EE663, Assignment 4

Question 2

Problem Statement. Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$, with average vector $\bar{\boldsymbol{x}}$. We have seen in class that the direction $\boldsymbol{e}$ such that $\sum_{i=1}^{N} \|\boldsymbol{x_i} - \bar{\boldsymbol{x}} - (\boldsymbol{e} \cdot (\boldsymbol{x_i} - \bar{\boldsymbol{x}}))\boldsymbol{e}\|^2$ is minimized, is obtained by maximizing $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{e}$, where $\boldsymbol{C}$ is the covariance matrix of the vectors in $\mathcal{X}$. This vector $\boldsymbol{e}$ is the eigenvector of matrix $\boldsymbol{C}$ with the highest eigenvalue. Prove that the direction $\boldsymbol{f}$ perpendicular to $\boldsymbol{e}$ for which $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$ is maximized, is the eigenvector of $\boldsymbol{C}$ with the second highest eigenvalue. For simplicity, assume that all non-zero eigenvalues of $\boldsymbol{C}$ are distinct and that $\text{rank}(\boldsymbol{C}) > 2$. [10 points]

Solution. We will use Lagrange Multipliers to show that the statement is true, similar to what was done in class, except with multiple constraints.

We want to maximize $g(\boldsymbol{f}) = \boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$, subject to the constraints that $h(\boldsymbol{f}) = \boldsymbol{f}^t\boldsymbol{f} - 1 = 0$ and $k(\boldsymbol{f}, \boldsymbol{e}) = \boldsymbol{e}^t\boldsymbol{f} = 0$. Note that the dimension of $\boldsymbol{C}$ is $d \times d$, and the dimensions of $\boldsymbol{e}, \boldsymbol{f}$ are $d \times 1$. Lagrange multipliers give us $d + 2$ relations for the maxima:

1. $h(\boldsymbol{f}) = 0$
2. $k(\boldsymbol{f}, \boldsymbol{e}) = 0$
3. $\nabla_{\boldsymbol{f}} g(\boldsymbol{f}) - \mu\nabla_{\boldsymbol{f}} h(\boldsymbol{f}) - \lambda\nabla_{\boldsymbol{f}} k(\boldsymbol{e}, \boldsymbol{f}) = 0$ - This represents d equations, since it is a vector derivative.

Simplifying the third expression, we get, using $\nabla_{\boldsymbol{f}} g(\boldsymbol{f}) = 2\boldsymbol{C}\boldsymbol{f}$, $\nabla_{\boldsymbol{f}} h(\boldsymbol{f}) = 2\boldsymbol{f}$, $\nabla_{\boldsymbol{f}} k(\boldsymbol{e}, \boldsymbol{f}) = \boldsymbol{e}$,

$$2\boldsymbol{C}\boldsymbol{f} - 2\mu\boldsymbol{f} - \lambda\boldsymbol{e} = 0$$

Pre-multiplying by $\boldsymbol{e}^t$, since $\boldsymbol{e}^t\boldsymbol{f} = 0$, and $\boldsymbol{e}^t\boldsymbol{e} = 1$, we get

$$2\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} - \lambda = 0 \implies \lambda = 2\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f}$$

We note that, since $\boldsymbol{C}$ is a symmetric matrix, $\boldsymbol{C} = \boldsymbol{C}^t$. Let us denote the eigenvalue of $\boldsymbol{e}$ wrt $\boldsymbol{C}$ as γ; i.e, $\boldsymbol{C}\boldsymbol{e} = \gamma\boldsymbol{e}$. Taking the transpose of both sides, we get that $\boldsymbol{e}^t\boldsymbol{C}^t = \boldsymbol{e}^t\boldsymbol{C} = \gamma\boldsymbol{e}^t$. Now, using this in the expression of λ,

$$\lambda = 2\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} = 2\gamma\boldsymbol{e}^t\boldsymbol{f} = 0,$$

because $\boldsymbol{f}$ and $\boldsymbol{e}$ are assumed to be orthogonal. Hence, we have that $2\boldsymbol{C}\boldsymbol{f} = 2\mu\boldsymbol{f}$, meaning that $\boldsymbol{f}$ is an eigenvector of $\boldsymbol{C}$. Since we want to maximize $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f} = \boldsymbol{f}^t\mu\boldsymbol{f} = \mu$, we should choose the eigenvector with the largest value that is perpendicular to $\boldsymbol{e}$. Since $\boldsymbol{C}$ is a square symmetric positive semi-definite matrix, this would be the second largest eigenvalue. ■