Bayesian HMM clustering of x-vector sequences (VBx) in speaker diarization: theory, implementation and analysis on standard tasks, technical report

Mireia Diez, Lukáš Burget

February 2, 2024

1 Introduction

This technical report is created as a complement to the paper [1]. The reader can find here all the derivations of the update formulas shown in the mentioned paper.

For the sake of completeness, we re-introduce in section 2 all the variables used in the technical report. Still, for a proper description and definition of all these variables, we refer the reader to the original paper.

2 Definition of variables

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T\}$ be the sequence of observed x-vectors and $\mathbf{Z} = \{z_1, z_2, ..., z_T\}$ the corresponding sequence of discrete latent variables defining the hard alignment of x-vectors to HMM states. In our notation, $z_t = s$ indicates that the speaker (HMM state) s is responsible for generating observation $\mathbf{x}_t$. Let $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_S\}$ be the set of all the speaker-specific latent variables.

The x-vectors that are used as input for the diarization algorithm are obtained as

$$\mathbf{X} = (\hat{\mathbf{X}} - \mathbf{m})\mathbf{E} \tag{1}$$

where $\mathbf{E}$ is the transformation matrix which transforms the x-vectors into the desired space. This matrix can be obtained by solving the standard generalized eigen-value problem

$$\mathbf{\Sigma}_b\mathbf{E} = \mathbf{\Sigma}_w\mathbf{E}\mathbf{\Phi} \tag{2}$$

where $\mathbf{E}$ is the matrix of eigen-vectors and $\mathbf{\Phi}$ is the diagonal matrix of eigen-values, which is also the between-speaker covariance matrix in the transformed space.

The speaker-specific means are:

$$p(\mathbf{m}_s) = \mathcal{N}(\mathbf{m}_s; \mathbf{0}, \mathbf{\Phi}). \tag{3}$$

For convenience we re-parametrize the speaker mean as

$$\mathbf{m}_s = \mathbf{V}\mathbf{y}_s, \tag{4}$$

where diagonal matrix $\mathbf{V} = \mathbf{\Phi}^{\frac{1}{2}}$ and $\mathbf{y}_s$ is a standard normal distributed random variable

$$p(\mathbf{y}_s) = \mathcal{N}(\mathbf{y}_s; \mathbf{0}, \mathbf{I}). \tag{5}$$

The speaker-specific distribution of x-vectors is

$$p(\mathbf{x}_t|\mathbf{y}_s) = \mathcal{N}(\mathbf{x}_t; \mathbf{V}\mathbf{y}_s, \mathbf{I}), \tag{6}$$

where $\mathbf{I}$ is identity matrix.

From the HMM model, state-specific distributions:

$$p(\mathbf{x}_t|z_t = s) = p(\mathbf{x}_t|s) = p(\mathbf{x}_t|\mathbf{y}_s) \tag{7}$$

Transition probabilities:

$$p(z_t = s|z_{t-1} = s') = p(s|s') \tag{8}$$

$$p(s|s') = (1 - P_{loop})\pi_s + \delta(s = s')P_{loop} \tag{9}$$

3 Inference

The joint probability distribution of all the random variables is:

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) &= p(\mathbf{X}|\mathbf{Z}, \mathbf{Y})p(\mathbf{Z})p(\mathbf{Y}) \\ &= \prod_t p\left(\mathbf{x}_t|z_t\right) \prod_t p\left(z_t|z_{t-1}\right) \prod_s p\left(\mathbf{y}_s\right), \end{aligned} \tag{10}$$

We will consider the following factorization for the approximate variational posteriors of the hidden variables (mean field approximation):

$$q(\mathbf{Z}, \mathbf{Y}) = q(\mathbf{Z})q(\mathbf{Y}). \tag{11}$$

The evidence lower bound objective (ELBO) is defined as

$$\mathcal{L}\left(q(\mathbf{Y}, \mathbf{Z})\right) = E_{q(\mathbf{Y},\mathbf{Z})}\left\{\ln\left(\frac{p(\mathbf{X}, \mathbf{Y}, \mathbf{Z})}{q(\mathbf{Y}, \mathbf{Z})}\right)\right\}. \tag{12}$$

Using the factorization (11), the ELBO can be split into three terms

$$\mathcal{L}\left(q(\mathbf{Y}, \mathbf{Z})\right) = E_{q(\mathbf{Y},\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + E_{q(\mathbf{Y})}\left[\ln\frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right] + E_{q(\mathbf{Z})}\left[\ln\frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right], \tag{13}$$

We modify the ELBO by scaling the first two terms by constant factors F_A and F_B.

$$\hat{\mathcal{L}}\left(q(\mathbf{Y}, \mathbf{Z})\right) = F_A E_{q(\mathbf{Y},\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + F_B E_{q(\mathbf{Y})}\left[\ln\frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right] + E_{q(\mathbf{Z})}\left[\ln\frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right], \tag{14}$$

3.1 Useful quantities

The speaker-specific log likelihoods are:

$$\begin{aligned} \ln p(\mathbf{x}_t|\mathbf{y}_s) &= \ln\mathcal{N}(\mathbf{x}_t; \mathbf{V}\mathbf{y}_s, \mathbf{I}) \\ &= \ln\frac{1}{(2\pi)^{\frac{D}{2}}} - \frac{1}{2}(\mathbf{x}_t - \mathbf{V}\mathbf{y}_s)^2 \\ &= \underbrace{-\frac{D}{2}\ln 2\pi - \frac{1}{2}\mathbf{x}_t\mathbf{x}_t^T}_{G(\mathbf{x}_t)} + \mathbf{y}_s^T\underbrace{\mathbf{V}^T\mathbf{x}_t}_{\boldsymbol{\rho}_t} - \frac{1}{2}\mathrm{tr}\left(\mathbf{y}_s\mathbf{y}_s^T\underbrace{\mathbf{V}^T\mathbf{V}}_{\boldsymbol{\Phi}}\right) \\ &= G(\mathbf{x}_t) + \mathbf{y}_s^T\boldsymbol{\rho}_t - \frac{1}{2}\mathrm{tr}\left(\mathbf{y}_s\mathbf{y}_s^T\boldsymbol{\Phi}\right) \end{aligned} \tag{15}$$

3.2 Updating $q(\mathbf{Y})$

To obtain $q(\mathbf{Y})$ we maximize the modified ELBO w.r.t. $q(\mathbf{Y})$ (given fixed $q(\mathbf{Z})$). To do so, we construct the corresponding Lagrangian and set its functional derivative w.r.t. $q(\mathbf{Y})$ equal to zero:

$$\begin{aligned} &\frac{\partial}{\partial q(\mathbf{Y})}\left[\hat{\mathcal{L}}\left(q\left(\mathbf{Y}, \mathbf{Z}\right)\right) + \lambda\left(\int q\left(\mathbf{Y}\right)d\mathbf{Y} - 1\right)\right] = 0 \\ &\frac{\partial\hat{\mathcal{L}}\left(q(\mathbf{Y}, \mathbf{Z})\right)}{\partial q(\mathbf{Y})} + \lambda = 0 \end{aligned} \tag{16}$$

Substituting eq. 14 gives us:

$$\begin{aligned} \frac{\partial\hat{\mathcal{L}}\left(q(\mathbf{Y}, \mathbf{Z})\right)}{\partial q(\mathbf{Y})} &= \frac{\partial}{\partial q(\mathbf{Y})}\left(F_A E_{q(\mathbf{Y}),q(\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + F_B E_{q(\mathbf{Y})}\left[\ln\frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right]\right) \\ &= F_A E_{q(\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + F_B \ln p(\mathbf{Y}) - F_B\left(\ln q(\mathbf{Y}) + 1\right) \end{aligned} \tag{17}$$

Substituting in 16 and solving for $q(\mathbf{Y})$ we obtain:

$$\ln q(\mathbf{Y}) = \frac{F_A}{F_B} E_{q(\mathbf{Z})} \left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + \ln p(\mathbf{Y}) + const. \tag{18}$$

We derive:

$$\ln q(\mathbf{Y}) = \frac{F_A}{F_B} E_{q(\mathbf{Z})} \left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + \ln p(\mathbf{Y}) + const. \tag{19}$$

$$\ln q(\mathbf{Y}) = \frac{F_A}{F_B} E_{q(\mathbf{Z})} \left[\sum_t \sum_s \ln p(\mathbf{x}_t | z_t = s)\right] + \sum_s \ln p(\mathbf{y}_s) + const. \tag{20}$$

$$E[a+b] = E[a] + E[b], \text{ therefore} \tag{21}$$

$$\ln q(\mathbf{Y}) = \sum_s \frac{F_A}{F_B} E_{q(\mathbf{Z})} \left[\sum_t \ln p(\mathbf{x}_t | z_t = s)\right] + \sum_s \ln p(\mathbf{y}_s) + const. \tag{22}$$

where we can see that we obtain the induced factorization

$$\ln q(\mathbf{Y}) = \sum_s \ln q(\mathbf{y}_s), \tag{23}$$

We can then derive the update formula for each speaker model as follows:

$$\begin{aligned}
\ln q(\mathbf{y}_s) =& \frac{F_A}{F_B} E_{q(\mathbf{Z})} \left[\sum_t \ln p(\mathbf{x}_t | z_t = s)\right] + \ln p(\mathbf{y}_s) + const. \\
=& \frac{F_A}{F_B} \sum_t \gamma_{ts} \ln p(\mathbf{x}_t | s) + \ln p(\mathbf{y}_s) + const. \\
=& \frac{F_A}{F_B} \sum_t \gamma_{ts} \left[\mathbf{y}_s^T \boldsymbol{\rho}_t - \frac{1}{2} \mathrm{tr}\left(\mathbf{y}_s \mathbf{y}_s^T \boldsymbol{\Phi}\right)\right] - \frac{1}{2} \mathbf{y}_s^T \mathbf{y}_s + const. \\
=& \frac{F_A}{F_B} \left[\sum_t \gamma_{ts} \boldsymbol{\rho}_t^T\right] \mathbf{y}_s - \frac{1}{2} \mathrm{tr}\left(\left[\frac{F_A}{F_B}\left(\sum_t \gamma_{ts}\right) \boldsymbol{\Phi} + \mathbf{I}\right] \mathbf{y}_s \mathbf{y}_s^T\right) + const.,
\end{aligned} \tag{24}$$

The completion of squares (see A.1.4 in [2]) gives us:

$$q^*(\mathbf{y}_s) = \mathcal{N}\left(\mathbf{y}_s | \boldsymbol{\alpha}_s, \mathbf{L}_s^{-1}\right) \tag{25}$$

which are Gaussians with the mean vector and precision matrix

$$\boldsymbol{\alpha}_s = \frac{F_A}{F_B} \mathbf{L}_s^{-1} \sum_t \gamma_{ts} \boldsymbol{\rho}_t \qquad \mathbf{L}_s = \mathbf{I} + \frac{F_A}{F_B} \left(\sum_t \gamma_{ts}\right) \boldsymbol{\Phi}. \tag{26}$$

3.3 Updating $q(\mathbf{Z})$

To maximize the modified ELBO w.r.t. $q(\mathbf{Z})$ (given fixed $q(\mathbf{Y})$), we solve an equation similar to (16), where symbols $\mathbf{Y}$ and $\mathbf{Z}$ are exchanged.

$$\begin{aligned}
&\frac{\partial}{\partial q(\mathbf{Z})} \left[\hat{\mathcal{L}}\left(q\left(\mathbf{Y}, \mathbf{Z}\right)\right) + \lambda \left(\int q\left(\mathbf{Z}\right) d\mathbf{Z} - 1\right)\right] = 0 \\
&\frac{\partial \hat{\mathcal{L}}\left(q(\mathbf{Y}, \mathbf{Z})\right)}{\partial q(\mathbf{Z})} + \lambda = 0
\end{aligned} \tag{27}$$

This time, solving for $q(\mathbf{Z})$ leads to

$$\begin{aligned}
\frac{\partial \hat{\mathcal{L}}\left(q(\mathbf{Y}, \mathbf{Z})\right)}{\partial q(\mathbf{Z})} =& \frac{\partial}{\partial q(\mathbf{Z})} \left(F_A E_{q(\mathbf{Y}), q(\mathbf{Z})} \left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + E_{q(\mathbf{Z})} \left[\ln \frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right]\right) \\
=& F_A E_{q(\mathbf{Y})} \left[\ln p(\mathbf{X}|\mathbf{Y}, \mathbf{Z})\right] + \ln p(\mathbf{Z}) - \left(\ln q(\mathbf{Z}) + 1\right)
\end{aligned} \tag{28}$$

$$
\begin{aligned}
\ln q(\mathbf{Z}) &= F_A E_{q(\mathbf{Y})}\left[\ln p(\mathbf{X}|\mathbf{Y},\mathbf{Z})\right] + \ln p(\mathbf{Z}) + const. \\
&= F_A E_{q(\mathbf{Y})}\left[\sum_t \ln p(\mathbf{x}_t|z_t)\right] + \ln p(\mathbf{Z}) + const. \\
&= \sum_t \ln \overline{p}(\mathbf{x}_t|z_t) + \ln p(\mathbf{Z}) + const.,
\end{aligned}
\tag{29}
$$

where $\overline{p}(\mathbf{x}_t|z_t = s)$ is defined (using (15) and (25)) as:

$$
\begin{aligned}
&E_{q(\mathbf{Y})}\left[F_A \ln p(\mathbf{x}_t|s)\right] = E_{q(\mathbf{y}_s)}\left[F_A \ln p(\mathbf{x}_t|s)\right] \\
&= F_A \left[\boldsymbol{\alpha}_s^T \boldsymbol{\rho}_t - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Phi}\left[\mathbf{L}_s^{-1} + \boldsymbol{\alpha}_s \boldsymbol{\alpha}_s^T\right]\right) + G(\mathbf{x}_t)\right] \\
&= \ln \overline{p}(\mathbf{x}_t|s)
\end{aligned}
\tag{30}
$$

3.4 The lower bound

Let us repeat here the expression for the ELBO eq. 14:

$$
\hat{\mathcal{L}}\left(q(\mathbf{X},\mathbf{Y})\right) = F_A E_{q(\mathbf{Y},\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y},\mathbf{Z})\right] + F_B E_{q(\mathbf{Y})}\left[\ln \frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right] + E_{q(\mathbf{Z})}\left[\ln \frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right],
$$

The first term of the modified ELBO (14) can be evaluated (using (30)) as

$$
\begin{aligned}
&F_A E_{q(\mathbf{Y},\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y},\mathbf{Z})\right] = \\
&= F_A E_{q(\mathbf{Y},\mathbf{Z})}\left[\sum_t \ln\ p(\mathbf{x}_t|z_t = s)\right] \\
&= F_A E_{q(\mathbf{Y})}\left[\sum_t \sum_s \gamma_{ts}\left(G(\mathbf{x}_t) + \mathbf{y}_s^T \boldsymbol{\rho}_t - \frac{1}{2}\mathrm{tr}\left(\mathbf{y}_s \mathbf{y}_s^T \boldsymbol{\Phi}\right)\right)\right] \\
&= F_A \left[\sum_t \sum_s \gamma_{ts}\left(G(\mathbf{x}_t) + E_{q(\mathbf{Y})}\left[\mathbf{y}_s^T\right]\boldsymbol{\rho}_t - \frac{1}{2}\mathrm{tr}\left(E_{q(\mathbf{Y})}\left[\mathbf{y}_s \mathbf{y}_s^T\right]\boldsymbol{\Phi}\right)\right)\right] \\
&= \sum_t \sum_s \gamma_{ts} \ln \overline{p}(\mathbf{x}_t|s)
\end{aligned}
\tag{31}
$$

Using the factorization (22), the second term of the ELBO (14) (excluding the scalar F_B) can be evaluated as follows. First, the expectation of $p(\mathbf{Y})$

$$
\begin{aligned}
E_{q(\mathbf{Y})}\left[\ln p(\mathbf{Y})\right] &= \sum_s \left\{-\frac{1}{2}\ln(2\pi) + E_{q(\mathbf{Y})}\left[-\frac{1}{2}\mathbf{y}_s^T \mathbf{y}_s\right]\right\} \\
&= \sum_s \left\{-\frac{1}{2}\ln(2\pi) - \frac{1}{2} tr\left(\mathbf{L}_s^{-1} + \boldsymbol{\alpha}_s \boldsymbol{\alpha}_s^T\right)\right\}
\end{aligned}
\tag{32}
$$

And the expectation of the log of the approximate posterior $q(\mathbf{Y})$ is:

$$\begin{aligned}
E_{q(\mathbf{Y})}\left[-\ln q(\mathbf{Y})\right] &= \sum_s E_{q(\mathbf{Y})}\left[\frac{1}{2}\left(\ln(2\pi)+\ln|\mathbf{L_s^{-1}}|+(\mathbf{y}_s-\boldsymbol{\alpha_s})^T\mathbf{L}_s(\mathbf{y}_s-\boldsymbol{\alpha_s})\right)\right] \\
&= \sum_s E_{q(\mathbf{Y})}\left[\underbrace{\frac{1}{2}\ln(2\pi)+\frac{1}{2}\ln|\mathbf{L_s^{-1}}|}_{C_s}+\frac{1}{2}(\mathbf{y}_s-\boldsymbol{\alpha_s})^T\mathbf{L}_s(\mathbf{y}_s-\boldsymbol{\alpha_s})\right] \\
&= \sum_s \left\{C_s+\frac{1}{2}E_{q(\mathbf{Y})}\left[tr\left(\mathbf{L}_s(\mathbf{y}_s\mathbf{y}_s^T-2\boldsymbol{\alpha_s}\mathbf{y}_s^T+\boldsymbol{\alpha_s}\boldsymbol{\alpha_s}^T)\right)\right]\right\} \\
&\quad \text{using the expression from [2] 6.2.2, for the expectation } E[\mathbf{y}_s\mathbf{y}_s^T]\text{:} \\
&= \sum_s \left\{C_s+\frac{1}{2}tr(\mathbf{L}_s\left[\mathbf{L}_s^{-1}+\boldsymbol{\alpha}_s\boldsymbol{\alpha}_s^T\right]-\mathbf{L}_s\boldsymbol{\alpha_s}\boldsymbol{\alpha_s}^T)\right\} \\
&= \sum_s \left\{C_s+\frac{1}{2}tr(\mathbf{I})\right\} \\
&= \sum_s \left\{\frac{1}{2}\ln(2\pi)+\frac{1}{2}\ln|\mathbf{L_s^{-1}}|+\frac{R}{2}\right\}
\end{aligned} \tag{33}$$

Therefore

$$\begin{aligned}
F_B E_{q(\mathbf{Y})}\left[\ln\frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right] &= -F_B\sum_s D_{KL}(q(\mathbf{y}_s)\|p(\mathbf{y}_s)) \\
&= \sum_s \frac{F_B}{2}\left(R+\ln|\mathbf{L_s^{-1}}|-tr(\mathbf{L}_s^{-1})-\boldsymbol{\alpha}_s^T\boldsymbol{\alpha}_s\right),
\end{aligned} \tag{34}$$

Finally, the third term in (14) is the negative KL divergence

$$E_{q(\mathbf{Z})}\left[\ln\frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right] = \sum_{s=1}^{S}\gamma_{1s}\ln\frac{\pi_s}{\gamma_{1s}}+\sum_{t=2}^{T}\sum_{m=1}^{S}\sum_{n=1}^{S}\xi_{tmn}\ln\frac{p(n|m)}{q(z_t{=}n|z_{t-1}{=}m)}, \tag{35}$$

where the approximate marginal probability of transitioning from state m to state n at time t

$$\xi_{tmn} = q(z_{t-1}=m, z_t=n) = \frac{A(t-1,m)\overline{p}(\mathbf{x}_t|n)p(n|m)B(t,n)}{\overline{p}(\mathbf{X})} \tag{36}$$

where $A(t-1,m)$, $B(t,n)$ and $\overline{p}(\mathbf{X})$ can be estimated using the forward-backward algorithm (see equations (19)-(22) in the original paper [1]), $\overline{p}(\mathbf{x}_t|n)$ can be estimated using (30), $p(n|m)$ is the transition probability as defined in (9) and the approximate posterior of transitioning to state n at time t given previous state m

$$q(z_t{=}n|z_{t-1}{=}m) = \frac{\xi_{tmn}}{\sum_s\xi_{tms}}. \tag{37}$$

It can also be seen that the separate expectations can be defined as:

$$E_{q(\mathbf{Z})}\left[\ln q(\mathbf{Z})\right] = \sum_{s=1}^{S}\gamma_{1s}\ln\gamma_{1s}+\sum_{t=2}^{T}\sum_{m=1}^{S}\sum_{n=1}^{S}\xi_{tmn}\ln\frac{\xi_{tmn}}{\sum_o\xi_{tmo}} \tag{38}$$

$$E_{q(\mathbf{Z})}\left[\ln p(\mathbf{Z})\right] = \sum_{s=1}^{S}\gamma_{1s}\ln\pi_s+\sum_{m=1}^{S}\sum_{n=1}^{S}\left(\sum_{t=2}^{T}\xi_{tmn}\right)\ln p(n|m) \tag{39}$$

The complete ELBO is therefore evaluated as:

$$\hat{\mathcal{L}} = \ln\overline{p}(\mathbf{X})+\sum_s\frac{F_B}{2}\left(R+\ln|\mathbf{L_s^{-1}}|-\mathrm{tr}(\mathbf{L}_s^{-1})-\boldsymbol{\alpha}_s^T\boldsymbol{\alpha}_s\right) \tag{40}$$

3.5 Updating π_s

We will obtain the updates for π_s from the ELBO eq. 14:

$$\hat{\mathcal{L}}\left(q(\mathbf{X},\mathbf{Y})\right) = F_A E_{q(\mathbf{Y},\mathbf{Z})}\left[\ln p(\mathbf{X}|\mathbf{Y},\mathbf{Z})\right] + F_B E_{q(\mathbf{Y})}\left[\ln \frac{p(\mathbf{Y})}{q(\mathbf{Y})}\right] + E_{q(\mathbf{Z})}\left[\ln \frac{p(\mathbf{Z})}{q(\mathbf{Z})}\right],$$

where only the term $E_{q(\mathbf{Z})}\left[\ln p(\mathbf{Z})\right]$ depends on π_s.

Given the constrain $\sum_{s=1}^{S}$ we construct the Lagrange multiplier and take the derivative with respect to π:

$$\begin{aligned}
&\frac{\partial}{\partial \pi_k}\left[E_{q(\mathbf{Z})}\left[\ln p(\mathbf{Z})\right] - \lambda\left(\sum_{s=1}^{S}\pi_s - 1\right)\right] = 0 \\
&\frac{\partial}{\partial \pi_k}\left[\sum_{s=1}^{S}\gamma_{1s}\ln \pi_s + \sum_{m=1}^{S}\sum_{n=1}^{S}\left(\sum_{t=2}^{T}\xi_{tmn}\right)\ln p(n|m) - \lambda\left(\sum_{s=1}^{S}\pi_s - 1\right)\right] = 0 \\
&\frac{\gamma_{1k}}{\pi_k} + \sum_{m=1}^{S}\left(\sum_{t=2}^{T}\xi_{tmk}\right)\frac{(1-P_{loop})}{p(k|m)} - \lambda = 0 \\
&\frac{\gamma_{1k}}{\pi_k} + \sum_{m=1}^{S}\left(\sum_{t=2}^{T}\frac{A(t-1,m)\bar{p}(\mathbf{x}_t|k)p(k|m)B(t,k)}{\bar{p}(\mathbf{X})}\right)\frac{(1-P_{loop})}{p(k|m)} - \lambda = 0 \\
&\lambda = \frac{\gamma_{1k}}{\pi_k} + \frac{(1-P_{loop})}{\bar{p}(\mathbf{X})}\sum_{m=1}^{S}\sum_{t=2}^{T}A(t-1,m)\bar{p}(\mathbf{x}_t|k)B(t,k) \\
&\lambda\pi_k = \gamma_{1k} + \frac{(1-P_{loop})\pi_k}{\bar{p}(\mathbf{X})}\sum_{m=1}^{S}\sum_{t=2}^{T}A(t-1,m)\bar{p}(\mathbf{x}_t|k)B(t,k) \\
&\pi_k \propto \gamma_{1k} + \frac{(1-P_{loop})\pi_k}{\bar{p}(\mathbf{X})}\sum_{m=1}^{S}\sum_{t=2}^{T}A(t-1,m)\bar{p}(\mathbf{x}_t|k)B(t,k)
\end{aligned} \tag{41}$$

which is a fixed point iteration, and would, in theory, require iterative updates to obtain the optimal value of π_k, which is not done in practice

References

[1] F. Landini, J. Profant, M. Diez, and L. Burget, "Bayesian HMM clustering of x-vector sequences (VBx) in speaker diarization: theory, implementation and analysis on standard tasks," 2020.

[2] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook." http://www.cs.toronto.edu/~bonner/courses/2012s/csc338/matrix_cookbook.pdf, 2006.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.