(Page-I)

* So, the principal component for 'm' data points is to find a direction 'w'

Where variance of the data along that direction is maximized.

→ Also, the each weight to be meaningful is constrained $\|w\| \leq 1$

$\boxed{\text{Variance}}$ of all points →

$$\boxed{\frac{1}{m}\sum_{i=1}^{m}\left(w^T x^i - w^T \mu\right)^2} \quad — (1)$$

Constraint → ~~max~~ $w;\ \|w\| < 1$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(w^T\left(x^i - \mu\right)\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m} w^T\left(x^i-\mu\right)\left(x^i-\mu\right)^T w$$

{ from linalg - $a^T b = b^T a$ }

$$= w^T \underbrace{\left(\frac{1}{m}\sum_{i=1}^{m}\left(x^i-\mu\right)\left(x^i-\mu\right)^T\right)}_{\text{Covariance Matrix } C} w \quad — (2)$$

So the optimization problem becomes -

$$\boxed{\max_{w:\ \|w\| \leq 1} w^T C w} \quad — (3)$$

(Page - 2)

For, constrained optimization, can be solved by "Lagrangian function" —

$$L(w, \lambda) = \underbrace{w^T C w}_{\text{original optimization}} + \lambda \big(\overbrace{1 - \|w\|^2}^{\text{constraint}}\big) \quad (4)$$

λ → Lagrangian multiplier

Which implies —

$$\frac{\partial L}{\partial w} = 0 = 2Cw + 2\lambda w \iff \boxed{Cw = \lambda w} \quad (5)$$

Which is in definition → w is eigen-vector & λ is eigen-value.

→ So, the objective function is λ →

$$w^T C w = \lambda w^T w = \underbrace{\lambda}_{\text{"eigen value"}} \|w\|^2 \quad (6)$$

So, to maximize the above, we need to find largest eigenvalue of C (covariance matrix).