Q2.) $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ where $x_i \in \mathbb{R}^d$

$\bar{x}$ = average vector $= \frac{1}{N}\sum_{i=1}^{N}(x_i)$

C = Covariance Matrix $= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x_i})(x_i - \bar{x})^T$

The unit vector $\hat{e}$ across which reconstruction loss is minimized is obtained by maximizing $e^t C e$.

We have seen that e is the eigenvector of C with the largest eigen value.

Now, we have another unit vector f such that

$f^T e = e^T f = 0$ (normal to each other)

and $f^t C f$ is maximum.

The 'f' satisfying these criteria is the eigenvector of C with second highest eigenvalue.

Assumption: ① All non-zero eigen values of C are distinct

② rank (C) > 2

proof: rank (C) > 2 implies that we have at least 2 non-zero eigen values [can be deduced from rank-nullity theorem]

Now, we need to maximize $J(f) = f^T C f$ given

$f^T f = 1$ and $f^T e = 0$

Lagrangian:

$$\tilde{J}(f) = f^T C f - \lambda_1 (f^T f - 1) - \lambda_2 (f^T e)$$

$$\Rightarrow \frac{\partial \tilde{J}(f)}{\partial f} = 0$$

$$\Rightarrow \frac{\partial}{\partial f}(f^T C f) - \lambda_1 \frac{\partial}{\partial f}(f^T f) - \lambda_2 \frac{\partial}{\partial f}(f^T e) = 0$$

$$\Rightarrow (C + C^T) f - 2\lambda_1 f - \lambda_2 e = 0$$

Since C is symmetric, $C^T = C$

$\Rightarrow \quad 2Cf - 2\lambda_1 f - \lambda_2 e = 0$

$\Rightarrow \quad Cf = \lambda_1 f + \frac{\lambda_2}{2} e$ —— ①

$\Rightarrow \quad f^T C f = \lambda_1 f^T f + \frac{\lambda_2}{2} f^T e$ (pre-multiply by f^T)

$\Rightarrow \quad f^T C f = \lambda_1$ —— Ⓘ

Premultiply ① by e^T :

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

$\Rightarrow \quad e^T C f = 0 + \frac{\lambda_2}{2}$

$\Rightarrow \quad f^T C e = \frac{\lambda_2}{2}$ (take transpose on both sides)

Since e is an eigenvector of C, let $Ce = \lambda^* e$ where λ^* is the maximum non-negative eigenvalue of covariance matrix C

$\Rightarrow \quad f^T \lambda^* e = \lambda_2/2$

$\Rightarrow \quad \lambda^* f^T e = \frac{\lambda_2}{2}$

$\Rightarrow \quad \lambda^* . 0 = \frac{\lambda_2}{2}$

$\therefore \boxed{\lambda_2 = 0}$

putting $\lambda_2 = 0$ in equⁿ ①, we get

$Cf = \lambda_1 f$

$\therefore$ f is an eigen-vector of C

Also, $f^T C f = \lambda_1$

So, to maximize $f^T C f$, we must choose the largest value of λ_1 (eigen-value of C).

Let $\lambda_1 = \lambda^*$ (eigenvalue corresponding to e)

But then both e and f have same eigenvalues.

$\Rightarrow$ $Cx = \lambda^* x$ has ~~to~~ at least two orthonormal (3)
solutions e and f

$\Rightarrow$ dimension of the eigen-space of eigenvalue λ^*
(i.e., the null-space of matrix $C - \lambda^* I$) is at least 2.

Since λ^* is non-zero, it is unique (given)

$\Rightarrow$ Algebraic multiplicity of $\lambda^* = 1$

Since Geometric multiplicity $\leq$ Algebraic multiplicity,
we can't have two different solutions to
$Cx = \lambda^* x$.

Hence, we can't choose $\lambda_1 = \lambda^*$.

Naturally, the second-best option is to choose
the second-highest eigen value of C.

$\text{rank}(C) > 2$ guarantees the existence of such a non-zero eigen-value.

$\therefore$ f is an eigen-vector of C with second highest eigen value λ^{**} (proved). $\Rightarrow f^t C f = \lambda^{**} \neq 0$

<u>Corollary</u>:- when we have a third unit-vector g normal to both e and f and we need to maximize $g^t C g$.

$g^t g = 1$; $g^t f = 0$; $g^t e = 0$

$\Rightarrow$ $\tilde{J}(g) = g^t C g - \lambda_1 (g^t g - 1) - \lambda_2 (g^t f) - \lambda_3 (g^t e)$

$\Rightarrow$ Lagrangian:

$$\frac{\partial \tilde{J}(g)}{\partial g} = 0$$

$$2Cg - 2\lambda_1 g - \lambda_2 f - \lambda_3 e = 0 \quad - \text{(i)}$$

$\Rightarrow$ premultiply by g^t:

$$2 g^t C g - 2\lambda_1 g^t g - \lambda_2 g^t f - \lambda_3 g^t e = 0$$

$$2 g^t C g - 2\lambda_1 = 0$$

$\Rightarrow$

$$\boxed{g^t C g = \lambda_1} \quad - \text{(ii)}$$

Premultiply eqn ① by f^t:

$$2 f^t C g - \lambda_2 = 0$$

$$\Rightarrow f^t C g = \frac{\lambda_2}{2}$$

$$\Rightarrow g^t C f = \frac{\lambda_2}{2}$$

$$\Rightarrow g^t \lambda^{**} f = \frac{\lambda_2}{2}$$

$$\Rightarrow \lambda^{**} g^t f = \lambda_2/2$$

$$\Rightarrow \boxed{\lambda_2 = 0}$$

Similarily, by premultiplying eqn ① by e^t, we get

$$\boxed{\lambda_3 = 0}$$

$$\therefore 2 C g - 2\lambda_1 g = 0 \quad \text{(from ①)}$$

$$\Rightarrow \boxed{Cg = \lambda_1 g}$$

$\Rightarrow$ g is an eigenvector of covariance matrix C

To maximize $g^t C g = \lambda_1$, we need to choose the maximum / largest eigen value of C.

letting $\lambda_1 = \lambda^*$ or $\lambda_1 = \lambda^{**}$ would lead to geometric multiplicity of λ^* or λ^{**} to exceed 2.

This is not valid since the eigen values are distinct with algebraic multiplicity $= 1$.

Hence, $\lambda_1 = \lambda^{***}$ which is the 3rd largest ~~eigenvector~~ eigenvalue of C. (proved)

By the same logic, we can project $\overline{x_i}$ into K orthonormal directions where $K < d$.