Probabilistic PCA

Let X be a standard Gaussian random variable in $\mathbb{R}^d$ and assume that conditionally on X, Z has a Gaussian distribution with mean $WX + \mu$ and variance $\sigma^2 I_d$.

1. Prove that Z has a Gaussian distribution with mean μ and variance $C = WW^\top + \sigma^2 I_d$.

   It is enough to write
   $$Z = WX + \mu + \sigma\varepsilon\,,$$
   where X and ε are independent Gaussian random variables. Therefore, $Z \sim \mathcal{N}(\mu, C)$.

2. Prove that conditionally on Z, X has a Gaussian distribution with mean $m = C^{-1}W^\top(Z-\mu)$ and variance $\Sigma = \sigma^2 C^{-1}$.

   The joint distribution of Z and X can be written, for all $(x, z) \in \mathbb{R}^d \times \mathbb{R}^d$,
   $$p(z, x) \propto \exp(-\frac{1}{2}x^\top x)\exp(-\frac{1}{2\sigma^2}(z - Wx - \mu)^\top(z - Wx - \mu))\,.$$
   Therefore,
   $$\begin{aligned} p(x|z) &\propto p(z, x) \\ &\propto \exp(-\frac{1}{2}x^\top x)\exp(-\frac{1}{2\sigma^2}(z - Wx - \mu)^\top(z - Wx - \mu))\,, \\ &\propto \exp(-\frac{1}{2\sigma^2}(z - Wx - \mu)^\top(z - Wx - \mu))\,, \\ &\propto \exp(-\frac{1}{2\sigma^2}(x - m)^\top\Sigma^{-1}(x - m))\,, \end{aligned}$$
   where $\Sigma = \sigma^2 C^{-1}$ and $m = C^{-1}W^\top(Z - \mu)$ which concludes the proof.

3. Assume that $\{Z_i\}_{1\leq i\leq n}$ are n i.i.d. observations with the same distribution has Z. Write the loglikelihood of $(Z_1, \ldots, Z_n)$.

   By definition,
   $$\begin{aligned} \log p_C(Z_1, \ldots, Z_n) &= \sum_{i=1}^n \log p_C(Z_i)\,, \\ &= \sum_{i=1}^n \left(-\frac{1}{2}\log\det(2\pi C) - \frac{1}{2}(Z_i - m)^\top\Sigma^{-1}(Z_i - m)\right)\,, \\ &= -\frac{n}{2}\log\det(2\pi C) - \frac{1}{2}\sum_{i=1}^n(Z_i - m)^\top\Sigma^{-1}(Z_i - m)\,, \\ &= -\frac{n}{2}\log\det(2\pi C) - \frac{n}{2}\mathrm{Trace}(C^{-1}S_n)\,, \end{aligned}$$
   where
   $$S_n = \frac{1}{n}\sum_{i=1}^n(Z_i - \mu)^\top(Z_i - \mu)\,.$$

4. Assuming that $C = WW^\top + \sigma^2 I_d$, show that $\widehat{C} = \widehat{W}\widehat{W}^\top + \sigma^2 I_d$ with $\widehat{W} = U_q(\Lambda_q - \sigma^2 I_q)^{1/2} R$ is a stationary point of the loglikelihood (and therefore maximizes the loglikelihood).

5.

6.