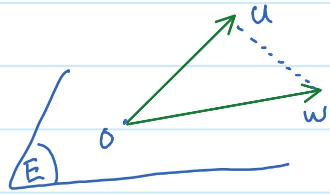
Lecture 25

Friday, March 6, 2020

Last time, we discussed an example where inner product can be used to solve an optimization problem. This is possible because inherent in the inner product is a "minimizing property": the distance between a vector $u \in V$ and a vector w in a subspace $E \subset V$ is smallest when w is the projection of u onto E. In other words, if we put



$$v = \mathrm{proj}_E u$$

then

$$\|u - v\| = \min_{w \in E} \|u - w\|.$$

When we say "distance between two vectors", we mean the distance between the tips of the vectors (assuming they are based at the origin).

Let us consider more examples of using inner product to solve an optimization problem.

Ex:

The equation $3x = 4$ has a unique solution $x = 4/3$.

The equation $4x = 5$ has a unique solution $x = 5/4$.

The system $\begin{cases} 3x = 4 \\ 4x = 5 \end{cases}$ therefore has no solutions.

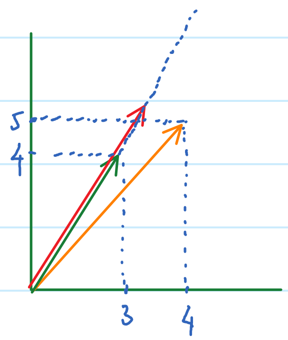
However, if we are to find the "best approximate solution" to this system, what could it be? Could it be one of $4/3$ and $5/4$, or the average of them, or something else? We need a method to quantify how "good" a solution candidate is. The least square method, which some of us may have seen before, proposes that the best solution is the one that minimizes the following quantity:

$$(3x-4)^2 + (4x-5)^2$$

This is not the only way to quantify the goodness of a candidate. However, this method ~~is~~ has a nice connection to inner product. It is equivalent to an intuitive geometric interpretation as follows. [We will see why they are equivalent to each other in the next example.]

The system of equation can be written in matrix form as

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} x = \begin{bmatrix} 4 \\ 5 \end{bmatrix}.$$



When x varies in $\mathbb{R}$, the vector $\begin{bmatrix} 3 \\ 4 \end{bmatrix} x$ varies on the line containing vector $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$. The most reasonable choice of x seems to be the one such that $\begin{bmatrix} 3 \\ 4 \end{bmatrix} x$ is closest to $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$. This happens when $\begin{bmatrix} 3 \\ 4 \end{bmatrix} x$ is equal to the projection of $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$ onto the line spanned by $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

$$\operatorname{proj}_{\{[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}]\}} \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \frac{\left(\begin{bmatrix} 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right)}{\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right)} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{32}{25} \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Thus, $x = 32/25$.

Ex: Let A be an $m \times n$ matrix and b be an $m \times 1$ column vector. Find the best solution $X \in F^n$ to the matrix equation $AX = b$. [Here F can be $\mathbb{R}$ or $\mathbb{C}$.]

If the equation $AX = b$ has a solution then X should be an exact solution. The issue comes when the equation has no solutions. If we write

$$A = \begin{bmatrix} \text{---} \; u_1 \; \text{---} \\ \text{---} \; u_2 \; \text{---} \\ \cdots \\ \text{---} \; u_m \; \text{---} \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

then matrix equation $AX = b$ is equivalent to a linear system of m equations

$$\begin{cases} u_1 X = b_1 \\ u_2 X = b_2 \\ \quad \cdots \\ u_m X = b_m \end{cases}$$

This system may be inconsistent (i.e. having no solutions), for example when $m > n$. We can ask what is the best candidate for solution. The least square method proposes that the best candidate is the one that minimizes the expression

$$(u_1 X - b_1)^2 + (u_2 X - b_2)^2 + \dots + (u_m X - b_m)^2. \qquad (*)$$

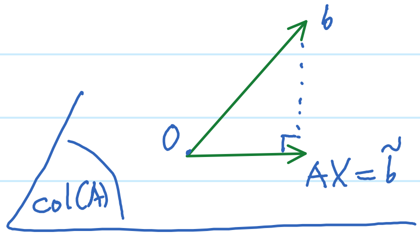
From an analytic point of view, one can rewrite this expression as a multivariable function $f(x_1, x_2, \dots, x_n)$ where $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and then try to solve the system $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$.

However, this approach is not of our interest here. We notice that

$$AX - b = \begin{bmatrix} \text{---}\; u_1 \;\text{---} \\ \text{---}\; u_2 \;\text{---} \\ \vdots \\ \text{---}\; u_m \;\text{---} \end{bmatrix} \begin{bmatrix} | \\ X \\ | \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} u_1 X - b_1 \\ u_2 X - b_1 \\ \vdots \\ u_m X - b_1 \end{bmatrix}$$

Thus, the expression $(*)$ is nothing but $\|AX - b\|^2$.

In other words, we are trying to find $X \in F^n$ that minimizes the norm $\|AX - b\|$. If X varies freely in F^n then vector AX will roam in the column space of A. The column space of A is the same as the range of the map $f(X) = AX$.



Therefore, we are searching for a vector $\tilde{b}$ in col(A) that is closest to b. Then X will be found by solving the equation $AX = \tilde{b}$. In short, to find $X \in F^n$ that is the best solution to $AX = b$, we follow two steps:

1) Find $\text{proj}_{col(A)} b$. Call it $\tilde{b}$.

2) Solve for (exact) solution X from the equation $AX = \tilde{b}$.

There is a trick that helps simplify this procedure. The trick comes from the following observations: write

$$A = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

Then $col(A) = \text{span}\{v_1, v_2, \dots, v_n\}$ by the definition of column spaces. We want to find $X \in F^n$ such that $\|AX - b\|$ is minimized. That is to find X such that $AX - b \perp col(A)$. That is to find X such that

$$AX - b \perp v_1, v_2, \dots, v_n.$$

That is to find X such that

$$\begin{cases} (AX - b, v_1) = 0 \\ (AX - b, v_2) = 0 \\ \quad \cdots\cdots \\ (AX - b, v_n) = 0. \end{cases} \qquad (**)$$

Recall that the natural inner product in $\mathbb{R}^m$ or $\mathbb{C}^m$ is given as

$$(x, y) = \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \right) = x_1 \overline{y_1} + x_2 \overline{y_2} + \cdots + x_m \overline{y_m},$$

which can be written as

$$(x, y) = \begin{bmatrix} \overline{y_1} & \overline{y_2} & \cdots & \overline{y_m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = y^* x$$

Here y^* denote $\overline{y^T}$ (taking the transpose of y, then take the conjugate of each entry). One can now rewrite the system (**) as

$$\begin{cases} v_1^* (AX - b) = 0 \\ v_2^* (AX - b) = 0 \\ \cdots\cdots \\ v_n^* (AX - b) = 0. \end{cases} \quad \text{or equivalently} \quad \begin{cases} v_1^* AX = v_1^* b \\ v_2^* AX = v_2^* b \\ \cdots\cdots \\ v_n^* AX = v_n^* b \end{cases}$$

We can combine these equations into a single matrix equation as

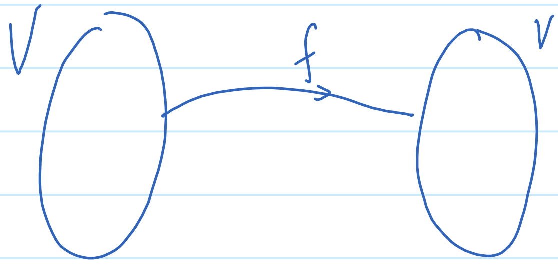
$$\underbrace{\begin{bmatrix} \text{---} & v_1^* & \text{---} \\ \text{---} & v_2^* & \text{---} \\ & \vdots & \\ \text{---} & v_n^* & \text{---} \end{bmatrix}}_{A^*} AX = \underbrace{\begin{bmatrix} \text{---} & v_1^* & \text{---} \\ \text{---} & v_2^* & \text{---} \\ & \vdots & \\ \text{---} & v_n^* & \text{---} \end{bmatrix}}_{A^*} b$$

which is $A^* A X = A^* b$.

In conclusion, the trick to find X that minimizes $\|AX - b\|$ is: solve for the exact solution X of the equation $A^* A X = A^* b$.

* Operators on inner product spaces:

Given a vector space V, there are of course many maps that go from V to V. Linear maps are a special type of maps that is compatible with the structure of vector spaces.



For example, summing in V and then taking to V through f is the same as taking through f and then summing in V. The same is true for scaling.

Now let V be an inner product space. One may ask: what kind of maps are compatible with the inner product on V? There is a special type of maps that preserve the inner product.

Def: Let V be an inner product space. A map $f: V \to V$ is said to be unitary if it is inner-product preserving, namely

$$(f(u), f(v)) = (u, v) \quad \forall\, u, v \in V.$$