Final solutions
ECE 271
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos Fall 2017

1. a) The BDR is to pick class 0 if

$$\begin{aligned}
P_{X|Y}(x|0)P_Y(0) &\geq P_{X|Y}(x|1)P_Y(1) \\
\frac{\beta_0^{\alpha_0} x^{\alpha_0-1} e^{-\beta_0 x}}{g(\alpha_0)}\pi_0 &\geq \frac{\beta_1^{\alpha_1} x^{\alpha_1-1} e^{-\beta_1 x}}{g(\alpha_1)}\pi_1 \\
\alpha_0 \log\beta_0 + (\alpha_0-1)\log x - \beta_0 x - \log g(\alpha_0) + \log\pi_0 &\geq \alpha_1 \log\beta_1 + (\alpha_1-1)\log x - \beta_1 x - \log g(\alpha_1) + \log\pi_1 \\
(\alpha_0-\alpha_1)\log x + (\beta_1-\beta_0)x &\geq \log\frac{\beta_1^{\alpha_1} g(\alpha_0)\pi_1}{\beta_0^{\alpha_0} g(\alpha_1)\pi_0}
\end{aligned}$$

This can also be written as

$$x^{(\alpha_0-\alpha_1)} e^{(\beta_1-\beta_0)x} \geq \frac{\beta_1^{\alpha_1} g(\alpha_0)\pi_1}{\beta_0^{\alpha_0} g(\alpha_1)\pi_0}$$

1. b) The log-likelihood function is

$$\mathcal{L}(\beta) = \sum_i \alpha\log\beta + (\alpha-1)\log x_i - \beta x_i - \log(\alpha). \tag{1}$$

It has a critical point when

$$\begin{aligned}
0 &= \frac{\partial\mathcal{L}}{\partial\beta} = \frac{n\alpha}{\beta} - \sum_i x_i \\
\beta &= \frac{n\alpha}{\sum_i x_i}
\end{aligned}$$

Since

$$\frac{\partial^2\mathcal{L}}{\partial\beta^2} = -\frac{n\alpha}{\beta^2} \leq 0$$

this is a maximum of the likelihood.

1.c) In the E-step, we compute the posterior probabilities of assignment of each point x_i to each mixture component j

$$h_{ij} = \frac{\Gamma(x_i,\alpha,\beta_j^{(i)})\pi_j^{(i)}}{\sum_k \Gamma(x_i,\alpha,\beta_k^{(i)})\pi_k^{(i)}} = \frac{(\beta_j^{(i)})^\alpha e^{-\beta_j^{(i)} x_i}\pi_j^{(i)}}{\sum_k (\beta_k^{(i)})^\alpha e^{-\beta_k^{(i)} x_i}\pi_k^{(i)}}$$

In the M-step we maximize the Q-function

$$Q(\Psi|\Psi^{(n)}) = \sum_{ij} h_{ij}\log(\Gamma(x_i,\alpha,\beta_j)\pi_j) \tag{2}$$

under the constraint that $\sum_j \pi_j = 1$. For this, we use the Lagrangian

$$L = \sum_{ij} h_{ij} \left(\alpha \log \beta_j + (\alpha - 1) \log x_i - \beta_j x_i - \log g(\alpha) + \log \pi_j\right) - \lambda \left(\sum_j \pi_j - 1 \right) \tag{3}$$

which has derivatives

$$\begin{aligned}
\frac{\partial L}{\partial \beta_j} &= \sum_i h_{ij} \frac{\alpha}{\beta_j} - \sum_i h_{ij} x_i \\
\frac{\partial L}{\partial \pi_j} &= \sum_i h_{ij} \frac{1}{\pi_j} - \lambda \\
\frac{\partial L}{\partial \lambda} &= -\sum_j \pi_j + 1
\end{aligned}$$

Setting to zero, we obtain

$$\begin{aligned}
\beta_j^{(i+1)} &= \alpha \frac{\sum_i h_{ij}}{\sum_i h_{ij} x_i} \\
\lambda \pi_j &= \sum_i h_{ij} \\
1 &= \sum_j \pi_j
\end{aligned}$$

from which $\lambda = \sum_{ij} h_{ij}$ and

$$\begin{aligned}
\beta_j^{(i+1)} &= \alpha \frac{\sum_i h_{ij}}{\sum_i h_{ij} x_i} \\
\pi_j^{(i+1)} &= \frac{\sum_i h_{ij}}{\sum_{ij} h_{ij}} = \frac{\sum_i h_{ij}}{n}
\end{aligned}$$

1.d) The posterior is

$$\begin{aligned}
P_{\alpha,\beta|T}(\alpha, \beta | \mathcal{D}) &\propto P_{T|\alpha,\beta}(\mathcal{D}|\alpha, \beta) P_{\alpha,\beta}(\alpha, \beta) \\
&= \prod_i P_{X|\alpha,\beta}(x_i|\alpha, \beta) P_{\alpha,\beta}(\alpha, \beta) \\
&= \prod_i \frac{\beta^\alpha x_i^{\alpha-1} e^{-\beta x_i}}{g(\alpha)} \frac{\beta^{\alpha s} p^{\alpha-1} e^{-\beta q}}{g(\alpha)^r} \\
&= \frac{\beta^{(n+s)\alpha}}{g(\alpha)^{n+r}} \left(\prod_i x_i \right)^{\alpha-1} e^{-\beta \sum_i x_i} p^{\alpha-1} e^{-\beta q} \\
&= \frac{\beta^{(n+s)\alpha}}{g(\alpha)^{n+r}} \left(p \prod_i x_i \right)^{\alpha-1} e^{-\beta\left(q + \sum_i x_i\right)}
\end{aligned}$$

Note that this has the same form as the prior, and the normalization constant is thus one, i.e. equality holds. The prior is thus a conjugate prior for the Gamma distribution. The MAP estimate of β is obtained by setting to zero the derivative of

$$\begin{aligned}
\mathcal{L} &= \log P_{\alpha,\beta|T}(\alpha,\beta|\mathcal{D}) \\
&= \log \frac{\beta^{(n+s)\alpha}}{g(\alpha)^{n+r}} \left(p\prod_i x_i\right)^{\alpha-1} e^{-\beta\left(q+\sum_i x_i\right)} \\
&= (n+s)\alpha\log\beta - \beta\left(q+\sum_i x_i\right)
\end{aligned}$$

where we eliminated the terms that do not depend on β. This leads to

$$\begin{aligned}
0 &= \frac{(n+s)\alpha}{\beta} - \left(q+\sum_i x_i\right) \\
\beta_{MAP} &= \frac{(n+s)\alpha}{q+\sum_i x_i}.
\end{aligned}$$

When compared to the ML estimate, we see that the MAP estimate is the same but estimated from an augmented dataset which has s additional examples whose values add up to q.

2. a) The complete data log-likelihood is

$$\begin{aligned}
&\log P_{Y,X}(y_1,\ldots,y_n,x_1,\ldots,x_n) = \\
&= \sum_i \log P_{Y_i,X_i}(y_i,x_i) = \sum_i \log P_{Y_i|X_i}(y_i|x_i)P_{X_i}(x_i) \\
&= \sum_i -\frac{1}{2}\log(2\pi\sigma_y^2) - \frac{(y_i-x_i)^2}{2\sigma_y^2} - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma_y^2) - \sum_i \frac{(y_i-x_i)^2}{2\sigma_y^2} - \frac{n}{2}\log(2\pi\sigma^2) - \sum_i \frac{(x_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma_y^2) - \sum_i \frac{y_i^2-2x_iy_i+x_i^2}{2\sigma_y^2} - \frac{n}{2}\log(2\pi\sigma^2) - \sum_i \frac{x_i^2-2\mu x_i+\mu^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi\sigma_y^2) - \frac{n}{2}\log(2\pi\sigma^2) - \sum_i \left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)x_i^2 + \sum_i \left(\frac{y_i}{\sigma_y^2}+\frac{\mu}{\sigma^2}\right)x_i - \sum_i \frac{y_i^2}{2\sigma_y^2} - \frac{n\mu^2}{2\sigma^2}
\end{aligned}$$

The incomplete data log-likelihood is

$$\begin{aligned}
&\log P_{Y,X}(y_1,\ldots,y_n) = \\
&= \sum_i \log P_{Y_i,X_i}(y_i,x_i) = \sum_i \log \int P_{Y_i|X_i}(y_i|x_i)P_{X_i}(x_i)dx_i \\
&= \sum_i \log \int \mathcal{G}(y_i,x_i,\sigma_y^2)\mathcal{G}(x_i,\mu,\sigma^2)dx_i \\
&= \sum_i \log \int \mathcal{G}(y_i-x_i,0,\sigma_y^2)\mathcal{G}(x_i,\mu,\sigma^2)dx_i \\
&= \sum_i \log \mathcal{G}(y_i,0,\sigma_y^2) \star \mathcal{G}(y_i,\mu,\sigma^2) \\
&= \sum_i \log \mathcal{G}(y_i,\mu,\sigma_y^2+\sigma^2)
\end{aligned}$$

where $\star$ denotes convolution.

2.b) The Q function is

$$
\begin{aligned}
E_{X|Y;\Psi^{(k)}}\left[\log P_{Y,X}(y_1,\ldots,y_n,x_1,\ldots,x_n)|\mathcal{D}\right] &= -\frac{n}{2}\log(2\pi\sigma_y^2) - \frac{n}{2}\log(2\pi\sigma^2) \\
- \quad \sum_i \left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right) E_{X_i|Y_i,\Psi^{(k)}}[x_i^2|y_i] &+ \sum_i \left(\frac{y_i}{\sigma_y^2}+\frac{\mu}{\sigma^2}\right) E_{X_i|Y_i,\Psi^{(k)}}[x_i|y_i] - \sum_i \frac{y_i^2}{2\sigma_y^2} - \frac{n\mu^2}{2\sigma^2} \\
= \quad & -\frac{n}{2}\log(2\pi\sigma_y^2) - \frac{n}{2}\log(2\pi\sigma^2) \\
- \quad \left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)\sum_i E_{X_i|Y_i,\Psi^{(k)}}[x_i^2|y_i] &+ \sum_i E_{X_i|Y_i,\Psi^{(k)}}[x_i|y_i]\left(\frac{y_i}{\sigma_y^2}+\frac{\mu}{\sigma^2}\right) - \sum_i \frac{y_i^2}{2\sigma_y^2} - \frac{n\mu^2}{2\sigma^2}
\end{aligned}
$$

It requires the computation of

$$
\begin{aligned}
P_{X|Y}(x|y) &\propto P_{Y|X}(y|x)P_X(x) \propto exp\left\{-\frac{(y-x)^2}{2\sigma_y^2} - \frac{(x-\mu)^2}{2\sigma^2}\right\} \\
&\propto exp\left\{-\left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)x^2 + 2\left(\frac{y}{2\sigma_y^2}+\frac{\mu}{2\sigma^2}\right)x\right\} \\
&\propto exp\left\{-\left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)\left(x^2 - 2\frac{\left(\frac{y}{2\sigma_y^2}+\frac{\mu}{2\sigma^2}\right)}{\left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)}x\right)\right\} \\
&\propto exp\left\{-\left(\frac{\sigma_y^2+\sigma^2}{2\sigma_y^2\sigma^2}\right)\left(x^2 - 2\left(\frac{y\sigma^2+\mu\sigma_y^2}{\sigma_y^2+\sigma^2}\right)x\right)\right\} \\
&= \mathcal{G}(x, \mu_{x|y}, \sigma_{x|y}^2)
\end{aligned}
$$

with

$$
\begin{aligned}
\mu_{x|y} &= \frac{y\sigma^2+\mu\sigma_y^2}{\sigma_y^2+\sigma^2} \\
\sigma_{x|y}^2 &= \frac{\sigma_y^2\sigma^2}{\sigma_y^2+\sigma^2}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\beta_i &= E_{X_i|Y_i,\Psi^{(k)}}[x_i|y_i] = \mu_{x_i|y_i}^{(k)} \\
&= \frac{y_i(\sigma^2)^{(k)}+\mu^{(k)}(\sigma_y^2)^{(k)}}{(\sigma_y^2)^{(k)}+(\sigma^2)^{(k)}} \\
\alpha_i &= E_{X_i|Y_i,\Psi^{(k)}}[x_i^2|y_i] = (\sigma_{x_i|y_i}^2)^{(k)} + (\mu_{x_i|y_i}^{(k)})^2 \\
&= \frac{(\sigma_y^2)^{(k)}(\sigma^2)^{(k)}}{(\sigma_y^2)^{(k)}+(\sigma^2)^{(k)}} + \beta_i^2
\end{aligned}
$$

and the Q function is

$$
\begin{aligned}
Q(\Psi|\Psi^{(k)}) &= -\frac{n}{2}\log(2\pi\sigma_y^2) - \frac{n}{2}\log(2\pi\sigma^2) \\
&- \left(\frac{1}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)\sum_i \alpha_i + \sum_i \beta_i\left(\frac{y_i}{\sigma_y^2}+\frac{\mu}{\sigma^2}\right) - \sum_i \frac{y_i^2}{2\sigma_y^2} - \frac{n\mu^2}{2\sigma^2}
\end{aligned}
$$

2. c) The M-step maximizes the Q function by setting gradients to zero

$$
\begin{aligned}
\frac{\partial Q}{\partial \mu} &= \frac{1}{\sigma^2}\sum_i \beta_i - \frac{n\mu}{\sigma^2} = 0 \\
\mu^{(k+1)} &= \frac{1}{n}\sum_i \beta_i \\
\frac{\partial Q}{\partial \sigma_y^2} &= -\frac{n}{2\sigma_y^2} + \frac{1}{2(\sigma_y^2)^2}\sum_i \alpha_i - \sum_i \beta_i 2\frac{y_i}{2(\sigma_y^2)^2} + \sum_i \frac{y_i^2}{2(\sigma_y^2)^2} = 0 \\
n &= \frac{1}{\sigma_y^2}\left(\sum_i \alpha_i - 2\sum_i \beta_i y_i + \sum_i y_i^2\right) \\
(\sigma_y^2)^{(k+1)} &= \frac{1}{n}\left(\sum_i \alpha_i - 2\sum_i \beta_i y_i + \sum_i y_i^2\right) \\
\frac{\partial Q}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_i \alpha_i - \sum_i \beta_i 2\frac{\mu}{2(\sigma^2)^2} + \frac{n\mu^2}{2(\sigma^2)^2} = 0 \\
n &= \frac{1}{\sigma^2}\left(\sum_i \alpha_i - 2\mu\sum_i \beta_i + n\mu^2\right) \\
(\sigma^2)^{(k+1)} &= \frac{1}{n}\left(\sum_i \alpha_i - 2\mu^{(k+1)}\sum_i \beta_i + n(\mu^{(k+1)})^2\right) \\
&= \frac{1}{n}\sum_i \alpha_i - \left(\frac{1}{n}\sum_i \beta_i\right)^2
\end{aligned}
$$

2. d) The μ-restricted incomplete data log-likelihood is

$$
\mathcal{L} = \sum_i (y_i - \mu)^2
$$

and has maximum

$$
\mu^* = \frac{1}{n}\sum_i y_i
$$

The μ-restricted Q function is

$$
\begin{aligned}
Q(\mu|\mu^{(k)}) &\propto 2\frac{\mu}{2\sigma^2}\frac{1}{n}\sum_i \beta_i - \frac{\mu^2}{2\sigma^2} \\
&\propto \left(\frac{1}{n}\sum_i \beta_i - \mu\right)^2
\end{aligned}
$$

and is maximum when

$$
\mu^{(k+1)} = \frac{1}{n}\sum_i \beta_i = \frac{(\sigma^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}}\frac{1}{n}\sum_i y_i + \frac{(\sigma_y^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}}\mu^{(k)}
$$

At equilibrium $\mu^{(k+1)} = \mu^{(k)} = \mu'$ where

$$\begin{aligned}
\mu' \left(1 - \frac{(\sigma_y^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}}\right) &= \frac{(\sigma^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}} \frac{1}{n} \sum_i y_i \\
\mu' \frac{(\sigma^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}} &= \frac{(\sigma^2)^{(k)}}{(\sigma_y^2)^{(k)} + (\sigma^2)^{(k)}} \frac{1}{n} \sum_i y_i \\
\mu' &= \frac{1}{n} \sum_i y_i = \mu^*
\end{aligned}$$

and the algorithm converges to the ML estimate.

3. a) The complete data log-likelihood is

$$
\begin{aligned}
\log P_{Y_1^n,X_1^n}(y_1^n,x_1^n) &= \\
&= \log P_{Y_1^n|X_1^n}(y_1^n|x_1^n) + \log P_{X_1^n}(x_1^n) \\
&= \sum_{k=1}^{n} \log P_{Y_k|X_k}(y_k|x_k) + \sum_{k=1}^{n} \log P_{X_k|X_1^{k-1}}(x_k|x_1^{k-1}) + \log P_{X_0}(x_0) \\
&= \sum_{k=1}^{n} \log P_{Y_k|X_k}(y_k|x_k) + \sum_{k=1}^{n} \log P_{X_k|X_{k-1}}(x_k|x_{k-1}) + \log P_{X_0}(x_0) \\
&= -\frac{n}{2}\log(2\pi\sigma_y^2) - \sum_{k=1}^{n}\frac{(y_k-bx_k)^2}{2\sigma_y^2} - \frac{n}{2}\log(2\pi\sigma^2) - \sum_{k=1}^{n}\frac{(x_k-ax_{k-1})^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma_0^2) - \frac{(x_0-\mu_0)^2}{2\sigma_0^2}
\end{aligned}
$$

Since $\mu_0, \sigma_0^2, \sigma^2$, and σ_y^2 are known, they are not going to have any impact in the EM optimization and we can immediately drop any terms that only depend on them. This makes the complete data log-likelihood

$$
\begin{aligned}
\log P_{Y_1^n,X_1^n}(y_1^n,x_1^n) &\propto -\sum_{k=1}^{n}\frac{(y_k-bx_k)^2}{2\sigma_y^2} - \sum_{k=1}^{n}\frac{(x_k-ax_{k-1})^2}{2\sigma^2} - \frac{(x_0-\mu_0)^2}{2\sigma_0^2} \\
&= -\sum_{k=1}^{n}\frac{y_k^2-2bx_ky_k+b^2x_k^2}{2\sigma_y^2} - \sum_{k=1}^{n}\frac{x_k^2-2ax_kx_{k-1}+a^2x_{k-1}^2}{2\sigma^2} - \frac{x_0^2-2\mu_0x_0+\mu_0^2}{2\sigma_0^2} \\
&= -\sum_{k=1}^{n}\frac{y_k^2}{\sigma_y^2} - \left(\frac{b^2}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)\sum_{k=1}^{n}x_k^2 - \frac{a^2}{2\sigma^2}\sum_{k=1}^{n}x_{k-1}^2 + \frac{2b}{2\sigma_y^2}\sum_{k=1}^{n}y_kx_k + \frac{2a}{2\sigma^2}\sum_{k=1}^{n}x_kx_{k-1} - \frac{x_0^2}{2\sigma_0^2} + \frac{2\mu_0}{2\sigma_0^2}x_0 - \frac{\mu_0^2}{2\sigma_0^2} \\
&\propto -\left(\frac{b^2}{2\sigma_y^2}+\frac{1}{2\sigma^2}\right)\sum_{k=1}^{n}x_k^2 - \frac{a^2}{2\sigma^2}\sum_{k=1}^{n}x_{k-1}^2 + \frac{2b}{2\sigma_y^2}\sum_{k=1}^{n}y_kx_k + \frac{2a}{2\sigma^2}\sum_{k=1}^{n}x_kx_{k-1} - \frac{x_0^2}{2\sigma_0^2} + \frac{2\mu_0}{2\sigma_0^2}x_0
\end{aligned}
$$

3. b) The Q-function requires the expectations

$$
E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_k|y_1^n] \quad E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_k^2|y_1^n] \quad E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_{k-1}^2|y_1^n] \quad E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_kx_{k-1}|y_1^n]
$$

and

$$
E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_ky_k|y_1^n] = y_kE_{X_1^n|Y_1^n,\Psi^{(t-1)}}[x_k|y_1^n] \quad E_{X_0|Y_1^n,\Psi^{(t-1)}}[x_0|y_1^n] \quad E_{X_0|Y_1^n,\Psi^{(t-1)}}[x_0^2|y_1^n]
$$

By the definition of expectation

$$
\begin{aligned}
E_{X_1^n|Y_1^n,\Psi^{(t-1)}}[f(x_k)|y_1^n] &= \int P_{X_1^n|Y_1^n}(x_1^n|y_1^n)f(x_k)dx_1^n \\
&= \int f(x_k)\left(\int P_{X_1^n|Y_1^n}(x_1^n|y_1^n)dx_1,\dots,dx_{k-1},dx_{k+1},\dots,dx_n\right)dx_k \\
&= \int f(x_k)P_{X_k|Y_1^n}(x_k|y_1^n)dx_k \\
&= E_{X_k|Y_1^n,\Psi^{(t-1)}}[f(x_k)|y_1^n]
\end{aligned}
$$

and a similar simplification can be made for the term in x_kx_{k-1}. Hence, the computation of the Q function requires

$$
\beta_k = E_{X_k|Y_1^n,\Psi^{(t-1)}}[x_k|y_1^n]
$$

$$
\begin{aligned}
\alpha_k &= E_{X_k|Y_1^n,\Psi^{(t-1)}}[x_k^2|y_1^n] \\
\alpha_{k-1} &= E_{X_{k-1}|Y_1^n,\Psi^{(t-1)}}[x_{k-1}^2|y_1^n] \\
\gamma_{k,k-1} &= E_{X_k,X_{k-1}|Y_1^n,\Psi^{(t-1)}}[x_k x_{k-1}|y_1^n] \\
y_k\alpha_k &= y_k E_{X_k|Y_1^n,\Psi^{(t-1)}}[x_k|y_1^n]
\end{aligned}
$$

for $k = 1, \ldots, n$. Note that

$$
\begin{aligned}
\alpha_0 &= E_{X_0|Y_1^n,\Psi^{(t-1)}}[x_0^2|y_1^n] = E_{X_0;\Psi^{(t-1)}}[x_0^2] = \sigma_0^2 + \mu_0^2 \\
\beta_0 &= E_{X_0|Y_1^n,\Psi^{(t-1)}}[x_0|y_1^n] = E_{X_0;\Psi^{(t-1)}}[x_0] = \mu
\end{aligned}
$$

The Q-function is

$$
\begin{aligned}
Q(\Psi,\Psi^{(t-1)}) &\propto -\left(\frac{b^2}{2\sigma_y^2} + \frac{1}{2\sigma^2}\right)\sum_{k=1}^n \alpha_k - \frac{a^2}{2\sigma^2}\sum_{k=1}^n \alpha_{k-1} + \frac{2b}{2\sigma_y^2}\sum_{k=1}^n y_k\beta_k + \frac{2a}{2\sigma^2}\sum_{k=1}^n \gamma_{k,k-1} - \frac{\sigma_0^2+\mu_0^2}{2\sigma_0^2} + \frac{2\mu_0^2}{2\sigma_0^2} \\
&\propto -\left(\frac{b^2}{2\sigma_y^2} + \frac{1}{2\sigma^2}\right)\sum_{k=1}^n \alpha_k - \frac{a^2}{2\sigma^2}\sum_{k=1}^n \alpha_{k-1} + \frac{2b}{2\sigma_y^2}\sum_{k=1}^n y_k\beta_k + \frac{2a}{2\sigma^2}\sum_{k=1}^n \gamma_{k,k-1}
\end{aligned}
$$

3. c) The M-step takes gradients wrt a, b and sets them to zero

$$
\begin{aligned}
\frac{\partial Q}{\partial a} &= -\frac{a}{\sigma^2}\sum_{k=1}^n \alpha_{k-1} + \frac{1}{\sigma^2}\sum_{k=1}^n \gamma_{k,k-1} = 0 \\
\frac{\partial Q}{\partial b} &= -\frac{b}{\sigma_y^2}\sum_{k=1}^n \alpha_k + \frac{1}{\sigma_y^2}\sum_{k=1}^n y_k\beta_k = 0
\end{aligned}
$$

leading to

$$
\begin{aligned}
a^{(t+1)} &= \frac{\sum_{k=1}^n \gamma_{k,k-1}}{\sum_{k=1}^n \alpha_{k-1}} \\
b^{(t+1)} &= \frac{\sum_{k=1}^n y_k\beta_k}{\sum_{k=1}^n \alpha_k}
\end{aligned}
$$

3. d)

$$
\begin{aligned}
P_{X_k|Y_1^n;\Psi^{(t-1)}}(x_k|y_1^n) &= \frac{P_{Y_1^n,X_k;\Psi^{(t-1)}}(y_1^n,x_k)}{P_{Y_1^n;\Psi^{(t-1)}}(y_1^n)} \quad \text{(definition of cond. pdf)} \\
&\propto P_{Y_{k+1}^n|X_k,Y_1^k;\Psi^{(t-1)}}(y_{k+1}^n|x_k,y_1^k)P_{X_k,Y_1^k;\Psi^{(t-1)}}(x_k,y_1^k) \quad \text{(denominator is a constant)} \\
&= P_{Y_{k+1}^n|X_k;\Psi^{(t-1)}}(y_{k+1}^n|x_k)P_{X_k,Y_1^k;\Psi^{(t-1)}}(x_k,y_1^k) \quad \text{(independence)} \\
&= \frac{P_{Y_{k+1}^n|X_k;\Psi^{(t-1)}}(y_{k+1}^n|x_k)P_{Y_1^k,X_k;\Psi^{(t-1)}}(y_1^k,x_k)}{\int P_{Y_{k+1}^n|X_k;\Psi^{(t-1)}}(y_{k+1}^n|x_k)P_{Y_1^k,X_k;\Psi^{(t-1)}}(y_1^k,x_k)dx_k} \quad \text{(normalization)}
\end{aligned}
$$

$$
P_{Y_1^k,X_k;\Psi^{(t-1)}}(y_1^k,x_k) =
$$

$$
\begin{aligned}
&= \int P_{Y_1^k, X_k, X_{k-1}; \Psi^{(t-1)}}(y_1^k, x_k, x_{k-1}) dx_{k-1} \\
&= \int P_{Y_k|Y_1^{k-1}, X_k, X_{k-1}; \Psi^{(t-1)}}(y_k|y_1^{k-1}, x_k, x_{k-1}) P_{Y_1^{k-1}, X_k, X_{k-1}; \Psi^{(t-1)}}(y_1^{k-1}, x_k, x_{k-1}) dx_{k-1} \\
&= P_{Y_k|X_k; \Psi^{(t-1)}}(y_k|x_k) \int P_{X_k|Y_1^{k-1}, X_{k-1}; \Psi^{(t-1)}}(x_k|y_1^{k-1}, x_{k-1}) P_{Y_1^{k-1}, X_{k-1}; \Psi^{(t-1)}}(y_1^{k-1}, x_{k-1}) dx_{k-1} \\
&= P_{Y_k|X_k; \Psi^{(t-1)}}(y_k|x_k) \int P_{X_k|X_{k-1}; \Psi^{(t-1)}}(x_k|x_{k-1}) P_{Y_1^{k-1}, X_{k-1}; \Psi^{(t-1)}}(y_1^{k-1}, x_{k-1}) dx_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
&P_{Y_{k+1}^n|X_k; \Psi^{(t-1)}}(y_{k+1}^n|x_k) = \\
&= \int P_{Y_{k+1}^n|X_{k+1}, X_k; \Psi^{(t-1)}}(y_{k+1}^n|x_{k+1}, x_k) P_{X_{k+1}|X_k; \Psi^{(t-1)}}(x_{k+1}|x_k) dx_{k+1} \\
&= \int P_{Y_{k+1}^n|X_{k+1}; \Psi^{(t-1)}}(y_{k+1}^n|x_{k+1}) P_{X_{k+1}|X_k; \Psi^{(t-1)}}(x_{k+1}|x_k) dx_{k+1} \\
&= \int P_{Y_{k+2}^n|Y_{k+1}, X_{k+1}; \Psi^{(t-1)}}(y_{k+2}^n|x_{k+1}, y_{k+1}) P_{Y_{k+1}|X_{k+1}; \Psi^{(t-1)}}(y_{k+1}|x_{k+1}) P_{X_{k+1}|X_k; \Psi^{(t-1)}}(x_{k+1}|x_k) dx_{k+1} \\
&= \int P_{Y_{k+2}^n|X_{k+1}; \Psi^{(t-1)}}(y_{k+2}^n|x_{k+1}) P_{Y_{k+1}|X_{k+1}; \Psi^{(t-1)}}(y_{k+1}|x_{k+1}) P_{X_{k+1}|X_k; \Psi^{(t-1)}}(x_{k+1}|x_k) dx_{k+1}
\end{aligned}
$$