Question 3.

Given N observations $\vec{x}_1, \vec{x}_2, \dots \vec{x}_N$ with each $\vec{x}_n \in R^D$,

Consider observations

$$\vec{x}_n = A\vec{z}_n + \vec{\varepsilon}_n \quad , \text{with } \vec{\varepsilon}_n \sim \mathcal{N}(0, \psi)$$

and $A = [\vec{a}_1, \vec{a}_2, \dots \vec{a}_K]$ $D \times K$ matrix,

and $z_n = [z_{n1}, z_{n2}, \dots z_{nK}]^T$, s.t. $Z = \{z_1, z_2, \dots z_N\}$.

1. Suppose Z is ~~given~~ known, and assuming $\vec{a}_k$ prior, $a_k \sim \mathcal{N}(0, D^{-1} I_D)$.

We have to derive $p(\vec{a}_k | X, Z)$.

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ x_{12} & x_{22} & & x_{N2} \\ x_{13} & x_{23} & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{1D} & x_{2D} & & x_{ND} \end{bmatrix} = [\vec{a}_1\ \vec{a}_2\ \vec{a}_3 \dots \vec{a}_K] Z + [\vec{\varepsilon}_1\ \vec{\varepsilon}_2 \dots \vec{\varepsilon}_N].$$

$$\Rightarrow [x_{1k}\ x_{2k} \dots x_{Nk}] = \vec{a}_k^T Z + [\varepsilon_{1k}\ \varepsilon_{2k} \dots \varepsilon_{Nk}].$$

Let $\vec{X}_k$ be $[x_{1k}\ x_{2k}\ x_{3k} \dots x_{Nk}]^T$.

and $E_k = [\varepsilon_{1k}\ \varepsilon_{2k} \dots \varepsilon_{Nk}]^T$.

This gives $\vec{X}_k = Z^T \vec{a}_k + \vec{E}_k$

Notice that each $\vec{a}_k$ depends upon only the k^{th} component of $\vec{x}_i$, that is, all other components have no effect on $\vec{a}_k$. Similarly, only the k^{th} component of ε has an effect on $\vec{a}_k$.

For this, marginalise components $i = 1:D$, $i \neq k$ of ε, such that the resulting probability distribution of ε_{-k} is $\mathcal{N}(0, \sigma_k^2)$, σ_k^2 is the marginalised variance of k^{th} component of ε.

Then, using the properties of probability distributions, we get:

$$p(a_k) = \mathcal{N}(a_k | 0, D^{-1} I_D), \quad p(X | \vec{a}_k, Z, \psi) = \mathcal{N}(Z^T\vec{a}_k + \vec{E}_k, \sigma_k^2 I_N)$$

$$\Rightarrow p(\vec{a}_k | X, Z) = \mathcal{N}\left(\vec{a}_k \,\middle|\, (DI_D + Z\sigma_k^{-2} I_N Z^T)^{-1} Z \sigma_k^{-2} I_N \vec{X}_k ,\ (DI_D + Z\sigma_k^{-2} I_N Z^T)^{-1}\right)$$