A Calculation of each term in the lower-bound

This section presents the calculation of for each term of the ELBO in (4). Note that the variational distribution q is defined in (3).

A.1 $\mathbb{E}_q\left[\ln p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right]$

$$\begin{aligned}
&\mathbb{E}_q\left[\ln p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\
&= \sum_{d=1}^{M} \sum_{\mathbf{z}} q(\mathbf{z}; \mathbf{r}) \ln \ln p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{k=1}^{K} q(z_{dcnk} = 1; r_{dcnk}) \ln p(\mathbf{x}_{dcn}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
&= -\frac{1}{2} \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{dcnk} \left[D \ln(2\pi) + \ln |\boldsymbol{\Sigma}_k| + (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)\right].
\end{aligned} \tag{7}$$

A.2 $\mathbb{E}_q\left[\ln p(\mathbf{z}|\boldsymbol{\theta})\right]$

$$\begin{aligned}
\mathbb{E}_q\left[\ln p(\mathbf{z}|\boldsymbol{\theta})\right] &= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{k=1}^{K} q(z_{dcnk} = 1; r_{dcnk}) \int q(\boldsymbol{\theta}_{dc}; \boldsymbol{\gamma}_{dc}) \ln p(z_{dcnk} = 1|\boldsymbol{\theta}_{dc}) \, \mathrm{d}\boldsymbol{\theta}_{dc} \\
&= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{dcnk} \int \mathrm{Dirichlet}(\boldsymbol{\theta}_{dc}; \boldsymbol{\gamma}_{dc}) \ln \theta_{dck} \, \mathrm{d}\theta_{dck} \\
&= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{dcnk} \ln \tilde{\theta}_{dck},
\end{aligned} \tag{8}$$

where:

$$\boxed{\ln \tilde{\theta}_{dc} = \psi\left(\gamma_{dck}\right) - \psi\left(\sum_{k=1}^{K} \gamma_{dck}\right)}, \tag{9}$$

and $\psi(.)$ is the digamma function.

A.3 $\mathbb{E}_q\left[\ln p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\alpha})\right]$

$$\begin{aligned}
\mathbb{E}_q\left[\ln p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\alpha})\right] &= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{l=1}^{L} q(y_{dcl} = 1; \eta_{dcl}) \int q(\boldsymbol{\theta}_{dc}; \boldsymbol{\gamma}_{dc}) \ln p(\boldsymbol{\theta}_{dc}|\boldsymbol{\alpha}_l) \, \mathrm{d}\boldsymbol{\theta}_{dcl} \\
&= \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{l=1}^{L} \eta_{dcl} \int \mathrm{Dirichlet}(\boldsymbol{\theta}_{dc}; \boldsymbol{\gamma}_{dc}) \ln \mathrm{Dirichlet}(\boldsymbol{\theta}_{dc}|\boldsymbol{\alpha}_l) \, \mathrm{d}\theta_{dcl}.
\end{aligned} \tag{10}$$

Note that the cross-entropy between 2 Dirichlet distributions can be expressed as:

$$\begin{aligned}
\mathcal{H}\left[\mathrm{Dir}\left(\mathbf{x}; \boldsymbol{\alpha}_0\right), \mathrm{Dir}\left(\mathbf{x}; \boldsymbol{\alpha}_1\right)\right] &= -\mathbb{E}_{\mathrm{Dir}(\mathbf{x}; \boldsymbol{\alpha}_0)}\left[\ln \mathrm{Dir}\left(\mathbf{x}; \boldsymbol{\alpha}_1\right)\right] \\
&= -\mathbb{E}_{\mathrm{Dir}(\mathbf{x}; \boldsymbol{\alpha}_0)}\left[-\ln B(\boldsymbol{\alpha_1}) + \sum_{k=1}^{K} (\alpha_{1k} - 1) \ln x_k\right] \\
&= \ln B(\boldsymbol{\alpha}_1) - \sum_{k=1}^{K} (\alpha_{1k} - 1) \left[\psi(\alpha_{0k}) - \psi\left(\sum_{k'=1}^{K} \alpha_{0k'}\right)\right],
\end{aligned} \tag{11}$$

where:

$$\ln B(\boldsymbol{\alpha}_1) = \sum_{k=1}^{K} \ln \Gamma\left(\alpha_{1k}\right) - \ln \Gamma\left(\sum_{j=1}^{K} \alpha_{1j}\right). \tag{12}$$

Hence:

$$\mathbb{E}_q\left[\ln p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\alpha})\right] = \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{l=1}^{L}\eta_{dcl}\left[-\ln B(\boldsymbol{\alpha}_l) + \sum_{k=1}^{K}(\alpha_{lk}-1)\ln\tilde{\theta}_{dck}\right], \tag{13}$$

where $\ln\tilde{\theta}_{dck}$ is defined in Eq. (9).

A.4 $\mathbb{E}_q\left[\ln p(\mathbf{y}|\boldsymbol{\phi})\right]$

$$\begin{aligned}
\mathbb{E}_q\left[\ln p(\mathbf{y}|\boldsymbol{\phi})\right] &= \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{l=1}^{L} q(y_{dcl}=1;\eta_{dcl})\int q(\boldsymbol{\phi}_d;\boldsymbol{\lambda}_d)\ln p(y_{dcl}=1|\phi_{dl})\,\mathrm{d}\phi_{dl} \\
&= \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{l=1}^{L}\eta_{dcl}\int \mathrm{Dirichlet}(\boldsymbol{\phi}_d;\boldsymbol{\lambda}_d)\ln\phi_{dl}\,\mathrm{d}\phi_{dl} \\
&= \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{l=1}^{L}\eta_{dcl}\ln\tilde{\phi}_{dl},
\end{aligned} \tag{14}$$

where:

$$\boxed{\ln\tilde{\phi}_{dl} = \psi\left(\lambda_{dl}\right) - \psi\left(\sum_{j=1}^{K}\lambda_{dl}\right)} \tag{15}$$

A.5 $\mathbb{E}_q\left[\ln p(\boldsymbol{\phi}|\boldsymbol{\delta})\right]$

$$\begin{aligned}
\mathbb{E}_q\left[\ln p(\boldsymbol{\phi}|\boldsymbol{\delta})\right] &= \sum_{d=1}^{M}\int q(\boldsymbol{\phi}_d;\boldsymbol{\lambda}_d)\ln p(\boldsymbol{\phi}_d|\boldsymbol{\delta})\,d\boldsymbol{\phi}_d \\
&= \sum_{d=1}^{M}\int \mathrm{Dirichlet}_L(\boldsymbol{\phi}_d;\boldsymbol{\lambda}_d)\ln\mathrm{Dirichlet}_L(\boldsymbol{\phi}_d|\boldsymbol{\delta})\,d\boldsymbol{\phi}_d \\
&= \sum_{d=1}^{M} -\ln B(\boldsymbol{\delta}) + \sum_{l=1}^{L}(\delta_l-1)\ln\tilde{\phi}_{dl},
\end{aligned} \tag{16}$$

where $\ln\tilde{\phi}_{dl}$ is defined in Eq. (15).

A.6 $\mathbb{E}_q\left[\ln q(\mathbf{z})\right]$

$$\mathbb{E}_q\left[\ln q(\mathbf{z})\right] = \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{dcnk}\ln r_{dcnk}. \tag{17}$$

A.7 $\mathbb{E}_q\left[\ln q(\boldsymbol{\theta})\right]$

$$\mathbb{E}_q\left[\ln q(\boldsymbol{\theta})\right] = \sum_{d=1}^{M}\sum_{c=1}^{C} -\ln B(\boldsymbol{\gamma}_{dc}) + \sum_{j=1}^{K}(\gamma_{dck}-1)\ln\tilde{\theta}_{dck}, \tag{18}$$

where $\ln\tilde{\theta}_{dck}$ is defined in Eq. (9).

A.8 $\mathbb{E}_q\left[\ln q(\mathbf{y})\right]$

$$\mathbb{E}_q\left[\ln q(\mathbf{y})\right] = \sum_{d=1}^{M}\sum_{c=1}^{C}\sum_{l=1}^{L}\eta_{dcl}\ln\eta_{dcl}. \tag{19}$$

A.9 $\mathbb{E}_q [\ln q(\boldsymbol{\phi})]$

$$\mathbb{E}_q [\ln q(\boldsymbol{\phi})] = \sum_{d=1}^{M} - \ln B(\boldsymbol{\lambda}_d) + \sum_{l=1}^{L} (\lambda_{dl} - 1) \ln \tilde{\phi}_{dl}, \tag{20}$$

where $\ln \tilde{\phi}_{dl}$ is defined in Eq. (15).

B Optimisation of the lower-bound

B.1 Variational categorical for z

The terms in the lower-bound that relates to r_{dcnk} are:

$$\begin{aligned} \mathsf{L} = & -\frac{1}{2} r_{dcnk} \left[D \ln(2\pi) + \ln |\boldsymbol{\Sigma}_k| + (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k) \right] + r_{dcnk} \ln \tilde{\theta}_{dck} \\ & - r_{dcnk} \ln r_{dcnk} + \zeta \left(\sum_{k=1}^{K} r_{dcnk} - 1 \right), \end{aligned} \tag{21}$$

where $\ln \tilde{\theta}_{dck}$ is defined in Eq. (9), and ζ is the Lagrange multiplier due to the assumption that $\mathbf{r}_{dcn}$ is the parameter of a categorical distribution, which requires:

$$\sum_{k=1}^{K} r_{dcnk} = 1. \tag{22}$$

Taking the derivative w.r.t. r_{dcnk} gives:

$$\begin{aligned} \frac{\partial \mathsf{L}}{\partial r_{dcnk}} = & -\frac{1}{2} \left[D \ln(2\pi) + \ln |\boldsymbol{\Sigma}_k| + (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k) \right] \\ & + \ln \tilde{\theta}_{dck} - \ln r_{dcnk} - 1 + \zeta \end{aligned} \tag{23}$$

Setting the derivative to zero yields the maximizing value of the variational parameter r_{dcnk} as:

$$\boxed{r_{dcnk} \propto \exp \left\{ \ln \tilde{\theta}_{dck} - \frac{1}{2} \left[D \ln(2\pi) + \ln |\boldsymbol{\Sigma}_k| + (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k) \right] \right\}.} \tag{24}$$

B.2 Variational Dirichlet for $\boldsymbol{\theta}$

The lower-bound isolating the terms for γ_{dck} is written as:

$$\begin{aligned} \mathsf{L} = & \sum_{n=1}^{N} \sum_{k=1}^{K} r_{dcnk} \ln \tilde{\theta}_{dck} + \sum_{l=1}^{L} \eta_{dcl} \sum_{k=1}^{K} (\alpha_{lk} - 1) \ln \tilde{\theta}_{dck} - \ln B(\boldsymbol{\gamma}_{dc}) \\ & + \sum_{k=1}^{K} (\gamma_{dck} - 1) \ln \tilde{\theta}_{dck} \\ = & - \ln B(\boldsymbol{\gamma}_{dc}) + \sum_{k=1}^{K} \ln \tilde{\theta}_{dck} \left[\sum_{n=1}^{N} r_{dcnk} + \sum_{l=1}^{L} \eta_{dcl} (\alpha_{lk} - 1) + \gamma_{dck} - 1 \right], \end{aligned} \tag{25}$$

where $\ln \tilde{\theta}_{dck}$ is defined in Eq. (9).

Taking derivative w.r.t. γ_{dck} gives:

$$\begin{aligned} \frac{\partial \mathsf{L}}{\partial \gamma_{dck}} = & \Psi(\gamma_{dck}) \left[\sum_{n=1}^{N} r_{dcnk} + \sum_{l=1}^{L} \eta_{dcl} (\alpha_{lk} - 1) - \gamma_{dck} + 1 \right] \\ & - \Psi \left(\sum_{j=1}^{K} \gamma_{dcj} \right) \sum_{j=1}^{K} \left[\sum_{n=1}^{N} r_{dcnj} + \sum_{l=1}^{L} \eta_{dcl} (\alpha_{lj} - 1) - \gamma_{dcj} + 1 \right]. \end{aligned} \tag{26}$$

Setting the derivative to zero and solve for γ_{dck} yields:

$$\boxed{\gamma_{dck} = 1 + \sum_{n=1}^{N} r_{dcnk} + \sum_{l=1}^{L} \eta_{dcl}(\alpha_{lk} - 1).} \tag{27}$$

B.3 Variational categorical for y

Note that the L-dimensional vector $\boldsymbol{\eta}_{dc}$ is the parameter of a categorical distribution for $\mathbf{y}_{dc}$, it satisfies the following constrain:

$$\sum_{l=1}^{L} \eta_{dcl} = 1. \tag{28}$$

The Lagrangian can be expressed as:

$$\begin{aligned} \mathsf{L}[\mathbf{y}_{dc}] = &\sum_{l=1}^{L} \eta_{dcl} \left[-\ln B(\boldsymbol{\alpha}_l) + \sum_{k=1}^{K} (\alpha_{lk} - 1) \ln \tilde{\theta}_{dck} \right] \\ &+ \sum_{l=1}^{L} \eta_{dcl} \ln \tilde{\phi}_{dl} - \sum_{l=1}^{L} \eta_{dcl} \ln \eta_{dcl} + \xi \left(\sum_{l=1}^{L} \eta_{dcl} - 1 \right), \end{aligned} \tag{29}$$

where ξ is the Lagrange multiplier, $\ln \tilde{\theta}_{dck}$ is defined in Eq. (9), and $\ln \tilde{\phi}_{dl}$ is defined in Eq. (15).

Taking the derivative w.r.t. η_{dcl} gives:

$$\frac{\partial \mathsf{L}}{\partial \eta_{dcl}} = -\ln B(\boldsymbol{\alpha}_l) + \sum_{k=1}^{K} (\alpha_{lk} - 1) \ln \tilde{\theta}_{dck} + \psi(\lambda_{dl}) - \psi \left(\sum_{j=1}^{L} \lambda_{dl} \right) - \ln \eta_{dcl} - 1 + \xi. \tag{30}$$

Setting the derivative to zero and solve for η_{dcl} yields:

$$\boxed{\eta_{dcl} \propto \exp \left[\ln \tilde{\phi}_{dl} - \ln B(\boldsymbol{\alpha}_l) + \sum_{k=1}^{K} (\alpha_{lk} - 1) \ln \tilde{\theta}_{dck} \right].} \tag{31}$$

B.4 Variational Dirchlet for ϕ

The lower-bound isolating the terms for λ_{dl} is written as:

$$\begin{aligned} \mathsf{L} &= \sum_{c=1}^{C} \sum_{l=1}^{L} \eta_{dcl} \ln \tilde{\phi}_{dl} + \sum_{l=1}^{L} (\delta_l - 1) \ln \tilde{\phi}_{dl} + \ln B(\boldsymbol{\lambda}_d) - \sum_{l=1}^{L} (\lambda_{dl} - 1) \ln \tilde{\phi}_{dl} \\ &= \ln B(\boldsymbol{\lambda}_d) + \sum_{l=1}^{L} \ln \tilde{\phi}_{dl} \left(\delta_l - \lambda_{dl} + \sum_{c=1}^{C} \eta_{dcl} \right), \end{aligned} \tag{32}$$

where $\ln \tilde{\phi}_{dl}$ is defined in Eq. (15).

Taking derivative w.r.t. λ_{dl} gives:

$$\frac{\partial \mathsf{L}}{\partial \lambda_{dl}} = \Psi(\lambda_{dl}) \left(\delta_l - \lambda_{dl} + \sum_{c=1}^{C} \eta_{dcl} \right) - \Psi \left(\sum_{j=1}^{L} \lambda_{dj} \right) \sum_{l=1}^{L} \left(\delta_l - \lambda_{dl} + \sum_{c=1}^{C} \eta_{dcl} \right). \tag{33}$$

Setting to zero and solving for λ_{dl} gives:

$$\boxed{\lambda_{dl} = \delta_l + \sum_{c=1}^{C} \eta_{dcl}.} \tag{34}$$

B.5 Maximum likelihood for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

The lower-bound isolating the terms for $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ is written as:

$$\mathsf{L} = -\frac{1}{2} \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk} \left[D \ln(2\pi) + \ln |\boldsymbol{\Sigma}_k| + (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k) \right]. \tag{35}$$

Taking derivative w.r.t. $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ gives:

$$\begin{cases} \dfrac{\partial \mathsf{L}}{\partial \boldsymbol{\mu}_k} &= \displaystyle\sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k) \\ \dfrac{\partial \mathsf{L}}{\partial \boldsymbol{\Sigma}_k} &= \displaystyle -\frac{1}{2} \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk} \left[\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)(\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} \right]. \end{cases} \tag{36}$$

Setting the derivative to zero yields the maximizing values at:

$$\boxed{\begin{aligned} \boldsymbol{\mu}_k &= \frac{1}{\sum_{d=1}^{M} N_{dk}} \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk} \mathbf{x}_{dcn} \\ \boldsymbol{\Sigma}_k &= \frac{1}{\sum_{d=1}^{M} N_{dk}} \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk} (\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)(\mathbf{x}_{dcn} - \boldsymbol{\mu}_k)^\top, \end{aligned}} \tag{37}$$

where:

$$N_{dk} = \sum_{c=1}^{C} \sum_{n=1}^{N} r_{dcnk}. \tag{38}$$

Note that the inference results for image-themes $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^{K}$ in (37) is very similar to the result of EM algorithm derived for a Gaussian mixture model (Bishop, 2006, Chapter 9). The result is, consequently, often suffered from the singularity issue happened in the MLE for a Gaussian mixture model. The issue is due to one of the Gaussian components collapses (or overfit) to a single data point, resulting in a zero covariance matrix. In the implementation, we add a small value (about 10^{-6}) diagonal matrix to the covariance matrices to avoid this problem.

B.6 MLE for Dirichlet parameter $\boldsymbol{\alpha}$

The terms in ELBO which contains $\boldsymbol{\alpha}$ are:

$$\mathsf{L}[\boldsymbol{\alpha}] = \sum_{d=1}^{M} \sum_{c=1}^{C} \sum_{l=1}^{L} \eta_{dcl} \left[\ln \Gamma \left(\sum_{k=1}^{K} \alpha_{lk} \right) - \sum_{k=1}^{K} \ln \Gamma (\alpha_{lk}) + \sum_{k=1}^{K} (\alpha_{lk} - 1) \ln \tilde{\theta}_{dck} \right]. \tag{39}$$

Taking the derivative w.r.t. α_{lk} gives:

$$\frac{\partial \mathsf{L}}{\partial \alpha_{lk}} = g_{lk} = M \left[\psi \left(\sum_{k=1}^{K} \alpha_{lk} \right) - \psi (\alpha_{lk}) \right] \sum_{c=1}^{C} \eta_{dcl} + \sum_{d=1}^{M} \sum_{c=1}^{C} \eta_{dcl} \ln \tilde{\theta}_{dck}. \tag{40}$$

The Hessian matrix can be calculated as:

$$\frac{\partial^2 \mathsf{L}}{\partial \alpha_{lk} \partial \alpha_{lj}} = \underbrace{M \left[\sum_{c=1}^{C} \eta_{dcl} \right] \Psi \left(\sum_{k=1}^{K} \alpha_{lk} \right)}_{u} \underbrace{- M \left[\sum_{c=1}^{C} \eta_{dcl} \right] \mathbb{1}[k = j] \, \Psi (\alpha_{lk})}_{q_{ljk}}. \tag{41}$$

According to (Minka, 2000), Newton-Raphson method can be used to infer $\boldsymbol{\alpha}_l$ as:

$$\boldsymbol{\alpha}_l \leftarrow \boldsymbol{\alpha}_l - \mathbf{H}_l^{-1} \mathbf{g}_l \tag{42}$$

$$\mathbf{H}_l^{-1} = \mathbf{Q}_l^{-1} - \frac{\mathbf{Q}_l^{-1} \mathbf{1} \mathbf{1}^\top \mathbf{Q}_l^{-1}}{1/u + \mathbf{1}^\top \mathbf{Q}_l^{-1} \mathbf{1}} \tag{43}$$

$$\left(\mathbf{H}_l^{-1} \mathbf{g}_l \right)_l = \frac{g_{lk} - b_l}{q_{lkk}}, \tag{44}$$

where:

$$b_l = \frac{\mathbf{1}^\top \mathbf{Q}_l^{-1}\mathbf{g}_l}{1/u + \mathbf{1}^\top \mathbf{Q}_l^{-1}\mathbf{1}} = \frac{\sum_j g_{lj} q_{ljj}}{1/u + \sum_j 1/q_{ljj}}. \tag{45}$$

C Learning algorithm

Algorithm 1 Online continuous LDCC

require Scalar hyper-parameters: e-step stopping criteria $\Delta\lambda$, learning rate parameters τ_0, τ_1, and symmetric Dirichlet prior parameter δ

procedure TRAINING
  Initialise $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$ and $\{\boldsymbol{\alpha}_l\}_{l=1}^L$
  for $d = 1, M$ **do**
    $\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \mathbf{r} \leftarrow$ E-STEP($\mathbf{x}_d, \Delta\lambda$) ▷ E-step
    Calculate "local" image-theme $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\}$ ▷ M-step - Eq. (37)
    Calculate the inverse of the Hessian times the gradient $\mathbf{H}^{-1}\mathbf{g}$ ▷ Eq. (44)
    Update learning rate: $\rho_d = (\tau_0 + d)^{-\tau_1}$
    $\boldsymbol{\mu} \leftarrow (1 - \rho_d)\boldsymbol{\mu} + \rho_d \tilde{\boldsymbol{\mu}}$ ▷ Eq. 6
    $\boldsymbol{\Sigma} \leftarrow (1 - \rho_d)\boldsymbol{\Sigma} + \rho_d \tilde{\boldsymbol{\Sigma}}$
    $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} - \rho_d \mathbf{H}^{-1}\mathbf{g}$
  end for
  return $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}$
end procedure

procedure E-STEP($\mathbf{x}, \Delta\lambda$)
  Initialise $\mathbf{r}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\lambda}$
  repeat
    calculate un-normalised r_{cnk}, where $n \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}$ ▷ Eq. (24)
    normalize $\mathbf{r}_{cn}$ such that $\sum_{k=1}^K r_{cnk} = 1$
    calculate γ_{ck} ▷ Eq. (27)
    calculate η_{cl}, where: $l \in \{1, \ldots, L\}$ ▷ Eq. (31)
    normalize $\boldsymbol{\eta}_c$ such that $\sum_{l=1}^L \eta_{cl} = 1$
    calculate λ_l ▷ Eq. (34)
  until $\frac{1}{L}$ |change in $\boldsymbol{\lambda}$| $< \Delta\lambda$
  return $\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \mathbf{r}$
end procedure