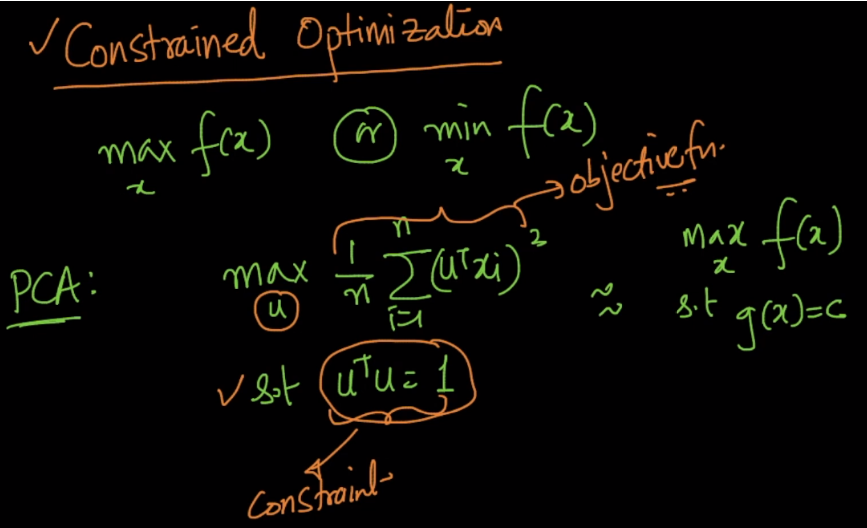
**Constrained Optimization:**

Whenever there is some constraint associated with objective function, then optimization of such function is said to be constrained optimization.

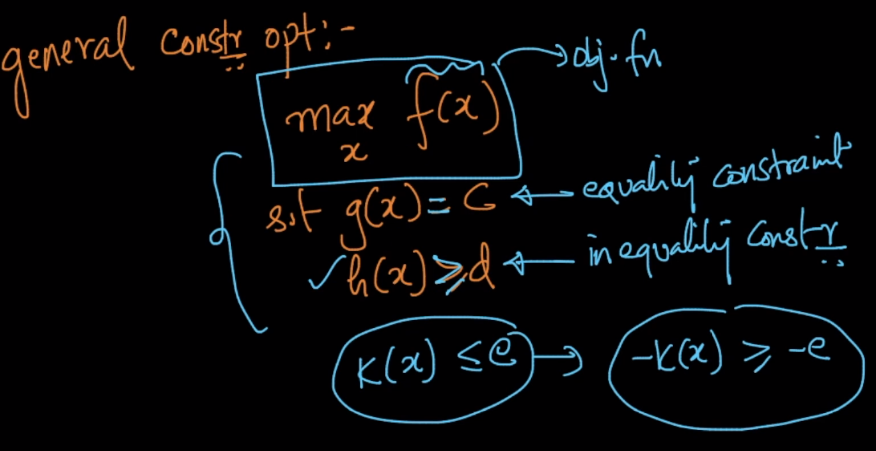
Example for PCA objective function we have constraint given in below image, therefore performing optimization on such function is said to be constraint optimization.



**General Constraint Optimization:**

Generally the constraint optimization has following things:

* An Objective Function
* Equality Constraint
* Inequality Constraint



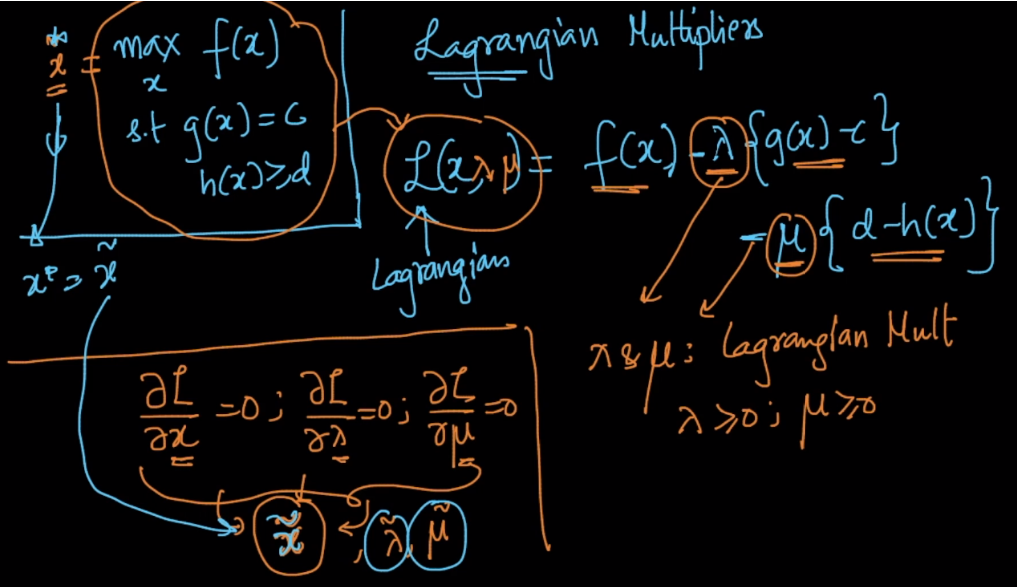
**How to find optimal value in case of Constrained Optimization:**

We use **Lagrangian Multipliers** which are always positive.

We’ll form an equation using objective function, and lagrangian multiplier multiplied to constraints.

Now we partial differentiation of this equation wrt x, and all lagrangian multipliers.

**Then the value we get while doing partial differentiation wrt x will be the optimal x value where the objective function will be minima or maxima.**



**How we find optimal value of u in PCA, for getting maxima of objective function.**

(UTxi)2can be written as UTSU, such that UTU = 1, where S is covariance matrix, when x is mean centric or standardized.

Now our objective function is to find U where UTSU is maxima, along with constraint UTU = 1.

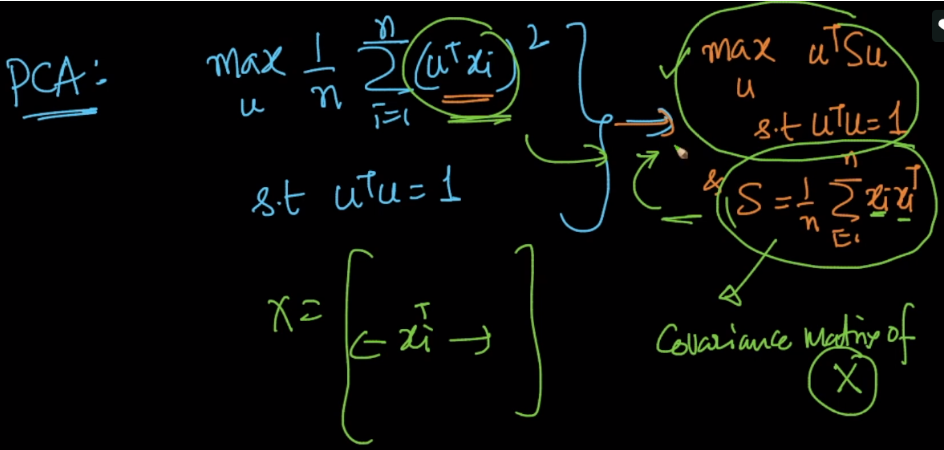
We make new equation given in below images.

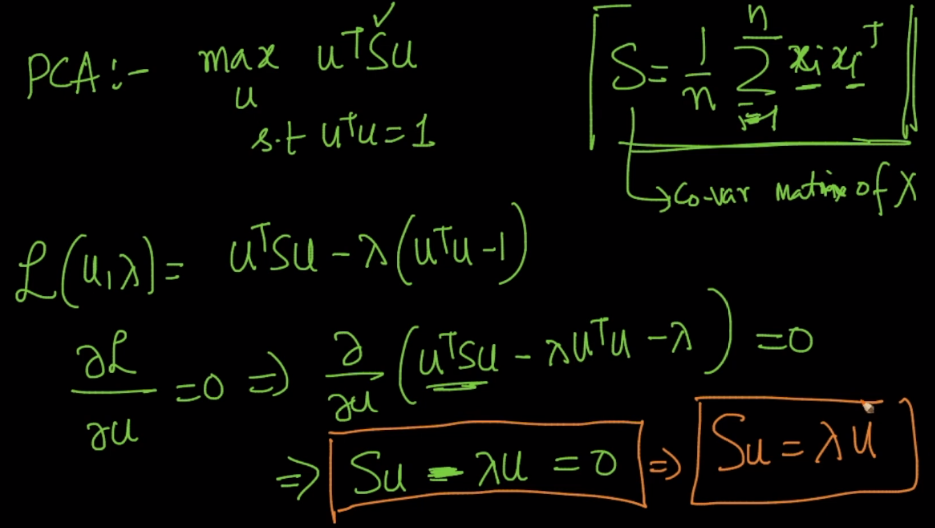
We’ll do parital differentiation wrt U. And we get SU = lambda U.

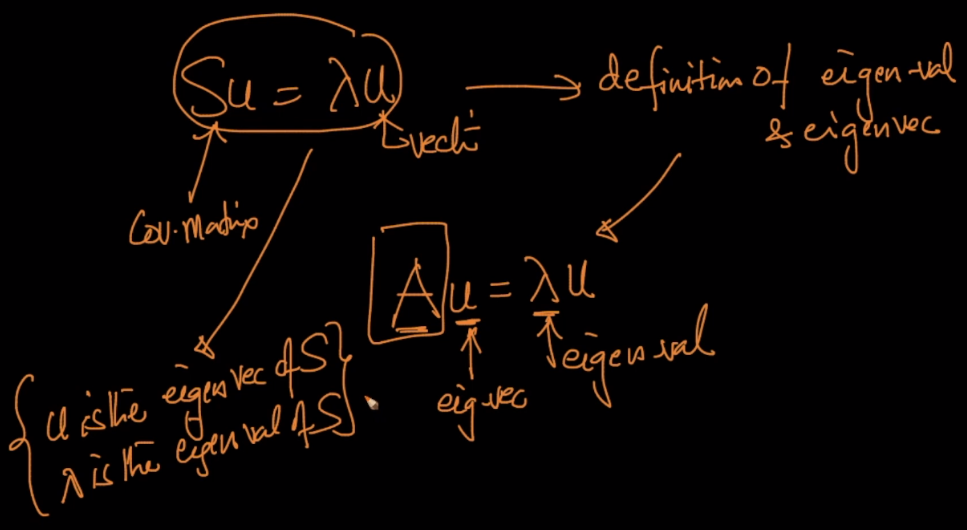
And when such equation come we say that U is eigen value and lambda is eigen vector.

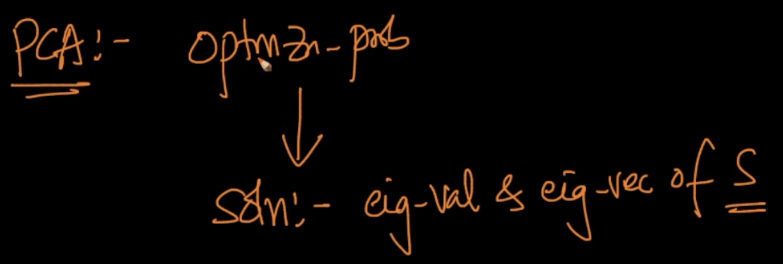
And since in PCA eigen vectors represent the vector where we’ll get max variance by applying GD on new equation we formed using objective function and constraints.

As we know that eigenvector corresponding to the maximal-eigenvalue(lambda) maximizes U\_T\*S\*U.









**How are we getting maximal lambda using this solution:**

Above we had:

Objective fun: armax UTSU

We get after differentiating eq: SU = lamdaU.

Therefore UTSU becomes:

* UT (lambda U)
* Lambda (UTU) {since lambda is scalar}
* Lambda { UTU = 1}

Hence solving above we were eventually getting lambda, and the largest lambda value will be associated to largest vector preserving maximum variance.

