Projections and Fixed Points

15

§1. Contraction mappings

We have been making use of the Implicit Function Theorem for a while now. To get ready to prove it, we need a result called the "Contraction mapping fixed point theorem".

Definition: Let $\mathcal{S} \subset \mathbb{R}^n$ be a subset, and $\vec{f}: \mathcal{S} \to \mathcal{S}$ be a map. It is called a contraction mapping if, for some $r \in (0,1)$, for all pairs $\vec{x}, \vec{y} \in \mathcal{S}$

$$\|\vec{f}(\vec{x}) - \vec{f}(\vec{y})\| \leq r \|\vec{x} - \vec{y}\|.$$

Notice that contraction mappings are (uniformly) continuous: given $\epsilon > 0$, taking $0 < \|\vec{x} - \vec{y}\| < \epsilon$ ensures that (i.e. taking $\delta := \epsilon$)

$$\|\vec{f}(\vec{x}) - \vec{f}(\vec{y})\| \leq r\|\vec{x} - \vec{y}\| < r\epsilon < \epsilon.$$

Ex 1/ If $f: [a,b] \to [a,b]$ is C^1, with $|f'| \leq r < 1$ on all of $[a,b]$, then for $a \leq x < y \leq b$ we have (by MVT)

$$f(x) - f(y) = f'(c)(x-y) \quad \text{for}$$

some $c \in (x,y) \subset [a,b]$. Hence

$$|f(x) - f(y)| = |f'(c)||x-y| \leq r|x-y|,$$

and f is a contraction mapping. //

The point here is that the MVT lets us use the derivative to bound the overall variation in the function f.

Definition: Let $S \subset \mathbb{R}^n$ be a subset, and $\vec{f}: S \to S$ be a map. It is called a contraction mapping if, for some $r \in (0,1)$, for all pairs $\vec{x}, \vec{y} \in S$

$$\|\vec{f}(\vec{x}) - \vec{f}(\vec{y})\| \leq r \|\vec{x} - \vec{y}\|.$$

Lemma: Let $\vec{a}_k \in \mathbb{R}^n$ be a sequence. If the series $\sum_k \|\vec{a}_k\|$ converges in $\mathbb{R}$, then the series $\sum_k \vec{a}_k$ converges in $\mathbb{R}^n$.

Proof: If $d_n := \sum_{k=1}^{n} \|\vec{a}_k\|$ converges, then it is Cauchy. So given $\epsilon > 0$, $\exists M$ for which $n > m \geq M \Rightarrow$

$$\epsilon > |d_n - d_m| = \sum_{k=m+1}^{n} \|\vec{a}_k\| \geq \left\| \sum_{k=m+1}^{n} \vec{a}_k \right\|.$$

Writing $\vec{A}_n := \sum_{k=1}^{n} \vec{a}_k$, this says $\epsilon > \|\vec{A}_n - \vec{A}_m\|$.

But then $\vec{A}_n$ is Cauchy, hence convergent. $\square$

Definition: Let $S \subset \mathbb{R}^n$ be a subset, and $\vec{f}: S \to S$ be a map. It is called a contraction mapping if, for some $r \in (0,1)$, for all pairs $\vec{x}, \vec{y} \in S$

$$\| \vec{f}(\vec{x}) - \vec{f}(\vec{y}) \| \leq r \| \vec{x} - \vec{y} \|.$$

Lemma: Let $\vec{a}_k \in \mathbb{R}^n$ be a sequence. If the series $\sum_k \|\vec{a}_k\|$ converges in $\mathbb{R}$, then the series $\sum_k \vec{a}_k$ converges in $\mathbb{R}^n$.

Theorem: Let $S \subset \mathbb{R}^n$ be closed, and $\vec{f}: S \to S$ a contraction mapping. Then $\vec{f}$ has a unique fixed point, i.e. $\vec{a} \in S$ such that $\vec{f}(\vec{a}) = \vec{a}$.

Proof: Pick any $\vec{x}_0 \in S$, and iteratively define $\vec{x}_k := \vec{f}(\vec{x}_{k-1})$. Then

$$(*) \quad \vec{x}_n = \vec{x}_0 + \sum_{k=1}^{n} (\vec{x}_k - \vec{x}_{k-1}) \quad \text{and}$$

$$\|\vec{x}_k - \vec{x}_{k-1}\| = \|\vec{f}(\vec{x}_{k-1}) - \vec{f}(\vec{x}_{k-2})\| \leq r \|\vec{x}_{k-1} - \vec{x}_{k-2}\|$$
$$\leq r^2 \|\vec{x}_{k-2} - \vec{x}_{k-3}\| \leq \cdots \leq r^{k-1} \|\vec{x}_1 - \vec{x}_0\|$$

$$\Downarrow$$

$$\sum_{k=1}^{n} \|\vec{x}_k - \vec{x}_{k-1}\| \leq \|\vec{x}_1 - \vec{x}_0\| \sum_{k=1}^{n} r^{k-1} = \|\vec{x}_1 - \vec{x}_0\| \frac{1-r^n}{1-r}$$

is increasing & bounded above by $\|\vec{x}_1 - \vec{x}_0\| \frac{1}{1-r}$

$\Downarrow$ Lemma

hence convergent by monotone sequence thm.

$(*)$ converges, to some $\vec{a} \in S$.

$\Downarrow$

$\vec{f}$ continuous

$$\vec{a} = \lim_{n\to\infty} \vec{x}_{n+1} = \lim_{n\to\infty} \vec{f}(\vec{x}_n) = \vec{f}\left(\lim_{n\to\infty} \vec{x}_n\right) = \vec{f}(\vec{a}).$$

To see the uniqueness, suppose that $\vec{a} \neq \vec{b}$ and $\vec{f}(\vec{a}) = \vec{a}$, $\vec{f}(\vec{b}) = \vec{b}$.

Then $\|\vec{a} - \vec{b}\| = \|\vec{f}(\vec{a}) - \vec{f}(\vec{b})\| \leq r \|\vec{a} - \vec{b}\|$
$< \|\vec{a} - \vec{b}\|$,

a contradiction. □

Definition: Let $S \subset \mathbb{R}^n$ be a subset, and $\vec{f}: S \to S$ be a map. It is called a contraction mapping if, for some $r \in (0,1)$, for all pairs $\vec{x}, \vec{y} \in S$

$$\|\vec{f}(\vec{x}) - \vec{f}(\vec{y})\| \leq r \|\vec{x} - \vec{y}\|.$$

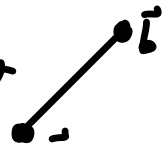
Theorem: Let $S \subset \mathbb{R}^n$ be closed, and $\vec{f}: S \to S$ a contraction mapping. Then $\vec{f}$ has a unique fixed point, i.e. $\vec{a} \in S$ such that $\vec{f}(\vec{a}) = \vec{a}$.

To produce examples of contraction maps, we will use the

Mean Value Inequality: If $U \subset \mathbb{R}^n$ is open, $\vec{f}: U \to \mathbb{R}^m$ is C^1, and $[\vec{a}, \vec{b}] \subset U$, then

$$\|\vec{f}(\vec{b}) - \vec{f}(\vec{a})\| \leq \left(\max_{\vec{x} \in [\vec{a}, \vec{b}]} \|D\vec{f}(\vec{x})\|\right) \|\vec{b} - \vec{a}\|.$$

Proof: $[\vec{a}, \vec{b}]$ denotes the line segment parametrized by $\vec{\varphi}(t) := \vec{a} + t(\vec{b} - \vec{a})$. $(t \in [0,1])$



Let $\vec{g}(t) := \vec{f}(\vec{\varphi}(t))$, so that $\vec{f}(\vec{b}) - \vec{f}(\vec{a}) = \vec{g}(1) - \vec{g}(0)$ and (by chain rule) $\vec{g}'(t) = D\vec{f}(\vec{\varphi}(t))(\vec{b} - \vec{a})$. Then

$$\|\vec{f}(\vec{b}) - \vec{f}(\vec{a})\| = \left\| \int_0^1 \vec{g}'(t)\,dt \right\| \leq \int_0^1 \|\vec{g}'(t)\|\,dt$$

$$\leq \max_{t \in [0,1]} \|\vec{g}'(t)\| = \max_{\vec{x} \in [\vec{a}, \vec{b}]} \|D\vec{f}(\vec{x})(\vec{b} - \vec{a})\|$$

$$\leq \left(\max_{\vec{x} \in [\vec{a}, \vec{b}]} \|D\vec{f}(\vec{x})\|\right) \|\vec{b} - \vec{a}\|. \qquad \square$$

Corollary: Let $S \subset \mathbb{R}^n$ is convex, and $U = \text{int}(S)$ has $\overline{U} = S$. Suppose $\vec{f}: S \to S$ is C^1 on U, with $r := \sup_{\vec{x} \in U} \|D\vec{f}(\vec{x})\| < 1$. Then $\vec{f}$ is a contraction mapping.

Ex 2 / Define $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ by $f(x) := \sqrt{x+7}$. Then

$$f'(x) = \frac{1}{2\sqrt{x+7}} \leq \frac{1}{2\sqrt{7}} < 1 \implies$$

f is a contraction mapping. You computed its fixed point on the midterm. //

Exercise / Ex 3: Show that

$$\vec{f}\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \sqrt{3 + \frac{y^2}{3}} \\ \sqrt{5 + \frac{x^2}{3}} \end{pmatrix}$$ defines a contraction mapping from $\mathbb{R}^2 \to \mathbb{R}^2$, & find its unique fixed point.

Corollary: Let $S \subset \mathbb{R}^n$ is convex, and $U = \text{int}(S)$ has $\overline{U} = S$. Suppose $\vec{f} : S \to S$ is C^1 on U, with $r := \sup_{\vec{x} \in U} \|D\vec{f}(\vec{x})\| < 1$. Then $\vec{f}$ is a contraction mapping.

Use what you know about matrix norms:

$$D\vec{f} = \begin{pmatrix} 0 & \frac{y/3}{\sqrt{3 + y^2/3}} \\ \frac{x/3}{\sqrt{5 + x^2/3}} & 0 \end{pmatrix}$$

$$\|D\vec{f}\| \leq \sqrt{\underbrace{\frac{x^2/9}{5 + x^2/3} + \frac{y^2/9}{3 + y^2/3}}_{\text{sum of squares of matrix entries}}}$$

$$= \sqrt{\frac{1}{3}\left(\frac{x^2}{15 + x^2}\right) + \frac{1}{3}\left(\frac{y^2}{9 + y^2}\right)}$$

$$\leq \frac{\sqrt{2}}{\sqrt{3}} < 1$$

r ← this is crucial

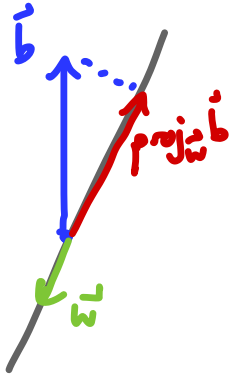
$\implies$ contraction.

Fixed point $\begin{pmatrix} a \\ b \end{pmatrix}$ must solve $a^2 = 3 + \frac{b^2}{3}$ and $b^2 = 5 + \frac{a^2}{3}$.

Turns out it is $\begin{pmatrix} \sqrt{21}/2 \\ 3\sqrt{3}/2 \end{pmatrix}$. □

§2. More on projections

Rather than starting from the end of the last Lecture here, let's start from the formula for the projection of $\vec{b} \in \mathbb{R}^n$ onto the line $L = \text{span}\{\vec{w}\}$



The formula was

$$\text{proj}_{\vec{w}}(\vec{b}) = \frac{\vec{b}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}$$

and works simply because

$\vec{b} - \frac{\vec{b}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}$ is $\perp$ to L:

$$\left(\vec{b} - \frac{\vec{b}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}\right)\cdot\vec{w} = \vec{b}\cdot\vec{w} - \frac{\vec{b}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\vec{w}\cdot\vec{w}$$

$$= 0.$$

Now suppose $W = \text{span}\{\vec{w}_1, \vec{w}_2\}$ and we know that $\vec{w}_1 \perp \vec{w}_2$.

I claim that the orthogonal projection of $\vec{b}$ to W is

$$\text{proj}_W(\vec{b}) = \frac{\vec{b}\cdot\vec{w}_1}{\|\vec{w}_1\|^2}\vec{w}_1 + \frac{\vec{b}\cdot\vec{w}_2}{\|\vec{w}_2\|^2}\vec{w}_2.$$

Check that $\vec{b} - \text{proj}_W(\vec{b})$ is $\perp$ to W, that is, to both $\vec{w}_1$ & $\vec{w}_2$:

$$\left(\vec{b} - \frac{\vec{b}\cdot\vec{w}_1}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_2}{\|\vec{w}_2\|^2}\vec{w}_2\right)\cdot\vec{w}_1$$

$$= \vec{b}\cdot\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1\cdot\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\underbrace{\vec{w}_2\cdot\vec{w}_1}_{0}$$

$$= \vec{b}\cdot\vec{w}_1 - \vec{b}\cdot\vec{w}_1 = 0.$$

Writing $A = \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix}$, so that $W = C(A)$,

that formula said that

$$\text{proj}_W(\vec{b}) = A({}^tAA)^{-1}{}^tA\,\vec{b}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \left\{ \underbrace{\begin{pmatrix} \leftarrow \vec{w}_1 \rightarrow \\ \leftarrow \vec{w}_2 \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix}}_{\begin{pmatrix} \|\vec{w}_1\|^2 & 0 \\ 0 & \|\vec{w}_2\|^2 \end{pmatrix}} \right\}^{-1} \begin{pmatrix} \leftarrow \vec{w}_1 \rightarrow \\ \leftarrow \vec{w}_2 \rightarrow \end{pmatrix} \vec{b}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} \|\vec{w}_1\|^{-2} & 0 \\ 0 & \|\vec{w}_2\|^{-2} \end{pmatrix} \begin{pmatrix} \vec{w}_1 \cdot \vec{b} \\ \vec{w}_2 \cdot \vec{b} \end{pmatrix}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} \frac{\vec{w}_1 \cdot \vec{b}}{\|\vec{w}_1\|^2} \\ \frac{\vec{w}_2 \cdot \vec{b}}{\|\vec{w}_2\|^2} \end{pmatrix} = \frac{\vec{w}_1 \cdot \vec{b}}{\|\vec{w}_1\|^2}\vec{w}_1 + \frac{\vec{w}_2 \cdot \vec{b}}{\|\vec{w}_2\|^2}\vec{w}_2.$$

I claim that the orthogonal projection of $\vec{b}$ to W is

$$\text{proj}_W(\vec{b}) = \frac{\vec{b}\cdot\vec{w}_1}{\|\vec{w}_1\|^2}\vec{w}_1 + \frac{\vec{b}\cdot\vec{w}_2}{\|\vec{w}_2\|^2}\vec{w}_2 .$$

Check that $\vec{b} - \text{proj}_W(\vec{b})$ is $\perp$ to W, that is, to both $\vec{w}_1$ & $\vec{w}_2$:

$$\left(\vec{b} - \frac{\vec{b}\cdot\vec{w}_1}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_2}{\|\vec{w}_2\|^2}\vec{w}_2\right)\cdot\vec{w}_1$$

$$= \vec{b}\cdot\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1\cdot\vec{w}_1 - \frac{\vec{b}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\underbrace{\vec{w}_2\cdot\vec{w}_1}_{0}$$

$$= \vec{b}\cdot\vec{w}_1 - \vec{b}\cdot\vec{w}_1 = 0 .$$

Of course $(\ldots)\cdot\vec{w}_2 = 0$ similarly, so it works... but you probably wonder the relation to the formula from last Lecture now — which was based on finding the closest vector to $\vec{b}$ in $W = \text{span}\{\vec{w}_1, \vec{w}_2\}$.

Writing $A = \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix}$, so that $W = C(A)$,

that formula said that

$$\text{proj}_W(\vec{b}) = A({}^tAA)^{-1}{}^tA\,\vec{b}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \left\{ \begin{pmatrix} \leftarrow \vec{w}_1 \rightarrow \\ \leftarrow \vec{w}_2 \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \right\}^{-1} \begin{pmatrix} \leftarrow \vec{w}_1 \rightarrow \\ \leftarrow \vec{w}_2 \rightarrow \end{pmatrix} \vec{b}$$

$$\begin{pmatrix} \|\vec{w}_1\|^2 & 0 \\ 0 & \|\vec{w}_2\|^2 \end{pmatrix}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} \|\vec{w}_1\|^{-2} & 0 \\ 0 & \|\vec{w}_2\|^{-2} \end{pmatrix} \begin{pmatrix} \vec{w}_1 \cdot \vec{b} \\ \vec{w}_2 \cdot \vec{b} \end{pmatrix}$$

$$= \begin{pmatrix} \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 \\ \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} \frac{\vec{w}_1 \cdot \vec{b}}{\|\vec{w}_1\|^2} \\ \frac{\vec{w}_2 \cdot \vec{b}}{\|\vec{w}_2\|^2} \end{pmatrix} = \boxed{\frac{\vec{w}_1 \cdot \vec{b}}{\|\vec{w}_1\|^2}\vec{w}_1 + \frac{\vec{w}_2 \cdot \vec{b}}{\|\vec{w}_2\|^2}\vec{w}_2} \quad (*)$$

So it's really the same formula.

But it had to be: as soon as you know that $\vec{b} - \text{proj}_W \vec{b}$ is $\perp$ to W, for any $\vec{w} \in W$

$$\vec{b} - \vec{w} = \underbrace{(\vec{b} - \text{proj}_W \vec{b})}_{\text{in } W^\perp} + \underbrace{(\text{proj}_W \vec{b} - \vec{w})}_{\text{in } W}$$

$\Rightarrow \perp \Leftarrow$

$\Leftarrow$ Pythagorean Thm.

$$\|\vec{b} - \vec{w}\|^2 = \|\vec{b} - \text{proj}_W \vec{b}\|^2 + \|\text{proj}_W \vec{b} - \vec{w}\|^2$$

$$\geq \|\vec{b} - \text{proj}_W \vec{b}\|^2, \text{ with equality} \Leftrightarrow \text{this} = 0.$$

Exercise: Find the closest point to $\vec{b} = \begin{pmatrix} 3 \\ 1 \\ 5 \\ 1 \end{pmatrix}$ in $W = \text{span}\left\{ \begin{pmatrix} 3 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \right\}$.

$\mathbb{R}^4$

$\vec{w}_1$ $\vec{w}_2$

It's essential to check that $\vec{w}_1 \cdot \vec{w}_2 = 0$ (otherwise you can't use $(*)$). Then the closest point is simply $\text{proj}_W \vec{b} =$

$$\frac{\vec{b} \cdot \vec{w}_1}{\vec{w}_1 \cdot \vec{w}_1}\vec{w}_1 + \frac{\vec{b} \cdot \vec{w}_2}{\vec{w}_2 \cdot \vec{w}_2}\vec{w}_2 = \frac{6}{12}\vec{w}_1 + \frac{6}{4}\vec{w}_2 = \begin{pmatrix} 3 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

Now suppose $\{\vec{u}_1, \vec{u}_2\}$ was an orthonormal basis for W, i.e. $\vec{u}_i \cdot \vec{u}_j = \delta_{ij} := \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$.

Then the formula becomes

$$\text{proj}_W \vec{b} = (\vec{b} \cdot \vec{u}_1)\vec{u}_1 + (\vec{b} \cdot \vec{u}_2)\vec{u}_2,$$

which looks much simpler. If

$$U = \begin{pmatrix} \uparrow & \uparrow \\ \vec{u}_1 & \vec{u}_2 \\ \downarrow & \downarrow \end{pmatrix},$$ we can also write it as $U \begin{pmatrix} \vec{b} \cdot \vec{u}_1 \\ \vec{b} \cdot \vec{u}_2 \end{pmatrix} = U\,{}^tU\,\vec{b}$, so that

$U\,{}^tU$ is the matrix of proj_W.

Of course, this is just the $A({}^tAA)^{-1}{}^tA$ formula from last Lecture with $A = U$, except that now ${}^tAA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ drops out. All of this works with W of dimension k, and k orthogonal basis vectors.

§3. Orthogonal bases

A set $\{\vec{v}_1, \ldots, \vec{v}_k\} \subset \mathbb{R}^n$ is orthogonal if $\vec{v}_i \cdot \vec{v}_j = 0$ for $i \neq j$. You showed in HW that orthogonal sets are linearly independent. So an orthogonal set $\{\vec{v}_1, \ldots, \vec{v}_n\} \subset \mathbb{R}^n$ must be a basis, and we must be able to write any $\vec{x} \in \mathbb{R}^n$ as a linear combination of them:

$$\vec{x} = \sum_{i=1}^n c_i \vec{v}_i \implies \vec{x} \cdot \vec{v}_j = \sum_{i=1}^n c_i \underbrace{\vec{v}_i \cdot \vec{v}_j}_{0 \text{ unless } i=j}$$

$$= c_j \vec{v}_j \cdot \vec{v}_j$$

$$\implies \vec{x} = \sum_{i=1}^n \frac{\vec{x} \cdot \vec{v}_i}{\vec{v}_i \cdot \vec{v}_i} \vec{v}_i.$$

This is the same formula as for projection to W, but where $W = \mathbb{R}^n$ itself!

If $\{\vec{u}_i\}$ is an orthonormal basis of $\mathbb{R}^n$, this becomes

$U = \left(\vec{u}_1 \cdots \vec{u}_n\right)$ square

$$\vec{x} = \sum_{i=1}^{n} (\vec{x}\cdot\vec{u}_i)\vec{u}_i = U\,{}^tU\,\vec{x} \ldots.$$

which is to say, $U\,{}^tU = I_n$ is the identity matrix.

Definition: We say an $n \times n$ matrix is orthogonal if any of the following properties (which are all equivalent) hold:

- $U\,{}^tU = I_n$
- ${}^tU\,U = I_n$
- columns form an o.n. basis of $\mathbb{R}^n$
- rows form an o.n. basis of $\mathbb{R}^n$

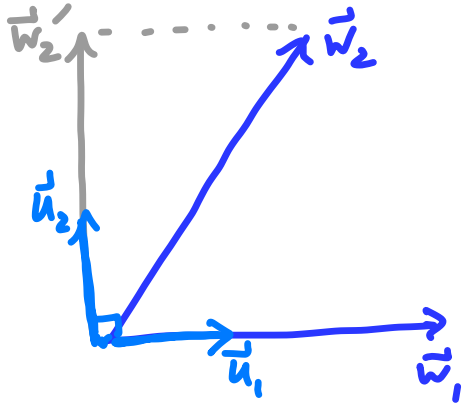
These matrices preserve the dot product:

$$U\vec{x}\cdot U\vec{y} = {}^t\vec{x}\,{}^tU U\,\vec{y} = {}^t\vec{x}\,\vec{y} = \vec{x}\cdot\vec{y}$$

hence in particular preserve lengths.

(They have determinant ± 1 too. Why?)

Going back to a subspace $W \subset \mathbb{R}^n$, it's clearly more useful (for projections) to have an orthogonal or o.n. basis of W. How do we get one?



Given $\vec{w}_1, \ldots, \vec{w}_k$ a basis of W, begin by normalizing $\vec{w}_1 \longrightarrow \vec{u}_1 := \dfrac{\vec{w}_1}{\|\vec{w}_1\|}$.

Next put $\vec{w}_2' := \vec{w}_2 - \text{proj}_{\vec{u}_1}\vec{w}_2$

$= \vec{w}_2 - (\vec{w}_2\cdot\vec{u}_1)\vec{u}_1$,

so that $\vec{w}_2'\cdot\vec{u}_1 = 0$,

and normalize to $\vec{u}_2 = \dfrac{\vec{w}_2'}{\|\vec{w}_2'\|}$.

Continuing, we write

$\vec{w}_3' := \vec{w}_3 - (\vec{w}_3\cdot\vec{u}_1)\vec{u}_1 - (\vec{w}_3\cdot\vec{u}_2)\vec{u}_2$

which is $\perp$ to $\vec{u}_1$ & $\vec{u}_2$, and

$\vec{u}_3 := \vec{w}_3'/\|\vec{w}_3'\|$.

... and ultimately get an o.n. basis of W. We can "invert" the equations defining the $\{\vec{u}_i\}$ like so:

$$\vec{w}_1 = \|\vec{w}_1\|\,\vec{u}_1$$

$$\vec{w}_2 = (\vec{w}_2\cdot\vec{u}_1)\vec{u}_1 + \|\vec{w}_2'\|\,\vec{u}_2$$

$$\vec{w}_3 = (\vec{w}_3\cdot\vec{u}_1)\,\vec{u}_1 + (\vec{w}_3\cdot\vec{u}_2)\vec{u}_2 + \|\vec{w}_3'\|\,\vec{u}_3$$

etc.,

which looks especially nice in matrix form:

$$\underbrace{\begin{pmatrix}\uparrow & & \uparrow\\ \vec{w}_1 & \cdots & \vec{w}_k\\ \downarrow & & \downarrow\end{pmatrix}}_{A} = \underbrace{\begin{pmatrix}\uparrow & & \uparrow\\ \vec{u}_1 & \cdots & \vec{u}_k\\ \downarrow & & \downarrow\end{pmatrix}}_{Q} \underbrace{\begin{pmatrix}\|\vec{w}_1\| & \vec{w}_2\cdot u_1 & \vec{w}_3\cdot u_1 & \cdots & \vec{w}_k\cdot u_1\\ & \|\vec{w}_2'\| & \vec{w}_3\cdot\vec{u}_2 & \cdots & \vec{w}_k\cdot\vec{u}_2\\ & & \|\vec{w}_3'\| & & \vdots\\ & 0 & & \ddots & \\ & & & & \|\vec{w}_k'\|\end{pmatrix}}_{R}$$

A: columns = original basis

Q: columns = o.n. basis

R: upper triangular with positive diagonal entries

If $k=n$, i.e. $W=\mathbb{R}^n$, then this is called the "Iwasawa decomposition" — it takes any invertible $n\times n$ matrix and writes it (uniquely) as the product of an orthogonal matrix and an upper triangular one.

Your book breaks the "Gram-Schmidt" algorithm just described into two pieces: first, it produces a merely orthogonal basis (avoiding square roots, & enough for some purposes) by

$$\vec{v}_1 = \vec{w}_1$$

$$\vec{v}_2 = \vec{w}_2 - \frac{\vec{w}_2\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1}\vec{v}_1$$

$$\vec{v}_3 = \vec{w}_3 - \frac{\vec{w}_3\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1}\vec{v}_1 - \frac{\vec{w}_3\cdot\vec{v}_2}{\vec{v}_2\cdot\vec{v}_2}\vec{v}_2$$

$$\vdots$$

which (if desired) can then be normalized (to an o.n. basis) by

$$\vec{u}_j = \frac{\vec{v}_j}{\|\vec{v}_j\|} \quad \text{for each } j.$$

You'll get a chance to apply Gram-Schmidt in the HW.