Bayesian Factor Analysis

Carlos Sevilla-Salcedo*

February 2021

In this document I will extend and explain the bayesian formulation of two well known factor analysis models, Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA). The main objective of this document is to develop the equations and justify the complete distributions of the random variables of both these models.

One of the most used extensions of the probabilistic formulation is the bayesian approach. This does not only provide prior information that include previous information to the model, but also develops the posterior distribution of the observed data to sample from that distributions. This implies that we are not only estimating the value of each parameter but also the distribution that describes each one of them. The process to estimate the distribution of these variables is known as bayesian variational inference.

First of all, in order to work with bayesian variational inference we need to know the posterior distribution of the variables, $p(\Theta|\,\mathbf{X})$, where Θ is the group of all the variables. However, this posterior distribution is not tractable. Nevertheless, we can approximate it by a distribution $q(\Theta)$, being $q(\Theta)$ a group of treatable variables.

In order to adjust $q(\Theta)$, we need to minimise the following equation:

$$KL(q||p) \quad = \quad \int q(\Theta)\ln\left(\frac{q(\Theta)}{p(\Theta|\,\mathbf{X})}\right)d\Theta, \tag{1}$$

where KL symbolizes the Kullback-Leibler divergence, which measures the difference between two distributions. This implies that we minimise the distance between these two distributions. If we develop this equation, we have that

$$\begin{aligned} KL(q||p) &= \int q(\Theta)\ln\left(\frac{q(\Theta)}{\frac{p(\mathbf{X},\Theta)}{p(\mathbf{X})}}\right)d\Theta \\ &= \int q(\Theta)\left(\ln\left(\frac{q(\Theta)}{p(\mathbf{X},\Theta)}\right)+\ln(p(\mathbf{X}))\right)d\Theta \\ &= \int q(\Theta)\ln\left(\frac{q(\Theta)}{p(\mathbf{X},\Theta)}\right)d\Theta+\int q(\Theta)\ln(p(\mathbf{X}))d\Theta \\ &= \int q(\Theta)\ln\left(\frac{q(\Theta)}{p(\mathbf{X},\Theta)}\right)d\Theta+\ln(p(\mathbf{X})). \end{aligned} \tag{2}$$

This way, we can rearrange the previous equation to get that

$$L(q) \quad = \quad \ln(p(\mathbf{X}))-KL(q||p)\leq\ln(p(\mathbf{X})), \tag{3}$$

*Corresponding author. Email address: sevisal@tsc.uc3m.es

having that

$$L(q) = -\int q(\Theta) \ln\left(\frac{q(\Theta)}{p(\mathbf{X}, \Theta)}\right) d\Theta \tag{4}$$

is a lower bound of $\ln(p(\mathbf{X}))$. For this reason, maximising the bound implies minimising $KL(q||p)$, keeping in mind $L(q)$ will have a maximum value when $p = q$. This simple trick allows us to approximate the original variable distribution to another that can be calculated, e.g. using the mean field method.

The mean field method

The mean field method raises as an approximation to the posterior of the model variables previously explained. Specifically, it considers that one can approximate the posterior distribution of all the model variables $q(\Theta)$ by factorising over all the variables

$$q(\Theta) \;\; = \;\; \prod_i q(\Theta_i) = \prod_i q_i. \tag{5}$$

Hence, we can use the lower bound to determine the distribution $q(\Theta)$, so that it maximizes $L(q)$. To do so, let's substitute Equation (5) in (4)

$$\begin{aligned}
L(q_j) &= \int q(\Theta) \ln\left(\frac{p(\mathbf{X}, \Theta)}{q(\Theta)}\right) d\Theta = \int \prod_i q_i \left[\ln(p(\mathbf{X}, \Theta)) - \sum_i \ln(q_i)\right] d\Theta \\
&= \int \prod_i q_i \ln(p(\mathbf{X}, \Theta)) d\Theta - \int \prod_i q_i \sum_i \ln(q_i) d\Theta \\
&= \int q_j \prod_{i \neq j} q_i \ln(p(\mathbf{X}, \Theta)) d\Theta - \int q_j \prod_{i \neq j} q_i \left(\ln(q_j) + \sum_{i \neq j} \ln(q_i)\right) d\Theta \\
&= \int q_j \prod_{i \neq j} q_i \ln(p(\mathbf{X}, \Theta)) d\Theta - \int q_j \prod_{i \neq j} q_i \sum_{i \neq j} \ln(q_i) d\Theta - \int q_j \prod_{i \neq j} q_i \ln(q_j) d\Theta \\
&= \int q_j \left[\int \prod_{i \neq j} q_i \ln(p(\mathbf{X}, \Theta)) d\Theta_i\right] d\Theta_j - \int q_j \ln(q_j) d\Theta_j + \text{const} \\
&= \int q_j \ln(f_j) d\Theta_j - \int q_j \ln(q_j) d\Theta_j + \text{const}
\end{aligned} \tag{6}$$

where

$$\ln(f_j) \;\; = \;\; \mathbb{E}_{-q_j}[\ln(p(\mathbf{X}, \Theta))] + \text{const} \tag{7}$$

and $-q_j$ means that we calculate this on all the variables except the j-th variable. It can be now seen that Equation (6) is a negative KL between $q_j(\Theta_j)$ and f_j, thus to maximise $L(q_j)$ we need to minimise $KL(q_j||f_j)$. Therefore, we have that the optimum solution has the following expression:

$$\ln(q_j^*) \;\; = \;\; \mathbb{E}_{-q_j}[\ln(p(\mathbf{X}, \Theta))] + \text{const}. \tag{8}$$

This constitutes the basis of the variational inference and is the approximation we will use in the following approaches.

1 Bayesian PCA

Here we explain the bayesian extension of the PPCA first presented in Bishop (1999). To do so, we adapt the probabilistic version by including a prior over the parameters of the model $(\mathbf{W}, \sigma^2)$. We will firstly define what is the generative model, including the variable distributions and the graphic model, to later present the variational inference result obtained.

Generative model

As we have seen before, PCA is a FA algorithm that combines the information of a set of observed data and determines a latent space. The particularity of this latent space is that its dimensions are orthogonal to one another. Keeping this in mind, we can start by defining the distribution over the observations marginalizing with bayes rules

$$p(\mathbf{X}) \sim \int p(\mathbf{X} \,|\, \mathbf{W}, \tau) p(\mathbf{W}, \tau |\, \mathbf{X}) d\mathbf{W} d\tau \tag{9}$$

where we have defined a new noise variable $\tau = \frac{1}{\sigma^2}$. This way, we can define our generative model presented in Figure 1, which includes the different variables of the model as well as the relation between them.

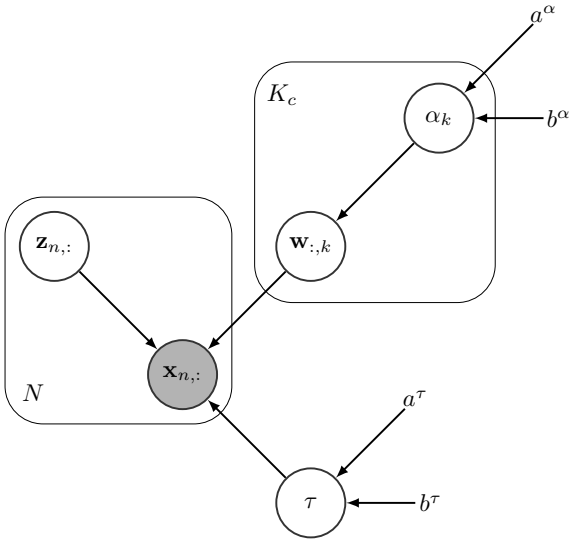

Figure 1: Plate diagram for the bayesian PCA graphical model. Gray circles denote observed variables, white circles unobserved random variables. Nodes without a circle correspond to the hyperparameters.

Furthermore, we can define the distribution of all the model variables included in the graphic model

$$\mathbf{z}_{\mathrm{n},:} \sim \mathcal{N}(0, \mathbf{I}_{\mathrm{K_c}}) \tag{10}$$

$$\mathbf{w}_{:,\mathrm{k}} \sim \mathcal{N}(0, \mathbf{I}_{\mathrm{K_c}}) \tag{11}$$

$$\mathbf{x}_{\mathrm{n},:} \,|\, \mathbf{z}_{\mathrm{n},:} \sim \mathcal{N}(\mathbf{z}_{\mathrm{n},:}\, \mathbf{W}^{\mathrm{T}}, \tau^{-1}\, \mathbf{I}_{\mathrm{D}}) \tag{12}$$

$$\tau \sim \Gamma(a^\tau, b^\tau) \tag{13}$$

where we are assuming the observed data to be independent, which implies that the random noise modeled is the same for every sample. Note that the distribution of the observed data, $\mathbf{x}_{\mathrm{n},:}$, connects the different random variables (r.v.).

1.1 Variational inference

Making use of the variational inference definition previously stated, the problem can be defined to apply those techniques to design the model. First of all, considering Figure 1 we can extract the probability distribution of the model

$$p(\Theta|\,\mathbf{X}) \;\approx\; q(\mathbf{W})q(\tau)\prod_{n=1}^{N} q(\mathbf{z}_{n,:}) \tag{14}$$

This way, we can determine the optimum distribution of the different variables, applying Equation (8) and then calculate the lower bound defined in Equation (6).

In this section the different distributions specified in Equation (14) will be developed making use of the variational inference.

1.1.1 Distribution of W

$$\ln(q^*(\mathbf{W})) \;=\; \mathbb{E}_{\mathbf{Z},\tau}[\ln(p(\mathbf{X},\mathbf{W},\mathbf{Z},\tau))] = \mathbb{E}_{\mathbf{Z},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] + \mathbb{E}[\ln(p(\mathbf{W}))] + \text{const} \tag{15}$$

We can now evaluate both terms independently and then sum the results:

$$\begin{aligned}
\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau)) \;&\approx\; \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln\Big(\mathcal{N}\Big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}},(\tau)^{-1}I\Big)\Big) + \text{const}\\
&=\; \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(-\frac{1}{2}\ln\left|(\tau)^{-1}I\right| -\frac{1}{2}\big(\mathbf{x}_{\mathrm{n},:}-\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\big)\tau\big(\mathbf{x}_{\mathrm{n},:}-\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\big)^{T}\right) + \text{const}\\
&=\; -\frac{\tau}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\big(\mathbf{x}_{\mathrm{n},:}\,\mathbf{x}_{\mathrm{n},:}^{\mathrm{T}} - 2\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} + \mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\big) + \text{const}\\
&=\; \sum_{\mathrm{n}=1}^{\mathrm{N}}\Big(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Big) + \text{const}\\
&=\; \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}}\big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\,\mathbf{w}_{\mathrm{d},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\big)\right) + \text{const}\\
&=\; \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}}\big(\mathbf{w}_{\mathrm{d},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)\right) + \text{const}
\end{aligned} \tag{16}$$

and applying the expectation we have that

$$\mathbb{E}_{\mathbf{Z},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] \;=\; \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\langle\tau\rangle\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\langle\mathbf{z}_{\mathrm{n},:}\rangle - \frac{\langle\tau\rangle}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}}\big(\mathbf{w}_{\mathrm{d},:}\langle\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathbf{z}_{\mathrm{n},:}\rangle\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)\right) \tag{17}$$

Regarding the second, we have that

$$\mathbb{E}[\ln(p(\mathbf{w}_{\mathrm{d},:}))] \;=\; \ln(p(\mathbf{w}_{\mathrm{d},:})) = \sum_{\mathrm{d}=1}^{\mathrm{D}}\left(-\frac{1}{2}\,\mathbf{w}_{\mathrm{d},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\right) + \text{const} \tag{18}$$

Finally, joining Equations (17) and (18), we obtain the mean field approximation

$$\begin{aligned}
&\ln(q^*(\mathbf{W})) \\
&= \sum_{d=1}^{D}\left(-\frac{1}{2}\,\mathbf{w}_{d,:}\,\mathbf{w}_{d,:}^{T}\right) + \sum_{n=1}^{N}\left(\langle\tau\rangle\,\mathbf{x}_{n,:}\,\mathbf{W}\langle\mathbf{z}_{n,:}\rangle - \frac{\langle\tau\rangle}{2}\sum_{d=1}^{D}\left(\mathbf{w}_{d,:}\langle\mathbf{z}_{n,:}^{T}\,\mathbf{z}_{n,:}\rangle\,\mathbf{w}_{d,:}^{T}\right)\right) + \text{const} \\
&= \sum_{d=1}^{D}\left(-\frac{1}{2}\,\mathbf{w}_{d,:}\,\mathbf{w}_{d,:}^{T} + \sum_{n=1}^{N}\left(\langle\tau\rangle\,x_{n,d}\,\mathbf{w}_{d,:}\langle\mathbf{z}_{n,:}\rangle - \frac{\langle\tau\rangle}{2}\,\mathbf{w}_{d,:}\langle\mathbf{z}_{n,:}^{T}\,\mathbf{z}_{n,:}\rangle\,\mathbf{w}_{d,:}^{T}\right)\right) + \text{const} \\
&= \sum_{d=1}^{D}\left(-\frac{1}{2}\,\mathbf{w}_{d,:}\left(\mathbf{I} + \langle\tau\rangle\langle\mathbf{z}_{n,:}^{T}\,\mathbf{z}_{n,:}\rangle\right)\mathbf{w}_{d,:}^{T} + \langle\tau\rangle\,\mathbf{w}_{d,:}\langle\mathbf{Z}^{T}\rangle\,\mathbf{x}_{:,d}^{T}\right) + \text{const} \qquad (19)
\end{aligned}$$

This way, comparing the results with the normal distribution, we can identify terms, extracting the following conclusions:

$$q^*(\mathbf{W}) = \prod_{d=1}^{D}\left(\mathcal{N}\left(\mathbf{w}_{d,:}\,|\mu_{\mathbf{w}_{d,:}}, \Sigma_{\mathbf{W}}\right)\right) \qquad (20)$$

where the variance is

$$\Sigma_{\mathbf{W}}^{-1} = \mathbf{I} + \langle\tau\rangle\langle\mathbf{Z}^{T}\,\mathbf{Z}\rangle \qquad (21)$$

and the mean is

$$\mu_{\mathbf{w}_{d,:}} = \langle\tau\rangle\Sigma_{\mathbf{W}}\langle\mathbf{Z}\rangle^{T}\,\mathbf{x}_{:,d}^{T} \qquad (22)$$

or in a matricial way

$$\langle\mathbf{W}\rangle = \langle\tau\rangle\,\mathbf{X}^{T}\langle\mathbf{Z}\rangle\Sigma_{\mathbf{W}} \qquad (23)$$

1.1.2 Distribution of Z

$$\ln(q^*(\mathbf{Z})) = \mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X},\mathbf{W},\mathbf{Z},\tau))] = \mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] + \mathbb{E}[\ln(p(\mathbf{Z}))] \qquad (24)$$

We will again start by calculating the first term (which is similar to equation (17)

$$\mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] = \sum_{n=1}^{N}\left(\langle\tau\rangle\,\mathbf{x}_{n,:}\langle\mathbf{W}\rangle\,\mathbf{z}_{n,:}^{T} - \frac{\langle\tau\rangle}{2}\,\mathbf{z}_{n,:}\langle\mathbf{W}^{T}\,\mathbf{W}\rangle\,\mathbf{z}_{n,:}^{T}\right) + \text{const} \qquad (25)$$

As the associated prior of this variable is the same as the previous one, the result of the second term is equivalent

$$\mathbb{E}[\ln(p(\mathbf{Z}))] = \ln(p(\mathbf{Z})) = \sum_{n=1}^{N}\left(-\frac{1}{2}\,\mathbf{z}_{n,:}\,\mathbf{z}_{n,:}^{T}\right) + \text{const} \qquad (26)$$

Therefore, joining the different terms in Equations (25) and (26), the mean field approximation has the form

$$\ln(q^*(\mathbf{Z})) = \sum_{n=1}^{N}\left(-\frac{1}{2}\,\mathbf{z}_{n,:}\,\mathbf{z}_{n,:}^{T} + \langle\tau\rangle\,\mathbf{x}_{n,:}\langle\mathbf{W}\rangle\,\mathbf{z}_{n,:}^{T} - \frac{\langle\tau\rangle}{2}\,\mathbf{z}_{n,:}\langle\mathbf{W}^{T}\,\mathbf{W}\rangle\,\mathbf{z}_{n,:}^{T}\right) + \text{const} \qquad (27)$$

Again, if we identify terms from the previous equation we get that

$$q^*(\mathbf{Z}) \;=\; \prod_{n=1}^{N} \left(\mathcal{N}\left(\mathbf{z}_{\mathrm{n},:} \,|\mu_{\mathbf{z}_{\mathrm{n},:}}, \Sigma_{\mathbf{Z}}\right)\right) \tag{28}$$

where the variance is

$$\Sigma_{\mathbf{Z}}^{-1} \;=\; I + \sum_{\mathrm{m}=1}^{\mathrm{M}} \left(\langle\tau\rangle \langle \mathbf{W}^{\mathrm{T}}\,\mathbf{W}\rangle\right) \tag{29}$$

and the mean is

$$\mu_{\mathbf{z}_{\mathrm{n},:}} \;=\; \sum_{\mathrm{m}=1}^{\mathrm{M}} (\langle\tau\rangle\, \mathbf{x}_{\mathrm{n},:} \langle \mathbf{W}\rangle \Sigma_{\mathbf{Z}}) \tag{30}$$

or in a matricial way

$$\langle \mathbf{Z}\rangle \;=\; \sum_{\mathrm{m}=1}^{\mathrm{M}} (\langle\tau\rangle\, \mathbf{X} \langle \mathbf{W}\rangle \Sigma_{\mathbf{Z}}) \tag{31}$$

1.1.3 Distribution of τ

Let's now calculate the approximate distribution of the noise variable, τ.

$$\ln(q^*(\tau)) \;=\; \mathbb{E}_{\mathbf{W},\mathbf{Z}}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] + \mathbb{E}[\ln(p(\tau))] \tag{32}$$

Similarly to what was done in equation (17) we can now calculate the first term, although the terms that will be constant for the expectation will vary

$$\begin{aligned}
\ln(p(\mathbf{X}|\,\mathbf{W},\mathbf{Z},\tau)) &= \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln\left(\mathcal{N}\left(\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}, (\tau)^{-1} I\right)\right) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} \left(\frac{1}{2}\ln|\tau| - \frac{\tau}{2}\left(\mathrm{x}_{\mathrm{n,d}} - \mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\right)^2\right) + \text{const} \\
&= \frac{DN}{2}\ln(\tau) - \frac{\tau}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} \left({\mathrm{x}_{\mathrm{n,d}}}^2 - 2\,\mathbf{w}_{\mathrm{d},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathrm{x}_{\mathrm{n,d}} + \left(\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\right)^2\right) + \text{const} \\
&= \frac{DN}{2}\ln(\tau) - \frac{\tau}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\sum_{\mathrm{d}=1}^{\mathrm{D_m}} \mathbf{w}_{\mathrm{d},:}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{x}_{:,\mathrm{d}} + \sum_{\mathrm{d}=1}^{\mathrm{D}} \mathbf{w}_{\mathrm{d},:}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\right) + \text{const} \\
&= \frac{DN}{2}\ln(\tau) - \frac{\tau}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\,\mathrm{Tr}\left\{\mathbf{W}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{X}\right\} + \mathrm{Tr}\left\{\mathbf{W}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\,\mathbf{W}^{\mathrm{T}}\right\}\right) + \text{const}
\end{aligned}$$

We can now calculate the expectation with respect the rest of the random variables

$$\begin{aligned}
&\mathbb{E}_{\mathbf{W},\mathbf{Z}}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] \\
&\quad= \frac{DN}{2}\ln(\tau) - \frac{\tau}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\,\mathrm{Tr}\left\{\langle\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}\right\} + \mathrm{Tr}\left\{\langle\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\right\}\right) + \text{const}
\end{aligned}$$

The second term can subsequently be determined as:

$$\mathbb{E}[\ln(p(\tau))] \;=\; \ln(p(\tau)) = -b_0^\tau \tau + (a_0^\tau - 1)\ln(\tau) + \text{const}$$

Joining them together, we have that:

$$\begin{aligned}\ln(q^*(\tau)) \;=\; & \frac{DN}{2}\ln(\tau) - \frac{\tau}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\,\mathrm{Tr}\big\{\langle\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}\big\} + \mathrm{Tr}\big\{\langle\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\big\}\right) \\ & - b_0^\tau\tau + (a_0^\tau - 1)\ln(\tau) + \text{const}\end{aligned}$$

So by identifying terms with the distribution, we would have that for this parameter:

$$q^*(\tau) \;=\; (Gamma(\tau|a^\tau, b^\tau)) \tag{33}$$

where the first parameter of the distribution is

$$a^\tau \;=\; \frac{DN}{2} + \alpha_0^\tau \tag{34}$$

and the second is

$$b^\tau \;=\; \beta_0^\tau + \frac{1}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\,\mathrm{Tr}\big\{\langle\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}\big\} + \mathrm{Tr}\big\{\langle\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\rangle\langle\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\big\}\right) \tag{35}$$

2 Bayesian PCA with ARD

2.1 Generative model

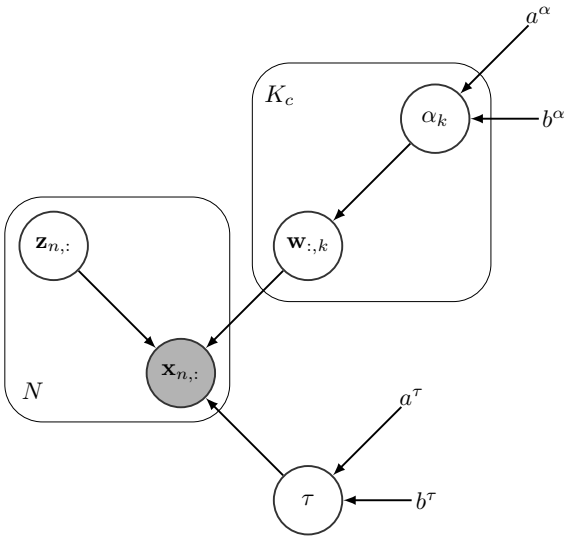

Figure 2: Plate diagram for the bayesian PCA with ARD graphical model. Gray circles denote observed variables, white circles unobserved random variables. Nodes without a circle correspond to the hyperparameters.

We can define the distribution over the observations by marginalizing using the bayes rule

$$p(\mathbf{X}) \sim \int p(\mathbf{X} \mid \mathbf{W}, \boldsymbol{\alpha}, \tau) p(\mathbf{W}, \boldsymbol{\alpha}, \tau \mid \mathbf{X}) d\,\mathbf{W}\, d\,\boldsymbol{\alpha}\, d\tau \tag{36}$$

Therefore, we can now establish the prior distributions over the different random variables in the problem

$$\begin{aligned}
\mathbf{z}_{\mathrm{n},:} &\sim \mathcal{N}(0, \mathbf{I}_{\mathrm{K_c}}) && (37)\\
\mathbf{w}_{:,\mathrm{k}} &\sim \mathcal{N}\left(0, {\alpha_k}^{-1} \mathbf{I}_{\mathrm{K_c}}\right) && (38)\\
\mathbf{x}_{\mathrm{n},:} \mid \mathbf{z}_{\mathrm{n},:} &\sim \mathcal{N}(\mathbf{z}_{\mathrm{n},:} \mathbf{W}^{\mathrm{T}}, \tau^{-1} \mathbf{I}_{\mathrm{D}}) && (39)\\
\alpha_k &\sim \Gamma(a^{\boldsymbol{\alpha}}, b^{\boldsymbol{\alpha}}) && (40)\\
\tau &\sim \Gamma(a^{\tau}, b^{\tau}) && (41)
\end{aligned}$$

2.2 Variational inference

Making use of the variational inference definition previously stated, the problem can be defined to apply those techniques to design the model. First of all, considering Figure 2 we can extract the probability distribution of the model

$$p(\Theta \mid \mathbf{X}) \approx \left(q(\mathbf{W}) q(\tau) \prod_{\mathrm{k}=1}^{\mathrm{K_c}} q(\alpha_k) \right) \prod_{\mathrm{n}=1}^{\mathrm{N}} q(\mathbf{z}_{\mathrm{n},:}) \tag{42}$$

This way, we can determine the optimum distribution of the different variables, applying Equation (8) and then calculate the lower bound defined in Equation (6).

In this section the different distributions specified in Equation (42) will be developed making use of the variational inference.

2.3 Distribution of W

$$\begin{aligned}
\ln(q^*(\mathbf{W})) &= \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha},\tau}[\ln(p(\mathbf{X}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \tau))] \\
&= \mathbb{E}_{\mathbf{Z},\tau}[\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau))] + \mathbb{E}_{\boldsymbol{\alpha}}[\ln(p(\mathbf{W} \mid \boldsymbol{\alpha}))] + \text{const} && (43)
\end{aligned}$$

We can now evaluate both terms independently and then sum the results:

$$\begin{aligned}
\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau)) &\approx \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln\Big(\mathcal{N}\Big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}, (\tau)^{-1} I\Big)\Big) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(-\frac{1}{2}\ln\left|\tau^{-1} I\right| - \frac{1}{2}\big(\mathbf{x}_{\mathrm{n},:} - \mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\big)\tau\big(\mathbf{x}_{\mathrm{n},:} - \mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\big)^{T}\right) + \text{const} \\
&= -\frac{\tau}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}} \big(\mathbf{x}_{\mathrm{n},:}\,\mathbf{x}_{\mathrm{n},:}^{\mathrm{T}} - 2\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} + \mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\big) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \Big(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{\mathrm{T}}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Big) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}} \big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\,\mathbf{w}_{\mathrm{d},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\big)\right) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\tau\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} - \frac{\tau}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}} \big(\mathbf{w}_{\mathrm{d},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)\right) + \text{const}
\end{aligned} \tag{44}$$

This way the expectation can be calculated as:

$$\mathbb{E}_{\mathbf{Z},\tau}[\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau))] = \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\langle\tau\rangle\,\mathbf{x}_{\mathrm{n},:}\,\mathbf{W}\langle\mathbf{z}_{\mathrm{n},:}\rangle - \frac{\langle\tau\rangle}{2}\sum_{\mathrm{d}=1}^{\mathrm{D}} \big(\mathbf{w}_{\mathrm{d},:}\langle\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathbf{z}_{\mathrm{n},:}\rangle\,\mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)\right) \tag{45}$$

Equivalently, the second term would be:

$$\begin{aligned}
\ln(p(\mathbf{W} \mid \boldsymbol{\alpha})) &= \sum_{\mathrm{d}=1}^{\mathrm{D}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}} \ln(p(\mathrm{w_{d,k}} \mid \alpha_k)) \\
&= \sum_{\mathrm{d}=1}^{\mathrm{D}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(\frac{1}{2}\ln(\alpha_k) - \frac{\alpha_k}{2}{\mathrm{w_{d,k}}}^2\right) + \text{const}
\end{aligned} \tag{46}$$

$$\mathbb{E}_{\boldsymbol{\alpha}}[\ln(p(\mathbf{W} \mid \boldsymbol{\alpha}))] = \sum_{\mathrm{d}=1}^{\mathrm{D}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(-\frac{\langle\alpha_k\rangle}{2}{\mathrm{w_{d,k}}}^2\right) + \text{const} \tag{47}$$

Finally, joining Equations (45) and (47), we would have that the optimum solution for

the variable $\mathbf{W}$ is:

$$\begin{aligned}\ln(q^*(\mathbf{W})) &= \sum_{d=1}^{D}\sum_{k=1}^{K_c}\left(-\frac{\langle\alpha_k\rangle}{2}{w_{d,k}}^2\right) + \sum_{n=1}^{N}\Big(\langle\tau\rangle\, x_{n,:}\, \mathbf{W}\langle z_{n,:}\rangle \\ &\quad -\frac{\langle\tau\rangle}{2}\sum_{d=1}^{D}\Big(w_{d,:}\langle z_{n,:}^T\, z_{n,:}\rangle\, w_{d,:}^T\Big)\Big) + \text{const} \\ &= \sum_{d=1}^{D}\left(-\frac{1}{2}\, w_{d,:}\, diag(\langle\alpha_k\rangle)\, w_{d,:}^T + \sum_{n=1}^{N}\Big(\langle\tau\rangle\, x_{n,d}\, w_{d,:}\langle z_{n,:}\rangle\right. \\ &\quad \left.-\frac{\langle\tau\rangle}{2}\, w_{d,:}\langle z_{n,:}^T\, z_{n,:}\rangle\, w_{d,:}^T\Big)\right) + \text{const} \\ &= \sum_{d=1}^{D}\Big(-\frac{1}{2}\, w_{d,:}\Big(diag(\langle\alpha_k\rangle) + \langle\tau\rangle\langle z_{n,:}^T\, z_{n,:}\rangle\Big)\, w_{d,:}^T + \langle\tau\rangle\, w_{d,:}\langle\mathbf{Z}^T\rangle\, x_{:,d}^T\Big) + \text{const}\end{aligned} \tag{48}$$

This way, comparing the results with the normal distribution, we can identify terms, extracting the following conclusions:

$$q^*(\mathbf{W}) = \prod_{d=1}^{D}\Big(\mathcal{N}\big(w_{d,:}\,|\mu_{w_{d,:}}, \Sigma_{\mathbf{W}}\big)\Big) \tag{49}$$

Where the variance would be:

$$\Sigma_{\mathbf{W}}^{-1} = diag(\langle\boldsymbol{\alpha}\rangle) + \langle\tau\rangle\langle\mathbf{Z}^T\,\mathbf{Z}\rangle \tag{50}$$

And the mean could be expressed as:

$$\begin{aligned}\mu_{w_{d,:}} &= \langle\tau\rangle\Sigma_{\mathbf{W}}\langle\mathbf{Z}\rangle^T\, x_{:,d}^T \\ \langle\mathbf{W}\rangle &= \langle\tau\rangle\, \mathbf{X}^T\langle\mathbf{Z}\rangle\Sigma_{\mathbf{W}}\end{aligned} \tag{51}$$

2.4 Distribution of Z

$$\ln(q^*(\mathbf{Z})) = \mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X},\mathbf{W},\mathbf{Z},\tau))] = \mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] + \mathbb{E}[\ln(p(\mathbf{Z}))] \tag{52}$$

The first term is similar to equation (17), changing the expectation dependency:

$$\mathbb{E}_{\mathbf{W},\tau}[\ln(p(\mathbf{X}\,|\,\mathbf{W},\mathbf{Z},\tau))] = \sum_{n=1}^{N}\left(\langle\tau\rangle\, x_{n,:}\langle\mathbf{W}\rangle\, z_{n,:}^T - \frac{\langle\tau\rangle}{2}\, z_{n,:}\langle\mathbf{W}^T\,\mathbf{W}\rangle\, z_{n,:}^T\right) + \text{const} \tag{53}$$

The second term could be calculated as:

$$\mathbb{E}[\ln(p(\mathbf{Z}))] = \ln(p(\mathbf{Z})) = \sum_{n=1}^{N}\left(-\frac{1}{2}\, z_{n,:}\, z_{n,:}^T\right) + \text{const} \tag{54}$$

Therefore, joining the different terms in Equations (53) and (54) , the optimum solution has the form:

$$\ln(q^*(\mathbf{Z})) = \sum_{n=1}^{N}\left(-\frac{1}{2}\, z_{n,:}\, z_{n,:}^T + \langle\tau\rangle\, x_{n,:}\langle\mathbf{W}\rangle\, z_{n,:}^T - \frac{\langle\tau\rangle}{2}\, z_{n,:}\langle\mathbf{W}^T\,\mathbf{W}\rangle\, z_{n,:}^T\right) + \text{const} \tag{55}$$

This way, we can now identify terms, having that:

$$q^*(\mathbf{Z}) = \prod_{n=1}^{N} \left(\mathcal{N}\left(\mathbf{z}_{\mathrm{n},:} \,|\mu_{\mathbf{z}_{\mathrm{n},:}}, \Sigma_{\mathbf{Z}} \right) \right) \tag{56}$$

Where the variance would be:

$$\Sigma_{\mathbf{Z}}^{-1} = I + \langle \tau \rangle \langle \mathbf{W}^{\mathrm{T}} \mathbf{W} \rangle \tag{57}$$

And the mean could be expressed as:

$$\begin{aligned} \mu_{\mathbf{z}_{\mathrm{n},:}} &= \langle \tau \rangle \, \mathbf{x}_{\mathrm{n},:} \langle \mathbf{W} \rangle \Sigma_{\mathbf{Z}} \\ \langle \mathbf{Z} \rangle &= \langle \tau \rangle \, \mathbf{X} \langle \mathbf{W} \rangle \Sigma_{\mathbf{Z}} \end{aligned} \tag{58}$$

2.5 Distribution of $\boldsymbol{\alpha}$

$$\ln(q^*(\boldsymbol{\alpha})) = \mathbb{E}_{\mathbf{W}}[\ln(p(\mathbf{W} \,|\, \boldsymbol{\alpha}))] + \mathbb{E}[\ln(p(\boldsymbol{\alpha}))] \tag{59}$$

Similarly to what we have done before, we can develop both terms independently:

$$\begin{aligned} \ln(p(\mathbf{W} \,|\, \boldsymbol{\alpha})) &= \sum_{\mathrm{d}=1}^{\mathrm{D}} \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(\frac{1}{2} \ln |\alpha_k| - \frac{1}{2} \alpha_k (\mathrm{w}_{\mathrm{d,k}})^2 \right) + \text{const} \\ &= \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(\frac{D}{2} \ln(\alpha_k) - \frac{1}{2} \alpha_k \, \mathbf{w}_{:,\mathrm{k}}^{\mathrm{T}} \, \mathbf{w}_{:,\mathrm{k}} \right) + \text{const} \end{aligned}$$

$$\mathbb{E}[\ln(p(\mathbf{W} \,|\, \boldsymbol{\alpha}))] = \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(\frac{D}{2} \ln(\alpha_k) - \frac{1}{2} \alpha_k \langle \mathbf{w}_{:,\mathrm{k}}^{\mathrm{T}} \, \mathbf{w}_{:,\mathrm{k}} \rangle \right) + \text{const} \tag{60}$$

The second term is deployed as follows:

$$\ln(p(\alpha_k)) = -\beta_0 \, \alpha_k + (\alpha_0 - 1) \ln(\alpha_k) + \text{const}$$

$$\mathbb{E}[\ln(p(\boldsymbol{\alpha}))] = \sum_{\mathrm{k}=1}^{\mathrm{K_c}} (\ln(p(\alpha_k))) = \sum_{\mathrm{k}=1}^{\mathrm{K_c}} (-\beta_0 \, \alpha_k + (\alpha_0 - 1) \ln(\alpha_k)) + \text{const} \tag{61}$$

Therefore, by joining both Equations (60) and (61) together, we can now determine the value f the expectation we were looking for

$$\ln(q^*(\boldsymbol{\alpha})) = \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \left(\frac{D}{2} \ln(\alpha_k) - \frac{1}{2} \alpha_k \langle \mathbf{w}_{:,\mathrm{k}}^{\mathrm{T}} \, \mathbf{w}_{:,\mathrm{k}} \rangle - \beta_0 \, \alpha_k + (\alpha_0 - 1) \ln(\alpha_k) \right) + \text{const} \tag{62}$$

This way, we can now identify terms, having

$$q^*(\boldsymbol{\alpha}) = \prod_{\mathrm{k}=1}^{\mathrm{K_c}} (Gamma(\alpha_k \,|a_{\alpha_k}, b_{\alpha_k})), \tag{63}$$

where the variance would be:

$$a_{\alpha_k} = \frac{D}{2} + \alpha_0 \tag{64}$$

and the mean is

$$b_{\alpha_k} = \beta_0 + \frac{1}{2}\langle \mathbf{w}_{:,\mathrm{k}}^{\mathrm{T}}\, \mathbf{w}_{:,\mathrm{k}} \rangle = \beta_0 + \frac{1}{2}\langle \mathbf{W}^{\mathrm{T}}\, \mathbf{W} \rangle_{k,k}, \tag{65}$$

where

$$\langle \mathbf{w}_{:,\mathrm{k}}^{\mathrm{T}}\, \mathbf{w}_{:,\mathrm{k}} \rangle = \sum_{\mathrm{d}=1}^{\mathrm{D}} (\mathrm{w}_{\mathrm{d,k}})^2 = \langle \mathbf{W}^{\mathrm{T}}\, \mathbf{W} \rangle_{k,k}.$$

2.6 Distribution of τ

$$\ln(q^*(\tau)) = \mathbb{E}_{\mathbf{W},\mathbf{Z}}[\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau))] + \mathbb{E}[\ln(p(\tau))] \tag{66}$$

Similarly to what was done in equation (45):

$$\begin{aligned}
\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau)) &= \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln\Big(\mathcal{N}\Big(\mathbf{z}_{\mathrm{n},:}\, \mathbf{W}^{\mathrm{T}}, (\tau)^{-1} I\Big)\Big) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} \left(\frac{1}{2} \ln|\tau| - \frac{\tau}{2} \big(\mathrm{x}_{\mathrm{n,d}} - \mathbf{z}_{\mathrm{n},:}\, \mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)^2 \right) + \text{const} \\
&= \frac{DN}{2} \ln(\tau) - \frac{\tau}{2} \sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} \left({\mathrm{x}_{\mathrm{n,d}}}^2 - 2\, \mathbf{w}_{\mathrm{d},:}\, \mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\, \mathrm{x}_{\mathrm{n,d}} + \big(\mathbf{z}_{\mathrm{n},:}\, \mathbf{w}_{\mathrm{d},:}^{\mathrm{T}}\big)^2 \right) + \text{const} \\
&= \frac{DN}{2} \ln(\tau) - \frac{\tau}{2} \left(\sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2 \sum_{\mathrm{d}=1}^{\mathrm{D}} \mathbf{w}_{\mathrm{d},:}\, \mathbf{Z}^{\mathrm{T}}\, \mathbf{x}_{:,\mathrm{d}} + \sum_{\mathrm{d}=1}^{\mathrm{D}} \mathbf{w}_{\mathrm{d},:}\, \mathbf{Z}^{\mathrm{T}}\, \mathbf{Z}\, \mathbf{w}_{\mathrm{d},:}^{\mathrm{T}} \right) + \text{const} \\
&= \frac{DN}{2} \ln(\tau) - \frac{\tau}{2} \left(\sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\, \mathrm{Tr}\big\{ \mathbf{W}\, \mathbf{Z}^{\mathrm{T}}\, \mathbf{X} \big\} + \mathrm{Tr}\big\{ \mathbf{W}\, \mathbf{Z}^{\mathrm{T}}\, \mathbf{Z}\, \mathbf{W}^{\mathrm{T}} \big\} \right) + \text{const}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\mathbf{W},\mathbf{Z}}[\ln(p(\mathbf{X} \mid \mathbf{W}, \mathbf{Z}, \tau))] = \frac{DN}{2} \ln(\tau) - \frac{\tau}{2} \Bigg(&\sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 - 2\, \mathrm{Tr}\big\{ \langle \mathbf{W} \rangle \langle \mathbf{Z}^{\mathrm{T}} \rangle\, \mathbf{X} \big\} \\
&+ \mathrm{Tr}\big\{ \langle \mathbf{W}^{\mathrm{T}}\, \mathbf{W} \rangle \langle \mathbf{Z}^{\mathrm{T}}\, \mathbf{Z} \rangle \big\} \Bigg) + \text{const}
\end{aligned}$$

The second term could subsequently be determined as:

$$\mathbb{E}[\ln(p(\tau))] = \ln(p(\tau)) = -\beta_0^{\tau} \tau + (\alpha_0^{\tau} - 1) \ln(\tau) + \text{const}$$

Joining them together, we have that:

$$\begin{aligned}
\ln(q^*(\tau)) = \frac{DN}{2} \ln(\tau) - \frac{\tau}{2} \Bigg(&\sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} {\mathrm{x}_{\mathrm{n,d}}}^2 \\
&- 2\, \mathrm{Tr}\big\{ \langle \mathbf{W} \rangle \langle \mathbf{Z}^{\mathrm{T}} \rangle\, \mathbf{X} \big\} + \mathrm{Tr}\big\{ \langle \mathbf{W}^{\mathrm{T}}\, \mathbf{W} \rangle \langle \mathbf{Z}^{\mathrm{T}}\, \mathbf{Z} \rangle \big\} \Bigg) \\
&- \beta_0^{\tau} \tau + (\alpha_0^{\tau} - 1) \ln(\tau) + \text{const}
\end{aligned}$$

So by identifying terms with the distribution, we would have that for this parameter:

$$q^*(\tau) \;=\; (Gamma(\tau|a_\tau, b_\tau)) \tag{67}$$

the variance would be:

$$a_\tau \;=\; \frac{DN}{2} + \alpha_0^\tau \tag{68}$$

and the mean could be expressed as:

$$b_\tau \;=\; \beta_0^\tau + \frac{1}{2}\left(\sum_{\mathrm{n=1}}^{\mathrm{N}}\sum_{\mathrm{d=1}}^{\mathrm{D}} \mathrm{x_{n,d}}^2 - 2\,\mathrm{Tr}\big\{\langle \mathbf{W}\rangle\langle \mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}\big\} + \mathrm{Tr}\big\{\langle \mathbf{W}^{\mathrm{T}}\,\mathbf{W}\rangle\langle \mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\big\}\right) \tag{69}$$

3 Bayesian CCA or Bayesian Inter-Battery Factor Analysis (BIBFA)

3.1 Generative model

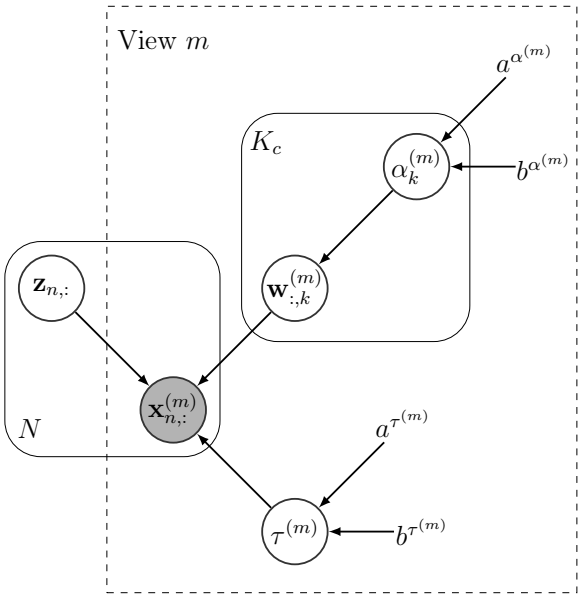

Figure 3: Graphic model of the CCA scheme for neuromarkers design.

In our case we have two views. The first one will be the input data ($\mathbf{X}^{(1)} \in \mathbb{R}^{N\times d}$) and the output labels ($\mathbf{X}^{(2)} \in \mathbb{R}^{N\times c}$). These two views share some latent variables ($Z \in \mathbb{R}^{N\times K}$) and have some specific ones ($Z^{(1)} \in \mathbb{R}^{N\times K_1}$ and $Z^{(2)} \in \mathbb{R}^{N\times K_2}$). This way, we have projection matrices of those shared latent variables for each view ($A^{(1)} \in \mathbb{R}^{d\times K}$ and $A^{(2)} \in \mathbb{R}^{d\times K}$) as well as other projection matrices for the specific latent variables ($B^{(1)} \in \mathbb{R}^{d\times K_1}$ and $B^{(2)} \in \mathbb{R}^{d\times K_2}$). Finally, we have the noise, which includes all we weren't able to model (the covariance matrices have the shape $\Sigma^{(1)} \in \mathbb{R}^{d\times d}$ and $\Sigma^{(2)} \in \mathbb{R}^{c\times c}$).

$$K_c = K + K_1 + K_2 \tag{70}$$

$$\left.\begin{array}{l}
\left.\begin{array}{l}\mathbf{X}^{(1)} \in \mathbb{R}^{N\times d}\\ \mathbf{X}^{(2)} \in \mathbb{R}^{N\times c}\end{array}\right\} \quad \rightarrow \quad X_n = [\mathbf{x}^{(1)}_{\mathrm{n},:}, \mathbf{x}^{(2)}_{\mathrm{n},:}] \in \mathbb{R}^{1\times(d+c)} \\
\left.\begin{array}{l}Z^{(com)} \in \mathbb{R}^{N\times K}\\ Z^{(1)} \in \mathbb{R}^{N\times K_1}\\ Z^{(2)} \in \mathbb{R}^{N\times K_2}\end{array}\right\} \quad \rightarrow \quad \mathbf{z}_{\mathrm{n},:} = [Z^{(com)}_n, Z^{(1)}_n, Z^{(2)}_n] \in \mathbb{R}^{1\times K_c} \\
\left.\begin{array}{l}A^{(1)} \in \mathbb{R}^{d\times K}\\ A^{(2)} \in \mathbb{R}^{c\times K}\\ B^{(1)} \in \mathbb{R}^{d\times K_1}\\ B^{(2)} \in \mathbb{R}^{c\times K_2}\end{array}\right\} \quad \rightarrow \quad W = \begin{bmatrix} A^{(1)} & B^{(1)} & 0 \\ A^{(2)} & 0 & B^{(2)} \end{bmatrix} \in \mathbb{R}^{(d+c)\times K_c} \\
\left.\begin{array}{l}\Sigma^{(1)} \in \mathbb{R}^{d\times d}\\ \Sigma^{(2)} \in \mathbb{R}^{c\times c}\end{array}\right\} \quad \rightarrow \quad \Sigma = \begin{bmatrix} \Sigma^{(1)} & 0 \\ 0 & \Sigma^{(2)} \end{bmatrix} \in \mathbb{R}^{(d+c)\times(d+c)}
\end{array}\right\} \quad \rightarrow \quad X_n = \mathbf{z}_{\mathrm{n},:} W^T + \epsilon \tag{71}$$

Let's assume that

$$\mathbf{z}_{\mathrm{n},:} \sim \mathcal{N}(0, I) \tag{72}$$

then, we have that

$$X_n | \, \mathbf{z}_{\mathrm{n},:} \sim \mathcal{N}(\mathbf{z}_{\mathrm{n},:} W^T, \Sigma) \tag{73}$$

$$\mathbf{x}^{(\mathrm{m})}_{\mathrm{n},:} \, | \, \mathbf{z}_{\mathrm{n},:} \sim \mathcal{N}(\mathbf{z}_{\mathrm{n},:} \, \mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}, \Sigma^{(m)}) \tag{74}$$

where $\mathbf{W}^{(\mathrm{m})}$ is a random variable defined as:

$$\mathbf{W}^{(\mathrm{m})} \sim ARD(\alpha_0, \beta_0) \tag{75}$$

If we force all columns to have the same prior while letting the rows have a different one this distribution can be written as:

$$\mathbf{w}^{(\mathrm{m})}_{:,\mathrm{k}} \sim \mathcal{N}\left(0, \left(\alpha^{(\mathrm{m})}_{\mathrm{k}}\right)^{-1} I\right) \tag{76}$$

This way, high values of α_k make the column tend to 0 (low variance). On the other hand, if α_k has a low value you allow the column to have any value (high variance).

$$\alpha^{(m)}_k \sim Gamma(\alpha_0, \beta_0) \tag{77}$$

where α_0 and β_0 are non-critical hyperparameters. This way we are capable of cancelling the components of W that we don't need. For this reason, the matrix W is learned disorganised. Intuitively, we have that:

- A 1 in the upper half of the matrix and a 0 in the lower $\rightarrow B^{(1)}$
- A 0 in the upper half of the matrix and a 1 in the lower $\rightarrow B^{(2)}$
- A 0 in the upper half of the matrix and a 0 in the lower $\rightarrow$ Noise
- A 1 in the upper half of the matrix and a 1 in the lower $\rightarrow A^{(1)}/A^{(2)}$

This can be achieved by forcing sparsity for the latent factors (K) when using ARD (Automatic Relevance Determination).

Conversely, the noise Σ can be modelled as:

$$\Sigma^{(m)} = \tau_m^{-1} I \tag{78}$$

where τ_m is, $\tau_m = Gamma(\alpha^\tau_0, \beta^\tau_0)$, being τ a model hyperparameter.

3.2 Variational inference

Making use of the variational inference definition previously stated, the problem can be defined to apply those techniques to design the model. First of all, considering Figure 3 we can extract the probability distribution of the model

$$p(\Theta|\,\mathbf{X}^{\{\mathcal{M}\}}) \;\approx\; \prod_{\mathrm{m}=1}^{\mathrm{M}}\left(q\Big(\mathbf{W}^{(\mathrm{m})}\Big)q\Big(\tau^{(\mathrm{m})}\Big)\prod_{\mathrm{k}=1}^{\mathrm{K_c}}q\Big(\alpha_{\mathrm{k}}^{(\mathrm{m})}\Big)\right)\prod_{n=1}^{N}q(\mathbf{z}_{\mathrm{n},:}) \tag{79}$$

This way, we can determine the optimum distribution of the different variables, applying Equation (8) and then calculate the lower bound defined in Equation (6).

In this section the different distributions specified in Equation (79) will be developed making use of the variational inference.

3.2.1 Distribution of $\mathbf{W}^{(\mathrm{m})}$

$$\begin{aligned}\ln\Big(q^*\Big(\mathbf{W}^{(\mathrm{m})}\Big)\Big) &= \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha}^{(\mathrm{m})},\tau^{(\mathrm{m})}}\left[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})},\mathbf{W}^{(\mathrm{m})},\mathbf{Z},\boldsymbol{\alpha}^{(\mathrm{m})},\tau^{(\mathrm{m})}\Big)\Big)\right]\\ &= \mathbb{E}_{\mathbf{Z},\tau^{(\mathrm{m})}}\left[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})}\,|\,\mathbf{W}^{(\mathrm{m})},\mathbf{Z},\tau^{(\mathrm{m})}\Big)\Big)\right]\\ &\quad+\mathbb{E}_{\boldsymbol{\alpha}^{(\mathrm{m})}}\left[\ln\Big(p\Big(\mathbf{W}^{(\mathrm{m})}\,|\,\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\right]+\text{const}\end{aligned} \tag{80}$$

We can now evaluate both terms independently and then sum the results:

$$\begin{aligned}p(\Theta,\tilde{\mathbf{T}}^{\{\mathcal{M}_\mathrm{t}\}},\tilde{\mathbf{X}}^{\{\mathcal{M}_\mathrm{r}\}}|\,\mathbf{T}^{\{\mathcal{M}_\mathrm{t}\}},\mathbf{X}^{\{\mathcal{M}_\mathrm{r}\}}) \;\approx&\\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\ln\left(\mathcal{N}\left(\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}},\Big(\tau^{(\mathrm{m})}\Big)^{-1}I\right)\right)+\text{const}\\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(-\frac{1}{2}\ln\left|\Big(\tau^{(\mathrm{m})}\Big)^{-1}I\right|\right.\\ &\quad\left.-\frac{1}{2}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}-\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\Big)\,\tau^{(\mathrm{m})}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}-\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\Big)^{T}\right)+\text{const}\\ &= -\frac{\tau^{(\mathrm{m})}}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})^{\mathrm{T}}}\\ &\quad-2\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}+\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Big)+\text{const}\\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\tau^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}-\frac{\tau^{(\mathrm{m})}}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\right)+\text{const}\\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\tau^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}-\frac{\tau^{(\mathrm{m})}}{2}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\Big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Big)\right)+\text{const}\\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\tau^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}-\frac{\tau^{(\mathrm{m})}}{2}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\Big(\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}}\Big)\right)+\text{const}\end{aligned} \tag{81}$$

This way the expectation can be calculated as:

$$\mathbb{E}_{\mathbf{Z},\tau^{(\mathrm{m})}}\left[\ln\left(p\left(\mathbf{X}^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\right)\right)\right] = \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\langle\tau^{(\mathrm{m})}\rangle\, \mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\, \mathbf{W}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}\rangle - \frac{\langle\tau^{(\mathrm{m})}\rangle}{2} \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\left(\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\, \mathbf{z}_{\mathrm{n},:}\rangle\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}}\right)\right) \tag{82}$$

Equivalently, the second term would be:

$$\begin{aligned}\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right) &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}} \ln\left(p\left(\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})} \mid \alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right) \\ &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{1}{2}\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \frac{\alpha_{\mathrm{k}}^{(\mathrm{m})}}{2}{\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})}}^2\right) + \text{const}\end{aligned} \tag{83}$$

$$\mathbb{E}_{\boldsymbol{\alpha}^{(\mathrm{m})}}\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] = \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(-\frac{\langle\alpha_{\mathrm{k}}^{(\mathrm{m})}\rangle}{2}{\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})}}^2\right) + \text{const} \tag{84}$$

Finally, joining Equations (82) and (84), we would have that the optimum solution for the variable $\mathbf{W}^{(\mathrm{m})}$ is:

$$\begin{aligned}\ln\left(q^*\left(\mathbf{W}^{(\mathrm{m})}\right)\right) &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(-\frac{\langle\alpha_{\mathrm{k}}^{(\mathrm{m})}\rangle}{2}{\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})}}^2\right) + \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\langle\tau^{(\mathrm{m})}\rangle\, \mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\, \mathbf{W}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}\rangle \right. \\ &\quad \left. - \frac{\langle\tau^{(\mathrm{m})}\rangle}{2} \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\left(\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\, \mathbf{z}_{\mathrm{n},:}\rangle\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}}\right)\right) + \text{const} \\ &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\left(-\frac{1}{2}\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}\, diag(\langle\alpha_{\mathrm{k}}^{(\mathrm{m})}\rangle)\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}} + \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\langle\tau^{(\mathrm{m})}\rangle\, \mathrm{x}_{\mathrm{n,d}}^{(\mathrm{m})}\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}\rangle \right. \right. \\ &\quad \left. \left. - \frac{\langle\tau^{(\mathrm{m})}\rangle}{2}\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \langle \mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\, \mathbf{z}_{\mathrm{n},:}\rangle\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}}\right)\right) + \text{const} \\ &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}(-\frac{1}{2}\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}\left(diag(\langle\alpha_{\mathrm{k}}^{(\mathrm{m})}\rangle) + \langle\tau^{(\mathrm{m})}\rangle\langle \mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\, \mathbf{z}_{\mathrm{n},:}\rangle\right) \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})^{\mathrm{T}}} \\ &\quad + \langle\tau^{(\mathrm{m})}\rangle\, \mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \langle\mathbf{Z}\rangle^{T}\, \mathbf{x}_{:,\mathrm{d}}^{(\mathrm{m})^{\mathrm{T}}}) + \text{const}\end{aligned} \tag{85}$$

This way, comparing the results with the normal distribution, we can identify terms, extracting the following conclusions:

$$q^*\left(\mathbf{W}^{(\mathrm{m})}\right) = \prod_{\mathrm{d}=1}^{\mathrm{D_m}}\left(\mathcal{N}\left(\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})} \mid \mu_{\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}}, \Sigma_{\mathbf{W}^{(\mathrm{m})}}\right)\right) \tag{86}$$

Where the variance would be:

$$\Sigma_{\mathbf{W}^{(\mathrm{m})}}^{-1} = diag(\langle\boldsymbol{\alpha}^{(\mathrm{m})}\rangle) + \langle\tau^{(\mathrm{m})}\rangle\langle\mathbf{Z}^{\mathrm{T}}\, \mathbf{Z}\rangle \tag{87}$$

And the mean could be expressed as:

$$\begin{aligned}\mu_{\mathbf{w}_{\mathrm{d},:}^{(\mathrm{m})}} &= \langle\tau^{(\mathrm{m})}\rangle \Sigma_{\mathbf{W}^{(\mathrm{m})}} \langle\mathbf{Z}\rangle^{T}\, \mathbf{x}_{:,\mathrm{d}}^{(\mathrm{m})^{\mathrm{T}}} \\ \langle\mathbf{W}^{(\mathrm{m})}\rangle &= \langle\tau^{(\mathrm{m})}\rangle\, \mathbf{X}^{(\mathrm{m})^{\mathrm{T}}} \langle\mathbf{Z}\rangle \Sigma_{\mathbf{W}^{(\mathrm{m})}}\end{aligned} \tag{88}$$

3.2.2 Distribution of Z

$$\begin{aligned}\ln(q^*(\mathbf{Z})) &= \mathbb{E}_{\mathbf{W}^{(\mathrm{m})},\tau^{(\mathrm{m})}}\left[\ln\left(p\left(\mathbf{X}^{(\mathrm{m})},\mathbf{W}^{(\mathrm{m})},\mathbf{Z},\tau^{(\mathrm{m})}\right)\right)\right] \\ &= \mathbb{E}_{\mathbf{W}^{(1)},\tau^{(1)}}\left[\ln\left(p\left(\mathbf{X}^{(1)}\mid\mathbf{W}^{(1)},\mathbf{Z},\tau^{(1)}\right)\right)\right]+\mathbb{E}_{\mathbf{W}^{(2)},\tau^{(2)}}\left[\ln\left(p\left(\mathbf{X}^{(2)}\mid\mathbf{W}^{(2)},\mathbf{Z},\tau^{(2)}\right)\right)\right] \\ &\quad+\mathbb{E}[\ln(p(\mathbf{Z}))] \end{aligned} \tag{89}$$

The first two terms are similar to equation (82), changing the expectation dependency:

$$\begin{aligned}\mathbb{E}_{\mathbf{W}^{(\mathrm{m})},\tau^{(\mathrm{m})}}\left[\ln\left(p\left(\mathbf{X}^{(\mathrm{m})}\mid\mathbf{W}^{(\mathrm{m})},\mathbf{Z},\tau^{(\mathrm{m})}\right)\right)\right] &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\Bigg(\langle\tau^{(\mathrm{m})}\rangle\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\langle\mathbf{W}^{(\mathrm{m})}\rangle\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} \\ &\quad-\frac{\langle\tau^{(\mathrm{m})}\rangle}{2}\,\mathbf{z}_{\mathrm{n},:}\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Bigg)+\text{const}\end{aligned} \tag{90}$$

The third term could be calculated as:

$$\ln(p(\mathbf{Z})) = \sum_{\mathrm{n}=1}^{\mathrm{N}}(\mathcal{N}(0,I)) = \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(-\frac{1}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\right)+\text{const}$$

$$\mathbb{E}[\ln(p(\mathbf{Z}))] = \ln(p(\mathbf{Z})) = \sum_{\mathrm{n}=1}^{\mathrm{N}}\left(-\frac{1}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\right)+\text{const} \tag{91}$$

Therefore, joining the different terms in Equations (90) and (91) , the optimum solution has the form:

$$\begin{aligned}\ln(q^*(\mathbf{Z})) &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\Bigg(-\frac{1}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}+\sum_{\mathrm{m}=1}^{\mathrm{M}}\Bigg(\langle\tau^{(\mathrm{m})}\rangle\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\langle\mathbf{W}^{(\mathrm{m})}\rangle\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} \\ &\quad-\frac{\langle\tau^{(\mathrm{m})}\rangle}{2}\,\mathbf{z}_{\mathrm{n},:}\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Bigg)\Bigg)+\text{const}\end{aligned} \tag{92}$$

This way, we can now identify terms, having that:

$$q^*(\mathbf{Z}) = \prod_{n=1}^{N}\left(\mathcal{N}\left(\mathbf{z}_{\mathrm{n},:}\,|\mu_{\mathbf{z}_{\mathrm{n},:}},\Sigma_{\mathbf{Z}}\right)\right) \tag{93}$$

Where the variance would be:

$$\Sigma_{\mathbf{Z}}^{-1} = I+\sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\langle\tau^{(\mathrm{m})}\rangle\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\right) \tag{94}$$

And the mean could be expressed as:

$$\begin{aligned}\mu_{\mathbf{z}_{\mathrm{n},:}} &= \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\langle\tau^{(\mathrm{m})}\rangle\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\langle\mathbf{W}^{(\mathrm{m})}\rangle\Sigma_{\mathbf{Z}}\right) \\ \langle\mathbf{Z}\rangle &= \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\langle\tau^{(\mathrm{m})}\rangle\,\mathbf{X}^{(\mathrm{m})}\langle\mathbf{W}^{(\mathrm{m})}\rangle\Sigma_{\mathbf{Z}}\right)\end{aligned} \tag{95}$$

3.2.3 Distribution of $\boldsymbol{\alpha}^{(\mathrm{m})}$

$$\ln\left(q^*\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right) = \mathbb{E}_{\mathbf{W}^{(\mathrm{m})}}\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] + \mathbb{E}\left[\ln\left(p\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] \tag{96}$$

Similarly to what we have done before, we can develop both terms independently:

$$\begin{aligned}\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right) &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{1}{2}\ln\left|\alpha_{\mathrm{k}}^{(\mathrm{m})}\right| - \frac{1}{2}\alpha_{\mathrm{k}}^{(\mathrm{m})}\left(\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})}\right)^2\right) + \text{const} \\ &= \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{D_m}{2}\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \frac{1}{2}\alpha_{\mathrm{k}}^{(\mathrm{m})}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\right) + \text{const}\end{aligned}$$

$$\mathbb{E}\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] = \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{D_m}{2}\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \frac{1}{2}\alpha_{\mathrm{k}}^{(\mathrm{m})}\langle\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\rangle\right) + \text{const} \tag{97}$$

The second term is deployed as follows:

$$\ln\left(p\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right) = -\beta_0\,\alpha_{\mathrm{k}}^{(\mathrm{m})} + (\alpha_0 - 1)\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) + \text{const}$$

$$\mathbb{E}\left[\ln\left(p\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] = \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\ln\left(p\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right)\right) = \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(-\beta_0\,\alpha_{\mathrm{k}}^{(\mathrm{m})} + (\alpha_0 - 1)\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right) + \text{const} \tag{98}$$

Therefore, by joining both Equations (97) and (98) together, we can now determine the value f the expectation we were looking for

$$\begin{aligned}\ln\left(q^*\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right) = \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\Bigg(&\frac{D_m}{2}\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \frac{1}{2}\alpha_{\mathrm{k}}^{(\mathrm{m})}\langle\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\rangle \\ &-\beta_0\,\alpha_{\mathrm{k}}^{(\mathrm{m})} + (\alpha_0 - 1)\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\Bigg) + \text{const}\end{aligned} \tag{99}$$

This way, we can now identify terms, having that:

$$q^*\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right) = \prod_{\mathrm{k}=1}^{\mathrm{K_c}}\left(Gamma\left(\alpha_{\mathrm{k}}^{(\mathrm{m})} \mid a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}, b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right) \tag{100}$$

where the variance would be:

$$a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}} = \frac{D_m}{2} + \alpha_0 \tag{101}$$

and the mean could be expressed as:

$$b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}} = \beta_0 + \frac{1}{2}\langle\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\rangle = \beta_0 + \frac{1}{2}\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle_{k,k} \tag{102}$$

where

$$\langle\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\rangle = \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\left(\mathrm{w}_{\mathrm{d,k}}^{(\mathrm{m})}\right)^2 = \langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle_{k,k}$$

3.2.4 Distribution of $\tau^{(\mathrm{m})}$

$$\ln\Big(q^*\Big(\tau^{(\mathrm{m})}\Big)\Big) \;=\; \mathbb{E}_{\mathbf{W}^{(\mathrm{m})},\mathbf{Z}}\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})} \,|\, \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\Big)\Big)\Big] + \mathbb{E}\Big[\ln\Big(p\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big] \tag{103}$$

Similarly to what was done in equation (82):

$$\begin{aligned}
\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})} \,|\, \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\Big)\Big) &= \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln\Big(\mathcal{N}\Big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}, \Big(\tau^{(\mathrm{m})}\Big)^{-1} I\Big)\Big) + \text{const} \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}} \Bigg(\frac{1}{2}\ln\Big|\tau^{(\mathrm{m})}\Big| \\
&\quad -\frac{\tau^{(\mathrm{m})}}{2}\Big(\mathrm{x}^{(\mathrm{m})}_{\mathrm{n,d}} - \mathbf{z}_{\mathrm{n},:}\,\mathbf{w}^{(\mathrm{m})^{\mathrm{T}}}_{\mathrm{d},:}\Big)^2\Bigg) + \text{const} \\
&= \frac{D_m N}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) - \frac{\tau^{(\mathrm{m})}}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\Bigg(\mathrm{x}^{(\mathrm{m})^2}_{\mathrm{n,d}} \\
&\quad -2\,\mathbf{w}^{(\mathrm{m})}_{\mathrm{d},:}\,\mathbf{z}^{\mathrm{T}}_{\mathrm{n},:}\,\mathrm{x}^{(\mathrm{m})}_{\mathrm{n,d}} + \Big(\mathbf{z}_{\mathrm{n},:}\,\mathbf{w}^{(\mathrm{m})^{\mathrm{T}}}_{\mathrm{d},:}\Big)^2\Bigg) + \text{const} \\
&= \frac{D_m N}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) - \frac{\tau^{(\mathrm{m})}}{2}\Bigg(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathrm{x}^{(\mathrm{m})^2}_{\mathrm{n,d}} \\
&\quad -2\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathbf{w}^{(\mathrm{m})}_{\mathrm{d},:}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{x}^{(\mathrm{m})}_{:,\mathrm{d}} + \sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathbf{w}^{(\mathrm{m})}_{\mathrm{d},:}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\,\mathbf{w}^{(\mathrm{m})^{\mathrm{T}}}_{\mathrm{d},:}\Bigg) + \text{const} \\
&= \frac{D_m N}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) - \frac{\tau^{(\mathrm{m})}}{2}\Bigg(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathrm{x}^{(\mathrm{m})^2}_{\mathrm{n,d}} \\
&\quad -2\,\mathrm{Tr}\Big\{\mathbf{W}^{(\mathrm{m})}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{X}^{(\mathrm{m})}\Big\} + \mathrm{Tr}\Big\{\mathbf{W}^{(\mathrm{m})}\,\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\Big\}\Bigg) + \text{const}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\mathbf{W}^{(\mathrm{m})},\mathbf{Z}}\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})} \,|\, \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\Big)\Big)\Big] &= \frac{D_m N}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) - \frac{\tau^{(\mathrm{m})}}{2}\Bigg(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathrm{x}^{(\mathrm{m})^2}_{\mathrm{n,d}} \\
&\quad -2\,\mathrm{Tr}\Big\{\langle\mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}^{(\mathrm{m})}\Big\} \\
&\quad +\mathrm{Tr}\Big\{\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\Big\}\Bigg) + \text{const}
\end{aligned}$$

The second term could subsequently be determined as:

$$\mathbb{E}\Big[\ln\Big(p\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big] \;=\; \ln\Big(p\Big(\tau^{(\mathrm{m})}\Big)\Big) = -\beta_0^{\tau}\,\tau^{(\mathrm{m})} + (\alpha_0^{\tau}-1)\ln\Big(\tau^{(\mathrm{m})}\Big) + \text{const}$$

Joining them together, we have that:

$$\begin{aligned}
\ln\Big(q^*\Big(\tau^{(\mathrm{m})}\Big)\Big) &= \frac{D_m N}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) - \frac{\tau^{(\mathrm{m})}}{2}\Bigg(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}}\mathrm{x}^{(\mathrm{m})^2}_{\mathrm{n,d}} \\
&\quad -2\,\mathrm{Tr}\Big\{\langle\mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}^{(\mathrm{m})}\Big\} + \mathrm{Tr}\Big\{\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\Big\}\Bigg) \\
&\quad -\beta_0^{\tau}\,\tau^{(\mathrm{m})} + (\alpha_0^{\tau}-1)\ln\Big(\tau^{(\mathrm{m})}\Big) + \text{const}
\end{aligned}$$

So by identifying terms with the distribution, we would have that for this parameter:

$$q^*\Big(\tau^{(\mathrm{m})}\Big) \;=\; \Big(Gamma\Big(\tau^{(\mathrm{m})}\,|a_{\tau^{(\mathrm{m})}}, b_{\tau^{(\mathrm{m})}}\Big)\Big) \tag{104}$$

the variance would be:

$$a_{\tau^{(\mathrm{m})}} \;=\; \frac{D_m N}{2} + \alpha_0^{\tau} \tag{105}$$

and the mean could be expressed as:

$$b_{\tau^{(\mathrm{m})}} \;=\; \beta_0^{\tau} + \frac{1}{2}\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{d}=1}^{\mathrm{D_m}} \mathrm{x}_{\mathrm{n,d}}^{(\mathrm{m})^2} - 2\,\mathrm{Tr}\Big\{\langle \mathbf{W}^{(\mathrm{m})}\rangle\langle \mathbf{Z}^{\mathrm{T}}\rangle\,\mathbf{X}^{(\mathrm{m})}\Big\} + \mathrm{Tr}\Big\{\langle \mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\rangle\langle \mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\Big\}\right) \tag{106}$$

3.3 Update of the lower bound ($L(q)$)

Once the optimisation of the different parameters has been determined, we can calculate the changes on the lower bound as:

$$\begin{aligned} L_q \;&=\; -\int q(\Theta)\ln\left(\frac{q(\Theta)}{p(X,\Theta)}\right)d\Theta = \int q(\Theta)\ln(p(X,\Theta))d\Theta - \int q(\Theta)\ln(q(\Theta))d\Theta \\ &=\; \mathbb{E}_q[\ln(p(X,\Theta))] - \mathbb{E}_q[\ln(q(\Theta))] \end{aligned} \tag{107}$$

To facilitate the analysis of the different parts of these equations we will evaluate in the following subsections the terms related to $\mathbb{E}_q[\ln(p(X,\Theta))]$ and to the entropy.

3.4 Terms associated to $\mathbb{E}_q[\ln(p(X,\Theta))]$

This first term of the lower bound would be composed by the following terms:

$$\begin{aligned} \mathbb{E}_q[\ln(p(X,\Theta))] \;=\; &\mathbb{E}_q[\ln(p(\mathbf{Z}))] + \sum_{\mathrm{m}=1}^{\mathrm{M}}\Big(\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{W}^{(\mathrm{m})}\,|\,\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] + \mathbb{E}_q\Big[\ln\Big(p\Big(\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] \\ &+\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})}\,|\,\mathbf{W}^{(\mathrm{m})},\mathbf{Z},\tau^{(\mathrm{m})}\Big)\Big)\Big] + \mathbb{E}_q\Big[\ln\Big(p\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big]\Big) \end{aligned} \tag{108}$$

Which, as we have done before, can be independently analysed. To facilitate the extraction of results, these new terms will be included in different subsubsections.

$\mathbb{E}_q[\ln(p(\mathbf{Z}))]$

$$\mathbb{E}_q[\ln(p(\mathbf{Z}))] \;=\; \sum_{\mathrm{n}=1}^{\mathrm{N}}\mathbb{E}_q\left[-\frac{K_c}{2}\ln(2\pi) - \frac{1}{2}\,\mathbf{z}_{\mathrm{n},:}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\right] = -\frac{NK_c}{2}\ln(2\pi) - \frac{1}{2}\,\mathrm{Tr}\{\langle \mathbf{Z}\,\mathbf{Z}^{\mathrm{T}}\rangle\} \tag{109}$$

$\mathbb{E}_q\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right]$

$$\begin{aligned}\mathbb{E}_q\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] &= \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \mathbb{E}_q\left[-\frac{D_m}{2}\ln(2\pi) + \frac{D_m}{2}\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \frac{1}{2}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\alpha_{\mathrm{k}}^{(\mathrm{m})}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\right] \\ &= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_{\mathrm{k}=1}^{\mathrm{K_c}} \mathbb{E}_q\left[\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right] \\ &\quad - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\mathbb{E}_q\left[\alpha_{\mathrm{k}}^{(\mathrm{m})}\right]\mathbb{E}_q\left[\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\right]\right) \end{aligned} \tag{110}$$

We can now calculate the expectation of the different variables independently to determine the value of the equation, having that:

$$\mathbb{E}_{q(\alpha)}\left[\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right] = \psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) \tag{111}$$

$$\mathbb{E}_{q(\alpha)}\left[\alpha_{\mathrm{k}}^{(\mathrm{m})}\right] = \frac{a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}{b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}} \tag{112}$$

$$\mathbb{E}_{q(W)}\left[\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\right] = \langle \mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})^{\mathrm{T}}}, \mathbf{w}_{:,\mathrm{k}}^{(\mathrm{m})}\rangle = \langle \mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}, \mathbf{W}^{(\mathrm{m})}\rangle_{k,k} \tag{113}$$

We can equate now this to Equation (102) and substituting all these equations on the previous one we have that:

$$\begin{aligned}\mathbb{E}_q\left[\ln\left(p\left(\mathbf{W}^{(\mathrm{m})} \mid \boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] &= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right) \\ &\quad - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}{b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}\left(2\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}} - \beta_0\right)\right)\right) \\ &= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right) \\ &\quad - \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) + \beta_0\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\frac{a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}{b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}\right) \end{aligned} \tag{114}$$

$\mathbb{E}_q\left[\ln\left(p\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right]$

$$\begin{aligned}\mathbb{E}_q\left[\ln\left(p\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] &= \sum_{\mathrm{k}=1}^{\mathrm{K_c}} \mathbb{E}_q\left[\ln(\beta_0) - \beta_0\,\alpha_{\mathrm{k}}^{(\mathrm{m})} + (\alpha_0 - 1)\ln(\beta_0) + (\alpha_0 - 1)\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right) - \ln(\Gamma(\alpha_0))\right] \\ &= \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(-\beta_0\mathbb{E}_q\left[\alpha_{\mathrm{k}}^{(\mathrm{m})}\right] + \alpha_0\ln(\beta_0) + (\alpha_0 - 1)\mathbb{E}_q\left[\ln\left(\alpha_{\mathrm{k}}^{(\mathrm{m})}\right)\right] - \ln(\Gamma(\alpha_0))\right) \end{aligned}$$

Using the expectations determined in Equations (111) and (112) we have:

$$\begin{aligned}\mathbb{E}_q\left[\ln\left(p\left(\boldsymbol{\alpha}^{(\mathrm{m})}\right)\right)\right] &= K_c(\alpha_0\ln(\beta_0) - \ln(\Gamma(\alpha_0))) \\ &\quad + \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(-\beta_0\frac{a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}}{b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}} + (\alpha_0 - 1)\left(\psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right)\right) \end{aligned} \tag{115}$$

$\mathbb{E}_q\big[\ln\big(p\big(\mathbf{W}^{(\mathrm{m})}, \boldsymbol{\alpha}^{(\mathrm{m})}\big)\big)\big]$

We can now calculate the joint distribution of the two previous variables by combining Equations (115) and (114):

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{W}^{(\mathrm{m})}, \boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] &= \Big(\frac{D_m}{2} + \alpha_0 - 1\Big)\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\Big(\psi\Big(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big) - \ln\Big(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big)\Big) \\ &\quad - \frac{K_c D_m}{2}\ln(2\pi) + K_c(\alpha_0\ln(\beta_0) - \ln(\Gamma(\alpha_0))) - \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\Big(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big)\end{aligned} \tag{116}$$

$\mathbb{E}_q\big[\ln\big(p\big(\mathbf{X}^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\big)\big)\big]$

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\Big)\Big)\Big] &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\mathbb{E}_q\Big[-\frac{D_m}{2}\ln(2\pi) + \frac{D_m}{2}\ln\Big(\tau^{(\mathrm{m})}\Big) \\ &\quad -\frac{1}{2}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})} - \mathbf{Z}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\Big)\,\tau^{(\mathrm{m})}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})} - \mathbf{Z}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\Big)^{T}\Big] \\ &= \sum_{\mathrm{n}=1}^{\mathrm{N}}\Big(-\frac{D_m}{2}\ln(2\pi) + \frac{D_m}{2}\mathbb{E}_q\Big[\ln\Big(\tau^{(\mathrm{m})}\Big)\Big] \\ &\quad -\frac{1}{2}\mathbb{E}_q\Big[\tau^{(\mathrm{m})}\Big]\mathbb{E}_q\Big[\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})^{\mathrm{T}}} \\ &\quad -2\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}} + \mathbf{z}_{\mathrm{n},:}\,\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}\,\mathbf{W}^{(\mathrm{m})}\,\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\Big]\Big) \\ &= -\frac{N D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\Big(\mathbb{E}_q\Big[\ln\Big(\tau^{(\mathrm{m})}\Big)\Big]\Big) \\ &\quad -\frac{1}{2}\mathbb{E}_q\Big[\tau^{(\mathrm{m})}\Big]\sum_{\mathrm{n}=1}^{\mathrm{N}}\Big(\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\,\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})^{\mathrm{T}}} \\ &\quad -2\,\mathrm{Tr}\Big\{\mathbf{x}_{\mathrm{n},:}^{(\mathrm{m})}\langle\mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}\rangle\Big\} + \mathrm{Tr}\Big\{\langle\mathbf{W}^{(\mathrm{m})^{\mathrm{T}}}, \mathbf{W}^{(\mathrm{m})}\rangle\langle\mathbf{z}_{\mathrm{n},:}^{\mathrm{T}}, \mathbf{z}_{\mathrm{n},:}\rangle\Big\}\Big)\end{aligned} \tag{117}$$

If we compare and substitute the value obtained in Equation (106):

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}, \tau^{(\mathrm{m})}\Big)\Big)\Big] &= -\frac{N D_m}{2}\ln(2\pi) + \frac{D_m}{2}(\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}})) - \frac{a_{\tau^{(\mathrm{m})}}}{b_{\tau^{(\mathrm{m})}}}(b_{\tau^{(\mathrm{m})}} - \beta_0^{\tau}) \\ &= -\frac{N D_m}{2}\ln(2\pi) + \frac{D_m}{2}(\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}})) - a_{\tau^{(\mathrm{m})}} + \frac{a_{\tau^{(\mathrm{m})}}}{b_{\tau^{(\mathrm{m})}}}\beta_0^{\tau}\end{aligned} \tag{118}$$

$\mathbb{E}_q\big[\ln\big(p\big(\tau^{(\mathrm{m})}\big)\big)\big]$

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(p\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big] &= \mathbb{E}_q\Big[\ln(\beta_0^\tau) - \beta_0^\tau\,\tau^{(\mathrm{m})} + (\alpha_0^\tau - 1)\ln(\beta_0^\tau) + (\alpha_0^\tau - 1)\ln\Big(\tau^{(\mathrm{m})}\Big) - \ln(\Gamma(\alpha_0^\tau))\Big] \\ &= -\beta_0^\tau \mathbb{E}_q\Big[\tau^{(\mathrm{m})}\Big] + \alpha_0^\tau \ln(\beta_0^\tau) + (\alpha_0^\tau - 1)\mathbb{E}_q\Big[\ln\Big(\tau^{(\mathrm{m})}\Big)\Big] - \ln(\Gamma(\alpha_0^\tau)) \\ &= \alpha_0^\tau \ln(\beta_0^\tau) - \ln(\Gamma(\alpha_0^\tau)) - \beta_0^\tau \frac{a_{\tau^{(\mathrm{m})}}}{b_{\tau^{(\mathrm{m})}}} + (\alpha_0^\tau - 1)(\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}}))\end{aligned} \tag{119}$$

$\mathbb{E}_q\big[\ln\big(p\big(\mathbf{X}^{(\mathrm{m})}, \tau^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}\big)\big)\big]$

In a similar way to what was done before, we can now compute the joint distribution of $\mathbf{X}^{(\mathrm{m})}$ and $\tau^{(\mathrm{m})}$ by combining Equations (119) and (118) we have:

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})}, \tau^{(\mathrm{m})} \mid \mathbf{W}^{(\mathrm{m})}, \mathbf{Z}\Big)\Big)\Big] &= -\frac{ND_m}{2}\ln(2\pi) - a_{\tau^{(\mathrm{m})}} + \alpha_0^\tau \ln(\beta_0^\tau) - \ln(\Gamma(\alpha_0^\tau)) + \\ &\quad \Big(\frac{D_m}{2} + \alpha_0^\tau - 1\Big)(\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}}))\end{aligned} \tag{120}$$

3.5 Terms of entropy, $\mathbb{E}_q[\ln(q(\Theta))]$

This first term of the lower bound would be composed by the following terms:

$$\begin{aligned}\mathbb{E}_q[\ln(q(\Theta))] &= \mathbb{E}_q[\ln(q(\mathbf{Z}))] + \sum_{\mathrm{m}=1}^{\mathrm{M}} \Big(\mathbb{E}_q\Big[\ln\Big(q\Big(\mathbf{W}^{(\mathrm{m})}\Big)\Big)\Big] \\ &\quad + \mathbb{E}_q\Big[\ln\Big(q\Big(\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] + \mathbb{E}_q\Big[\ln\Big(q\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big]\Big)\end{aligned}$$

Which, as we have done before, can be independently analysed. To facilitate the extraction of results, these new terms will be included in different subsections.

$\mathbb{E}_q[\ln(q(\mathbf{Z}))]$

$$\begin{aligned}\mathbb{E}_q[\ln(q(\mathbf{Z}))] &= \sum_{\mathrm{n}=1}^{\mathrm{N}} \Big(\frac{K_c}{2}\ln(2\pi e) + \frac{1}{2}\ln|\Sigma_{\mathbf{Z}}|\Big) \\ &= \frac{NK_c}{2}\ln(2\pi e) + \frac{N}{2}\ln|\Sigma_{\mathbf{Z}}|\end{aligned} \tag{121}$$

$\mathbb{E}_q\big[\ln\big(q\big(\mathbf{W}^{(\mathrm{m})}\big)\big)\big]$

$$\begin{aligned}\mathbb{E}_q\Big[\ln\Big(q\Big(\mathbf{W}^{(\mathrm{m})}\Big)\Big)\Big] &= \sum_{\mathrm{d}=1}^{\mathrm{D_m}} \Big(\frac{K_c}{2}\ln(2\pi e) + \frac{1}{2}\ln|\Sigma_{\mathbf{W}^{(\mathrm{m})}}|\Big) \\ &= \frac{D_m K_c}{2}\ln(2\pi e) + \frac{D_m}{2}\ln|\Sigma_{\mathbf{W}^{(\mathrm{m})}}|\end{aligned} \tag{122}$$

$\mathbb{E}_q\big[\ln\big(q\big(\boldsymbol{\alpha}^{(\mathrm{m})}\big)\big)\big]$

$$\mathbb{E}_q\Big[\ln\Big(q\Big(\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] \;=\; \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\Big(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}} + \ln\Big(\Gamma\Big(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big)\Big) - \Big(1 - a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big)\psi\Big(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big) - \ln\Big(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\Big)\Big) \tag{123}$$

$\mathbb{E}_q\big[\ln\big(q\big(\tau^{(\mathrm{m})}\big)\big)\big]$

$$\mathbb{E}_q\Big[\ln\Big(q\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big] \;=\; a_{\tau^{(\mathrm{m})}} + \ln(\Gamma(a_{\tau^{(\mathrm{m})}})) - (1 - a_{\tau^{(\mathrm{m})}})\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}}) \tag{124}$$

3.6 Complete definition of the lower bound ($L(q)$)

Once the different terms have been calculated, we can now determine how the lower bound is going to update by joining together all the previously calculated terms:

$$\begin{aligned}
L_q \;&=\; \mathbb{E}_q[\ln(p(X,\Theta))] - \mathbb{E}_q[\ln(q(\Theta))] \\
&=\; \mathbb{E}_q[\ln(p(\mathbf{Z}))] + \sum_{\mathrm{m}=1}^{\mathrm{M}}\Big(\mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{W}^{(\mathrm{m})},\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] \\
&\quad + \mathbb{E}_q\Big[\ln\Big(p\Big(\mathbf{X}^{(\mathrm{m})},\tau^{(\mathrm{m})}\,|\,\mathbf{W}^{(\mathrm{m})},\mathbf{Z}\Big)\Big)\Big]\Big) \\
&\quad - \mathbb{E}_q[\ln(q(\mathbf{Z}))] - \sum_{\mathrm{m}=1}^{\mathrm{M}}\Big(\mathbb{E}_q\Big[\ln\Big(q\Big(\mathbf{W}^{(\mathrm{m})}\Big)\Big)\Big] + \mathbb{E}_q\Big[\ln\Big(q\Big(\boldsymbol{\alpha}^{(\mathrm{m})}\Big)\Big)\Big] \\
&\quad + \mathbb{E}_q\Big[\ln\Big(q\Big(\tau^{(\mathrm{m})}\Big)\Big)\Big]\Big)
\end{aligned}$$

$$
\begin{aligned}
= & -\frac{NK_c}{2}\ln(2\pi) - \frac{1}{2}\mathrm{Tr}\{\langle \mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\} \\
& + \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(-\frac{K_c D_m}{2}\ln(2\pi) + \left(\frac{D_m}{2}+\alpha_0-1\right)\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right)\right. \\
& \left. + K_c(\alpha_0\ln(\beta_0) - \ln(\Gamma(\alpha_0))) - \sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right) \\
& + \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(-\frac{ND_m}{2}\ln(2\pi) - a_{\tau^{(\mathrm{m})}} + \alpha_0^{\tau}\ln(\beta_0^{\tau}) - \ln(\Gamma(\alpha_0^{\tau}))\right. \\
& \left. + \left(\frac{D_m}{2}+\alpha_0^{\tau}-1\right)(\psi(a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}}))\right) \\
& - \frac{NK_c}{2}\ln(2\pi e) - \frac{N}{2}\ln|\Sigma_{\mathbf{Z}}| \\
& - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\frac{D_m K_c}{2}\ln(2\pi e) + \frac{D_m}{2}\ln|\Sigma_{\mathbf{W}^{(\mathrm{m})}}|\right) \\
& - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}} + \ln\left(\Gamma\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right) - \left(1-a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\psi\left(a_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right) - \ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right)\right) \\
& - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(a_{\tau^{(\mathrm{m})}} + \ln(\Gamma(a_{\tau^{(\mathrm{m})}})) - (1-a_{\tau^{(\mathrm{m})}}) - \ln(b_{\tau^{(\mathrm{m})}})\right) \qquad (125)
\end{aligned}
$$

As this equation is going to be iterated through to check the state of the bound, we can simplify it by eliminating all the terms that are constant through the iteration process:

$$
\begin{aligned}
L_q = & -\frac{1}{2}\mathrm{Tr}\{\langle \mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\rangle\} \\
& - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\left(\frac{D_m}{2}+\alpha_0-1\right)\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right)\right) \\
& - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\left(\frac{D_m}{2}+\alpha_0^{\tau}-1\right)(\ln(b_{\tau^{(\mathrm{m})}}))\right) \\
& - \frac{N}{2}\ln|\Sigma_{\mathbf{Z}}| - \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\frac{D_m}{2}\ln|\Sigma_{\mathbf{W}^{(\mathrm{m})}}|\right) + \sum_{\mathrm{m}=1}^{\mathrm{M}}\left(\sum_{\mathrm{k}=1}^{\mathrm{K_c}}\left(\ln\left(b_{\alpha_{\mathrm{k}}^{(\mathrm{m})}}\right)\right)\right) \\
& + \sum_{\mathrm{m}=1}^{\mathrm{M}}(\ln(b_{\tau^{(\mathrm{m})}})) \qquad (126)
\end{aligned}
$$

References

Bishop, C.M., 1999. Bayesian pca. Advances in neural information processing systems , 382–388.