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?

I will address the first two questions!
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 \( \log R = \log(Z/Z_{\text{EP}}) \) 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
  - Asymptotics of memoryless dynamics
Example 1 - Gaussian process (GP) in a box

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

\[ K_{ij} = k(s_i, s_j) = \exp\left(- \frac{|s_i - s_j|}{\ell}\right), \quad \ell = 1 \]

\[
p_a(x) = \frac{1}{Z} \prod_n \mathbb{I}(|x_n| < a) \mathcal{N}(x; 0, K)
\]
Example 2 - Ising model

• Ising model

\[ p(x) = \frac{1}{Z} \prod_k \left[ \delta(x_k + 1) + \delta(x_k - 1) \right] \exp \left\{ x^\top J x / 2 + \theta^\top x \right\} . \]
Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

\[ p(x) = \frac{1}{Z} \prod_a f_a(x_a) \]
Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

\[ p(x) = \frac{1}{Z} \prod_a f_a(x_a) \]

- Exponential family approx., restricted Gaussian

\[ q(x) = \frac{1}{Z_q} \prod_a g_a(x_a), \quad g_a(x_a) = \exp \left( \gamma_a^T x_a - x_a^T \Lambda_a x_a / 2 \right) \]
Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:
  \[ p(x) = \frac{1}{Z} \prod_a f_a(x_a) \]

- Exponential family approx., restricted Gaussian
  \[ q(x) = \frac{1}{Z_q} \prod_a g_a(x_a), \quad g_a(x_a) = \exp \left( \gamma_a^T x_a - x_a^T \Lambda_a x_a / 2 \right) \]

- Tilted distribution tractable! Note subscript \( a \)
  \[ q_a(x) = \frac{1}{Z_a} \frac{q(x)f_a(x_a)}{g_a(x_a)} \]
Expectation propagation (EP) in a nutshell

- Model of interest has a certain factorization:

\[ p(x) = \frac{1}{Z} \prod_a f_a(x_a) \]

- Exponential family approx., restricted Gaussian

\[ q(x) = \frac{1}{Z_q} \prod_a g_a(x_a) , \quad g_a(x_a) = \exp \left( \gamma_a^T x_a - x_a^T \Lambda_a x_a / 2 \right) \]

- Tilted distribution tractable! Note subscript \( a \)

\[ q_a(x) = \frac{1}{Z_a} \frac{q(x)f_a(x_a)}{g_a(x_a)} \]

- Moment matching \( \phi_a(x_a) = [x_a \ x_a x_a^T]^T \) determines \( g_a(x_a) \):

\[ \langle \phi_a(x_a) \rangle_{q_a} = \langle \phi_a(x_a) \rangle_q \]
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

\[ q(\mathbf{x}) = \frac{1}{Z_q} \prod_a g_a(\mathbf{x}_a), \quad g_a(\mathbf{x}_a) = \exp \left( \gamma_a^T \mathbf{x}_a - \mathbf{x}_a^T \Lambda_a \mathbf{x}_a / 2 \right) \]

- 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 \( \phi_a(\mathbf{x}_a) = [\mathbf{x}_a \ \mathbf{x}_a \mathbf{x}_a^T]^T \) 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: \( Z \approx Z_{\text{EP}} = Z_q \prod_a Z_a \).
EP for Ising model

- Ising model

\[
p(\mathbf{x}) = \frac{1}{Z} \prod_k \left[ \delta(x_k + 1) + \delta(x_k - 1) \right] \exp \left\{ \mathbf{x}^\top \mathbf{J} \mathbf{x} / 2 + \mathbf{\theta}^\top \mathbf{x} \right\} f_k(x_k) f_0(\mathbf{x})
\]
EP for Ising model

• Ising model

\[ p(x) = \frac{1}{Z} \prod_k \left[ \delta(x_k + 1) + \delta(x_k - 1) \right] \frac{\exp\left\{ x^\top J x / 2 + \theta^\top x \right\}}{f_0(x)} \]

\[ f_k(x_k) \]

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

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

with \( \Sigma = (\Lambda - J)^{-1} \) and \( \mu = \Sigma(\gamma + \theta) \).
EP for Ising model

- Ising model

\[
p(x) = \frac{1}{Z} \prod_k \left[ \delta(x_k + 1) + \delta(x_k - 1) \right] f_k(x_k) \exp\left\{ \frac{x^\top J x}{2} + \theta^\top x \right\}.
\]

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

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

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

- Tilted distribution \( q_k(x_k) = \int q_k(x)dx_k \)

\[
q_k(x_k) = \frac{1}{Z_k} \frac{f_k(x_k)}{g_k(x_k)} \int q(x)dx_k \Rightarrow m_k = \tanh\left( \frac{\mu_k}{\Sigma_{kk}} - \gamma_k \right)
\]
EP algorithmic recipe

- Loop over $k$:
  1. Tilted distribution $q_k(x_k) = \int q_k(x) dx_k$

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

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 $\Sigma$ and $\mu = \Sigma(\gamma + \theta)$.  

Part 1 – EP convergence
EP for GP in a box

- **GP in a box**

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

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

  \[ f_n(x_n) = \mathbb{I}(|x_n| < a) \]
EP for GP in a box

- GP in a box

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

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

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

- Approximating factors \( g_0(x) = f_0(x) \) 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 \( q_n(x_n) = \int q_n(x) \, dx \setminus n \)
     \[
     q_n(x_n) = \frac{1}{Z_n} \frac{f_n(x_n)}{g_n(x_n)} q(x_n)
     \]
     \[
     = \frac{1}{Z_n} \mathbb{I}(|x_n| < a) \exp \left(-\frac{\lambda_n}{2} x_n^2 \right)
     \]
     \[
     \lambda_n \equiv \frac{1}{\Sigma_{nn}} - \Lambda_{nn} = \frac{1}{[(\Lambda - J)^{-1}]_{nn}} - \Lambda_{nn}
     \]
  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 \( \Sigma \).
Analysis of EP mapping

- The sequence of $N$ updates defines maps
  $$\Lambda_{nn} \leftarrow T_n(\Lambda)$$

- Fixed-point theorem: if map is differentiable in a neighborhood of $T(\Lambda^*) = \Lambda^*$ and
  $$\left| \frac{dT_N(\Lambda^*)}{d\Lambda_{NN}} \right| < 1$$
  then attraction is guaranteed.

- Use chain to calculate $T'_N \equiv \frac{dT_N}{d\Lambda_{NN}}$. 
Analysis of EP mapping cont.

- Zero mean EP:
  \[
  \Lambda_{nn} \leftarrow \frac{1}{\langle x^2 \rangle_q n} - \frac{1}{\Sigma_{nn}} + \Lambda_{nn} = F(\lambda_n)
  \]
  \[
  \lambda_n \equiv \frac{1}{\Sigma_{nn}} - \Lambda_{nn} = \frac{1}{[(\Lambda - J)^{-1}]_{nn}} - \Lambda_{nn}
  \]

- 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 \( \Sigma \), 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 $q_n(x_k) = \int q_n(x)dx \_n$
     
     \[ q_n(x_k) = \frac{1}{Z_n} f(x) - q(x) \]

  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 $\Sigma$ and $\mu = \Sigma(\gamma + \theta)$. 
Equivalence with adaptive TAP equations

- EP moment matching:
  \[ m = \mu, \quad 1 - m_k^2 = \Sigma_{kk}, \forall k \]

- with

  \[ m_k = \tanh \left( \frac{\mu_k}{\Sigma_{kk}} - \gamma_k \right) \]

  \[ \Sigma = (\Lambda - J)^{-1} \]

  \[ \mu = \Sigma (\gamma + \theta) \]
Equivalence with adaptive TAP equations

- EP moment matching:
  \[ m = \mu \quad 1 - m_k^2 = \Sigma_{kk}, \forall k \]

- with

  \[ m_k = \tanh \left( \frac{\mu_k}{\Sigma_{kk}} - \gamma_k \right) \]
  \[ \Sigma = (\Lambda - J)^{-1} \]
  \[ \mu = \Sigma (\gamma + \theta) \]

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

  \[ m = \mu = (\Lambda - J)^{-1} (\gamma + \theta) \quad \Leftrightarrow \quad -\gamma = \theta + (J - \Lambda)m \]
  \[ m_k = \tanh \left( \frac{\mu_k}{\Sigma_{kk}} - \gamma_k \right) = \tanh \left( [Jm]_k - \nu_k m_k + \theta_k \right) \]
  \[ \nu_k = \Lambda_{kk} - \frac{1}{\Sigma_{kk}} \]
Marginal distribution theorem

• Exact result for marginal distribution

\[ p(x_k) = \frac{1}{Z} f_k(x_k) \int \exp \left( \frac{1}{2} x^T J x + \theta^T x \right) \prod_{k' \neq k} f_{k'}(x_{k'}) \, dx_{\setminus k} \]

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

\[ \int \exp(x_k \sum_{k' \neq k} J_{kk', x_{k'}}) e^{\frac{1}{2} x_{\setminus k}^T J_{\setminus k} x_{\setminus k} + \theta_{\setminus k}^T x_{\setminus k}} \prod_{k' \neq k} f_{k'}(x_{k'}) \, dx_{\setminus k} \]

\[ \equiv h_k \]

\( \alpha p(x_{\setminus k} | J_{\setminus k}, \theta_{\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 \( \theta_k \)

\[ p(h_k) \equiv \int \delta(h_k - \sum_{k' \neq k} J_{kk', x_{k'}}) p(x_{\setminus k} | J_{\setminus k}, \theta_{\setminus k}) \, dx_{\setminus k} \]
Gaussian cavity assumption $\leadsto$ TAP

- Gaussian cavity assumption

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

- leads to **tilted distribution** form:

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

$$\approx \frac{1}{Z} f_k(x_k) e^{\frac{1}{2}(J_{kk} + \langle h_k^2 \rangle_k - \langle h_k \rangle_k^2)x_k^2 + (\langle h_k \rangle_k + \theta_k)x_k} \propto q_k(x_k)$$

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

$$m_k = \tanh((\langle h_k \rangle_k + \theta_k))$$

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

$$\langle h_k \rangle_k = [Jm]_k - v_k m_k$$

$$v_k = \langle h_k^2 \rangle_k - \langle h_k \rangle_k^2$$
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:
  \[ m_i(t + 1) = f \left( \phi_i(t) + \theta_i(t) \right) \]

\[
\phi_i(t) = \sum_j J_{ij} m_j(t) - \sum_{s < t} \hat{K}_{i}(t, s) m_i(s).
\]

- Set parameters \( \hat{K}_i(t, s) \) to remove memory.
- \( \theta_i(t) = \theta_i \) in actual dynamics
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:
  \[ m_i(t + 1) = f(\phi_i(t) + \theta_i(t)) \]
  \[ \phi_i(t) = \sum_j J_{ij} m_j(t) - \sum_{s< t} \hat{K}_i(t, s) m_i(s) . \]

- 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 function:
  \[ G_{ij}(t, \tau) \equiv \frac{\partial m_i(t)}{\partial \theta_j(\tau)} = \text{change in } m_i(t) \text{ due to change in } \theta_j(\tau) \]
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 \( g_i(t + 1) = \frac{\partial f(z)}{\partial z} \bigg|_{z=\phi_i(t)+\theta_i(t)} \)

- 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 \( O(N^3) \)-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

Approximately memoryless TAP for SK-model $N=100 \, \beta=0.8$
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:

\[ g_i(t) \to g_i \quad \chi_{ij}(t) \to \chi_{ij} \]

• 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

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

• Other approaches to proofs?

• Part 2: New memoryless parallel updates.

• High complexity!