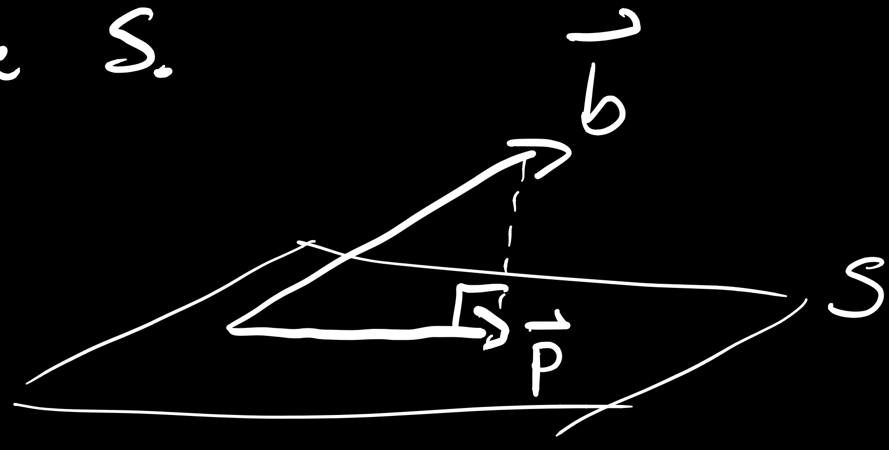
4.2 Projections

We have considered subspaces and the linear transformation viewpoint of matrices. We want to find a projection matrix P that projects some vector $\vec{b}$ onto a subspace S.



__Recall__ In problem set 2, we saw the projection matrix of the form

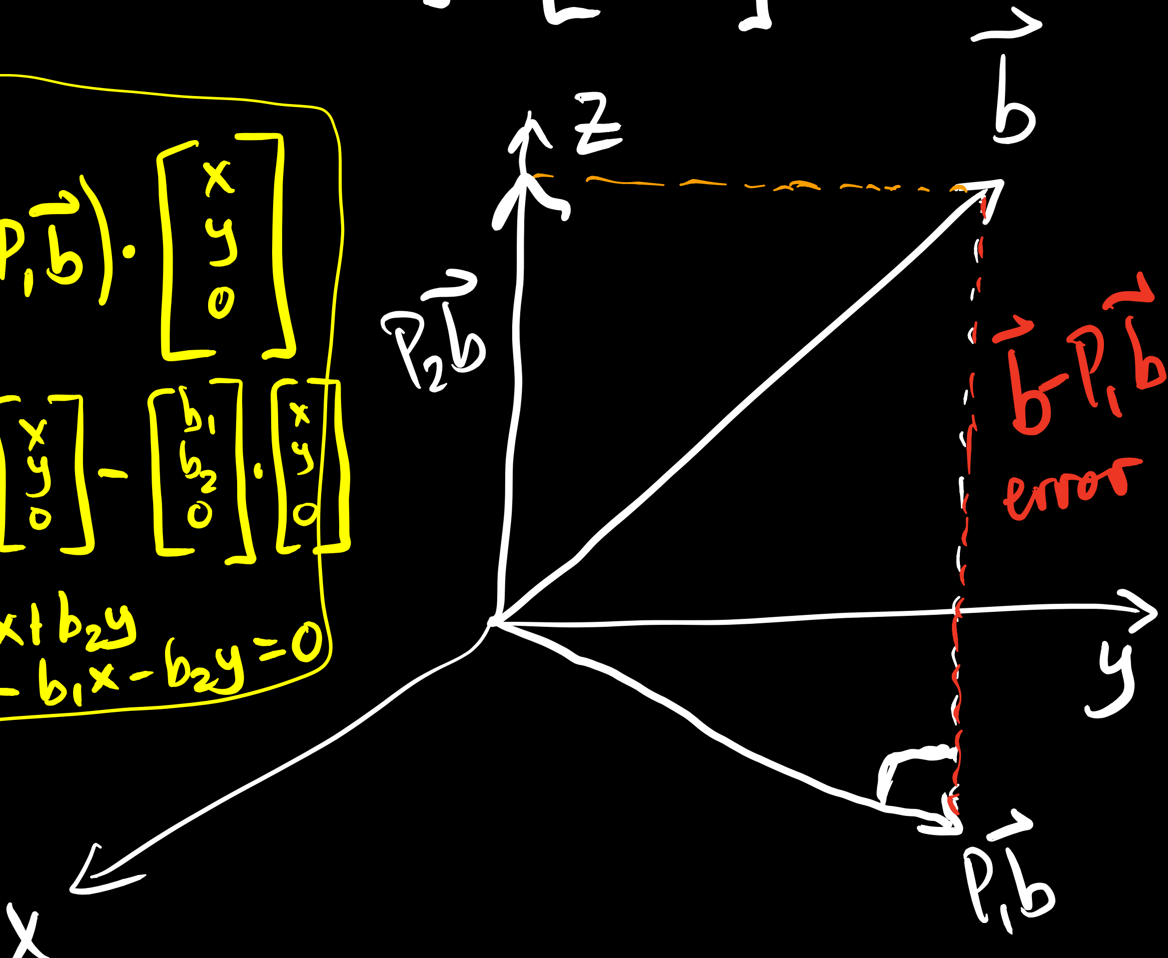
$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P_1 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}$$

$$\left(\vec{b} - P_1\vec{b}\right) \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

$$= \vec{b} \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

$$= b_1x + b_2y - b_1x - b_2y = 0$$



$$P_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_2 \vec{b} = \begin{bmatrix} 0 \\ 0 \\ b_3 \end{bmatrix}$$

We know the x-y plane and z-axis are orthogonal subspaces, and they are infact orthogonal complements.

Notice that

$$P_1 \vec{b} + P_2 \vec{b} = \vec{b}$$

and $$P_1 + P_2 = I$$

Now, let's find the projection matrix for an arbitrary subspace

Projection Onto a Subspace

Let $\vec{a}_1, \dots, \vec{a}_n$ be a basis of a subspace S of $\mathbb{R}^m$.

We want to project $\vec{b} \in \mathbb{R}^m$ onto the subspace spanned by the $\vec{a}_i$'s.

Since the projection $\vec{p} \in S$,

$$\vec{p} = \hat{x}_1 \vec{a}_1 + \dots + \hat{x}_n \vec{a}_n$$

and we choose $\hat{x}_1, \dots, \hat{x}_n$ so $\vec{p}$ is closest to $\vec{b}$. Now, writing

$$A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}$$

we have

$$\vec{p} = A\hat{\vec{x}} \quad , \quad \hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}.$$

The error vector is

$$\vec{b} - \vec{p} = \vec{b} - A\vec{x}$$

The error vector is orthogonal to S, so

$$\vec{a}_1^{\,T}(b - A\hat{\vec{x}}) = 0$$

$$\vdots$$

$$\vec{a}_n^{\,T}(b - A\hat{\vec{x}}) = 0$$

so

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

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

$$A^T A \hat{\vec{x}} = A^T \vec{b} \quad \left(\text{called the normal equation}\right)$$

A is $m \times n$, so $\underset{(n\times m)}{A^T}\underset{(m\times n)}{A}$ is $n \times n$

and we have n equations for n unknowns.

<u>Fact</u> A^TA is invertible if and only if A has linearly independent columns.

Now, inverting A^TA, we find

$$\hat{\vec{x}} = (A^TA)^{-1}A^T\vec{b}$$

so

$$\begin{aligned}\vec{p} &= A\hat{\vec{x}} \\ &= A(A^TA)^{-1}A^T\vec{b}.\end{aligned}$$

Thus, the projection $\vec{p}$ of $\vec{b}$ onto S is

$$\vec{p} = A(A^TA)^{-1}A^T\vec{b}$$

and the projection matrix is

$$P = A(A^TA)^{-1}A^T$$

<u>Note</u>: A is rectangular, so we cannot compute $(A^TA)^{-1}$ as $A^{-1}(A^T)^{-1}$

Ex 1 (Projection onto a line)

Project $\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\vec{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$$\vec{p} = \underbrace{\underbrace{A \underbrace{(\underbrace{A^T A}_{①})^{-1}}_{②} A^T}_{③} \vec{b}}_{④}$$

① $A^T A = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$= 9$

② $(A^T A)^{-1} = \frac{1}{9}$

③ $A(A^T A)^{-1} A^T = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}$

$$= \frac{1}{9}\begin{bmatrix}1\\2\\2\end{bmatrix}\begin{bmatrix}1 & 2 & 2\end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}$$

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

$$\vec{p} = P\vec{b}$$

$$= \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix}5\\10\\10\end{bmatrix}$$

$$\text{proj}_{\vec{a}}\vec{b} = \vec{a}\,\frac{\vec{b}\cdot\vec{a}}{\|\vec{a}\|^2} = \vec{a}\,\frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$$

Ex 2 Find the projection $\vec{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ onto $\text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right)$

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

① $A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$

$$= \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

② $(A^TA)^{-1} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1}$

$$= \frac{1}{15-9} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}$$

$AB \neq BA$

③ $P = A(A^TA)^{-1}A^T$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

④ $\vec{p} = P\vec{b}$

$$= \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 30 \\ 12 \\ -6 \end{bmatrix}$$

$$= \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

4.3 Least Squares Approximations

We have seen that $A\vec{x}=\vec{b}$ may have no solution. Often, we run into this situation when we have more equations than unknowns. We will instead find the "best" solution $\hat{\vec{x}}$.

Q: What do we mean by "best"?

A: By "best", we mean minimal error. We previously mentioned the error vector

$$\vec{e} = \vec{b} - A\vec{x}.$$

When the length of $\vec{e}$ is as

small as possible, $\hat{x}$ is a least squares solution.

We want to minimize $\|\vec{e}\|^2$ and $\vec{x}$ is unknown. Let

$$E(x_1, \ldots, x_n) = \|\vec{e}\|^2.$$

To minimize E, we want to find $x_1, \ldots, x_n$, so that $\frac{\partial E}{\partial x_k} = 0$ for $k = 1, \ldots, n$.

A $m \times n$

$A\vec{x} = \vec{b}$

$$\frac{\partial E}{\partial x_k} = \frac{\partial}{\partial x_k} \|\vec{e}\|^2$$

$$= \frac{\partial}{\partial x_k} \|\vec{b} - A\vec{x}\|^2$$

$$= \frac{\partial}{\partial x_k} \sum_{i=1}^{m} \left(b_i - (A\vec{x})_i \right)^2$$

$$= \frac{\partial}{\partial x_k} \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2$$

$$= \sum_{i=1}^{m} 2 \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) \left(-a_{ik} \right)$$

Since $\frac{\partial E}{\partial x_k} = 0$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j a_{ik} = \sum_{i=1}^{m} a_{ik} b_i$$

$$\sum_{j=1}^{n} \sum_{i=1}^{m} a^T_{ki} a_{ij} x_j = \sum_{i=1}^{m} a^T_{ki} b_i$$

$$\sum_{j=1}^{n} (A^T A)_{kj} x_j = \sum_{i=1}^{m} a^T_{ki} b_i$$

$$(A^T A \vec{x})_k = (A^T \vec{b})_k$$

Then, to minimize E, we

need to solve

$$A^T A \vec{x} = A^T \vec{b} .$$