Supplemental Material
Evolving Cluster Mixed-Membership Blockmodel for Time-Varying Networks

A Variational EM Algorithm

Our goal is to find the posterior distribution of the latent variables μ, c, γ, z given the observed sequence network $E^{(1)}, \ldots, E^{(T)}$, under the maximum likelihood model parameters $B, \delta, \nu, \Phi, \Sigma$. Finding the posterior (inference) or solving for the maximum likelihood parameters (learning) are both intractable under our original model. Hence we resort to a Variational EM algorithm, which locally optimizes the model parameters with respect to a lower bound on the true marginal log-likelihood, while simultaneously finding a variational distribution that approximates the latent variable posterior. The marginal log-likelihood lower bound being optimized is

$$\begin{aligned}
\log p(E \mid \Theta) &= \log \int_X p(E, X \mid \Theta)\, dX \\
&= \log \int_X q(X) \frac{p(E, X \mid \Theta)}{q(X)} dX \\
&\geq \int_X q(X) \log \frac{p(E, X \mid \Theta)}{q(X)} dX \quad \text{(Jensen's inequality)} \\
&= \mathbb{E}_q \left[\log p(E, X \mid \Theta) - \log q(X)\right] =: \mathcal{L}(q, \Theta)
\end{aligned}$$

where X denotes the latent variables $\{\mu, c, \gamma, z\}$, Θ denotes the model parameters $\{B, \delta, \nu, \Phi, \Sigma\}$, and q is the variational distribution. This lower bound is iteratively maximized with respect to q's parameters (E-step) and the model parameters Θ(M-step).

In principle, the lower bound $\mathcal{L}(q, \Theta)$ holds for any distribution q; ideally q should closely approximate the true posterior $p(X \mid E, \Theta)$. In the next section, we define a factored form for q and derive its optimal solution.

B Variational Distribution q

We assume a factorized form for q:

$$q = q_\mu\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)}\right) \prod_{t,i=1}^{T,N} \left[q_\gamma\left(\gamma_i^{(t)}\right) q_c\left(c_i^{(t)}\right) \prod_{j \neq i}^{N} q_z\left(z_{i\rightarrow j}^{(t)}, z_{i\leftarrow j}^{(t)}\right) \right]$$

We now make use of Generalized Mean Field (GMF) theory (Xing et al. 2003) to determine each factor's form. GMF theory optimizes a lower bound on the marginal distribution $p(E \mid \Theta)$ over arbitrary choices of $q_\mu, q_\gamma, q_c, q_z$. In particular, the optimal solution to q_X is $p\left(X \mid E, \mathbb{E}_q\left[\phi\left(\mathrm{MB}_X\right)\right]\right)$, the distribution of the latent variable set X conditioned on the observed variables E and the *expected exponential family sufficient statistics* (under q) of X's Markov Blanket variables. More precisely, q_X has the same functional form as $p(X \mid E, \mathrm{MB}_X)$, but where a variational parameter $\mathcal{V}$ replaces $\phi(Y)$ for each $Y \in \mathrm{MB}_X$, with optimal solution $\mathcal{V} := \mathbb{E}_q[\phi(Y)]$. In general, if $Y \in \mathrm{MB}_X$, then we shall use $\langle \phi(Y) \rangle$ to denote the variational parameter corresponding to Y.

We begin by deriving optimal solutions to $q_\mu, q_\gamma, q_c, q_z$ in terms of the the variational parameters $\langle \phi(Y) \rangle$. After we have derived all factors, we will present closed-form solutions to $\langle \phi(Y) \rangle$. These solutions form a set of fixed-point equations which, when iterated, converge to a local optimum in the space of variational parameters (thus completing the E-step).

B.1 Distribution of q_z

q_z is a discrete distribution since the zs are indicator vectors. We begin by deriving the distribution of the zs conditioned on their Markov Blanket:

$$
\begin{aligned}
& p\left(z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)} \mid \mathrm{MB}_{z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}}\right) \\
\propto\ & p\left(E_{ij}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}\right) p\left(z_{i\to j}^{(t)} \mid \gamma_i^{(t)}\right) p\left(z_{i\leftarrow j}^{(t)} \mid \gamma_j^{(t)}\right) \\
=\ & \left(\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right)^{E_{ij}^{(t)}} \left(1-\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right)^{1-E_{ij}^{(t)}} \prod_{k=1}^{K} \left(\frac{\exp \gamma_{i,k}^{(t)}}{\sum_{l=1}^{K} \exp \gamma_{i,l}^{(t)}}\right)^{z_{i\to j,k}^{(t)}} \left(\frac{\exp \gamma_{j,k}^{(t)}}{\sum_{l=1}^{K} \exp \gamma_{j,l}^{(t)}}\right)^{z_{i\leftarrow j,k}^{(t)}} \\
\propto\ & \exp\left\{E_{ij}^{(t)} \log\left(\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right) + \left(1-E_{ij}^{(t)}\right)\log\left(1-\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right) + \left(z_{i\to j}^{(t)}\right)^\top \gamma_i^{(t)} + \left(z_{i\leftarrow j}^{(t)}\right)^\top \gamma_j^{(t)}\right\}
\end{aligned}
$$

The variables $\gamma_i^{(t)}, \gamma_j^{(t)}$ belong to other variational factors, and their exponential family sufficient statistics are just $\gamma_i^{(t)}$ and $\gamma_j^{(t)}$ themselves. Hence

$$
\begin{aligned}
& q_z\left(z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}\right) \\
:\propto\ & \exp\left\{E_{ij}^{(t)} \log\left(\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right) + \left(1-E_{ij}^{(t)}\right)\log\left(1-\left(z_{i\to j}^{(t)}\right)^\top B z_{i\leftarrow j}^{(t)}\right) + \left(z_{i\to j}^{(t)}\right)^\top \left\langle\gamma_i^{(t)}\right\rangle + \left(z_{i\leftarrow j}^{(t)}\right)^\top \left\langle\gamma_j^{(t)}\right\rangle\right\}
\end{aligned}
$$

with variational parameters $\left\langle\gamma_i^{(t)}\right\rangle$ and $\left\langle\gamma_j^{(t)}\right\rangle$. We can also express q_z in terms of indices k, l:

$$
q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right) \quad :\propto \quad \exp\left\{E_{ij}^{(t)} \log B_{k,l} + \left(1-E_{ij}^{(t)}\right)\log\left(1-B_{k,l}\right) + \left\langle\gamma_{i,k}^{(t)}\right\rangle + \left\langle\gamma_{j,l}^{(t)}\right\rangle\right\}
$$

B.2 Distribution of q_γ

q_γ is a continuous distribution. The distribution of $\gamma_i^{(t)}$ conditioned on its Markov Blanket is

$$
\begin{aligned}
& p\left(\gamma_i^{(t)} \mid \mathrm{MB}_{\gamma_i^{(t)}}\right) \\
\propto\ & p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}\right) \prod_{j\neq i}^{N} p\left(z_{i\to j}^{(t)} \mid \gamma_i^{(t)}\right) p\left(z_{j\leftarrow i}^{(t)} \mid \gamma_i^{(t)}\right) \\
\propto\ & \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right\} \\
& \prod_{j\neq i}^{N} \prod_{k=1}^{K} \left(\frac{\exp \gamma_{i,k}^{(t)}}{\sum_{l=1}^{K} \exp \gamma_{i,l}^{(t)}}\right)^{z_{i\to j,k}^{(t)}} \left(\frac{\exp \gamma_{i,k}^{(t)}}{\sum_{l=1}^{K} \exp \gamma_{i,l}^{(t)}}\right)^{z_{j\leftarrow i,k}^{(t)}} \\
=\ & \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right. \\
& \left. + \sum_{j\neq i}^{N} \sum_{k=1}^{K} \left(z_{i\to j,k}^{(t)} \gamma_{i,k}^{(t)} + z_{j\leftarrow i,k}^{(t)} \gamma_{i,k}^{(t)}\right) - (2N-2) \log \sum_{l=1}^{K} \exp \gamma_{i,l}^{(t)}\right\} \\
\propto\ & \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \left[\left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1} \gamma_i^{(t)} - \left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1} \mu_h^{(t)} - \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1} \gamma_i^{(t)}\right]\right. \\
& \left. + \left(\sum_{j\neq i}^{N} z_{i\to j}^{(t)} + z_{j\leftarrow i}^{(t)}\right)^\top \gamma_i^{(t)} - (2N-2) \log \sum_{l=1}^{K} \exp \gamma_{i,l}^{(t)}\right\}
\end{aligned}
$$

The variables $c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}, z_{i\to 1}^{(t)}, \ldots, z_{i\to N}^{(t)}, z_{1\leftarrow i}^{(t)}, \ldots, z_{N\leftarrow i}^{(t)}$ belong to other variational factors. The sufficient statistics for variables z are just $z_{i\to j}^{(t)}$ and $z_{j\leftarrow i}^{(t)}$ themselves. For variables c and μ, their sufficient statistics are $c_{i,h}^{(t)}$ and

$c_{i,h}^{(t)} \left(\mu_h^{(t)}\right)^\top$. However, since c is marginally independent of μ under q, we can take their expectations independently, hence the variational parameters are just $\left\langle c_{i,h}^{(t)} \right\rangle$ and $\left\langle \mu_h^{(t)} \right\rangle$. Hence

$$
\begin{aligned}
q_\gamma\left(\gamma_i^{(t)}\right) \quad :\propto \quad & \exp\left\{\sum_{h=1}^{C} -\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle \left[\left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\gamma_i^{(t)} - \left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\left\langle \mu_h^{(t)}\right\rangle - \left\langle \mu_h^{(t)}\right\rangle^\top \Sigma_h^{-1}\gamma_i^{(t)}\right] \right. \\
& \left. + \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top \gamma_i^{(t)} - (2N-2)\log\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}\right\}
\end{aligned}
$$

with variational parameters $\left\langle c_i^{(t)}\right\rangle, \left\langle \mu_h^{(t)}\right\rangle, \left\langle z_{i\to j}^{(t)}\right\rangle, \left\langle z_{j\leftarrow i}^{(t)}\right\rangle$.

B.2.1 Laplace Approximation to q_γ

The term $\mathcal{Z}_\gamma\left(\gamma_i^{(t)}\right) := \log\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}$ makes the exponent analytically un-integrable, which prevents us from computing the normalizer for $q_\gamma\left(\gamma_i^{(t)}\right)$. Thus, we approximate $\mathcal{Z}_\gamma\left(\gamma_i^{(t)}\right)$ with its second-order Taylor expansion around a chosen point $\hat{\gamma}_i^{(t)}$:

$$
\begin{aligned}
\mathcal{Z}_\gamma\left(\gamma_i^{(t)}\right) \quad &\approx \quad \mathcal{Z}_\gamma\left(\hat{\gamma}_i^{(t)}\right) + \left(g_i^{(t)}\right)^\top\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right) + \frac{1}{2}\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right)^\top H_i^{(t)}\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right) \qquad (1) \\
g_{i,k}^{(t)} \quad &:= \quad \frac{\exp\hat{\gamma}_{i,k}^{(t)}}{\sum_{k'=1}^{K}\exp\hat{\gamma}_{i,k'}^{(t)}} \\
H_{i,kl}^{(t)} \quad &:= \quad \frac{\mathbb{I}\left[k=l\right]\exp\hat{\gamma}_{i,k}^{(t)}}{\sum_{k'=1}^{K}\exp\hat{\gamma}_{i,k'}^{(t)}} - \frac{\exp\hat{\gamma}_{i,k}^{(t)}\exp\hat{\gamma}_{i,l}^{(t)}}{\left(\sum_{k'=1}^{K}\exp\hat{\gamma}_{i,k'}^{(t)}\right)^2}
\end{aligned}
$$

Note that $H_i^{(t)} = \operatorname{diag}\left(g_i^{(t)}\right) - g_i^{(t)}\left(g_i^{(t)}\right)^\top$. Because the Variational EM algorithm is iterative, we set $\hat{\gamma}_i^{(t)}$ to $\tilde{\gamma}_i^{(t)} := \mathbb{E}_q\left[\gamma_i^{(t)}\right]$ from the previous iteration, which should keep the point of expansion close to $\mathbb{E}_q\left[\gamma_i^{(t)}\right]$ for the current iteration. The point of this Taylor expansion is to approximate q_γ with a normal distribution — consider the exponent of q_γ,

$$
\begin{aligned}
& -\left\{\sum_{h=1}^{C}\frac{\left\langle c_{i,h}^{(t)}\right\rangle}{2}\left[\left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\gamma_i^{(t)} - \left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\left\langle \mu_h^{(t)}\right\rangle - \left\langle \mu_h^{(t)}\right\rangle^\top \Sigma_h^{-1}\gamma_i^{(t)}\right]\right\} \\
& + \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top \gamma_i^{(t)} - (2N-2)\,\mathcal{Z}_\gamma\left(\gamma_i^{(t)}\right) \\
= \quad & \mathrm{const}^{(1)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) + \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top \gamma_i^{(t)} - (2N-2)\,\mathcal{Z}_\gamma\left(\gamma_i^{(t)}\right)
\end{aligned}
$$

where $\mathrm{const}^{(i)}$ denotes a constant independent of $\gamma_i^{(t)}$, $S := \sum_{h=1}^{C}\Sigma_h^{-1}\left\langle c_{i,h}^{(t)}\right\rangle$ and $u := S^{-1}\left(\sum_{h=1}^{C}\Sigma_h^{-1}\left\langle c_{i,h}^{(t)}\right\rangle\left\langle \mu_h^{(t)}\right\rangle\right)$. Applying the Taylor expansion in eq (1) gives

$$
\begin{aligned}
\approx \quad & \mathrm{const}^{(1)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) + \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top \gamma_i^{(t)} \\
& - (2N-2)\left[\mathcal{Z}_\gamma\left(\hat{\gamma}_i^{(t)}\right) + \left(g_i^{(t)}\right)^\top\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right) + \frac{1}{2}\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right)^\top H_i^{(t)}\left(\gamma_i^{(t)} - \hat{\gamma}_i^{(t)}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \text{const}^{(2)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) + \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top \gamma_i^{(t)} \\
&\quad - (2N-2)\left[\left(g_i^{(t)}\right)^\top \gamma_i^{(t)} + \frac{1}{2}\left(\gamma_i^{(t)}\right)^\top H_i^{(t)}\gamma_i^{(t)} - \left(\hat{\gamma}_i^{(t)}\right)^\top H_i^{(t)}\gamma_i^{(t)}\right] \\
&= \text{const}^{(2)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) \\
&\quad + \left[\left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top - (2N-2)\left(\left(g_i^{(t)}\right)^\top - \left(\hat{\gamma}_i^{(t)}\right)^\top H_i^{(t)}\right)\right]\gamma_i^{(t)} - (N-1)\left(\gamma_i^{(t)}\right)^\top H_i^{(t)}\gamma_i^{(t)}
\end{aligned}
$$

Define $A := \left(\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)^\top - (2N-2)\left(\left(g_i^{(t)}\right)^\top - \left(\hat{\gamma}_i^{(t)}\right)^\top H_i^{(t)}\right)$ and $B := -(N-1)\,H_i^{(t)}$, so that we obtain

$$
\begin{aligned}
&= \text{const}^{(2)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) + A\gamma_i^{(t)} + \left(\gamma_i^{(t)}\right)^\top B\gamma_i^{(t)} \\
&= \text{const}^{(2)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top S\left(\gamma_i^{(t)} - u\right) + A\left(\gamma_i^{(t)} - u + u\right) + \left(\gamma_i^{(t)} - u + u\right)^\top B\left(\gamma_i^{(t)} - u + u\right) \\
&= \text{const}^{(3)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top (S - 2B)\left(\gamma_i^{(t)} - u\right) + \left(A + 2u^\top B\right)\left(\gamma_i^{(t)} - u\right)
\end{aligned}
$$

Finally, define $D := A + 2u^\top B$ and $E := S - 2B$, resulting in

$$
\begin{aligned}
&= \text{const}^{(3)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top E\left(\gamma_i^{(t)} - u\right) + D\left(\gamma_i^{(t)} - u\right) \\
&= \text{const}^{(4)} - \frac{1}{2}\left(\gamma_i^{(t)} - u\right)^\top E\left(\gamma_i^{(t)} - u\right) + \left(E^{-1}D^\top\right)^\top E\left(\gamma_i^{(t)} - u\right) - \frac{1}{2}\left(E^{-1}D^\top\right)^\top E\left(E^{-1}D^\top\right) \\
&= \text{const}^{(4)} - \frac{1}{2}\left(\gamma_i^{(t)} - u - E^{-1}D^\top\right)^\top E\left(\gamma_i^{(t)} - u - E^{-1}D^\top\right)
\end{aligned}
$$

Hence $q_\gamma\left(\gamma_i^{(t)}\right)$ is approximately $\text{Normal}\left(\tau_i^{(t)}, \Lambda_i^{(t)}\right)$ with variance and mean

$$
\begin{aligned}
\Lambda_i^{(t)} &:= E^{-1} \\
&= \left(\left[\sum_{h=1}^{C}\Sigma_h^{-1}\left\langle c_{i,h}^{(t)}\right\rangle\right] + (2N-2)\,H_i\right)^{-1} \\
\tau_i^{(t)} &:= u + E^{-1}D^\top \\
&= u + \Lambda_i^{(t)}\left\{\left[\sum_{j\neq i}^{N}\left\langle z_{i\to j}^{(t)}\right\rangle + \left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right] - (2N-2)\left[g_i^{(t)} + H_i^{(t)}\left(u - \hat{\gamma}_i^{(t)}\right)\right]\right\} \\
u &:= \left(\sum_{h=1}^{C}\Sigma_h^{-1}\left\langle c_{i,h}^{(t)}\right\rangle\right)^{-1}\left(\sum_{h=1}^{C}\Sigma_h^{-1}\left\langle c_{i,h}^{(t)}\right\rangle\left\langle \mu_h^{(t)}\right\rangle\right)
\end{aligned}
$$

B.3 Distribution of q_c

q_c is a discrete distribution. The distribution of $c_i^{(t)}$ conditioned on its Markov Blanket is

$$
\begin{aligned}
& p\left(c_i^{(t)} \mid \text{MB}_{c_i^{(t)}}\right) \\
\propto\; & p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}\right) p\left(c_i^{(t)}\right) \\
\propto\; & \left(\prod_{h=1}^{C}\left[|\Sigma_h|^{-1/2}\right]^{c_{i,h}^{(t)}}\right)\exp\left\{\sum_{h=1}^{C} -\frac{1}{2}c_{i,h}^{(t)}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right\}\left(\prod_{h=1}^{C}\delta_h^{c_{i,h}^{(t)}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1} \left(\gamma_i^{(t)} - \mu_h^{(t)}\right) + \sum_{h=1}^{C} c_{i,h}^{(t)} \log \frac{\delta_h}{|\Sigma_h|^{1/2}}\right\} \\
&= \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \left[\left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\gamma_i^{(t)} - \left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)} - \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\gamma_i^{(t)} + \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)}\right]\right. \\
&\quad \left. + \sum_{h=1}^{C} c_{i,h}^{(t)} \log \frac{\delta_h}{|\Sigma_h|^{1/2}}\right\} \\
&= \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \mathrm{tr}\left[\Sigma_h^{-1}\left(\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top - \mu_h^{(t)}\left(\gamma_i^{(t)}\right)^\top - \gamma_i^{(t)}\left(\mu_h^{(t)}\right)^\top + \mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right)\right]\right. \\
&\quad \left. + \sum_{h=1}^{C} c_{i,h}^{(t)} \log \frac{\delta_h}{|\Sigma_h|^{1/2}}\right\}
\end{aligned}
$$

The variables $\gamma_1^{(t)}, \ldots, \gamma_N^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}$ belong to other variational factors. The sufficient statistics of γ and μ are $\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top$, $\mu_h^{(t)}\left(\gamma_i^{(t)}\right)^\top$ $\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top$, but since γ and μ are marginally independent under q, we can take their expectations separately. Hence

$$
\begin{aligned}
q_c\left(c_i^{(t)}\right) \quad :\propto \quad & \exp\left\{\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)} \mathrm{tr}\left[\Sigma_h^{-1}\left(\left\langle\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle^\top - \left\langle\gamma_i^{(t)}\right\rangle\left\langle\mu_h^{(t)}\right\rangle^\top + \left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)\right]\right. \\
& \left. + \sum_{h=1}^{C} c_{i,h}^{(t)} \log \frac{\delta_h}{|\Sigma_h|^{1/2}}\right\}
\end{aligned}
$$

with variational parameters $\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle, \left\langle\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle, \left\langle\mu_h^{(t)}\right\rangle, \left\langle\gamma_i^{(t)}\right\rangle$. We can also express q_c in terms of indices h:

$$
\begin{aligned}
& q_c\left(c_i^{(t)} = h\right) \\
:\propto \quad & \frac{\delta_h}{|\Sigma_h|^{1/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma_h^{-1}\left(\left\langle\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle^\top - \left\langle\gamma_i^{(t)}\right\rangle\left\langle\mu_h^{(t)}\right\rangle^\top + \left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)\right]\right\}
\end{aligned}
$$

B.4 Distribution of q_μ

q_μ is a continuous distribution. The distribution of $\mu_1^{(1)}, \ldots, \mu_C^{(T)}$ conditioned on its Markov Blanket is

$$
\begin{aligned}
& p\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)} \mid \mathrm{MB}_{\mu_1^{(1)}, \ldots, \mu_C^{(T)}}\right) \\
\propto \quad & \left[\prod_{t=1}^{T}\prod_{i=1}^{N} p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}\right)\right]\left[\prod_{h=1}^{C} p\left(\mu_h^{(1)}\right)\prod_{t=2}^{T} p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}\right)\right] \\
\propto \quad & \exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right. \\
& \left. + \sum_{h=1}^{C}\left[-\frac{1}{2}\left(\mu_h^{(1)} - \nu\right)^\top \Phi^{-1}\left(\mu_h^{(1)} - \nu\right) + \sum_{t=2}^{T} -\frac{1}{2}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)^\top \Phi^{-1}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)\right]\right\} \\
\propto \quad & \exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{h=1}^{C} -\frac{1}{2} c_{i,h}^{(t)}\left[-\left(\gamma_i^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)} - \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\gamma_i^{(t)} + \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)}\right]\right. \\
& \left. + \sum_{h=1}^{C}\left[-\frac{1}{2}\left(\mu_h^{(1)} - \nu\right)^\top \Phi^{-1}\left(\mu_h^{(1)} - \nu\right) + \sum_{t=2}^{T} -\frac{1}{2}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)^\top \Phi^{-1}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)\right]\right\}
\end{aligned}
$$

The variables $\gamma_1^{(1)}, \ldots, \gamma_N^{(T)}, c_1^{(1)}, \ldots, c_N^{(T)}$ belong to other variational factors. The sufficient statistic of γ and c is $c_{i,h}^{(t)} \left(\gamma_u^{(t)}\right)^\top$, but since γ and c are marginally independent under q, we can take their expectations separately. Hence

$$
\begin{aligned}
q_\mu\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)}\right) \quad &:\propto \quad \exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{h=1}^{C} -\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle\left[-\left\langle\gamma_i^{(t)}\right\rangle^\top \Sigma_h^{-1}\mu_h^{(t)} - \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left\langle\gamma_i^{(t)}\right\rangle + \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)}\right]\right. \\
&\qquad \left. + \sum_{h=1}^{C}\left[-\frac{1}{2}\left(\mu_h^{(1)} - \nu\right)^\top \Phi^{-1}\left(\mu_h^{(1)} - \nu\right) + \sum_{t=2}^{T} -\frac{1}{2}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)^\top \Phi^{-1}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)\right]\right\} \\
&\propto \quad \prod_{h=1}^{C}\exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{N} -\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle\left[-\left\langle\gamma_i^{(t)}\right\rangle^\top \Sigma_h^{-1}\mu_h^{(t)} - \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left\langle\gamma_i^{(t)}\right\rangle + \left(\mu_h^{(t)}\right)^\top \Sigma_h^{-1}\mu_h^{(t)}\right]\right. \\
&\qquad \left. -\frac{1}{2}\left(\mu_h^{(1)} - \nu\right)^\top \Phi^{-1}\left(\mu_h^{(1)} - \nu\right) + \sum_{t=2}^{T} -\frac{1}{2}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)^\top \Phi^{-1}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)\right\}
\end{aligned}
$$

with variational parameters $\left\langle\gamma_i^{(t)}\right\rangle, \left\langle c_i^{(t)}\right\rangle$.

B.4.1 Kalman Smoother for q_μ

We can apply the Kalman Smoother to compute the mean and covariance of each $\mu_h^{(t)}$ under q_μ. Let $\Psi(a, b, C) := \exp\left\{-\frac{1}{2}(a-b)^\top C^{-1}(a-b)\right\}$, then with some manipulation we obtain

$$
\begin{aligned}
q_\mu\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)}\right) \quad &\propto \quad \prod_{h=1}^{C}\left[\Psi\left(\mu_h^{(1)}, \nu, \Phi\right)\prod_{i=1}^{N}\Psi\left(\left\langle\gamma_i^{(1)}\right\rangle, \mu_h^{(1)}, \Sigma_h\right)^{\left\langle c_{i,h}^{(1)}\right\rangle}\right] \\
&\qquad \left[\prod_{t=2}^{T}\Psi\left(\mu_h^{(t)}, \mu_h^{(t-1)}, \Phi\right)\prod_{i=1}^{N}\Psi\left(\left\langle\gamma_i^{(t)}\right\rangle, \mu_h^{(t)}, \Sigma_h\right)^{\left\langle c_{i,h}^{(t)}\right\rangle}\right] \\
&\propto \quad \prod_{h=1}^{C}\left[\Psi\left(\mu_h^{(1)}, \nu, \Phi\right)\Psi\left(\frac{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle\left\langle\gamma_i^{(1)}\right\rangle}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle}, \mu_h^{(1)}, \frac{\Sigma_h}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle}\right)\right] \\
&\qquad \left[\prod_{t=2}^{T}\Psi\left(\mu_h^{(t)}, \mu_h^{(t-1)}, \Phi\right)\Psi\left(\frac{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle}, \mu_h^{(t)}, \frac{\Sigma_h}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle}\right)\right]
\end{aligned}
$$

Notice that q_μ factorizes across cluster indices h:

$$
\begin{aligned}
q_\mu\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)}\right) \quad &= \quad \prod_{h=1}^{C} q_{\mu_h}\left(\mu_h^{(1)}, \ldots, \mu_h^{(T)}\right) \\
q_{\mu_h}\left(\mu_h^{(1)}, \ldots, \mu_h^{(T)}\right) \quad &:\propto \quad \Psi\left(\mu_h^{(1)}, \nu, \Phi\right)\Psi\left(\frac{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle\left\langle\gamma_i^{(1)}\right\rangle}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle}, \mu_h^{(1)}, \frac{\Sigma_h}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(1)}\right\rangle}\right) \\
&\qquad \left[\prod_{t=2}^{T}\Psi\left(\mu_h^{(t)}, \mu_h^{(t-1)}, \Phi\right)\Psi\left(\frac{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle}, \mu_h^{(t)}, \frac{\Sigma_h}{\sum_{i=1}^{N}\left\langle c_{i,h}^{(t)}\right\rangle}\right)\right]
\end{aligned}
$$

Observe that each factor $q_{\mu_h}\left(\mu_h^{(1)}, \ldots, \mu_h^{(T)}\right)$ is a linear system of the form

$$
\begin{aligned}
\mu_h^{(t+1)} &= \mu_h^{(t)} + w_h^{(t)} \\
\alpha_h^{(t)} &= \mu_h^{(t)} + v_h^{(t)}
\end{aligned}
$$

where $\mu_h^{(t)}$ are latent variables, and $\alpha_h^{(t)}$ are observed variables with value $\alpha_h^{(t)} = \frac{\sum_{i=1}^N \left\langle c_{i,h}^{(t)} \right\rangle \left\langle \gamma_i^{(t)} \right\rangle}{\sum_{i=1}^N \left\langle c_{i,h}^{(t)} \right\rangle}$. Furthermore, $w_h^{(t)} \sim N(0, \Phi)$, $v_h^{(t)} \sim N\left(0, \Xi_h^{(t)}\right)$ with $\Xi_h^{(t)} = \frac{\Sigma_h}{\sum_{i=1}^N \left\langle c_{i,h}^{(t)} \right\rangle}$, and $\mu_h^{(1)} \sim N(\nu, \Phi)$. Hence the distribution of each $\mu_h^{(t)}$ under q_μ is Gaussian, and its mean and covariance can be computed using the Kalman Smoother equations

$$\begin{aligned}
\hat{\mu}_h^{(t+1)|(t)} &= \hat{\mu}_h^{(t)|(t)} \\
P_h^{(t+1)|(t)} &= P_h^{(t)|(t)} + \Phi \\
K_h^{(t+1)} &= P_h^{(t+1)|(t)} \left(P_h^{(t+1)|(t)} + \Xi_h^{(t+1)} \right)^{-1} \\
\hat{\mu}_h^{(t+1)|(t+1)} &= \hat{\mu}_h^{(t+1)|(t)} + K_h^{(t+1)} \left(\alpha_h^{(t+1)} - \hat{\mu}_h^{(t+1)|(t)} \right) \\
P_h^{(t+1)|(t+1)} &= \left(\mathbb{I} - K_h^{(t+1)} \right) P_h^{(t+1)|(t)}
\end{aligned}$$

and

$$\begin{aligned}
L_h^{(t)} &= P_h^{(t)|(t)} \left(P_h^{(t+1)|(t)} \right)^{-1} \\
\hat{\mu}_h^{(t)|(T)} &= \hat{\mu}_h^{(t)|(t)} + L_h^{(t)} \left(\hat{\mu}_h^{(t+1)|(T)} - \hat{\mu}_h^{(t+1)|(t)} \right) \\
P_h^{(t)|(T)} &= P_h^{(t)|(t)} + L_h^{(t)} \left(P_h^{(t+1)|(T)} - P_h^{(t+1)|(t)} \right) \left(L_h^{(t)} \right)^\top
\end{aligned}$$

Thus μ_h has mean $\hat{\mu}_h^{(t)|(T)}$ and covariance $P_h^{(t)|(T)}$ under q_μ.

B.5 E-Step: Solutions to Variational Parameters

In the E-step, we find locally optimal variational parameters for each factor of q. The solutions to the continuous parameters are

$$\begin{aligned}
\left\langle \mu_h^{(t)} \right\rangle &= \hat{\mu}_h^{(t)|(T)} \\
\left\langle \mu_h^{(t)} \left(\mu_h^{(t)} \right)^\top \right\rangle &= \mathbb{E}_{q_\mu} \left[\mu_h^{(t)} \left(\mu_h^{(t)} \right)^\top \right] \\
&= \mathbb{V}_{q_\mu} \left[\mu_h^{(t)} \right] + \mathbb{E}_{q_\mu} \left[\mu_h^{(t)} \right] \mathbb{E}_{q_\mu} \left[\mu_h^{(t)} \right]^\top \\
&= P_h^{(t)|(T)} + \hat{\mu}_h^{(t)|(T)} \left(\hat{\mu}_h^{(t)|(T)} \right)^\top \\
\left\langle \gamma_i^{(t)} \right\rangle &= \tau_i^{(t)} \\
\left\langle \gamma_i^{(t)} \left(\gamma_i^{(t)} \right)^\top \right\rangle &= \mathbb{E}_{q_\gamma} \left[\gamma_i^{(t)} \left(\gamma_i^{(t)} \right)^\top \right] \\
&= \mathbb{V}_{q_\gamma} \left[\gamma_i^{(t)} \right] + \mathbb{E}_{q_\gamma} \left[\gamma_i^{(t)} \right] \mathbb{E}_{q_\mu} \left[\gamma_i^{(t)} \right]^\top \\
&= \Lambda_i^{(t)} + \tau_i^{(t)} \left(\tau_i^{(t)} \right)^\top
\end{aligned}$$

while the solutions to the discrete parameters are

$$\begin{aligned}
\left\langle c_{h,i}^{(t)} \right\rangle &= q\left(c_i^{(t)} = h \right) \\
\left\langle z_{(i \to j),k}^{(t)} \right\rangle &= \sum_{l=1}^K q_z \left(z_{i \to j}^{(t)} = k, z_{i \leftarrow j}^{(t)} = l \right) \\
\left\langle z_{(i \leftarrow j),l}^{(t)} \right\rangle &= \sum_{k=1}^K q_z \left(z_{i \to j}^{(t)} = k, z_{i \leftarrow j}^{(t)} = l \right)
\end{aligned}$$

These solutions are used to update the variational parameters in each factor of q. Note that they form a set of fixed-point equations that converges to a local optimum in the space of variational parameters. Hence the E-step involves iterating these equations until some convergence threshold has been reached.

C M-Step

In the M-step, we maximize $\mathcal{L}(q,\Theta)$ with respect to the model parameters $\Theta = \{B, \Sigma, \delta, \nu, \Phi\}$. Recall that

$$\mathcal{L}(q,\Theta) \quad := \quad \mathbb{E}_q\left[\log p\left(E, X \mid \Theta\right) - \log q\left(X\right)\right]$$

Note that the variational distribution q is not actually a function of the model parameters Θ; the model parameters that appear in the q's optimal solution come from the previous M-step, similar to regular EM. Hence it suffices to maximize

$$\begin{aligned}
\mathcal{L}'(q,\Theta) := \mathbb{E}_q\left[\log p\left(E, X \mid \Theta\right)\right] \quad &= \quad \mathbb{E}_q\left[\log\left(\prod_{t,i=1}^{T,N}\prod_{j\neq i}^{N} p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right) p\left(z_{i\to j}^{(t)} \mid \gamma_i^{(t)}\right) p\left(z_{i\leftarrow j}^{(t)} \mid \gamma_j^{(t)}\right)\right)\right. \\
&\quad \left(\prod_{t,i=1}^{T,N} p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}; \Sigma_1, \ldots, \Sigma_C\right) p\left(c_i^{(t)}; \delta\right)\right) \\
&\quad \left.\left(\prod_{h=1}^{C} p\left(\mu_h^{(1)}; \nu, \Phi\right) \prod_{t=2}^{T} p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}; \Phi\right)\right)\right] \\
&= \quad \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N} \log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N} \log p\left(z_{i\to j}^{(t)} \mid \gamma_i^{(t)}\right) p\left(z_{i\leftarrow j}^{(t)} \mid \gamma_j^{(t)}\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N} \log p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}; \Sigma_1, \ldots, \Sigma_C\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N} \log p\left(c_i^{(t)}; \delta\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{h=1}^{C} \log p\left(\mu_h^{(1)}; \nu, \Phi\right) + \sum_{h=1}^{C}\sum_{t=2}^{T} \log p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}; \Phi\right)\right]
\end{aligned}$$

C.1 Maximizing B

Consider the B-dependent terms in $\mathcal{L}'(q,\Theta)$,

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N} \log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right)\right] \\
= \quad &\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N} \mathbb{E}_q\left[\log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right)\right] \\
= \quad &\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{z_{i\to j}^{(t)}}\sum_{z_{i\leftarrow j}^{(t)}} q_z\left(z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}\right) \log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right) \quad (z\text{s indep. of other latent vars under } q)
\end{aligned}$$

Since $z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}$ are indicator variables, we index their possible values with $k \in \{1, \ldots, K\}$ and $l \in \{1, \ldots, K\}$ respectively:

$$= \quad \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k,l=1}^{K,K} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right) \log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l; B\right)$$

$$= \sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} \sum_{k,l=1}^{K,K} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)\left(E_{i,j}^{(t)} \log B_{k,l} + \left(1 - E_{i,j}^{(t)}\right)\log\left(1 - B_{k,l}\right)\right) \quad (2)$$

Setting the first derivative wrt $B_{k,l}$ to zero yields the maximizer $\hat{B}_{k,l}$ for $\mathcal{L}'(q,\Theta)$:

$$\begin{aligned}
0 &= \frac{\partial}{\partial B_{k,l}} \sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} \sum_{k',l'=1}^{K,K} q_z\left(z_{i\to j}^{(t)} = k', z_{i\leftarrow j}^{(t)} = l'\right)\left(E_{i,j}^{(t)} \log B_{k',l'} + \left(1 - E_{i,j}^{(t)}\right)\log\left(1 - B_{k',l'}\right)\right) \\
0 &= \sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)\left(\frac{E_{i,j}^{(t)}}{B_{k,l}} - \frac{1 - E_{i,j}^{(t)}}{1 - B_{k,l}}\right) \\
0 &= \sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)\left(E_{i,j}^{(t)} - B_{k,l}\right) \\
\hat{B}_{k,l} := B_{k,l} &= \frac{\sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right) E_{i,j}^{(t)}}{\sum_{t,i=1}^{T,N} \sum_{j\neq i}^{N} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)}
\end{aligned}$$

C.2 Maximizing Σ

Consider the $\Sigma_1, \ldots, \Sigma_C$-dependent terms in $\mathcal{L}'(q,\Theta)$,

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{t,i=1}^{T,N} \log p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}; \Sigma_1, \ldots, \Sigma_C\right)\right] \\
= &\sum_{t,i=1}^{T,N} \mathbb{E}_q\left[\log p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}; \Sigma_1, \ldots, \Sigma_C\right)\right] \\
= &\sum_{t,i=1}^{T,N} \mathbb{E}_q\left[\log \prod_{h=1}^{C}\left((2\pi)^{-K/2}\left|\Sigma_h\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right\}\right)^{c_{i,h}^{(t)}}\right] \\
= &\sum_{t,i=1}^{T,N} \sum_{h=1}^{C} \mathbb{E}_q\left[c_{i,h}^{(t)} \log\left((2\pi)^{-K/2}\left|\Sigma_h\right|^{-1/2}\right) - \frac{1}{2} c_{i,h}^{(t)}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)^\top \Sigma_h^{-1}\left(\gamma_i^{(t)} - \mu_h^{(t)}\right)\right] \\
= &\sum_{t,i=1}^{T,N} \sum_{h=1}^{C} -\log\left((2\pi)^{K/2}\left|\Sigma_h\right|^{1/2}\right)\mathbb{E}_q\left[c_{i,h}^{(t)}\right] \\
&-\frac{1}{2}\mathbb{E}_q\left[c_{i,h}^{(t)} \mathrm{tr}\left[\Sigma_h^{-1}\left(\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top - \mu_h^{(t)}\left(\gamma_i^{(t)}\right)^\top - \gamma_i^{(t)}\left(\mu_h^{(t)}\right)^\top + \mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right)\right]\right]
\end{aligned}$$

Since c, μ, γ are independent of each other (and other latent variables) under q,

$$\begin{aligned}
= &\sum_{t,i=1}^{T,N} \sum_{h=1}^{C} -\log\left((2\pi)^{K/2}\left|\Sigma_h\right|^{1/2}\right)\left\langle c_{i,h}^{(t)}\right\rangle \quad (3)\\
&-\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle \mathrm{tr}\left[\Sigma_h^{-1}\left(\left\langle\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle^\top - \left\langle\gamma_i^{(t)}\right\rangle\left\langle\mu_h^{(t)}\right\rangle^\top + \left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)\right]
\end{aligned}$$

where we have defined $\langle X\rangle := \mathbb{E}_q[X]$, and the solutions to $\langle X\rangle$ are identical to Section B.5. Setting the first derivative wrt Σ_h to zero yields the maximizer $\hat{\Sigma}_h$ for $\mathcal{L}'(q,\Theta)$:

$$\begin{aligned}
0 = &\nabla_{\Sigma_h} \sum_{t,i=1}^{T,N} \sum_{h=1}^{C} -\log\left((2\pi)^{K/2}\left|\Sigma_h\right|^{1/2}\right)\left\langle c_{i,h}^{(t)}\right\rangle \\
&-\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle \mathrm{tr}\left[\Sigma_h^{-1}\left(\left\langle\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\right\rangle\left\langle\gamma_i^{(t)}\right\rangle^\top - \left\langle\gamma_i^{(t)}\right\rangle\left\langle\mu_h^{(t)}\right\rangle^\top + \left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)\right]
\end{aligned}$$

$$
\begin{aligned}
0 &= \sum_{t,i=1}^{T,N} -\frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle \Sigma_h^{-1} + \frac{1}{2}\left\langle c_{i,h}^{(t)}\right\rangle \Sigma_h^{-1}\left(\left\langle \gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle \gamma_i^{(t)}\right\rangle\left\langle \mu_h^{(t)}\right\rangle^\top - \left\langle \mu_h^{(t)}\right\rangle\left\langle \gamma_i^{(t)}\right\rangle^\top + \left\langle \mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)\Sigma_h^{-1} \\
0 &= \sum_{t,i=1}^{T,N} -\left\langle c_{i,h}^{(t)}\right\rangle \Sigma_h + \left\langle c_{i,h}^{(t)}\right\rangle\left(\left\langle \gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle \gamma_i^{(t)}\right\rangle\left\langle \mu_h^{(t)}\right\rangle^\top - \left\langle \mu_h^{(t)}\right\rangle\left\langle \gamma_i^{(t)}\right\rangle^\top + \left\langle \mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right) \\
\hat{\Sigma}_h &:= \Sigma_h = \frac{\sum_{t,i=1}^{T,N}\left\langle c_{i,h}^{(t)}\right\rangle\left(\left\langle \gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right\rangle - \left\langle \gamma_i^{(t)}\right\rangle\left\langle \mu_h^{(t)}\right\rangle^\top - \left\langle \mu_h^{(t)}\right\rangle\left\langle \gamma_i^{(t)}\right\rangle^\top + \left\langle \mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle\right)}{\sum_{t,i=1}^{T,N}\left\langle c_{i,h}^{(t)}\right\rangle}
\end{aligned}
$$

C.3 Maximizing δ

Consider the δ-dependent terms in $\mathcal{L}'(q,\Theta)$,

$$
\begin{aligned}
&\mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\log p\left(c_i^{(t)};\delta\right)\right] \\
&= \sum_{t,i=1}^{T,N}\mathbb{E}_q\left[\log\prod_{h=1}^{C}\delta_h^{c_{i,h}^{(t)}}\right] \\
&= \sum_{t,i=1}^{T,N}\sum_{h=1}^{C}\mathbb{E}_q\left[c_{i,h}^{(t)}\log\delta_h\right] \\
&= \sum_{t,i=1}^{T,N}\sum_{h=1}^{C}\left\langle c_{i,h}^{(t)}\right\rangle\log\delta_h \\
&= \left(\sum_{t,i=1}^{T,N}\left\langle c_i^{(t)}\right\rangle\right)^\top\log\delta
\end{aligned}
\tag{4}
$$

where $\left\langle c_{i,h}^{(t)}\right\rangle := \mathbb{E}_q\left[c_{i,h}^{(t)}\right]$, and the solution to $\left\langle c_{i,h}^{(t)}\right\rangle$ is identical to Section B.5. Taking the first derivative with respect to $\delta_1,\ldots,\delta_{C-1}$,

$$
\begin{aligned}
\frac{\partial}{\partial\delta_h}\sum_{t,i=1}^{T,N}\sum_{h'=1}^{C}\left\langle c_{i,h'}^{(t)}\right\rangle\log\delta_{h'} &= \sum_{t,i=1}^{T,N}\frac{\partial}{\partial\delta_h}\left\langle c_{i,h}^{(t)}\right\rangle\log\delta_h + \frac{\partial}{\partial\delta_h}\left\langle c_{i,C}^{(t)}\right\rangle\log\left(1-\sum_{h'=1}^{C-1}\delta_{h'}\right) \\
&= \sum_{t,i=1}^{T,N}\frac{\left\langle c_{i,h}^{(t)}\right\rangle}{\delta_h} - \frac{\left\langle c_{i,C}^{(t)}\right\rangle}{1-\sum_{h'=1}^{C-1}\delta_{h'}}
\end{aligned}
$$

By setting all the derivatives to zero and performing some manipulation, we obtain the maximizer $\hat{\delta}$ for $\mathcal{L}'(q,\Theta)$:

$$
\hat{\delta} = \frac{\sum_{t,i=1}^{T,N}\left\langle c_i^{(t)}\right\rangle}{TN}
$$

C.4 Maximizing ν, Φ

Consider the ν, Φ-dependent terms in $\mathcal{L}'(q,\Theta)$,

$$
\begin{aligned}
&\mathbb{E}_q\left[\sum_{h=1}^{C}\log p\left(\mu_h^{(1)};\nu,\Phi\right) + \sum_{h=1}^{C}\sum_{t=2}^{T}\log p\left(\mu_h^{(t)}\mid\mu_h^{(t-1)};\Phi\right)\right] \\
&= \sum_{h=1}^{C}\mathbb{E}_q\left[\log p\left(\mu_h^{(1)};\nu,\Phi\right)\right] + \sum_{h=1}^{C}\sum_{t=2}^{T}\mathbb{E}_q\left[\log p\left(\mu_h^{(t)}\mid\mu_h^{(t-1)};\Phi\right)\right]
\end{aligned}
$$

We begin by maximizing wrt ν, which only requires us to focus on the first term:

$$\begin{aligned}
& \sum_{h=1}^{C} \mathbb{E}_q \left[\log p\left(\mu_h^{(1)}; \nu, \Phi\right)\right] \\
= & \sum_{h=1}^{C} \mathbb{E}_q \left[\log \left((2\pi)^{-K/2} |\Phi|^{-1/2} \exp\left\{ -\frac{1}{2}\left(\mu_h^{(1)} - \nu\right)^\top \Phi^{-1} \left(\mu_h^{(1)} - \nu\right)\right\}\right)\right] \\
= & \sum_{h=1}^{C} \mathbb{E}_q \left[\log\left((2\pi)^{-K/2} |\Phi|^{-1/2}\right) - \frac{1}{2}\left(\left(\mu_h^{(1)}\right)^\top \Phi^{-1}\mu_h^{(1)} - \left(\mu_h^{(1)}\right)^\top \Phi^{-1}\nu - \nu^\top\Phi^{-1}\mu_h^{(1)} + \nu^\top \Phi^{-1}\nu\right)\right]
\end{aligned}$$

Dropping terms that do not depend on ν,

$$\begin{aligned}
= & \sum_{h=1}^{C} \mathbb{E}_q \left[-\frac{1}{2}\left(- \left(\mu_h^{(1)}\right)^\top \Phi^{-1}\nu - \nu^\top\Phi^{-1}\mu_h^{(1)} + \nu^\top \Phi^{-1}\nu\right)\right] \\
= & \sum_{h=1}^{C} \frac{1}{2}\left\langle \mu_h^{(1)}\right\rangle^\top \Phi^{-1}\nu + \frac{1}{2}\nu^\top\Phi^{-1}\left\langle\mu_h^{(1)}\right\rangle - \frac{1}{2}\nu^\top\Phi^{-1}\nu
\end{aligned}$$

where $\left\langle \mu_h^{(1)}\right\rangle := \mathbb{E}_q\left[\mu_h^{(1)}\right]$, and the solution to $\left\langle \mu_h^{(1)}\right\rangle$ is identical to Section B.5. Setting the first derivative wrt ν to zero yields the maximizer $\hat{\nu}$ for $\mathcal{L}'\left(q, \Theta\right)$:

$$\begin{aligned}
0 &= \nabla_\nu \sum_{h=1}^{C} \frac{1}{2}\left\langle \mu_h^{(1)}\right\rangle^\top \Phi^{-1}\nu + \frac{1}{2}\nu^\top\Phi^{-1}\left\langle\mu_h^{(1)}\right\rangle - \frac{1}{2}\nu^\top\Phi^{-1}\nu \\
0 &= \sum_{h=1}^{C} \Phi^{-1}\left\langle\mu_h^{(1)}\right\rangle - \Phi^{-1}\nu \\
\hat{\nu} := \nu &= \frac{\sum_{h=1}^{C}\left\langle\mu_h^{(1)}\right\rangle}{C}
\end{aligned}$$

We now substitute $\nu = \hat{\nu}$ and consider the Φ-dependent terms in $\mathcal{L}'\left(q, \Theta\right)$:

$$\begin{aligned}
& \mathbb{E}_q\left[\sum_{h=1}^{C} \log p\left(\mu_h^{(1)}; \hat{\nu}, \Phi\right) + \sum_{h=1}^{C}\sum_{t=2}^{T} \log p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}; \Phi\right)\right] \\
= & \sum_{h=1}^{C} \mathbb{E}_q\left[\log p\left(\mu_h^{(1)}; \hat{\nu}, \Phi\right)\right] + \sum_{h=1}^{C}\sum_{t=2}^{T} \mathbb{E}_q\left[\log p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}; \Phi\right)\right] \\
= & \sum_{h=1}^{C} -\log\left((2\pi)^{K/2}|\Phi|^{1/2}\right) - \frac{1}{2}\mathbb{E}_q\left[\left(\mu_h^{(1)} - \hat{\nu}\right)^\top \Phi^{-1}\left(\mu_h^{(1)} - \hat{\nu}\right)\right] \\
& + \sum_{h=1}^{C}\sum_{t=2}^{T} -\log\left((2\pi)^{K/2}|\Phi|^{1/2}\right) - \frac{1}{2}\mathbb{E}_q\left[\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)^\top \Phi^{-1}\left(\mu_h^{(t)} - \mu_h^{(t-1)}\right)\right] \\
= & -TC\log\left((2\pi)^{K/2}|\Phi|^{1/2}\right) - \sum_{h=1}^{C}\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(\left\langle\mu_h^{(1)}\left(\mu_h^{(1)}\right)^\top\right\rangle - \hat{\nu}\left\langle\mu_h^{(1)}\right\rangle^\top - \left\langle\mu_h^{(1)}\right\rangle\hat{\nu}^\top + \hat{\nu}\hat{\nu}^\top\right)\right] \qquad (5) \\
& - \sum_{h=1}^{C}\sum_{t=2}^{T}\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle + \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle\right)\right]
\end{aligned}$$

where $\langle X\rangle := \mathbb{E}_q\left[X\right]$. The solutions to $\left\langle\mu_h^{(1)}\right\rangle, \left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle, \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle$ are identical to Section B.5. The remaining expectations are

$$\left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle = \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle^\top = P_h^{(t)|(T)}\left(L_h^{(t-1)}\right)^\top + \left\langle\mu_h^{(t)}\right\rangle\left\langle\mu_h^{(t-1)}\right\rangle^\top$$

where P and L are defined in Section B.4.1. Setting the first derivative wrt Φ to zero yields the maximizer $\hat{\Phi}$ for $\mathcal{L}'(q,\Theta)$:

$$
\begin{aligned}
0 &= \nabla_\Phi - TC\log\left((2\pi)^{K/2}|\Phi|^{1/2}\right) - \sum_{h=1}^{C}\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(\left\langle\mu_h^{(1)}\left(\mu_h^{(1)}\right)^\top\right\rangle - \hat{\nu}\left\langle\mu_h^{(1)}\right\rangle^\top - \left\langle\mu_h^{(1)}\right\rangle\hat{\nu}^\top + \hat{\nu}\hat{\nu}^\top\right)\right] \\
&\quad - \sum_{h=1}^{C}\sum_{t=2}^{T}\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle + \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle\right)\right] \\
0 &= -\frac{TC}{2}\Phi^{-1} + \sum_{h=1}^{C}\frac{1}{2}\Phi^{-1}\left(\left\langle\mu_h^{(1)}\left(\mu_h^{(1)}\right)^\top - \left\langle\mu_h^{(1)}\right\rangle\hat{\nu}^\top - \hat{\nu}\left\langle\mu_h^{(1)}\right\rangle^\top + \hat{\nu}\hat{\nu}^\top\right\rangle\right)\Phi^{-1} \\
&\quad + \sum_{h=1}^{C}\sum_{t=2}^{T}\frac{1}{2}\Phi^{-1}\left(\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle - \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle + \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle\right)\Phi^{-1} \\
0 &= -TC\Phi + \sum_{h=1}^{C}\left\langle\mu_h^{(1)}\left(\mu_h^{(1)}\right)^\top\right\rangle - \left\langle\mu_h^{(1)}\right\rangle\hat{\nu}^\top - \hat{\nu}\left\langle\mu_h^{(1)}\right\rangle^\top + \hat{\nu}\hat{\nu}^\top \\
&\quad + \sum_{h=1}^{C}\sum_{t=2}^{T}\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle - \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle + \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle \\
\hat{\Phi} := \Phi &= \frac{\sum_{h=1}^{C}\left\langle\mu_h^{(1)}\left(\mu_h^{(1)}\right)^\top\right\rangle - \left\langle\mu_h^{(1)}\right\rangle\hat{\nu}^\top - \hat{\nu}\left\langle\mu_h^{(1)}\right\rangle^\top + \hat{\nu}\hat{\nu}^\top}{TC} \\
&\quad + \frac{\sum_{h=1}^{C}\sum_{t=2}^{T}\left\langle\mu_h^{(t)}\left(\mu_h^{(t)}\right)^\top\right\rangle - \left\langle\mu_h^{(t)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle - \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t)}\right)^\top\right\rangle + \left\langle\mu_h^{(t-1)}\left(\mu_h^{(t-1)}\right)^\top\right\rangle}{TC}
\end{aligned}
$$

D Computing the Variational Lower Bound $\mathcal{L}(q,\Theta)$

The marginal likelihood lower bound $\mathcal{L}(q,\Theta)$ can be used to test for convergence in the Variational EM algorithm. It also functions as a surrogate for the true marginal likelihood $p(E \mid \Theta)$; this is useful when taking random restarts, as it enables us to select the highest likelihood restart. Recall that

$$
\begin{aligned}
\mathcal{L}(q,\Theta) &= \mathbb{E}_q\left[\log p(E,X\mid\Theta) - \log q(X)\right] \\
&= \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\log p\left(E_{i,j}^{(t)} \mid z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}; B\right)\right] + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\log p\left(z_{i\to j}^{(t)} \mid \gamma_i^{(t)}\right)p\left(z_{i\leftarrow j}^{(t)} \mid \gamma_j^{(t)}\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\log p\left(\gamma_i^{(t)} \mid c_i^{(t)}, \mu_1^{(t)}, \ldots, \mu_C^{(t)}; \Sigma_1, \ldots, \Sigma_C\right)\right] + \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\log p\left(c_i^{(t)}; \delta\right)\right] \\
&\quad + \mathbb{E}_q\left[\sum_{h=1}^{C}\log p\left(\mu_h^{(1)}; \nu, \Phi\right) + \sum_{h=1}^{C}\sum_{t=2}^{T}\log p\left(\mu_h^{(t)} \mid \mu_h^{(t-1)}; \Phi\right)\right] \\
&\quad - \mathbb{E}_q\left[\log q_\mu\left(\mu_1^{(1)}, \ldots, \mu_C^{(T)}\right)\right] - \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\log q_\gamma\left(\gamma_i^{(t)}\right)\right] \\
&\quad - \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\log q_c\left(c_i^{(t)}\right)\right] - \mathbb{E}_q\left[\sum_{t,i,j=1}^{T,N,N}\log q_z\left(z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}\right)\right]
\end{aligned}
$$

It turns out that we cannot compute $\mathcal{L}(q,\Theta)$ exactly because of term 2, but we can lower bound the latter to produce a lower bound $\mathcal{L}_{lower}(q,\Theta)$ on $\mathcal{L}(q,\Theta)$.

Closed forms for terms 1,3,4,5 are in eqs (2,3,4,5) respectively. We shall now provide closed forms for terms 6,7,8,9, as well as the aforementioned lower bound for term 2.

D.1 Lower Bound for Term 2

$$
\begin{aligned}
& \mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\log p\left(z_{i\rightarrow j}^{(t)}\mid\gamma_i^{(t)}\right)p\left(z_{i\leftarrow j}^{(t)}\mid\gamma_j^{(t)}\right)\right] \\
= & \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\mathbb{E}_q\left[\log\prod_{k=1}^{K}\left(\frac{\exp\gamma_{i,k}^{(t)}}{\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}}\right)^{z_{i\rightarrow j,k}^{(t)}}\left(\frac{\exp\gamma_{j,k}^{(t)}}{\sum_{l=1}^{K}\exp\gamma_{j,l}^{(t)}}\right)^{z_{i\leftarrow j,k}^{(t)}}\right] \\
= & \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k=1}^{K}\mathbb{E}_q\left[z_{i\rightarrow j,k}^{(t)}\gamma_{i,k}^{(t)}-z_{i\rightarrow j,k}^{(t)}\log\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}+z_{i\leftarrow j,k}^{(t)}\gamma_{j,k}^{(t)}-z_{i\leftarrow j,k}^{(t)}\log\sum_{l=1}^{K}\exp\gamma_{j,l}^{(t)}\right]
\end{aligned}
$$

Since z,γ are independent of each other under q,

$$
= \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k=1}^{K}\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{i,k}^{(t)}\right\rangle-\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\mathbb{E}_q\left[\log\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}\right]+\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{j,k}^{(t)}\right\rangle-\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\mathbb{E}_q\left[\log\sum_{l=1}^{K}\exp\gamma_{j,l}^{(t)}\right]
$$

Applying Jensen's inequality to the log-sum-exp terms,

$$
\begin{aligned}
\geq & \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k=1}^{K}\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{i,k}^{(t)}\right\rangle-\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\log\mathbb{E}_q\left[\sum_{l=1}^{K}\exp\gamma_{i,l}^{(t)}\right]+\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{j,k}^{(t)}\right\rangle-\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\log\mathbb{E}_q\left[\sum_{l=1}^{K}\exp\gamma_{j,l}^{(t)}\right] \\
= & \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k=1}^{K}\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{i,k}^{(t)}\right\rangle-\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\log\left(\sum_{l=1}^{K}\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle\right)+\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\left\langle\gamma_{j,k}^{(t)}\right\rangle-\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\log\left(\sum_{l=1}^{K}\left\langle\exp\gamma_{j,l}^{(t)}\right\rangle\right) \\
= & \sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k=1}^{K}\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle\left(\left\langle\gamma_{i,k}^{(t)}\right\rangle-\log\sum_{l=1}^{K}\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle\right)+\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle\left(\left\langle\gamma_{j,k}^{(t)}\right\rangle-\log\sum_{l=1}^{K}\left\langle\exp\gamma_{j,l}^{(t)}\right\rangle\right) \\
= & \sum_{t,i=1}^{T,N}\sum_{k=1}^{K}\left(\left\langle\gamma_{i,k}^{(t)}\right\rangle-\log\sum_{l=1}^{K}\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle\right)\left(\sum_{j\neq i}^{N}\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle+\left\langle z_{j\leftarrow i,k}^{(t)}\right\rangle\right) \\
= & \sum_{t,i=1}^{T,N}\left(\left\langle\gamma_{i}^{(t)}\right\rangle-\log\sum_{l=1}^{K}\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle\right)^{\top}\left(\sum_{j\neq i}^{N}\left\langle z_{i\rightarrow j}^{(t)}\right\rangle+\left\langle z_{j\leftarrow i}^{(t)}\right\rangle\right)
\end{aligned}
$$

where $\langle X\rangle := \mathbb{E}_q[X]$. The solutions to $\left\langle z_{i\rightarrow j,k}^{(t)}\right\rangle,\left\langle z_{i\leftarrow j,k}^{(t)}\right\rangle,\left\langle\gamma_{i,k}^{(t)}\right\rangle$ are in Section B.5. As for $\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle$, observe that for a univariate Gaussian random variable X with mean μ and variance σ^2,

$$
\begin{aligned}
\mathbb{E}[\exp X] &= \int_x\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\exp\{x\}\,dx \\
&= \int_x\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2-2x\left(\mu+\sigma^2\right)+\mu^2}{2\sigma^2}\right\}dx \\
&= \int_x\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2-2x\left(\mu+\sigma^2\right)+\left(\mu^2+2\mu\sigma^2+\sigma^4\right)}{2\sigma^2}\right\}\exp\left\{\mu+\frac{\sigma^2}{2}\right\}dx \\
&= \exp\left\{\mu+\frac{\sigma^2}{2}\right\}\int_x\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left(x-\left(\mu+\sigma^2\right)\right)^2}{2\sigma^2}\right\}dx \\
&= \exp\left\{\mu+\frac{\sigma^2}{2}\right\}
\end{aligned}
$$

Hence

$$
\left\langle\exp\gamma_{i,l}^{(t)}\right\rangle = \exp\left\{\left\langle\gamma_{i,l}^{(t)}\right\rangle+\frac{1}{2}\Lambda_{i,ll}^{(t)}\right\}
$$

where Λ_i is defined in Section B.2.1.

D.2 Term 6

Define

$$\mathcal{N}(a,b,C): \quad = \quad (2\pi)^{-\dim(C)/2}|C|^{-1/2}\exp\left\{-\frac{1}{2}(a-b)^\top C^{-1}(a-b)\right\}$$

Thus

$$\begin{aligned}
& -\mathbb{E}_q\left[\log q_\mu\left(\mu_1^{(1)},\ldots,\mu_C^{(T)}\right)\right] \\
= & -\mathbb{E}_q\left[\log\prod_{h=1}^{C} q_{\mu_h}\left(\mu_h^{(1)},\ldots,\mu_h^{(T)}\right)\right] \\
= & -\sum_{h=1}^{C}\mathbb{E}_q\left[\log\mathcal{N}\left(\mu_h^{(1)},\nu,\Phi\right)\mathcal{N}\left(\alpha_h^{(1)},\mu_h^{(1)},\Xi_h^{(1)}\right)\right. \\
& \left.\prod_{t=2}^{T}\mathcal{N}\left(\mu_h^{(t)},\mu_h^{(t-1)},\Phi\right)\mathcal{N}\left(\alpha_h^{(t)},\mu_h^{(t)},\Xi_h^{(t)}\right)\right] \\
= & -\sum_{h=1}^{C}\mathbb{E}_q\left[\log\mathcal{N}\left(\mu_h^{(1)},\nu,\Phi\right)+\log\mathcal{N}\left(\alpha_h^{(1)},\mu_h^{(1)},\Xi_h^{(1)}\right)\right. \\
& \left.+\sum_{t=2}^{T}\log\mathcal{N}\left(\mu_h^{(t)},\mu_h^{(t-1)},\Phi\right)+\log\mathcal{N}\left(\alpha_h^{(t)},\mu_h^{(t)},\Xi_h^{(t)}\right)\right]
\end{aligned}$$

where $\alpha_h^{(t)},\Xi_h^{(t)}$ are from Section B.4.1. Also note our abuse of notation: ν,Φ refer to the values used to compute $\mu_h^{(t)|(T)},P_h^{(t)|(T)},L_h^{(t)}$ in the E-step (Section B.4.1), and not their current values (recall that q_μ is *not* a function of ν,Φ). Now define

$$\begin{aligned}
\mathcal{Z}_N(C): \quad &= \quad \log(2\pi)^{-\dim(C)/2}|C|^{-1/2} \\
\Psi(a,b,C): \quad &= \quad \exp\left\{-\frac{1}{2}(a-b)^\top C^{-1}(a-b)\right\}
\end{aligned}$$

so we have

$$\begin{aligned}
= & -\left[CT\mathcal{Z}_N(\Phi)+\sum_{h=1}^{C}\sum_{t=1}^{T}+\mathcal{Z}_N\left(\Xi_h^{(t)}\right)\right] \\
& -\sum_{h=1}^{C}\left[\mathbb{E}_q\left[\log\Psi\left(\mu_h^{(1)},\nu,\Phi\right)\right]+\mathbb{E}_q\left[\log\Psi\left(\alpha_h^{(1)},\mu_h^{(1)},\Xi_h^{(1)}\right)\right]\right. \\
& \left.+\sum_{t=2}^{T}\mathbb{E}_q\left[\log\Psi\left(\mu_h^{(t)},\mu_h^{(t-1)},\Phi\right)\right]+\mathbb{E}_q\left[\log\Psi\left(\alpha_h^{(t)},\mu_h^{(t)},\Xi_h^{(t)}\right)\right]\right]
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_q\left[\log\Psi\left(\mu_h^{(1)},\nu,\Phi\right)\right] &= -\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(m_h^{(1)}-\nu\left(\hat{\mu}_h^{(1)|(T)}\right)^\top-\left(\hat{\mu}_h^{(1)|(T)}\right)\nu^\top+\nu\nu^\top\right)\right] \\
\mathbb{E}_q\left[\log\Psi\left(\mu_h^{(t)},\mu_h^{(t-1)},\Phi\right)\right] &= -\frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\left(m_h^{(t)}-\left(V_h^{(t,t-1)}\right)^\top-V_h^{(t,t-1)}+m_h^{(t-1)}\right)\right] \quad \forall t\in\{2,\ldots T\} \\
\mathbb{E}_q\left[\log\Psi\left(\alpha_h^{(t)},\mu_h^{(t)},\Xi_h^{(t)}\right)\right] &= -\frac{1}{2}\mathrm{tr}\left[\left(\Xi_h^{(t)}\right)^{-1}\left(\alpha_h^{(t)}\left(\alpha_h^{(t)}\right)^\top-\hat{\mu}_h^{(t)|(T)}\left(\alpha_h^{(t)}\right)^\top-\alpha_h^{(t)}\left(\hat{\mu}_h^{(t)|(T)}\right)^\top+m_h^{(t)}\right)\right] \\
& \quad \forall t\in\{1,\ldots,T\} \\
m_h^{(t)} &:= P_h^{(t)|(T)}+\hat{\mu}_h^{(t)|(T)}\left(\hat{\mu}_h^{(t)|(T)}\right)^\top, \quad V_h^{(t,t-1)}:=P_h^{(t)|(T)}\left(L_h^{(t-1)}\right)^\top+\hat{\mu}_h^{(t)|(T)}\left(\hat{\mu}_h^{(t-1)|(T)}\right)^\top
\end{aligned}$$

and where $\mu_h^{(t)|(T)},P_h^{(t)|(T)},L_h^{(t)}$ are from Section B.4.1.

D.3 Term 7

Using definitions from the previous section,

$$
\begin{aligned}
& -\mathbb{E}_q\left[\sum_{t,i=1}^{T,N} \log q_\gamma\left(\gamma_i^{(t)}\right)\right] \\
= & -\sum_{t,i=1}^{T,N} \mathbb{E}_q\left[\log \mathcal{N}\left(\gamma_i^{(t)}, \tau_i^{(t)}, \Lambda_i^{(t)}\right)\right] \\
= & -\sum_{t,i=1}^{T,N} \mathcal{Z}_N\left(\Lambda_i^{(t)}\right) - \sum_{t,i=1}^{T,N} \mathbb{E}_q\left[\log \Psi\left(\gamma_i^{(t)}, \tau_i^{(t)}, \Lambda_i^{(t)}\right)\right] \\
= & -\sum_{t,i=1}^{T,N} \mathcal{Z}_N\left(\Lambda_i^{(t)}\right) + \sum_{t,i=1}^{T,N} \frac{1}{2}\mathrm{tr}\left[\left(\Lambda_i^{(t)}\right)^{-1}\left(\mathbb{E}_q\left[\gamma_i^{(t)}\left(\gamma_i^{(t)}\right)^\top\right] - \tau_i^{(t)}\mathbb{E}_q\left[\gamma_i^{(t)}\right]^\top - \mathbb{E}_q\left[\gamma_i^{(t)}\right]\left(\tau_i^{(t)}\right)^\top + \tau_i^{(t)}\left(\tau_i^{(t)}\right)^\top\right)\right] \\
= & -\sum_{t,i=1}^{T,N} \mathcal{Z}_N\left(\Lambda_i^{(t)}\right) + \frac{TNK}{2}
\end{aligned}
$$

where $\Lambda_i^{(t)}$ is from Section B.2.1.

D.4 Term 8

Term 8 is trivial to compute since q_c is discrete:

$$
-\mathbb{E}_q\left[\sum_{t,i=1}^{T,N} \log q_c\left(c_i^{(t)}\right)\right] \quad = \quad -\sum_{t,i=1}^{T,N}\sum_{h=1}^{C} q_c\left(c_i^{(t)} = h\right)\log q_c\left(c_i^{(t)} = h\right)
$$

D.5 Term 9

Term 9 is also trivial to compute since q_z is discrete:

$$
-\mathbb{E}_q\left[\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N} \log q_z\left(z_{i\to j}^{(t)}, z_{i\leftarrow j}^{(t)}\right)\right] \quad = \quad -\sum_{t,i=1}^{T,N}\sum_{j\neq i}^{N}\sum_{k,l=1}^{K,K} q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)\log q_z\left(z_{i\to j}^{(t)} = k, z_{i\leftarrow j}^{(t)} = l\right)
$$