Appendix

A.1 Deviation of the Lower Bound

Given the assumption of the Markov property and fixed state distributions we can first split the upper bound into two integrals given by

$$\begin{aligned}
&\iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p(\boldsymbol{\tau}, \boldsymbol{\theta})}{p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta})} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&= \iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p\left(\boldsymbol{\theta}\right) \prod_{t=1}^{T} \pi(a_t|\boldsymbol{\theta}, s_t)}{q(\boldsymbol{\theta})} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&+ \iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p(s_1) \prod_{t=1}^{T} p(s_{t+1}|a_t, s_i)}{p(s_1) \prod_{t=1}^{T} p(s_{t+1}|a_t, s_t) \pi'(a_t|s_t)} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau},
\end{aligned} \tag{1}$$

where $\pi'(a_t|s_t)$ is the old policy. Simplifications and the fixed nature of $\pi'(a_t|s_t)$ leads then directly to the proposed simple form given by

$$\begin{aligned}
&= \iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p\left(\boldsymbol{\theta}\right) \prod_{t=1}^{T} \pi(a_t|\boldsymbol{\theta}, s_t)}{q(\boldsymbol{\theta})} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&- \iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\prod_{t=1}^{T} \pi'(a_t|s_t) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&= \iint p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p\left(\boldsymbol{\theta}\right) \prod_{t=1}^{T} \pi(a_t|\boldsymbol{\theta}, s_t)}{q(\boldsymbol{\theta})} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&+ const.
\end{aligned} \tag{2}$$

A.2 Applying the Idea of Variational Bayes

Given the simplified form from A.1 we can apply the techniques of Variational Bayes for the estimation of the parameters $\boldsymbol{\theta}$. In this section we will derive the approximated distributions step by step from the original integral. For simplicity the parameters are labelled with $\boldsymbol{\theta}_i$ and we assume that the factorization $q(\boldsymbol{\theta}) = \prod_i q(\boldsymbol{\theta}_i)$ holds. The derivations are then given by

$$
\begin{aligned}
&\int\int p_{\text{old}}(\boldsymbol{\tau}) q(\boldsymbol{\theta}) \log \left(\frac{p(\boldsymbol{\theta}) \prod_{t=1}^{T} \pi(a_t|\boldsymbol{\theta}, s_t)}{q(\boldsymbol{\theta})} \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\tau} \\
&= \int \prod_i q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \left(\log \left(p(\boldsymbol{\theta}) \prod_{t=1}^{T} \pi(a_t|\boldsymbol{\theta}, s_t) \right) - \sum_k \log q_k(\theta_k) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta} \\
&= \int_{\theta_j} \int_{\boldsymbol{\theta}_{-j}} q_j(\theta_j) \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \left(\log \left(p(\boldsymbol{\theta}) \prod_{t=1}^{T} p(a_t|\boldsymbol{\theta}, s_t) \right) - \sum_k \log q_k(\theta_k) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j} \mathrm{d}\boldsymbol{\theta}_j \\
&= \int_{\theta_j} \int_{\boldsymbol{\theta}_{-j}} q_j(\theta_j) \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \log \left(p(\boldsymbol{\theta}) \prod_{t=1}^{T} p(a_t|\boldsymbol{\theta}, s_t) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j} \mathrm{d}\boldsymbol{\theta}_j \\
&\quad - \int_{\theta_j} \int_{\boldsymbol{\theta}_{-j}} q_j(\theta_j) \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \left(\sum_{k \neq j} \log q_k(\theta_k) + \log q_j(\theta_j) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j} \mathrm{d}\boldsymbol{\theta}_j \\
&= \int_{\theta_j} q_j(\theta_j) \int_{\boldsymbol{\theta}_{-j}} \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \log \left(p(\boldsymbol{\theta}) \prod_{t=1}^{T} p(a_t|\boldsymbol{\theta}, s_t) \right) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j} \mathrm{d}\boldsymbol{\theta}_j \\
&\quad - \int_{\theta_j} q_j(\theta_j) \int p_{\text{old}}(\boldsymbol{\tau}) \log q_j(\theta_j) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_j + const.
\end{aligned}
\tag{3}
$$

Omitting the constant term we find that the maximization of this formula is given when the inner parts of the two separated integrals over θ_j are equal. Thus, we can find that the maximization is given by

$$
\begin{aligned}
&\int p_{\text{old}}(\boldsymbol{\tau}) \log q_j(\theta_j) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} = \int_{\boldsymbol{\theta}_{-j}} \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \log \prod_{t=1}^{T} \pi(a_t, \boldsymbol{\theta}|s_t) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j} \\
\Leftrightarrow &\log q_j(\theta_j) = \left(\int p_{\text{old}}(\boldsymbol{\tau}) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \right)^{-1} \int_{\boldsymbol{\theta}_{-j}} \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \log \prod_{t=1}^{T} \pi(a_t, \boldsymbol{\theta}|s_t) p(r=1|\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}_{-j},
\end{aligned}
\tag{4}
$$

where the last line is the final approximation.

A.3 Parameter Distribution for M

The distribution of $q_{\mathbf{M}}(\mathbf{M})$ can be derived by

$$
\begin{aligned}
\log q_{\mathbf{M}}(\mathbf{M}) &= \mathbb{E}_{-\mathbf{M}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\log\left(p\left(\boldsymbol{\theta}\right)\prod_{t=1}^{T}\pi(a_t|\boldsymbol{\theta},s_t)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{M}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\log\prod_{t=1}^{T}\pi(a_t|\boldsymbol{\theta},s_t)\right.\right.\right. \\
&\quad \left.\left.\left.+\log\prod_{m=1}^{M}\mathcal{N}\left(\mathbf{m}^{(m)}|\mathbf{m}_{\text{old}}^{(m)},\sigma^2\mathbf{I}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{M}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{m=1}^{M}\sum_{j=1}^{D_m}\sum_{t=1}^{T}\log\pi(a_{tj}^{(m)}|\boldsymbol{\theta},s_t)+\right.\right.\right. \\
&\quad \left.\left.\left.\sum_{m=1}^{M}\sum_{j=1}^{D_m}\mathcal{N}\left(\mathbf{m}_{j,:}^{(m)^{\mathrm{T}}}|\mathbf{m}_{\text{old}_{j,:}}^{(m)^{\mathrm{T}}},\sigma^2\mathbf{I}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{M}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{m=1}^{M}\sum_{j=1}^{D_m}\sum_{t=1}^{T}\right.\right.\right. \\
&\quad \left(-\frac{1}{2}\left(\left(a_{t,j}^{(m)}-\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t\right)-\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)^{\mathrm{T}}\left(\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\tilde{\tau}_m^{-1}\mathbf{I}\right)^{-1}\right. \\
&\quad \left.\left.\cdot\left(\left(a_{t,j}^{(m)}-\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t\right)-\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)\right)\right) \\
&\quad \left.\left.-\sum_{m=1}^{M}\sum_{j=1}^{D_m}\frac{1}{2}\left(\mathbf{m}_{j,:}^{(m)}-\mathbf{m}_{\text{old}_{j,:}}^{(m)}\right)\sigma^{-2}\left(\mathbf{m}_{j,:}^{(m)}-\mathbf{m}_{\text{old}_{j,:}}^{(m)}\right)^{\mathrm{T}}\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{M}}\left[\sum_{m=1}^{M}\sum_{j=1}^{D_m}\mathbb{E}_{p(\boldsymbol{\tau})}\left[R(\boldsymbol{\tau})\left(\sum_{t=1}^{T}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\tilde{\tau}_m\right.\right.\right. \\
&\quad \left.\left.\left.\cdot\left(\boldsymbol{\phi}^{\mathrm{T}}\mathbf{m}_{j,:}^{(m)^{\mathrm{T}}}\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}-2\left(a_{t,j}^{(m)}-\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t\right)^{\mathrm{T}}\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)\right)\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1}\right. \\
&\quad \left.-\frac{1}{2}\left(\mathbf{m}_{j,:}^{(m)}\mathbf{m}_{j,:}^{(m)^{\mathrm{T}}}-2\mathbf{m}_{j,:}^{(m)}\mathbf{m}_{\text{old}_{j,:}}^{(m)}\right)\sigma^{-2}\right]
\end{aligned}
\tag{5}
$$

A.4 Parameter Distribution for W

The parameter distribution $q_{\mathbf{W}}(\mathbf{W})$ can be found with

$$
\begin{aligned}
\log q_{\mathbf{W}}(\mathbf{W}) &= \mathbb{E}_{-\mathbf{W}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\log\left(p(\boldsymbol{\theta})\prod_{t=1}^{T}\pi(a_t|\boldsymbol{\theta},s_t)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{W}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right(\right. \\
&\quad \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{j=1}^{D_m}\log\mathcal{N}\left(a_{t,j}^{(m)}|\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t+\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi},\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\tilde{\tau}_m^{-1}\right) \\
&\quad \left.\left.+\sum_{m=1}^{M}\sum_{j=1}^{D_m}\mathcal{N}\left({\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}|\mathbf{0},\bar{\bar{\boldsymbol{\alpha}}}_{m,K}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{W}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{j=1}^{D_m}-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\tilde{\tau}_m\right.\right.\right. \\
&\quad \cdot\left(\left(a_{t,j}^{(m)}-\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)-\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t\right)^{\mathrm{T}}\left(\left(\left(a_{t,j}^{(m)}-\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)-\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t\right)\right. \\
&\quad \left.\left.\left.-\sum_{m=1}^{M}\sum_{j=1}^{D_m}\frac{1}{2}\mathbf{w}_{j,:}^{(m)}\bar{\bar{\boldsymbol{\alpha}}}_{m,K}{\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\mathbf{W}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{j=1}^{D_m}-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\tilde{\tau}_m\right.\right.\right. \\
&\quad \cdot\left(-2\left(a_{t,j}^{(m)}-\mathbf{m}_{j,:}^{(m)}\boldsymbol{\phi}\right)\mathbf{w}_{j,:}^{(m)}\tilde{\mathbf{z}}_t+\mathbf{w}_{j,:}^{(m)}+\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^{\mathrm{T}}{\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}\right) \\
&\quad \left.\left.\left.-\sum_{m=1}^{M}\sum_{j=1}^{D_m}\frac{1}{2}\mathbf{w}_{j,:}^{(m)}\bar{\bar{\boldsymbol{\alpha}}}_{m,K}^{(m)}{\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(-2\sum_{t=1}^{T}\frac{\left(a_{t,j}^{(m)}-\mathbb{E}_{\mathbf{m}}\left[\mathbf{m}_{j,:}^{(m)}\right]\boldsymbol{\phi}\right)\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]^{\mathrm{T}}{\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbb{E}_{\tilde{\tau}}\left[\tilde{\tau}_m^{-1}\right]}\right.\right. \\
&\quad \left.\left.\mathbf{w}_{j,:}^{(m)}\left(\sum_{t=1}^{T}\frac{\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^{\mathrm{T}}\right]}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbb{E}_{\tilde{\tau}}\left[\tilde{\tau}_m\right]^{-1}}+\bar{\bar{\boldsymbol{\alpha}}}_{m,K}\right){\mathbf{w}_{j,:}^{(m)}}^{\mathrm{T}}\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1}
\end{aligned}
\tag{6}
$$

A.5 Distribution of Z

The distribution $q_{\tilde{\mathbf{Z}}}\left(\tilde{\mathbf{Z}}\right)$ can be derived with

$$
\begin{aligned}
\log q_{\tilde{\mathbf{Z}}}\left(\tilde{\mathbf{Z}}\right) &= \mathbb{E}_{-\tilde{\mathbf{z}}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\log\left(p\left(\boldsymbol{\theta}\right)\prod_{t=1}^{T}\pi\left(a_t|\boldsymbol{\theta},s_t\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\tilde{\mathbf{z}}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{t=1}^{T}\sum_{m=1}^{M}\log\mathcal{N}\left(a_t^{(m)}|\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t+\mathbf{M}^{(m)}\boldsymbol{\phi},\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\tilde{\tau}_m^{-1}\mathbf{I}\right)\right.\right.\right. \\
&\quad \left.\left.\left. + \sum_{t=1}^{T}\log\mathcal{N}\left(\tilde{\mathbf{z}}_t|\mathbf{0},\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbf{I}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\tilde{\mathbf{z}}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\tilde{\tau}_m\left(-2\left(a_t^{(m)}-\mathbf{M}^{(m)}\boldsymbol{\phi}\right)^{\mathrm{T}}\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t\right.\right.\right.\right. \\
&\quad \left.\left.\left.\left. +\tilde{\mathbf{z}}_t^{\mathrm{T}}\mathbf{W}^{(m)\mathrm{T}}\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t\right)-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\tilde{\mathbf{z}}_t^{\mathrm{T}}\tilde{\mathbf{z}}_t\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(-\frac{1}{2}\sum_{t=1}^{T}\left(-2\left(\sum_{m=1}^{M}\frac{\left(a_t^{(m)}-\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)}\right]\boldsymbol{\phi}\right)^{\mathrm{T}}\mathbf{W}^{(m)}}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbb{E}_{\tilde{\tau}}\left[\tilde{\tau}_m\right]^{-1}}\right)\tilde{\mathbf{z}}_t\right.\right.\right. \\
&\quad \left.\left.\left. +\tilde{\mathbf{z}}_t^{\mathrm{T}}\left(\sum_{m=1}^{M}\frac{\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)\mathrm{T}}\mathbf{W}^{(m)}\right]}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbb{E}_{\tilde{\tau}}\left[\tilde{\tau}_m\right]^{-1}}\right)\tilde{\mathbf{z}}_t+\frac{\tilde{\mathbf{z}}_t^{\mathrm{T}}\tilde{\mathbf{z}}_t}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)}\right)\right)\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1}
\end{aligned}
\tag{7}
$$

A.6 Distribution of Tau

The distribution $q_{\tilde{\tau}}(\tilde{\tau})$ of the precision can be found with

$$
\begin{aligned}
&\log q_{\tilde{\tau}}(\tilde{\tau}) \\
&= \mathbb{E}_{-\tilde{\tau}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\log\left(p\left(\boldsymbol{\theta}\right)\prod_{t=1}^{T}\pi\left(a_t|\boldsymbol{\theta},s_t\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\tilde{\tau}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\log\prod_{t=1}^{T}\pi\left(a_t|\boldsymbol{\theta},s_t\right)+\log\mathcal{G}\left(\tilde{\tau}|a^{\tilde{\tau}},b^{\tilde{\tau}}\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\tilde{\tau}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\sum_{t=1}^{T}\log\pi\left(a_t|\boldsymbol{\theta},s_t\right)+\sum_{m=1}^{M}\log\mathcal{G}\left(\tilde{\tau}_m|a^{\tilde{\tau}},b^{\tilde{\tau}}\right)\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \mathbb{E}_{-\tilde{\tau}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\left(\left(\sum_{m=1}^{M}\sum_{t=1}^{T}\log\left(|\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\tilde{\tau}_m^{-1}\mathbf{I}|^{-\frac{1}{2}}\right)\right.\right.\right.\right. \\
&\quad\left.\left.\left.\left.-\frac{1}{2}\left(\left(a_t^{(m)}-\mathbf{M}^{(m)}\boldsymbol{\phi}\right)-\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t\right)^{\mathrm{T}}\left(\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\tilde{\tau}_m^{-1}\mathbf{I}\right)^{-1}\left(a_t^{(m)}-\mathbf{M}^{(m)}\boldsymbol{\phi}\right)-\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t\right)\right)\right)\right. \\
&\quad\left.+\sum_{m=1}^{M}\left(\log\left(\tilde{\tau}_m^{a^{\tilde{\tau}}-1}\right)-b^{\tilde{\tau}}\tilde{\tau}_m\right)\right)\Bigg]\Bigg]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \sum_{m=1}^{M}\frac{1}{2}D_mT\log\tilde{\tau}_m+\sum_{m=1}^{M}\left(\log\left(\tilde{\tau}_m\right)\left(a^{\tilde{\tau}}-1\right)-b^{\tilde{\tau}}\tilde{\tau}_m\right)+ \\
&\quad\mathbb{E}_{-\tilde{\tau}}\left[\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\sum_{m=1}^{M}\sum_{t=1}^{T}\frac{\tilde{\tau}_m}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)}\left(a_t^{(m)^{\mathrm{T}}}a_t^{(m)}-2a_t^{(m)^{\mathrm{T}}}\mathbf{M}^{(m)}\boldsymbol{\phi}\right.\right.\right. \\
&\quad+2\tilde{\mathbf{z}}_t^{\mathrm{T}}\mathbf{W}^{(m)^{\mathrm{T}}}\mathbf{M}^{(m)}\boldsymbol{\phi}-2a_t^{(m)}\mathbf{W}^{(m)}\tilde{\mathbf{z}}_t+\boldsymbol{\phi}^{\mathrm{T}}\mathbf{M}^{(m)^{\mathrm{T}}}\mathbf{M}^{(m)}\boldsymbol{\phi} \\
&\quad\left.\left.\left.+\tilde{\mathbf{z}}_t^{\mathrm{T}}\mathbf{W}^{(m)^{\mathrm{T}}}\mathbf{W}^{(m)}\tilde{\mathbf{z}}\right)\right]\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1} \\
&= \sum_{m=1}^{M}\frac{1}{2}D_mT\log\tilde{\tau}_m+\sum_{m=1}^{M}\left(\log\left(\tilde{\tau}_m\right)\left(a^{\tilde{\tau}}-1\right)-b^{\tilde{\tau}}\tilde{\tau}_m\right)+ \\
&\quad\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\sum_{m=1}^{M}\sum_{t=1}^{T}\frac{\tilde{\tau}_m}{\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)}\left(a_t^{(m)^{\mathrm{T}}}a_t^{(m)}-2a_t^{(m)^{\mathrm{T}}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)}\right]\boldsymbol{\phi}\right.\right. \\
&\quad+2\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]^{\mathrm{T}}\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)}\right]^{\mathrm{T}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)}\right]\boldsymbol{\phi}-2a_t^{(m)^{\mathrm{T}}}\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)}\right]\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]+\boldsymbol{\phi}^{\mathrm{T}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)^{\mathrm{T}}}\mathbf{M}^{(m)}\right]\boldsymbol{\phi} \\
&\quad\left.\left.+\mathbb{E}_{\tilde{\mathbf{z}},\mathbf{W}}\left[\tilde{\mathbf{z}}_t^{\mathrm{T}}\mathbf{W}^{(m)^{\mathrm{T}}}\mathbf{W}^{(m)}\tilde{\mathbf{z}}\right]\right)\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1}
\end{aligned}
\tag{8}
$$

If we apply now the quadratic expectation

$$
\mathbb{E}\left[\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right]=\mathrm{Tr}\left(\mathbf{A}\mathrm{Cov}\left[\mathbf{x}\right]\right)+\mathbb{E}\left[\mathbf{x}\right]^{\mathrm{T}}\mathbf{A}\mathbb{E}\left[\mathbf{x}\right], \tag{9}
$$

then we get

$$
\begin{aligned}
&= \sum_{m=1}^{M}\frac{1}{2}D_mT\log\tilde{\tau}_m+\sum_{m=1}^{M}\left(\log\left(\tilde{\tau}_m\right)\left(a^{\tilde{\tau}}-1\right)-b^{\tilde{\tau}}\tilde{\tau}_m\right)- \\
&\quad\frac{1}{2}\sum_{m=1}^{M}\tilde{\tau}_m\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\sum_{t=1}^{T}\mathrm{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\left(a_t^{(m)^{\mathrm{T}}}a_t^{(m)}-2a_t^{(m)^{\mathrm{T}}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)}\right]\boldsymbol{\phi}\right.\right. \\
&\quad+2\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]^{\mathrm{T}}\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)}\right]^{\mathrm{T}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)}\right]\boldsymbol{\phi}-2a_t^{(m)^{\mathrm{T}}}\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)}\right]\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]+\boldsymbol{\phi}^{\mathrm{T}}\mathbb{E}_{\mathbf{M}}\left[\mathbf{M}^{(m)^{\mathrm{T}}}\mathbf{M}^{(m)}\right]\boldsymbol{\phi} \\
&\quad\left.\left.+\mathrm{Tr}\left(\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)^{\mathrm{T}}}\mathbf{W}^{(m)}\right]\mathrm{Cov}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]\right)+\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]^{\mathrm{T}}\mathbb{E}_{\mathbf{W}}\left[\mathbf{W}^{(m)^{\mathrm{T}}}\mathbf{W}^{(m)}\right]\mathbb{E}_{\tilde{\mathbf{z}}}\left[\tilde{\mathbf{z}}_t\right]\right)\right]\left(\mathbb{E}_{p(\boldsymbol{\tau})}\left[p(r=1|\boldsymbol{\tau})\right]\right)^{-1}
\end{aligned}
\tag{10}
$$