Problem 1.a.

Solution: We want to find the explicit form of the complete data log-likelihood function.

$$\log p(X, Z|\theta) = \sum_{n=1}^{N} \{\log p(x_n|z_n;\theta) + \log p(z_n|\theta)\} \ . \tag{1}$$

The first term in the sum is defined to be

$$\log p(x_n|z_n;\theta) = \log \mathcal{N}(Wz_n + \mu, \sigma^2 \mathbb{I}_d) \, .$$

Using equation (2) from the exercise sheet we can write this explicitly as

$$\log p(x_n|z_n;\theta) = -\frac{1}{2}\left(d \log 2\pi + \log \sigma^{2d} + \frac{1}{\sigma^2}\sum_{i=1}^{d}(x_{n,i} - \mu_i - \sum_{j=1}^{m} w_{ij} z_{n,j})^2\right) \ , \tag{2}$$

where w_{ij} is the element in the i-th row and j-th column of W and we have used that $\det(\sigma^2 \mathbb{I}_d) = \sigma^{2d}$, $(\sigma^2 \mathbb{I}_d)^{-1} = \frac{1}{\sigma^2}\mathbb{I}_d$ and

$$\operatorname{tr}\left\{(\sigma^2 \mathbb{I}_d)^{-1}(x - (Wz + \mu))(x - (Wz + \mu))^T\right\} = \frac{1}{\sigma^2}\sum_{i=1}^{d}(x_i - \mu_i - \sum_{j=1}^{m} w_{ij} z_j)^2 \, .$$

The second term of eq. (1) is defined as

$$\log p(z_n|\theta) = \log \mathcal{N}(0, \mathbb{I}_m) \, .$$

Again using equation (2) from the exercise sheet we obtain

$$\log p(z_n|\theta) = -\frac{1}{2}\left(m \log 2\pi + \sum_{i=1}^{m} z_{n,i}^2\right) \ , \tag{3}$$

since $\log|\mathbb{I}_m| = \log 1 = 0$ and $\operatorname{tr}\left(\mathbb{I}_m (z-0)(z-0)^T\right) = \sum_{i=1}^{m} z_i^2$.

Inserting eq. (2) and (3) into the original formulation (1) gives the desired expression. □

Problem 2.a.

Solution:
Given a marginal Gaussian distribution for x and a conditional Gaussian distribution for y given x in the form:

$$p(x) = \mathcal{N}(x|\mu_1,\ \sigma_1^{-1}) \tag{4}$$

$$p(y|x) = \mathcal{N}(y|Wx + \mu_2,\ \sigma_2^{-1}) \tag{5}$$

the marginal distribution of y and the conditional distribution of x given y are given by:

$$p(y) = \mathcal{N}(y|W_1 + \mu_2,\ \sigma_2^{-1} + W\sigma_1^{-1}W^T) \tag{6}$$

$$p(x|y) = \mathcal{N}(x|\Sigma W^T\sigma_2(y - \mu_2) + \sigma_1\mu_1, \Sigma) \tag{7}$$

, where $\Sigma = (\sigma_1 + W^T\sigma_2 W)^{-1}$.
In our case $\Sigma = \mathbb{I}_M + \sigma^2(W^TW)^{-1}$.
Let's introduce new variable M s.t : $\sigma^2M^{-1} = \Sigma$. It follows that $M := W^TW + \sigma^2\mathbb{I}_M$
Applying (4) and (5) and definition of M to $p(x|z)$ yields:

$$p(z|x) = \mathcal{N}(z|M^{-1}W^T(x - \mu),\ \sigma^2M^{-1}). \tag{8}$$

If $p(z|x)$ is written in this form it is easy to compute the expected value of z_n given x_n:

$$\mathbb{E}[z_n|x_n] = \mathbb{E}[p(z_n|x_n)] = (W^TW + \sigma^2\mathbb{I}_M)^{-1}W^T(x_n - \mu) \tag{9}$$

Having the value of $\mathbb{E}[z_n|x_n]$ allows us to compute $\mathbb{E}[z_nz_n^T|x_n]$:

$$\mathbb{E}[z_nz_n^T|x_n] = \mathbb{E}[z_n|x_n]\mathbb{E}[z_n|x_n]^T + Cov(z_n) = \sigma^2(W^TW + \sigma^2\mathbb{I}_M)^{-1} + \mathbb{E}[z_n|x_n]\mathbb{E}[z_n|x_n]^T \tag{10}$$

□

Problem 3.a.

Solution: we have

$$p(x_n) \sim \mathcal{N}(0, WW^\top + \sigma^2)$$

and

$$p(x|z) \sim \mathcal{N}(Wz + \mu, \sigma^2\mathbb{I})$$

with the marginal distribution over the latent variables also Gaussian and conventially defined by $z \sim \mathcal{N}(0, \mathbb{I})$, the marginal distribution for x is obtained by integrating out the latent variables $x \sim \mathcal{N}(\mu, C)$
Where the Covariance Matrix is :

$$C = WW^\top + \sigma^2\mathbb{I}$$

The corresponding log-likelihood is then:

$$\mathcal{L} = -\frac{N}{2}\Bigg[d\log(2\pi) + ln|C| + tr(C^{-1}S)\Bigg]$$

Where

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^\top$$

S is the sample covariance Matrix of the observations x_n. we make use of the conditional-distribution of the latent variable $z|x$:

$$z|x \sim \mathcal{N}(M^{-1}W^\top(x-\mu), \sigma^2 M^{-1})$$

Where $M = W^\top W + \sigma^2\mathbb{I}$, The Corresponding complete data log-likelihood is then:

$$\mathcal{L} = \sum_{n=1}^{N}\ln\big[p(x_n, z_n)\big]$$

Where in PCCA, we have:

$$\mathrm{p}(\mathrm{x}_n, z_n) = (2\pi\sigma^2)^{-\frac{d}{2}}.\exp\Bigg[-\tfrac{||x_n - Wz_n - \mu||^2}{2\sigma^2}\Bigg](2\pi)^{\frac{-q}{2}}.exp\big[-\tfrac{||z_n||^2}{2}\big]$$

$$\log p(x_n, z_n|W, \sigma^2) = -\sum_{n=1}^{N}\Bigg[\frac{d}{2}\ln\sigma^2 + \frac{1}{2}tr(z_n z_n^\top) + \frac{1}{2\sigma^2}(x_n-\mu)^\top(x_n-\mu) - \frac{1}{\sigma^2}(z_n^\top)W^\top(x_n-\mu) + \frac{1}{2\sigma^2}tr(W^\top W(z_n z_n^\top))\Bigg]$$

$$\mathbb{E}\Bigg[\log p(x_n, z_n|W, \sigma^2)\Bigg] = -\sum_{n=1}^{N}\Bigg[\frac{d}{2}\log\sigma^2 + \frac{1}{2}tr(\mathbb{E}[z_n z_n^\top]) + \frac{1}{2\sigma^2}(x_n-\mu)^\top(x_n-\mu) - \frac{1}{\sigma^2}\mathbb{E}(z_n^\top)W^\top(x_n-\mu) + \frac{1}{2\sigma^2}tr(\mathbb{E}(z_n z_n^\top)W^\top W)\Bigg]$$

Where we have proved before in 2.a:

$$\mathbb{E}[z_n z_n^T] = \sigma^2 M^{-1} + \mathbb{E}[z_n]\mathbb{E}[z_n]^T \tag{11}$$

$$\mathbb{E}[z_n] = M^{-1}W^T(x_n - \mu) \tag{12}$$

Where $M = (W^\top.W + \sigma^2\mathbb{I})$ In M-step, we Re-estimate W and σ^2(taking the derivative w.r.t W and σ^2), which gives:

$$W_{new} = \Bigg[\sum_{n=1}^{N}(x_n-\mu)\mathbb{E}[z_n^T]\Bigg].\Bigg[\sum_{n=1}^{N}\mathbb{E}[z_n z_n^T]\Bigg]^{-1}$$

and

$$\sigma_{new}^2 = \frac{1}{ND}\sum_{n=1}^{N}\left[||x_n - \mu||^2 - 2\mathbb{E}[z_n^T].W_{new}^\top (x_n - \mu) + tr(\mathbb{E}(z_n z_n^\top) W_{new}^\top W_{new}) \right]$$

By substitution of $\mathbb{E}[z_n]$ and $\mathbb{E}[z_n z_n^T]$(the E-step) into the expression of W_{new} and σ_{new}^2(the M-step), we get:

$$W_{new} = SW(\sigma^2\mathbb{I} + M^{-1}W^\top SW)^{-1}$$

$$\sigma_{new}^2 = \frac{1}{D} tr(S - SWM^{-1}W_{new}^\top)$$

where:

$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^\top$ □