Ashutosh Kumar Singh 19CS30008

4. Least Squares Problem :-

$$\min_{x \in R^n} \|Ax - b\|_2^2$$

Given that the columns of $A$ are linearly independent.

So, we can write :-

$$Ax = x_1 \begin{bmatrix} A_1 \\ \\ \end{bmatrix} + x_2 \begin{bmatrix} A_2 \\ \\ \end{bmatrix} + \cdots + x_n \begin{bmatrix} A_n \\ \\ \end{bmatrix}$$

where $A_i$ is the $i^{th}$ column of $A$.

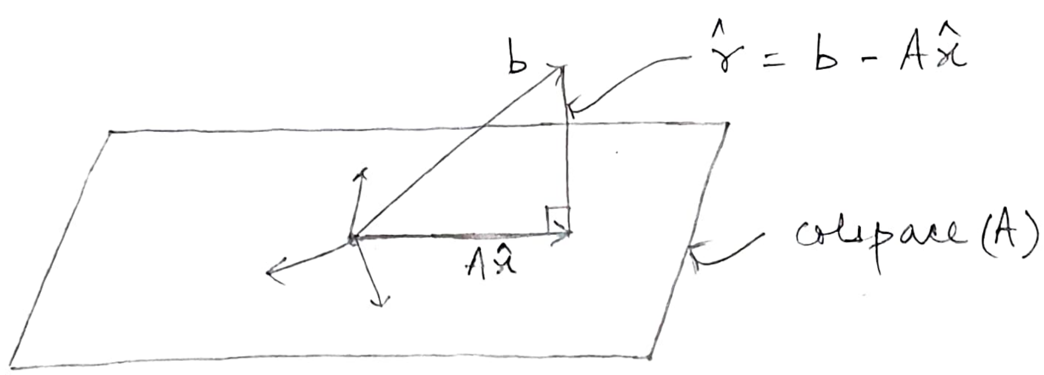
and we want to find the $x$ that minimizes

$$\|x_1 A_1 + x_2 A_2 + \cdots + x_n A_n - b\|_2^2$$

Now, we know that colspace $(A)$ is the set of all vectors of the form $Ax$.

So, geometrically, the colspace $(A)$ formed by all vectors $Ax$ can be considered as a plane.

Now, the least squares solution $\hat{x}$ will be that vector for which the vector $A\hat{x}$ is closest to the vector $b$, in other words, $A\hat{x}$ is the linear combination of the columns of $A$ that is closest to $b$.



Geometrically, we can see that the closest vector $A\hat{x}$ to $b$ is the orthogonal projection of $b$ onto colspace $(A)$.
This is the geometrical interpretation of the least squares problem.

---

Now, the optimal residual $\hat{r} = A\hat{x} - b$ satisfies a property known as the orthogonality principle.
It is orthogonal to the the columns of $A$, therefore it is orthogonal to any linear combination of the columns of $A$, hence, it is orthogonal to colspace $(A)$.
This orthogonality principle can also be written as:-

$$A^T(A\hat{x} - b) = 0$$

$$\Rightarrow A^T A\hat{x} = A^T b.$$

These in fact, are the normal equations, and we can understand that the name normal equations arises from the orthogonal (normal) principle of the optimal residual vector.

---

When the matrix $A$ does not have linearly independent columns, then we can say that the least squares problem $Ax = b$ has infinitely many solutions.
The least squares solution $\hat{x}$ is such that $\|A\hat{x} - b\|_2^2$ is minimum for all $x$.

Now, if the columns of $A$ are linearly dependent, then, we can find a vector $y$ such that $Ay = 0$.

Now, we claim that any vector of the form $(\hat{x} + \lambda y)$ will also be a least squares solution to $Ax = b$, because,

$$\|A(\hat{x} + \lambda y) - b\|_2^2$$

$$= \|A\hat{x} + \lambda Ay - b\|_2^2$$

$$= \|A\hat{x} - b\|_2^2 \quad (\text{as } Ay = 0),$$

which is the minimum possible value since we already know that $\hat{x}$ is a least squares solution.

Hence, we can say that when the columns of $A$ are linearly dependent, the least squares solution to $Ax = b$ has infinitely many solutions.