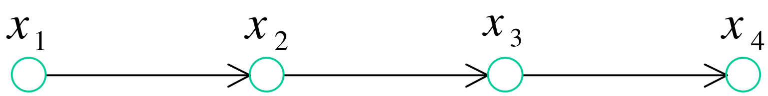
Expectation Propagation for approximate Bayesian inference

Tom Minka
Carnegie Mellon University

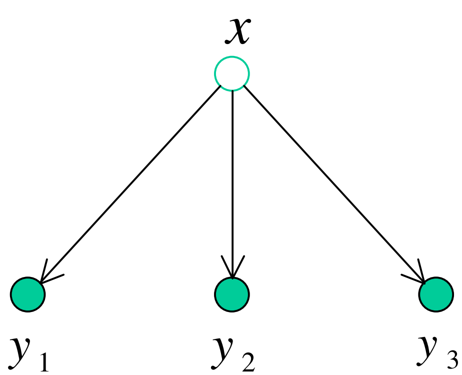
"An important open problem is to devise approximation schemes that are suitable for hybrid networks where Lauritzen's algorithm cannot be applied."

Lerner, Segal, and Koller, UAI'01

Graph topology is a red herring in general nets



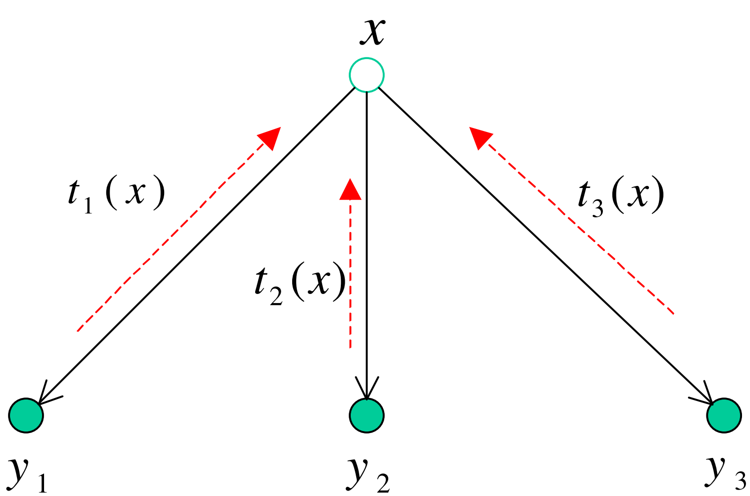
$$p(x_t \mid x_{t-1}) = 0.5N(x_{t-1}, \sigma^2) + 0.5N(x_{t-1}+1, \sigma^2)$$



$$p(y_i \mid x) = 0.5N(x, \sigma^2) + 0.5N(0, \sigma^2)$$

Both intractable!

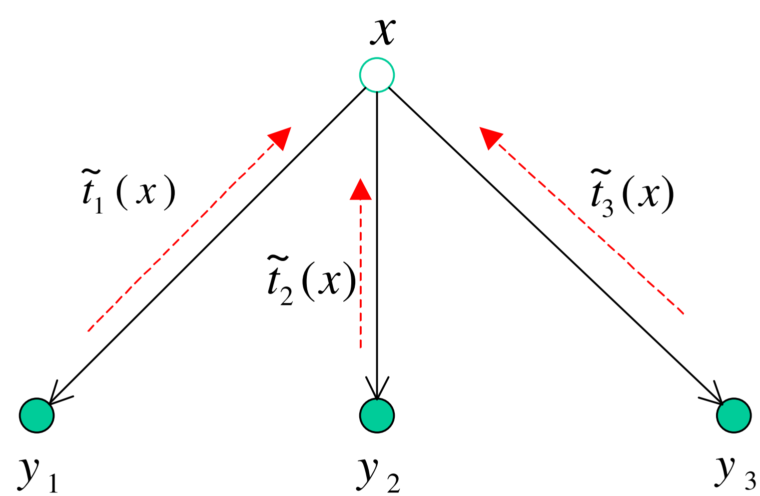
Belief Propagation



$$p(x \mid y_1, y_2, y_3) = \prod_i t_i(x)$$

Arbitrary function of x

Expectation Propagation



$$\tilde{t}_i(x) = N(m_i, v_i)$$

$$p(x \mid y_1, y_2, y_3) \approx \prod_i \tilde{t}_i(x)$$

Gaussian in x

(or other exponential family)

General problem: Approximate a distribution in product form

$$
\begin{aligned}
p(\mathbf{x}) = \quad & t_1(\mathbf{x}) \quad t_2(\mathbf{x}) \quad \cdots \quad t_n(\mathbf{x}) \\
\approx \quad & \tilde{t}_1(\mathbf{x}) \quad \tilde{t}_2(\mathbf{x}) \quad \cdots \quad \tilde{t}_n(\mathbf{x}) \quad = q(\mathbf{x})
\end{aligned}
$$

(exponential family)

Graphical models are special case, e.g. directed network:

$$p(\mathbf{x}) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_2) \cdots$$

Undirected network:

$$p(\mathbf{x}) = \Psi(x_1, x_2)\Psi(x_2, x_3)\Psi(x_3, x_4) \cdots$$

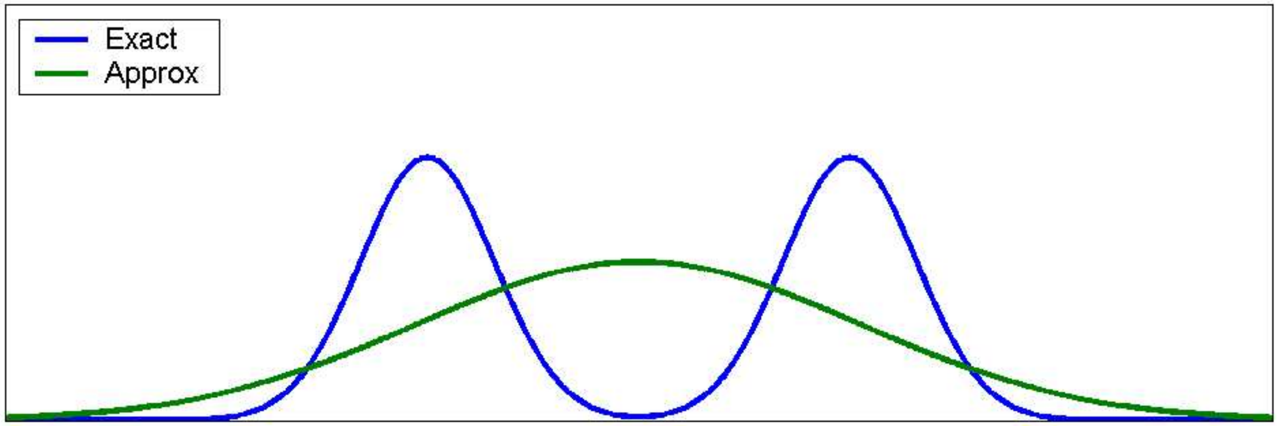
Naïve algorithm

Approximate each term in isolation:

$$\min_{\mathbf{x}} \int (t_i(\mathbf{x}) - \tilde{t}_i(\mathbf{x}))^2 \, d\mathbf{x} \qquad \text{(for example)}$$

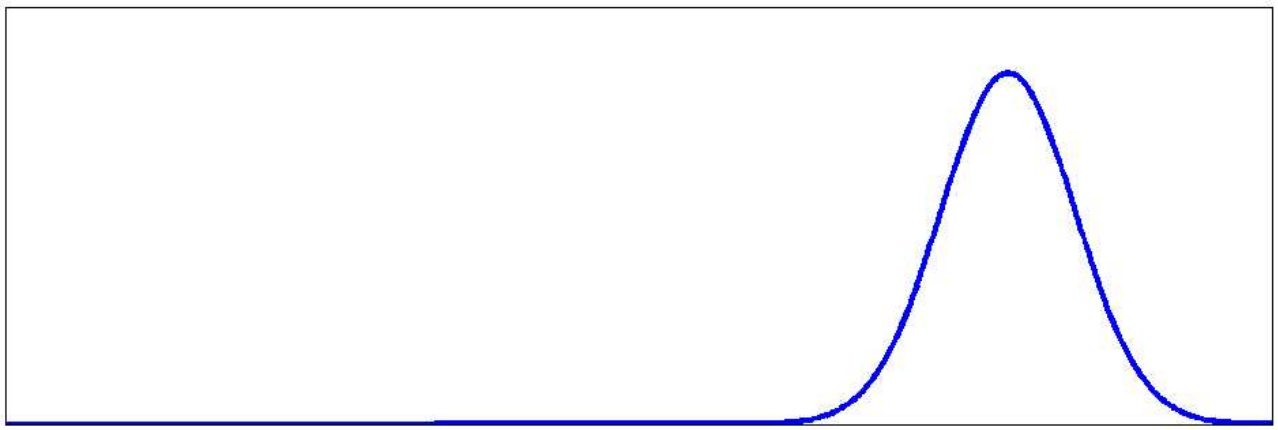
$\boxed{\text{Bad idea}}$

(naïve)



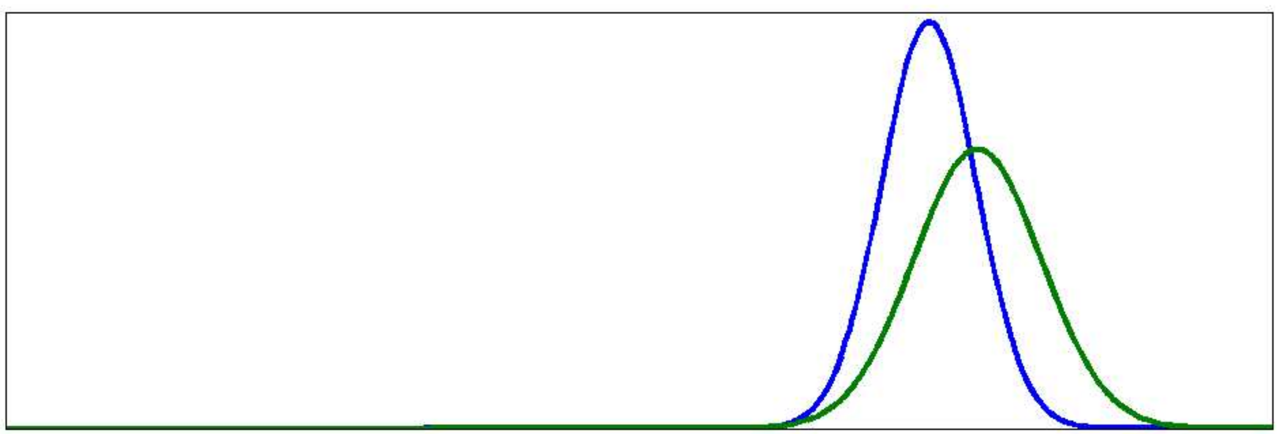
$t_1(x)$

$\times$

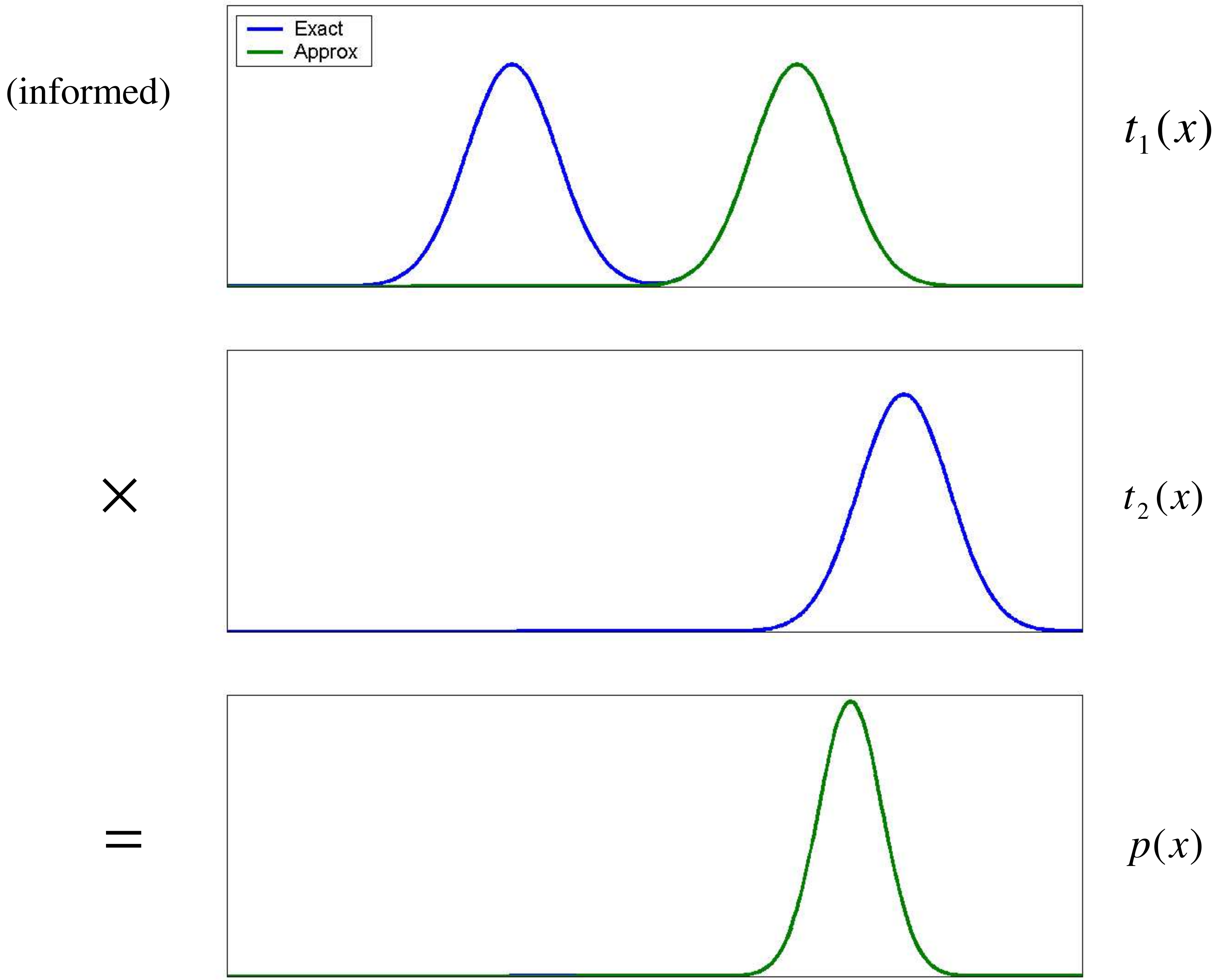


$t_2(x)$

$=$



$p(x)$


(informed)
Exact
Approx
$t_1(x)$
$\times$
$t_2(x)$
$=$
$p(x)$

Filtering algorithm

- One pass, left to right

- Make approximations *in context:* $q^{\setminus i}(\mathbf{x}) = \prod_{j<i} \tilde{t}_j(\mathbf{x})$

$$\min_{\tilde{t}_i} KL(q^{\setminus i}(\mathbf{x}) t_i(\mathbf{x}) \,\|\, q^{\setminus i}(\mathbf{x}) \tilde{t}_i(\mathbf{x}))$$

- Sensitive to ordering

KL minimization

- Factorized $\tilde{t}_i(\mathbf{x}) = \tilde{t}_{i1}(x_1)\tilde{t}_{i2}(x_2)\cdots\tilde{t}_{id}(x_d)$
  - min KL = preserve marginals
- Gaussian $\tilde{t}_i(\mathbf{x}) = \exp(\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + \mathbf{c})$
  - min KL = preserve moments
- Exponential family $\tilde{t}_i(\mathbf{x}) = \exp(\sum_j f_j(\mathbf{x})\lambda_j)$
  - min KL = preserve expectations $E[f_j(\mathbf{x})]$

Assumed-density filtering

- Generalizes Extended Kalman filtering (Kushner)
- Boyen-Koller, Factored Frontier
- “Moment matching” (West et al, 1985) (Bernardo&Giron,1988) (Spiegelhalter&Cowell,1992)
- “Online learning” (Opper,1999)
- “Inclusive trees” (Frey et al, 2000)
- “Gaussian fields” (Barber&Sollich,1999)

Expectation Propagation

- Round-robin algorithm

- Context is all other terms: $q^{\backslash i}(\mathbf{x}) = \prod_{j \neq i} \tilde{t}_j(\mathbf{x})$

- Coupled consistency conditions:

$$\tilde{t}_i(\mathbf{x}) = \arg\min\ KL(q^{\backslash i}(\mathbf{x}) t_i(\mathbf{x}) \,\|\, q^{\backslash i}(\mathbf{x}) \tilde{t}_i(\mathbf{x}))$$

Loopy BP = Factorized EP

- Directed net : $t_i(\mathbf{x})$ is CPT for family i
- Context is approx marginals, excluding messages from i:

$$q^{\setminus i}(x_j) = q(x_j) / \tilde{t}_{ij}(x_j)$$

- Message to x_k from family i :

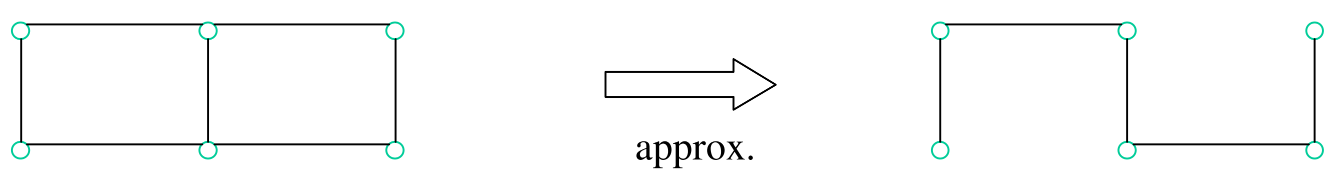
$$\arg\min \; KL(q^{\setminus i}(\mathbf{x}) t_i(\mathbf{x}) \,\|\, q^{\setminus i}(\mathbf{x}) \tilde{t}_i(\mathbf{x}))$$

$$\Rightarrow \tilde{t}_{ik}(x_k) = \sum_{\mathbf{x} \setminus x_k} t_i(\mathbf{x}) \prod_{j \neq k} q^{\setminus i}(x_j)$$

Insights

1. Loopy belief propagation is a round-robin factorization of each CPT in the context of other factorizations.

2. Using EP, other approximations are possible (not completely factorized)



Benchmark results

Dataset	EP	Billiard	SVM
Heart	.203	.207	.232
Thyroid	.037	.037	.053
Ionosphere	.099	.113	.115
Sonar	.140	.147	.129

Future work/Open questions

- Different networks
- Different approximating families
  - Mixture approximations
- Alternatives to KL
- Alternative fixed-point schemes
- Error estimates