Computer Science

CSC 665-1: Advanced Topics in Probabilistic Graphical Models

Expectation Propagation

Instructor: Prof. Jason Pacheco

Outline

- Belief Propagation Review

- Expectation propagation
  - Unnormalized Exponential Families
  - EP Algorithm

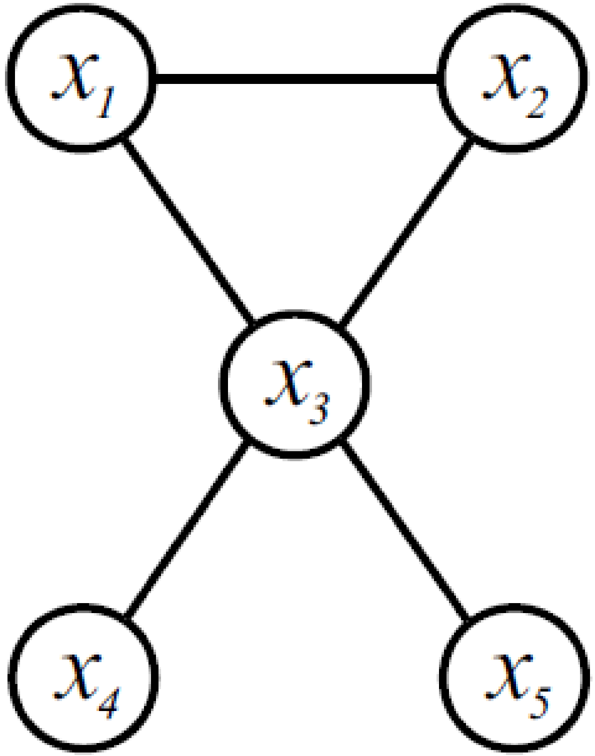
- Variational Optimization Perspective
  - Bethe Variational Problem
  - EP-Bethe Variational Problem

Pairwise MRF Notation

- Consider the pairwise Markov Random Field $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$p(x) = \frac{1}{Z} \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t)$$

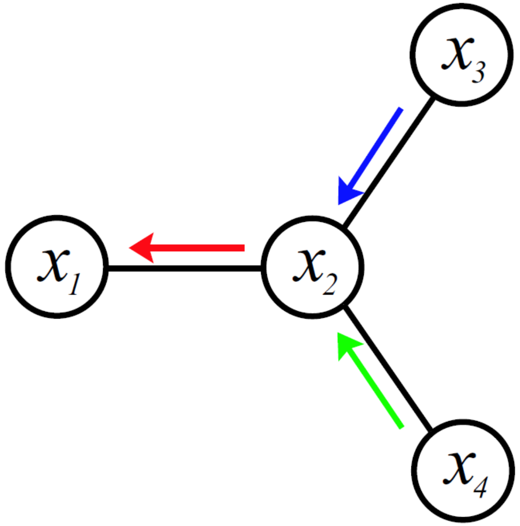
with vertices $\mathcal{V}$ and edges $\mathcal{E}$



- Without loss of generality we ignore observations *y*
- They can be considered as part of unary potentials $\psi_s(x_s, y)$

Message Passing

Global inference decomposes into local computations via graph structure…

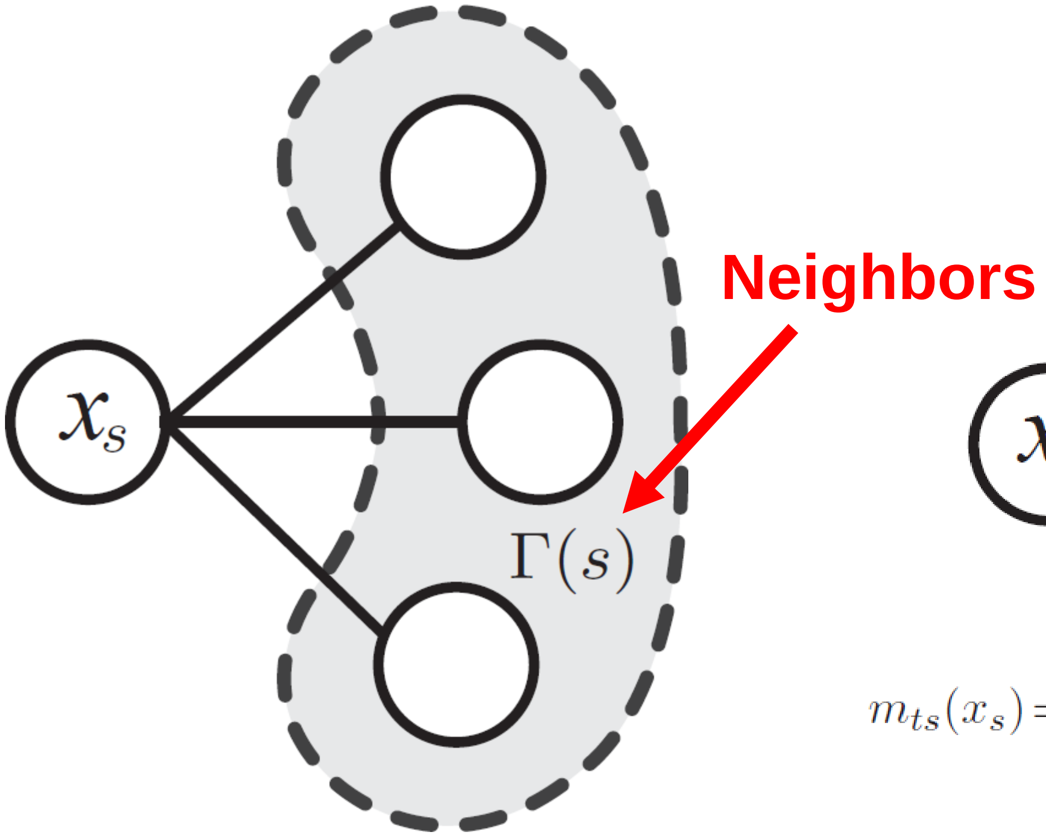


$$p(x_1) \propto \iiint \psi_1(x_1)\psi_{12}(x_1,x_2)\psi_2(x_2)\psi_{23}(x_2,x_3)\psi_3(x_3)\psi_{24}(x_2,x_4)\psi_4(x_4)\, dx_4\, dx_3\, dx_2$$

$$\propto \psi_1(x_1) \underbrace{\int \psi_{12}(x_1,x_2)\psi_2(x_2) \underbrace{\left[\int \psi_{23}(x_2,x_3)\psi_3(x_3)\, dx_3\right]}_{m_{32}(x_2)} \cdot \underbrace{\left[\int \psi_{24}(x_2,x_4)\psi_4(x_4)\, dx_4\right]}_{m_{42}(x_2)} dx_2}_{m_{21}(x_1) \propto \int \psi_{12}(x_1,x_2)\psi_2(x_2) m_{32}(x_2)\, m_{42}(x_2)\; dx_2}$$

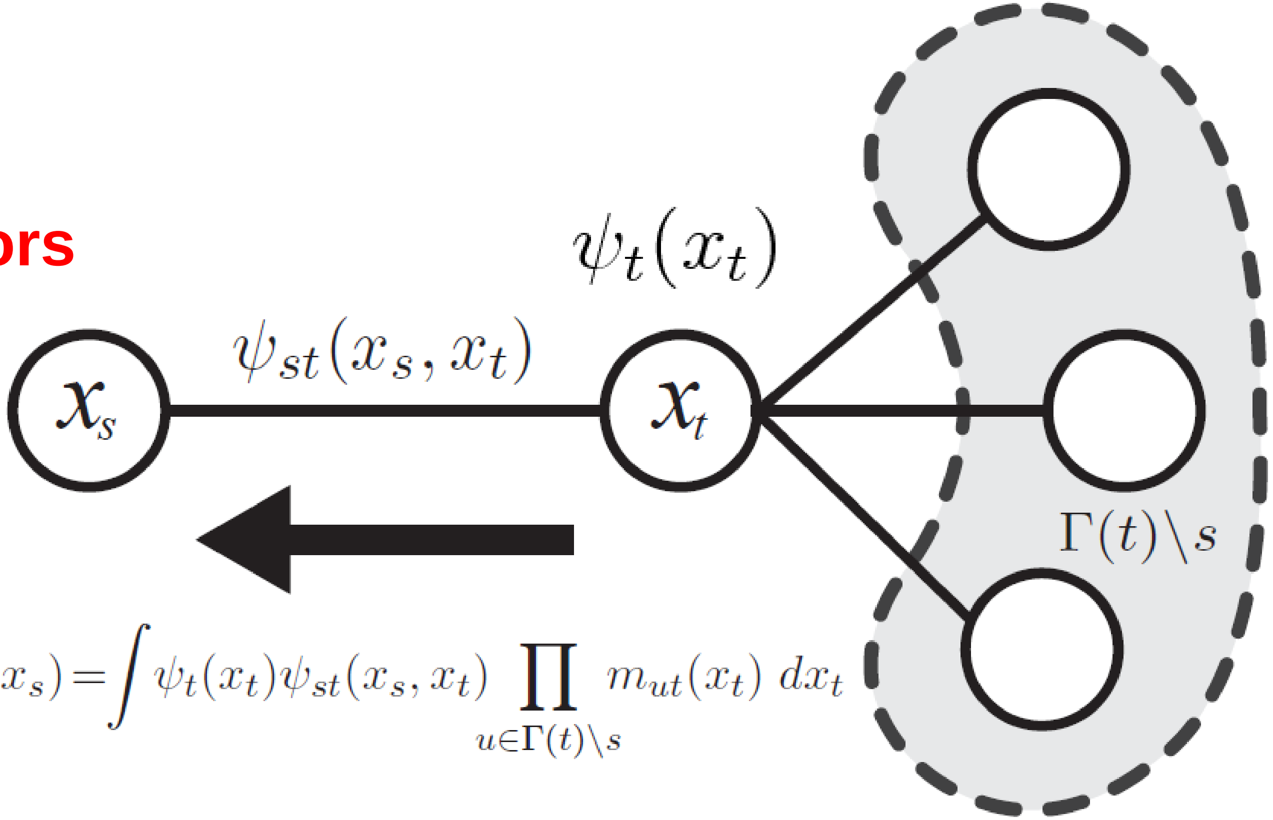
Belief Propagation (for Pairwise MRFs)

Marginal Belief



$$q_s(x_s) \propto \psi_s(x_s) \prod_{k \in \Gamma(s)} m_{ks}(x_s)$$

Message



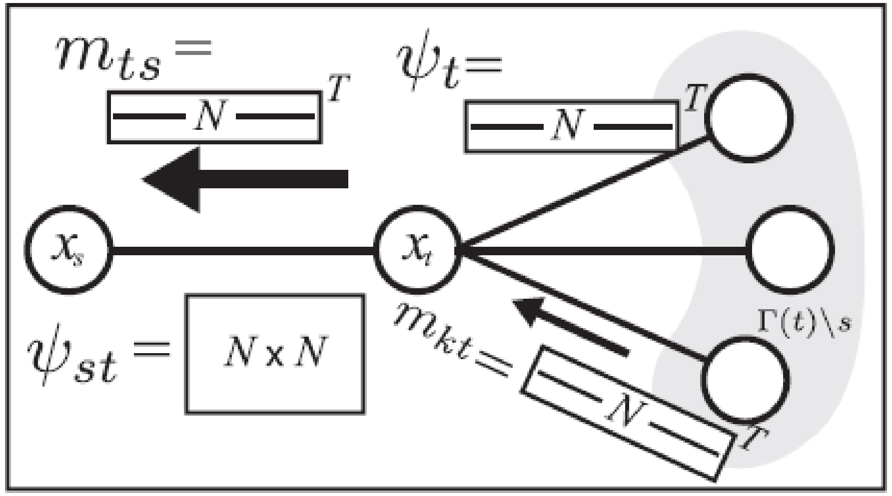
$$m_{ts}(x_s) = \int \psi_t(x_t) \psi_{st}(x_s, x_t) \prod_{u \in \Gamma(t) \backslash s} m_{ut}(x_t)\, dx_t$$

***Loopy* BP**: **Update messages in some order until fixed point (i.e. they stop changing)**

Belief Propagation

Discrete → Sum-Product

$$x \in \{1, \ldots, N\}^D$$



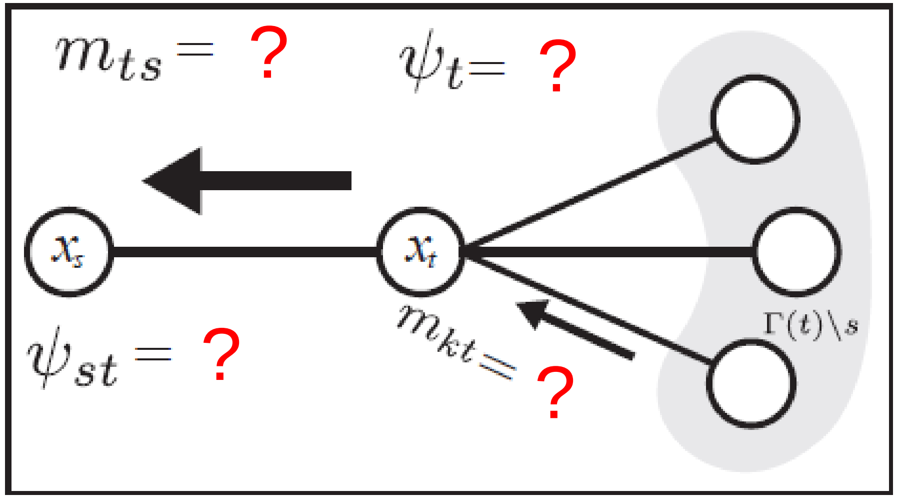
Message Update:

$$m_{ts} = \sum_{x_t} \boxed{\psi_{st}} \; \psi_t \prod m_{kt}$$

Matrix-vector multiplication

Continuous

$$x \in \mathcal{R}^D$$



Message Update:

$$m_{ts}(x_s) = \int_{\mathcal{X}_t} \psi_{st}(x_s, x_t)\psi_t(x_t) \prod_{k \in \Gamma(t)\backslash s} m_{kt}(x_t)\, dx_t$$

Integral requires conjugacy (e.g. jointly Gaussian)

(Unnormalized) Exponential Families

- We consider exponential family variational distributions:

$$q(x) = \exp\{\eta^T \phi(x) - A(\eta)\}$$

$$A(\eta) = \log \int \exp\{\eta^T \phi(x)\} \, h(dx)$$

- EP makes frequent use of an *unnormalized* version:

$$\mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta) = \exp\{\eta^T \phi(x)\}$$

- Messages and intermediate quantities are all in $\mathrm{ExpFam}^{\mathrm{U}}(\cdot)$
- **Note:** Members of $\mathrm{ExpFam}^{\mathrm{U}}(\cdot)$ need not be normalizable!

Unnormalized ExpFam Arithmetic

Multiplication is easy with natural parameters

$$\mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta) \cdot \mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta_2)$$

$$= \exp\{\eta_1^T \phi(x)\} \exp\{\eta_2^T \phi(x)\} = \exp\{(\eta_1 + \eta_2)^T \phi(x)\}$$

$$= \mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta_1 + \eta_2)$$

Multiplication → Add natural parameters

Division easy too!

$$\frac{\mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta_1)}{\mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta_2)} = \exp\{(\eta_1 - \eta_2)^T \phi(x)\}$$

$$= \mathrm{ExpFam}^{\mathrm{U}}(x \mid \eta_1 - \eta_2)$$

Division → Subtract natural parameters

Unnormalizability

- Products / ratios of $\mathrm{ExpFam}^{\mathrm{U}}(\cdot)$ may not be normalizable!
- **Example:** Consider two Gaussians

$$\mathcal{N}(x \mid 0, \Lambda_1^{-1}), \quad \text{where } \Lambda_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

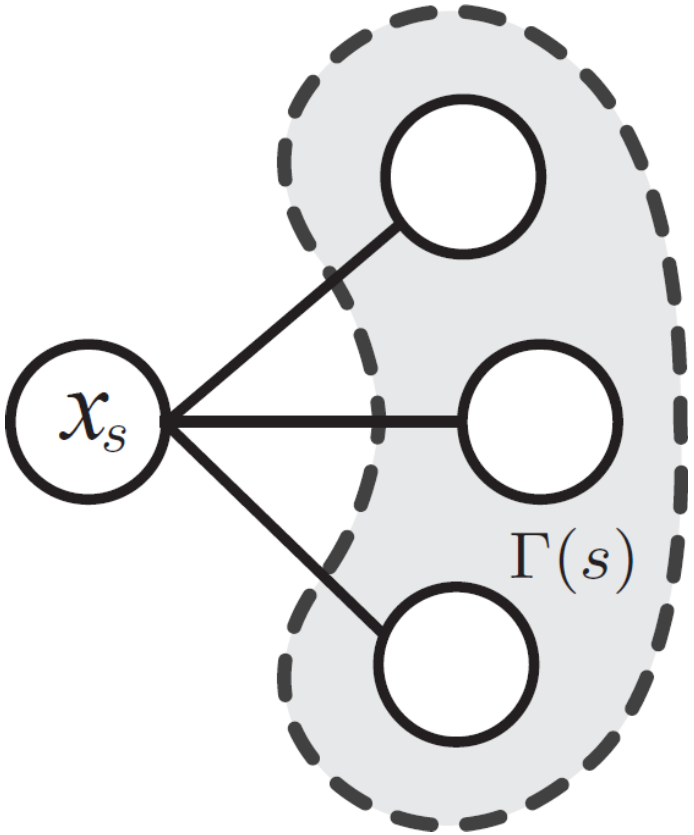
$$\mathcal{N}(x \mid 0, \Lambda_2^{-1}), \quad \text{where } \Lambda_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The ratio yields negative definite inverse covariance:

$$\frac{\mathcal{N}(x \mid 0, \Lambda_1^{-1})}{\mathcal{N}(x \mid 0, \Lambda_2^{-1})} \Rightarrow \Lambda_1 - \Lambda_2 = \begin{pmatrix} -0.5 & 0 \\ 0 & -0.5 \end{pmatrix}$$

Expectation Propagation (for Pairwise MRFs)

Marginal Belief



$$q_s(x_s) \propto \psi_s(x_s) \prod_{k \in \Gamma(s)} m_{ks}(x_s)$$

Step 1: Cavity Function

- Remove effect of <u>single message:</u>

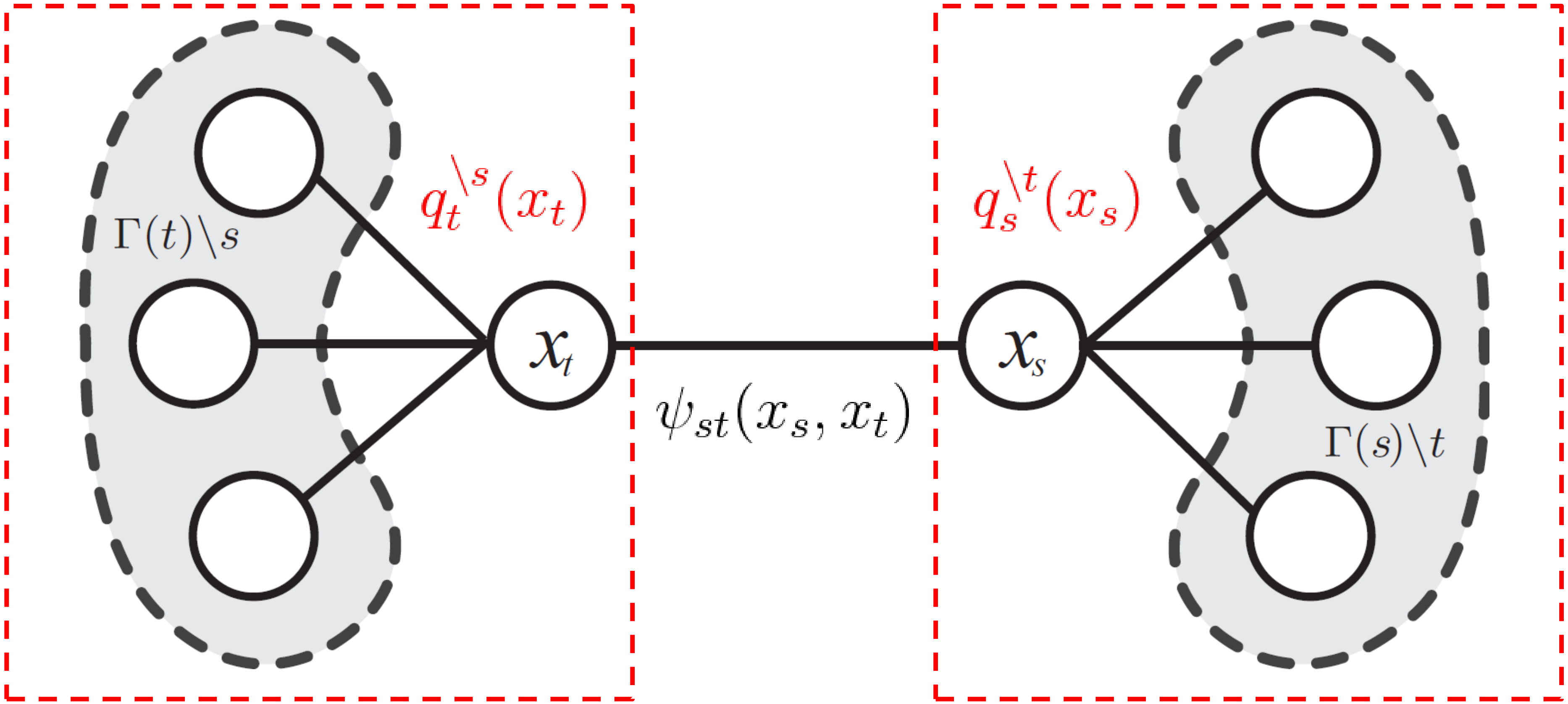
$$q_s^{\setminus t}(x_s) = \frac{q_s(x_s)}{m_{ts}(x_s)}$$

- Equivalent form:

$$q_s^{\setminus t}(x_s) = \psi_s(x_s) \prod_{k \in \Gamma(s) \setminus t} m_{ks}(x_s)$$

- Belief about x_s without knowledge from x_t
- Messages / cavity in $\mathrm{ExpFam}^{\mathrm{U}}(\cdot)$

Expectation Propagation



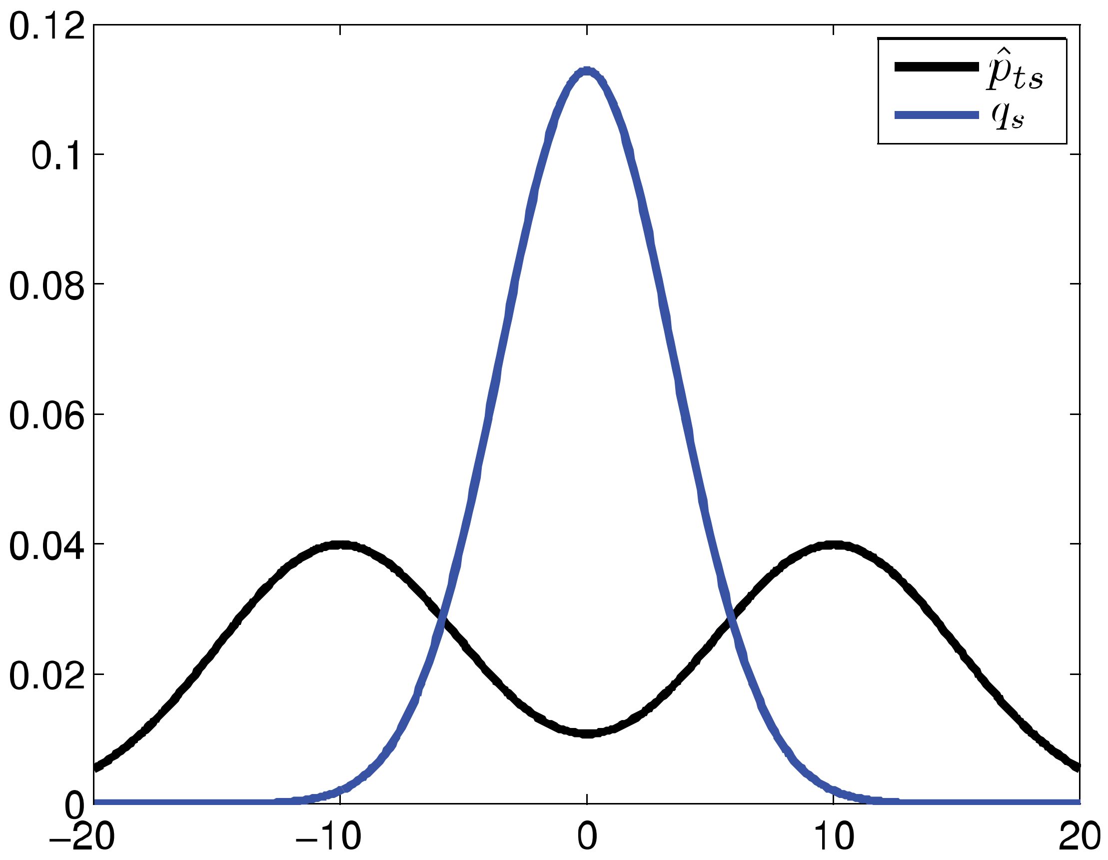
Step 2: Augmented Distribution

This must normalize!

$$\hat{p}_{ts}(x_s) \propto \int_{\mathcal{X}_t} q_t^{\backslash s}(x_t)\psi_{st}(x_s, x_t)q_s^{\backslash t}(x_s)\, dx_t$$

Expectation Propagation

Step 3: Kullback-Leibler Projection



$$q_s^{\text{new}}(x_s) = \arg\min_{q_s} \mathrm{KL}(\hat{p}_{ts} \| q_s)$$

Recall: For exponential families KL-projection solved via moment-matching:

$$\mathbb{E}_{q^{\text{new}}}[\phi(x)] = \mathbb{E}_{\hat{p}}[\phi(x)]$$

(e.g. for Gaussian q(x) match mean / variance)

Expectation Propagation

Step 4: Update message

- Message update is then:

$$m_{ts}^{\text{new}}(x_s) = \frac{\psi_s(x_s) m_{ts}^{\text{new}}(x_s) \prod_{k \in \Gamma(s) \setminus t} m_{ks}(x_s)}{\psi_s(x_s) \prod_{k \in \Gamma(s) \setminus t} m_{ks}(x_s)} = \boxed{\frac{q_s^{\text{new}}(x_s)}{q_s^{\setminus t}(x_t)}}$$

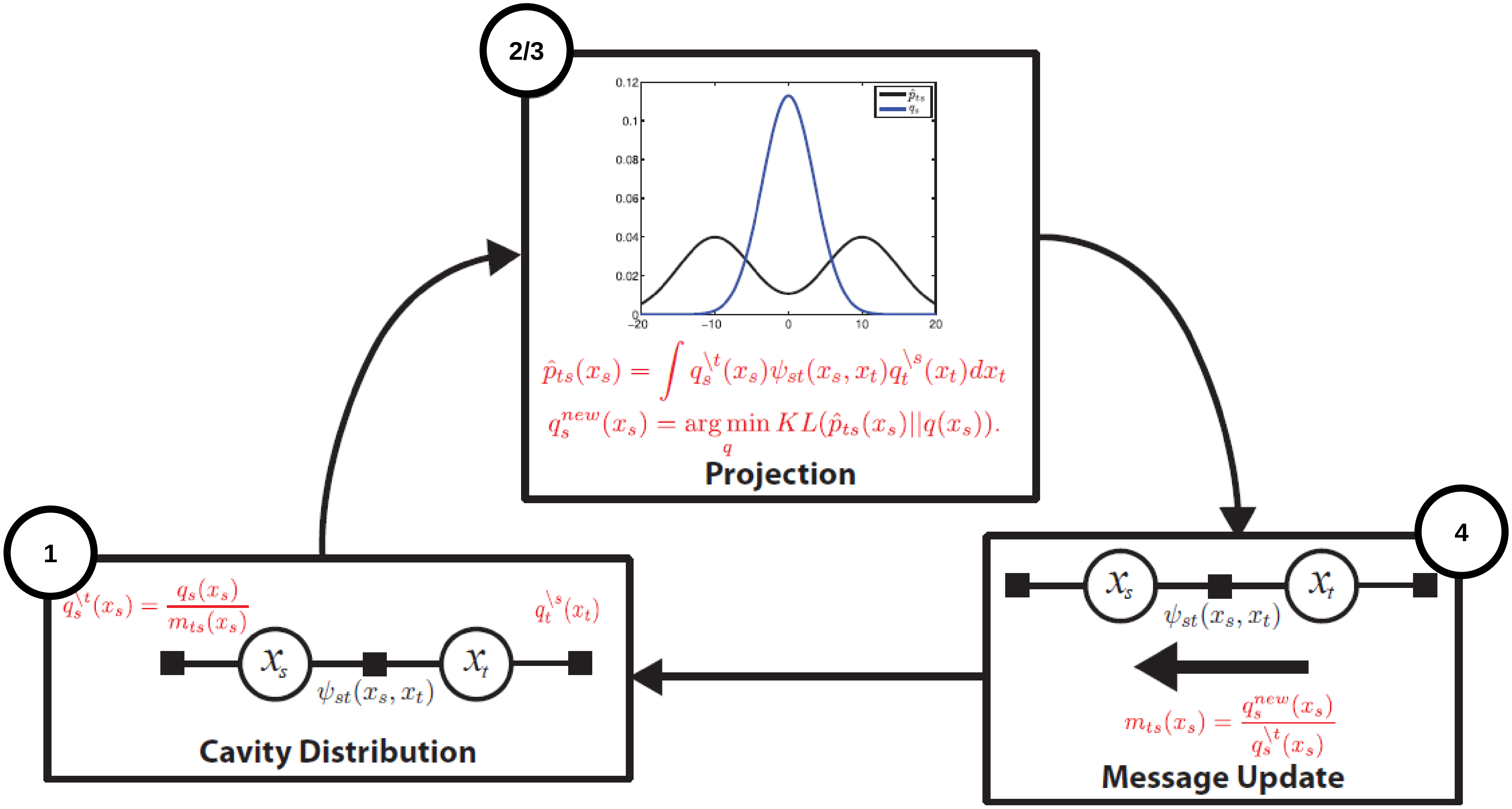
- Recall, we can do this by subtracting natural parameters

Repeat updates until fixed point reached:

$$m_{ts}^{\text{new}}(x_s) = m_{ts}^{\text{old}}(x_s)$$

For all messages

Expectation Propagation



Some Known Properties of EP

- When p(x) discrete or Gaussian then:
  - EP equivalent to BP
  - Guaranteed to converge (when p(x) is tree structured)
  - Exact marginal inference
- For non-discrete / non-Gaussian p(x)
  - Approximate, but computable updates
  - Not guaranteed to converge, even for tree-structured p(x)
- Only requires computable moments for each $\hat{p}_{ts}(x_s)$

Numerical Stability Issues

- EP updates require mean & natural parameter conversion:

$$\mu = \mathbb{E}[\phi(x)] \Leftrightarrow \eta$$

  - Moment-matching = Mean parameters
  - Multiplication / Division = Natural parameters
- Conversion can be numerically unstable (e.g. Gaussian):

$$\Sigma = \mathrm{Cov}(X) \Leftrightarrow \eta = \Sigma^{-1}$$

- Augmented distribution may become unnormalizable:

$$\mathrm{Cov}_{\hat{p}}(X) \prec 0$$

Care must be taken during implementation…

Variational Lower Bound

- Recall for any distribution $q(x)$ marginal likelihood equals:

$$\log p(y) = \underbrace{\mathrm{KL}[q(x) \| p(x \mid y)]}_{\text{Can't compute in general}} - \mathrm{KL}[q(x) \| p(x, y)]$$

- By Gibbs' inequality we have:

$$\mathrm{KL}[q(x) \| p(x \mid y)] \geq 0$$

- Variational lower bound is then:

$$\log p(y) \geq \max_{q \in \mathcal{Q}} -\mathrm{KL}[q(x) \| p(x, y)] \triangleq \mathcal{L}(q)$$

Variational Families

$$q(x) = \frac{1}{Z} e^{-E(x,y)}$$

$\mathcal{Q}^{\text{Gibbs}}$

$$q(x) = \prod_{s \in \mathcal{V}} q_s(x_s)$$

$\mathcal{Q}^{\text{MF}}$

$$\log p(y) = \max_{q \in \mathcal{Q}^{\text{Gibbs}}} \mathcal{L}(q) \geq \max_{q \in \mathcal{Q}^{\text{Bethe}}} \mathcal{L}(q) \geq \max_{q \in \mathcal{Q}^{\text{MF}}} \mathcal{L}(q)$$

$$q(x) \propto \prod_{s \in \mathcal{V}} q_s(x_s) \prod_{(s,t) \in \mathcal{E}} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

$\mathcal{Q}^{\text{Bethe}}$

Bethe Variational Problem

Minima of the Bethe variational problem correspond to BP fixed-points (log-messages = Lagrange multipliers)

$$\max_q \mathcal{L}(q)$$

$$\text{s.t.} \int q_s(x_s)\, dx_s = 1 \qquad \forall s \in \mathcal{V}$$

$$\int\int q_{st}(x_s, x_t)\, dx_s\, dx_t = 1 \qquad \forall (s,t) \in \mathcal{E}$$

Normalization

$$q_s(x_s) = \int q_{st}(x_s, x_t)\, dx_t \qquad \forall s \in \mathcal{V}, t \in \Gamma(s)$$

Local Marginal Consistency

EP Variational Problem

Minima of the Bethe variational problem correspond to **EP fixed-points** (log-messages = Lagrange multipliers)

$$\max_q \mathcal{L}(q)$$

$$\text{s.t.} \int q_s(x_s)\, dx_s = 1 \qquad \forall s \in \mathcal{V}$$

$$\int\int q_{st}(x_s, x_t)\, dx_s\, dx_t = 1 \qquad \forall (s,t) \in \mathcal{E}$$

Normalization

$$\mathbb{E}_{q_s}[\phi_s(x_s)] = \mathbb{E}_{\hat{p}_{ts}}[\phi_s(x_s)] \qquad \forall s \in \mathcal{V}, t \in \Gamma(s)$$

Local Expectation Consistency

Summary

- BP updates only computable for discrete or Gaussian models
- EP generalizes BP to models where moments of auxiliary distribution $\hat{p}_{ts}(x_s)$ can be computed (or estimated)
- EP = BP for discrete or Gaussian models (even loopy ones)
- EP optimizes Bethe variational problem over modified local consistency constraints