1 Appendix: KL minimization

To optimize the parameters $\theta = (\sigma_1, \sigma_2, p)$, we need to maximize the following objective function:

$$p(W|\theta) = \prod_{ij} \left(p_i \mathcal{N}(W_{ij}|0, {\sigma_1}_i^2) + (1 - p_i)\mathcal{N}(W_{ij}|0, {\sigma_2}_i^2)\right) \tag{1}$$

$$\mathbb{E}_q \log p(W|\theta) \to \max_\theta \tag{2}$$

To facilitate this optimization, we introduce latent variables z_{ij}, which denote the assignment of each weight W_{ij} to one of the two Gaussian components:

$$p(W, z|\theta) = \prod_{ij} \left(p_i \mathcal{N}(W_{ij}|0, {\sigma_1}_i^2)\right)^{z_{ij}} \left((1 - p_i)\mathcal{N}(W_{ij}|0, {\sigma_2}_i^2)\right)^{1-z_{ij}} \tag{3}$$

We can then derive a lower bound for the objective function:

$$\begin{aligned} \mathbb{E}_q \log p(W|\theta) = \mathbb{E}_q\mathbb{E}_r \left(\log p(W, z|\theta) - \log r(z)\right) + \mathbb{E}_q KL(r(z)\|p(z|W, \theta)) \geq \\ \mathbb{E}_r \left(\mathbb{E}_q \log p(W, z|\theta) - \log r(z)\right) = \mathbb{E}_r \left(\log p(\tilde{W}, z|\theta) - \log r(z)\right) = \mathcal{L}_{GM}(\theta, r) \end{aligned} \tag{4}$$

where $\tilde{W}_{ij} = \sqrt{\mu_{ij}^2 + s_{ij}^2}$. This lower bound is equivalent to the objective function of a Gaussian mixture model applied to $\tilde{W}$:

$$p(\tilde{W}|\theta) = \prod_{ij} \left(p_i \mathcal{N}(\tilde{W}_{ij}|0, {\sigma_1}_i^2) + (1 - p_i)\mathcal{N}(\tilde{W}_{ij}|0, {\sigma_2}_i^2)\right) \to \max_\theta \tag{5}$$

Thus, we can solve this optimization problem using the same EM algorithm that is applied to a Gaussian mixture model. After applying the EM algorithm, the lower bound $\mathcal{L}_{GM}(\theta, r)$ will coincide with $\log p(\tilde{W}|\theta)$. Consequently, in the optimization step with respect to q, we will use the lower bound $p(\tilde{W}|\theta)$ instead of $\mathbb{E}_q \log p(W|\theta)$.