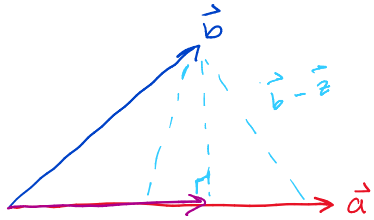
tinyurl.com/miyuki-feedback

Projection onto a vector



$\vec{z} = \mathrm{proj}_{\vec{a}}(\vec{b})$

"projection of $\vec{b}$ onto $\vec{a}$"

$$\underbrace{\mathrm{proj}_{\vec{a}}(\vec{b})}_{\vec{z}} = \underbrace{\frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}}_{\text{scalar}} \vec{a}$$

$\vec{z} = \mathrm{proj}_{\vec{a}}(\vec{b})$ is the closest vector to $\vec{b}$ along $\vec{a}$ $\in \mathrm{span}\{\vec{a}\}$

$\hookrightarrow$ $\|\vec{b}-\vec{z}\|^2$ is the smallest

$\hookrightarrow$ $\|\vec{b}-\vec{z}\|^2$ is smallest when $\perp$ to $\vec{a}$

1. Mechanical Projection

In $\mathbb{R}^n$, the vector valued projection of vector $\vec{b}$ onto vector $\vec{a}$ is defined as:

$$\mathrm{proj}_{\vec{a}}\left(\vec{b}\right) = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}\vec{a}.$$

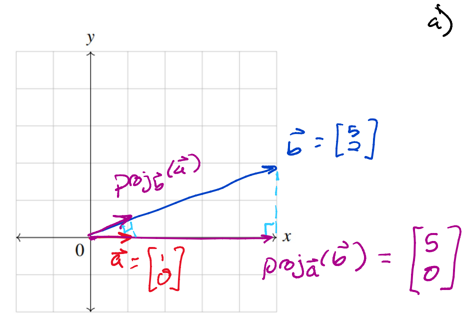
Recall $\|\vec{a}\|^2 = \langle \vec{a}, \vec{a} \rangle$.

(a) Project $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ – that is, onto the x-axis. Graph these two vectors and the projection.

$\vec{b} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ $\quad \vec{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$\mathrm{proj}_{\vec{a}}(\vec{b}) = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}\vec{a} = \frac{(5)(1)+(2)(0)}{(1)(1)+(0)(0)}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \frac{5}{1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$



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

$\vec{a}\hat{x} = \vec{b}$

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}\hat{x} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

$\uparrow$ can't find unique sol

best estimate

$\hat{x} = 5$

$\vec{a}\hat{x} = \vec{b}$

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}\hat{x} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$

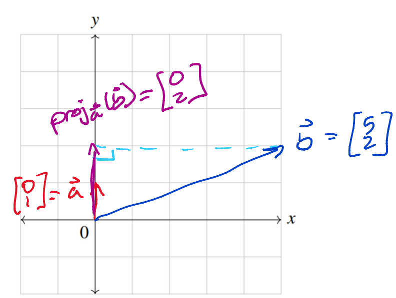
$\uparrow$ unique sol

(b) Project $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ onto $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ – that is, onto the y-axis. Graph these two vectors and the projection.

$\vec{b} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ $\quad \vec{a} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$$\mathrm{proj}_{\vec{a}}(\vec{b}) = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}\vec{a} = \frac{(5)(0)+(2)(1)}{(0)(0)+(1)(1)}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{2}{1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \checkmark$$

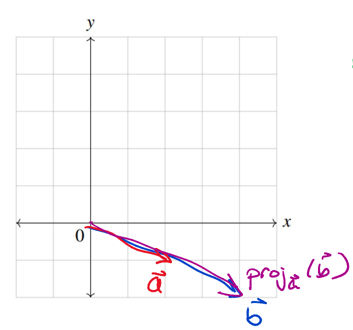


(c) Project $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$ onto $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Graph these two vectors and the projection.

$\vec{b} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$ $\quad \vec{a} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

$$\mathrm{proj}_{\vec{a}}(\vec{b}) = \frac{(4)(2)+(-2)(-1)}{(2)(2)+(-1)(-1)}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{10}{5}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \vec{b} \checkmark$$



projection is "closest" vector to $\vec{b}$ along $\vec{a}$ ($\in \mathrm{span}\{\vec{a}\}$)

if it is already $\in \mathrm{span}\{\vec{a}\}$

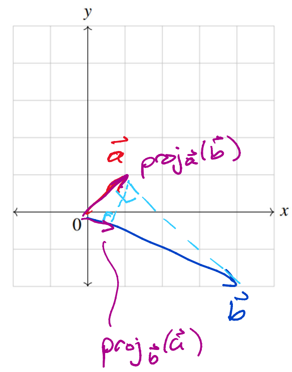
$\mathrm{proj}_{\vec{a}}(\vec{b}) = \vec{b} = \vec{b}$

(d) Project $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Graph these two vectors and the projection.

$\vec{b} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$ $\quad \vec{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\text{proj}_{\vec{a}}(\vec{b}) = \frac{(1)(4) + (1)(-2)}{(1)(1) + (1)(1)} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{2}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \vec{a}$$



(e) (Practice) Project $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto the span of the vectors $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ – that is, onto the x-y plane in $\mathbb{R}^3$.

(Hint: From least squares, the matrix $A(A^\top A)^{-1}A^\top$ projects a vector into $C(A)$.)

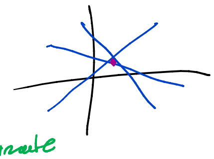
(f) (Practice) What is the geometric/physical interpretation of projection? Justify using the previous parts.

Least Squares

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

sol cases

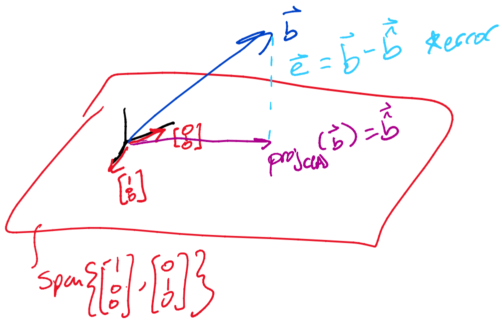
inf, unique, no sol



Ex: $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ $\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

→ when no sol, want to find a $\underbrace{\text{best estimate}}_{\hat{\vec{x}}}$



$A\vec{x} = \vec{b}$ → no sol $\vec{b} \notin C(A)$

$A\hat{\vec{x}} = \hat{\vec{b}}$ → unique sol $\hat{\vec{b}} \in C(A)$

closest vector to $\vec{b}$ in $C(A)$

↳ best estimate $\hat{\vec{x}}$

by minimizing $\|\vec{e}\|^2$ we can solve for $\hat{\vec{x}}$ using

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

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

$* \langle \vec{a}, \vec{b} \rangle = \vec{a}^T\vec{b}$

*if A is only a vector: $\vec{a}(\vec{a}^T\vec{a})^{-1}\vec{a}^T\vec{b}$

$= \vec{a}\dfrac{\langle \vec{a}, \vec{b}\rangle}{\langle \vec{a}, \vec{a}\rangle} = \dfrac{\langle \vec{a}, \vec{b}\rangle}{\|\vec{a}\|^2}\vec{a}$

2. Least Squares with Orthogonal Columns

Suppose we would like to solve the least squares problem for $\mathbf{A} \in \mathbb{R}^{3\times 2}$ and $\vec{b} \in \mathbb{R}^3$; that is, find an optimal vector $\vec{x} \in \mathbb{R}^2$ which gets $\mathbf{A}\vec{x}$ closest to $\vec{b}$ such that the distance $\|\vec{e}\| = \|\vec{b} - \mathbf{A}\vec{x}\|$ is minimized. Call this optimal vector $\hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$. Mathematically, we can express this as:

$$\|\vec{b} - \mathbf{A}\hat{\vec{x}}\|^2 = \min_{\vec{x} \in \mathbb{R}^2} \|\vec{b} - \mathbf{A}\vec{x}\|^2 = \min_{\vec{x} \in \mathbb{R}^2} \left\| \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} - \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2.$$

To identify the solution $\hat{\vec{x}}$, we may recall the least squares formula: $\hat{\vec{x}} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\vec{b}$, which is applicable when $\mathbf{A}$ has linearly independent columns. We would now like to walk through the intuition behind this formula for the case when $\mathbf{A}$ has orthogonal columns: $\langle \vec{a}_1, \vec{a}_2 \rangle = 0$.

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

$\min \|\vec{e}\|^2 = \min \|\vec{b} - A\hat{\vec{x}}\|^2$

→ use least squares to find $\hat{\vec{x}}$ that minimizes this error

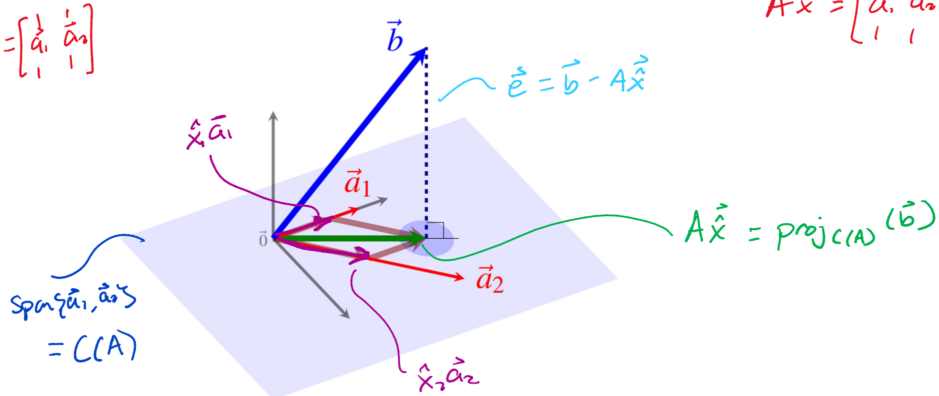
(a) On the diagram below, please label the following elements:

NOTE: For this sub-part only, the matrix $\mathbf{A}$ does not have orthogonal columns.

span $\{\vec{a}_1, \vec{a}_2\}$ $\quad \mathbf{A}\,\hat{\vec{x}}$ $\quad \hat{x}_1\,\vec{a}_1$ $\quad \hat{x}_2\,\vec{a}_2$ $\quad C(\mathbf{A})$ $\quad \vec{e} = \vec{b} - \mathbf{A}\,\hat{\vec{x}}$ $\quad \text{proj}_{C(\mathbf{A})}(\vec{b})$.

$A\hat{\vec{x}} = \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \hat{x}_1\vec{a}_1 + \hat{x}_2\vec{a}_2$

$A = \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix}$



given this special case

(b) Now suppose we assume a special case of the least squares problem where the columns of $\mathbf{A}$ are orthogonal (illustrated in the figure below). Given that $\vec{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\vec{b}$, and $\text{proj}_{C(\mathbf{A})}(\vec{b}) = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\vec{b} = \mathbf{A}\vec{x}$, show the following statement holds.

$\rightarrow \langle \vec{a}_1, \vec{a}_2 \rangle = 0$

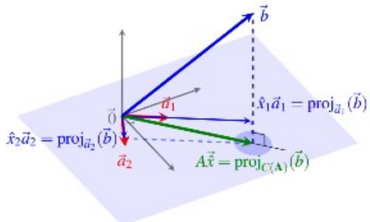
show this

$$\langle \vec{a}_1, \vec{a}_2 \rangle = 0 \quad \longrightarrow \quad \boxed{\vec{x} = \begin{bmatrix} \frac{\langle \vec{a}_1, \vec{b}\rangle}{\|\vec{a}_1\|^2} \\ \frac{\langle \vec{a}_2, \vec{b}\rangle}{\|\vec{a}_2\|^2} \end{bmatrix} \quad \text{and} \quad \text{proj}_{C(\mathbf{A})}(\vec{b}) = \text{proj}_{\vec{a}_1}(\vec{b}) + \text{proj}_{\vec{a}_2}(\vec{b})}$$

In words, the statement says that when the columns of $\mathbf{A}$ are orthogonal, the entries of the least squares solution vector $\vec{x}$ can be computed by using $\vec{b}$ and only the single other vector $\vec{a}_i$, and that the projection of $\vec{b}$ onto $C(\mathbf{A})$ can be computed by summing the projections of $\vec{b}$ onto the $\vec{a}_i$.

RECALL... $\quad \text{proj}_{\vec{a}_1}(\vec{b}) = \frac{\langle \vec{a}_1, \vec{b}\rangle}{\|\vec{a}_1\|^2}\vec{a}_1, \quad \|\vec{a}\|^2 = \langle \vec{a}, \vec{a}\rangle \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix}$

$$\underset{n\times 2}{A}\ \underset{2\times 1}{\vec{x}} = \underset{n\times 1}{\vec{b}}$$



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

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \Bigg(\underset{2\times n}{\begin{bmatrix} - \vec{a}_1^\top - \\ - \vec{a}_2^\top - \end{bmatrix}} \underset{n\times 2}{\begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix}} \Bigg)^{-1} \underset{2\times n}{\begin{bmatrix} - \vec{a}_1^\top - \\ - \vec{a}_2^\top - \end{bmatrix}} \underset{n\times 1}{\vec{b}}$$

$$= \underset{2\times 2}{\begin{bmatrix} \vec{a}_1^\top \vec{a}_1 & \vec{a}_1^\top \vec{a}_2 \\ \vec{a}_2^\top \vec{a}_1 & \vec{a}_2^\top \vec{a}_2 \end{bmatrix}}^{-1} \underset{2\times 1}{\begin{bmatrix} \vec{a}_1^\top \vec{b} \\ \vec{a}_2^\top \vec{b} \end{bmatrix}}$$

$$= \begin{bmatrix} \langle \vec{a}_1, \vec{a}_1 \rangle & \cancel{\langle \vec{a}_1, \vec{a}_2 \rangle}^{0} \\ \cancel{\langle \vec{a}_2, \vec{a}_1 \rangle}^{0} & \langle \vec{a}_2, \vec{a}_2 \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle \vec{a}_1, \vec{b} \rangle \\ \langle \vec{a}_2, \vec{b} \rangle \end{bmatrix} = \begin{bmatrix} \|\vec{a}_1\|^2 & 0 \\ 0 & \|\vec{a}_2\|^2 \end{bmatrix}^{-1} \begin{bmatrix} \langle \vec{a}_1, \vec{b} \rangle \\ \langle \vec{a}_2, \vec{b} \rangle \end{bmatrix}$$

* because orthogonal

$$= \begin{bmatrix} \frac{1}{\|\vec{a}_1\|^2} & 0 \\ 0 & \frac{1}{\|\vec{a}_2\|^2} \end{bmatrix} \begin{bmatrix} \langle \vec{a}_1, \vec{b} \rangle \\ \langle \vec{a}_2, \vec{b} \rangle \end{bmatrix} = \begin{bmatrix} \frac{\langle \vec{a}_1, \vec{b} \rangle}{\|\vec{a}_1\|^2} = \hat{x}_1 \\ \frac{\langle \vec{a}_2, \vec{b} \rangle}{\|\vec{a}_2\|^2} = \hat{x}_2 \end{bmatrix} = \hat{\vec{x}} \quad ✓$$

* for the special case where cols of A are orthogonal

$$\text{proj}_{C(A)}(\vec{b}) = A\hat{\vec{x}} = \underset{\text{scalar}}{\hat{x}_1}\vec{a}_1 + \hat{x}_2\vec{a}_2 = \frac{\langle \vec{a}_1, \vec{b} \rangle}{\|\vec{a}_1\|^2}\vec{a}_1 + \frac{\langle \vec{a}_2, \vec{b} \rangle}{\|\vec{a}_2\|^2}\vec{a}_2$$

$$= \text{proj}_{\vec{a}_1}(\vec{b}) + \text{proj}_{\vec{a}_2}(\vec{b}) \quad ✓$$

(c) Compute the least squares solution $\vec{x} \in \mathbb{R}^2$ to the following system:

$\hat{\vec{x}}$

$$\min_{\vec{x}\in\mathbb{R}^2} \left\| \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2.$$

* notice cols of A are orthogonal!

HINT: Notice that the columns of **A** are orthogonal!!

$\min \| \vec{b} - A\vec{x} \|^2$

$\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$

$\vec{a}_1 \quad \vec{a}_2$

$\langle \vec{a}_1, \vec{a}_2 \rangle = 0$

$(1)(0) + (0)(1) + (0)(1) = 0$

$$\hat{\vec{x}} = \begin{bmatrix} \dfrac{\langle \vec{a}_1, \vec{b} \rangle}{\|\vec{a}_1\|^2} \\[2ex] \dfrac{\langle \vec{a}_2, \vec{b} \rangle}{\|\vec{a}_2\|^2} \end{bmatrix} = \begin{bmatrix} \dfrac{(1)(1) + (0)(2) + (0)(3)}{(1)(1) + (0)(0) + (0)(0)} \\[2ex] \dfrac{(0)(1) + (1)(2) + (1)(3)}{(0)(0) + (1)(1) + (1)(1)} \end{bmatrix} = \begin{bmatrix} 1 \\[1ex] \dfrac{5}{2} \end{bmatrix}$$

could've used least squares formula

$\hat{\vec{x}} = (A^TA)^{-1}A^T\vec{b}$ but it would've taken longer!