1 Question 5

We need to maximize $\mathbf{a}^T C\mathbf{a}$, where C is the co-variance matrix. Using method of Lagrange multipliers we set the constraint that $\mathbf{a}^T\mathbf{a} = 1$ and write the objective function:

$$J(\mathbf{a}) = \mathbf{a}^T C\mathbf{a} - \lambda(\mathbf{a}^T\mathbf{a} - 1)$$

Taking derivative and setting to zero, we get:

$$C\mathbf{a} = \lambda\mathbf{a}$$

So solutions are restricted to the unit eigenvectors of the covariance matrix. Also, $\mathbf{a}^T C\mathbf{a} = \lambda$. Clearly, the eigenvector corresponding to the largest eigenvalue maximizes the value of $\mathbf{a}^T C\mathbf{a}$. Let's call this eigenvector e.

Since co-variance matrix is a symmetric matrix, all its eigenvalues must be real and all its eigenvectors must be orthogonal to each other. So all the eigenvectors apart from e are orthogonal to it.

Repeating the above maximization procedure with the additional constraint that the vector must be perpendicular to e simply eliminates e itself from the possible solutions, and we are left with the rest of the eigenvectors as possible solutions. Hence, the eigenvector of C with the second-highest eigenvalue is the required vector.

Mathematically, new objective function is:

$$\begin{aligned} J(\mathbf{f}) &= \mathbf{f}^T C\mathbf{f} - \lambda(\mathbf{f}^T\mathbf{f} - 1) - \mu\mathbf{f}^T\mathbf{e} \\ J'(\mathbf{f}) &= 2C\mathbf{f} - 2\lambda\mathbf{f} - \mu\mathbf{e} \end{aligned}$$

Setting derivative to zero:

$$\begin{aligned} 2C\mathbf{f} - 2\lambda\mathbf{f} - \mu\mathbf{e} &= 0 \\ 2\mathbf{f}^T C\mathbf{f} - 2\lambda\mathbf{f}^T\mathbf{f} - \mu\mathbf{f}^T\mathbf{e} &= 0 \\ 2\mathbf{f}^T C\mathbf{f} - 2\lambda &= 0 \\ \implies \mathbf{f}^T C\mathbf{f} &= \lambda \end{aligned}$$

Alternatively,

$$\begin{aligned} 2C\mathbf{f} - 2\lambda\mathbf{f} - \mu\mathbf{e} &= 0 \\ 2\mathbf{e}^T C\mathbf{f} - 2\lambda\mathbf{e}^T\mathbf{f} - \mu\mathbf{e}^T\mathbf{e} &= 0 \\ 2\mathbf{e}^T C\mathbf{f} - \mu &= 0 \end{aligned}$$

Since $\mathbf{e}^T C\mathbf{f}$ is a scalar, it doesn't matter if we take its transpose. So we have $\mathbf{e}^T C\mathbf{f} = \mathbf{f}^T C\mathbf{e} \propto \mathbf{f}^T\mathbf{e} = 0$ This means that $\mu = 0$ and hence $C\mathbf{f} = \lambda\mathbf{f}$.