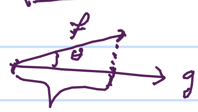
Goal:

- Continue the discussion of projection
- Present the projection theorem
- Introduce the Gram matrix
- Do a real-world example

Recall: last time we discussed the idea of a generalized angle between two vectors f, g in an inner product space:

$$\cos(\text{"}\theta\text{"}) = \frac{(f,g)}{\|f\|\|g\|} = \frac{(f,g)}{(f,f)^{1/2}(g,g)^{1/2}}$$ and used this to

define a generalized projection operator where we can project f onto g by:



$$\text{Proj}_g(f) = \left(\cos(\text{"}\theta\text{"})\,\|f\|\right)\left(\frac{g}{\|g\|}\right)$$ ← unit vector in the direction of g

"projection of ___ onto ___"

$$= \left(\frac{(f,g)}{(g,g)^{1/2}}\right)\frac{g}{(g,g)^{1/2}} = \frac{(f,g)}{(g,g)}\,g$$

Then we saw that we could define $\tilde{f}$ to be the component of f orthogonal to g as: $\tilde{f} = f - \text{proj}_g(f)$ and that $(\tilde{f}, g) = 0$.

The Projection theorem (3.4.2)

Now we want to discuss how to use the idea of projections to build "the best" approximations possible in a vector space V.

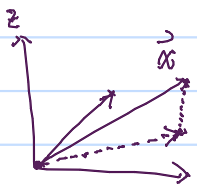
Let V be a vector space with inner product $(\cdot,\cdot)_V$. Let $W \subseteq V$ be a finite dimensional subspace of V with $\dim(W) \leq \dim(V)$. We want to find a vector $w \in W$ such that "w is the closest vector in the subspace W to the vector v".

Recall: by "close" we are implicitly talking about distance as measured by the inner product norm: $\|x\|_V = (x,x)_V^{1/2}$

So in mathematical terms we want to find $w \in W$ that <u>minimizes</u> $\|v - y\|$ where y is any vector in W. That is we want to find $w \in W$ such that

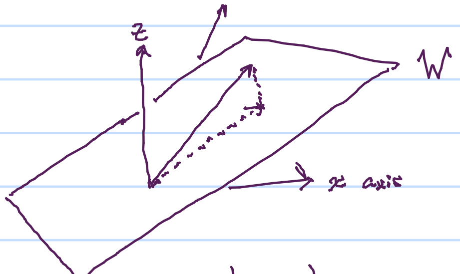
$$\|v - w\| \leq \|v - y\| \text{ for any } y \in W.$$

How can we understand this intuitively? Lets go to our good friend and constant companion $V = \mathbb{R}^3$. The dot product $\langle x, y \rangle = x \cdot y$ gives rise to the notion of geometry to which we are physically accustomed. Let $W = \mathbb{R}^2$ be our subspace: we will use the x-y plane but any 2d surface in $\mathbb{R}^2$ would work.



Suppose we have $x = (1, 2, 7)$ and we want to find the best approximation $w \in \mathbb{R}^2$ to x how would you find the closest vector on the x-y axis?

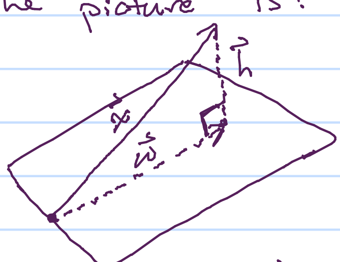
Thats right! you would "project" x onto the x-y axis. What if we took W to be a plane oriented diffeently



The approach, intuitively, would be the same. We would want to drop a horizontal line from the tip of $\vec{x}$ down to W and take our approximation w to be the vector whose tail was at the origin and whose tip was at the point where this "projection" of $\vec{x}$ touched W.

<u>Key idea</u>: we can find the vector $w \in W$ that best approximates $\vec{x}$ if we can find the <u>horizontal</u> (perpendicular) vector $\vec{h}$ that forms a $90°$ angle between W and $\vec{x}$.

The picture is:



- Notice that if we find h we can write: $\vec{x} = \vec{w} + h \Rightarrow \vec{x} - \vec{w} = \vec{h}$
- Notice that such an $\vec{h}$ is <u>perpendicular</u> to <u>all</u> of the subspace W
- Combining these observations means that

The projection of x onto W is orthogonal to $\vec{w}$

eg: $\vec{h} \cdot \vec{y} = (\vec{x} - \vec{w}) \cdot \vec{y} = 0$ for <u>all</u> $y \in W$

Another way to write this is: $(x-w, y) = 0$ for every $y \in W$.

Here $\vec{x} - \vec{h} = \vec{w} \in W$ is called the <u>projection of $\vec{x}$ onto the space W</u> and is written as $w = x - h = \text{proj}_W(x)$.

▷ So when $V = \mathbb{R}^3$ and $W = \mathbb{R}^2$ we can make nice intuitive sense of the problem of finding the closest vector $w \in W$ to x.

Q: Can we recast this idea into a result for general inner product spaces?

A: Theorem: (the projection theorem): Let V be a vector space with inner product $(\cdot,\cdot)_V$ and let W be a finite dimensional subspace. Let $v \in V$ be fixed.

1) There is a <u>unique</u> $w \in W$ satisfying

$$\|v - w\| = \underset{y \in W}{\text{minimum}} \|v - y\|$$

2) A vector $w \in W$ is the best approximation in W to $v \in W$ if and only if: $(v-w, y)_V = 0$ for every $y \in W$.

3) Let $\{w_1, w_2, \dots, w_j\}$ be a basis for W. Then the projection of v onto W (e.g. the best approximation to v in W) is given by the formula:

$$w = \text{Proj}_W(v) = \sum_{i=1}^{j} x_i \vec{w}_i$$

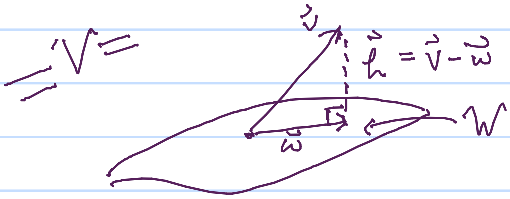
where the x_i solve the matrix equation $G\vec{x} = \vec{b}$. The matrix G is $G_{ij} = (w_i, w_j)$ and the vector b is $b_i = (v, w_i)$

The matrix

$$G = \begin{bmatrix} (w_1,w_1) & (w_1,w_2) & \cdots & (w_1,w_j) \\ \vdots & (w_2,w_2) & & \\ \vdots & & \ddots & \\ (w_j,w_1) & (w_j,w_2) & \cdots & (w_j,w_j) \end{bmatrix}$$

is called the <u>Grahm matrix</u>.

Where does the Grahm matrix come from? It is a direct result of part (2) of the theorem. We know we should have that $\vec{v} - \vec{w}$ is orthogonal to every vector in W



So a consequence is that $\vec{v} - \vec{w} = \vec{h}$ must be orthogonal to each of the basis elements $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_j\}$ of W.

- We also know that the vector $\vec{w}$ can be written as a linear combination of the basis vectors $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_j\}$

write $\vec{w} = x_1 \vec{w}_1 + x_2 \vec{w}_2 + \cdots + x_n \vec{w}_n$

- We get:

$$(v - (x_1 \vec{w}_1 + \cdots + x_j \vec{w}_j), w_1) = 0$$
$$(v - (x_1 \vec{w}_1 + \cdots + x_j \vec{w}_j), w_2) = 0$$
$$\vdots$$
$$(v - (x_1 \vec{w}_1 + \cdots + x_j \vec{w}_j), w_j) = 0$$

Rearranging these equations gives

$$\left.\begin{aligned}
&\overbrace{x_1(w_1,w_1) + x_2(w_2,w_2) + \cdots + x_j(w_1,w_j)}^{j \text{ unknowns}} = (v,w_1)\\
&x_1(w_2,w_1) + x_2(w_2,w_2) + \cdots + x_j(w_2,w_j) = (v,w_2)\\
&\vdots\\
&x_1(w_j,w_1) + x_2(w_j,w_2) + \cdots + x_j(w_j,w_j) = (v,w_j)
\end{aligned}\right\} j \text{ equations}$$

Written in matrix form this is

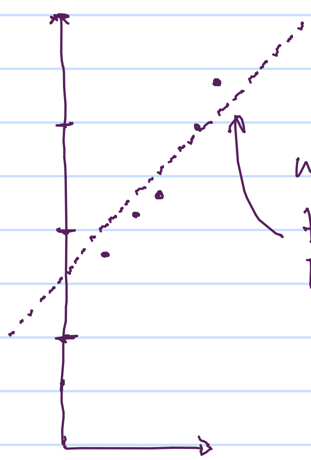
$$\begin{bmatrix} (w_1,w_1) & (w_1,w_2) & \cdots & (w_1,w_j) \\ \vdots & (w_2,w_2) & & \\ & & \ddots & \\ (w_j,w_1) & (w_j,w_2) & \cdots & (w_j,w_j) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \end{bmatrix} = \begin{bmatrix} (v,w_1) \\ (v,w_2) \\ \vdots \\ (v,w_j) \end{bmatrix}$$

Which is exactly the equation $G\vec{x} = \vec{b}$ where G is the Grahm matrix.

Example: (pg 65 of text) Linear regression

Suppose 5 datapoints are given:

$(x_1,y_1) = (0.1, 1.7805)$
$(x_2,y_2) = (0.3, 2.2285)$
$(x_3,y_3) = (0.4, 2.3941)$
$(x_4,y_4) = (0.75, 3.2226)$
$(x_5,y_5) = (0.9, 3.5697)$

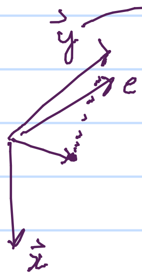


Want to find the best fit Line

Suppose we hypothesize that the relationship between x and y should be approximately linear. Eg we want to find a, b such that $y_i = ax_i + b_i$

The dataset and our hypothesis can be written together as $\vec{y} = a\vec{x} + b\vec{e}$ where $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$ $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$ $\vec{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

The idea is to pick a, b minimizing the distance between the vector $\vec{y} \in \mathbb{R}^5$ and the two-dimensional subspace spanned by the vectors $\{\vec{x}, \vec{e}\}$.



$\rightarrow$ our vector $\vec{y}$ is some vector floating in $\mathbb{R}^5$

- x and e are both vectors in $\mathbb{R}^5$ as well but together they span a two dimensional subspace

- The vector $\vec{w}$ minimizing the distance between $\vec{y}$ and $W = \text{Span}\{\vec{x}, \vec{e}\}$ will be of the form $w = a\vec{x} + b\vec{e}$ for some coefficients a, b.

- The line $f(x) = ax+b$ will then be the best fit line to the datapoints since the points $f(x_i)$ will be the entries to the vector $\vec{w}$. i.e. $f(x_i) = w_i$

$\longrightarrow$ So the abstract setup for our problem is: Let $V = \mathbb{R}^5$ and $(\cdot, \cdot)_V$ be the usual dot product. Find $w = \text{proj}_W(v)$ where $W = \text{span}\{\vec{x}, \vec{e}\}$

Solution: We know $G\vec{w} = \vec{r}$ where $G = \begin{bmatrix} (\vec{x}, \vec{x}) & (\vec{x}, \vec{e}) \\ (\vec{e}, \vec{x}) & (\vec{e}, \vec{e}) \end{bmatrix} = \begin{bmatrix} \vec{x}\cdot\vec{x} & \vec{x}\cdot\vec{e} \\ \vec{e}\cdot\vec{x} & \vec{e}\cdot\vec{e} \end{bmatrix}$

and $\vec{r} = \begin{bmatrix} \vec{y}\cdot\vec{x} \\ \vec{y}\cdot\vec{e} \end{bmatrix}$. We have $\begin{bmatrix} 1.6325 & 2.45 \\ 2.45 & 5 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 7.4370 \\ 13.196 \end{bmatrix}$

The solution is: $w = \begin{bmatrix} 2.2411 \\ 1.5409 \end{bmatrix}$. Hence by taking $a = w_1$ and $b = w_2$ we get that the best fit line is $f(x) = ax+b = 2.2411x + 1.5409$