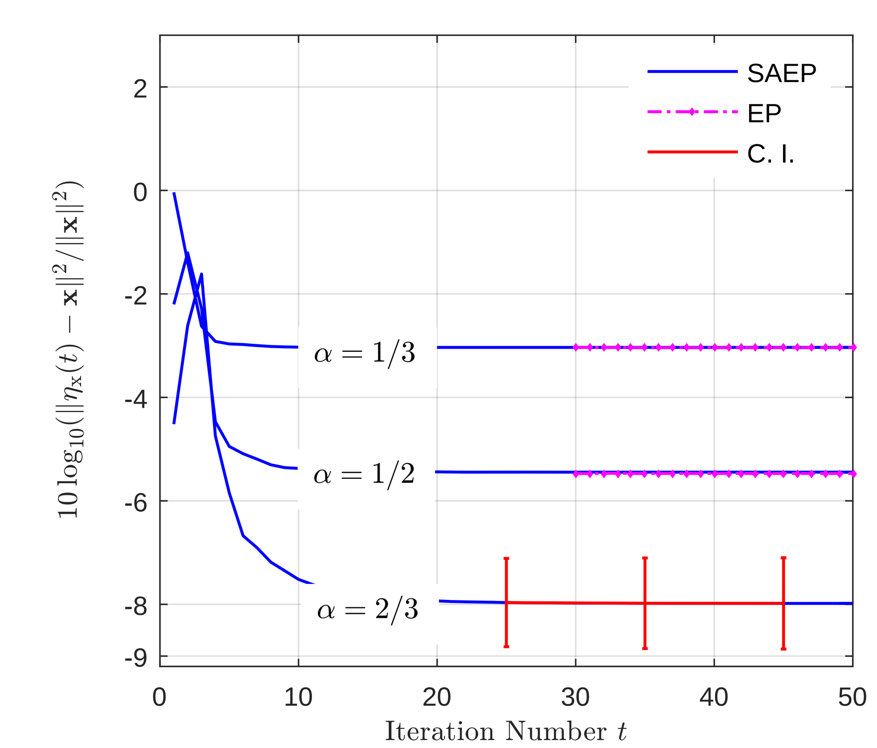


Figure 2: Mean-square-error of EP and self-averaging EP (SAEP) versus number of iterations: $\boldsymbol{\eta}_\mathrm{x}(t)$ denotes the estimate of $\boldsymbol{x}$ computed by an algorithm at iteration number t, the size of $\boldsymbol{A}$ is $\alpha 1200 \times 1200$, $\rho = 0.1$ and $\tau = 1$. 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.