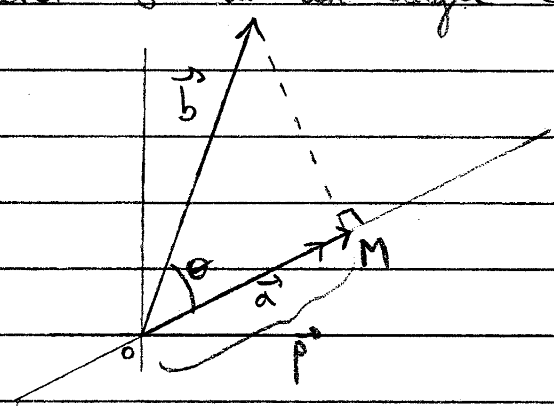
MATH - UA 140 - Linear Algebra

Lecture 15: Projections

I] Projection onto a line

1) Geometric description

Consider a line through the origin with direction vector $\vec{a}$, and a vector $\vec{b}$ at an angle θ compared to this line:



The projection vector $\vec{p}$ of $\vec{b}$ onto the line is the vector $\vec{OM}$ where M is the point on the line which is closest to the tip of $\vec{b}$. Geometrically, M is the intersection of the line with direction vector $\vec{a}$ and the line through the tip of $\vec{b}$ which is orthogonal to that line (dashed line in figure)

2) Vector expression for $\vec{p}$

$\vec{p}$ is aligned with $\vec{a}$. We may write $\vec{p} = \hat{x}\vec{a}$, where $\hat{x}$ is a scalar. You may have seen in previous courses that $\|\vec{p}\| = \|\vec{b}\|\cos\theta$, so $\vec{p} = \dfrac{\|\vec{b}\|\cos\theta}{\|\vec{a}\|}\vec{a}$ so $\hat{x} = \dfrac{\|\vec{b}\|\cos\theta}{\|\vec{a}\|}$

We will now construct a more general, more "Linear Algebra"-like expression for $\vec{p}$ and $\hat{x}$.

The key is to consider the error vector $\vec{e} = \vec{b} - \vec{p}$, represented by the dashed line in the figure.

We have $\vec{a} \cdot \vec{e} = 0 = \vec{a} \cdot (\vec{b} - \vec{p}) = \vec{a} \cdot (\vec{b} - \hat{x}\vec{a})$
$= \vec{a} \cdot \vec{b} - \hat{x}\, \vec{a} \cdot \vec{a}$

$$\Leftrightarrow \hat{x} = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}$$

Or, using more general notation: $\hat{x} = \dfrac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}$

Conclusion: The projection of $\vec{b}$ onto the line with direction vector $\vec{a}$ is the vector $\boxed{\vec{p} = \hat{x}\vec{a} = \dfrac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}\vec{a}}$

Note 1: If $\vec{b} = \vec{a}$, $\hat{x} = 1$, $\vec{p} = \vec{a}$. The projection of $\vec{a}$ onto $\vec{a}$ is $\vec{a}$ itself.

Note 2: If $\vec{b}$ and $\vec{a}$ are orthogonal, $\vec{p} = \vec{0}$

Example: What is the projection $\vec{p}$ of $\vec{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ onto $\vec{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$$\vec{a}^T \vec{b} = 7; \quad \vec{a}^T \vec{a} = 14 \Rightarrow \vec{p} = \frac{7}{14}\vec{a} = \frac{1}{2}\vec{a} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{3}{2} \end{bmatrix}$$

As a check, we can compute $\vec{e} = \vec{b} - \vec{p} = \begin{bmatrix} \frac{3}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}$

We have $\vec{a}^T \vec{e} = \frac{3}{2} - \frac{3}{2} = 0$, as expected.

3) Projection matrix

At the end of this lecture, we will want to not only project vectors onto lines, but also more general subspaces of $\mathbb{R}^n$. The most convenient way to do this will be to represent the projection operation as a projection matrix P acting on the vector $\vec{b}$ to be projected. What would be P in the case of a line

We can rewrite $\vec{p} = \underbrace{\frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}}_{\text{scalar}} \vec{a}$ as $\vec{p} = \vec{a} \frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}$

Hence, in this case, the projection matrix P such that $P\vec{b} = \vec{p}$ is $P = \frac{\vec{a}\vec{a}^T}{\vec{a}^T \vec{a}}$

$\vec{a}^T\vec{a}$ is a scalar. So we see that P is a column ($\vec{a}$) times a row ($\vec{a}^T$), divided by a scalar. It is an $m \times m$ rank 1 matrix. The column space of P is $\vec{a}$

Example: What is the projection matrix onto the line through $\vec{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$?

$$P = \frac{1}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{14} & \frac{2}{14} & \frac{3}{14} \\ \frac{2}{14} & \frac{4}{14} & \frac{6}{14} \\ \frac{3}{14} & \frac{6}{14} & \frac{9}{14} \end{bmatrix}$$

• We can verify that $\begin{bmatrix} \frac{1}{14} & \frac{2}{14} & \frac{3}{14} \\ \frac{2}{14} & \frac{4}{14} & \frac{6}{14} \\ \frac{3}{14} & \frac{6}{14} & \frac{9}{14} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{3}{2} \end{bmatrix} = \vec{p}$

• Note also that $P^2 = \frac{1}{14^2} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$

$$= \frac{1}{14} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} = P!$$

• Projecting a second time does not change anything, as expected.

• $(I - P)\vec{b} = \vec{b} - \vec{p} = \vec{e}$
In other words, $I - P$ is the projection matrix of $\vec{b}$ onto the line through the tip of $\vec{b}$ and M.
When P projects onto one subspace, $I - P$ projects onto the perpendicular subspace (Here it is the plane perpendicular to $\vec{a}$.)

II) Projection onto a subspace

1) Projection and projection matrix

Let us consider the subspace of $\mathbb{R}^m$ spanned by the n linearly independent vectors $\vec{a_1}, \dots, \vec{a_n}$
We take a vector $\vec{b}$ in $\mathbb{R}^m$ and look for the projection

$\vec{p}$ in the subspace spanned by $\vec{a}_1, \dots, \vec{a}_n$, i.e the vector $\vec{p}$ which minimizes $\|\vec{e}\| = \|\vec{b} - \vec{p}\|$

The idea here is given by geometry: $\|\vec{e}\|$ is minimized if $\vec{e} = \vec{b} - \vec{p}$ is orthogonal to the subspace.

Since $\vec{p}$ is in the subspace, we write $\vec{p} = \hat{x}_1 \vec{a}_1 + \dots + \hat{x}_n \vec{a}_n$, with $\hat{x}_1, \dots, \hat{x}_n$ scalars. In other words, $\vec{p} = A\hat{\vec{x}}$, with $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ an $m \times n$ matrix, and $\hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}$

$\vec{e}$ is orthogonal to the subspace, so it is orthogonal to all the vectors $\vec{a}_1, \dots, \vec{a}_n$:

$$\begin{cases} \vec{a}_1^T \vec{e} = 0 \\ \vdots \\ \vec{a}_n^T \vec{e} = 0 \end{cases} \iff \begin{cases} \vec{a}_1^T (\vec{b} - A\hat{\vec{x}}) = 0 \\ \vdots \\ \vec{a}_n^T (\vec{b} - A\hat{\vec{x}}) = 0 \end{cases} \iff \begin{bmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_n^T \end{bmatrix} \begin{bmatrix} \vec{b} - A\hat{\vec{x}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

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

This is the matrix equation which determines $\hat{\vec{x}}$, hence $\vec{p}$. We will soon prove that $\underline{A^T A \text{ is invertible}}$ because the $\vec{a}_i$ are independent. We are now ready to write:

The projection $\vec{p}$ of $\vec{b}$ onto the subspace spanned by $\vec{a}_1, \dots, \vec{a}_n$ is
$$\vec{p} = A(A^T A)^{-1} A^T \vec{b}$$

This means that the $n \times n$ projection matrix that produces $\vec{p} = P\vec{b}$ is
$$\underline{\underline{P = A(A^T A)^{-1} A^T}}$$

• Note 1: The formula agrees with the formula for projection onto a line. In that case, A is just the column vector $\vec{a}$, and $\vec{a}^T\vec{a}$ is just a number, so that $(A^TA)^{-1}$ just becomes $\frac{1}{\vec{a}^T\vec{a}}$

• Note 2: Let us consider what we have just learned in the language of the four subspaces:

A. Our subspace of interest is $C(A)$

B. $\vec{e} = \vec{b} - A\hat{\vec{x}}$ is orthogonal to that space, so it belongs to $N(A^T)$, the left nullspace

C. This means that $A^T(\vec{b} - A\hat{\vec{x}}) = \vec{0}$, which is exactly the equation for $\hat{\vec{x}}$

In other words, the vector $\vec{b}$ is being split into the projection $\vec{p}$ in $C(A)$, and the error $\vec{e}$ in $N(A^T)$, the left nullspace.

QUESTION: Let $A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$

Find $\hat{x}$, $\vec{p}$, and P. Compute P^2.

2) Proof that A^TA is invertible in our case

Observe first that the formula $P = A(A^TA)^{-1}A^T$ can be misleading if it is not read the right way. Indeed, the temptation is high to expand $(A^TA)^{-1}$ and find $P = I$ at the end.

This however cannot be done, because A is rectangular (not square) in general, so A^{-1} does not exist.

However, $A^T A$, a $n \times n$ symmetric matrix, does have an inverse, because of the following property:

$A^T A$ is invertible if and only if A linearly independent columns.

The linear independence of the columns is precisely the assumption we made at the beginning of this section. Now, the proof of the property.

Let us first show that A and $A^T A$ have the same nullspace:

- Let $\vec{x}$ be in $N(A)$. Then $A\vec{x} = \vec{0}$, and multiplying by A^T, $A^T A \vec{x} = \vec{0}$
$\vec{x}$ is also in $N(A^T A)$. So $N(A)$ is included in $N(A^T A)$
- Let $\vec{x}$ be in $N(A^T A)$. Then $A^T A \vec{x} = \vec{0}$
Multiplying by $\vec{x}^T$ on both sides,

$$\vec{x}^T A^T A \vec{x} = 0 \iff (A\vec{x})^T A\vec{x} = 0$$
$$\iff \|A\vec{x}\| = 0$$

$A\vec{x}$ is a vector with zero magnitude, it must be the zero vector.
So $A\vec{x} = \vec{0}$, which means that $\vec{x} \in N(A)$
$N(A^T A)$ is included in $N(A)$

We conclude that $N(A) = N(A^T A)$

Now, when the columns of A are linearly independent, $N(A) = \{\vec{0}\}$
so $N(A^T A) = \{\vec{0}\}$ and $A^T A$ is invertible.
Conversely, when $A^T A$ is invertible, $N(A^T A) = \{\vec{0}\}$, so $N(A) = \{\vec{0}\}$ and the columns of A are linearly independent. This completes our proof.