Problem 1. Let $X = \{x_1, \dots, x_N\}$, $Z = \{z_1, \dots, z_N\}$. We summarize the EM algorithm as follows:

1. Initialize W_0.
2. [1]For iteration $t = 1, \dots, T$
   (a) E-step: Calculate $q_t(Z) = p(Z|X, W_{t-1}) = \prod_{n=1}^N p(z_n|x_n, W_{t-1})$.
   (b) M-step: Update $W_t = \arg\max_W \mathcal{L}_t(W) = \arg\max_W \mathbb{E}_{q_t(Z)}[\ln p(X, Z, W) - \ln q_t(Z)]$.
   (c) [2] Calculate $\ln p(X, W_{t-1}) = \mathcal{L}_t(W_{t-1})$.

(a). For E-step, we have

$$q_t(Z) = p(Z|X, W_{t-1}) = \prod_{n=1}^N p(z_n|x_n, W_{t-1}),$$

where

$$p(z_n|x_n, W_{t-1}) \propto \underbrace{p(x_n|z_n, W_{t-1})}_{\mathcal{N}(W_{t-1}z_n, \sigma^2 I)} \cdot \underbrace{p(z_n)}_{\mathcal{N}(0,I)} = \mathcal{N}(z_n; \mu_n, \Sigma),$$

where

$$\Sigma = (I + \frac{1}{\sigma^2} W_{t-1}^T W_{t-1})^{-1}$$
$$\mu_n = \Sigma \cdot W_{t-1}^T x_n/\sigma^2 = (I + \frac{1}{\sigma^2} W_{t-1}^T W_{t-1})^{-1} W_{t-1}^T x_n/\sigma^2.$$

Some of the useful expectations under the posterior will be useful later:

$$\mathbb{E}_{q_t(z_n)}[z_n] = \mu_n$$
$$\mathbb{E}_{q_t(z_n)}[z_n z_n^T] = \mu_n \mu_n^T + \Sigma,$$

where we denote $q_t(z_n) = p(z_n|x_n, W_{t-1})$.
(b). For M-step, we have

$$\begin{aligned}
\mathcal{L}_t(W) &= \mathbb{E}_{q_t(Z)}[\ln p(X, Z, W)] + \text{const} \\
&= \mathbb{E}_{q_t(Z)}[\ln p(W) + \sum_{n=1}^N \ln p(z_n) + \sum_{n=1}^N \ln p(x_n|z_n, W)] + \text{const} \\
&= \ln p(W) + \sum_{n=1}^N \mathbb{E}_{q_t(z_n)}[\ln p(x_n|z_n, W)] + \text{const} \\
&= -\frac{\lambda}{2}\text{tr}(W^T W) + \sum_{n=1}^N \mathbb{E}_{q_t(z_n)}\left[-\frac{1}{2\sigma^2}(x_n - Wz_n)^T(x_n - Wz_n)\right] + \text{const} \\
&= -\frac{\lambda}{2}\text{tr}(W^T W) + \sum_{n=1}^N -\frac{1}{2\sigma^2}\left\{-2\text{tr}(\mu_n x_n^T W) + \text{tr}(W^T W(\mu_n\mu_n^T + \Sigma))\right\} + \text{const},
\end{aligned}$$

where all the terms free from W are absorbed into one "const" term, and we make use of the expectations calculated above and the following "trace" trick:

$$\mathbb{E}_{q_t(z_n)}[z_n^T W^T W z_n] = \mathbb{E}_{q_t(z_n)}[\text{tr}(z_n^T W^T W z_n)] = \text{tr}\Big(W^T W \cdot \mathbb{E}_{q_t(z_n)}[z_n z_n^T]\Big).$$

Then the derivative of $\mathcal{L}_t(W)$ w.r.t. W and set it to 0:

$$\frac{\partial \mathcal{L}_t(W)}{\partial W} = 0 \quad \to \quad \frac{1}{\sigma^2}\sum_{n=1}^N x_n \mu_n^T - W\Big\{\frac{1}{\sigma^2}\sum_{n=1}^N \Big(\mu_n\mu_n^T + \Sigma\Big) + \lambda I\Big\} = 0$$

[1]We may stop when $\ln p(X, W_t) - \ln p(X, W_{t-1}) < \varepsilon$.
[2]We may do this before the M-step.

$$W_t = \Big(\sum_{n=1}^{N} x_n \mu_n^T\Big)\Big(\sum_{n=1}^{N} \mu_n \mu_n^T + N \cdot \Sigma + \lambda\sigma^2 \cdot I\Big)^{-1}.$$

(c). The marginal objective is

$$\begin{aligned}
\ln p(X, W_{t-1}) = \mathcal{L}_t(W_{t-1}) &= \mathbb{E}_{q_t(Z)}[\ln p(X, Z, W_{t-1}) - \ln p(Z|X, W_{t-1})] \\
&= \ln p(W_{t-1}) + \sum_{n=1}^{N} \mathbb{E}_{q_t(z_n)}[\ln p(x_n|z_n, W_{t-1}) + \ln p(z_n) - \ln q_t(z_n)].
\end{aligned}$$

We may compute term-by-term as:

$$\begin{aligned}
\ln p(W_{t-1}) &= \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} \mathrm{tr}(W_{t-1}^T W_{t-1}), \\
\mathbb{E}_{q_t(z_n)}[\ln p(x_n|z_n, W_{t-1})] &= -\frac{d}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} x_n^T x_n + \frac{1}{\sigma^2} \mathrm{tr}(\mu_n x_n^T W_{t-1}) - \frac{1}{2\sigma^2} \mathrm{tr}(W_{t-1}^T W_{t-1}(\mu_n \mu_n^T + \Sigma)), \\
\mathbb{E}_{q_t(z_n)}[\ln p(z_n)] &= -\frac{k}{2} \ln(2\pi) - \frac{\mathrm{tr}(\mu_n \mu_n^T + \Sigma)}{2}, \\
\mathbb{E}_{q_t(z_n)}[-\ln q_t(z_n)] &= H(q_t(z_n)) = \frac{1}{2} \ln \det(2\pi e \Sigma),
\end{aligned}$$

where $H(\cdot)$ denotes the entropy of a distribution, and $\det(\cdot)$ represents the matrix determinant.