Advanced Susceptibility Propagation

Muneki Yasuda

(collaborator: Kazuyuki Tanaka)

Graduate School of Information Sciences,
Tohoku University , Japan.

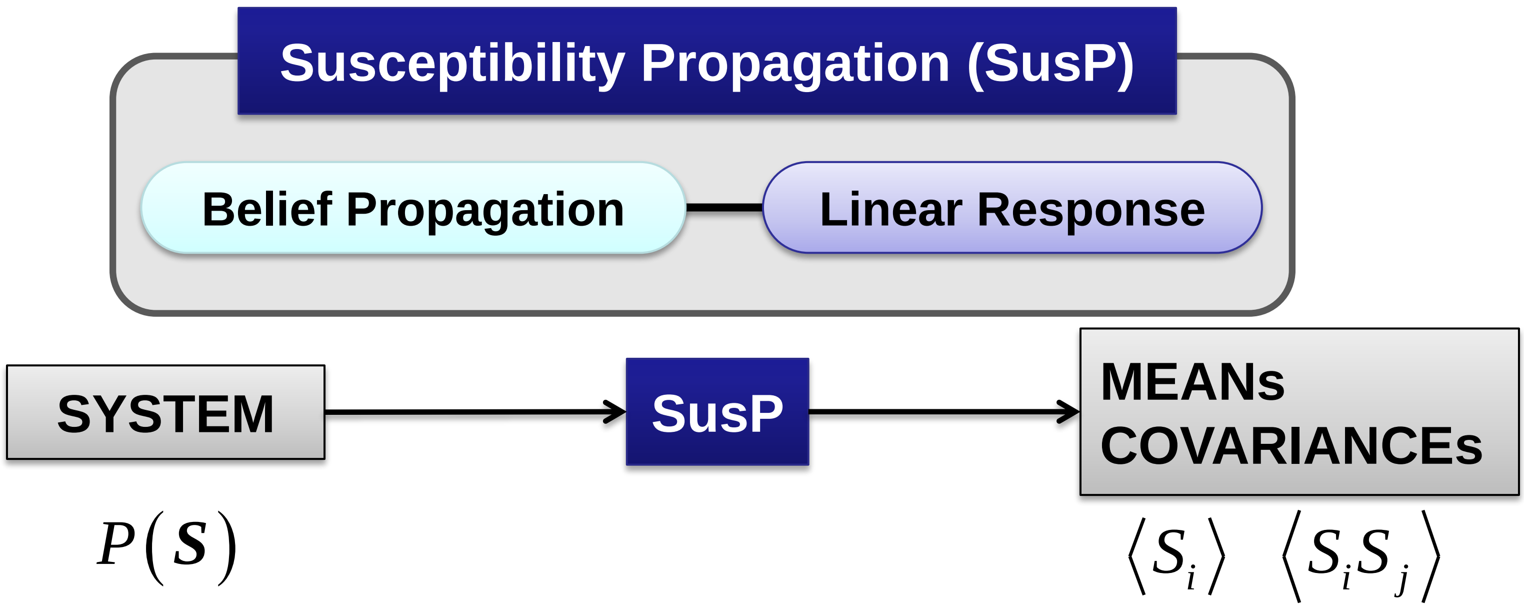
CONTENTS

INTRODUCTION

- Calculating mean values and covariances in **Markov random fields** (MRFs) is generally NP-hard problem.

- **Belief propagations** (BPs) are one of the most well-known approximate methods on MRFs.

- Combining BPs with **linear response methods** leads to **susceptibility propagations** (SusPs) that can give approximate values of covariances with a high degree of accuracy.
  (K. Tanaka, 2003; M. Welling & Y. W. The, 2004; M. Mézard & T. Mora, 2009)



Aim of This Presentation

- **Susceptibility propagations** are techniques to compute approximate covariances on Markov random fields using **belief propagations** and **linear response methods**.

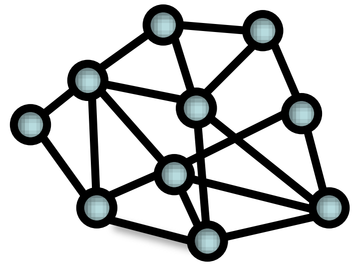
- In this presentation, **I develop a scheme of susceptibility propagations** using concepts of a ***variance matching technique***.

Susceptibility Propagation

On a given undirected graph $\boldsymbol{G}(\boldsymbol{V}, \boldsymbol{E})$,

we define a graphical model (an Ising model) expressed by

$$P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) = \frac{1}{Z(\boldsymbol{h}, \boldsymbol{J})} \exp\left(\sum_{i \in V} h_i S_i + \sum_{(i,j) \in E} J_{ij} S_i S_j \right). \quad \boldsymbol{S} \in \{+1, -1\}^n$$



Free Energy

$$F(\boldsymbol{h}, \boldsymbol{J}) := -\ln Z(\boldsymbol{h}, \boldsymbol{J})$$

The derivatives of the free energy give statistical quantities of the MRF:

$$\frac{\partial F(\boldsymbol{h}, \boldsymbol{J})}{\partial h_i} = -\sum_{\boldsymbol{S}} S_i P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J})$$

means

$$\frac{\partial^2 F(\boldsymbol{h}, \boldsymbol{J})}{\partial h_i \partial h_j} = -\sum_{\boldsymbol{S}} S_i S_j P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) + \left(\sum_{\boldsymbol{S}} S_i P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) \right) \left(\sum_{\boldsymbol{S}} S_j P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) \right)$$

$$\vdots$$

covariances

Belief Propagation (1)

I introduce a Belief propagation by a **Bethe free energy**.

Bethe Free Energy

$\partial(i)$: set of nodes connecting to node i.

$$F_{\mathrm{B}}(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J}) := -\sum_{i\in V} h_i m_i - \sum_{(i,j)\in E} J_{ij}\xi_{ij} + \sum_{i\in V}\left(1-|\partial(i)|\right)\sum_{\sigma_i=\pm 1}\frac{1+m_i\sigma_i}{2}\ln\frac{1+m_i\sigma_i}{2}$$

$$+\sum_{(i,j)\in E}\sum_{\sigma_i,\sigma_j=\pm 1}\frac{1+m_i\sigma_i+m_j\sigma_j+\xi_{ij}\sigma_i\sigma_j}{4}\ln\frac{1+m_i\sigma_i+m_j\sigma_j+\xi_{ij}\sigma_i\sigma_j}{4}$$

where

$$\xi_{ij} := \coth(2J_{ij})\left(1-\sqrt{1-\left(1-m_i^2-m_j^2\right)\tanh(2J_{ij})-2m_i m_j\tanh(2J_{ij})}\right).$$

Bethe Approximation

The true free energy is approximated by minimizing the Bethe free energy w.r.t. $\boldsymbol{m}$.

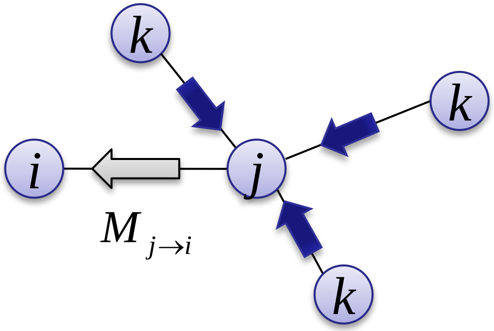
$$F(\boldsymbol{h},\boldsymbol{J}) \approx \min_{\boldsymbol{m}} F_{\mathrm{B}}(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J})$$

Belief Propagation (2)

The minimum condition of the Bethe free energy is equivalent to a **message-passing rule** (equations of *effective fields*) of BP.

Message-Passing Rule

$$M_{i\to j} = \tanh^{-1}\left(\tanh\left(J_{ij}\right)\tanh\left(h_i + \sum_{k\in\partial(i)\setminus\{j\}} M_{k\to i}\right)\right)$$



Using the messages satisfying the message-passing rule, we obtain $\boldsymbol{m}$ that minimize the Bethe free energy as follows:

$$\hat{m}_i = \tanh\left(h_i + \sum_{j\in\partial(i)} M_{j\to i}\right) \quad \text{where} \quad \hat{\boldsymbol{m}} := \arg\min_{\boldsymbol{m}} F_{\mathrm{B}}\left(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J}\right).$$

The quantities $\boldsymbol{m}$ given by these relations are approximations of the mean values:

$$\sum_{\boldsymbol{S}} S_i P\left(\boldsymbol{S} \mid \boldsymbol{h},\boldsymbol{J}\right) = -\frac{\partial F\left(\boldsymbol{h},\boldsymbol{J}\right)}{\partial h_i} \approx -\frac{\partial}{\partial h_i}\left(\min_{\boldsymbol{m}} F_{\mathrm{B}}\left(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J}\right)\right) = \hat{m}_i$$

Susceptibility Propagation (1)

I define the covariant matrix by

$$\chi_{ij} := \sum_{\boldsymbol{S}} S_i S_j P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) - \left(\sum_{\boldsymbol{S}} S_i P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) \right) \left(\sum_{\boldsymbol{S}} S_j P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) \right).$$

These quantities are sometime called **susceptibilities**.

Linear Response Relation

We approximate the susceptibilities using the Bethe free energy:

$$\chi_{ij} = -\frac{\partial^2 F(\boldsymbol{h}, \boldsymbol{J})}{\partial h_i \partial h_j} \approx -\frac{\partial^2}{\partial h_i \partial h_j} \left(\min_{\boldsymbol{m}} F_{\mathrm{B}}(\boldsymbol{m}, \boldsymbol{h}, \boldsymbol{J}) \right) = \frac{\partial \hat{m}_i}{\partial h_j}.$$

The SusP is a message-passing algorithm to compute $\hat{\chi}_{ij} := \partial \hat{m}_i / \partial h_j$.

Susceptibility Propagation (2)

Message-Passing Rule of SusP

After the BP, we compute the following message-passing:

$$\hat{\chi}_{ij} = \left(1 - \hat{m}_i^2\right)\left(\delta_{ij} + \sum_{k \in \partial(i)} \eta_{k \to j,i}\right),$$

$$\eta_{i \to j,k} = \frac{\sinh\left(2J_{ij}\right)\left(\delta_{ik} + \sum_{l \in \partial(i) \setminus \{j\}} \eta_{l \to i,k}\right)}{\cosh\left(2J_{ij}\right) + \cosh\left(2h_i + 2\sum_{l \in \partial(i) \setminus \{j\}} M_{l \to i}\right)},$$

where $\eta_{i \to j,k} := \partial M_{i \to j} / \partial h_k$.

Above equations are closed w.r.t. the approximate susceptibilities $\hat{\chi}_{ij}$.

The computational complexity of the SusP is $O\left(|V||E|\right)$.

(with *synchronous* updating rule)

Susceptibility Propagation (3)

Summary of SusP

BP

$$M_{i\to j} = \tanh^{-1}\left(\tanh\left(J_{ij}\right)\tanh\left(h_i + \sum_{k\in\partial(i)\setminus\{j\}} M_{k\to i}\right)\right)$$

$$\hat{m}_i = \tanh\left(h_i + \sum_{j\in\partial(i)} M_{j\to i}\right)$$

local magnetization

$$\begin{cases} \hat{\chi}_{ij} := \partial \hat{m}_i / \partial h_j \\ \eta_{i\to j,k} := \partial M_{i\to j} / \partial h_k \end{cases}$$

susceptibility $\approx \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle$

SusP

$$\hat{\chi}_{ij} = \left(1 - \hat{m}_i^2\right)\left(\delta_{ij} + \sum_{k\in\partial(i)} \eta_{k\to j,i}\right)$$

$$\eta_{i\to j,k} = \frac{\sinh\left(2J_{ij}\right)\left(\delta_{ij} + \sum_{l\in\partial(i)\setminus\{j\}} \eta_{l\to i,k}\right)}{\cosh\left(2J_{ij}\right) + \cosh\left(2h_i + 2\sum_{l\in\partial(i)\setminus\{j\}} M_{l\to i}\right)}$$

Advanced Susceptibility Propagation

Extended Bethe Free Energy

$$\tilde{F}_{\mathrm{B}}(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J},\boldsymbol{\Lambda}) := F_{\mathrm{B}}(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J}) + \frac{1}{2}\sum_{i\in V}\Lambda_i m_i^2$$

If $\Lambda_i > 0$, this additive term corresponds to **the L_2 regularization**.

Extended BP

The additive term changes the message-passing rule in the BP as

$$\tilde{M}_{i\to j} = \tanh^{-1}\left(\tanh(J_{ij})\tanh\left(h_i - \Lambda_i \tilde{m}_i + \sum_{k\in\partial(i)\setminus\{j\}} \tilde{M}_{k\to i}\right)\right),$$

$$\tilde{m}_i = \tanh\left(h_i - \Lambda_i \tilde{m}_i + \sum_{j\in\partial(i)} \tilde{M}_{j\to i}\right)$$ where $$\tilde{\boldsymbol{m}} := \arg\min_{\boldsymbol{m}} \tilde{F}_{\mathrm{B}}(\boldsymbol{m},\boldsymbol{h},\boldsymbol{J},\boldsymbol{\Lambda}).$$

For a given $\boldsymbol{\Lambda}$, these equations are closed.

Advanced Susceptibility Propagation (2)

Extended SusP

The additive term changes the message-passing rule in the SusP as

$$\tilde{\chi}_{ij} = \frac{1-\tilde{m}_i^2}{1+\Lambda_i\left(1-\tilde{m}_i^2\right)}\left(\delta_{ij} + \sum_{k\in\partial(i)} \tilde{\eta}_{k\to j,i}\right),$$

$$\tilde{\eta}_{i\to j,k} = \frac{\sinh\left(2J_{ij}\right)\left(\delta_{ik} - \Lambda_i\tilde{\chi}_{ik} + \sum_{l\in\partial(i)\setminus\{j\}} \tilde{\eta}_{l\to i,k}\right)}{\cosh\left(2J_{ij}\right)+\cosh\left(2h_i - 2\Lambda_i\tilde{m}_i + 2\sum_{l\in\partial(i)\setminus\{j\}} \tilde{M}_{l\to i}\right)},$$

where $\hat{\chi}_{ij} := \partial\tilde{m}_i/\partial h_j$ and $\tilde{\eta}_{i\to j,k} := \partial\tilde{M}_{i\to j}/\partial h_k$.

For a given $\mathbf{\Lambda}$, above message-passing rules are closed.

The computational complexity of the extended SusP is the same as the original SusP.

How to determine suitable values of $\mathbf{\Lambda}$?

Advanced Susceptibility Propagation (3)

Variance Matching

On binary MRFs, the relations

$$\chi_{ii} + \left(\sum_{\boldsymbol{S}} S_i P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) \right)^2 = \sum_{\boldsymbol{S}} S_i^2 P(\boldsymbol{S} \mid \boldsymbol{h}, \boldsymbol{J}) = 1$$

are **always** hold.

However, the SusP no longer keeps the consistencies due to approximation.
(M. Yasuda & K. Tanaka, 2007)

We determine values of Λ so as to satisfy the relations that are trivially hold on binary MRFs, say, match true variances and variances obtained through the SusP. *Variance Matching* !

This requirement corresponds to the conditions : $\tilde{\chi}_{ii} + \tilde{m}_i^2 = 1.$

This conditions hold by setting

$$\Lambda_i = \frac{1}{1 - \tilde{m}_i^2} \sum_{j \in \partial(i)} \tilde{\eta}_{j \to i, i}.$$

Algorithm of Advanced Susceptibility Propagation

$$\tilde{M}_{i\to j} \leftarrow \tanh^{-1}\left(\tanh\left(J_{ij}\right)\tanh\left(h_i - \Lambda_i \tilde{m}_i + \sum_{k\in\partial(i)\setminus\{j\}} \tilde{M}_{k\to i}\right)\right)$$

$$\tilde{m}_i \leftarrow \tanh\left(h_i - \Lambda_i \tilde{m}_i + \sum_{j\in\partial(i)} \tilde{M}_{j\to i}\right)$$

Extended BP

$$\tilde{\chi}_{ij} \leftarrow \frac{1-\tilde{m}_i^2}{1+\Lambda_i\left(1-\tilde{m}_i^2\right)}\left(\delta_{ij} + \sum_{k\in\partial(i)} \tilde{\eta}_{k\to j,i}\right)$$

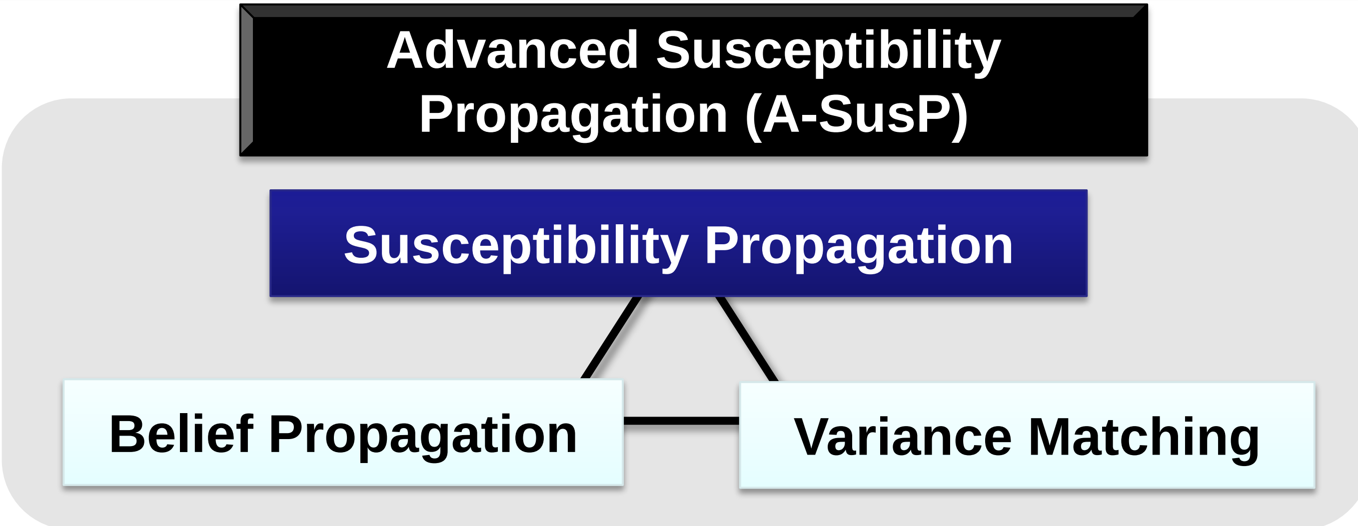
$$\tilde{\eta}_{i\to j,k} \leftarrow \frac{\sinh\left(2J_{ij}\right)\left(\delta_{ij} - \Lambda_i \tilde{\chi}_{ik} + \sum_{l\in\partial(i)\setminus\{j\}} \tilde{\eta}_{l\to i,k}\right)}{\cosh\left(2J_{ij}\right) + \cosh\left(2h_i - 2\Lambda_i \tilde{m}_i + 2\sum_{l\in\partial(i)\setminus\{j\}} \tilde{M}_{l\to i}\right)}$$

Extended SusP

$$\Lambda_i \leftarrow \frac{1}{1-\tilde{m}_i^2} \sum_{j\in\partial(i)} \tilde{\eta}_{j\to i,i}.$$

Variance Matching

Overview of Advanced Susceptibility Propagation



The SusP and the A-SusP have **the same computational cost**.

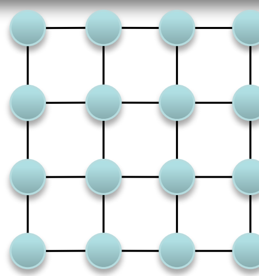
The variance matching technique introduced here is known as the **diagonal trick method** in learning in inverse Ising problems.
(H. J. Kappen & F. B. Rodríguez, 1998; T. Tanaka, 1998; M. Yasuda & K. Tanaka, 2009)

If one employs the naïve mean-field free energy instead of the Bethe free energy, the present framework gives **the adaptive TAP equation** (M. Opper & O. Winther, 2001).

The A-SusP is interpreted as **an extension of the adaptive TAP approach**.

Numerical Experiment (1)

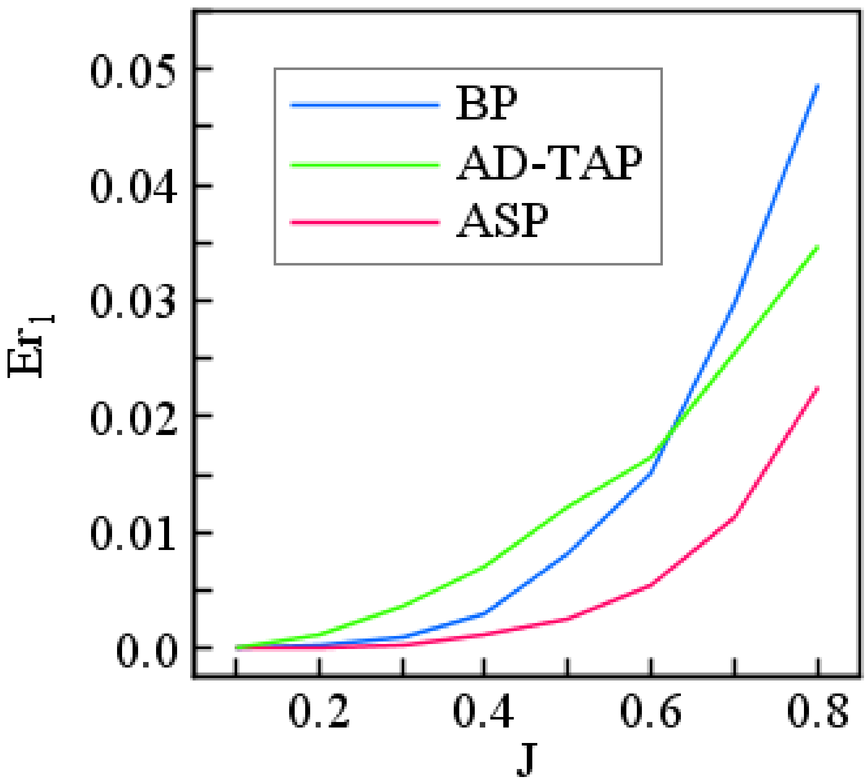
Consider systems on **the 4 × 4 square grid**.

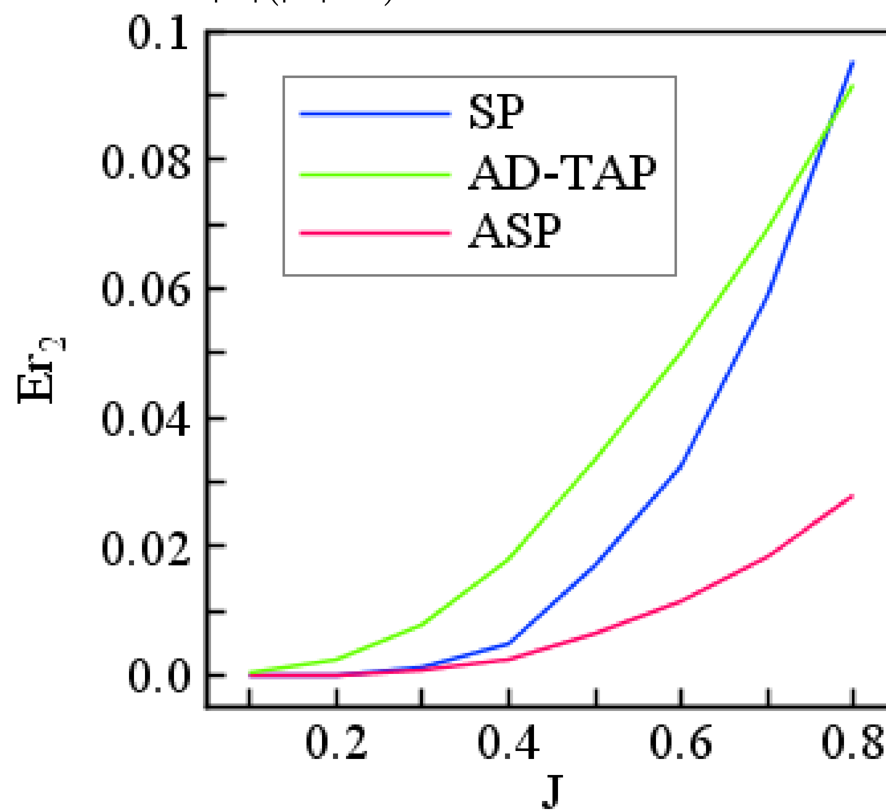


The parameters h_i and J_{ij} are independently drawn from distributions $N(0, 0.1^2)$ and $N(0, J^2)$, respectively .
$N(a, b)$: Gaussian with mean a and variance b.

$$\mathrm{Er}_1 := \frac{1}{|V|}\sum_{i\in V}\left|\langle S_i\rangle_{\mathrm{exact}} - \langle S_i\rangle_{\mathrm{approx}}\right|$$

$$\mathrm{Er}_2 := \frac{2}{|V|(|V|-1)}\sum_{i<j}\left|\langle S_i S_j\rangle_{\mathrm{exact}} - \langle S_i S_j\rangle_{\mathrm{approx}}\right|$$



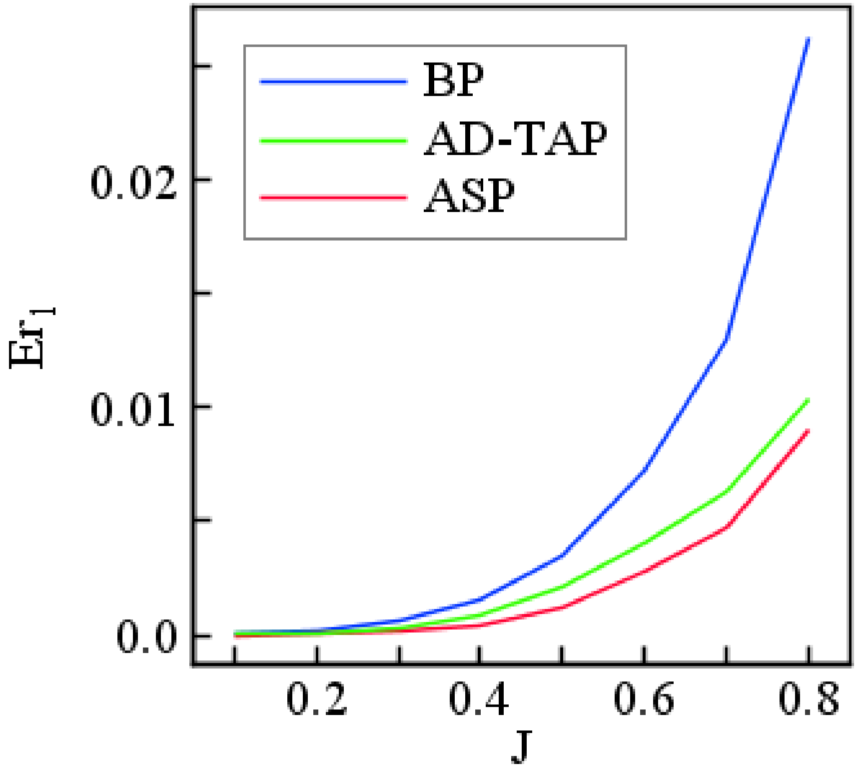


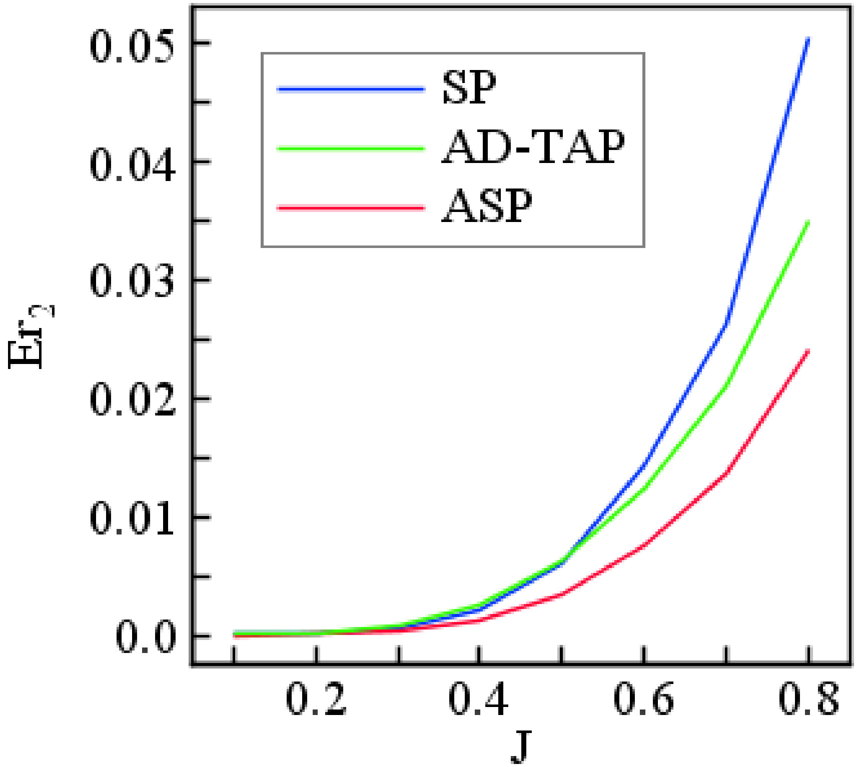
Numerical Experiment (2)

Next, consider systems on **the fully-connected graph with 16 vertices**.

The parameters h_i and J_{ij} are independently drawn from distributions $N(0, 0.1^2)$ and $N(0, J^2/n)$, respectively .

$$\mathrm{Er}_1 := \frac{1}{|V|}\sum_{i\in V}\left|\langle S_i\rangle_{\mathrm{exact}} - \langle S_i\rangle_{\mathrm{approx}}\right| \qquad \mathrm{Er}_2 := \frac{2}{|V|\left(|V|-1\right)}\sum_{i<j}\left|\langle S_i S_j\rangle_{\mathrm{exact}} - \langle S_i S_j\rangle_{\mathrm{approx}}\right|$$



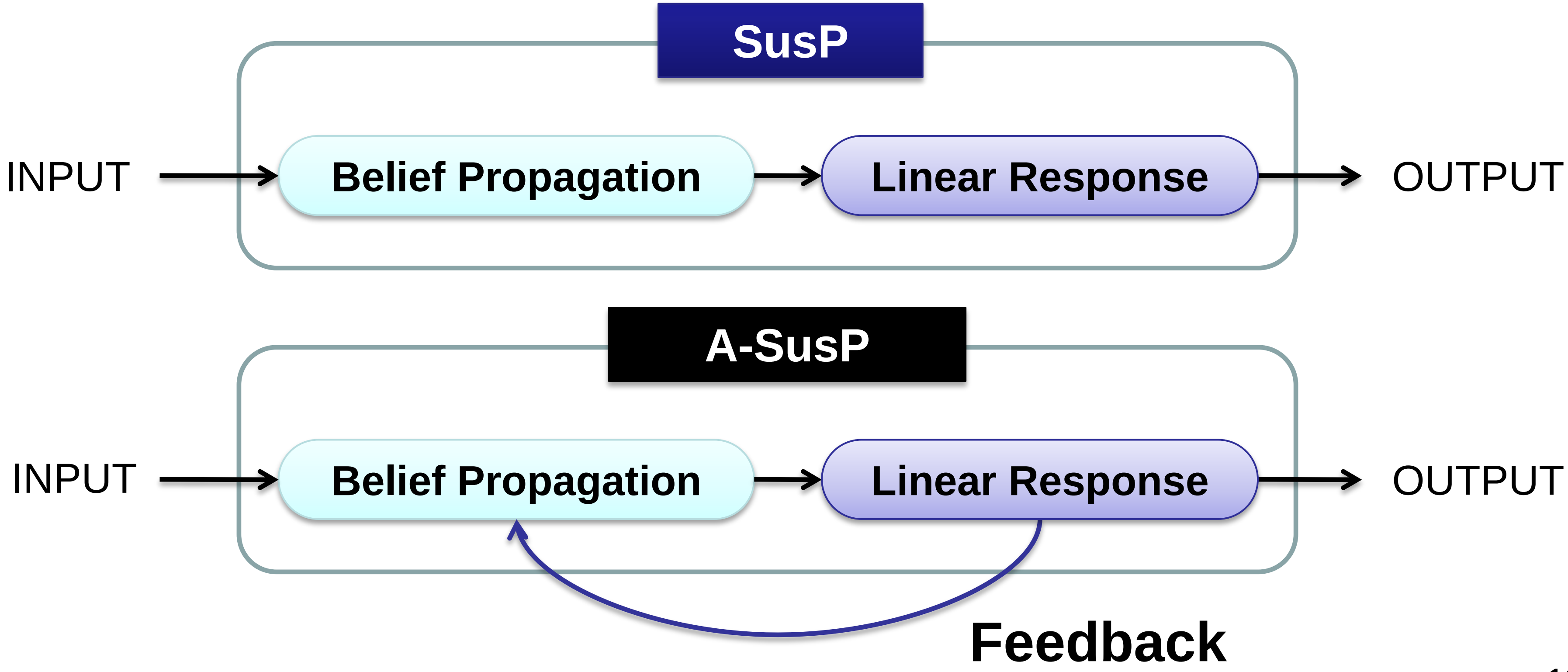


CONCLUSION

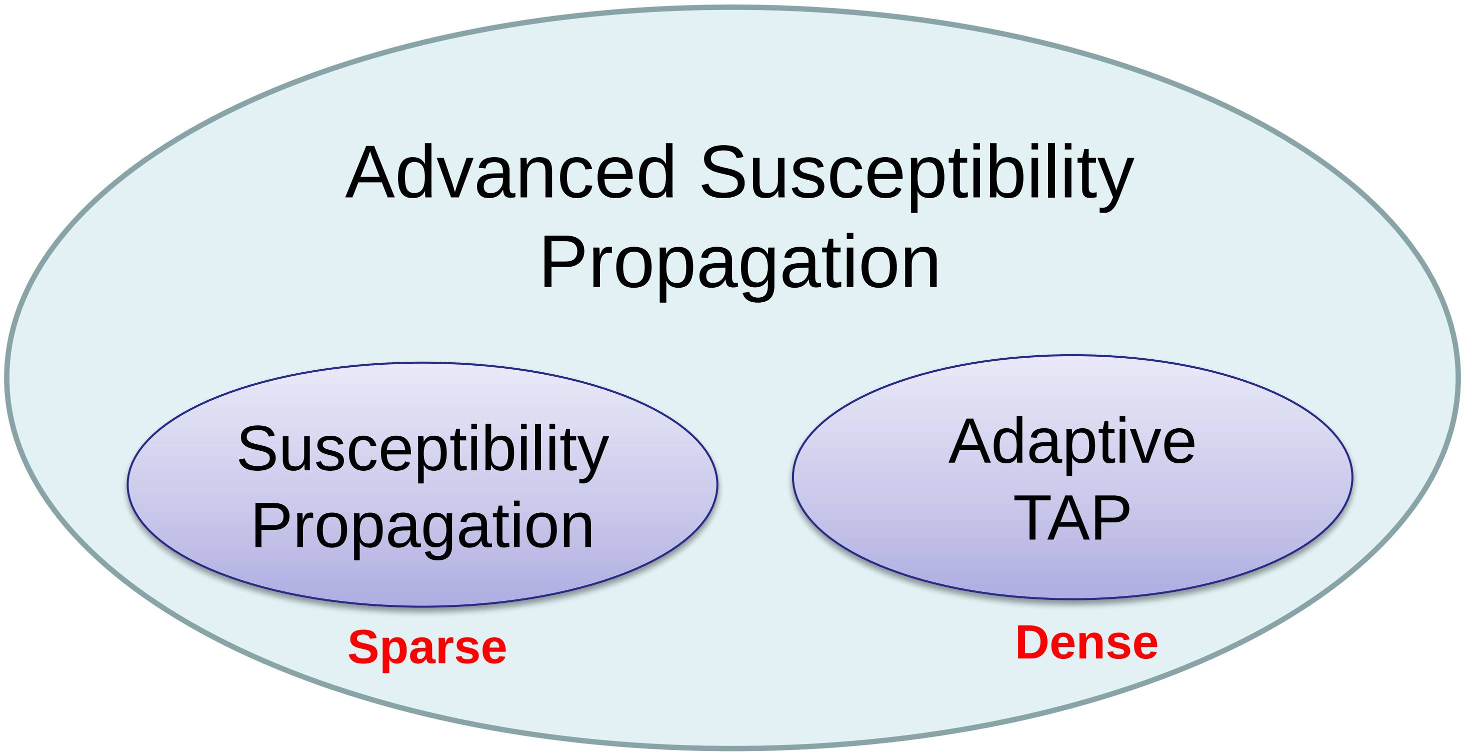
We have proposed the improved SusP algorithm.

The new SusP has the same computational cost as the conventional SusP.

Since the A-SusP has **a feedback scheme to the BP**,
it improves not only **covariances** but **means**.



Thank you for your kindly attentions !



The proposed method is strong for both **dense** and **sparse** systems !

What are $\boldsymbol{\Lambda}$?

- The parameters $\boldsymbol{\Lambda}$ force $\langle S_i^2 \rangle = \chi_{ii} - m_i^2$,

  obtained through susceptibility propagations, to be one.

- The condition for $\boldsymbol{\Lambda}$ can be also interpreted as a *Hessian matching*.

Introduction of Gibbs Free Energy (GFE)

$$H(\boldsymbol{S}) := -\sum_{i \in V} h_i S_i - \sum_{(i,j) \in E} J_{ij} S_i S_j, \qquad \boldsymbol{S} \in \{+1, -1\}^n$$

$$G(\boldsymbol{m}) := \underset{\{\lambda,\gamma\}}{\mathrm{extr}} \min_{Q} \left\{ \sum_{\boldsymbol{S}} H(\boldsymbol{S}) Q(\boldsymbol{S}) + \sum_{\boldsymbol{S}} Q(\boldsymbol{S}) \ln Q(\boldsymbol{S}) - \gamma \left(\sum_{\boldsymbol{S}} Q(\boldsymbol{S}) - 1 \right) - \sum_{i \in V} \lambda_i \left(\sum_{\boldsymbol{S}} S_i Q(\boldsymbol{S}) - m_i \right) \right\}$$

$$= -\sum_{i \in V} h_i m_i + \max_{\lambda} \left\{ \sum_{i \in V} \lambda_i m_i + F(\boldsymbol{\lambda}, \boldsymbol{J}) \right\}.$$

Properties of Gibbs Free Energy

- minimum of the GFE is equal to the free energy,
- values of $\boldsymbol{m}$ that minimize the GFE are equal to exact magnetizations of the original Ising model:

$$-\ln Z(\boldsymbol{h},\boldsymbol{J}) = \min_{\boldsymbol{m}} G(\boldsymbol{m}), \quad \langle \boldsymbol{S} \rangle = \arg\min_{\boldsymbol{m}} G(\boldsymbol{m}).$$

Approximate Gibbs Free Energy

By using an approximation, for example the Bethe approximation, we can approximate the exact GFE:

$$G(\boldsymbol{m}) \approx G_{\mathrm{app}}(\boldsymbol{m}).$$

And, let us extend the approximate GFE as

$$\hat{G}_{\mathrm{app}}(\boldsymbol{m},\boldsymbol{\Lambda}) \approx G_{\mathrm{app}}(\boldsymbol{m}) + \frac{1}{2}\sum_{i\in V}\Lambda_i m_i^2.$$

Hessian Matrices of Gibbs Free Energies

Let us define Hessian matrices of the exact GFE and the approximate GFE as

$$\left[\boldsymbol{G}(\boldsymbol{m})\right]_{ij} := \frac{\partial^2 G(\boldsymbol{m})}{\partial m_i \partial m_j}, \quad \left[\hat{\boldsymbol{G}}_{\mathrm{app}}(\boldsymbol{m},\boldsymbol{\Lambda})\right]_{ij} := \frac{\partial^2 \hat{G}_{\mathrm{app}}(\boldsymbol{m},\boldsymbol{\Lambda})}{\partial m_i \partial m_j}.$$

We want to find optimal values of $\boldsymbol{\Lambda}$ which make the Hessian matrix of approximate GFE the best approximation of that of exact GFE:

$$\min_{\boldsymbol{\Lambda}}\left(\text{distance between } \boldsymbol{G}(\boldsymbol{m}) \text{ and } \hat{\boldsymbol{G}}_{\mathrm{app}}(\boldsymbol{m},\boldsymbol{\Lambda})\right)$$

A Measure of *Similarity* of Matrices

Given two (positive definite and symmetric) matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, let us measure a similarity between these matrices, using a *Kullback-Leibler divergence* (KLD), as

$$D(\boldsymbol{A} \| \boldsymbol{B}) := \int N_0(\boldsymbol{x} \mid \boldsymbol{A}) \ln \frac{N_0(\boldsymbol{x} \mid \boldsymbol{A})}{N_0(\boldsymbol{x} \mid \boldsymbol{B})} \,\mathrm{d}\boldsymbol{x},$$

where $N_0(\boldsymbol{x} \mid \boldsymbol{A})$ is a multivariate Gaussian

$$N_0(\boldsymbol{x} \mid \boldsymbol{A}) := \sqrt{\frac{\det \boldsymbol{A}}{(2\pi)^n}} \exp\left(-\frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}\right).$$

Properties of the KLD

$$D(\boldsymbol{A} \| \boldsymbol{B}) \geq 0, \quad D(\boldsymbol{A} \| \boldsymbol{B}) = 0 \quad \text{iff} \quad \boldsymbol{A} = \boldsymbol{B}.$$

Let us regard values of $\boldsymbol{\Lambda}$, which minimize the KLD between the Hessian matrices, give the best approximation of the Hessian matrix of exact GFE:

$$\min_{\boldsymbol{\Lambda}}\left(\text{distance between } \boldsymbol{G}(\boldsymbol{m}) \text{ and } \hat{\boldsymbol{G}}_{\text{app}}(\boldsymbol{m},\boldsymbol{\Lambda})\right)$$
$$\approx \min_{\boldsymbol{\Lambda}} D(\boldsymbol{G}(\boldsymbol{m}) \,||\, \hat{\boldsymbol{G}}_{\text{app}}(\boldsymbol{m},\boldsymbol{\Lambda}))$$

The minimum condition of above KLD is equivalent to the condition for $\boldsymbol{\Lambda}$ in the proposed framework.