Probabilistic PCA :

We can formulate PCA in a probabilistic / Bayesian setting :

Likelihood : $p(x|z) = \mathcal{N}(x \mid zW^T + \mu, \sigma^2 I)$, $x \in \mathbb{R}^d$

$\Rightarrow x_i = z_i W^T + \mu + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$, $\underset{\text{data mat}}{X \in \mathbb{R}^{n \times d}}$

Prior : $p(z) = \mathcal{N}(z \mid 0, I)$, $z \in \mathbb{R}^q$, $q \ll d$

$$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

Parameters : $\theta := \{W, \mu, \sigma^2\}$ to be determined via MLE/MAP.

Recall : $p(z|x) = \dfrac{p(x|z)\,p(z)}{p(x)}$

Because x depends on z in a linear fashion, and $p(x|z)$, $p(z)$ are Gaussian :

$$p(x) = \int p(x|z)\,p(z)\,dz$$

we can $\Longrightarrow$ derive

$$\boxed{p(x) = \mathcal{N}(x \mid \mu, C), \quad \text{where} \quad C := \underset{d\times q}{W}\,\underset{q \times d}{W^T} + \sigma^2 I}$$

since $\mathbb{E}[x] = \mathbb{E}[\overset{\downarrow}{z}W^T + \mu + \overset{\downarrow}{\varepsilon}] = \mu$

$$\mathrm{Cov}[x] = \mathbb{E}\left[(zW^T + \mu + \varepsilon)(zW^T + \mu + \varepsilon)^T\right]$$

$$= \mathbb{E}[Wzz^TW^T] + \mathbb{E}[\varepsilon\varepsilon^T] = WW^T + \sigma^2 I$$

Moreover, we can obtain the posterior over the latent variables :

$$\boxed{\begin{gathered} p(z|x) = \mathcal{N}(z \mid M^{-1}(X-\mu)W, \ \sigma^2 M^{-1}), \\ \text{where} \quad M := W^TW + \sigma^2 I \end{gathered}}$$

Let's now estimate the PPCA model parameters via MLE:

$$-\log p(x \mid \mu, W, \sigma^2) \overset{i.i.d.}{=} -\sum_{i=1}^{n} \log p(x_i \mid \mu, W, \sigma^2)$$

$$= \frac{nd}{2} \log 2\pi + \frac{n}{2} \log |C| + \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T C^{-1} (x_i - \mu)$$

$\Longrightarrow$ Then we can obtain:

$$\mu_{MLE} = \bar{x}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad S := \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$$

Then: $$-\log p(x \mid W, \mu, \sigma^2) = \frac{n}{2} \left\{ d \log 2\pi + \log |C| + \mathrm{Tr}(C^{-1} \underset{\text{sample covarian}}{\underset{\downarrow}{S}} \right.$$

This expression can be minimizd wtr W and σ^2:

$$\Rightarrow \quad W_{MLE} = U(\Lambda - \sigma^2 I)^{\frac{1}{2}} R,$$

$$\sigma^2_{MLE} = \frac{1}{d-q} \sum_{i=q+1}^{d} \lambda_i$$

where:

U: is a $d \times q$ matrix whose columns o the eigenvectors of S.

Λ: is a diagonal matrix containing the eigenvalues of S.

R: is an arbitrary orthogonal matrix