Probabilistic Graphical Models

Variational inference III: Mean-Field

Siamak Ravanbakhsh

Fall 2019

Learning objectives

- naive mean-field method
  - its derivation as I-projection
  - its update equations

Previously: variational inference

$$D(q\|p) = \sum_{\mathbf{x}} q(\mathbf{x})(\ln q(x) - \ln p(x))$$

$$= -H(q) - \mathbb{E}_q[\sum_{i,j} \ln \phi_{i,j}(x_i, x_j)] + \ln Z$$

entropy

expected energy

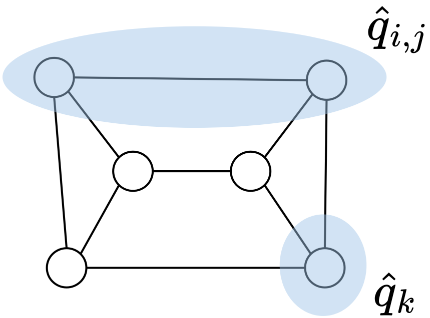
ignore: does not depend on q

negative of variational free energy

the optimization is defined such that

- marginals of interest can be read from $q \in \mathcal{Q}$
- entropy and expected energy are easy to calculate/approximate

Previously: loopy BP



$$\arg\min_q D(q||p)$$

$$p(x) = \frac{1}{Z}\prod_k \phi_{i,j}(x_i, x_j)$$

$\mathcal{Q}$

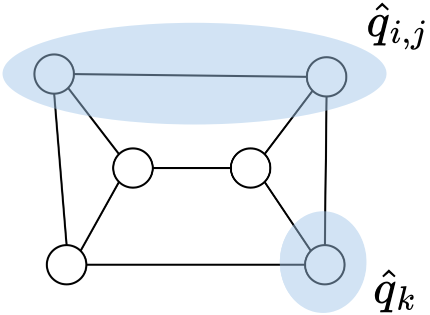
$$q(x) \propto \frac{\prod_{i,j\in\mathcal{E}} \hat{q}_{i,j}(x_i,x_j)}{\prod_i \hat{q}_i(x_i)^{|Nb_i|-1}}$$

such that $\sum_{x_i} \hat{q}_{i,j}(x_i, x_j) = \hat{q}_j(x_j) \quad \forall i,j \in \mathcal{E}, x_j$

Previously: loopy BP

$\arg\min_q D(q||p)$

$p(x) = \frac{1}{Z} \prod_k \phi_{i,j}(x_i, x_j)$



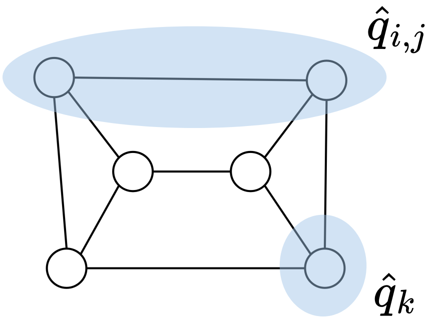
$\mathcal{Q}$

$$q(x) \propto \frac{\prod_{i,j \in \mathcal{E}} \hat{q}_{i,j}(x_i, x_j)}{\prod_i \hat{q}_i(x_i)^{|Nb_i|-1}}$$

such that $\sum_{x_i} \hat{q}_{i,j}(x_i, x_j) = \hat{q}_j(x_j) \quad \forall i, j \in \mathcal{E}, x_j$

- pseudo-marginals can be read from q

Previously: loopy BP



$$\arg\min_q D(q||p)$$

$$p(x) = \frac{1}{Z}\prod_k \phi_{i,j}(x_i, x_j)$$

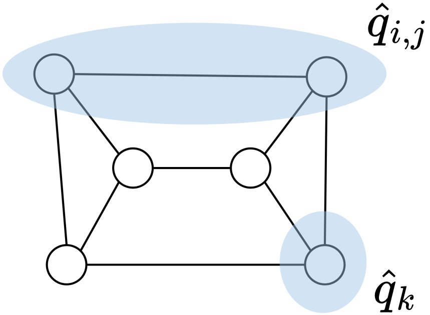
$\mathcal{Q}$

$$q(x) \propto \frac{\prod_{i,j\in\mathcal{E}} \hat{q}_{i,j}(x_i,x_j)}{\prod_i \hat{q}_i(x_i)^{|Nb_i|-1}}$$

such that $\sum_{x_i} \hat{q}_{i,j}(x_i, x_j) = \hat{q}_j(x_j) \quad \forall i,j \in \mathcal{E}, x_j$

- pseudo-marginals can be read from q
- entropy term is approximated (Bethe approximation)

Previously: loopy BP



$$\arg\min_q D(q\|p)$$

$$p(x) = \frac{1}{Z}\prod_k \phi_{i,j}(x_i, x_j)$$

$\mathcal{Q}$

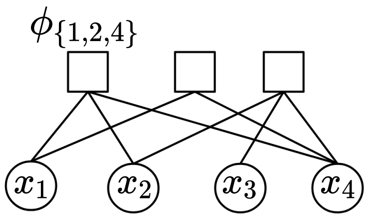
$$q(x) \propto \frac{\prod_{i,j\in\mathcal{E}} \hat{q}_{i,j}(x_i,x_j)}{\prod_i \hat{q}_i(x_i)^{|Nb_i|-1}}$$

such that $\sum_{x_i} \hat{q}_{i,j}(x_i, x_j) = \hat{q}_j(x_j) \quad \forall i,j \in \mathcal{E}, x_j$

- pseudo-marginals can be read from q
- entropy term is approximated (Bethe approximation)
- expected energy term is exact

Naive mean-field: objective

I-project p into the product form



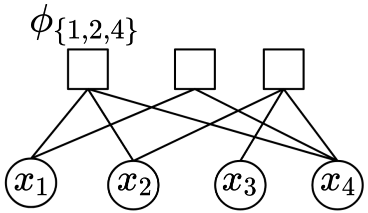
$$\arg\min_{q \in \mathcal{Q}} D(q \| p)$$

$$p(x) = \frac{1}{Z} \prod_I \phi_I(x_I)$$

a subset of vars.

$$q(x) = \prod_i q_i(x_i)$$

Naive mean-field: objective



I-project p into the product form

$$\arg\min_{q\in\mathcal{Q}} D(q\|p)$$

$$p(x) = \frac{1}{Z}\prod_I \phi_I(x_I)$$

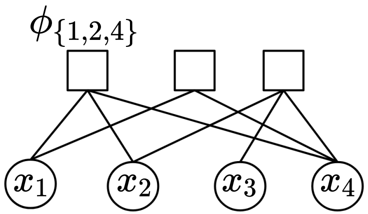
a subset of vars.

$$q(x) = \prod_i q_i(x_i)$$

$$= \arg\min_{q\in\mathcal{Q}} -H(q) - \mathbb{E}_q[\sum_I \ln \phi_I(x_I))] + \ln Z$$

$$\sum_i H(q_i) \qquad \sum_I \sum_{x_I} (\prod_{i\in I} q_i(x_i)) \ln \phi_I(x_I)$$

Naive mean-field: objective



I-project p into the product form

$$\arg\min_{q \in \mathcal{Q}} D(q \| p)$$

$$p(x) = \frac{1}{Z} \prod_I \phi_I(x_I)$$

a subset of vars.

$$q(x) = \prod_i q_i(x_i)$$

$$= \arg\min_{q \in \mathcal{Q}} -H(q) - \mathbb{E}_q[\sum_I \ln \phi_I(x_I))] + \ln Z$$

$$\sum_i H(q_i) \qquad \sum_I \sum_{x_I} (\prod_{i \in I} q_i(x_i)) \ln \phi_I(x_I)$$

- both terms are *tractable* for family Q
- the objective is non-convex
- lower-bound on Z (this is not true for loopy BP, where both entropy and constraints were approximations)

Is I-projection the right choice in MF?

M-projection of p into a q with **factorized** form $q(x) = \prod_k q(x_k)$

gives $q^M(x) = \prod_k p(x_k)$

Proof $D(p\|q) = \mathbb{E}_p[\ln p(x)] - \sum_k \mathbb{E}_p[\ln q(x_k)]$

$= \mathbb{E}_p[\ln \frac{p(x)}{\prod_k p(x_k)}] + \sum_k \mathbb{E}_p[\ln \frac{p(x_k)}{q(x_k)}]$

$= D(p\|q^M) + \sum_k D(p(x_k)\|q(x_k))$

minimized when this is zero! $q = q^M$

Is I-projection the right choice in MF?

M-projection of p into a q with **factorized** form $q(x) = \prod_k q(x_k)$

gives $q^M(x) = \prod_k p(x_k)$

calculating the M-projection:

$$D(p \| \prod_k q(x_k)) = -\mathbb{E}_p[\ln q] + const. = -\mathbb{E}_p[\sum_k \ln q_k] + const. = -\sum_k \mathbb{E}_{p_k}[\ln q_k] + const.$$

Is I-projection the right choice in MF?

M-projection of p into a q with **factorized** form $q(x) = \prod_k q(x_k)$

gives $q^M(x) = \prod_k p(x_k)$

calculating the M-projection:

$$D(p \| \prod_k q(x_k)) = -\mathbb{E}_p[\ln q] + const. = -\mathbb{E}_p[\sum_k \ln q_k] + const. = -\sum_k \mathbb{E}_{p_k}[\ln q_k] + const.$$

...but we are projecting into product form to obtain p_i in the first place

we do I-projection because it is easier

Optimization

objective: $\arg\max_q \sum_i H(q_i) + \sum_I \left(\prod_{i \in I} q_i(x_i)\right) \ln \phi_I(x_I)$

(block) coordinate descent:

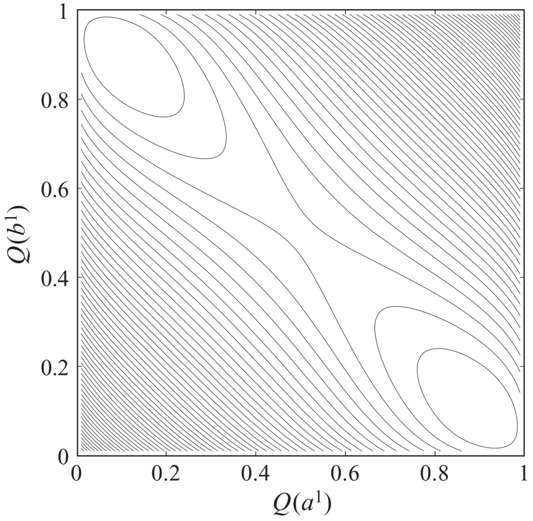
- optimize qi, one at a time
- non-convex
- *stable* convergence points are *local* optima

Optimization

objective: $\arg\max_q \sum_i H(q_i) + \sum_I (\prod_{i \in I} q_i(x_i)) \ln \phi_I(x_I)$

(block) coordinate descent:

- optimize qi, one at a time
- non-convex
- *stable* convergence points are *local* optima



Example

$$p(a,b) = \mathbb{I}(a=b)(.5-\epsilon) + \mathbb{I}(a \neq b)\epsilon$$

Derivation of updates

objective: $\arg\max_q \sum_i H(q_i) + \sum_I \sum_{x_I} \left(\prod_{i\in I} q_i(x_i)\right) \ln \phi_I(x_I)$

optimizing individual $q_i(x_i)$

$$\arg\max_{q_i} H(q_i) + \sum_{I|i\in I} \sum_{x_I} \left(\prod_{j\in I} q_j(x_j)\right) \ln \phi_I(x_I) + C(q_{-i})$$

only the Markov blanket of "i" appears

ignore: terms that do not depend on qi

Derivation of updates

objective: $\arg\max_q \sum_i H(q_i) + \sum_I \sum_{x_I} \left(\prod_{i \in I} q_i(x_i)\right) \ln \phi_I(x_I)$

optimizing individual $q_i(x_i)$

$$\arg\max_{q_i} H(q_i) + \sum_{I|i \in I} \sum_{x_I} \left(\prod_{j \in I} q_j(x_j)\right) \ln \phi_I(x_I) + C(q_{-i})$$

only the Markov blanket of "i" appears

ignore: terms that do not depend on qi

$$\arg\max_{q_i} H(q_i) + \sum_{x_i} q_i(x_i) \left[\sum_{I|i \in I} \sum_{x_{I-i}} \left(\prod_{j \in I, j \neq i} q_j(x_j)\right) \ln \phi_I(x_I)\right]$$

$\ln g(x_i)$

Derivation of updates

objective: $\arg\max_q \sum_i H(q_i) + \sum_I \sum_{x_I} \left(\prod_{i \in I} q_i(x_i)\right) \ln \phi_I(x_I)$

optimizing individual $q_i(x_i)$

$$\arg\max_{q_i} H(q_i) + \sum_{I|i \in I} \sum_{x_I} \left(\prod_{j \in I} q_j(x_j)\right) \ln \phi_I(x_I) + C(q_{-i})$$

only the Markov blanket of "i" appears

ignore: terms that do not depend on qi

$$\arg\max_{q_i} H(q_i) + \sum_{x_i} q_i(x_i) \underbrace{\left[\sum_{I|i \in I} \sum_{x_{I-i}} \left(\prod_{j \in I, j \neq i} q_j(x_j)\right) \ln \phi_I(x_I)\right]}_{\ln g(x_i)}$$

$$\arg\max_{q_i} -D\left(q_i(x_i) \| \frac{1}{\sum_{x_i} g(x_i)} g(x_i)\right)$$

minimized by $q_i(x_i) \propto g(x_i)$

Derivation of updates

objective: $\arg\max_q \sum_i H(q_i) + \sum_I \sum_{x_I} (\prod_{i \in I} q_i(x_i)) \ln \phi_I(x_I)$

optimizing individual $q_i(x_i)$

$$\arg\max_{q_i} H(q_i) + \sum_{I|i \in I} \sum_{x_I} (\prod_{j \in I} q_j(x_j)) \ln \phi_I(x_I) + C(q_{-i})$$

only the Markov blanket of "i" appears

ignore: terms that do not depend on qi

$$\arg\max_{q_i} H(q_i) + \sum_{x_i} q_i(x_i) \left[\sum_{I|i \in I} \sum_{x_{I-i}} (\prod_{j \in I, j \neq i} q_j(x_j)) \ln \phi_I(x_I)\right] \ln g(x_i)$$

$$\arg\max_{q_i} -D(q_i(x_i) \| \frac{1}{\sum_{x_i} g(x_i)} g(x_i))$$ minimized by $q_i(x_i) \propto g(x_i)$

closed form update: $q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}} \sum_{I|i \in I} \ln \phi_I(x_I))$

expectation is wrt the current estimate qj for the markov blanket of i

Closed form update

for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}} \textstyle\sum_{I|i\in I} \ln \phi_I(x_I))$$

$$q_{MB(i)}(x_{MB(i)}) = q_j(x_j)q_k(x_k)q_l(x_l)q_m(x_m)q_n(x_n)q_o(x_o)$$

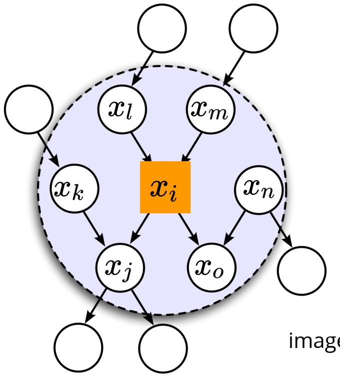


image: wikipedia

three factors that involve xi:

$$\phi_{i,l,m}(x_i, x_l, x_m) = p(x_i \mid x_l, x_m)$$

$$\phi_{i,n,o}(x_i, x_n, x_o) = p(x_o \mid x_i, x_n)$$

$$\phi_{i,j,k}(x_i, x_j, x_k) = p(x_j \mid x_i, x_k)$$

Closed form update

for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}} \textstyle\sum_{I|i \in I} \ln \phi_I(x_I))$$

$$q_{MB(i)}(x_{MB(i)}) = q_j(x_j)q_k(x_k)q_l(x_l)q_m(x_m)q_n(x_n)q_o(x_o)$$

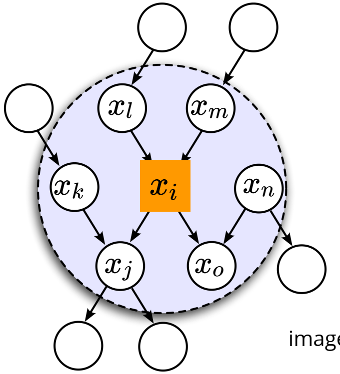


image: wikipedia

three factors that involve xi:

$$\phi_{i,l,m}(x_i, x_l, x_m) = p(x_i \mid x_l, x_m)$$

$$\phi_{i,n,o}(x_i, x_n, x_o) = p(x_o \mid x_i, x_n)$$

$$\phi_{i,j,k}(x_i, x_j, x_k) = p(x_j \mid x_i, x_k)$$

- initialize qi (random or uniform)
- iteratively update qi for each node i
- until convergence

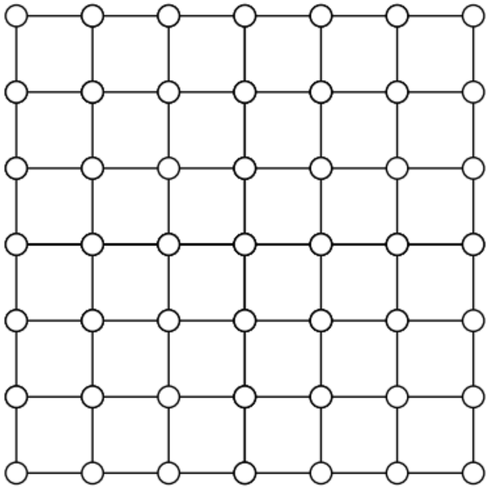
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

local field



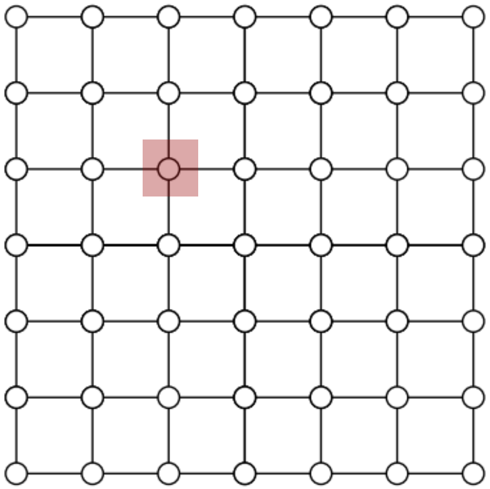
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

local field



for each node i:

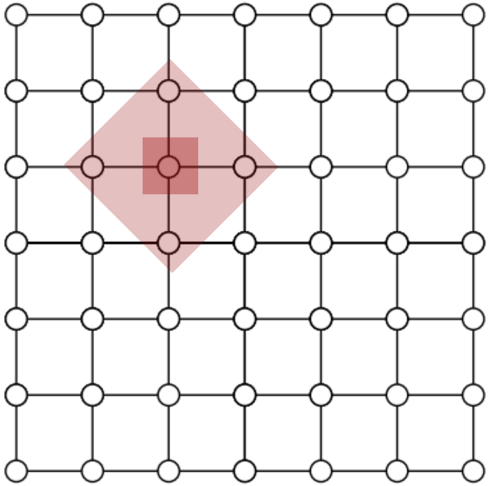
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$$x_i \in \{-1, +1\}$$

local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

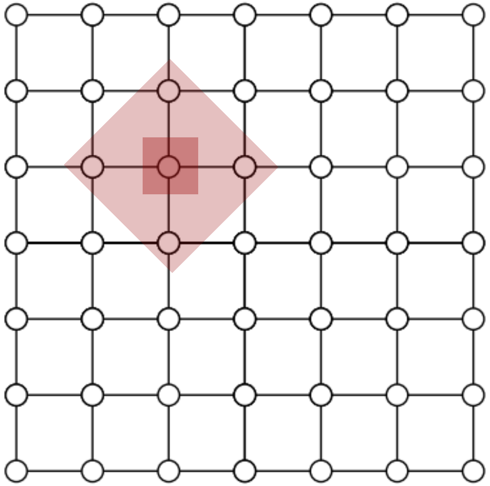
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

h_i → local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$$x_i \in \{-1, +1\}$$

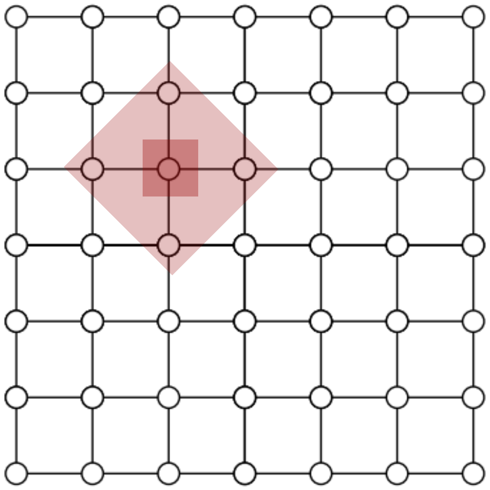
local field

for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$



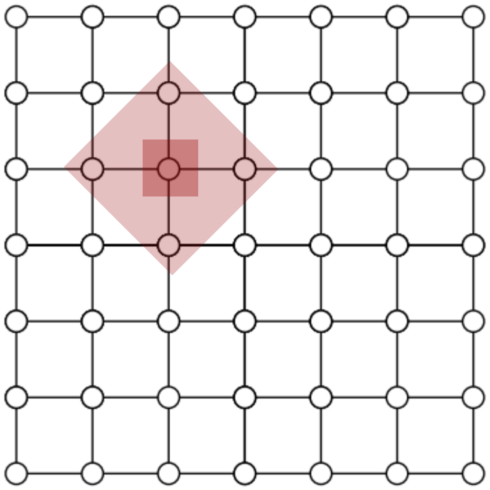
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mu_j))$$

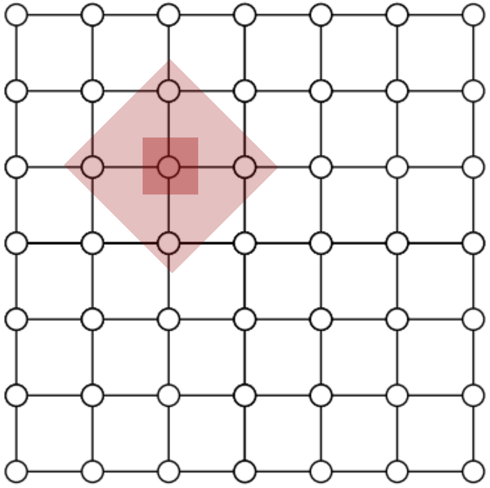
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$$x_i \in \{-1, +1\}$$

local field (h_i)



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mu_j))$$

mean-field! m_i

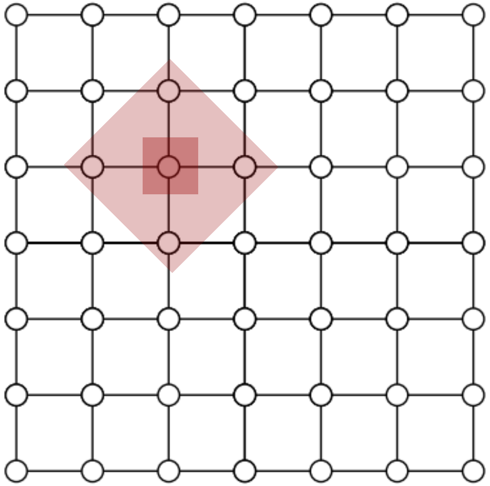
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mu_j))$$

mean-field! m_i

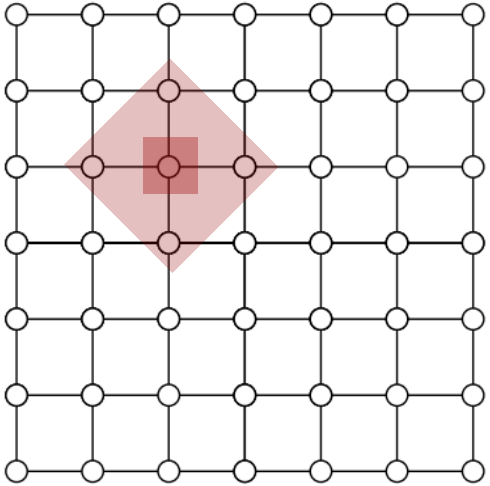
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$x_i \in \{-1, +1\}$

local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mu_j))$$

mean-field! m_i

$$q_i(x_i = +1) = \frac{\exp(m_i)}{\exp(m_i) + \exp(-m_i)} = \sigma(2m_i)$$

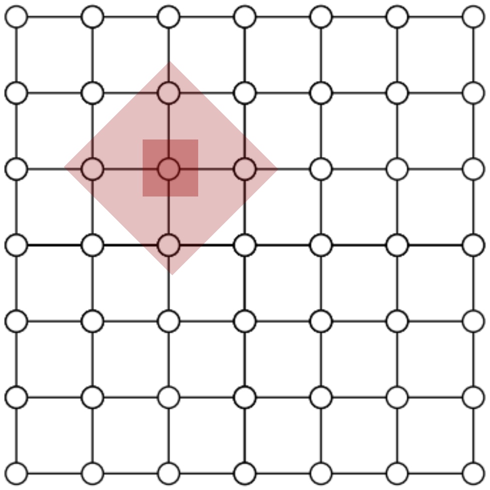
Example: MF in Ising grid

recall the Ising model:

$$p(x) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$$

$$x_i \in \{-1, +1\}$$

local field



for each node i:

$$q_i(x_i) \propto \exp(\mathbb{E}_{q_{MB(i)}}[h_i x_i + \sum_j x_i x_j J_{i,j}])$$

$$= \exp(x_i(h_i + \mathbb{E}_{q_{MB(i)}} \sum_j x_j J_{i,j}))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mathbb{E}_{q_j}[x_j]))$$

$$= \exp(x_i(h_i + \sum_j J_{i,j} \mu_j))$$

mean-field! m_i

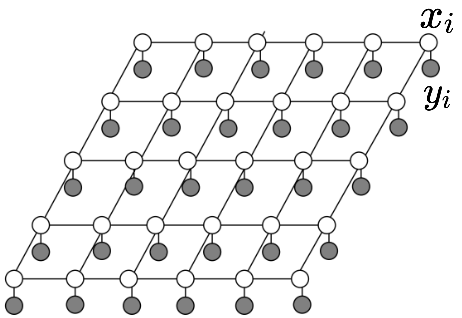
$$q_i(x_i = +1) = \frac{\exp(m_i)}{\exp(m_i) + \exp(-m_i)} = \sigma(2m_i)$$

$$\mu_i = q_i(x_i = +1) - q_i(x_i = -1)$$

Example: MF in the Ising grid

apply MF to image denoising

prior $p(x) \propto \exp(\sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$ where $J_{i,j} > 0$



iter. 1 iter. 3 iter. 15

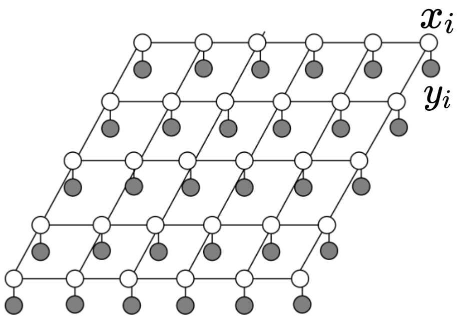
image credit: Murphy

Example: MF in the Ising grid

apply MF to image denoising

prior $p(x) \propto \exp(\sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$ where $J_{i,j} > 0$

likelihood $p(y \mid x) \propto \exp(\sum_i x_i y_i J_i)$ where $J_i > 0$



iter. 1 iter. 3 iter. 15

image credit: Murphy

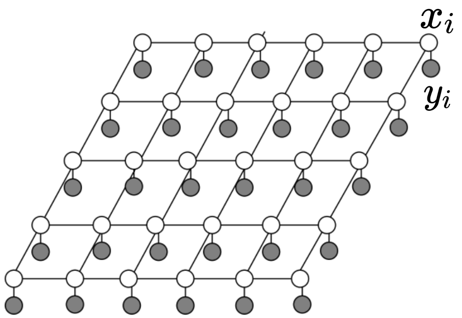
Example: MF in the Ising grid

apply MF to image denoising

prior $p(x) \propto \exp(\sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$ where $J_{i,j} > 0$

likelihood $p(y \mid x) \propto \exp(\sum_i x_i y_i J_i)$ where $J_i > 0$

posterior $p(x \mid y) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$



iter. 1 iter. 3 iter. 15

image credit: Murphy

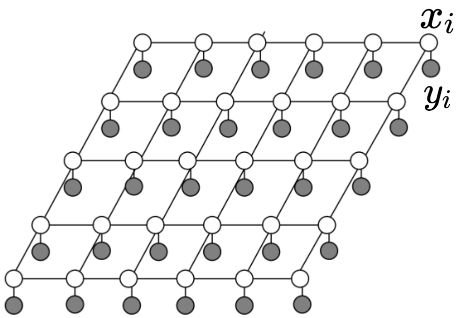
Example: MF in the Ising grid

apply MF to image denoising

prior $p(x) \propto \exp(\sum_{i,j\in\mathcal{E}} x_i x_j J_{i,j})$ where $J_{i,j} > 0$

likelihood $p(y \mid x) \propto \exp(\sum_i x_i y_i J_i)$ where $J_i > 0$

posterior $p(x \mid y) \propto \exp(\sum_i x_i h_i + \sum_{i,j\in\mathcal{E}} x_i x_j J_{i,j})$



iter. 1 iter. 3 iter. 15

image credit: Murphy

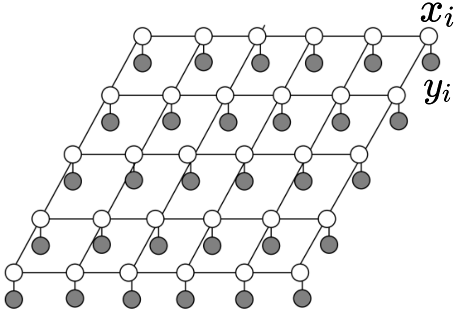
Example: MF in the Ising grid

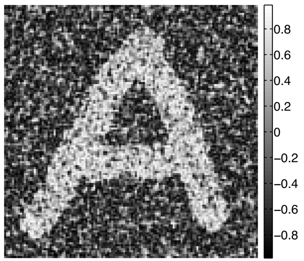
apply MF to image denoising

prior $p(x) \propto \exp(\sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$ where $J_{i,j} > 0$

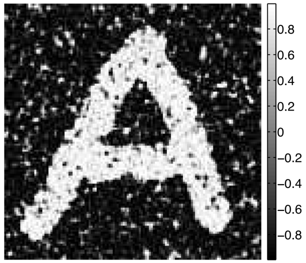
likelihood $p(y \mid x) \propto \exp(\sum_i x_i y_i J_i)$ where $J_i > 0$

posterior $p(x \mid y) \propto \exp(\sum_i x_i h_i + \sum_{i,j \in \mathcal{E}} x_i x_j J_{i,j})$

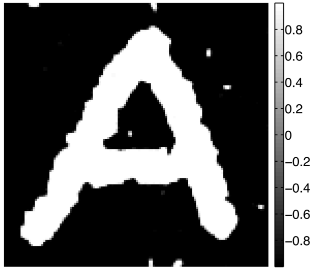




iter. 1



iter. 3



iter. 15

image credit: Murphy

Example: MF for multivariate Gaussian

Recap

mean parametrization $p(\mathbf{x}; \mu, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right)$

Example: MF for multivariate Gaussian

Recap

mean parametrization $p(\mathbf{x}; \mu, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right)$

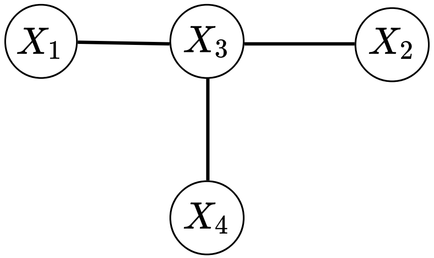
$$\eta = \Sigma^{-1}\mu$$
$$\Lambda = \Sigma^{-1}$$

Example: MF for multivariate Gaussian

Recap

mean parametrization $p(\mathbf{x};\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$

$$\Lambda = \begin{bmatrix} \Lambda_{11}, & 0, & \Lambda_{1,3}, & 0 \\ 0, & \Lambda_{2,2}, & \Lambda_{2,3}, & 0 \\ \Lambda_{3,1}, & \Lambda_{3,2}, & \Lambda_{3,3}, & \Lambda_{3,4} \\ 0, & 0, & \Lambda_{4,3}, & \Lambda_{4,4} \end{bmatrix}$$



$$\eta = \Sigma^{-1}\mu$$
$$\Lambda = \Sigma^{-1}$$

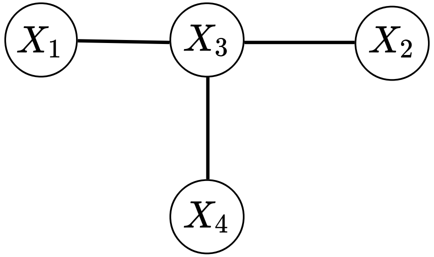
canonical parametrization $p(\mathbf{x};\eta,\Lambda) = \sqrt{\frac{|\Lambda|}{(2\pi)^n}} \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda\mathbf{x} + \eta\mathbf{x} - \frac{1}{2}\eta^T\Lambda\eta\right)$

Example: MF for multivariate Gaussian

Recap

mean parametrization $p(\mathbf{x};\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$

$$\Lambda = \begin{bmatrix} \Lambda_{11}, & 0, & \Lambda_{1,3}, & 0 \\ 0, & \Lambda_{2,2}, & \Lambda_{2,3}, & 0 \\ \Lambda_{3,1}, & \Lambda_{3,2}, & \Lambda_{3,3}, & \Lambda_{3,4} \\ 0, & 0, & \Lambda_{4,3}, & \Lambda_{4,4} \end{bmatrix}$$



$\eta = \Sigma^{-1}\mu$

$\Lambda = \Sigma^{-1}$

$\mu = \Lambda^{-1}\eta$

$\Sigma = \Lambda^{-1}$

canonical parametrization $p(\mathbf{x};\eta,\Lambda) = \sqrt{\frac{|\Lambda|}{(2\pi)^n}} \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda\mathbf{x} + \eta\mathbf{x} - \frac{1}{2}\eta^T\Lambda\eta\right)$

Example: MF for multivariate Gaussian

given a multivariate Gaussian (η, Λ)

I-project it into product of univariates of "any" form

Example: MF for multivariate Gaussian

given a multivariate Gaussian (η, Λ)

I-project it into product of univariates of "any" form

mean-field update:

$$\ln q_i(x_i) = \mathbb{E}_{q_{MB(i)}} \left[-\sum_{j \in MB(i)} x_i x_j \Lambda_{ij} - \frac{1}{2} \Lambda_{ii} x_i^2 + \eta_i x_i \right] + const.$$

from the precision matrix

terms that do not depend on xi

Example: MF for multivariate Gaussian

given a multivariate Gaussian (η, Λ)

I-project it into product of univariates of "any" form

mean-field update:

$$\ln q_i(x_i) = \mathbb{E}_{q_{MB(i)}} \left[-\sum_{j \in MB(i)} x_i x_j \Lambda_{ij} - \frac{1}{2} \Lambda_{ii} x_i^2 + \eta_i x_i \right] + const.$$

from the precision matrix

terms that do not depend on xi

μ_j

$$= -\frac{1}{2} \Lambda_{ii} x_i^2 - \sum_{j \in MB(i)} x_i \mathbb{E}[x_j] \Lambda_{ij} + \eta_i x_i + const.$$

Example: MF for multivariate Gaussian

given a multivariate Gaussian (η, Λ)

I-project it into product of univariates of "any" form

mean-field update:

$$\ln q_i(x_i) = \mathbb{E}_{q_{MB(i)}} \left[-\sum_{j \in MB(i)} x_i x_j \Lambda_{ij} - \frac{1}{2} \Lambda_{ii} x_i^2 + \eta_i x_i \right] + const.$$

from the precision matrix

terms that do not depend on xi

$$\mu_j$$

$$= -\frac{1}{2} \Lambda_{ii} x_i^2 - \sum_{j \in MB(i)} x_i \mathbb{E}[x_j] \Lambda_{ij} + \eta_i x_i + const.$$

it follows that $q_i(x_i)$ has a univariate Gaussian form: $q_i(x_i) = \mathcal{N}(x_i; \mu_i, \sigma^2)$

$$\mu_i \leftarrow \Lambda_{ii}^{-1} (\eta_i - \sum_{j \in MB(i)} \Lambda_{ij} \mu_j)$$

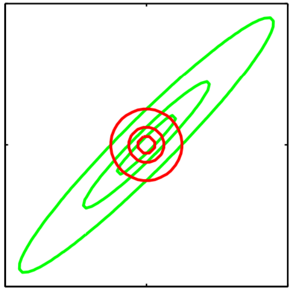
$$\sigma_i^2 \leftarrow \Lambda_{ii}^{-1}$$

Example: MF for multivariate Gaussian

given a multivariate Gaussian (η, Λ)

I-project it into product of univariates of "any" form

it follows that $q_i(x_i)$ has a univariate Gaussian form: $q_i(x_i) = \mathcal{N}(x_i; \mu_i, \sigma^2)$



$$\mu_i \leftarrow \Lambda_{ii}^{-1}(\eta_i - \sum_{j \in MB(i)} \Lambda_{ij}\mu_j)$$

$$\sigma_i^2 \leftarrow \Lambda_{ii}^{-1}$$

since the projection has the same mean

this gives an iterative solution for $\mu = \Lambda^{-1}\eta$

- with (η, Λ) as input

Structured mean-field

replace the product form $q(x) = \prod_k q_k(x_k)$

with tractable sub-structures $q(x) = \prod_I q_I(x_I)$

allow efficient exact inference

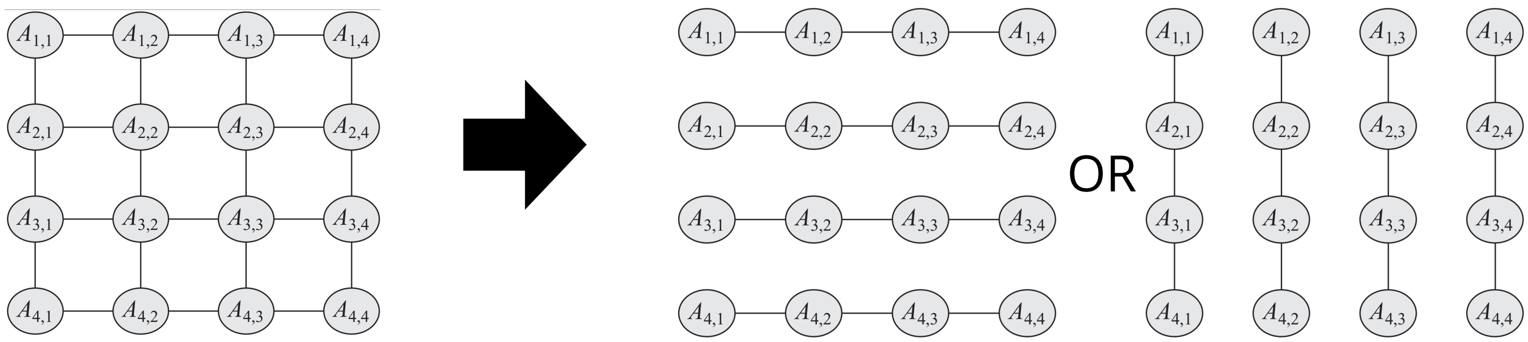
Structured mean-field

replace the product form $q(x) = \prod_k q_k(x_k)$

with tractable sub-structures $q(x) = \prod_I q_I(x_I)$

allow efficient exact inference

updates are produced by coordinate descent in each substructure



Summary

I-project into tractable sub-graphs:

- naive mean-field
- perform coordinate descent

Inherits the mode-seeking behavior of I-projection

optimal in special settings (e.g., some dense graphs with weak interactions)

less restricted than BP in the choice of dists.

in practice, LBP often performs better