Problem 1. *Eigenvalues and eigenvectors of a symmetric matrix.* Show that the eigenvalues of a symmetric matrix with real coefficients are real. In addition, show that the eigenvectors can be chosen orthonormal.

Problem 2. *Finding principal components.* Let Σ be a $d \times d$ symmetric positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d > 0$, and associated eigenvectors $u_1, \ldots, u_d$.

(i) Show that

$$\max_{x \neq 0} \frac{x^t \Sigma x}{x^t x} = \lambda_1 \,, \qquad \text{attained at } x = u_1 \,,$$

and

$$\min_{x \neq 0} \frac{x^t \Sigma x}{x^t x} = \lambda_d \,, \qquad \text{attained at } x = u_d \,.$$

(ii) Similarly, show that

$$\max_{x \perp u_1, \ldots, u_\ell} \frac{x^t \Sigma x}{x^t x} = \lambda_{\ell+1} \,, \qquad \text{attained at } x = u_{\ell+1} \,.$$

(iii) Deduce from *(i)* and *(ii)* that the ℓ-th principal component of a random vector $X \in \mathbb{R}^d$ with covariance matrix Σ corresponds to the eigenvector u_ℓ associated with the ℓ-th largest eigenvalue λ_ℓ of Σ. In addition, show that the variance of the projection $u_\ell^t X$ on the ℓ-th principal component is λ_ℓ, and that the projections of X onto the principal components are uncorrelated with each other.

(iv) Based on *(iii)*, how would you estimate the principal components in practice, based on a random sample $\{x_1, \ldots, x_n\}$ of size n?

Problem 3. *Large Sample Properties.* Assume that $X_1, \ldots, X_n \in \mathbb{R}^d$ are independent multivariate normal random variables, with mean zero and covariance matrix Σ. The eigenvalues of Σ are all assumed distinct and such that $\lambda_1 > \lambda_2 > \ldots > \lambda_d > 0$. Put

$$X = \begin{pmatrix} X_1^t \\ \vdots \\ X_n^t \end{pmatrix} \in \mathbb{R}^{n \times d} \,.$$

Denote by $S = X^t X / n$ the sample covariance matrix of X, with eigenvalue-eigenvector pairs $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$. Let Λ be the diagonal matrix with elements $\lambda_1 > \lambda_2 > \ldots > \lambda_d > 0$. Put $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d)$ and $\lambda = (\lambda_1, \ldots, \lambda_d)$. Then

$$\sqrt{n}(\hat{\lambda} - \lambda) \to \mathcal{N}(0, 2\Lambda^2) \,, \quad \text{as } n \to \infty \,.$$

Deduce an approximate confidence interval for λ_i with nominal coverage $(1 - \alpha)$, assuming n large enough.

Problem 4. *pPCA.* The probabilistic PCA model assumes that $x \in \mathbb{R}^d$ is such that

$$x = \mathbf{W}z + \mu + \epsilon\,, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{d\times r}$, $z \in \mathbb{R}^r$, with $z \sim \mathcal{N}(0, I_r)$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_d)$, independent.

(i) Explain why the covariance matrix of ϵ is taken to be diagonal.

(ii) Show that $x \sim \mathcal{N}(\mu, \mathbf{C})$, with $\mathbf{C} = \mathbf{W}\mathbf{W}^t + \sigma^2 I_d \in \mathbb{R}^{d\times d}$.

(iii) Show that the posterior distribution $z|x \sim \mathcal{N}(\mathbf{M}^{-1}\mathbf{W}^t(x-\mu), \sigma^2\mathbf{M}^{-1})$, with $\mathbf{M} = \mathbf{W}^t\mathbf{W} + \sigma^2 I_r \in \mathbb{R}^{r\times r}$.

(iv) Write down the expression of the log-likelihood $\mathcal{L}$ of a sample $x_1, \ldots, x_n$ of size n, in terms of $\mathbf{C}$ and

$$\tilde{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^t\,.$$

(v) Show that the maximum likelihood of μ is $\bar{x} = \frac{1}{n}\sum_i x_i$. In the remainder, the covariance matrix defined is question *(iii)* is replaced by the sample covariance matrix $\mathbf{S}$, obtained by plugging in $\bar{x}$ instead of μ in the expression of $\tilde{\mathbf{S}}$.

(vi) Show that the derivative of $\mathcal{L}$ with respect to $\mathbf{W}$ is given by

$$\frac{\partial\mathcal{L}}{\partial\mathbf{W}} = n(\mathbf{C}^{-1}\mathbf{S}\mathbf{C}\mathbf{W} - \mathbf{C}^{-1}\mathbf{W})\,.$$

The maximum of the log-likelihood function can be written

$$\mathbf{W}_{ML} = \mathbf{U}_r(\mathbf{K}_r - \sigma^2_{ML} I)^{1/2}\mathbf{R}\,,$$

where $\mathbf{U}_r$ is a $(d \times r)$ matrix comprising r eigenvectors of the covariance matrix $\mathbf{S}$ associated with its largest r eigenvalues $\lambda_1 \geq \ldots \geq \lambda_r$, $\mathbf{K}_r = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, and $\mathbf{R}$ is an arbitrary (orthogonal) rotation matrix. In addition,

$$\sigma^2_{ML} = \frac{1}{d-r}\sum_{j=r+1}^{d}\lambda_j\,.$$

(vii) For applications, such as visualisation or data compression, you need to reverse the mapping from the latent space into the data space, given by (1). Explain how you would do that.

(viii) Explain then how you would reconstruct the data from the latent variable. Then show that in the limit $\sigma^2 \to 0$, we recover the standard PCA model.

Problem 5. *pPCA with no reduction of dimensionality.* Consider the model (1) with $r = d$, corresponding to the case of no reduction of dimensionality. Compute in this case the maximum likelihood estimator of the covariance matrix $\mathbf{C}$ of x, and show that you obtain the MLE of the covariance matrix for an unconstrained multivariate Gaussian distribution.

Problem 6. *Factor Analysis.* Factor Analysis assumes that observations $x_1, \ldots, x_n$ are generated from

$$x_i = \mathbf{W} z_i + \mu + \epsilon_i \,,$$

where $\mathbf{W} \in \mathbb{R}^{d\times r}$, $z \in \mathbb{R}^r$, with $z \sim \mathcal{N}(0, I_r)$ and $\epsilon \sim \mathcal{N}(0, \mathbf{\Psi})$, independent, where Ψ is a $d \times d$ diagonal matrix.

(i) Derive the marginal distribution of x_i under this model, as we as the conditional distributions of x_i given z_i, and of z_i given x_i.

(ii) Show that the maximum likelihood estimate of μ is given by the sample mean of the x_is.

(iii) Show that the complete log-likelihood $\mathcal{L}_c$ can be written

$$\mathcal{L}_c = -\frac{1}{2}\sum_{i=1}^{n}\left\{\log|\mathbf{\Psi}| - 2z_i^t\mathbf{W}^t(x_i-\mu) + \mathrm{Tr}(z_i z_i^t \mathbf{W}^t\mathbf{\Psi}^{-1}\mathbf{W})\right\} - \frac{n}{2}\mathrm{Tr}(\mathbf{\Psi}^{-1}\mathbf{S})\,,$$

where $\mathbf{S}$ is the sample covariance matrix.

(iv) E-step. Show that the conditional mean of the latent variables z_i and $z_i z_i^t$ given x_i and the current model estimates $\mathbf{W}, \mathbf{\Psi}$ are

$$\begin{aligned}\langle z_i\rangle &= \mathbf{G}\mathbf{W}^t\psi^{-1}(x-\mu)\\ \langle z_i z_i^t\rangle &= \mathbf{G} + \langle z_i\rangle\langle z_i^t\rangle\,,\end{aligned}$$

where $\mathbf{G} = (I + \mathbf{W}^t\mathbf{\Psi}^{-1}\mathbf{W})^{-1}$.

(v) M-step. Show that maximisation of the expected conditional complete log-likelihood with respect to $\mathbf{W}$ yields the update

$$\tilde{\mathbf{W}} = \left(\sum_{i=1}^{n}(x_i-\mu)\langle z_i^t\rangle\right)\left(\sum_{i=1}^{n}\langle z_i z_i^t\rangle\right)^{-1}\,.$$

(vi) M-step. Show that maximisation of the expected conditional complete log-likelihood with respect to $\mathbf{\Psi}^{-1}$ yields the update

$$\tilde{\mathbf{\Psi}} = \mathrm{diag}\left[\mathbf{S} - \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)\langle z_i^t\rangle\right)\tilde{\mathbf{W}}^t\right]\,.$$