11/12

5.1 Inner Product Spaces

- Recall that for $z = x+iy \in \mathbb{C}$, $\bar{z} = x - iy$ and $|z|^2 = \bar{z}z = x^2+y^2$.
- For $\vec{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$, $\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{C}^n$ we define the inner product of $\vec{z}$ and $\vec{w}$ by

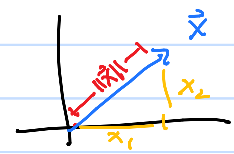
$$\langle \vec{z}, \vec{w} \rangle := \sum_{i=1}^{n} \bar{w}_i z_i$$

- For $\vec{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^n$ we define the norm of $\vec{z}$ by

$$\|\vec{z}\| = \langle \vec{z}, \vec{z} \rangle^{1/2} = \left(\sum_{i=1}^{n} \bar{z}_i z_i \right)^{1/2} = \left(\sum_{i=1}^{n} |z_i|^2 \right)^{1/2}$$

Ex Let $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$. Then

$$\|\vec{x}\| = \sqrt{|x_1|^2 + |x_2|^2} = \underbrace{\sqrt{x_1^2 + x_2^2}}_{\text{length of vector}}$$



Def For $A \in M_{m\times n}(\mathbb{C})$, the adjoint of A is the matrix $A^* \in M_{n\times m}(\mathbb{C})$ with entries $(A^*)_{i,j} = \overline{(A)_{j,i}}$.

- For $\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{C}^n$, if we think of $\vec{w} \in M_{n\times 1}(\mathbb{C})$, then

$$\vec{w}^* = (\bar{w}_1, \dots, \bar{w}_n) \in M_{1\times n}(\mathbb{C})$$

and for $\vec{z} \in \mathbb{C}^n$ ← matrix multiplication.

$$\langle \vec{z}, \vec{w} \rangle = \bar{w}_1 z_1 + \cdots + \bar{w}_n z_n = \vec{w}^* \vec{z}$$

Rem Note that the inner product satisfies the following properties

1. $\langle \vec{w}, \vec{z} \rangle = \overline{\langle \vec{z}, \vec{w} \rangle}$
2. $\langle \alpha \vec{z}_1 + \beta \vec{z}_2, \vec{w} \rangle = \alpha \langle \vec{z}_1, \vec{w} \rangle + \beta \langle \vec{z}_2, \vec{w} \rangle$ for all $\alpha, \beta \in \mathbb{C}$
3. $\langle \vec{z}, \vec{z} \rangle = \|\vec{z}\|^2 \geq 0$
4. $\langle \vec{z}, \vec{z} \rangle = 0 \Rightarrow \vec{z} = \vec{0}$.

Def Let V be vector space. A inner product on V is a map that assigns to each pair of vectors $\vec{v}, \vec{w} \in V$ a scalar denoted $\langle \vec{v}, \vec{w} \rangle$ with the following properties:

1. Conjugate symmetry: $\langle \vec{w}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{w} \rangle}$ for all $\vec{v}, \vec{w} \in V$.
   (If $\mathbb{F} = \mathbb{R}$, then this replaced with symmetry: $\langle \vec{w}, \vec{v} \rangle = \langle \vec{v}, \vec{w} \rangle$ for all $\vec{v}, \vec{w} \in V$.)
2. Linearity: $\langle \alpha \vec{u} + \beta \vec{v}, \vec{w} \rangle = \alpha \langle \vec{u}, \vec{w} \rangle + \beta \langle \vec{v}, \vec{w} \rangle$ for scalars α, β and all $\vec{u}, \vec{v}, \vec{w} \in V$.
3. Non-negativity: $\langle \vec{v}, \vec{v} \rangle \geq 0$ for all $\vec{v} \in V$.
4. Non-degeneracy: $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$.

We call V together with the map $\langle \cdot, \cdot \rangle$ an inner product space. For each $\vec{v} \in V$, the norm of $\vec{v}$ is the quantity:

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}.$$

Ex 1 $\mathbb{C}^n$ (and $\mathbb{R}^n$) is an inner product space with inner product
$$\langle \vec{z}, \vec{w} \rangle = \sum_{i=1}^{n} \overline{w_i}\, z_i .$$

2 Define
$$\ell^2(\mathbb{N}) = \{ (a_n)_{n\in\mathbb{N}} : a_n \in \mathbb{C}, \sum_{n=1}^{\infty} |a_n|^2 < \infty \}$$
Then this is a vector space with operations
$$(a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} = (a_n + b_n)_{n\in\mathbb{N}}$$
$$\alpha (a_n)_{n\in\mathbb{N}} = (\alpha a_n)_{n\in\mathbb{N}}$$
Moreover, it has an inner product:
$$\langle (a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \rangle_2 := \sum_{n=1}^{\infty} \overline{b_n}\, a_n,$$
but one has to justify why the series on the right converges. Note that
$$\| (a_n)_{n\in\mathbb{N}} \|_2 := \left(\sum_{n=1}^{\infty} |a_n|^2 \right)^{1/2}.$$

3 Let $C[0,1]$ be the vector space of cts functions $f: [0,1] \to \mathbb{F}$ with operations
$$(f+g)(t) = f(t) + g(t)$$
$$(\alpha f)(t) = \alpha f(t)$$
This is an inner product space with inner product:
$$\langle f, g \rangle_2 := \int_0^1 f(t)\, \overline{g(t)}\, dt .$$
Note that for each $n \in \mathbb{N}$, $\mathbb{P}_n \subseteq C[0,1]$.

4 $M_{m\times n}$ can be made into an inner product space:
$$\langle A, B \rangle_2 := \mathrm{Tr}(B^* A) = \sum_{i=1}^{m} \sum_{j=1}^{n} \overline{B_{ij}}\, A_{ij}$$
Note that
$$\|A\|_2 = \left(\sum_{i=1} \sum_{j=1} |A_{ij}|^2 \right)^{1/2}. \qquad \square$$

Properties of Inner Product Spaces

Fix an inner product space $(V, \langle \cdot,\cdot \rangle)$. Note that 1 and 2 imply:
$$\langle \vec{u}, \alpha\vec{v} + \beta\vec{w} \rangle = \overline{\langle \alpha\vec{v} + \beta\vec{w}, \vec{u} \rangle} = \overline{\alpha \langle \vec{v}, \vec{u} \rangle + \beta \langle \vec{w}, \vec{u} \rangle}$$
$$= \overline{\alpha}\, \overline{\langle \vec{v}, \vec{u} \rangle} + \overline{\beta}\, \overline{\langle \vec{w}, \vec{u} \rangle} = \overline{\alpha} \langle \vec{u}, \vec{v} \rangle + \overline{\beta} \langle \vec{u}, \vec{w} \rangle.$$
Thus the inner product is conjugate linear in the second coordinate.

2' $\quad \langle \vec{u}, \alpha\vec{v} + \beta\vec{w} \rangle = \overline{\alpha} \langle \vec{u}, \vec{v} \rangle + \overline{\beta} \langle \vec{u}, \vec{w} \rangle$

2 also implies (Exercise)

$$* \qquad \langle \vec{v}, \vec{0} \rangle = \langle \vec{0}, \vec{v} \rangle = 0 \qquad \forall \vec{v} \in V.$$

Lemma For $\vec{v} \in V$, $\vec{v} = \vec{0}$ if and only if
$$\langle \vec{v}, \vec{w} \rangle = 0 \qquad \forall \vec{w} \in V.$$

Proof ($\Rightarrow$) This follows from (*).
($\Leftarrow$) Note $\|\vec{v}\| = (\langle \vec{v}, \vec{v} \rangle)^{1/2} = 0$. So $\vec{v} = \vec{0}$ by 4. $\square$

Cor For $\vec{x},\vec{y}\in V$, $\vec{x}=\vec{y}$ if and only if

$$\langle \vec{x},\vec{w}\rangle = \langle \vec{y},\vec{w}\rangle \qquad \forall \vec{w}\in V.$$

Moreover, if $A,B: U\to V$ are linear transformations, then $A=B$ if and only if

$$\langle A\vec{u},\vec{w}\rangle = \langle B\vec{u},\vec{w}\rangle \qquad \forall \vec{u}\in U,\ \vec{w}\in W.$$

Proof For the first statement, apply the previous Lemma to $\vec{v}:=\vec{x}-\vec{y}$ and use linearity. For the second statement, the only if direction ($\Rightarrow$) is obvious. For the other direction, fix $\vec{u}\in U$. Then the first statement implies $A\vec{u}=B\vec{u}$. Since $\vec{u}\in U$ was arbitrary, we have $A=B$. $\square$

Thm (Cauchy-Schwarz Inequality)
Let V be an inner product space. Then for any $\vec{x},\vec{y}\in V$

$$|\langle \vec{x},\vec{y}\rangle| \le \|\vec{x}\|\,\|\vec{y}\|.$$

Proof Let $t\in\mathbb{C}$ and consider

$$\begin{aligned}
0\le \|\vec{x}-t\vec{y}\|^2 &= \langle \vec{x}-t\vec{y},\vec{x}-t\vec{y}\rangle\\
&= \langle \vec{x},\vec{x}-t\vec{y}\rangle - t\langle \vec{y},\vec{x}-t\vec{y}\rangle\\
&= \langle \vec{x},\vec{x}\rangle - \bar{t}\langle \vec{x},\vec{y}\rangle - t\langle \vec{y},\vec{x}\rangle + |t|^2\langle \vec{y},\vec{y}\rangle\\
&= \|\vec{x}\|^2 - \bar{t}\langle \vec{x},\vec{y}\rangle - t\overline{\langle \vec{x},\vec{y}\rangle} + |t|^2\|\vec{y}\|^2.
\end{aligned}$$

Set $t:=\frac{\langle \vec{x},\vec{y}\rangle}{\|\vec{y}\|^2}$. Then the above becomes

$$\begin{aligned}
0\le \cdots &= \|\vec{x}\|^2 - \frac{|\langle \vec{x},\vec{y}\rangle|^2}{\|\vec{y}\|^2} - \frac{|\langle \vec{x},\vec{y}\rangle|^2}{\|\vec{y}\|^2} + \frac{|\langle \vec{x},\vec{y}\rangle|^2}{\|\vec{y}\|^4}\cdot\|\vec{y}\|^2\\
&= \|\vec{x}\|^2 - \frac{|\langle \vec{x},\vec{y}\rangle|^2}{\|\vec{y}\|^2}
\end{aligned}$$

Thus

$$\frac{|\langle \vec{x},\vec{y}\rangle|^2}{\|\vec{y}\|^2} \le \|\vec{x}\|^2 \implies |\langle \vec{x},\vec{y}\rangle|^2 \le \|\vec{x}\|^2\cdot\|\vec{y}\|^2.$$ $\square$

EX For $V=\mathbb{R}^2$, you might recall

$$\langle \vec{v},\vec{w}\rangle = \cos\theta\,\|\vec{v}\|\cdot\|\vec{w}\|$$

where θ is the angle between $\vec{v}$ and $\vec{w}$. Since $-1\le\cos\theta\le 1$, we have $|\cos\theta|\le 1$, so the above formula implies the Cauchy-Schwarz inequality:

$$|\langle \vec{v},\vec{w}\rangle| = |\cos\theta|\cdot\|\vec{v}\|\cdot\|\vec{w}\| \le \|\vec{v}\|\,\|\vec{w}\|.$$ $\square$

Thm Let V be an inner product space, and let $\vec{x},\vec{y}\in V$.

1. (Triangle Inequality) $\|\vec{x}+\vec{y}\|\le\|\vec{x}\|+\|\vec{y}\|$.
2. (Polarization Identity) $$\langle \vec{x},\vec{y}\rangle = \frac{1}{4}\Big(\underbrace{\|\vec{x}+\vec{y}\|^2-\|\vec{x}-\vec{y}\|^2}_{\text{just these if } V \text{ is real vector space}}+i\|\vec{x}+i\vec{y}\|^2-i\|\vec{x}-i\vec{y}\|^2\Big) = \frac{1}{4}\sum_{k=1}^{4} i^k\|\vec{x}+i^k\vec{y}\|^2$$
3. (Parallelogram Identity) $\|\vec{x}+\vec{y}\|^2+\|\vec{x}-\vec{y}\|^2 = 2(\|\vec{x}\|^2+\|\vec{y}\|^2)$

Proof All three parts are proven by direct computation.

① We compute

$$\|\vec{x}+\vec{y}\|^2 = \langle \vec{x}+\vec{y}, \vec{x}+\vec{y}\rangle = \|\vec{x}\|^2 + \underbrace{\langle \vec{x},\vec{y}\rangle + \overline{\langle \vec{x},\vec{y}\rangle}}_{=2\cdot\text{real part of }\langle \vec{x},\vec{y}\rangle} + \|\vec{y}\|^2$$
$$\le \|x\|^2 + 2|\langle \vec{x},\vec{y}\rangle| + \|\vec{y}\|^2$$
$$\text{Cauchy-Schwarz} \le \|x\|^2 + 2\|\vec{x}\|\cdot\|\vec{y}\| + \|\vec{y}\|^2 = (\|\vec{x}\| + \|\vec{y}\|)^2$$

② We first compute

$$\|\vec{x}+\vec{y}\|^2 + i\|\vec{x}+i\vec{y}\|^2 = \|\vec{x}\|^2 + \langle \vec{x},\vec{y}\rangle + \overline{\langle \vec{x},\vec{y}\rangle} + \|\vec{y}\|^2 + i\left(\|\vec{x}\|^2 - i\langle \vec{x},\vec{y}\rangle + i\overline{\langle \vec{x},\vec{y}\rangle} + \|\vec{y}\|^2\right)$$
$$= (1+i)(\|\vec{x}\|^2 + \|\vec{y}\|^2) + 2\langle \vec{x},\vec{y}\rangle$$

Swapping $-\vec{y}$ for $\vec{y}$ yields

$$\|\vec{x}-\vec{y}\|^2 + i\|\vec{x}-i\vec{y}\|^2 = (1+i)(\|\vec{x}\|^2 + \|\vec{y}\|^2) - 2\langle \vec{x},\vec{y}\rangle.$$

Thus

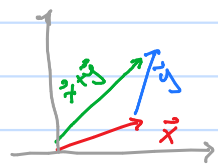
$$\sum_{k=1}^{4} (i)^k \|\vec{x} + (i)^k \vec{y}\|^2 = 4\langle \vec{x},\vec{y}\rangle.$$

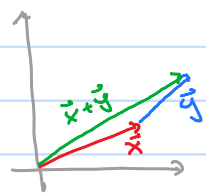
③ We compute

$$\|\vec{x}+\vec{y}\|^2 + \|\vec{x}-\vec{y}\|^2 = \|\vec{x}\|^2 + \langle \vec{x},\vec{y}\rangle + \overline{\langle \vec{x},\vec{y}\rangle} + \|\vec{y}\|^2 + \|\vec{x}\|^2 - \langle \vec{x},\vec{y}\rangle - \overline{\langle \vec{x},\vec{y}\rangle} + \|\vec{y}\|^2$$
$$= 2(\|\vec{x}\|^2 + \|\vec{y}\|^2)$$

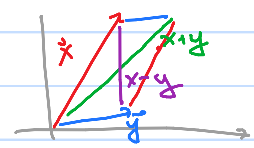
□

Ex For $V = \mathbb{R}^2$, ① says the length of $\vec{x}+\vec{y}$ is at most the length of $\vec{x}$ plus the length of $\vec{y}$. In other words, its shorter to travel straight along $\vec{x}+\vec{y}$, rather than along $\vec{x}$ then along $\vec{y}$.





② is a fact from geometry about parallelograms: the sum of the squares of the diagonals equals the sum of the squares of all four side lengths.



□

11/19

Normed Vector Spaces

- We have seen that all inner products give rise to norms, but these are objects that can be studied independently.

Def For a vector space V, a norm is a map from V to $\mathbb{R}$ satisfying:

① (Homogeneity) $\|\alpha\vec{v}\| = |\alpha|\,\|\vec{v}\|$ for all scalars α and $\vec{v} \in V$.

② (Triangle Inequality) $\|\vec{u}+\vec{v}\| \le \|\vec{u}\| + \|\vec{v}\|$ for all $\vec{u},\vec{v} \in V$.

③ (Non-negativity) $\|\vec{v}\| \ge 0$ for all $\vec{v} \in V$

④ (Non-degeneracy) $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0}$.

We call V along with the map $\|\cdot\|$ a normed vector space.

• Clearly every inner product space is a normed vector space, but the converse is <u>not</u> true.

$\boxed{\text{Ex}}$ Fix $1 \le p < \infty$. The following norms come from inner products iff $p=2$.

① For $\vec{x} = (x_1, \dots, x_n)^T \in \mathbb{F}^n$ define

$$\|\vec{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \leftarrow \text{to ensure homogeneity}$$

② Define

$$\ell^p(\mathbb{N}) := \{ (a_n)_{n \in \mathbb{N}} : a_n \in \mathbb{C}, \ \sum_{n=1}^{\infty} |a_n|^p < \infty \}$$

with

$$\|(a_n)_{n \in \mathbb{N}}\|_p := \left(\sum_{n=1}^{\infty} |a_n|^p \right)^{1/p}$$

③ For $f \in C[0,1]$ define

$$\|f\|_p := \left(\int_0^{\infty} |f(t)|^p dt \right)^{1/p}$$

④ For $A \in M_{m \times n}$ define

$$\|A\|_p := \mathrm{Tr}\left((A^* A)^{p/2} \right)^{1/p}$$

One can also make sense of $p = \infty$ for each of the above. For example

$$\|\vec{x}\|_\infty = \max_{1 \le i \le n} |x_i| \quad \text{and} \quad \|f\|_\infty = \max_{0 \le t \le 1} |f(t)|.$$ $\square$

1. • One way to tell if a norm comes from an inner product is the following theorem.

$\boxed{\text{Thm}}$ Let V be a normed vector space with norm $\|\cdot\|$. Then the norm comes from an inner product if and only if it satisfies the parallelogram identity:

$$\|\vec{x} + \vec{y}\|^2 + \|\vec{x} - \vec{y}\|^2 = 2(\|\vec{x}\|^2 + \|\vec{y}\|^2) \qquad \forall \vec{x}, \vec{y} \in V.$$

<u>Proof</u> ($\Longrightarrow$) This is ③ from the previous theorem.

($\Longleftarrow$) Homework 10. $\square$

5.2 Orthogonality

$\boxed{\text{Def}}$ In an inner product space V, we say $\vec{v}, \vec{w} \in V$ are orthogonal if $\langle \vec{v}, \vec{w} \rangle = 0$, in which case we write $\vec{v} \perp \vec{w}$.

$\boxed{\text{Ex}}$ ① In $\mathbb{F}^3$,

$$\left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ a \\ b \end{pmatrix} \right\rangle = 1 \cdot 0 + 0 \cdot a + 0 \cdot b = 0$$

In particular $\vec{e}_1 \perp \vec{e}_2$ and $\vec{e}_1 \perp \vec{e}_3$

② For the std. basis $\vec{e}_1, \ldots, \vec{e}_n \in \mathbb{F}^n$, $\vec{e}_i \perp \vec{e}_j$ whenever $i \neq j$.

③ $C[-1,1]$ has inner product

$$\langle f, g \rangle_2 = \int_{-1}^{1} f(t) \overline{g(t)} \, dt$$

Let $f(t) = t^{2n}$, $g(t) = t^{2m+1}$. Then

$$\langle f, g \rangle_2 = \int_{-1}^{1} t^{2n} t^{2m+1} dt = \int_{-1}^{1} t^{\overbrace{2(n+m)+1}^{\text{odd}}} dt = 0$$

So $f \perp g$. $\square$

- Observe that $\vec{v} \perp \vec{v} \iff \langle \vec{v}, \vec{v} \rangle = 0 \iff \vec{v} = \vec{0}$.
- If $\vec{v} \perp \vec{w}$ then

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 \qquad \text{(Pythagorean Identity)}$$

Indeed:

$$\|\vec{v} + \vec{w}\|^2 = \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle = \|\vec{v}\|^2 + \cancelto{0}{\langle \vec{v}, \vec{w} \rangle} + \cancelto{0}{\langle \vec{w}, \vec{v} \rangle} + \|\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2.$$

$\boxed{\text{Def}}$ Let V be an inner product space. We say $\vec{v} \in V$ is orthogonal to a subspace $E \subset V$ if $\vec{v} \perp \vec{w}$ for all $\vec{w} \in E$. We write $\vec{v} \perp E$.

We say two subspaces $E, F \subset V$ are orthogonal if $\vec{v} \perp \vec{w}$ for all $\vec{v} \in E$ and $\vec{w} \in F$.

$\boxed{\text{Ex}}$ Let $A \in M_{m \times n}(\mathbb{C})$. Then $\text{Ker}(A) \perp \text{Ran}(A^*)$. Indeed, let $\vec{v} \in \text{Ker}(A)$ and $\vec{w} \in \text{Ran}(A^*)$. Then $\vec{w} = A^* \vec{u}$ for some $\vec{u} \in \mathbb{C}^m$. Now

$$\langle \vec{v}, \vec{w} \rangle = \vec{w}^* \vec{v} = (A^* \vec{u})^* \vec{v} = (\vec{u}^* A) \vec{v} = \vec{u}^* (A \vec{v}) = \vec{u}^* \vec{0} = 0.$$

Thus $\vec{v} \perp \vec{w}$. $\square$

$\boxed{\text{Lemma}}$ Let $\vec{v}_1, \ldots, \vec{v}_r \in V$ and let $E = \text{span}\{\vec{v}_1, \ldots, \vec{v}_r\}$. Then for $\vec{v} \in V$, we have $\vec{v} \perp E$ if and only if $\vec{v} \perp \vec{v}_k$ for each $k = 1, \ldots, r$.

Proof ($\Longrightarrow$) Since $\vec{v}_1, \ldots, \vec{v}_r \in E$, this follows by definition of $\vec{v} \perp E$.

($\Longleftarrow$) Suppose $\vec{v} \perp \vec{v}_k$ for each $k = 1, \ldots, r$. Let $\vec{w} \in E$, then

$$\vec{w} = \alpha_1 \vec{v}_1 + \cdots + \alpha_r \vec{v}_r$$

for some scalars $\alpha_1, \ldots, \alpha_r$. We have

$$\langle \vec{v}, \vec{w} \rangle = \langle \vec{v}, \alpha_1 \vec{v}_1 + \cdots + \alpha_r \vec{v}_r \rangle = \bar{\alpha}_1 \langle \vec{v}, \vec{v}_1 \rangle + \cdots + \bar{\alpha}_r \langle \vec{v}, \vec{v}_r \rangle = 0$$

So $\vec{v} \perp \vec{w}$, and since $\vec{w} \in E$ was arbitrary we have $\vec{v} \perp E$. $\square$

$\boxed{\text{Def}}$ In an inner product space V, a system $\vec{v}_1, \ldots, \vec{v}_n \in V$ is called orthogonal if $\vec{v}_i \perp \vec{v}_j$ whenever $i \neq j$. The system is said to be orthonormal if it is orthogonal and $\|\vec{v}_i\| = 1$ for each $i = 1, \ldots, n$.

$\boxed{\text{Ex}}$ In $\mathbb{R}^n$, $\vec{e}_1, \ldots, \vec{e}_n$ is orthonormal. $\square$

11/21

$\boxed{\text{Lemma}}$ (Generalized Pythagorean Identity) Let $\vec{v}_1, \ldots, \vec{v}_n$ be an orthogonal system. Then

$$\| \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n \|^2 = |\alpha_1|^2 \|\vec{v}_1\|^2 + \cdots + |\alpha_n|^2 \|\vec{v}_n\|^2$$

for any scalars $\alpha_1, \ldots, \alpha_n$.

Proof We compute

$$\begin{aligned} \Big\| \sum_{j=1}^n \alpha_j \vec{v}_j \Big\|^2 &= \Big\langle \sum_{j=1}^n \alpha_j \vec{v}_j , \sum_{k=1}^n \alpha_k \vec{v}_k \Big\rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n \alpha_j \bar{\alpha}_k \langle \vec{v}_j, \vec{v}_k \rangle \\ &= \sum_{j=1}^n \alpha_j \bar{\alpha}_j \langle \vec{v}_j, \vec{v}_j \rangle = \sum_{j=1}^n |\alpha_j|^2 \|\vec{v}_j\|^2. \end{aligned}$$ $\square$

$\boxed{\text{Cor}}$ Any orthogonal system $\vec{v}_1, \ldots, \vec{v}_n$ of non-zero vectors is linearly independent.

Proof Suppose

$$\alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \vec{0}$$

for some scalars $\alpha_1, \ldots, \alpha_n$. Then using the lemma we have

$$0 = \|\vec{0}\|^2 = \| \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n \|^2 = |\alpha_1|^2 \|\vec{v}_1\|^2 + \cdots + |\alpha_n|^2 \|\vec{v}_n\|^2$$

Note $\|\vec{v}_j\| \neq 0$ since $\vec{v}_j \neq \vec{0}$ for each $j = 1, \ldots, n$. So the only way the sum of non-negative numbers on the right equals zero is if $|\alpha_1|^2 = \cdots = |\alpha_n|^2 = 0 \Rightarrow \alpha_1 = \cdots = \alpha_n = 0$.

Thus $\vec{v}_1, \ldots, \vec{v}_n$ are linearly independent. $\square$

$\boxed{\text{Ex}}$ In $C[-1,1]$, consider

$$f_n(t) := e^{i\pi n t} \qquad n \in \mathbb{Z}$$

We claim $f_n \perp f_m$ for $n \neq m$. Fact $e^{ix} = \cos(x) + i\sin(x)$ for $x \in \mathbb{R}$. Thus

$$\langle f_n, f_m \rangle_2 = \int_{-1}^1 f_n(t) \overline{f_m(t)}\, dt = \int_{-1}^1 e^{i\pi n t} e^{-i\pi m t} dt = \int_{-1}^1 e^{i\pi(n-m)t} dt$$

$$= \int_{-1}^1 \cos(\pi(n-m)t) + i\sin(\pi(n-m)t)\, dt$$

$n - m \neq 0$

$$= \left[\frac{\sin(\pi(n-m)t)}{\pi(n-m)} - i \frac{\cos(\pi(n-m)t)}{\pi(n-m)} \right]_{-1}^{1} = 0$$

Exercise: compute $\langle f_n, f_n \rangle_2$.

Thus for any $N \in \mathbb{N}$, $f_{-N}, f_{-(N-1)}, \ldots, f_{-1}, f_0, f_1, \ldots, f_N$ is an orthogonal system, and so by the lemma it is linearly independent. Since we can increase N to however large we want, it follows that $C[-1,1]$ is infinite dimensional.

Fact "Fourier Analysis" says $\{f_n : n \in \mathbb{Z}\}$ is generating for $C[-1,1]$ (in a certain sense). $\square$

Orthonormal Bases

Def An orthogonal (resp. orthonormal) basis is an orthogonal (resp. orthonormal) system that is also a basis (i.e. generating).

- Recall that given an arbitrary basis $\mathcal{B}$ for V, computing coordinate vectors can be hard. E.g. for $\vec{x} \in \mathbb{F}^n$

$$[\vec{x}]_{\mathcal{B}} = [I]_S^{\mathcal{B}} [\vec{x}]_S = \left([I]_{\mathcal{B}}^S\right)^{-1} \vec{x}$$

What makes orthogonal basis nice is that these are much easier to find.

Prop Let V be an inner product space, and let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be an orthogonal basis. Then for any $\vec{x} \in V$,

$$[\vec{x}]_{\mathcal{B}} = (\alpha_1, \dots, \alpha_n)^T$$

where

$$* \quad \alpha_k = \frac{\langle \vec{x}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \qquad k = 1, \dots, n.$$

Proof Since $\mathcal{B}$ is a basis we know

$$\vec{x} = \alpha_1 \vec{v}_1 + \dots + \alpha_n \vec{v}_n$$

for some scalars $\alpha_1, \dots, \alpha_n$. We must show these scalars are as in $(*)$. For $1 \le k \le n$, we have

$$\langle \vec{x}, \vec{v}_k \rangle = \left\langle \sum_{j=1}^n \alpha_j \vec{v}_j, \vec{v}_k \right\rangle = \sum_{j=1}^n \alpha_j \langle \vec{v}_j, \vec{v}_k \rangle = \alpha_k \langle \vec{v}_k, \vec{v}_k \rangle = \alpha_k \|\vec{v}_k\|^2.$$

Solving for α_k yields $(*)$. $\square$

- Note that this proposition implies

$$\vec{x} = \sum_{k=1}^n \frac{\langle \vec{x}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \vec{v}_k$$

for all $\vec{x} \in \vec{V}$. If we further assume that $\vec{v}_1, \dots, \vec{v}_n$ is orthonormal, then this simplifies to

$$\vec{x} = \sum_{k=1}^n \langle \vec{x}, \vec{v}_k \rangle \vec{v}_k.$$

5.3 Orthogonal projections and Gram-Schmidt orthogonalization

Def Let V be an inner product space and let $E \subset V$ be a subspace. An orthogonal projection onto E is a linear transformation $P_E : V \to V$ satisfying

1. $P_E(\vec{v}) \in E$ for all $\vec{v} \in V$.
2. $(\vec{v} - P_E(\vec{v})) \perp E$ for all $\vec{v} \in V$.

Ex Let $E = \text{span}\{\vec{e}_2, \vec{e}_3\} \le \mathbb{F}^3$. Then $P_E\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}$. Indeed, it clearly satisfies 1 and

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} - P_E\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} = x\vec{e}_1 \perp E.$$

$\square$

Prop Let V be an inner product space and let E be a subspace. Suppose $\vec{v}_1, \dots, \vec{v}_r$ is an orthogonal basis for E. Then the orthogonal projection onto E is given by

$$P_E(\vec{v}) = \sum_{k=1}^{r} \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \vec{v}_k.$$

Proof First note that P_E is linear by the linearity of the inner product. Also,

$$P_E(\vec{v}) \in \text{span}\{\vec{v}_1, \dots, \vec{v}_r\} = E$$

so P_E satisfies 1. So it remains to check 2. By the first Lemma in Section 5.2, it suffices to show $(\vec{v} - P_E(\vec{v})) \perp \vec{v}_k$ for each $k = 1, \dots, n$. We have

$$\begin{aligned}
\langle \vec{v} - P_E(\vec{v}), \vec{v}_k \rangle &= \langle \vec{v}, \vec{v}_k \rangle - \langle P_E(\vec{v}), \vec{v}_k \rangle \\
&= \langle \vec{v}, \vec{v}_k \rangle - \left\langle \sum_{j=1}^{r} \frac{\langle \vec{v}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2}, \vec{v}_k \right\rangle \\
&= \langle \vec{v}, \vec{v}_k \rangle - \sum_{j=1}^{n} \frac{\langle \vec{v}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2} \langle \vec{v}_j, \vec{v}_k \rangle \\
&= \langle \vec{v}, \vec{v}_k \rangle - \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \langle \vec{v}_k, \vec{v}_k \rangle \\
&= \langle \vec{v}, \vec{v}_k \rangle - \langle \vec{v}, \vec{v}_k \rangle = 0.
\end{aligned}$$

$\square$

Ex Recall that in $\mathbb{C}^n$, $\langle \vec{x}, \vec{y} \rangle = \vec{y}^* \vec{x}$. Thus for a subspace $E \subset \mathbb{C}^n$ with orthogonal basis $\vec{v}_1, \dots, \vec{v}_r$ the Proposition says:

$$P_E(\vec{x}) = \sum_{k=1}^{n} \frac{1}{\|\vec{v}_k\|^2} (\vec{v}_k^* \vec{x}) \vec{v}_k = \sum_{k=1}^{n} \frac{1}{\|\vec{v}_k\|^2} \vec{v}_k (\vec{v}_k^* \vec{x})$$

Observe that $\vec{v}_k \vec{v}_k^* \in M_{n \times 1} \cdot M_{1 \times n} = M_{n \times n}$. It follows that

$$[P_E] = \sum_{k=1}^{n} \frac{1}{\|\vec{v}_k\|^2} \vec{v}_k \vec{v}_k^* \in M_{n \times n}.$$

$\square$

Thm Let V be an inner product space and let P_E be an orthogonal projection onto a subspace E. Then for all $\vec{v} \in V$ and all $\vec{x} \in E$

$$\|\vec{v} - P_E(\vec{v})\| \le \|\vec{v} - \vec{x}\|$$

That is, $P_E(\vec{v})$ minimizes the distance from $\vec{v}$ to E. Moreover, if for some $\vec{x} \in E$
$$\| \vec{v} - P_E(\vec{v}) \| = \| \vec{v} - \vec{x} \|$$
Then $P_E(\vec{v}) = \vec{x}$. Consequently, any orthogonal projection onto E is unique.

Proof For $\vec{v} \in V$ and $\vec{x} \in E$, note that
$$\vec{v} - \vec{x} = \underbrace{\vec{v} - P_E(\vec{v})}_{\perp E} + \underbrace{P_E(\vec{v})}_{\in E} - \vec{x}$$
So by the Pythagorean theorem
$$\begin{aligned} \| \vec{v} - \vec{x} \|^2 &= \| (\vec{v} - P_E(\vec{v})) + (P_E(\vec{v}) - \vec{x}) \|^2 \\ &= \| \vec{v} - P_E(\vec{v}) \|^2 + \| P_E(\vec{v}) - \vec{x} \|^2 \geq \| \vec{v} - P_E(\vec{v}) \|^2 \end{aligned}$$
Which proves the first part. Note that $\| \vec{v} - \vec{x} \| = \| \vec{v} - P_E(\vec{v}) \|$ is only possible if
$$\| P_E(\vec{v}) - \vec{x} \| = 0 \iff P_E(\vec{v}) - \vec{x} = \vec{0} \iff P_E(\vec{v}) = \vec{x}.$$
Finally, let P_E' be another orthogonal projection onto E. Then by the first part of the theorem
$$\| \vec{v} - P_E(\vec{v}) \| \leq \| \vec{v} - P_E'(\vec{v}) \|$$
and reversing the roles of P_E and P_E' we get $\| \vec{v} - P_E(\vec{v}) \| = \| \vec{v} - P_E'(\vec{v}) \|$. So by the second part of the theorem we have $P_E(\vec{v}) = P_E'(\vec{v})$. Since $\vec{v} \in V$ was arbitrary, this implies $P_E = P_E'$. $\square$

11/26

Gram-Schmidt Orthogonalization Algorithm

- From the first proposition in this section, we see that we can always define P_E provided E has an orthogonal basis. In this subsection we show this is always the case.

The following algorithm takes in a linearly independent system $\vec{x}_1, \dots, \vec{x}_n$ and constructs an orthogonal system $\vec{v}_1, \dots, \vec{v}_n$ s.t.
$$\operatorname{span}\{\vec{x}_1, \dots, \vec{x}_n\} = \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_n\}$$
Moreover, for each $r \leq n$ it satisfies
$$\operatorname{span}\{\vec{x}_1, \dots, \vec{x}_r\} = \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_r\}$$

$\boxed{\text{Step 1}}$ Set $\vec{v}_1 := \vec{x}_1$ and denote $E_1 := \operatorname{span}\{\vec{x}_1\} = \operatorname{span}\{\vec{v}_1\}$

$\boxed{\text{Step 2}}$ Set
$$\vec{v}_2 := \vec{x}_2 - P_{E_1}\vec{x}_2 = \vec{x}_2 - \frac{\langle \vec{x}_2, \vec{v}_1 \rangle}{\| \vec{v}_1 \|^2} \vec{v}_1$$
Note that $\vec{v}_2 \perp E_1$ by definition of P_{E_1}. In particular, $\vec{v}_2 \perp \vec{v}_1$.

Denote $E_2 := \operatorname{span}\{\vec{v}_1, \vec{v}_2\} = \operatorname{span}\{\vec{x}_1, \vec{x}_2\}$

$\boxed{\text{Step 3}}$ Set
$$\vec{v}_3 := \vec{x}_3 - P_{E_2}\vec{x}_3 = \vec{x}_3 - \frac{\langle \vec{x}_3, \vec{v}_1 \rangle}{\| \vec{v}_1 \|^2} \vec{v}_1 - \frac{\langle \vec{x}_3, \vec{v}_2 \rangle}{\| \vec{v}_2 \|^2} \vec{v}_2$$

Then $\vec{v}_3 \perp E_2$ and in particular $\vec{v}_3 \perp \vec{v}_1$, $\vec{v}_3 \perp \vec{v}_2$. Note that $\vec{x}_3 \notin E_2$ since $E_2 = \operatorname{span}\{\vec{x}_1, \vec{x}_2\}$ and $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are lin. indep. Thus $\vec{v}_3 \neq \vec{0}$.

$\vdots$

Step r+1 Suppose we have already iterated the above process r times to construct an orthogonal system $\vec{v}_1, \dots, \vec{v}_r$ (of non-zero vectors) such that

$$E_r := \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_r\} = \operatorname{span}\{\vec{x}_1, \dots, \vec{x}_r\}$$

Define

$$\vec{v}_{r+1} := \vec{x}_{r+1} - P_{E_r}\vec{x}_{r+1} = \vec{x}_{r+1} - \sum_{k=1}^{r} \frac{\langle \vec{x}_{r+1}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \vec{v}_k$$

Then $\vec{v}_{r+1} \perp E_r$ and therefore $\vec{v}_{r+1} \perp \vec{v}_k$ for $k = 1, \dots, r$. Also, since $\vec{x}_{r+1} \notin E_r$ by lin. indep. of $\vec{x}_1, \dots, \vec{x}_r, \vec{x}_{r+1}$, we have $\vec{v}_{r+1} \neq \vec{0}$.

Iterating this process n times yields the desired orthogonal system $\vec{v}_1, \dots, \vec{v}_n$. $\square$

Ex In $\mathbb{R}^3$

$$\vec{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \vec{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \qquad \vec{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

Step 1 $\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Step 2 $$\vec{v}_2 = \vec{x}_2 - \frac{\langle \vec{x}_2, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} \vec{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{0+1+2}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Step 3 $$\vec{v}_3 = \vec{x}_3 - \frac{\langle \vec{x}_3, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} \vec{v}_1 - \frac{\langle \vec{x}_3, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2} \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} - \frac{1+0+2}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{-1+0+2}{2} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -1/2 \\ 0 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix}$$

So

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad \vec{v}_3 = \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix}$$

Let $E := \operatorname{span}\{\vec{x}_1, \vec{x}_2\} = \operatorname{span}\{\vec{v}_1, \vec{v}_2\}$. Then $\vec{v}_1, \vec{v}_2$ is an orthogonal basis for E and so we know the formula for P_E. Let's write out its matrix representation (with respect to the standard basis S):

$$[P_E]_S^S = \frac{1}{\|\vec{v}_1\|^2} \vec{v}_1 \cdot \vec{v}_1^* + \frac{1}{\|\vec{v}_2\|^2} \vec{v}_2 \cdot \vec{v}_2^* = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot (1\ 1\ 1) + \frac{1}{2} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \cdot (-1\ 0\ 1)$$

$$= \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{pmatrix}$$ $\square$

Rem Note that if $\vec{u}_k := \dfrac{\vec{v}_k}{\|\vec{v}_k\|}$ for $k = 1, \dots, n$ then $\vec{u}_1, \dots, \vec{u}_n$ is an ortho<u>normal</u> system which satisfies

$$\operatorname{span}\{\vec{u}_1, \dots, \vec{u}_r\} = \operatorname{span}\{\vec{x}_1, \dots, \vec{x}_r\} \qquad r = 1, \dots, n.$$ $\square$

Orthogonal Complement

Def For a subset $S \subset V$, its orthogonal complement is the set
$$S^\perp := \{\vec{x} \in V : \vec{x} \perp S\}$$

Prop $S^\perp$ is always a subspace (even if S is not). Also
$$(S^\perp)^\perp = \operatorname{span} S$$
In particular, $S = (S^\perp)^\perp$ if and only if S is a subspace.

Proof Homework 11. □

Lemma Let $E \subset V$ be a subspace. Then $E \cap E^\perp = \{\vec{0}\}$.

Proof Let $\vec{x} \in E \cap E^\perp$ then
$$\langle \underbrace{\vec{x}}_{E \ni}, \underbrace{\vec{x}}_{\in E^\perp} \rangle = 0$$
so $\vec{x} = \vec{0}$ by non-degeneracy. □

Thm Let V be an inner product space and let $E \subset V$ be a subspace. Then for every $\vec{v} \in V$ there exists unique $\vec{v}_1 \in E$ and $\vec{v}_2 \in E^\perp$ s.t.
$$\vec{v} = \vec{v}_1 + \vec{v}_2.$$

Proof Let P_E be the orthogonal projection onto E. Given $\vec{v} \in V$, set $\vec{v}_1 := P_E(\vec{v})$ and $\vec{v}_2 := \vec{v} - P_E(\vec{v})$. Then $\vec{v}_1 \in E$ and $\vec{v}_2 \in E^\perp$ by def. of P_E. Also
$$\vec{v}_1 + \vec{v}_2 = P_E(\vec{v}) + \vec{v} - P_E(\vec{v}) = \vec{v}.$$
Suppose $\vec{w}_1 \in E$ and $\vec{w}_2 \in E^\perp$ also satisfy $\vec{w}_1 + \vec{w}_2 = \vec{v}$. Then
$$\vec{v}_1 + \vec{v}_2 = \vec{w}_1 + \vec{w}_2$$
$$\vec{v}_1 - \vec{w}_1 = \vec{w}_2 - \vec{v}_2$$
Note that $\vec{v}_1 - \vec{w}_1 \in E$ while $\vec{w}_2 - \vec{v}_2 \in E^\perp$. Thus $\vec{v}_1 - \vec{w}_1 = \vec{w}_2 - \vec{v}_2 \in E \cap E^\perp = \{\vec{0}\}$ by the Lemma.
So $\vec{v}_1 - \vec{w}_1 = \vec{0} \Rightarrow \vec{v}_1 = \vec{w}_1$ and $\vec{w}_2 - \vec{v}_2 = \vec{0} \Rightarrow \vec{w}_2 = \vec{v}_2$. Hence $\vec{v}_1$ and $\vec{v}_2$ are unique. □

- Whenever $E_1, E_2 \subset V$ are subspaces s.t. every $\vec{v} \in V$ can be written uniquely as $\vec{v} = \vec{v}_1 + \vec{v}_2$ for $\vec{v}_1 \in E_1$ and $\vec{v}_2 \in E_2$, we write
$$V = E_1 \oplus E_2.$$
The above theorem says
$$V = E \oplus E^\perp.$$

Cor For a subspace $E \leq V$,
$$\dim(E) + \dim(E^\perp) = \dim(V)$$

Proof Using Gram-Schmidt, we can find orthogonal bases $\mathcal{B}$ and $\mathcal{C}$ for E and $E^\perp$ respectively. Then $\mathcal{B} \cup \mathcal{C}$ is orthogonal and therefore lin. indep. The theorem tells us $\mathcal{B} \cup \mathcal{C}$ is generating. Hence $\mathcal{B} \cup \mathcal{C}$ is a basis for V. □

$\boxed{5.4}$ Method of Least Squares

- Let $A\vec{x}=\vec{b}$ be an inconsistent linear system. Since there is no solution, we cannot solve it exactly, but we can solve it approximately. That is, we can find $\vec{x}$ s.t. $\|A\vec{x}-\vec{b}\|$ is as small as possible

- Note:

$$\|A\vec{x}-\vec{b}\|^2 = \sum_{i=1}^{n} |(A\vec{x})_i - b_i|^2 = \underbrace{\sum_{i=1}^{n} \Big| \sum_{j=1}^{n} (A)_{ij}\, x_j - b_i \Big|^2}_{\text{"least squares"}}$$

- Obviously $A\vec{x}\in \mathrm{Ran}(A)$ for all $\vec{x}\in \mathbb{F}^n$. Let P be the orthogonal projection onto $\mathrm{Ran}(A)$. Then by a theorem from Section 5.3 we have:

$$\|\vec{b}-P\vec{b}\| \le \|\vec{b}-A\vec{x}\| \qquad \forall \vec{x}\in \mathbb{F}^n.$$

So we want to find $\vec{x}$ s.t. $A\vec{x}=P\vec{b}$.

$\boxed{\text{Thm}}$ Let $A\in M_{m\times n}$ and let P the orthogonal projection onto $\mathrm{Ran}(A)$. For $\vec{x}\in\mathbb{F}^n$, $A\vec{x}=P\vec{b}$ if and only if $A^*A\vec{x}=A^*\vec{b}$.

Proof Since $A\vec{x}\in\mathrm{Ran}(A)$, $A\vec{x}=P\vec{b}$ iff $(\vec{b}-A\vec{x})\perp \mathrm{Ran}(A)$. Let $\vec{a}_1,\dots,\vec{a}_n$ be the columns of A. Then $\mathrm{Ran}(A)$ (the column space of A) is generated by $\vec{a}_1,\dots,\vec{a}_n$, and so $(\vec{b}-A\vec{x})\perp\mathrm{Ran}(A)$ iff

$$\langle \vec{b}-A\vec{x}, \vec{a}_j\rangle = 0 \quad \text{for each } j=1,\dots,n.$$
$$\vec{a}_j^{\,*}(\vec{b}-A\vec{x})=0$$

iff

$$\vec{0} = \begin{pmatrix} \vec{a}_1^{\,*}(\vec{b}-A\vec{x}) \\ \vdots \\ \vec{a}_n^{\,*}(\vec{b}-A\vec{x}) \end{pmatrix} = A^*(\vec{b}-A\vec{x})$$

iff $A^*A\vec{x}=A^*\vec{b}$. $\square$

12/3

- Suppose we are given data points (x_i,y_i), $i=1,\dots,n$ and want to find the line $y=ax+b$ that "best fits" our data. That means find the values a,b that minimize:

$$\sum_{i=1}^{n} |y_i-(ax_i+b)|^2$$

This is equivalent to applying the method of least squares to the linear system

$$\underbrace{\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\vec{b}}$$

The theorem tells us this is equivalent to solving $A^*A\vec{x}=A^*\vec{b}$.

$\boxed{\text{Ex}}$ Suppose our data set is:

$$(-2,4),\ (-1,2),\ (0,1),\ (2,1),\ (3,1)$$

So we will find the least squares solution to

$$\begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

we have

$$A^*A = \begin{pmatrix} -2 & -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix}$$

and

$$A^*\vec{b} = \begin{pmatrix} -2 & -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \end{pmatrix}$$

So we must solve:

$$\begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \end{pmatrix} \rightsquigarrow \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1/2 \\ 2 \end{pmatrix}$$

Thus the line $y = -\frac{1}{2}x + 2$ best fits the data. $\square$

- Note that a typical reason for a lin. sys. $A\vec{x} = \vec{b}$ to be inconsistent is that it is "over determined"; that is, it has more equations than unknowns: $m > n$ for $A \in M_{m \times n}$. That is, A is tall. But then A^*A is $n \times n$ and so has been compressed.

- There was not anything special about trying to fit a line to our data. If we wanted to find a parabola $y = ax^2 + bx + c$ that best fits our data, we would apply the least squares method to

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ & \vdots & \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$\boxed{\text{Ex}}$ Using the same data as before, we have

$$\begin{pmatrix} 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Then

$$A^*A = \begin{pmatrix} 114 & 26 & 18 \\ 26 & 18 & 2 \\ 18 & 2 & 5 \end{pmatrix} \quad \text{and} \quad A^*b = \begin{pmatrix} 31 \\ -5 \\ 9 \end{pmatrix}$$

And $A^*A\vec{x} = A^*b$ has solution

$$\begin{pmatrix} 43/154 \\ -62/77 \\ 86/77 \end{pmatrix}$$

So that the parabola best fitting our data is $y = \frac{43}{154}x^2 + \frac{-62}{77}x + \frac{86}{77}$. $\square$

- More generally, if we think our data should be best modeled by a curve

of the form
$$y = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_\ell f_\ell(x)$$
for functions $f_1, \ldots, f_n$, then we would apply the least squares method to
$$\begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_\ell(x_1) \\ \vdots & \vdots & & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_\ell(x_\ell) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_\ell \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

- One issue we might run into is that A^*A need not be invertible. However, since it is square, we know it is invertible iff it has a pivot in every column iff $\operatorname{Ker}(A^*A) = \{\vec{0}\}$.

$\boxed{\text{Thm}}$ For $A \in M_{m\times n}$, $\operatorname{Ker}(A) = \operatorname{Ker}(A^*A)$

Proof: Clearly $\operatorname{Ker}(A) \subset \operatorname{Ker}(A^*A)$. Now, suppose $\vec{x} \in \operatorname{Ker}(A^*A)$. Then
$$\|A\vec{x}\|^2 = \langle A\vec{x}, A\vec{x}\rangle = \langle \vec{x}, A^*A\vec{x}\rangle = \langle \vec{x}, \vec{0}\rangle = 0$$
where we have used an exercise on Homework 12. Thus $A\vec{x} = \vec{0} \Rightarrow \vec{x} \in \operatorname{Ker}(A)$. It follows that $\operatorname{Ker}(A^*A) \subset \operatorname{Ker}(A)$ and so we have equality. $\square$

$\boxed{\text{Cor}}$ A^*A is invertible if and only if $\operatorname{rank}(A) = n$.

Proof: From the discussion preceding the theorem, we know A^*A is invertible iff $\operatorname{Ker}(A^*A) = \{\vec{0}\}$. By the theorem, this is equivalent to $\operatorname{Ker}(A) = \{\vec{0}\}$, and by rank-nullity this is in turn equivalent to $\operatorname{rank}(A) = n$. $\square$

$\boxed{\text{Cor}}$ If $\operatorname{rank}(A) = n$, then $P_{\operatorname{Ran}(A)} = A(A^*A)^{-1}A^*$.

Proof: By the previous corollary, A^*A is invertible. By the first theorem in this section, for any $\vec{b} \in \mathbb{F}^m$ we have $P_{\operatorname{Ran}(A)}\vec{b} = A\vec{x}$ where $\vec{x}$ is a solution of $A^*A\vec{x} = A^*\vec{b}$. Thus
$$\vec{x} = (A^*A)^{-1}A\vec{b}$$
and so
$$P_{\operatorname{Ran}(A)}\vec{b} = A\vec{x} = A(A^*A)^{-1}A^*\vec{b}.$$
Since $\vec{b} \in \mathbb{F}^m$ was arbitrary, we obtain $P_{\operatorname{Ran}(A)} = A(A^*A)^{-1}A^*\vec{b}$. $\square$

Polar Decomposition and Singular Values

Def A self-adjoint matrix $A \in M_{n\times n}(\mathbb{C})$ is called <u>positive semi-definite</u> if

$$\langle A\vec{x}, \vec{x}\rangle \geq 0 \qquad \forall \vec{x} \in \mathbb{C}^n$$

In this case we write $A \geq 0$. A is called <u>positive definite</u> if

$$\langle A\vec{x}, \vec{x}\rangle > 0 \qquad \forall \vec{x} \in \mathbb{C}^n \text{ with } \vec{x} \neq \vec{0}.$$

In this case we write $A > 0$.

Ex Let $A \in M_{m\times n}(\mathbb{C})$ (not necessarily square). Then by Homework 12 Exercise 4, for any $\vec{x} \in \mathbb{C}^n$

$$\langle A^*A\vec{x}, \vec{x}\rangle = \langle A\vec{x}, A\vec{x}\rangle = \|A x\|^2 \geq 0.$$

Thus $A^*A \in M_{n\times n}(\mathbb{C})$ is positive semi-definite. $\square$

Thm Let $A \in M_{n\times n}(\mathbb{C})$ be self-adjoint.

1. A is positive semi-definite if and only if $\sigma(A) \subset [0,\infty)$
2. A is positive definite if and only if $\sigma(A) \subset (0,\infty)$

Proof By Homework 12 Exercise 4, A is diagonalizable

$$A = U \underbrace{\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}}_{=:D} U^*$$

with U a unitary matrix.

By Homework 12 Exercise 2,

$$\langle A\vec{x}, \vec{x}\rangle = \langle UDU^*\vec{x}, \vec{x}\rangle = \langle DU^*\vec{x}, U^*\vec{x}\rangle$$

Thus A is positive (semi-)definite if and only if D is positive (semi-) definite.

For $\vec{x} = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$ we have

$$\langle D\vec{x}, \vec{x}\rangle = \left\langle \begin{pmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\rangle = \lambda_1 x_1 \bar{x}_1 + \cdots + \lambda_n x_n \bar{x}_n$$
$$= \lambda_1 |x_1|^2 + \cdots + \lambda_n |x_n|^2$$

This is > 0 (≥ 0) for <u>all</u> $\vec{x}$ if and only if $\lambda_1, \ldots, \lambda_n > 0$ (≥ 0). $\square$

Cor If $A \in M_{n\times n}(\mathbb{C})$ is positive semi-definite then there exists a positive semi-definite matrix $\sqrt{A} \in M_{n\times n}(\mathbb{C})$ such that $A = (\sqrt{A})^2$.

Proof With U and D as in the theorem, take

$$\sqrt{A} := U \begin{pmatrix} \lambda_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{1/2} \end{pmatrix} U^*.$$

$\square$

not square

Def For $A \in M_{m\times n}(\mathbb{C})$, the <u>absolute value</u> of A is

$$|A| := \sqrt{A^*A} \in M_{n\times n}(\mathbb{C})$$

The eigenvalues of $|A|$ are called the <u>singular values</u> of A.

Rem Note that the eigenvalues of $|A| = \sqrt{A^*A}$ are the square-roots of the eigenvalues of A^*A.

- Suppose A has singular values $\sigma_1, \dots, \sigma_n$. Let $\vec{v}_1, \dots, \vec{v}_n$ be an orthonormal basis of eigenvectors for $|A|$ corresponding to eigenvalues $\sigma_1, \dots, \sigma_n$ (respectively). Observe

$$A^*A\vec{v}_i = |A|^2\vec{v}_i = |A|(|A|\vec{v}_i) = |A|(\sigma_i \vec{v}_i) = \sigma_i^2 \vec{v}_i.$$

So $\vec{v}_i$ is an eigenvector of A^*A with eigenvalue σ_i^2.

Lemma 1 Assume $\sigma_1, \dots, \sigma_r$ are the non-zero singular values of A for some $r \leq n$. If

$$\vec{w}_k := \frac{1}{\sigma_k} A\vec{v}_k \qquad k = 1, \dots, r$$

then $\vec{w}_1, \dots, \vec{w}_r$ is an orthonormal system.

Proof For $j, k = 1, \dots, r$ we have

$$\langle \vec{w}_j, \vec{w}_k \rangle = \frac{1}{\sigma_j \sigma_k} \langle A\vec{v}_j, A\vec{v}_k \rangle = \frac{1}{\sigma_j \sigma_k} \langle A^*A\vec{v}_j, \vec{v}_k \rangle = \frac{1}{\sigma_j \sigma_k} \langle \sigma_j^2 \vec{v}_j, \vec{v}_k \rangle = \frac{\sigma_j^2}{\sigma_j \sigma_k} \langle \vec{v}_j, \vec{v}_k \rangle.$$

If $j \neq k$, this is zero. Otherwise $j = k$ and we obtain $\frac{\sigma_j^2}{\sigma_j \sigma_j} \|\vec{v}_j\|^2 = 1$ $\square$

Lemma 2 For any $\vec{x} \in \mathbb{C}^n$

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

Consequently, $\ker(|A|) = \ker(A)$

Proof We compute

$$\| |A|\vec{x} \|^2 = \langle |A|\vec{x}, |A|\vec{x} \rangle = \langle |A|^2 \vec{x}, \vec{x} \rangle = \langle A^*A\vec{x}, \vec{x} \rangle = \langle A\vec{x}, A\vec{x} \rangle = \| A\vec{x} \|^2.$$

In particular, $|A|\vec{x} = \vec{0}$ iff $\| |A|\vec{x} \| = 0$ iff $\|A\vec{x}\| = 0$ iff $A\vec{x} = \vec{0}$. $\square$

Thm (Polar Decomposition) For a square matrix $A \in M_{n\times n}(\mathbb{C})$ there exists a unitary matrix $U \in M_{n \times n}(\mathbb{C})$ such that

$$A = U|A|$$

12/5

Proof Let $\vec{v}_1, \dots, \vec{v}_n$ be an orthonormal basis of eigenvectors of $|A|$ corresponding to eigenvalues $\sigma_1, \dots, \sigma_n$. Assume $\sigma_1, \dots, \sigma_r$ are the non-zero eigenvalues. Then $\vec{v}_{r+1}, \dots, \vec{v}_n$ is a basis for $\ker(|A|) = \ker(A)$ (by Lemma 2). By the rank-nullity theorem, since A is square

$$\dim(\ker(A^*)) = \dim(\ker(A)) = n - r$$

Let $\vec{w}_{r+1}, \dots, \vec{w}_n$ be an orthonormal basis for $\ker(A^*)$ and for $k = 1, \dots, r$ let

$$\vec{w}_k := \frac{1}{\sigma_k} A\vec{v}_k.$$

Then $\vec{w}_1, \dots, \vec{w}_r$ is an orthonormal system by Lemma 1 and are orthogonal to $\ker(A^*)$ by an example from Section 5.2. Hence $\vec{w}_1, \dots, \vec{w}_r, \vec{w}_{r+1}, \dots, \vec{w}_n$ is an orthonormal basis for $\mathbb{C}^n$. Define $U : \mathbb{C}^n \to \mathbb{C}^n$ by $U(\vec{v}_i) = \vec{w}_i$. Since U takes an orthonormal basis to an orthonormal basis, U is unitary (Exercise: prove this).

It remains to show $U|A| = A$. For any $\vec{x} \in \mathbb{C}^n$ we have

$$\vec{x} = \sum_{i=1}^{n} \alpha_i \vec{v}_i$$

for some scalars α_i. Recall the $\vec{v}_i$'s are eigenvectors of $|A|$. Thus

$$U|A|\vec{x} = U\Big(\sum_{i=1}^{n} \alpha_i |A|\vec{v}_i\Big) = U\Big(\sum_{i=1}^{n} \alpha_i \sigma_i \vec{v}_i\Big) = U\Big(\sum_{i=1}^{r} \alpha_i \sigma_i v_i\Big)$$

$$= \sum_{i=1}^{r} \alpha_i \sigma_i U v_i = \sum_{i=1}^{r} \alpha_i \sigma_i w_i = \sum_{i=1}^{r} \alpha_i A v_i = A\Big(\sum_{i=1}^{r} \alpha_i v_i\Big) = A\vec{x}$$

where we have used $\vec{v}_{r+1}, \dots, \vec{v}_n \in \ker(A)$ in the last equality. □

$\boxed{\text{Rem}}$ Compare the above theorem to the fact that for any $z \in \mathbb{C}$, $z = e^{i\theta}|z|$ for some $\theta \in [0, 2\pi)$. Note that $\overline{e^{i\theta}} = e^{-i\theta} = (e^{i\theta})^{-1}$.