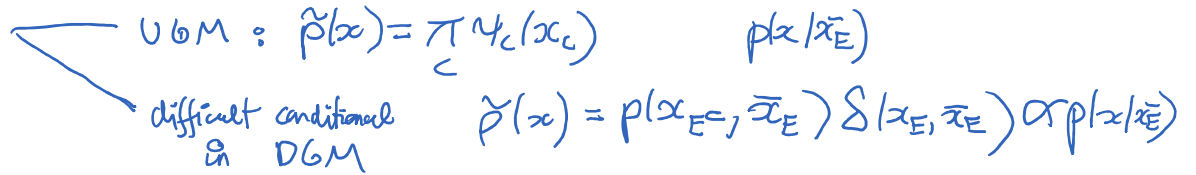


today: • Gibbs sampling
• variational methods

Gibbs sampling alg.

↳ M.H. with a clever choice of proposal $q_t(x'|z)$ [clever: $a(x'|x) = 1$]

example of applications:



(UGM)

cyclic Gibbs sampling alg.: nodes $i=1, \dots, n$

start at some $x^{(0)}$

for $t=1, \dots, \dots$

- pick $i = (t \text{ mod } n) + 1$
- sample $x_i^{(t)} \sim p(x_i = \cdot | x_{\neq i} = x_{\neq i}^{(t-1)})$
True conditional on x_i as proposal
- set $x_j^{(t)} = x_j^{(t-1)}$ for $j \neq i$

end

⊛ Gibbs sampling is M.H. with a time varying proposal

suppose we pick i at time t

then proposal is $q_t(x' | x^{(t-1)}) = p(x_i' | x_{\neq i}^{(t-1)}) \delta(x_{\neq i}', x_{\neq i}^{(t-1)})$
for $x_{\neq i}$ to stay constant

M.H. acceptance ratio:

$$a(x' | x^{(t-1)}) = \frac{q_t(x^{(t-1)} | x') p(x')}{q_t(x' | x^{(t-1)}) p(x^{(t-1)})}$$

$\rightarrow \frac{p(x_i^{(t-1)} | x_{\neq i}^{(t-1)}) \delta(x_{\neq i}^{(t-1)}, x_{\neq i}^{(t-1)}) p(x')}{p(x_i' | x_{\neq i}^{(t-1)}) \delta(x_{\neq i}', x_{\neq i}^{(t-1)}) p(x^{(t-1)})}$
 $\rightarrow \frac{p(x_i^{(t-1)}) p(x_{\neq i}^{(t-1)})}{p(x_i') p(x_{\neq i}^{(t-1)})}$

$$p(x_i | x_{-i}^{(t-1)}) \delta(x_i, x_{-i}^{(t-1)})$$

= 1 \rightarrow always accepts!

convergence of GS:

• let A be a Markov transition kernel of one full cycle of GS. (ie: n steps)
 \rightarrow homogeneous M.C.

if suppose that $\underline{p(x) > 0}$ then
 $p(x_i | x_{-i}) > 0$
 $\forall x_i \ \& \ x_{-i}$

$\Rightarrow A$ is irreducible & aperiodic
 because $A_{ii} > 0$
 \downarrow
 $A_{ij} > 0$
 $\forall i, j$
 \rightarrow since can get to any state with n steps

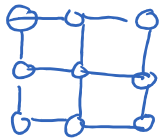
$$\Rightarrow A^t \pi_0 \xrightarrow{t \rightarrow \infty} p$$

⊗ also works for a random scan (pick $i \sim \text{Unif}(1:n)$ at each step)

example: GS for Ising model

Ising model $X_i \in \{0, 1\}$

UGM:



for GS,

want to compute $p(x_i | x_{-i})$ $\stackrel{\text{by cond. indep.}}{=} p(x_i | x_{\text{neighbors}(i)})$

$$\propto \exp(n_i x_i + \sum_{j \in N(i)} m_{ij} x_i x_j + \text{rest})$$

\Rightarrow renormalize to get conditional:

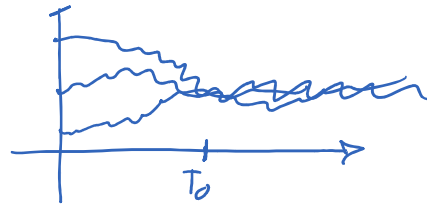
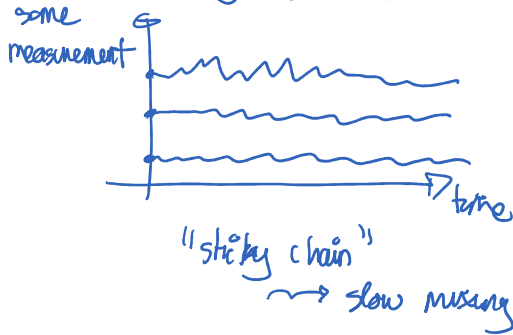
$$p(x_i=1 | x_{-i}) = \frac{\exp(n_i + \sum_{j \in N(i)} m_{ij} x_j) \exp(\text{rest})}{(1 + \exp(n_i + \sum_{j \in N(i)} m_{ij} x_j)) \exp(\text{rest})}$$

$$= \sigma(n_i + \sum_{j \in N(i)} m_{ij} x_j)$$

diagnostic of mixing

monitor mixing by running independent chains

monitor mixing by running independent chains



usually slow mixing comes from difficulty to move between high prob regions



→ annealing methods help this: proposal looks like $\frac{1}{Z} \exp(-\frac{E(x)}{T_k})$

example: "Annealed importance sampling"

↑ "temperature"
high temperature
⇒ more exploring

15h34

Variational methods

general idea: say we want to approximate Θ^*

then, express it as solution to opt. problem

$$\Theta^* = \underset{\Theta \in \Theta}{\text{argmin}} f(\Theta) \quad] \text{ OPT}$$

idea: approximate Θ^* via an approximation to OPT

linear algebra example

say want solⁿ to $Ax=b$ (ie. $x^* = A^{-1}b$)

$$x^* = \underset{x}{\text{argmin}} \|Ax-b\|^2$$

variational EM (motivation for objective)

recall EM trick:

latent variable $p(x, z | \theta)$
↓
unobserved

$$\log p(x | \theta) \geq \mathbb{E}_q[\log p(x, z | \theta)] \triangleq \mathcal{J}(q, \theta)$$

$$\log p(x|\theta) \geq \mathbb{E}_q[\log \frac{p(z|x;\theta)}{q(z)}] \triangleq \mathcal{J}(q, \theta)$$

$$\log p(x|\theta) - \mathcal{J}(q, \theta) = KL(q(\cdot) \| p(z|x;\theta))$$

$$E \text{ step: } \underset{q \in \text{all distributions on } z}{\text{argmax}} \mathcal{J}(q, \theta^{(t)}) \Leftrightarrow \underset{q}{\text{argmin}} KL(q \| p(z|x;\theta^{(t)}))$$

⊕ a variational approx for E step:

$$\text{do } q_{\text{approx}}^{(t+1)} = \underset{q \in \mathcal{Q}_{\text{simple}}}{\text{argmin}} KL(q \| p(\cdot|x;\theta^{(t+1)}))$$

$\mathcal{Q}_{\text{simple}}$
 ↑
 source of approx → b approximate $p(z|x, \theta^{(t)})$

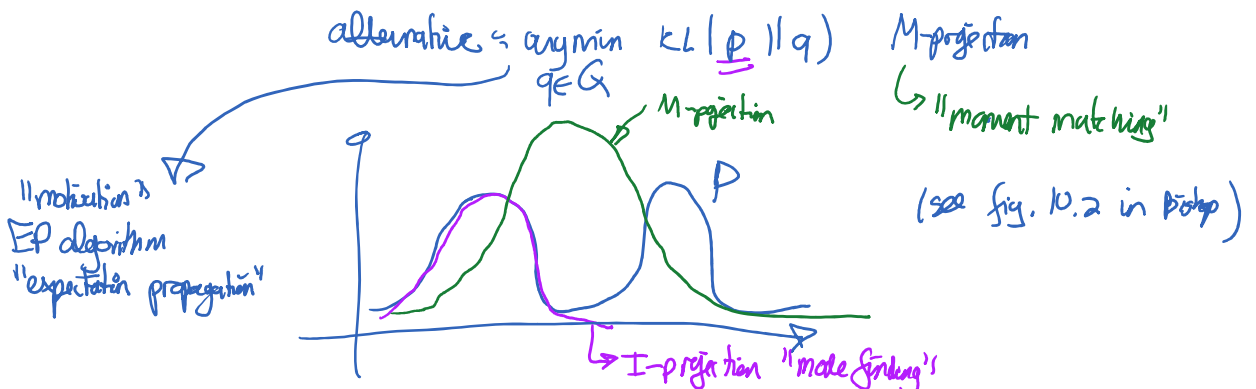
approximate M step:

$$\theta^{(t+1)} = \underset{\theta \in \Theta}{\text{argmax}} \underbrace{\mathbb{E}_{q_{\text{approx}}^{(t+1)}} [\log p(z, x|\theta)]}_{\mathcal{J}(q_{\text{approx}}^{(t+1)}, \theta)}$$

still a lower bound on $\log p(x|\theta)$
 but lose monotonicity guarantee
 on $\log p(x|\theta^{(t)})$ as a fun of t

more generally, using $\underset{q \in \mathcal{Q}}{\text{argmin}} KL(q \| p)$ is a variational approach to approximate p

note: I-projection; if q is simple, $\mathbb{E}_q[\log p]$
 can compute



Mean-field approximation (section 10.1 in Bishop)

Mean-field approximation (section 10.1 in Bishop)

let's suppose that $p(z)$ is in exp. family

$$z_1, \dots, z_p \quad p(z) = \exp(n^T T(z) - A(n))$$

mean field approximation $Q_{MF} = \{q(z) = \prod_i q_i(z_i)\}$
 set of fully factored dist.

$$\begin{aligned} KL(q \| p) &= \mathbb{E}_q [\log \frac{q}{p}] \\ &= -n^T \mathbb{E}_q [T(z)] + A(n) + \sum_z q(z) \log q(z) \\ &= -n^T \mathbb{E}_q [T(z)] + A(n) + \sum_i \sum_{z_i} q_i(z_i) \log q_i(z_i) \end{aligned}$$

coordinate descent on q_i 's

fix q_j for $j \neq i$; min w.r. to q_i $KL(q_i q_{-i} \| p)$
 $= -\mathbb{E}_{q_i} [n^T \mathbb{E}_{q_{-i}} [T(z)]] + \text{const.} + \sum_{z_i} q_i(z_i) \log q_i(z_i)$
 $\triangleq f_i(z_i)$

(like MaxENT)

add Lagrange multiplier for $\sum_{z_i} q_i(z_i) = 1 \rightarrow \lambda (1 - \sum_{z_i} q_i(z_i))$

$$\begin{aligned} \frac{\partial}{\partial q_i(z_i)} = 0 &\Rightarrow -f_i(z_i) + \log q_i(z_i) + 1 - \lambda = 0 \\ &\Rightarrow q_i^*(z_i) \propto \exp(f_i(z_i)) \end{aligned}$$

* general mean field update when target p is in exp. family:

$$q_i^{(t+1)}(z_i) \propto \exp(n^T \mathbb{E}_{q_{-i}^{(t)}} T(z))$$

Ising model:

$$T(z) \Rightarrow \begin{matrix} (z_i)_{i \in V} \\ (z_i z_j)_{i,j \in E} \end{matrix} \quad z_i \in \{-1, 1\}$$

$$\mathbb{E}_{q_i^{(t)}}(z_j) = q_j^{(t)}(z_j = 1) \triangleq \mu_j^{(t)}$$

$$\mathbb{E}_{q_i^{(t)}}(z_i z_j) = z_i \mu_j^{(t)} \quad z_i = \mu_i^{(t)}$$

$$\mathbb{E}q_i(z_j) = \mathbb{1}_{\{z_j=1\}} - \mu_j$$

$$\mathbb{E}q_i^{(t)}(z_i z_j) = z_i \mu_j^{(t)}$$

$$n^T \mathbb{E}q_i^{(t)}(z) = n_i z_i + \sum_{j \neq i} n_j \underbrace{\mathbb{E}q_i^{(t)}(z_j)}_{\mu_j^{(t)}} + \sum_{j \in N(i)} n_{ij} \underbrace{\mathbb{E}q_i^{(t)}(z_i z_j)}_{z_i \mu_j^{(t)}} + \text{rest (no } z_i)$$

result: $q_i^{(t+1)}(z_i) \propto \exp(n_i z_i + z_i \sum_{j \in N(i)} n_{ij} \mu_j^{(t)})$
 & parameter for $q_j^{(t)}$

$$\mu_j^{(t+1)} = \sigma\left(n_j + \sum_{i \in N(j)} n_{ij} \mu_i^{(t)}\right)$$

MF update for $q_i(z_i)$
 [with parameter] μ_i

compare with

G.S. update where $z_i^{(t+1)} = 1$ with prob. $\sigma\left(n_i + \sum_{j \in N(i)} n_{ij} z_j^{(t)}\right)$