Applications

- Least-squares (we will see this in detail next lecture)
- Discrete cosine and Fourier Transforms (you will see this in ESE 2240)

Topics

- Orthogonal & Orthonormal Bases (ALA 4.1)
- The Gram-Schmidt Process (ALA 4.2)
- Orthogonal Matrices (ALA 4.3)
    - QR-factorization

Orthogonality

Orthogonality is a generalization/abstraction of perpendicularity (right angles) to general inner product spaces. Algorithms using orthogonality are at the core of modern linear algebra, & include the Gram-Schmidt algorithm, the QR decomposition, and the least-squares algorithm, all of which we shall see in this lecture.

More abstract applications of orthogonality, that you will see for example, in ESE 2240, include the Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT), algorithms that lie at the heart of modern digital media (e.g., JPG image compression and MP3 audio compression).

Orthogonal and Orthonormal Bases

Let V be an inner product space (as usual we assign that the scalars over which V is defined are real valued). Remember that $\underline{v}, \underline{w} \in V$ are orthogonal if $\langle \underline{v}, \underline{w} \rangle = 0$. If $\underline{v}, \underline{w} \in \mathbb{R}^n$ and $\langle \underline{v}, \underline{w} \rangle = \underline{v} \cdot \underline{w}$ is the dot product, this simply means that $\underline{v}$ and $\underline{w}$ are perpendicular (meet at a right angle).

Orthogonal vectors are useful, because they point in completely different directions making them particularly well-suited for defining bases.

A basis $\underline{b}_1, \ldots, \underline{b}_n$ of an n-dimensional inner product space V is called orthogonal if $\langle \underline{b}_i, \underline{b}_j \rangle = 0$ for all $i \neq j$. In this case, the collection $\underline{b}_i$ are said to be mutually orthogonal. Furthermore, if each $\underline{b}_i$ has unit length, i.e., if $\|\underline{b}_i\| = 1$ for all $i = 1, \ldots, n$, then the basis is called orthonormal.

A simple way to construct an orthonormal basis from an orthogonal one is to normalize each of its elements, that is, to replace each basis element $\underline{b}_i$ with its normalized counterpart $\frac{\underline{b}_i}{\|\underline{b}_i\|}$. Can you formally verify that $\frac{\underline{b}_1}{\|\underline{b}_1\|}, \ldots, \frac{\underline{b}_n}{\|\underline{b}_n\|}$ is an orthonormal basis if $\underline{b}_1, \ldots, \underline{b}_n$ is an orthogonal one? Can you explain why rescaling each entry does not affect the mutual orthogonality of the set?

Example: A familiar example of an orthonormal basis for $\mathbb{R}^n$ equipped with the standard inner product is the collection of standard basis elements:

$$\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \underline{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

A very useful property of a collection of mutually orthogonal vectors is that they are automatically linearly independent. In particular, if $\underline{v}_1, \dots, \underline{v}_k$ satisfy $\langle \underline{v}_i, \underline{v}_j \rangle = 0$ for all $i \neq j$ (and $\underline{v}_i \neq 0$ for all i), then they are linearly independent.

To see this, we take an arbitrary linear combination of the $\underline{v}_i$ and set it to 0:

$$c_1 \underline{v}_1 + c_2 \underline{v}_2 + \cdots + c_k \underline{v}_k = \underline{0} \qquad (*)$$

Let's take the inner product of both sides of this equation with any $\underline{v}_i$:

$$0 = \langle \underline{0}, \underline{v}_i \rangle = \langle c_1 \underline{v}_1 + c_2 \underline{v}_2 + \cdots + c_k \underline{v}_k, \underline{v}_i \rangle$$

(linearity of $\langle \cdot, \underline{v}_i \rangle$) $\quad = c_1 \langle \underline{v}_1, \underline{v}_i \rangle + \cdots + c_i \langle \underline{v}_i, \underline{v}_i \rangle + \cdots + c_k \langle \underline{v}_k, \underline{v}_i \rangle$

(orthogonality) $\quad = c_i \langle \underline{v}_i, \underline{v}_i \rangle = c_i \|\underline{v}_i\|^2$

Since $\underline{v}_i \neq 0$, $\|\underline{v}_i\|^2 > 0$, which means $c_i = 0$. We can repeat this same with all $\underline{v}_i$, $i = 1, \dots, k$, to conclude $(*)$ holds if and only if $c_1 = c_2 = \cdots = c_k = 0$. Hence, the mutually orthogonal collection $\underline{v}_1, \dots, \underline{v}_k$ is linearly independent.

Example: The vectors

$$\underline{b}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \underline{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \underline{b}_3 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$$

are a basis for $\mathbb{R}^3$. One easy way to check this is to confirm that $\underline{b}_i \cdot \underline{b}_j = 0$ for all $i \neq j$ (this is indeed true). Since $\dim \mathbb{R}^3 = 3$, and $\underline{b}_1, \underline{b}_2, \underline{b}_3$ are linearly independent, they must be a basis.

To turn them from an orthogonal basis into an orthonormal basis, we simply divide every vector by its length to obtain:

$$\underline{v}_1 = \frac{\underline{b}_1}{\|\underline{b}_1\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \underline{v}_2 = \frac{\underline{b}_2}{\|\underline{b}_2\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \underline{v}_3 = \frac{\underline{b}_3}{\|\underline{b}_3\|} = \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$$

This example highlights a more general principle, which is again quite useful: if $\underline{v}_1, \dots, \underline{v}_n$ are mutually orthogonal, then they form a basis for their span $W = \operatorname{span}\{\underline{v}_1, \dots, \underline{v}_n\} \subseteq V$, which is thus a subspace of $\dim W = n$.

It then follows that if $\dim V = n$, then $\underline{v}_1, \dots, \underline{v}_n$ are an orthogonal basis for V (this is precisely the observation we used in the example above.

Working in Orthogonal Bases

So why do we care about orthogonal (or even better, orthonormal) bases? Turns out they make a lot of the computations that we've been doing so far MUCH easier.

We'll start with some important properties of computing a vector's coordinates with respect to an orthogonal basis:

Theorem: Let $\underline{u}_1, \ldots, \underline{u}_n$ be an orthonormal basis for an inner product space V. Then we can write any $\underline{v} \in V$ as a linear combination

$$\underline{v} = c_1 \underline{u}_1 + \cdots + c_n \underline{u}_n$$

in which its coordinates are given by

$$c_i = \langle \underline{v}, \underline{u}_i \rangle, \quad i = 1, \ldots, n.$$

Moreover, its norm is given by the Pythagorean formula

$$\|\underline{v}\|^2 = c_1^2 + \cdots + c_n^2 = \sum_{i=1}^{n} \langle \underline{v}, \underline{u}_i \rangle^2.$$

Proof: The trick here is to exploit that $\langle \underline{u}_i, \underline{u}_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$.

Let's compute

$$\begin{aligned}
\langle \underline{v}, \underline{u}_i \rangle &= \langle c_1 \underline{u}_1 + \cdots + c_n \underline{u}_n, \underline{u}_i \rangle \\
\text{(linearity of } \langle \cdot, \underline{u}_i \rangle) \quad &= c_1 \langle \underline{u}_1, \underline{u}_i \rangle + \cdots + c_i \langle \underline{u}_i, \underline{u}_i \rangle + \cdots + c_n \langle \underline{u}_n, \underline{u}_i \rangle \\
\text{(orthogonality)} \quad &= c_i \|\underline{u}_i\|^2 \\
(\|\underline{u}_i\| = 1) \quad &= c_i.
\end{aligned}$$

So we have $c_i = \langle \underline{v}, \underline{u}_i \rangle$. Now to compute the norm, we again use a similar trick:

$$\begin{aligned}
\|\underline{v}\|^2 = \langle \underline{v}, \underline{v} \rangle &= \left\langle \sum_{i=1}^{n} c_i \underline{u}_i, \sum_{j=1}^{n} c_j \underline{u}_j \right\rangle \\
&= \sum_{i=1}^{n} c_i \left\langle \underline{u}_i, \sum_{j=1}^{n} c_j \underline{u}_j \right\rangle && \left(\text{linearity of } \left\langle \cdot, \sum_{j=1}^{n} c_j \underline{u}_j \right\rangle\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \langle \underline{u}_i, \underline{u}_j \rangle && (\text{linearity of } \langle \underline{u}_i, \cdot \rangle) \\
&= \sum_{i=1}^{n} c_i^2 \|\underline{u}_i\|^2 && (\text{orthogonality}) \\
&= \sum_{i=1}^{n} c_i^2 && (\|\underline{u}_i\| = 1)
\end{aligned}$$

Example: Let's rewrite $\underline{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ in terms of the orthonormal basis

$$\underline{u}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \underline{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \underline{u}_3 = \frac{1}{\sqrt{30}}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}.$$

We saw earlier, All we need to do is compute dot products!

$$\underline{v}\cdot\underline{u}_1 = \frac{2}{\sqrt{6}}, \quad \underline{v}\cdot\underline{u}_2 = \frac{3}{\sqrt{5}}, \quad \underline{v}\cdot\underline{u}_3 = \frac{4}{\sqrt{30}}$$

to then write: $\underline{v} = \frac{2}{\sqrt{6}}\underline{u}_1 + \frac{3}{\sqrt{5}}\underline{u}_2 + \frac{4}{\sqrt{30}}\underline{u}_3$.

This is much simpler than solving the system of linear equations

$$\begin{bmatrix} | & | & | \\ \underline{u}_1 & \underline{u}_2 & \underline{u}_3 \\ | & | & | \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

for the coordinates c_1, c_2, c_3.

A small change to the above allows us to extend these ideas to orthogonal, but not orthonormal, bases:

Theorem: If $\underline{v}_1, \dots, \underline{v}_n$ are an orthogonal basis, then $\underline{v} \in V$ can be written

$$\underline{v} = a_1\underline{v}_1 + \cdots + a_n\underline{v}_n \quad \text{with } a_i = \frac{\langle \underline{v}, \underline{v}_i \rangle}{\|\underline{v}_i\|^2}$$

and its norm is given by

$$\|\underline{v}\|^2 = a_1^2\|\underline{v}_1\|^2 + \cdots + a_n^2\|\underline{v}_n\|^2.$$

This is derived using our previous theorem by rescaling the $\underline{v}_i$ to $\frac{\underline{v}_i}{\|\underline{v}_i\|}$.

Example. Even though our focus in this class will be mostly on $\mathbb{R}^n$ (or vector spaces that "behave like" $\mathbb{R}^n$), all of these ideas apply to general inner product spaces, including function spaces.

As a simple example, let's consider the space of quadratic polynomials of degree ≤ 2 $P^{(2)}$ over $[0,1]$ equipped with the integral inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

The standard monomials do NOT form an orthogonal basis:

$$\langle 1, x\rangle = \tfrac{1}{2}, \quad \langle 1, x^2\rangle = \tfrac{1}{3}, \langle x, x^2\rangle = \tfrac{1}{4}.$$

One orthogonal basis for $P^{(2)}$ is:

$$p_1(x) = 1, \quad p_2(x) = x - \tfrac{1}{2}, \quad p_3(x) = x^2 - x + \tfrac{1}{6}.$$

For example, $\langle p_1, p_2\rangle = \int_0^1 1 \cdot (x - \tfrac{1}{2})\,dx = \int_0^1 x\,dx - \tfrac{1}{2}\int_0^1 dx$

$$= \frac{x^2}{2}\Big|_0^1 - \frac{1}{2}x\Big|_0^1$$

$$= \frac{1}{2} - 0 - \frac{1}{2} + 0 = 0.$$

With a little bit more calculus, you can check that $\langle p_1, p_3\rangle = \langle p_2, p_3\rangle = 0.$ and that

$$\|p_1\| = 1, \quad \|p_2\| = \frac{1}{2\sqrt{3}}, \quad \|p_3\| = \frac{1}{6\sqrt{5}}.$$

If we now want to compute the coordinates c_1, c_2, c_3 of a quadratic polynomial

$$p(x) = c_1 p_1(x) + c_2 p_2(x) + c_3 p(x)$$

we simply compute some inner products:

$$c_1 = \frac{\langle p, p_1\rangle}{\|p_1\|^2}, \quad c_2 = \frac{\langle p, p_2\rangle}{\|p_2\|^2}, \quad c_3 = \frac{\langle p, p_3\rangle}{\|p_3\|^2}.$$

So, for example, if $p(x) = x^2 + x + 1$, then

$$c_1 = \frac{\int_0^1 (x^2+x+1)\cdot 1\,dx}{1} = \frac{11}{6}, \quad c_2 = \frac{\int_0^1 (x^2+x+1)(x-\tfrac{1}{2})\,dx}{\left(\tfrac{1}{12}\right)} = 2$$

$$c_3 = \frac{\int_0^1 (x^2+x+1)(x^2-x+\tfrac{1}{6})\,dx}{\left(\tfrac{1}{180}\right)} = 1$$

So that $p(x) = x^2 + x + 1 = \frac{11}{6} + 2(x - \tfrac{1}{2}) + (x^2 - x + \tfrac{1}{6})$.

While this may look very abstract, this is exactly the same mechanism underpinning things like the Discrete Fourier Transform, which is a change of basis of a signal to a (complex) orthonormal basis in function space, where each basis element is a complex sinusoid.

Gram-Schmidt Process

Hopefully we've convinced you that orthogonal bases are useful, so now the natural question becomes: how do I compute one? That's where the famed Gram-Schmidt Process (GSP) comes into play.

The idea behind GSP is fairly straightforward: given an initial basis for a vector space, iteratively modify it until it is orthogonal. Let's start with a simple concrete example, and then introduce the general algorithm:

Example: Let $W = \text{Span}\{\underline{x}_1, \underline{x}_2\}$, where $\underline{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\underline{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

Since $\underline{x}_1$ and $\underline{x}_2$ are linearly independent (why?), they form a basis for the subspace $W \subset \mathbb{R}^3$, and $\dim W = 2$.

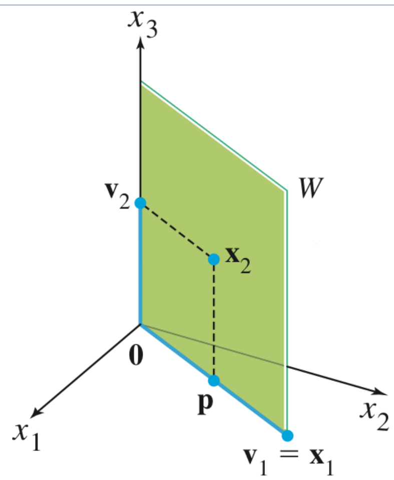
However, $\underline{x}_1$ and $\underline{x}_2$ are not orthogonal because

$$\langle \underline{x}_1, \underline{x}_2 \rangle = 3 \cdot 1 + 6 \cdot 2 + 0 \cdot 2 = 15.$$

Let's use $\{\underline{v}_1, \underline{v}_2\}$ for our new basis, and set $\underline{v}_1 = \underline{x}_1$. We need to find a vector $\underline{v}_2$ that is orthogonal to $\underline{v}_1$ such that $\text{Span}\{\underline{v}_1, \underline{v}_2\} = W$.

Our idea here is to "extract out" the components of $\underline{x}_2$ that are parallel to $\underline{v}_1$ and subtract them off of $\underline{x}_2$, so that what's left is orthogonal.

Let's look at a picture first



From this picture, we observe that we can write $\underline{x}_2 = \underline{p} + \underline{v}_2$ where $\underline{p}$ is the component of $\underline{x}_2$ parallel with $\underline{x}_1$ and $\underline{v}_2$ is what's left over, i.e., the part of $\underline{x}_2$ that is orthogonal to $\underline{x}_1 = \underline{v}_1$.

If $\underline{p}$ is parallel with $\underline{v}_1$, then we must have that $\underline{p} = c\underline{v}_1$ for some constant c, and therefore $\underline{v}_2 = \underline{x}_2 - \underline{p} = \underline{x}_2 - c\underline{v}_1$.

Now from our previous discussion, we know that $c = \frac{\langle \underline{x}_2, \underline{v}_1 \rangle}{\|\underline{v}_1\|^2}$ (why?) but let's see a different way of computing c. We want $\langle \underline{v}_2, \underline{v}_1 \rangle = 0$, so we must have:

$$\langle \underline{v}_2, \underline{v}_1 \rangle = \langle \underline{x}_2 - c\underline{v}_1, \underline{v}_1 \rangle = \langle \underline{x}_2, \underline{v}_1 \rangle - c\|\underline{v}_1\|^2 = 0$$

or equivalently, $c = \frac{\langle \underline{x}_2, \underline{v}_1 \rangle}{\|\underline{v}_1\|^2}$. therefore, $\underline{v}_2 = \underline{x}_2 - \frac{\langle \underline{x}_2, \underline{v}_1 \rangle}{\|\underline{v}_1\|^2} \underline{v}_1$.

By construction, we have that $\langle \underline{v}_1, \underline{v}_2 \rangle = 0$, and since $\underline{v}_1 = \underline{x}_1$ and

$\underline{v}_2 = \underline{x}_2 - c\underline{x}_1$, both $\underline{v}_1, \underline{v}_2 \in W$. So $\underline{v}_1$ and $\underline{v}_2$ are linearly independent and contained in W, so form a basis for W.

The Gram-Schmidt Process simply repeats this process over and over if there are more than two vectors, but the idea remains the same: at each step you subtract off the directions of the current vector that are parallel with previous ones.

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p$$

Farhad: please transcribe with $\langle \underline{x}_i, \underline{v}_j \rangle$ notation in online notes.

Example: A typical use case is to find an orthonormal basis, with respect to the standard dot product, for the subspace $W \subset \mathbb{R}^4$ consisting of all vectors that are orthogonal the vector $\underline{a} = (1, 2, -1, -3)$. The first task is to find a basis for the subspace. A vector $\underline{x} = (x_1, x_2, x_3, x_4)$ is orthogonal to $\underline{a}$ if and only if

$$\underline{x} \cdot \underline{a} = x_1 + 2x_2 - x_3 - 3x_4 = 0.$$

Solving this in the usual way, we observe that the free variables are x_2, x_3, x_4, so that a (non-orthogonal) basis for the subspace is:

$$\underline{w_1} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \underline{w_2} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{w_3} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

because

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Now we apply Gram-Schmidt to obtain an orthogonal basis: first we set $\underline{v_1} = \underline{w_1}$. To get $\underline{v_2}$, we compute:

$$\underline{v_2} = \underline{w_2} - \frac{\langle \underline{w_2}, \underline{v_1} \rangle}{\|\underline{v_1}\|^2} \underline{v_1} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \left(\frac{-2}{5}\right) \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 2/5 \\ 1 \\ 0 \end{bmatrix}$$

Finally, we compute $\underline{v_3}$:

$$\underline{v_3} = \underline{w_3} - \frac{\langle \underline{w_3}, \underline{v_1} \rangle}{\|\underline{v_1}\|^2} \underline{v_1} - \frac{\langle \underline{w_3}, \underline{v_2} \rangle}{\|\underline{v_2}\|^2} \underline{v_2} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \left(\frac{-6}{5}\right) \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{3/5}{6/5} \begin{bmatrix} 1/5 \\ 2/5 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \\ 1 \end{bmatrix}.$$

To get our hands on an orthonormal basis, we simply normalize the $\underline{v_i}$ by dividing them by their norms.

NOTE: The orthogonal basis you obtain from the GSP does depend on the order of the vectors in the original basis — different orderings will produce different bases, but they will all span the same space as the original basis.

FACT: The GSP tells us something very important: given any basis for a finite dimensional inner-prod. space, we can always "orthogonalize" it. That is, every finite dimensional inner-product space has an orthonormal basis!

Orthogonal Matrices

Rotations and reflections play key roles in geometry, physics, robotics, quantum mechanics, airplanes, computer graphics, data science, and more. As we'll explore today and later in the semester, such transformations are encoded via orthogonal matrices, that is matrices whose columns form an orthonormal basis for $\mathbb{R}^n$. They also play a central role in one of the most important methods of linear algebra, the QR factorization.

We start with a definition. A square matrix Q is called orthogonal if it satisfies $QQ^T = Q^TQ = I$.

This means that $Q^{-1} = Q^T$ (in fact, we could define orthogonal matrices this way instead), and that solving linear systems of the form $Q\underline{x} = \underline{b}$ is very easy: simply set $\underline{x} = Q^T\underline{b}$!

Notice that $Q^TQ = I$ implies that the columns of Q are orthonormal. If $Q = [\underline{q}_1, \dots, \underline{q}_n]$, then $(Q^TQ)_{ij} = \underline{q}_i^T\underline{q}_j = I_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j \end{cases}$

which is exactly the definition of an orthonormal collection of vectors. Further, since there are n such vectors, they must form an orthonormal basis for $\mathbb{R}^n$.

Now, let's explore some of the consequences of this definition.

Example: 2x2 orthogonal matrices.

A 2x2 matrix $Q = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is orthogonal if and only if

$$Q^TQ = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or equivalently}$$

$$a^2+c^2 = 1, \quad ab+cd = 0, \quad b^2+d^2 = 1$$

The first and last equations say that $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\begin{bmatrix} b \\ d \end{bmatrix}$ lie on the unit circle in $\mathbb{R}^2$: a convenient and revealing way of writing this is by setting

$$a = \cos\theta, \quad c = \sin\theta, \quad b = \cos\varphi, \quad d = \sin\varphi$$

since $\cos^2\theta + \sin^2\theta = 1$ for all $\theta \in \mathbb{R}$.

Our last condition is $0 = ab+cd = \cos\theta\cos\varphi + \sin\theta\sin\varphi = \cos(\theta-\varphi)$

Now $\cos(\theta - \varphi) = 0$ iff $\theta - \varphi = \frac{1}{2}\pi$ or $\theta - \varphi = -\frac{1}{2}\pi$, i.e.

$$\varphi = \theta \pm \tfrac{1}{2}\pi.$$

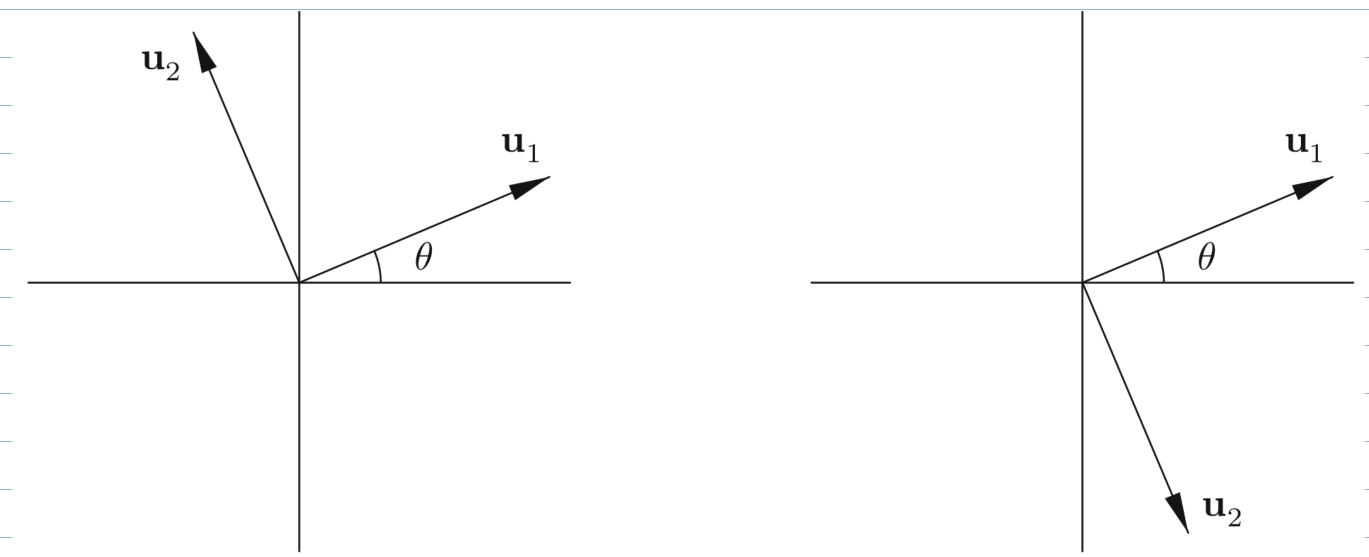
This means either $b = -\sin\theta$ and $d = \cos\theta$
or $b = \sin\theta$ and $d = -\cos\theta$.

As a result, every 2×2 orthogonal matrix has one of two possible forms:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

where by convention, we restrict $\theta \in [0, 2\pi)$.

The columns of both matrices form an orthonormal basis for $\mathbb{R}^2$. The first is obtained by rotating the standard basis $\underline{e}_1, \underline{e}_2$ through angle θ; the second by first reflecting about the x-axis and then rotating.



If we think about the map $\underline{x} \mapsto Q\underline{x}$ defined by multiplication with an orthonormal matrix as rotating and/or reflecting the vector $\underline{x}$, then the following property should not be too suprising:

FACT: the product of two orthogonal matrices is also orthogonal.

Before grinding through some algebra, let's think about this through the lens of rotations & reflections. Multiplying $\underline{x}$ by a product of orthogonal matrices Q_2Q_1 is the same as first rotating/reflecting $\underline{x}$ by Q_1 to obtain $Q_1\underline{x}$, and then rotating/reflecting $Q_1\underline{x}$ by Q_2 to get $Q_2Q_1\underline{x}$. Now a sequence of rotations and reflections is still ultimately a rotation and/or reflection, so we must have $Q_2Q_1\underline{x} = Q\underline{x}$ for some orthogonal $Q = Q_2Q_1$.

Let's check that this intuition carries over in the math. Since Q_1 and Q_2 are orthogonal, we have that

$$Q_1^T Q_1 = I = Q_2^T Q_2.$$

Let's check that $(Q_1 Q_2)^T (Q_1 Q_2) = I$:

$$(Q_1 Q_2)^T (Q_1 Q_2) = Q_2^T \underbrace{Q_1^T Q_1}_{=I} Q_2 = \underbrace{Q_2^T Q_2}_{=I} = I$$

Therefore $(Q_1 Q_2)^{-1} = (Q_1 Q_2)^T$, and we indeed have $Q_1 Q_2$ is orthogonal.

FACT: This multiplicative property combined with the fact that the inverse of an orthogonal matrix is orthogonal (why?) says that the set of all orthogonal matrices forms a group. Group theory underlies much of modern physics and quantum mechanics, and plays a central role in robotics. Although we will not spend too much time on groups in this class, you are sure to see them again in the future. The orthogonal group in particular is central to rigid body mechanics, atomic structure and chemistry, and computer graphics, among many other applications.

The QR Factorization

The GSP, when applied to orthonormalize a basis of $\mathbb{R}^n$, in fact gives us a famous the incredibly useful QR factorization of a matrix.

Let us start with a basis $\underline{b}_1, \ldots, \underline{b}_n$ for $\mathbb{R}^n$, and let $\underline{u}_1, \ldots, \underline{u}_n$ be the result of applying the GSP to it. Define the matrices:

$$A = [\underline{b}_1 \; \underline{b}_2 \cdots \underline{b}_n] \quad \text{and} \quad Q = [\underline{u}_1 \; \underline{u}_2 \cdots \underline{u}_n].$$

Q is an orthogonal matrix because the $\underline{u}_i$ form an orthonormal basis.

Now, let's revisit the GSP equations:

$$\underline{v}_1 = \underline{b}_1$$

$$\underline{v}_2 = \underline{b}_2 - \frac{\underline{b}_2 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1} \underline{v}_1$$

$$\underline{v}_3 = \underline{b}_3 - \frac{\underline{b}_3 \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1} \underline{v}_1 - \frac{\underline{b}_3 \cdot \underline{v}_2}{\underline{v}_2 \cdot \underline{v}_2} \underline{v}_2$$

$$\vdots$$

$$\underline{v}_n = \underline{b}_n - \frac{\underline{b}_n \cdot \underline{v}_1}{\underline{v}_1 \cdot \underline{v}_1} \underline{v}_1 - \cdots - \frac{\underline{b}_n \cdot \underline{v}_{n-1}}{\underline{v}_{n-1} \cdot \underline{v}_{n-1}} \underline{v}_{n-1}$$

We start by replacing each element $\underline{v}_i$ w/ its normalized form, $\underline{u}_i = \frac{\underline{v}_i}{\|\underline{v}_i\|}$.

Rearranging the above, we can write the original basis elements $\underline{b}_i$ in terms of the orthonormal basis $\underline{u}_i$ via the triangular system:

$$\begin{aligned}
\underline{b}_1 &= r_{11}\underline{u}_1 \\
\underline{b}_2 &= r_{12}\underline{u}_1 + r_{22}\underline{u}_2 \\
\underline{b}_3 &= r_{13}\underline{u}_1 + r_{23}\underline{u}_2 + r_{33}\underline{u}_3 \\
&\vdots \\
\underline{b}_n &= r_{1n}\underline{u}_1 + r_{2n}\underline{u}_2 + \cdots + r_{nn}\underline{u}_n.
\end{aligned} \qquad (*).$$

Using our usual trick of taking inner products with both sides we see that

$$\langle \underline{b}_j, \underline{u}_i \rangle = \langle r_{1j}\underline{u}_1 + \cdots + r_{jj}\underline{u}_j, \underline{u}_i \rangle = r_{1j}\langle \underline{u}_1, \underline{u}_i\rangle + \cdots + r_{ij}\langle \underline{u}_i, \underline{u}_i\rangle + \cdots + r_{jj}\langle \underline{u}_i, \underline{u}_j\rangle$$
$$= r_{ij}$$

So we conclude that $r_{ij} = \langle \underline{b}_j, \underline{u}_i \rangle$.

Now, returning to $(*)$, we observe that if we define the upper triangular matrix

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

we can write $A = QR$. Since the GSP works on any basis, the only requirement for A to have a QR factorization is that its columns form a basis for $\mathbb{R}^n$, i.e, that A be non singular.

The online notes will include both pseudocode & a numpy implementation.

QR Factorization of a Matrix A

```
start
    for j = 1 to n
        set r_jj = sqrt(a_1j^2 + ... + a_nj^2)
        if r_jj = 0, stop; print "A has linearly dependent columns"
            else for i = 1 to n
                set a_ij = a_ij / r_jj
            next i
        for k = j+1 to n
            set r_jk = a_1j a_1k + ... + a_nj a_nk
            for i = 1 to n
                set a_ik = a_ik - a_ij r_jk
            next i
        next k
    next j
end
```

Solving linear systems using a QR decomposition is easy. Observe that if our goal is to solve $A\underline{x}=\underline{b}$ and a QR decomposition is available we first notice that

$$QR\underline{x}=\underline{b} \Leftrightarrow R\underline{x}=Q^T\underline{b}=\tilde{\underline{b}}$$

since $Q^TQ=I$. Now, solving $R\underline{x}=\tilde{\underline{b}}$ can be easily accomplised via Back Substitution since R is an upper triangular matrix!