$(V, \langle\,,\rangle)$ an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$.

$W \subseteq V$ fin-dim'l subspace $\Rightarrow$ $V = W \oplus W^\perp$, so every $y \in V$ has a unique expression of the form $y = u + z$, $u \in W$, $z \in W^\perp$. If $\{u_1, \dots, u_k\}$ is an orthonormal basis for W, then

$$u = \sum_{i=1}^{k} \langle y, u_i \rangle u_i$$

and u is the point in W closest to y.

e.g. In $\mathbb{R}^2$, find the line closest to the three points $(0,6)$, $(1,0)$, $(2,0)$. For $y = ax + b$ to pass through the pts, we would need

$$\underset{A}{\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}} \underset{x}{\begin{pmatrix} a \\ b \end{pmatrix}} = \underset{y}{\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}}$$

There is no x satisfying this eq'n, so instead we look for $x = (a, b)$ minimizing the <u>error</u> $e := \|y - Ax\|$.

Define $W := \operatorname{im}(A)$. Then to minimize e, we need to compute the projection of $y = (6,0,0)$ onto W. For this, we need an orthonormal basis of W. Begin w/ columns of A & apply Gram-Schmidt: $v_1 = (0,1,2)$

$$v_2 = (1,1,1) - \frac{(1,1,1)\cdot(0,1,2)}{(0,1,2)\cdot(0,1,2)}(0,1,2)$$

$$= (1,1,1) - \tfrac{3}{5}(0,1,2)$$

$$= \left(1, \tfrac{2}{5}, -\tfrac{1}{5}\right).$$

Then $u_1 = \dfrac{v_1}{\|v_1\|} = \dfrac{1}{\sqrt{5}}(0,1,2)$

$u_2 = \dfrac{v_2}{\|v_2\|} = \sqrt{\dfrac{5}{6}}\left(1, \tfrac{2}{5}, -\tfrac{1}{5}\right)$ form an orthonormal basis of W.

The projection of y onto W is

$$u = \langle y, u_1 \rangle u_1 + \langle y, u_2 \rangle u_2$$
$$= (6,0,0)\cdot \tfrac{1}{\sqrt5}(0,1,2)\, u_1 + (6,0,0)\cdot \sqrt{\tfrac{5}{6}}\left(1,\tfrac{2}{5},-\tfrac{1}{5}\right) u_2$$
$$= 6\sqrt{\tfrac{5}{6}}\, u_2$$
$$= 6\sqrt{\tfrac{5}{6}}\sqrt{\tfrac{5}{6}}\left(1,\tfrac{2}{5},-\tfrac{1}{5}\right)$$
$$= (5,2,-1).$$

Since $(5,2,-1) \in W = \operatorname{im} A$, we can solve

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$ and get $a = -3$, $b = 5$.

So the line of best fit is $\boxed{y = -3x + 5.}$

<u>Adjoints</u> (w/o proof)

If $f \in \mathcal{L}(V)$, $\exists!\, f^\dagger \in \mathcal{L}(V)$, the <u>adjoint</u> of f satisfying

$$\langle f(x), y \rangle = \langle x, f^\dagger(y) \rangle$$

$\forall x, y \in V$. If f is ~~rep~~ represented by a matrix A wrt some ordered basis, then $f^\dagger$ is rep'd by $A^\dagger$ (aka A^*), the conjugate transpose of A: $\langle Ax, y \rangle = \langle x, A^\dagger y \rangle$.

<u>Utility</u>: given $A \in M_{m\times n}(F)$ and $y \in F^m$, want to compute $x \in F^n$ minimizing $\|y - Ax\|$.

<u>Lemma</u> $\operatorname{rank}(A^\dagger A) = \operatorname{rank}(A)$.

<u>Pf</u> Note $A^\dagger A \in M_{n\times n}(F)$. By rank-nullity,

$$\operatorname{rank}(A) = n - \dim(\ker A)$$
$$\operatorname{rank}(A^\dagger A) = n - \dim(\ker A^\dagger A)$$

so it suffices to show $\dim \ker A = \dim \ker A^\dagger A$.

If $x \in \ker A$, then $Ax = 0 \Rightarrow A^{+}Ax = A^{+}0 = 0$ so $x \in \ker A^{+}A$.
Thus $\ker A \subseteq \ker A^{+}A$.
If $x \in \ker A^{+}A$, then $A^{+}Ax = 0 \Rightarrow 0 = \langle x, 0\rangle = \langle x, A^{+}Ax\rangle = \langle Ax, Ax\rangle$.
By pos def, $Ax = 0$, so $x \in \ker A$, proving the opposite inclusion. $\square$

<u>Cor</u> If $A \in M_{m\times n}(F)$ has rank n, then $A^{+}A$ is invertible. $\square$

<u>Prop</u> Given $A \in M_{m\times n}(F)$ and $y \in F^m$, there exists $x_0 \in F^n$ such that $\|y - Ax_0\| \le \|y - Ax\|$ $\forall x \in F^n$. For this x_0, we have $A^{+}A\,x_0 = A^{+}y$. If $\operatorname{rank}(A) = n$, then $x_0 = (A^{+}A)^{-1}A^{+}y$.

<u>Pf</u> Want Ax_0 closest to y so looking for the proj'n of y onto $\operatorname{im}(A) \subseteq F^m$. This proves existence. Now want to find $x_0 \in F^n$ s.t. $y = Ax_0 + z$ with $z = y - Ax_0 \in (\operatorname{im} A)^{\perp}$.

$$
\begin{aligned}
y - Ax_0 \in (\operatorname{im} A)^{\perp} &\iff \langle Ax, y - Ax_0\rangle = 0 \quad \forall x \in F^n \\
&\iff \langle x, A^{+}(y - Ax_0)\rangle = 0 \quad \forall x \in F^n \\
&\iff A^{+}(y - Ax_0) = 0 \\
&\iff A^{+}Ax_0 = A^{+}y.
\end{aligned}
$$
$\square$

<u>e.g.</u> $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$, $\operatorname{rank} A = 2$.

$$A^{+}A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}, \quad (A^{+}A)^{-1} = \frac{1}{6}\begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}$$

$$\Rightarrow x_0 = (A^{+}A)^{-1}A^{+}y = \frac{1}{6}\begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}. \checkmark$$

<u>Rmk</u> This method avoids computing an orthonormal basis (as before).

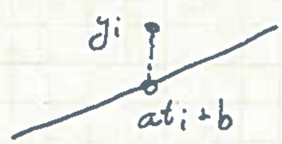
<u>Least Squares</u> Minimizing $\|y - Ax\|$ is called the method of least squares. Imagine that at time t_i we are measuring a quantity $y_i \in F$, $i = 1, \ldots, n$. Want the "best" line $y = ax+b$. Then we want $a, b \in F$ s.t. $y_i = at_i + b$ for each i, i.e.

$$\underbrace{\begin{pmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{y}$$

Seek to minimize error $\|y - Ax\|$, or equivalently $\|y - Ax\|^2$. But

$$\|y - Ax\|^2 = \sum_{i=1}^{n} (y_i - (at_i + b))^2 .$$

The terms $y_i - (at_i + b)$ are vertical distances



and we are minimizing their squares.