5) Istead of projecting z_i's onto a single $\bar{e}$, lets project it onto two $\bar{e}_a$ & $\bar{e}_b$'s - both are orthonormal to each other.

(Here $z_i = x_i - \bar{x}_i$)

Consider,

$$\hat{z}_i = (z_i \cdot e_a) e_a + (e_b \cdot z_i) e_b$$

Our goal is to find the $\hat{z}_i$ that minimize the MSE

$$MSE = \sum_{i=1}^{N} \|z_i - \hat{z}_i\|^2 \times \frac{1}{N-1}$$ (let there be N samples & let $z_i \in \mathbb{R}^d$)

$$MSE = \frac{1}{N-1} \sum_{i=1}^{N} \left(z_i - (z_i \cdot e_a) e_a - (z_i \cdot e_b) e_b\right) \cdot \left(z_i - (z_i \cdot e_a) e_a - (z_i \cdot e_b) e_b\right)$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} \|z_i\|^2 - (z_i \cdot e_a)^2 - (z_i \cdot e_b)^2$$

$$= \left(\frac{1}{N-1} \sum_{i=1}^{N} \|z_i\|^2\right) - \left(\frac{1}{N-1} \sum_{i=1}^{N} (z_i \cdot e_a)^2\right) - \frac{1}{N-1}\left(\sum_{i=1}^{N} (z_i \cdot e_b)^2\right)$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} \|z_i\|^2 - \frac{1}{N-1}\left(e_a^T Z Z^T e_a\right) - \frac{1}{N-1}\left(e_b^T Z Z^T e_b\right)$$

For minimizing MSE, we need to maximise

$$e_a^T \frac{\left(Z Z^T\right)}{N-1} e_a + \frac{e_b^T \left(Z Z^T\right) e_b}{N-1}$$

$$= \underbrace{e_a^T (C) e_a}_{T_1} + \underbrace{e_b^T (C) e_b}_{T_2}$$

Because each of the terms is independent, we need to maximize them seperately.

$$T_1 = e_a^T C e_a \quad , \quad T_2 = e_b^T C e_b$$

We would like to impose restrictions like $\|e_a\| = 1$ & $\|e_b\| = 1$
So, we can use the lagrange multiplier method

$$\mathcal{L}_{e_1}(e_a, \lambda) = T_1 - \lambda_1(e_a^T e_a - 1) \qquad [e_a^T e_a = \|e_a\|^2 = 1]$$
is the constraint

$$\mathcal{L}_{e_2}(e_b, \lambda) = T_2 - \lambda_2(e_b^T e_b - 1) \qquad [e_b^T e_b = \|e_b\|^2 = 1$$ is the constraint]

$$\frac{\partial \mathcal{L}_{e_1}}{\partial \lambda_1} = -(e_a^T e_a - 1) \qquad \frac{\partial \mathcal{L}_{e_2}}{\partial \lambda_2} = -(e_b^T e_b - 1)$$

$$\frac{\partial \mathcal{L}_{e_1}}{\partial e_a} = 2Ce_a - 2\lambda_1 e_a \qquad \frac{\partial \mathcal{L}_{e_2}}{\partial e_b} = 2Ce_b - 2\lambda_2 e_b$$

$$\Rightarrow \quad e_a^T e_a = \|e_a\|^2 = 1 \qquad e_b^T e_b = \|e_b\| = 1$$

$$Ce_a = \lambda_1 e_a \qquad Ce_b = \lambda_2 e_b$$

So, e_a & e_b are both eigen vectors of 'C' matrix

$$T_1 = \lambda_1(e_a^T e_a) \qquad T_2 = \lambda_2(e_b^T e_b)$$
$$= \lambda_1 \qquad\qquad = \lambda_2$$

So, if we want to maximize $T_1 + T_2$, we need to choose λ_1 & λ_2 values in such a way they are the largest two eigen values.

Hence, in the question f will be the eigen vector with the second highest eigen value