Q42***: AR Model

Part I

Consider the following model:

$$\begin{aligned} A &\sim \mathcal{N}(A; 0, 1.2) \\ R &\sim \mathcal{IG}(R; 0.4, 250) \\ x_k | x_{k-1}, A, R &\sim \mathcal{N}(x_k; A x_{k-1}, R) \\ x_0 &= 1 \qquad\qquad x_1 = -6 \end{aligned}$$

1. Draw the directed graphical model and the factor graph
2. Write the expression for the full joint distribution and assign terms to the individual factors on the factor graph
3. Derive the full conditional distributions $p(A|R, x_0, x_1)$ and $p(R|A, x_0, x_1)$
4. Derive the joint distribution $p(A, R, x_0 = 1, x_1 = -6)$ and create a contour plot.

Part II

Consider the following model discussed in detail during the lectures.

$$\begin{aligned} A &\sim \mathcal{N}(A; 0, P) \\ R &\sim \mathcal{IG}(R; \nu, \nu/\beta) \\ x_k | x_{k-1}, A, R &\sim \mathcal{N}(x_k; A x_{k-1}, R) \end{aligned}$$

where $\mathcal{N}$ is a Gaussian and

$$\mathcal{IG}(R; a, b) = \exp\left(-(a+1)\log R - \frac{b}{R} - \log\Gamma(a) + a\log b\right)$$

Caution: (This definition is different from the definition of $\mathcal{IG}$ given in some of the earlier lectures.)
We are given the hyperparameters $\theta = (\nu, \beta, P)$

$$\begin{aligned} \nu &= 0.4 \qquad \beta = 100 \qquad P = 1.2 \\ x_0 &= 1 \qquad x_1 = -6 \end{aligned}$$

1. Derive and implement an EM algorithm to find the MAP estimate

$$R^* = \underset{R}{\operatorname{argmax}}\, p(R|x_0, x_1, \theta)$$

2. Derive and implement an EM algorithm to find the MAP estimate

$$A^* = \underset{A}{\operatorname{argmax}}\, p(A|x_0, x_1, \theta)$$

3. Derive and implement an ICM (Iterative conditional modes) algorithm to find

$$(R^*, A^*) \quad = \quad \operatorname*{argmax}_{A,R} p(A, R|x_0, x_1, \theta)$$

4. In the lectures, we have shown that the unnormalised posterior is

$$\begin{aligned}
\phi &= p(A, R, x_1 = \hat{x}_1 | x_0 = \hat{x}_0, \theta) = \mathcal{N}(x_1; Ax_0, R)\mathcal{N}(A; 0, P)\mathcal{IG}(R; \nu, \nu/\beta) \\
&\propto \exp\left(-\frac{1}{2}\frac{x_1^2}{R} + x_0 x_1 \frac{A}{R} - \frac{1}{2}\frac{x_0^2 A^2}{R} - \frac{1}{2}\log 2\pi R\right) \\
&\quad \exp\left(-\frac{1}{2}\frac{A^2}{P} - \frac{1}{2}\log|2\pi P|\right) \\
&\quad \exp\left(-(\nu+1)\log R - \frac{\nu}{\beta}\frac{1}{R} - \log\Gamma(\nu) + \nu\log(\nu/\beta)\right)
\end{aligned}$$

We know also that the marginal log-likelihood

$$\log Z = \log p(x_1 = \hat{x}_1 | x_0 = \hat{x}_0, \theta)$$

is lower bounded by

$$\mathcal{B}_{VB} \quad = \quad \langle \log \phi \rangle_Q + H[Q]$$

where

$$\begin{aligned}
Q &= q(A)q(R) \\
q(A) &= \mathcal{N}(A; m, \Sigma) \\
q(R) &= \mathcal{IG}(R; a, b)
\end{aligned}$$

Extend the VB algorithm given in the slides so that you compute this bound at every iteration and plot the bound $\mathcal{B}$ as a function of iterations. You should observe that the VB fixed point **monotonically** increases this lower bound. Restart your algorithm several times and compare the largest bound you find with the bound you find with importance sampling.