MAT 216

Linear Algebra and Fourier Analysis

Lecture 14

Prepared by : Zadid Hasan (MZH)

Reference Book:

- Introduction to Linear Algebra, 5th Edition by Gilbert Strang.

natbib graphicx

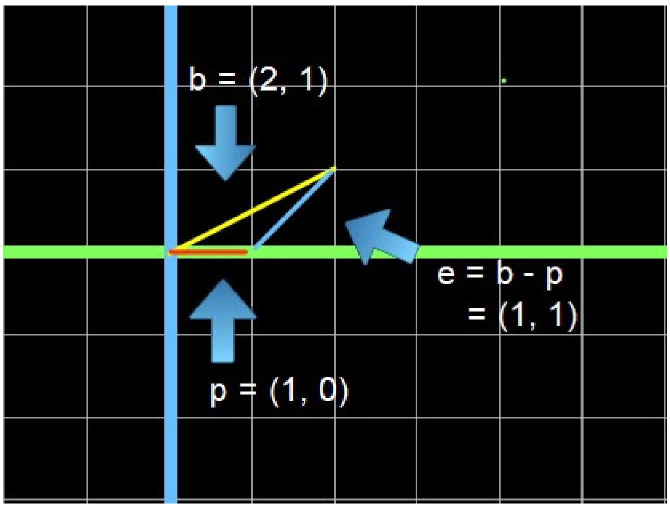
Table of Contents

Projection

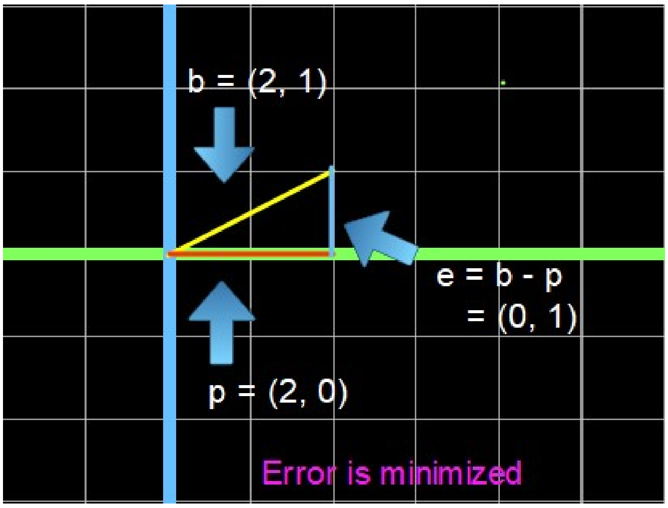
Definition: A projection is a linear transformation P from a vector space to itself such that $P^n = P$ for any $n \geq 1$. That is, whenever P is applied more than once to any value, it gives the same result as if it were applied once (idempotent).
Motivation: In Linear Algebra, we spend a lot of time trying to solve a system of linear equation $Ax = b$. But this equation is solvable only when $b \in col(A)$. If b is not in the column space of A, we can't have a solution. But in some cases, we only want the vector in $col(A)$ that is closest to b. That vector is the orthogonal projection of b onto the column space of A. As we learn about projection, we will gradually understand what this means.

Projection onto a line

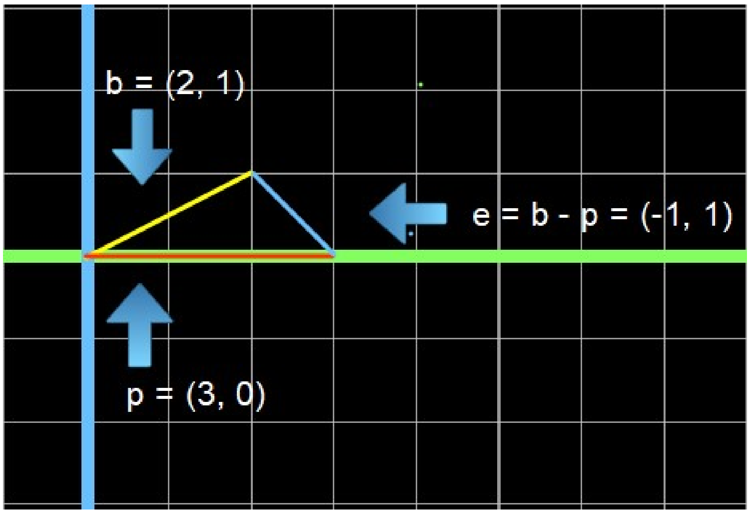
There are oblique projections, orthogonal projections. But, in this lecture when we say, "projection", we only mean "orthogonal projection". Suppose, we want to solve $Ax = b$ and $A = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \end{bmatrix}$.
So, there is only 1 column in A. And the column space of A is the span of the vector $(1, 0)$ which is the x-axis and any vector in that space will be of the form $(x, 0)$. There is no way we can get a solution for this equation where $b = (2, 1) \neq (x, 0)$. But we may take different vectors p along the x-axis and see if the error $e = b - p$ is minimized.
Case 1: p = (1, 0):



Case 2: p = (2, 0):



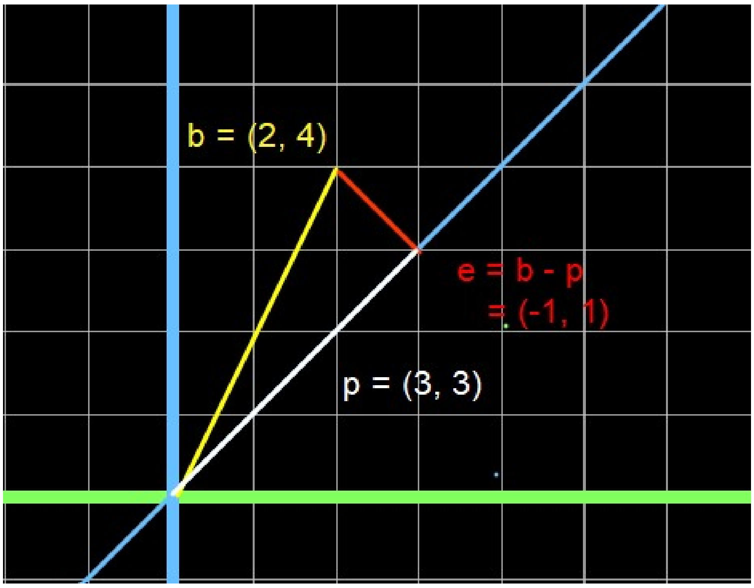
Case 3: p = (3, 0):



From these examples, we can see that the error is minimized when the error vector e is orthogonal(perpendicular in easy English) to the column space of A. This particular projection vector p is called the orthogonal projection of b onto $col(A)$.

Now, let's say we want to solve a different equation where $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. To find the vector p on $col(A)$ which is basically the line $x - y = 0$, that minimizes $e = b - p$, we drop a perpendicular from the tip of the vector b onto the line $x - y = 0$. That point is the tip of the vector p. In the next page, we can see that $p = (3, 3)$. So, now let's derive equations to find this p. Since p is along the only column a of A, so for some scalar $\hat{x}$, $p = a\hat{x}$. Since, the error vector $e = b - p = b - a\hat{x}$ is perpendicular to the column space of A, so it is perpendicular to a. So, we get,

$$a^T(b - a\hat{x}) = 0$$
$$a^Tb = a^Ta\hat{x}$$
$$\hat{x} = \frac{a^Tb}{a^Ta}$$



If we use this formula for the example in the diagram, $\hat{x} = \dfrac{\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}}{\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}}$

$$\hat{x} = \frac{2+4}{1+1} = 3$$

So, $p = a\hat{x} = \begin{bmatrix}1 & 1\end{bmatrix} 3 = \begin{bmatrix}3 & 3\end{bmatrix}$
Now, if we want to find the matrix that transformed b to p,
$p = a\hat{x} = \frac{aa^T}{a^Ta}b = Mb$

So, $M = \frac{aa^T}{a^Ta} = \dfrac{\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}}{\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$ is the projection matrix for this transformation.

Projection onto a plane

Suppose, we have a 3-D vector $\begin{bmatrix}2\\3\\4\end{bmatrix}$. If we wanted to project this vector onto the z-axis, then the projection matrix would be $P_z = \begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}$. And If we wanted to project this vector onto the xy-plane, the projection matrix would be $P_{xy} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$. So, $P_z + P_{xy} = I$. Actually this holds true for any two orthogonal sub-spaces.

Now, let's find the projection of the vector $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ onto the sub-space spanned by the following two vectors $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$. In the next section, we will derive a formula for the projection of a vector onto a sub-space which is $P = A(A^TA)^{-1}A^T$.
The projection matrix should be,
$P = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{9} & \frac{-4}{9} \\ \frac{-4}{9} & \frac{5}{9} \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = P_{xy}.$
This was meant to be. Because the column vectors of A, they both lie on the xy-plane. So, projecting onto the subspace spanned by them is equivalent to projecting a vector on the xy-plane.

Projection onto a sub-space

Suppose we have a special vector b and we have n linearly independent vectors $a_1, a_2, \cdots, a_n$, all from $\mathbb{R}^m$. Now we want to find a combination $p = \hat{x_1}a_1 + \hat{x_2}a_2 + \cdots + \hat{x_n}a_n = A\hat{x}$ where. $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$ such that the error $e = b - A\hat{x}$ is minimized.
As we have learned earlier, the error will be minimized only when the error vector is perpendicular to all vectors in $col(A)$. So, it will be perpendicular/orthogonal to the columns of A as well. So,

$$\begin{gathered} \mathrm{a}_1^T(b - A\hat{x}) = 0 \\ \mathrm{a}_2^T(b - A\hat{x}) = 0 \\ \cdots \\ \mathrm{a}_n^T(b - A\hat{x}) = 0 \end{gathered}$$

And, this can be written in the matrix form, like this,

$$\begin{gathered} \begin{bmatrix} a_1^T \\ a_2^T \\ \cdots \\ a_n^T \end{bmatrix} (b - A\hat{x}) = 0. \\ \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T (b - A\hat{x}) = 0. \\ A^T(b - A\hat{x}) = 0. \\ A^Tb - A^TA\hat{x} = 0. \\ \hat{x} = (A^TA)^{-1}A^Tb \end{gathered}$$

So, the projection is $p = A\hat{x} = A(A^TA)^{-1}A^Tb = Pb$
And the projection matrix is $P = A(A^TA)^{-1}A^T$

The Gram Matrix

A^TA is called the Gram matrix of A. In the previous section, we have written, $P = A(A^TA)^{-1}A^T$.
But, can we really assume that A^TA is invertible and it is okay to write $(A^TA)^{-1}$? Since, we assumed that the columns of A are linearly independent, we can.
Theorem: A and A^TA have the same null space.
Proof:

1. First we assume $Ax = 0$. If we multiply A^T on both sides, we get, $A^TAx = 0$. So, any vector from the null space of A belongs to the null space of A^TA.

2. Then, let $A^TAx = 0$. We can multiply both sides by x^T and get, $x^TA^TAx = 0$. which can also be written as $(Ax)^TAx = 0$., so $||Ax||^2 = 0$, so $Ax = 0$. This means that, any vector from the nullspace of A^TA also belongs to the null space of A.

Since, we assumed that, the column vectors of A are linearly independent, that means $x = 0$ is the only vector in the null space of A. So, $x = 0$is also the only vector in the null space of A^TA. And since, A^TA is a square matrix and $A^TAx = 0$ has no other solution than $x = 0$, A^TA is invertible. So, we can safely write $P = A(A^TA)^{-1}A^T$.