Francisco Javier Arceo
EECS 6892
Homework 2

Problem 1.)

$$x_n \in \mathbb{R}^d \quad x_n \sim N(Wz_n, \sigma^2 I)$$
$$W \in \mathbb{R}^{d \times k} \quad W \sim N(0, \lambda^{-1})$$
$$z_n \in \mathbb{R}^k \quad z_n \sim N(0, I)$$

$$p(x \mid Wz_n, \sigma^2 I) = (2\pi\sigma^2 I)^{-\frac{d}{2}} \exp\left[-\frac{1}{2\sigma^2 I}(X_n - Wz_n)^T (X_n - Wz_n)\right]$$

$$p(W) = \left(\frac{\lambda}{2\pi}\right)^{\frac{dk}{2}} \exp\left[-\frac{\lambda}{2}\,\mathrm{trace}(W^T W)\right]$$

$$p(z_n) = (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z_n^T z_n\right]$$

In order to find

$$W' = \arg\max_W \ln p(x_1, \dots, x_n, W)$$

Note that

$$\ln p(x_1, \dots, x_n, W) = \int q(z_n) \ln \frac{p(x_1, \dots, x_n, W, z_n)}{q(z_n)} \, dz_n$$

$$q(z) = p(z \mid W, x) \propto \prod_{i=1}^{n} p(x_n \mid W, z_n)\, p(z_n)\, p(Wz_n)$$

$$q(z) \propto \prod_{i=1}^{n} (2\pi\sigma^2 I)^{-\frac{d}{2}} \exp\left[-\frac{1}{2\sigma^2 I}(X_n - Wz_n)^T (X_n - Wz_n)\right] (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z^T z\right]$$

$$= (2\pi\sigma^2 I)^{-\frac{dn}{2}} (2\pi)^{-\frac{nk}{2}} \exp\left[-\frac{1}{2\sigma^2 I}(X_n - Wz_n)^T (X_n - Wz_n) - \frac{1}{2} z^T z\right]$$

$$= \underbrace{(2\pi)^{-\frac{n(d+k)}{2}} (\sigma^2 I)^{-\frac{dn}{2}}}_{\text{constants}} \exp\left[\sum_{i=1}^{n} -\frac{1}{2\sigma^2 I}\left[X_n^T X_n - X_n^T W z_n - (Wz_n)^T X_n + (Wz_n)^T W z_n\right] - \frac{1}{2} z^T z\right]$$

$$= \text{constants} \times \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left[\frac{X_n^T X_n}{\sigma^2} - \frac{X_n^T W z_n}{\sigma^2} - \frac{X_n^T W z_n}{\sigma^2} + \frac{(Wz_n)^T W z_n}{\sigma^2} + \frac{\sigma^2 z^T z}{\sigma^2}\right]\right]$$

$$\propto \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left[\frac{X_n^T X_n}{\sigma^2} - \frac{2 X_n^T W z_n}{\sigma^2} + \frac{(Wz_n)^T W z_n}{\sigma^2} + \frac{\sigma^2 z_n^T z_n}{\sigma^2}\right]\right]$$

$$= \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left[\frac{X_n^T X_n}{\sigma^2} - \frac{2 X_n^T W z_n}{\sigma^2} + \frac{z_n^T \mathrm{tr}(W^T W) z_n}{\sigma^2} + \frac{\sigma^2 z_n^T z_n}{\sigma^2}\right]\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(X_n^T X_n - 2 X_n^T W z_n + z_n^T z \left(\mathrm{tr}(W^T W) + \sigma^2\right)\right)\right]$$

$$= \exp\left[-\frac{\mathrm{tr}(W^T W) + \sigma^2}{2\sigma^2} \sum_{i=1}^{n} \left(\frac{X_n^T X_n - 2 X_n^T W z_n}{\mathrm{tr}(W^T W) + \sigma^2} + z_n^T z\right)\right]$$

doesn't depend on z

$$q(z) \propto \exp\left[-\frac{tr(W^TW)+\sigma^2}{2\sigma^2}\sum_{n=1}^{n}\left(\cancel{\frac{X_n^TX_n}{tr(W^TW)+\sigma^2}} - \frac{2X_n^T W z_n}{tr(W^TW)+\sigma^2} + z_n^T z_n\right)\right]$$

$$\propto \exp\left[-\frac{trace(W^TW)+\sigma^2}{2\sigma^2}\sum_{n=1}^{n}\left(z_n^T - \frac{X_n^TW}{tr(W^TW)+\sigma^2}\right)\left(z_n - \frac{X_n^TW}{tr(W^TW)+\sigma^2}\right)\right]$$

Which is a Normal distribution with

$$p(z|w,x) \sim N\left(\frac{\sum_{n=1}^N X_n^TW \frac{1}{N}}{trace(W^TW)+\sigma^2 I} = \mu^*, \ \frac{\sigma^2 I}{N(trace(W^TW)+\sigma^2 I)} = \Sigma^*\right)$$

Then, we need to take the expectation of the joint likelihood and maximize it.

$$\mathcal{L}(W) = E_q[\ln p(z,x,w)]$$

$$p(z,x,w) \propto p(w)\prod_{n=1}^N p(x_n|w,z_n)\,p(z_n)$$

$$\ln p(z,x,w) \propto \ln\left[\left(\frac{\lambda}{2\pi}\right)^{\frac{dk}{2}}\exp\left[-\frac{\lambda}{2}trace(W^TW)\right](2\pi\sigma^2 I)^{-\frac{dn}{2}}\exp\left[-\frac{1}{2\sigma^2 I}\sum_{n=1}^N (x_n - Wz_n)^T(x_n - Wz_n)\right] \times (2\pi)^{-\frac{kn}{2}}\exp\left[\sum_{n=1}^N -\frac{1}{2}z_n^Tz_n\right]\right]$$

$$= \ln\left[[2\pi]^{-\frac{dk}{2}}(\lambda)^{\frac{dk}{2}}(2\pi)^{-\frac{dn}{2}}(2\pi)^{-\frac{kn}{2}}(\sigma)^{-d}\exp\left[-\frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2 I}\left(\sum_{n=1}^N(x_n - Wz_n)^2\right) - \frac{1}{2}\sum_{n=1}^N z_n^Tz_n\right]\right]$$

$$= \ln\left[\lambda^{\frac{dk}{2}}(2\pi)^{-\frac{n(d+k)}{2}}\sigma^{-d}\exp\left[-\frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2 I}\left(\sum_{n=1}^N(x_n - Wz_n)^2\right) - \frac{1}{2}\sum_{n=1}^N z_n^Tz_n\right]\right]$$

$$= \underbrace{\left(\frac{dk}{2}\right)\ln\lambda - \frac{n(d+k)}{2}\ln 2\pi - d\ln\sigma}_{\text{constants wrt } W} - \frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2 I}\left(\sum_{n=1}^N(x_n - Wz_n)^2\right) - \frac{1}{2}\sum_{n=1}^N z_n^Tz_n$$

$$= E_q\left[\text{constants} - \frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2 I}\left(\sum_{n=1}^N(x_n - Wz_n)^T(x_n - Wz_n)\right) - \frac{1}{2}\sum_{n=1}^N z_n^Tz_n\right.$$

$$= E\left[\text{constants} - \frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2}\sum_{n=1}^N X_n^TX_n + \frac{1}{\sigma^2 I}\sum_{n=1}^N X_n^TWz_n - \frac{1}{2\sigma^2}trace(W^TW)\sum_{n=1}^N z_nz_n^T\right]$$

$$\mathcal{L}_t(W) = \text{constants} - \frac{\lambda}{2}trace(W^TW) - \frac{1}{2\sigma^2}\sum_{n=1}^N X_n^TX_n + \frac{1}{\sigma^2 I}\sum_{n=1}^N X_n^TWE[z_n] - \frac{1}{2\sigma^2}trace(WW^T)\sum_{n=1}^N E[z_nz_n^T]$$

$$\nabla_W\mathcal{L}(W) = 0 = -\lambda W - \frac{W}{\sigma^2}\sum_{n=1}^N E(z_nz_n^T) + \sum_{n=1}^N X_nE(z_n)$$

$$W\left(\lambda + \frac{1}{\sigma^2}\sum_{n=1}^N E(z_nz_n^T)\right) = \sum_{n=1}^N X_nE(z_n)$$

$$W^*_{t+1} = \frac{\sum_{n=1}^N X_nE(z_n)}{\lambda + \frac{1}{\sigma^2}E(z_nz_n^T)} = \frac{\sum_{n=1}^N X_n\mu^*}{\lambda + \frac{1}{\sigma^2}\Sigma^*}$$

//

Thus, we have both the E-step & the M-step. The next step that is left is to verify convergence by calculating $\ln p_t(x_1,\dots,x_n,W)$ and ensuring that as $t \to \infty$ the term is monotonically increasing. To calculate we do,

$$\ln p_t(x_1,\dots,x_n,W) = \ln \int q(z_n) \frac{p(x_1,\dots,x_n,W,z_n)}{q(z_n)} dz_n.$$

$$= E_q[\ln p(x_1,\dots,x_n,W,z_n)]$$

$$= -\frac{\lambda}{2}\,\mathrm{trace}(W^TW) - \frac{1}{2\sigma^2}\sum_{n=1}^{N} X_n^T X_n + \frac{1}{\sigma^2}\sum_{n=1}^{N} X_n^T W \mu^* - \frac{1}{2\sigma^2}\,\mathrm{trace}(WW^T)\sum_{n=1}^{N} \Sigma^*$$

$$\ln p_{t+1}(x_1,\dots,x_n,W_t) = -\frac{\lambda}{2}\,\mathrm{trace}(W_t^TW_t) - \frac{1}{2\sigma^2}\sum_{n=1}^{N} X_n^T X_n + \frac{1}{\sigma^2}\sum_{n=1}^{N} X_n^T W_t\left(\frac{\sum_{n=1}^{N} X_n^T W_t \frac{1}{N}}{\mathrm{trace}(W_t^TW_t)+\sigma^2 I}\right) - \frac{1}{2\sigma^2}\,\mathrm{trace}(W_tW_t^T)\, N\left(\frac{\sigma^2 I}{N\cdot \mathrm{trace}(W_t^TW_t)+\sigma^2 I}\right)$$

This EM algorithm can be implemented by the following:

1. Initialize W_0 as a $(d \times k)$ matrix with zeroes.
2. For iterations $t = 1, 2, \dots, T$
   a. Calculate $E_{q_t}[z] = \frac{N^{-1}\sum X_i^T W}{\mathrm{trace}(W^TW)+\sigma^2 I}$ (E-step)
   b. Update W_t^* as $W_{t+1} = \frac{\sum_{n=1}^{N} X_n E(z_n)}{\lambda + \frac{1}{\sigma^2} E(z_n z_n^T)}$ (M-step).
   c. Calculate $\ln p_t(x_1,\dots,x_n,W)$ and stop when $\ln p_t(\cdot) - \ln p_{t-1}(\cdot) < \epsilon$ for some small $\epsilon > 0$.