# Convergence and dynamics of expectation propagation and adaptive TAP

Ole Winther

Technical University of Denmark (DTU)

Collaborators: Manfred Opper, Ulrich Paquet, Burak Cakmak and Bernard Fleury

May 21, 2015

# Motivation – important open questions in EP

1. EP most often converges – but not always

   Can we derive conditions for convergence?

# Motivation – important open questions in EP

1. EP most often converges – but not always

   Can we derive conditions for convergence?

2. Parallel update EP – needs damping

   Can we derive better algorithms (from stat. physics)?

# Motivation – important open questions in EP

1. EP most often converges – but not always

   Can we derive conditions for convergence?

2. Parallel update EP – needs damping

   Can we derive better algorithms (from stat. physics)?

3. Empirically EP marginal likelihood often lower bound?

   Can we derive conditions for bound?

# Motivation – important open questions in EP

1. EP most often converges – but not always

   Can we derive conditions for convergence?

2. Parallel update EP – needs damping

   Can we derive better algorithms (from stat. physics)?

3. Empirically EP marginal likelihood often lower bound?

   Can we derive conditions for bound?

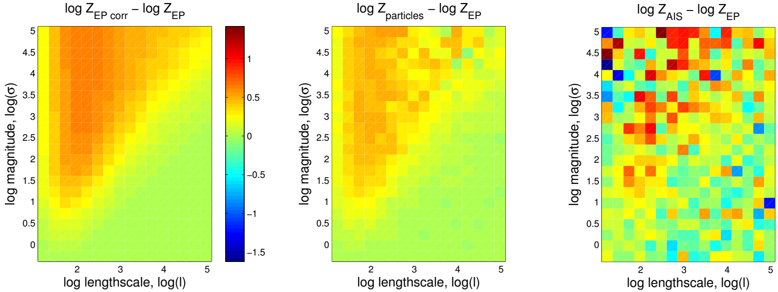
I will address the first two questions!

# Simulations - Gaussian process classification

- Marginal likelihood (Paquet, Winther & Opper, 2008+2013)
- USPS digits 3-vs-5, $N = 767$ and kernel

$$k(\xi, \xi') = \sigma^2 \exp(-||\xi - \xi||^2/2\ell^2) \; .$$

- Note correction $\log R = \log(Z/Z_{\mathrm{EP}})$ is always positive - EP a bound in this case?



# Outline

- Running examples – GP in a box and Ising model
- Expectation propagation (EP) in a nutshell

# Outline

- Running examples – GP in a box and Ising model
- Expectation propagation (EP) in a nutshell
- Part 1: EP Convergence
  - Sequential EP map
  - Convergence proof and conditions for specific models

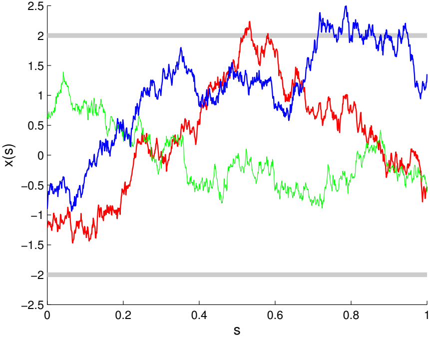
# Outline

- Running examples – GP in a box and Ising model
- Expectation propagation (EP) in a nutshell
- Part 1: EP Convergence
  - Sequential EP map
  - Convergence proof and conditions for specific models
- Part 2: Memoryless adaptive TAP dynamics
  - EP fixed-point = adaptive TAP equations
  - Marginal distribution theorem
  - Gaussian cavity field approximation $\rightarrow$ TAP
  - Construction of memoryless dynamics
  - Asymptottics of memoryless dynamics

# Example 1 - Gaussian process (GP) in a box

- GP prior over functions $x(s)$: $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}\,;\,\mathbf{0},\,\mathbf{K})$
- Take inputs $s_i = (i-1)/(N-1)$, $i = 0, \ldots, N-1$.
- Kernel matrix $K_{ij} = [\mathbf{K}]_{ij}$ from Kernel function $k(s, s')$

$$K_{ij} = k(s_i, s_j) = \exp(-|s_i - s_j|/\ell)\ , \qquad \ell = 1$$




$$p_a(\mathbf{x}) = \frac{1}{Z}\prod_n \mathbb{I}(|x_n| < a)\,\mathcal{N}(\mathbf{x}\,;\,\mathbf{0},\,\mathbf{K})$$

# Example 2 - Ising model

- Ising model

$$p(\mathbf{x}) = \frac{1}{Z} \prod_k \underbrace{[\delta(x_k + 1) + \delta(x_k - 1)]}_{f_k(x_k)} \underbrace{\exp\{\mathbf{x}^\top \mathbf{J} \mathbf{x}/2 + \boldsymbol{\theta}^\top \mathbf{x}\}}_{f_0(\mathbf{x})} .$$

# Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_a f_a(\mathbf{x}_a)$$

# Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_a f_a(\mathbf{x}_a)$$

- Exponential family approx., restricted Gaussian

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_a g_a(\mathbf{x}_a)\,, \qquad g_a(\mathbf{x}_a) = \exp\left(\gamma_a^T \mathbf{x}_a - \mathbf{x}_a^T \mathbf{\Lambda}_a \mathbf{x}_a / 2\right)$$

# Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_a f_a(\mathbf{x}_a)$$

- Exponential family approx., restricted Gaussian

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_a g_a(\mathbf{x}_a) \,, \qquad g_a(\mathbf{x}_a) = \exp\left(\gamma_a^T \mathbf{x}_a - \mathbf{x}_a^T \mathbf{\Lambda}_a \mathbf{x}_a / 2\right)$$

- Tilted distribution tractable! Note subscript $a$

$$q_a(\mathbf{x}) = \frac{1}{Z_a} \frac{q(\mathbf{x}) f_a(\mathbf{x}_a)}{g_a(\mathbf{x}_a)}$$

# Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_a f_a(\mathbf{x}_a)$$

- Exponential family approx., restricted Gaussian

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_a g_a(\mathbf{x}_a) \,, \qquad g_a(\mathbf{x}_a) = \exp\left(\gamma_a^T \mathbf{x}_a - \mathbf{x}_a^T \mathbf{\Lambda}_a \mathbf{x}_a / 2\right)$$

- Tilted distribution tractable! Note subscript $a$

$$q_a(\mathbf{x}) = \frac{1}{Z_a} \frac{q(\mathbf{x}) f_a(\mathbf{x}_a)}{g_a(\mathbf{x}_a)}$$

- Moment matching $\phi_a(\mathbf{x}_a) = \left[\mathbf{x}_a \;\; \mathbf{x}_a \mathbf{x}_a^T\right]^T$ determines $g_a(\mathbf{x}_a)$:

$$\langle \phi_a(\mathbf{x}_a) \rangle_{q_a} = \langle \phi_a(\mathbf{x}_a) \rangle_q$$

# Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_a f_a(\mathbf{x}_a)$$

- Exponential family approx., restricted Gaussian

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_a g_a(\mathbf{x}_a)\,, \qquad g_a(\mathbf{x}_a) = \exp\left(\gamma_a^T \mathbf{x}_a - \mathbf{x}_a^T \mathbf{\Lambda}_a \mathbf{x}_a / 2\right)$$

- Tilted distribution tractable! Note subscript $a$

$$q_a(\mathbf{x}) = \frac{1}{Z_a} \frac{q(\mathbf{x}) f_a(\mathbf{x}_a)}{g_a(\mathbf{x}_a)}$$

- Moment matching $\phi_a(\mathbf{x}_a) = \left[\mathbf{x}_a \;\; \mathbf{x}_a \mathbf{x}_a^T\right]^T$ determines $g_a(\mathbf{x}_a)$:

$$\langle \phi_a(\mathbf{x}_a) \rangle_{q_a} = \langle \phi_a(\mathbf{x}_a) \rangle_q$$

- Marginal likelihood: $Z \approx Z_{\mathrm{EP}} = Z_q \prod_a Z_a$ .

# EP for Ising model

- Ising model

$$p(\mathbf{x}) = \frac{1}{Z} \prod_k \underbrace{[\delta(x_k + 1) + \delta(x_k - 1)]}_{f_k(x_k)} \underbrace{\exp\{\mathbf{x}^\top \mathbf{J} \mathbf{x}/2 + \boldsymbol{\theta}^\top \mathbf{x}\}}_{f_0(\mathbf{x})} .$$

# EP for Ising model

- Ising model

$$p(\mathbf{x}) = \frac{1}{Z} \prod_k \underbrace{[\delta(x_k + 1) + \delta(x_k - 1)]}_{f_k(x_k)} \underbrace{\exp\{\mathbf{x}^\top \mathbf{J}\mathbf{x}/2 + \boldsymbol{\theta}^\top \mathbf{x}\}}_{f_0(\mathbf{x})} .$$

- Factorise: $g_0(\mathbf{x}) = f_0(\mathbf{x})$ & $g_k(x_k) = \exp\left(\gamma_k x_k - \Lambda_{kk} x_k^2/2\right)$:

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_{k=0}^{K} g_k(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} - \mathbf{J})^{-1}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\gamma + \boldsymbol{\theta})$.

# EP for Ising model

- Ising model

$$p(\mathbf{x}) = \frac{1}{Z} \prod_k \underbrace{[\delta(x_k+1) + \delta(x_k-1)]}_{f_k(x_k)} \underbrace{\exp\{\mathbf{x}^\top \mathbf{J}\mathbf{x}/2 + \boldsymbol{\theta}^\top \mathbf{x}\}}_{f_0(\mathbf{x})} .$$

- Factorise: $g_0(\mathbf{x}) = f_0(\mathbf{x})$ & $g_k(x_k) = \exp\left(\gamma_k x_k - \Lambda_{kk} x_k^2/2\right)$:

$$q(\mathbf{x}) = \frac{1}{Z_q} \prod_{k=0}^{K} g_k(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} - \mathbf{J})^{-1}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\gamma + \boldsymbol{\theta})$.

- Tilted distribution $q_k(x_k) = \int q_k(\mathbf{x}) d\mathbf{x}_{\setminus k}$

$$q_k(x_k) = \frac{1}{Z_k} \frac{f_k(x_k)}{g_k(x_k)} \int q(\mathbf{x}) d\mathbf{x}_{\setminus k} \quad \Rightarrow \quad m_k = \tanh\left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right)$$

# EP algorithmic recipe

- Loop over $k$:
  1. Tilted distribution $q_k(x_k) = \int q_k(\mathbf{x}) d\mathbf{x}_{\setminus k}$

$$m_k \leftarrow \tanh\left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right)$$

  2. Moment matching

$$\mu_k = m_k \quad \text{and} \quad \Sigma_{kk} = 1 - m_k^2$$

     Solve wrt $\gamma_k$ and $\Lambda_{kk}$:

$$\gamma_k \leftarrow \frac{m_k}{1 - m_k^2} - \frac{\mu_k}{\Sigma_{kk}} + \gamma_k$$
$$\Lambda_{kk} \leftarrow \frac{1}{1 - m_k^2} - \frac{1}{\Sigma_{kk}} + \Lambda_{kk}$$

  3. Rank-one update of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\gamma + \boldsymbol{\theta})$.

# Part 1 – EP convergence

# EP for GP in a box

- GP in a box

$$p_a(\mathbf{x}) = \frac{1}{Z}\prod_n \mathbb{I}(|x_n| < a)\,\mathcal{N}(\mathbf{x}\,;\,\mathbf{0},\,\mathbf{K})$$

- Factors: $f_0(\mathbf{x}) = \mathcal{N}(\mathbf{x}\,;\,\mathbf{0},\,\mathbf{K})$ and

$$f_n(x_n) = \mathbb{I}(|x_n| < a)$$

# EP for GP in a box

- GP in a box

$$p_a(\mathbf{x}) = \frac{1}{Z} \prod_n \mathbb{I}(|x_n| < a)\, \mathcal{N}(\mathbf{x}\,;\, \mathbf{0},\, \mathbf{K})$$

- Factors: $f_0(\mathbf{x}) = \mathcal{N}(\mathbf{x}\,;\, \mathbf{0},\, \mathbf{K})$ and

$$f_n(x_n) = \mathbb{I}(|x_n| < a)$$

- Approximating factors $g_0(\mathbf{x}) = f_0(\mathbf{x})$ and by symmetry:

$$g_n(x_n) = \exp\left(-\Lambda_{nn} x_n^2 / 2\right)$$

# EP algorithmic recipe - GP in a box

- Loop over $n$:
  1. Tilted distribution $q_n(x_n) = \int q_n(\mathbf{x}) d\mathbf{x}_{\setminus n}$

$$q_n(x_n) = \frac{1}{Z_n} \frac{f_n(x_n)}{g_n(x_n)} q(x_n)$$
$$= \frac{1}{Z_n} \mathbb{I}(|x_n| < a) \exp\left(-\frac{\lambda_n}{2} x_n^2\right)$$
$$\lambda_n \equiv \frac{1}{\Sigma_{nn}} - \Lambda_{nn} = \frac{1}{[(\mathbf{\Lambda} - \mathbf{J})^{-1}]_{nn}} - \Lambda_{nn}$$

  2. Moment matching

$$\Sigma_{nn} = \langle x^2 \rangle_{q_n}$$

     Solve wrt $\Lambda_{nn}$:

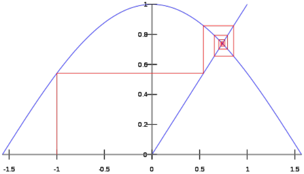
$$\Lambda_{nn} \leftarrow \frac{1}{\langle x^2 \rangle_{q_n}} - \frac{1}{\Sigma_{nn}} + \Lambda_{nn} = F(\lambda_n)$$

  3. Rank-one update of $\mathbf{\Sigma}$.

# Analysis of EP mapping

- The sequence of $N$ updates defines maps

$$\Lambda_{nn} \leftarrow T_n(\mathbf{\Lambda})$$




- Fixed-point theorem: if map is differentiable in a neighborhood of $\mathbf{T}(\mathbf{\Lambda}^*) = \mathbf{\Lambda}^*$ and

$$\left| \frac{dT_N(\mathbf{\Lambda}^*)}{d\Lambda_{NN}} \right| < 1$$

then attraction is guarenteed.

- Use chain to calculate $T'_N \equiv \frac{dT_N}{d\Lambda_{NN}}$.

# Analysis of EP mapping cont.

- Zero mean EP:

$$\Lambda_{nn} \leftarrow \frac{1}{\langle x^2 \rangle_{q_n}} - \frac{1}{\Sigma_{nn}} + \Lambda_{nn} = F(\lambda_n)$$

$$\lambda_n \equiv \frac{1}{\Sigma_{nn}} - \Lambda_{nn} = \frac{1}{[(\mathbf{\Lambda} - \mathbf{J})^{-1}]_{nn}} - \Lambda_{nn}$$

- Update order $1, \ldots, N$:

$$T_i' \equiv \frac{dT_i}{d\Lambda_{NN}} = F_i'(\lambda_i) \left( \frac{\partial \lambda_i}{\partial \Lambda_{NN}} + \sum_{l<i} \frac{\partial \lambda_i}{\partial \Lambda_{ll}} T_l' \right)$$

$$F_i'(\lambda_i) = \frac{1}{2} \left( \frac{\langle x^4 \rangle_{q_n}}{\langle x^2 \rangle_{q_n}^2} - 3 \right) = \frac{1}{2} \times \text{excess kurtosis}$$

$$\frac{\partial \lambda_i}{\partial \Lambda_{jj}} = \frac{\Sigma_{ij}^2}{\Sigma_{jj}^2} - \delta_{ij}$$

- Iteration index omitted for simplicity.

# Fixed-point analysis

- At fixed-point we can simplify to

$$\Delta_i \equiv \frac{T_i'}{F_i'(\lambda_i)} \frac{\Sigma_{ii}}{\Sigma_{NN}} = \rho_{iN}^2(1-\delta_{iN}) + \sum_{l<i} \rho_{il}^2 F_l'(\lambda_l)\Delta_l$$

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

$$T_N' = F_N'(\lambda_N)\Delta_N$$

# Fixed-point analysis

- At fixed-point we can simplify to

$$\Delta_i \equiv \frac{T_i'}{F_i'(\lambda_i)} \frac{\Sigma_{ii}}{\Sigma_{NN}} = \rho_{iN}^2(1 - \delta_{iN}) + \sum_{l<i} \rho_{il}^2 F_l'(\lambda_l)\Delta_l$$

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

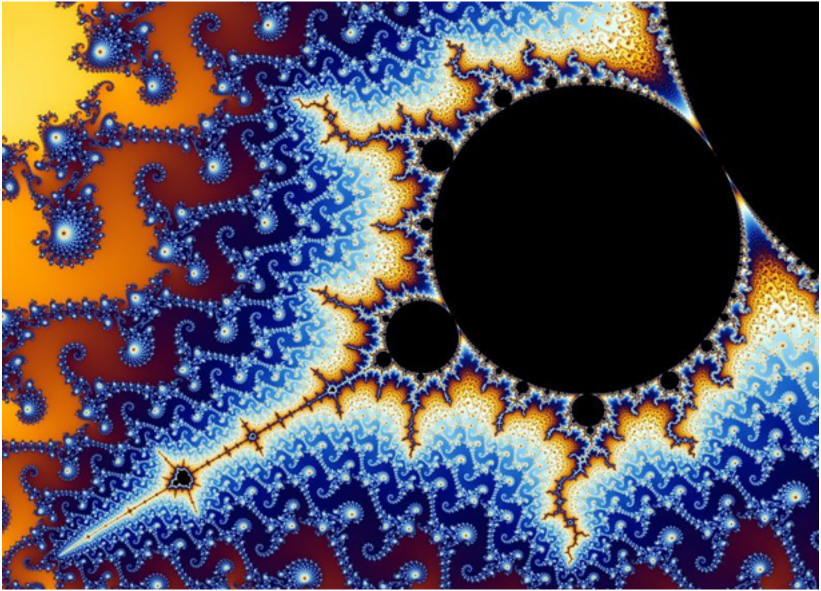
$$T_N' = F_N'(\lambda_N)\Delta_N$$

- Special case $\rho_{ij}^2 = \rho^2$, $i \neq j$

$$T_N' = \rho^2 F_N'(\lambda_N) \left( \prod_{l=1}^{i-1} \left(1 + \rho^2 F_l'\right) - 1 \right)$$

- For GP in a box (and log concave factors?) $F_i'(\lambda) \in [-1, 0]$.
- Thus $|T_N'| \leq 1$.
- Not proved for general $\boldsymbol{\Sigma}$, but for other special cases. e.g. repeated box factors (Cunningham et. al.)

# Part 2 – Memoryless dynamics



# EP algorithmic recipe

- Loop over $n$:
  1. Tilted distribution $q_n(x_k) = \int q_n(\mathbf{x}) d\mathbf{x}_{\setminus n}$

$$q_n(x_k) = \frac{1}{Z_n} \frac{f}{-} q(\mathbf{x})$$

  2. Moment matching

$$\mu_k = m_k \quad \text{and} \quad \Sigma_{kk} = 1 - m_k^2$$

Solve wrt $\gamma_k$ and $\Lambda_{kk}$:

$$\gamma_k \leftarrow \frac{m_k}{1 - m_k^2} - \frac{\mu_k}{\Sigma_{kk}} + \gamma_k$$

$$\Lambda_{kk} \leftarrow \frac{1}{1 - m_k^2} - \frac{1}{\Sigma_{kk}} + \Lambda_{kk}$$

  3. Rank-one update of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\boldsymbol{\gamma} + \boldsymbol{\theta})$.

# Equivalence with adaptive TAP equations

- EP moment matching:

$$\mathbf{m} = \boldsymbol{\mu} \qquad 1 - m_k^2 = \Sigma_{kk}, \forall k$$

- with

$$\begin{aligned} m_k &= \tanh\left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right) \\ \Sigma &= (\boldsymbol{\Lambda} - \mathbf{J})^{-1} \\ \mu &= \boldsymbol{\Sigma}(\gamma + \boldsymbol{\theta}) \end{aligned}$$

# Equivalence with adaptive TAP equations

- EP moment matching:

$$\mathbf{m} = \boldsymbol{\mu} \qquad\qquad 1 - m_k^2 = \Sigma_{kk}\,, \forall k$$

- with

$$\begin{aligned} m_k &= \tanh\left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right) \\ \Sigma &= (\boldsymbol{\Lambda} - \mathbf{J})^{-1} \\ \mu &= \boldsymbol{\Sigma}(\gamma + \boldsymbol{\theta}) \end{aligned}$$

- Fixed-points are the adaptive TAP equations (Opper and Winther, Neural Comp 2000, PRL and PRE 2001):

$$\begin{aligned} \mathbf{m} &= \boldsymbol{\mu} = (\boldsymbol{\Lambda} - \mathbf{J})^{-1}(\gamma + \boldsymbol{\theta}) \quad \Leftrightarrow -\gamma = \boldsymbol{\theta} + (\mathbf{J} - \boldsymbol{\Lambda})\mathbf{m} \\ m_k &= \tanh\left(\frac{\mu_k}{\Sigma_{kk}} - \gamma_k\right) = \tanh\left([\mathbf{Jm}]_k - v_k m_k + \theta_k\right) \\ v_k &= \Lambda_{kk} - \frac{1}{\Sigma_{kk}} \end{aligned}$$

# Marginal distribution theorem

- Exact result for marginal distribution

$$p(x_k) = \frac{1}{Z} f_k(x_k) \int \exp\left(\frac{1}{2}\mathbf{x}^T \mathbf{J}\mathbf{x} + \boldsymbol{\theta}^T \mathbf{x}\right) \prod_{k' \neq k} f_{k'}(x_{k'}) d\mathbf{x}_{\setminus k}$$

$$= \frac{1}{Z} f_k(x_k)\, e^{\frac{1}{2}J_{kk}x_k^2 + \theta_k x_k} \times$$

$$\int \exp(\underbrace{x_k \sum_{k' \neq k} J_{kk'} x_{k'}}_{\equiv h_k}) \underbrace{e^{\frac{1}{2}\mathbf{x}_{\setminus k}^T \mathbf{J}_{\setminus k} \mathbf{x}_{\setminus k} + \boldsymbol{\theta}_{\setminus k}^T \mathbf{x}_{\setminus k}} \prod_{k' \neq k} f_{k'}(x_{k'})}_{\propto p(\mathbf{x}_{\setminus k}|\mathbf{J}_{\setminus k}, \boldsymbol{\theta}_{\setminus k})} d\mathbf{x}_{\setminus k}$$

$$= \frac{1}{Z} f_k(x_k)\, e^{\frac{1}{2}J_{kk}x_k^2 + \theta_k x_k} \int e^{x_k h_k} p(h_k) dh_k$$

- $p(h_k)$ has no memory of $J_{kk}$ and $\theta_k$

$$p(h_k) \equiv \int \delta(h_k - \sum_{k' \neq k} J_{kk'} x_{k'}) p(\mathbf{x}_{\setminus k}|\mathbf{J}_{\setminus k}, \boldsymbol{\theta}_{\setminus k}) d\mathbf{x}_{\setminus k}$$

# Gaussian cavity assumption $\rightarrow$ TAP

- Gaussian cavity assumption

$$p(h_k) = \mathcal{N}(h_k | \langle h_k \rangle_{\backslash k}, \langle h_k^2 \rangle_{\backslash k} - \langle h_k \rangle_{\backslash k}^2)$$

- leads to tilted distribution form:

$$p(x_k) = \frac{1}{Z} f_k(x_k)\, e^{\frac{1}{2} J_{kk} x_k^2 + \theta_k x_k} \int e^{x_k h_k} p(h_k) dh_k$$
$$\approx \frac{1}{Z} f_k(x_k)\, e^{\frac{1}{2}(J_{kk} + \langle h_k^2 \rangle_{\backslash k} - \langle h_k \rangle_{\backslash k}^2)\, x_k^2 + (\langle h_k \rangle_{\backslash k} + \theta_k)\, x_k} \propto q_k(x_k)$$

- For Ising ($x_k^2 = 1$):

$$m_k = \tanh((\langle h_k \rangle_{\backslash k} + \theta_k)$$

- From $p(x_k, h_k)$ (Mezard, Parisi and Virasoro 1987):

$$\langle h_k \rangle_{\backslash k} = [\mathbf{Jm}]_k - v_k m_k$$
$$v_k = \langle h_k^2 \rangle_{\backslash k} - \langle h_k \rangle_{\backslash k}^2$$

# Memoryless dynamics

- Dynamics wanted that should:
  - converge to adaptive TAP fixed-point
  - be memoryless in the same sense as fixed-point.

# Memoryless dynamics

- Dynamics wanted that should:
  - converge to adaptive TAP fixed-point
  - be memoryless in the same sense as fixed-point.
- Propose (parallel) update on the form:

$$
\begin{aligned}
m_i(t+1) &= f\left(\phi_i(t) + \theta_i(t)\right) \\
\phi_i(t) &= \sum_j J_{ij} m_j(t) - \sum_{s<t} \hat{K}_i(t,s) m_i(s) \ .
\end{aligned}
$$

- Set parameters $\hat{K}_i(t,s)$ to remove memory.
- $\theta_i(t) = \theta_i$ in actual dynamics

# Memoryless dynamics

- Dynamics wanted that should:
  - converge to adaptive TAP fixed-point
  - be memoryless in the same sense as fixed-point.
- Propose (parallel) update on the form:

$$m_i(t+1) = f\left(\phi_i(t) + \theta_i(t)\right)$$

$$\phi_i(t) = \sum_j J_{ij} m_j(t) - \sum_{s<t} \hat{K}_i(t,s) m_i(s)\ .$$

- Set parameters $\hat{K}_i(t,s)$ to remove memory.
- $\theta_i(t) = \theta_i$ in actual dynamics
- Condition for memoryless dynamics $\tau < t$:

$$\frac{\partial \phi_i(t)}{\partial \theta_i(\tau)} = \sum_k J_{ik} G_{ki}(t,\tau) - \sum_{\tau<s<t} \hat{K}_i(t,s) G_{ii}(s,\tau) = 0\ .$$

- Response funtion:

$$G_{ij}(t,\tau) \equiv \frac{\partial m_i(t)}{\partial \theta_j(\tau)} = \text{change in } m_i(t) \text{ due to change in } \theta_j(\tau)$$

# Dynamics for response

- Dynamics

$$m_i(t+1) = f\left(\sum_j J_{ij} m_j(t) - \sum_{s<t} \hat{K}_i(t,s) m_i(s) + \theta_i(t)\right)$$

- Response dynamics – differentiate dynamics:

$$G_{ij}(t+1,\tau) = \frac{\partial m_i(t+1)}{\partial \theta_j(\tau)} =$$

$$g_i(t+1)\left(\delta_{ij}\delta_{t\tau} + \sum_k J_{ik} G_{kj}(t,\tau) - \sum_{\tau<s<t} \hat{K}_i(t,s) G_{ij}(s,\tau)\right)$$

- with $g_i(t+1) = \frac{\partial f(z)}{\partial z}\Big|_{z=\phi_i(t)+\theta_i(t)}$
- Complexity for step $t$: $\mathcal{O}(N^3 t)$. Not feasible!

# Dynamics for cumulative response

- Define cumulative response

$$\chi_{ij}(t) \equiv \sum_{\tau<t} G_{ij}(t,\tau)$$

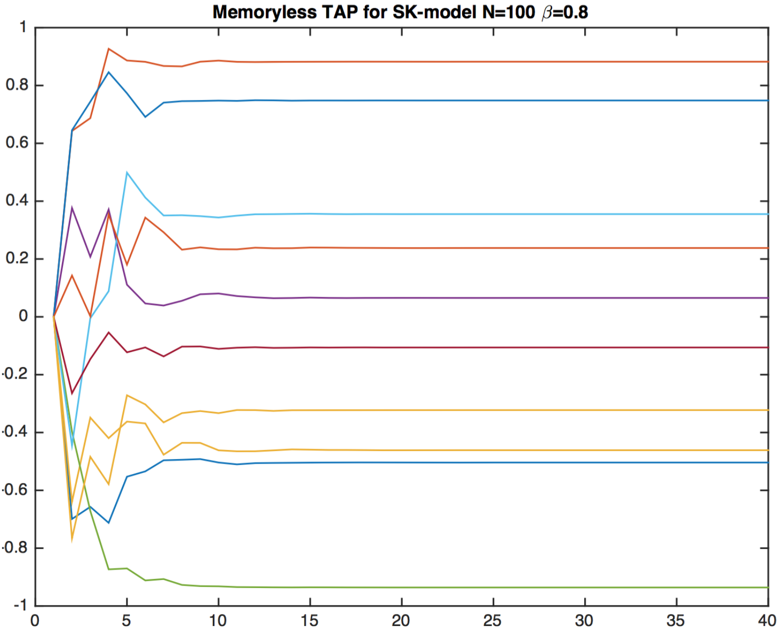
- and sum response recursion

$$G_{ij}(t+1,\tau) = g_i(t+1)\left(\delta_{ij}\delta_{t\tau} + \sum_k J_{ik}G_{kj}(t,\tau) - \sum_{\tau<s<t}\hat{K}_i(t,s)G_{ij}(s,\tau)\right)$$

- to get $\mathcal{O}(N^3)$-update:

$$\chi_{ij}(t{+}1) = g_i(t{+}1)\left(\delta_{ij} + \sum_k J_{ik}\chi_{kj}(t) - \sum_{\tau<t}\hat{K}_i(t,\tau)\chi_{ij}(\tau)\right) .$$

# Memoryless for SK model



# Approximately memoryless

- Replace $\frac{\partial \phi_i(t)}{\partial \theta_i(\tau)} = 0$ with $\sum_{\tau<t} \frac{\partial \phi_i(t)}{\partial \theta_i(\tau)} = 0$:

$$\sum_k J_{ik} \chi_{ki}(t) - \sum_{\tau<t} \hat{K}_i(t,\tau) \chi_{ii}(\tau) = 0$$

- Leaves considerable freedom to choose $\hat{K}_i(t,\tau)$

# Approximately memoryless

- Replace $\frac{\partial \phi_i(t)}{\partial \theta_i(\tau)} = 0$ with $\sum_{\tau<t} \frac{\partial \phi_i(t)}{\partial \theta_i(\tau)} = 0$:

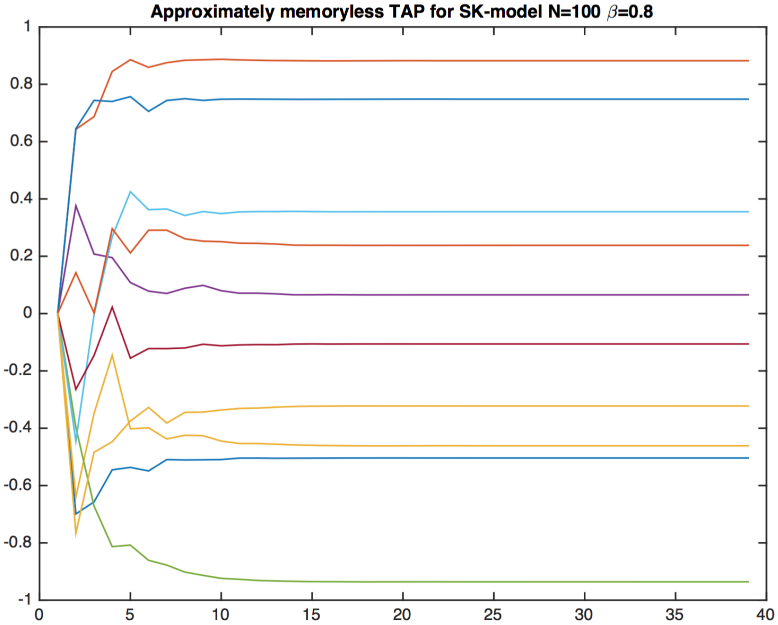
$$\sum_k J_{ik}\chi_{ki}(t) - \sum_{\tau<t} \hat{K}_i(t,\tau)\chi_{ii}(\tau) = 0$$

- Leaves considerable freedom to choose $\hat{K}_i(t,\tau)$
- Approximate single step memory: $\hat{K}_i(t,\tau) = 0$ for $\tau < t-1$
- Recursion simplifies:

$$\chi_{ij}(t+1) = g_i(t+1)\left(\delta_{ij} + \sum_k J_{ik}\chi_{kj}(t) - \hat{K}_i(t,t-1)\chi_{ij}(t-1)\right)$$

$$\hat{K}_i(t,t-1) = \frac{1}{\chi_{ii}(t-1)} \sum_k J_{ik}\chi_{ki}(t)$$

# Approximately memoryless for SK model



# Convergence to adaptive TAP

- If dynamics

$$\chi_{ij}(t+1) = g_i(t+1)\left(\delta_{ij} + \sum_k J_{ik}\chi_{kj}(t) - \hat{K}_i(t, t-1)\chi_{ij}(t-1)\right)$$

- converges:

$$g_i(t) \to g_i \qquad \chi_{ij}(t) \to \chi_{ij}$$

- then we recover the adaptive TAP response (= covariance)

$$\chi_{ij} = g_i\left(\delta_{ij} + \sum_k J_{ik}\chi_{kj} - v_i\chi_{ij}\right)$$

- with $\hat{K}_i(t, t-1) \to v_i$

# Simulations

- Cool simulations here

# Summary and outlook

- Questions: Convergence, parallel algorithms, marginal likelihood bound, assessing accuracy?
- Part 1: Sequential EP convergent for

$$\frac{\langle x^4 \rangle_{q_n}}{\langle x^2 \rangle_{q_n}^2} - 3 \in [-2, 0] \quad ?$$

- Other approaches to proofs?
- Part 2: New memoryless parallel updates.
- High complexity!