Exercise 2

We are given the first principal component direction $\boldsymbol{e}$, which is the eigen vector of $\boldsymbol{S}$ (a constant multiple of the covariance matrix $\boldsymbol{C}$, i.e. $\boldsymbol{S} = (N-1)\boldsymbol{C}$) with the largest eigenvalue, say λ_e and is obtained by maximizing $Var(\boldsymbol{e}^T\boldsymbol{x}) = \boldsymbol{e}^T\boldsymbol{S}\boldsymbol{e}$. Let $\boldsymbol{f}$ be the direction perpendicular to $\boldsymbol{e}$ which maximizes $\boldsymbol{f}^T\boldsymbol{S}\boldsymbol{f}$, subject to the conditions that $\boldsymbol{f}$ is a unit vector and is perpendicular to $\boldsymbol{e}$. The Lagrangian for this problem can then be written as:

$$\mathcal{L} = \boldsymbol{f}^T\boldsymbol{S}\boldsymbol{f} - a(\boldsymbol{f}^T\boldsymbol{f} - 1) - b(\boldsymbol{f}^T\boldsymbol{e} - 0)$$

Derivative of the Lagrangian can then be computed as:

$$\frac{d}{d\boldsymbol{f}}(\mathcal{L}) = \frac{d}{d\boldsymbol{f}}(\boldsymbol{f}^T\boldsymbol{S}\boldsymbol{f} - a(\boldsymbol{f}^T\boldsymbol{f} - 1) - b(\boldsymbol{f}^T\boldsymbol{e} - 0)) = 2\boldsymbol{S}\boldsymbol{f} - 2a\boldsymbol{f} - b\boldsymbol{e} = 0$$

Pre-multiplying the equation above by $\boldsymbol{e}^T$, we get:

$$\boldsymbol{e}^T\boldsymbol{S}\boldsymbol{f} - a\boldsymbol{e}^T\boldsymbol{f} - b\boldsymbol{e}^T\boldsymbol{e} = 0$$

The second term in the equation above is zero because $\boldsymbol{e}$ and $\boldsymbol{f}$ are perpendicular. The first term in the expression above is $\boldsymbol{e}^T\boldsymbol{S}\boldsymbol{f}$. This can further be manipulated as follows:

$$\boldsymbol{e}^T\boldsymbol{S}\boldsymbol{f} = (\boldsymbol{S}^T\boldsymbol{e})^T\boldsymbol{f}$$

Since $\boldsymbol{C}$, and thus $\boldsymbol{S}$, is a symmetric matrix, i.e. $\boldsymbol{S}^T = \boldsymbol{S}$, we have:

$$\boldsymbol{e}^T\boldsymbol{S}\boldsymbol{f} = (\boldsymbol{S}\boldsymbol{e})^T\boldsymbol{f}$$

Now, we also know that $\boldsymbol{e}$ is an eigen vector of $\boldsymbol{S}$. Therefore, $\boldsymbol{S}\boldsymbol{e} = \lambda_e\boldsymbol{e}$. Substituting this in the expression above, we have:

$$\boldsymbol{e}^T\boldsymbol{S}\boldsymbol{f} = (\lambda_e\boldsymbol{e})^T\boldsymbol{f} = \lambda_e\boldsymbol{e}^T\boldsymbol{f} = 0$$

The last equality follows from the fact that $\boldsymbol{e}$ and $\boldsymbol{f}$ are perpendicular.

Therefore, the first and the second terms in the original equation both are zero. Thus, we are left with $b\boldsymbol{e}^T\boldsymbol{e} = 0$. Since $\boldsymbol{e}$ is a unit vector, we have $\boldsymbol{e}^T\boldsymbol{e} = 1$. Thus the only way that this equation is satisfied is if we have $b = 0$.

Now, coming back to the expression for the derivative of Lagrangian, substituting $b = 0$, we have:

$$2\boldsymbol{S}\boldsymbol{f} - 2a\boldsymbol{f} = 0$$

$$\boldsymbol{S}\boldsymbol{f} = a\boldsymbol{f}$$

Therefore, $\boldsymbol{f}$ is an eigenvector of $\boldsymbol{S}$, and hence $\boldsymbol{C}$. Now, we also want to maximize the variance captured in the direction of this eigenvector. So, we want to maximize the following expression:

$$\boldsymbol{f}^T\boldsymbol{S}\boldsymbol{f} = \boldsymbol{f}^T(\boldsymbol{S}\boldsymbol{f}) = \boldsymbol{f}^T(\lambda_f\boldsymbol{f}) = \lambda_f\boldsymbol{f}^T\boldsymbol{f} = \lambda_f$$

In the first equality, we have used the fact that $\boldsymbol{f}$ is an eigenvector of $\boldsymbol{S}$ with an eigenvalue of λ_f and in the last equality, we have used the fact that $\boldsymbol{f}$ is a unit vector. Therefore, we want to maximize λ_f. Since λ_e is the largest eigenvalue corresponding to the eigenvector $\boldsymbol{e}$, λ_f has to be necessarily the second largest eigenvalue.