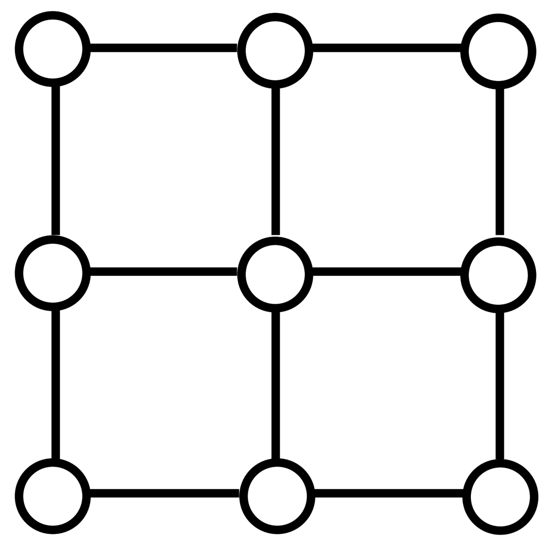
Mean Field Networks

Yujia Li, Richard S. Zemel

UNIVERSITY OF TORONTO

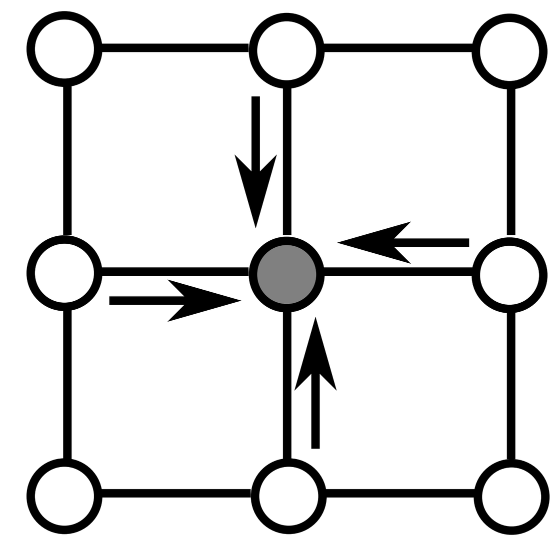
Pairwise MRF



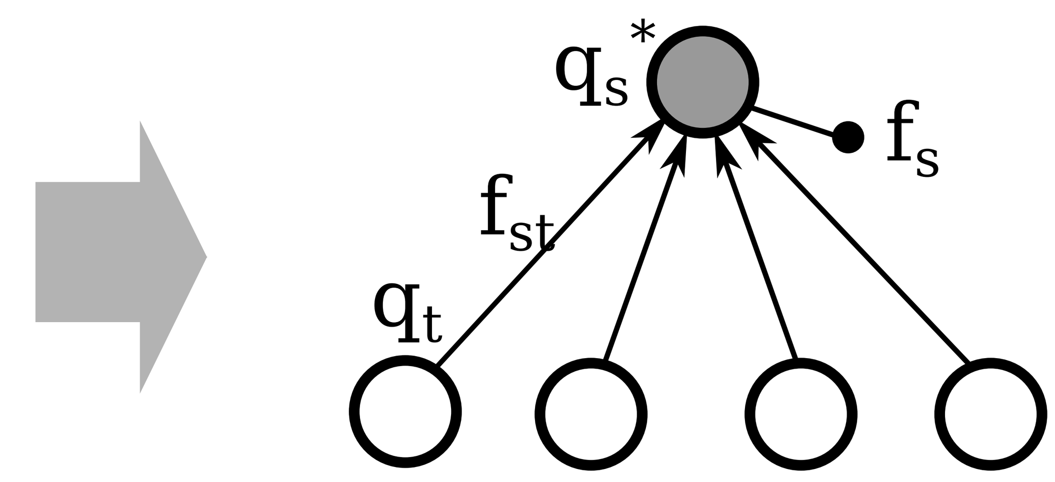
$$p(\mathbf{x}) = \frac{1}{Z}\exp(E(\mathbf{x};\theta))$$

$$E(\mathbf{x};\theta) = \sum_{s\in\mathcal{V}} f_s(x_s;\theta) + \sum_{(s,t)\in\mathcal{E}} f_{st}(x_s, x_t;\theta)$$

Mean Field Inference



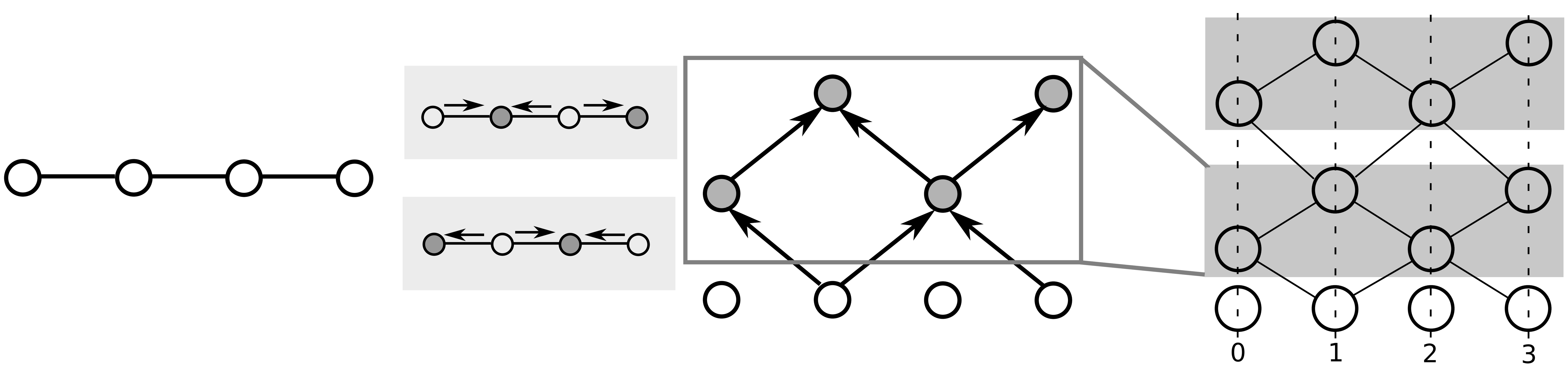
Feed Forward Operation



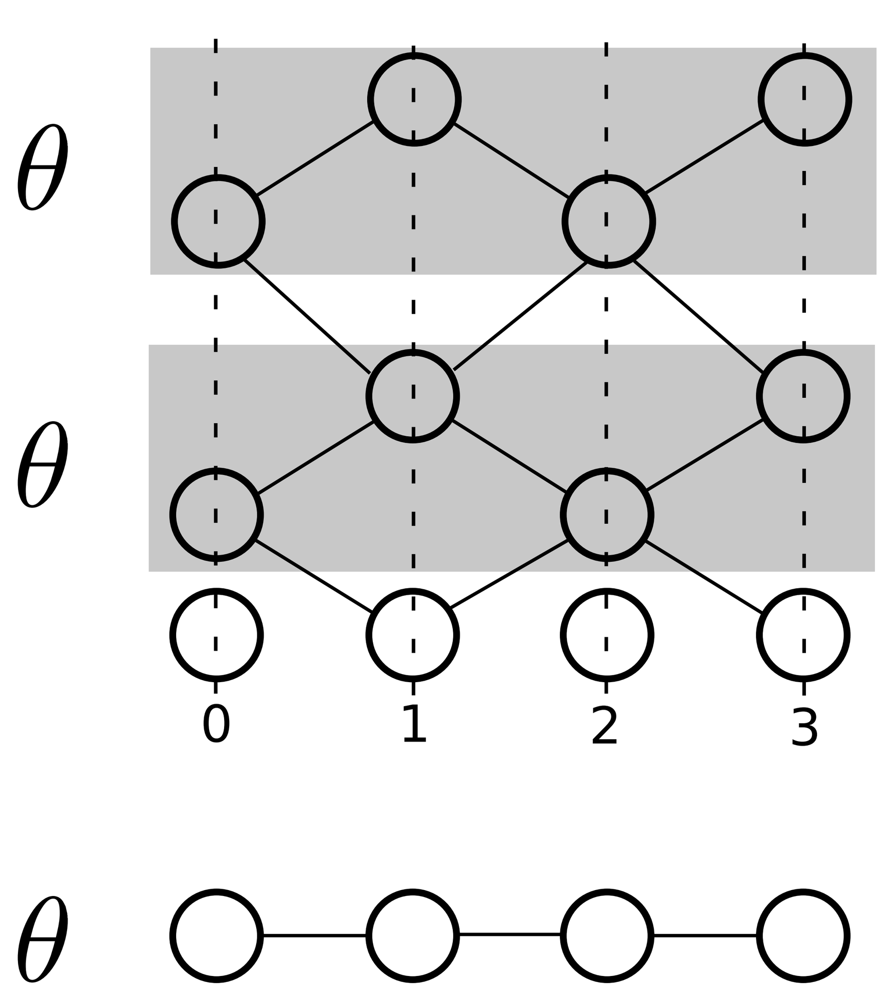
Minimize KL($q||p$) for factorial q

$$q_s^*(x_s) = \frac{1}{Z_s}\exp\left(f_s(x_s;\theta) + \sum_{t\in\mathcal{N}(s)}\sum_{x_t} q_t(x_t) f_{st}(x_s, x_t;\theta)\right)$$

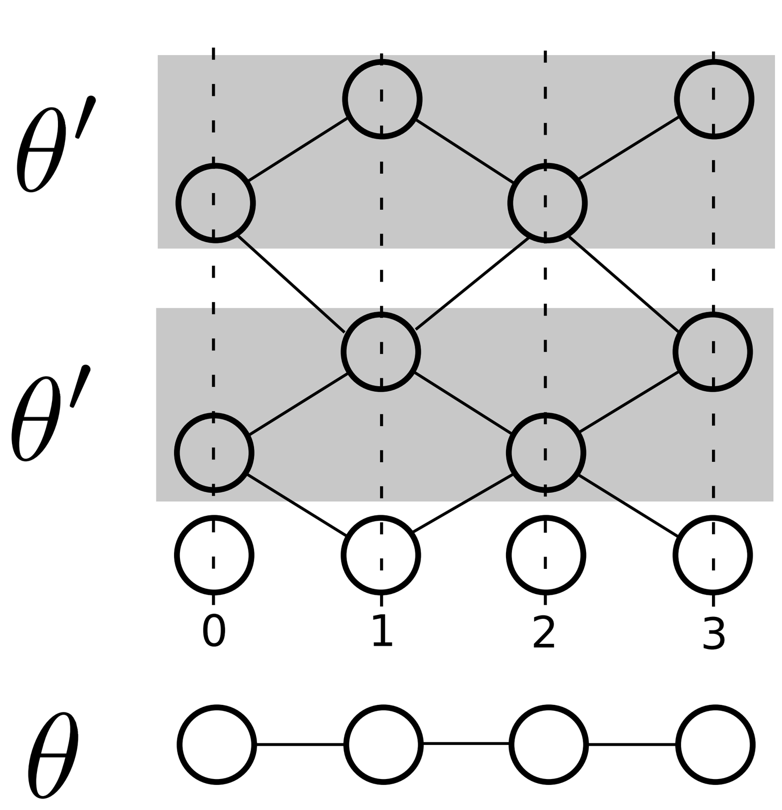
Mean Field Network



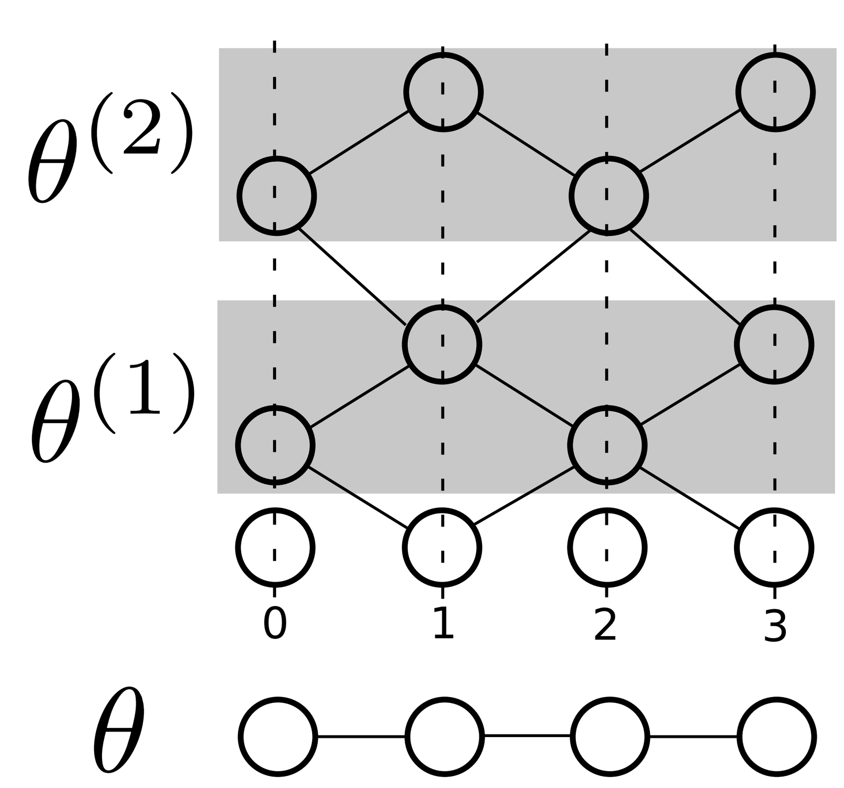
Restrictions and Relaxations



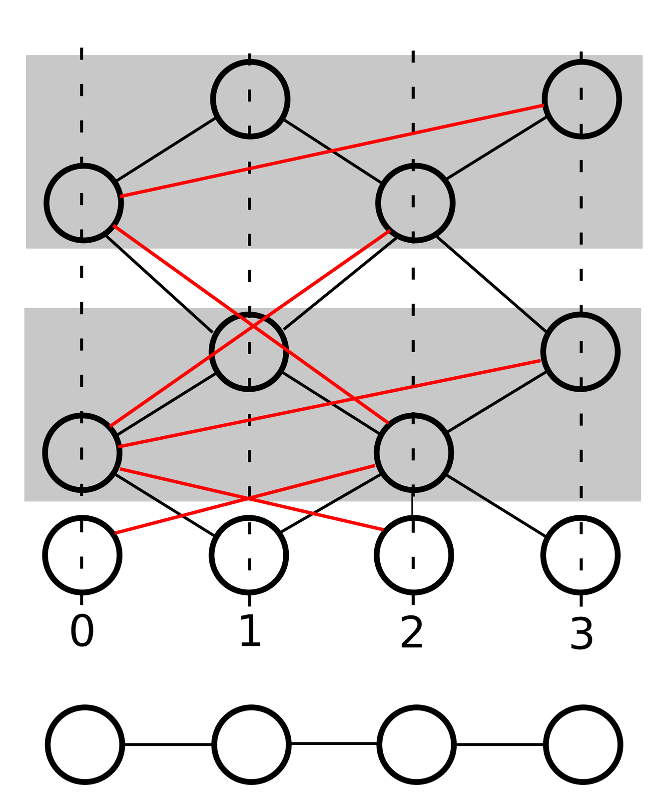
(1) Parameters in the network tied with the graphical model.



(2) Parameters on all layers tied together.



(3) Structure of the network tied with the graphical model.



Mean Field Networks as Stand-Alone Models

- A range of loss functions can be used
- Hinge loss, other more structured task loss
- Better models when trained discriminatively

Experiments

Denoising/Binary Labeling

Base model: pairwise CRF trained with mean field inference runned for 30 steps.

MFNs as Inference Tools

MF-1	MF-3	MF-10	MF-30
-12779.05	-12881.50	-12904.43	-12908.54
MFN-1	MFN-3	MFN-10	MFN-30
-12837.87	-12893.52	-12908.80	-12909.34

KL($q||p$)

MFNs as Stand-Alone Models

MF-30	MF-3	MFN-3-t	MFN-3
81.09%	80.65%	81.34%	81.51%

Test accuracy, MFN trained with element-wise hinge loss