CS 663, Fall 2023

Assignment 4

Jahanvi Rajput 23D0378
Badri Vishal Kasuba 22M2119
Abhishek Kumar Singh 22M2104

Question 3:

Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in 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 rank$(\boldsymbol{C}) > 2$. Extend the derivation to handle the case of a unit vector $\boldsymbol{g}$ which is perpendicular to both $\boldsymbol{e}$ and $\boldsymbol{f}$ which maximizes $\boldsymbol{g}^t\ \boldsymbol{C}\boldsymbol{g}$.

Solution:

Here, we have optimization problem which is given below:

$$\begin{aligned} \text{maximize} \quad & f^T C f \\ \text{subject to} \quad & f^T f = 1 \\ & f^T e = 0 \end{aligned} \tag{1}$$

Here, f is perpendicular to e implies $f^T e$ is equal to zero as.
Now applying method of Lagrange multipliers to solve this optimization problem.

$$\nabla_f \left(f^T C f - \lambda_1 (f^T f - 1) - \lambda_2 (f^T e)\right) = 0 \tag{2}$$

where λ_1 and λ_2 are Lagrange multipliers for the first and second constraints, respectively. Applying term by term gradient with respect to f.

$$\nabla_f (f^T f) = 2Cf$$

$$\nabla_f (f^T f - 1) = 2f$$

$$\nabla_f (f^T e) = e$$

Substituting these values in equation (2):

$$2Cf - 2\lambda_1 f - \lambda_2 e = 0 \tag{3}$$

$$\Rightarrow Cf = \lambda_1 f + \frac{\lambda_2}{2} e$$

Applying the transpose to both sides and multiplying e to the right:

$$f^T C^T e = \lambda_1 f^T e + \frac{\lambda_2}{2} e^T e$$

Noting that C is symmetric and e is an eigenvector (unit magnitude) of C with the highest eigenvalue:

$$C^T = C, \quad e^T e = 1, \quad Ce = \Lambda e$$

where Λ is the highest eigenvalue, and using the constraints, we get:

$$f^T \Lambda e = \lambda_1(0) + \frac{\lambda_2}{2} e^T e$$

$$f^T \Lambda e = 0 + \frac{\lambda_2}{2}$$

$$f^T \Lambda e = \frac{\lambda_2}{2}$$

$$\Lambda f^T e = \frac{\lambda_2}{2}$$

$$\frac{\lambda_2}{2} = \Lambda(0)$$

$$\lambda_2 = 2\Lambda(0)$$

$$\lambda_2 = 0$$

Substituting the value of λ_2 in equation (3):

$$Cf = \lambda_1 f$$

which implies that f is an eigenvector of C.

Now, $f^T C f = \lambda_1$, hence we want to maximize λ_1. λ_1 is not Λ (the highest eigenvalue) because the space of eigenvectors for which $C\nu = \Lambda\nu$ has a single dimension. This is because all non-zero eigenvalues of C are assumed to be different ($\Lambda \neq 0$ because $rank(C) > 2$), so Λ has algebraic multiplicity $= 1$, and the geometric multiplicity (dimension of that eigenspace) can never be greater than the algebraic multiplicity, which implies the dimension of the eigenspace of Λ is 1. This implies that there are no two orthogonal vectors in this space. So if $Cf = \Lambda f$, then e and f wouldn't be orthogonal. Hence, λ_1 has to be the second highest eigenvalue.

To extend the derivation to handle the case of a unit vector g that is perpendicular to both e and f and maximizes $g^T C g$, you can follow a similar approach with the additional constraint. The optimization problem will became:

$$\begin{aligned} \text{maximize} \quad & g^T C g \\ \text{subject to} \quad & g^T g = 1 \\ & g^T e = 0 \\ & g^T f = 0 \end{aligned} \tag{4}$$

Using Lagrange multipliers to find the maximum of $g^T C g$ subject to the constraints. The Lagrangian can be defined as:

$$\mathcal{L}(g, \lambda_1, \lambda_2, \lambda_3) = g^T C g - \lambda_1(g^T g - 1) - \lambda_2(g^T e) - \lambda_3(g^T f)$$

Here, λ_1, λ_2, and λ_3 are Lagrange multipliers.
To find the extremum, set the gradient of the Lagrangian with respect to g equal to zero:

$$\nabla_g \mathcal{L} = 2Cg - 2\lambda_1 g - \lambda_2 e - \lambda_3 f = 0$$

Solving for g, we get:

$$Cg = \lambda_1 g + \frac{\lambda_2}{2} e + \frac{\lambda_3}{2} f$$

Now, let's find the eigenvalue of C corresponding to the vector g. Taking the tranpose and multiply both sides of the eigenvalue equation by g:

$$g^T (C) g = \lambda_1 g^T g + \frac{\lambda_2}{2} e^T g + \frac{\lambda_3}{2} f^T g$$

$$g^T (C) g = \lambda_1 g^T g + \frac{\lambda_2}{2} (g^T e)^T + \frac{\lambda_3}{2} (g^T f)^T$$

$$g^T (C) g = \lambda_1 (1) + \frac{\lambda_2}{2} (0) + \frac{\lambda_3}{2} (0)$$

$$g^T C g = \lambda_1$$

$$Cg = \lambda g$$

Here, λ is the eigenvalue corresponding to g. Since we assumed that all non-zero eigenvalues of C are distinct and the rank$(C) > 2$, the second highest eigenvalue of C corresponds to g.

So, the direction g that is perpendicular to both e and f and maximizes $g^T C g$ is the eigenvector of C with the second highest eigenvalue.