Math 2270-002 Week 12 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material in 6.1-6.4

Chapter 6 is about orthogonality and related topics. We'll spend maybe two weeks plus a day in this chapter. The ideas we develop start with the dot product, which we've been using algebraically to compute individual entries in matrix products, but which has important geometric meaning. By the end of the Chapter we will see applications to statistics, discuss generalizations of the dot product, "inner products", which can apply to function vector spaces and which lie at the heart of physics applications that use Fourier series, and more recent applications such as image and audio compression, see e.g.
https://en.wikipedia.org/wiki/Discrete_cosine_transform

Mon Nov 12
- 6.1-6.2 dot product, length, orthogonality, projection onto the span of a single vector, and angles - in $\mathbb{R}^n$.

Announcements:

Warm-up Exercise:

Recall, for any two vectors $\underline{\boldsymbol{v}}, \underline{\boldsymbol{w}} \in \mathbb{R}^n$, the dot product $\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}$ is the scalar computed by the definition

$$\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}} := \sum_{i=1}^{n} v_i\, w_i\,.$$

We don't care if $\underline{\boldsymbol{v}}, \underline{\boldsymbol{w}}$ are row vectors or column vectors, or one of each, for the dot product.

We've been using the dot product algebraically to compute entries of matrix products $A\,B$, since

$$entry_{i\,j}\,[A\,B] = \left[row_i\,A\right]\left[col_j\,B\right] = \left[row_i\,A\right] \cdot \left[col_j\,B\right].$$

The algebra for dot products is a mostly a special case of what we already know for matrices, but worth writing down and double-checking, so we're ready to use it in the rest of Chapters 6 and 7.

Exercise 1 Check why

1a) dot product is commutative:

$$\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}} = \underline{\boldsymbol{w}} \cdot \underline{\boldsymbol{v}}\,.$$

1b) dot product distributes over addition:

$$(\underline{\boldsymbol{u}} + \underline{\boldsymbol{v}}) \cdot \underline{\boldsymbol{w}} = \underline{\boldsymbol{u}} \cdot \underline{\boldsymbol{w}} + \underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}$$

$$\underline{\boldsymbol{u}} \cdot (\underline{\boldsymbol{v}} + \underline{\boldsymbol{w}}) = \underline{\boldsymbol{u}} \cdot \underline{\boldsymbol{v}} + \underline{\boldsymbol{u}} \cdot \underline{\boldsymbol{w}}$$

1c) for $k \in \mathbb{R}$,

$$(k\,\underline{\boldsymbol{v}}) \cdot \underline{\boldsymbol{w}} = k\,(\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}) = \underline{\boldsymbol{v}} \cdot (k\,\underline{\boldsymbol{w}}).$$

1d) dot product distributes over linear combinations:

$$\underline{\boldsymbol{v}} \cdot \left(c_1\,\underline{\boldsymbol{w}}_1 + c_2\,\underline{\boldsymbol{w}}_2 + \ldots + c_k\,\underline{\boldsymbol{w}}_k\right) = c_1\left(\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}_1\right) + c_2\left(\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}_2\right) + \ldots + c_k\left(\underline{\boldsymbol{v}} \cdot \underline{\boldsymbol{w}}_k\right).$$

$$\underline{v} \cdot \underline{w} := \sum_{i=1}^{n} v_i\, w_i$$

1e)

$$\underline{v} \cdot \underline{v} > 0 \quad \text{for each } \underline{v} \neq \underline{\mathbf{0}} \quad (\text{and } \underline{\mathbf{0}} \cdot \underline{\mathbf{0}} = \underline{\mathbf{0}}.\,)$$

Chapter 6 is about topics related to the geometry of the dot product. It begins now, with definitions and consequences that generalize what you learned for $\mathbb{R}^2$, $\mathbb{R}^3$ in your multivariable Calculus class, to $\mathbb{R}^n$.

2) Geometry of the dot product, stage 1. We'll add examples with pictures as we go throught these definitions.

2a) For $\underline{v} \in \mathbb{R}^n$ we define the *length* or *norm* or *magnitude* of $\underline{v}$ by

$$\|\underline{v}\| := \sqrt{\sum_{i=1}^{n} v_i^2} = (\underline{v} \cdot \underline{v})^{\frac{1}{2}} .$$

Notice that the length of a scalar multiple of a vector is what you'd expect:

$$\|\, t\,\underline{v}\| = (t\,\underline{v} \cdot t\,\underline{v})^{\frac{1}{2}} = \left(t^2\, \underline{v} \cdot \underline{v}\right)^{\frac{1}{2}} = |\,t\,|\, \|\underline{v}\| .$$

Also notice that $\|\underline{v}\| > 0$ unless $\underline{v} = \underline{\mathbf{0}}$.

2b) The distance between points (with position vectors) $\underline{\boldsymbol{P}}, \underline{\boldsymbol{Q}}$ is defined to be the magnitude of the displacement vector(s) between them, $\|\underline{\boldsymbol{Q}} - \underline{\boldsymbol{P}}\|$ (or $\|\underline{\boldsymbol{P}} - \underline{\boldsymbol{Q}}\|$).

2c) For $\underline{\mathbf{v}}, \underline{\mathbf{w}} \in \mathbb{R}^n$, we define $\underline{\mathbf{v}}$ to be *orthogonal* (or *perpendicular*) to $\underline{\mathbf{w}}$ if and only if

$$\underline{\mathbf{v}} \bullet \underline{\mathbf{w}} = 0.$$

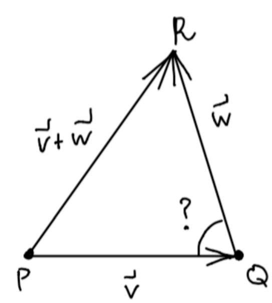
And in this case we write $\underline{\mathbf{v}} \perp \underline{\mathbf{w}}$.

Note: In $\mathbb{R}^2$ or $\mathbb{R}^3$ and in your multivariable calculus class, this definition was a special case of the identity

$$\underline{\mathbf{v}} \bullet \underline{\mathbf{w}} = \|\underline{\mathbf{v}}\| \; \|\underline{\mathbf{w}}\| \cos(\theta)$$

where θ is the angle between $\underline{\mathbf{v}}, \underline{\mathbf{w}}$. (Because $\cos(\theta) = 0$ when $\theta = \frac{\pi}{2}$.) That identity followed from the law of cosines, although you probably don't recall the details. Today we'll use the identity above to *define* angles between vectors, in $\mathbb{R}^n$, and show that it makes sense. (Next week, in section 6.6 we'll see that what we call "$\cos(\theta)$" in Math 2270 is known as the "correlation coefficient" in linear regression problems in statistics. In about two weeks, we will use the same identity to define angles between functions and perpendicular functions, in inner product function spaces, sections 6.8-6.9.)

2d) The $\mathbb{R}^n$ reason for defining orthogonality as in 2c is that the Pythagorean Theorem holds for the triangle with displacement vectors $\underline{\mathbf{v}}, \underline{\mathbf{w}}$ and hypotenuse $\underline{\mathbf{v}} + \underline{\mathbf{w}}$ if and only if $\underline{\mathbf{v}} \bullet \underline{\mathbf{w}} = 0$. Check!



2e) A vector $\underline{\boldsymbol{u}} \in \mathbb{R}^n$ is called a *unit vector* if and only if $||\underline{\boldsymbol{u}}|| = 1$.

2f) If $\underline{\boldsymbol{v}} \in \mathbb{R}^n$ then the unit vector in the direction of $\underline{\boldsymbol{v}}$ is given by

$$\underline{\boldsymbol{u}} = \frac{1}{||\underline{\boldsymbol{v}}||}\underline{\boldsymbol{v}}.$$

2g) Projection onto a line. Let $\underline{\boldsymbol{v}} \in \mathbb{R}^n$ be a non-zero vector, let $L = span\{\underline{\boldsymbol{v}}\}$ be a line through the origin. Then for any $\underline{\boldsymbol{x}} \in \mathbb{R}^n$ the projection of $\underline{\boldsymbol{x}}$ onto L is defined by the formula

$$proj_L\, \underline{\boldsymbol{x}} := (\underline{\boldsymbol{x}} \cdot \underline{\boldsymbol{u}})\, \underline{\boldsymbol{u}}$$

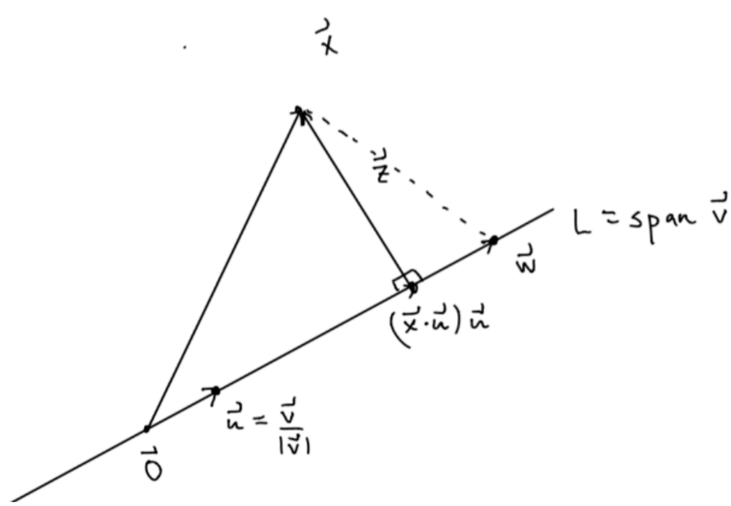
for $\underline{\boldsymbol{u}}$ the unit vector in the direction of $\underline{\boldsymbol{v}}$, $\underline{\boldsymbol{u}} = \frac{1}{||\underline{\boldsymbol{v}}||}\underline{\boldsymbol{v}}$. Equivalently

$$proj_L\, \underline{\boldsymbol{x}} := \frac{(\underline{\boldsymbol{x}} \cdot \underline{\boldsymbol{v}})}{||\underline{\boldsymbol{v}}||^2}\underline{\boldsymbol{v}}.$$

Then the vector

$$\underline{\boldsymbol{z}} := \underline{\boldsymbol{x}} - (\underline{\boldsymbol{x}} \cdot \underline{\boldsymbol{u}})\, \underline{\boldsymbol{u}}$$

is perpendicular to every vector in $span\{\underline{\boldsymbol{v}}\} = span\{\underline{\boldsymbol{u}}\}$. Thus every triangle containing (the point with position vector) $\underline{\boldsymbol{x}}$, (the point with position vector) $proj_L\, \underline{\boldsymbol{x}}$, and another point (with position vector) $\underline{\boldsymbol{w}}$ on the line L is a Pythagorean triangle. Consequently, $proj_L \underline{\boldsymbol{x}}$ is the (position vector of) nearest point on L to (the point with position vector) $\underline{\boldsymbol{x}}$.



2h) Refer to the same diagram as in 2g, which is an $\mathbb{R}^n$ picture. Using the Pythagorean triangle with edges $(\underline{x} \cdot \underline{u})\underline{u}$, $\underline{z}$, $\underline{x}$ we have

$$\| (\underline{x} \cdot \underline{u})\underline{u} \|^2 + \|\underline{z}\|^2 = \|\underline{x}\|^2, \text{ i.e.}$$

$$(\underline{x} \cdot \underline{u})^2 + \| \underline{z} \|^2 = \|\underline{x}\|^2.$$

The quantity $\underline{x} \cdot \underline{u}$ is called *the component of* $\underline{x}$ in the direction of $\underline{u}$, and from the formula above,

$$-\|x\| \le \underline{x} \cdot \underline{u} \le \| x\| \ .$$

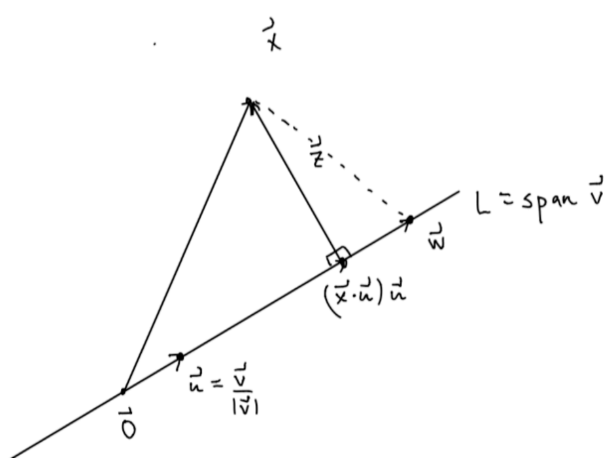
Define the angle θ between $\underline{v}$ and $\underline{w}$ the same way we would in $\mathbb{R}^2$, using the congruent triangle in the figure below, namely

$$\cos(\theta) = \frac{(\underline{x} \cdot \underline{u})}{\|\underline{x}\|}.$$

Notice that $-1 \le \cos(\theta) \le 1$ and so there is a unique θ with $0 \le \theta \le \pi$ for which the cos θ equation can hold. Substituting $\underline{u} = \frac{\underline{v}}{\|\underline{v}\|}$ gives the familiar formulas that you learned in multivariable Calculus for $\mathbb{R}^2$, $\mathbb{R}^3$, which now holds in $\mathbb{R}^n$.

$$\cos(\theta) = \frac{\left(\underline{x} \cdot \frac{\underline{v}}{\|\underline{v}\|}\right)}{\|\underline{x}\|} = \frac{(\underline{x} \cdot \underline{v})}{\|\underline{x}\| \ \|\underline{v}\|}, \text{ i.e.}$$

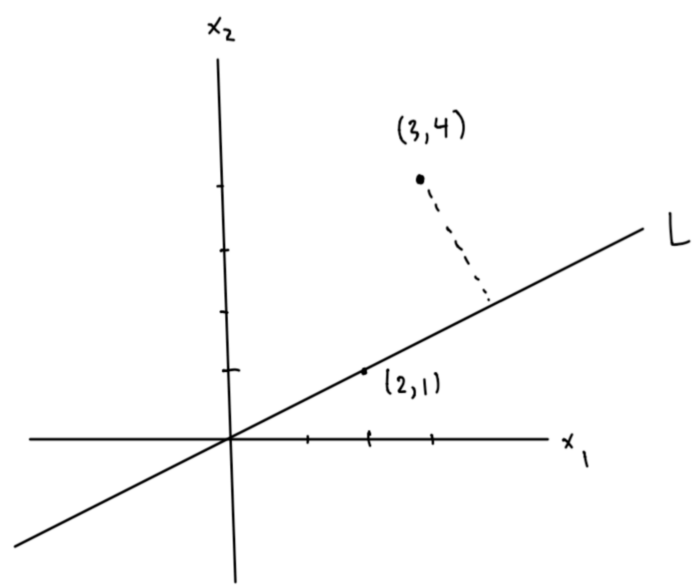
$$(\underline{x} \cdot \underline{v}) = \|\underline{x}\| \ \|\underline{v}\| \cos(\theta)$$



3) Summary exercise

a) In $\mathbb{R}^2$, let $L = span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$. Find $proj_L \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Illustrate.

b) Verify the Pythagorean Theorem for some triple of points where two of them are $(3, 4)$, the point with position vector $proj_L \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, and the third one is any other point on $L = span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$.



Tues Nov 13

- 6.2-6.3 Orthogonal complements, and the four fundamental subspaces of a matrix revisited.

Announcements:

Warm-up Exercise:

Orthogonal complements, and the four subspaces associated with a matrix transformation, revisited more carefully than our first time through.

Let $W \subseteq \mathbb{R}^n$ be a subspace of dimension $1 \leq p \leq n$. The *orthogonal complement to W* is the collection of all vectors perpendicular to every vector in W. We write the orthogonal complement to W as $W^{\perp}$, and say "W perp". Let $B = \{\underline{w}_1, \underline{w}_2, \ldots \underline{w}_p\}$ be a basis for W. Let $\underline{v} \in W^{\perp}$. This means

$$\left(c_1 \underline{w}_1 + c_2 \underline{w}_2 + \ldots + c_p \underline{w}_p\right) \cdot \underline{v} = 0$$

for all linear combinations of the spanning vectors. Since the dot product distributes over linear combinations, the identity above expands as

$$c_1\left(\underline{w}_1 \cdot \underline{v}\right) + c_2\left(\underline{w}_2 \cdot \underline{v}\right) + \ldots + c_p\left(\underline{w}_p \cdot \underline{v}\right) = 0$$

for all possible weights. This is always true as soon as we check the special cases

$$\underline{w}_1 \cdot \underline{v} = \underline{w}_2 \cdot \underline{v} = \ldots = \underline{w}_p \cdot \underline{v} = 0.$$

In other words, $\underline{v} \in Nul\, A$ where A is the $m \times n$ matrix having the spanning vectors as rows:

$$A\,\underline{v} = \begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_p^T \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \underline{\mathbf{0}}.$$

So

$$W^{\perp} = Nul\, A.$$

<u>Exercise 1</u> Find $W^{\perp}$ for $W = span\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \right\}$.

Theorem: Let A be any $m \times n$ matrix. Then $(Row\ A)^{\perp} = Nul\ A$.

Theorem. Conversely, let A be an $m \times n$ matrix. Then $(Nul\ A)^{\perp} = Row\ A$.

Corollary Let $W \subseteq \mathbb{R}^n$ be a subspace. Then $(W^{\perp})^{\perp} = W$.

Exercise 2 For $W = span\left\{\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}\right\}$ as in Exercise 1, verify that $(W^{\perp})^{\perp} = W$.

Theorem (fill in details).

1a) Let $W \subseteq \mathbb{R}^n$ be a subspace with $dim\ W = r,\ 1 \le r \le n$. Then $dim(W^\perp) = n - r$, so

$$dim(W) + dim(W^\perp) = n$$

Hint: Use reduced row echelon form ideas.

1b) $W \cap W^\perp = \{\mathbf{0}\}$

Hint: Let $\mathbf{x} \in W \cap W^\perp$. Compute $\mathbf{x} \cdot \mathbf{x}$.

1c) Let $B = \{\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_p\}$ be a basis for W and $C = \{\mathbf{z}_1, \mathbf{z}_2, \dots \mathbf{z}_{n-p}\}$ be a basis for $W^\perp$. Then their union, $B \cup C$, is a basis for $\mathbb{R}^n$.

Hint: Show $B \cup C$ is linearly independent.

1d) Every $\mathbf{x} \in \mathbb{R}^n$ can we written as $\mathbf{x} = \mathbf{w} + \mathbf{z}$ with unique choices of $\mathbf{w} \in W$ and $\mathbf{z} \in W^\perp$.

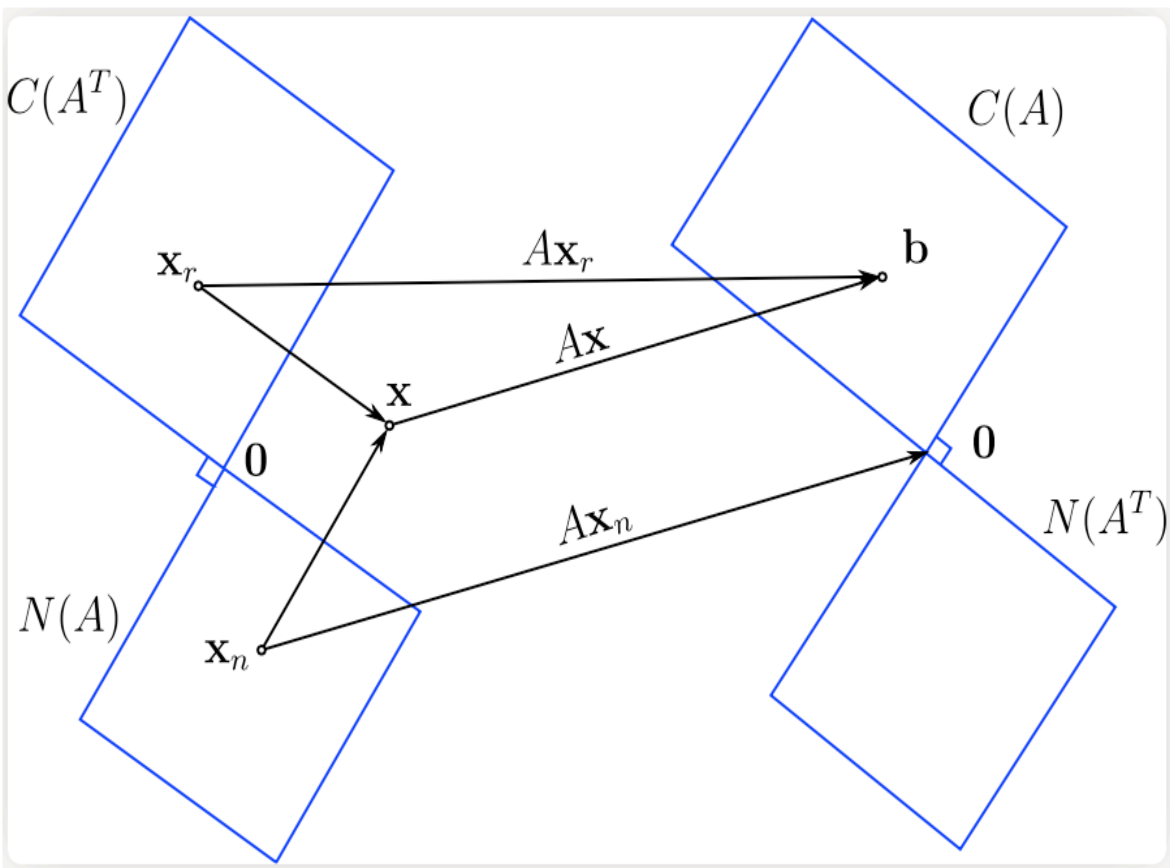
Remark: From the discussion above and our previous knowledge, we see that for any $m \times n$ matrix A, $(Row\ A)^{\perp} = Nul\ A$; $(Nul\ A)^{\perp} = Row\ A$; $dim\ (Row\ A) = r$, $dim\,(Nul\ A) = n - r$, where r is the rank of the matrix. Furthermore from 1d on the previous page, each $\underline{x} \in \mathbb{R}^n$ can we written uniquely as $\underline{x} = \underline{w} + \underline{z}$ with $\underline{w} \in Row\ A$ and $\underline{z} \in Nul\ A$. So for the linear transformation

$$T(\underline{x}) := A\,\underline{x}$$

we have

$$T(\underline{x}) = A\,\underline{x} = A(\underline{w} + \underline{z}) = A\,\underline{w} + A\,\underline{z} = A\,\underline{w}.$$

By the same reasoning applied to the transpose transformation from $\mathbb{R}^m \to \mathbb{R}^n$, the codomain of T decomposes into $Col\ A = Row\ A^T$ and $(Col\ A)^{\perp} = Nul\ A^T$, with $dim\,(Col\ A) = r$ and $dim\ \left(Nul\ A^T\right) = m - r$. In other words, we have justified the diagram we really only waved our hands at back in Chapter 4:



Wed Apr 4

- 6.2-6.3 very good bases revisited: orthogonal and orthonormal bases. Projection onto multi-dimensional subspaces.

Announcements:

Warm-up Exercise:

Definitions: a The set $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p\} \subseteq \mathbb{R}^n$ is called *orthogonal* if and only if all the vectors are non-zero and

$$\underline{v}_i \cdot \underline{v}_j = 0, \quad i \neq j, \; i, j = 1 \dots p$$

(The vectors in an orthogonal set are mutually perpendicular to each other.)

b The set $\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\} \subseteq \mathbb{R}^n$ is called *orthonormal* if and only if

$$\underline{u}_i \cdot \underline{u}_j = 0, \quad i \neq j, \; i, j = 1 \dots p.$$

$$\underline{u}_i \cdot \underline{u}_i = 1, \; i = 1, 2, \dots p$$

So this is a set of mutually orthogonal vectors that are all unit vectors.

Remark: If $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p\} \subseteq \mathbb{R}^n$ is an orthogonal set, then there is a corresponding orthonormal set

$$\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\} = \left\{ \frac{\underline{v}_1}{\|\underline{v}_1\|}, \frac{\underline{v}_2}{\|\underline{v}_2\|}, \dots, \frac{\underline{v}_p}{\|\underline{v}_p\|} \right\}$$

which spans the same subspace as the original orthogonal set.

Examples of ortho-normal sets you know already:

1) The standard basis $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\} \subseteq \mathbb{R}^n$, or any subset of the standard basis vectors.

2) Rotated bases in $\mathbb{R}^2$. $\{\underline{u}_1, \underline{u}_2\} = \left\{ \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix} \right\}$.

Theorem (why orthonormal sets are very good bases): Let $B = \{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\} \subseteq \mathbb{R}^n$ be orthonormal. Let $W = span\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\}$. Then

a) $\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\}$ is linearly independent, so a basis for W.

b) For $\underline{w} \in W$, the coordinate vector $[\underline{w}]_B = \begin{bmatrix} \underline{u}_1 \cdot \underline{w} \\ \underline{u}_2 \cdot \underline{w} \\ \vdots \\ \underline{u}_p \cdot \underline{w} \end{bmatrix}$ is directly computable. In other words,

$$\underline{\boldsymbol{w}} = \left(\underline{\boldsymbol{u}}_1 \bullet \underline{\boldsymbol{w}}\right)\underline{\boldsymbol{u}}_1 + \left(\underline{\boldsymbol{u}}_2 \bullet \underline{\boldsymbol{w}}\right)\underline{\boldsymbol{u}}_2 + \ldots + \left(\underline{\boldsymbol{u}}_p \bullet \underline{\boldsymbol{w}}\right)\underline{\boldsymbol{u}}_p$$

c) Let $\underline{\boldsymbol{x}} \in \mathbb{R}^n$. Then there is a unique nearest point to $\underline{\boldsymbol{x}}$ in W, which we call $proj_W\underline{\mathbf{x}}$, ("the projection of $\underline{\boldsymbol{x}}$ onto W.") The formula for this projection is given by

$$proj_W\underline{\mathbf{x}} = (\underline{\boldsymbol{u}}_1 \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{u}}_2 \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_2 + ... + (\underline{\boldsymbol{u}}_p \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_p .$$

(As should be the case, projection onto W leaves elements of W fixed.)

Proof: We will use the Pythagorean Theorem to show that the formula above for $proj_W\underline{\mathbf{x}}$ yields the nearest point in W to $\underline{\boldsymbol{x}}$:

Define

$$\underline{\boldsymbol{z}} = \underline{\mathbf{x}} - proj_W\underline{\mathbf{x}}$$

$$\underline{\boldsymbol{z}} = \underline{\mathbf{x}} - (\underline{\boldsymbol{u}}_1 \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_1 - (\underline{\boldsymbol{u}}_2 \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_2 - ... - (\underline{\boldsymbol{u}}_p \bullet \underline{\mathbf{x}})\underline{\boldsymbol{u}}_p.$$

Then for $j = 1, 2, ... p$,

$$\underline{\boldsymbol{z}} \bullet \underline{\boldsymbol{u}}_j = \underline{\mathbf{x}} \bullet \underline{\boldsymbol{u}}_j - \underline{\mathbf{x}} \bullet \underline{\boldsymbol{u}}_j = 0 .$$

So $\underline{\boldsymbol{z}} \perp W$, i.e.

$$\underline{\boldsymbol{z}} \bullet (t_1 \underline{\boldsymbol{u}}_1 + t_2 \underline{\boldsymbol{u}}_2 + ... t_p \underline{\boldsymbol{u}}_p) = 0$$

for all choices of the weight vector $\underline{\boldsymbol{t}}$.

Let $\underline{\boldsymbol{w}} \in W$. Then

$$\|\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2 = \|(\underline{\mathbf{x}} - proj_W\underline{\mathbf{x}}) + (proj_W\underline{\mathbf{x}} - \underline{\mathbf{w}})\|^2 .$$

Since $(\underline{\mathbf{x}} - proj_W\underline{\mathbf{x}}) = \underline{\boldsymbol{z}}$ and since $(proj_W\underline{\mathbf{x}} - \underline{\mathbf{w}}) \in W$, we have the Pythagorean Theorem

$$\|\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2 = \|\underline{x} - proj_W\underline{\mathbf{x}}\|^2 + \|proj_W\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2$$

$$\|\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2 = \|\underline{\boldsymbol{z}}\|^2 + \|proj_W\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2 .$$

So $\|\underline{\mathbf{x}} - \underline{\mathbf{w}}\|^2$ is always at least $\|\underline{\boldsymbol{z}}\|^2$, with equality if and only if $\underline{\boldsymbol{w}} = proj_W\underline{\mathbf{x}}$.

QED

Exercise 1

1a) Check that the set

$$B = \left\{ \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} \right\}$$

is an orthonormal basis for $\mathbb{R}^3$.

1b) For $\underline{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ find the coordinate vector $[\underline{\boldsymbol{x}}]_B$ and check your answer.

solution $[\underline{\boldsymbol{x}}]_B = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$

Exercise 2 Consider the plane from Tuesday

$$W = span\left\{\begin{bmatrix}1\\1\\3\end{bmatrix}, \begin{bmatrix}1\\0\\-2\end{bmatrix}\right\}$$

which is also given implicitly as a nullspace,

$$W = Nul\begin{bmatrix}2 & -5 & 1\end{bmatrix}.$$

2a) Verify that

$$B = \left\{\frac{1}{\sqrt{5}}\begin{bmatrix}1\\0\\-2\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}2\\1\\1\end{bmatrix}\right\}$$

is an ortho-normal basis for W.

2b) Find $proj_W\underline{\mathbf{x}}$ for $\underline{\mathbf{x}} = \begin{bmatrix}7\\-3\\1\end{bmatrix}$. Then verify that $\underline{\mathbf{z}} = \underline{\mathbf{x}} - proj_W\underline{\mathbf{x}}$ is perpendicular to W.

solution $proj_W\underline{\mathbf{x}} = \begin{bmatrix}5\\2\\0\end{bmatrix}$

Remark: As we mentioned, if $\{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots, \underline{\boldsymbol{v}}_p\} \subseteq \mathbb{R}^n$ is an *orthogonal* set (of mutually perpendicular vectors), then there is the corresponding orthonormal basis obtained from that set by normalizing, namely

$$\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots, \underline{\boldsymbol{u}}_p\} = \left\{ \frac{\underline{\boldsymbol{v}}_1}{\|\underline{\boldsymbol{v}}_1\|}, \frac{\underline{\boldsymbol{v}}_2}{\|\underline{\boldsymbol{v}}_2\|}, \ldots, \frac{\underline{\boldsymbol{v}}_p}{\|\underline{\boldsymbol{v}}_p\|} \right\} \subseteq \mathbb{R}^n .$$

One can avoid square roots if one uses the original orthogonal basis rather than the ortho-normal one. This is the approach the text prefers. For example, for orthogonal bases, the very good basis theorem reads

Theorem (why orthogonal bases are very good bases): Let $B = \{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots, \underline{\boldsymbol{v}}_p\} \subseteq \mathbb{R}^n$ be orthogonal. Let $W = span\{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots, \underline{\boldsymbol{v}}_p\}$. Then

a) $\{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots, \underline{\boldsymbol{v}}_p\}$ is linearly independent, so a basis for W.

b) For $\underline{\boldsymbol{w}} \in W$,

$$\underline{\boldsymbol{w}} = (\underline{\boldsymbol{u}}_1 \bullet \underline{\boldsymbol{w}})\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{u}}_2 \bullet \underline{\boldsymbol{w}})\underline{\boldsymbol{u}}_2 + \ldots + (\underline{\boldsymbol{u}}_p \bullet \underline{\boldsymbol{w}})\underline{\boldsymbol{u}}_p$$

$$\underline{\boldsymbol{w}} = \frac{(\underline{\boldsymbol{v}}_1 \bullet \underline{\boldsymbol{w}})}{\|\underline{\boldsymbol{v}}_1\|^2}\underline{\boldsymbol{v}}_1 + \frac{(\underline{\boldsymbol{v}}_2 \bullet \underline{\boldsymbol{w}})}{\|\underline{\boldsymbol{v}}_2\|^2}\underline{\boldsymbol{v}}_2 + \ldots + \frac{(\underline{\boldsymbol{v}}_p \bullet \underline{\boldsymbol{w}})}{\|\underline{\boldsymbol{v}}_p\|^2}\underline{\boldsymbol{v}}_p$$

c) Let $\underline{\boldsymbol{x}} \in \mathbb{R}^n$. Then there is a unique nearest point to $\underline{\boldsymbol{x}}$ in W, which we call $proj_W\underline{\boldsymbol{x}}$, ("the projection of $\underline{\boldsymbol{x}}$ onto W.") The formula for this projection is given by

$$proj_W\underline{\boldsymbol{x}} = (\underline{\boldsymbol{u}}_1 \bullet \underline{\boldsymbol{x}})\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{u}}_2 \bullet \underline{\boldsymbol{x}})\underline{\boldsymbol{u}}_2 + \ldots + (\underline{\boldsymbol{u}}_p \bullet \underline{\boldsymbol{x}})\underline{\boldsymbol{u}}_p .$$

$$proj_W\underline{\boldsymbol{x}} = \frac{(\underline{\boldsymbol{v}}_1 \bullet \underline{\boldsymbol{x}})}{\|\underline{\boldsymbol{v}}_1\|^2}\underline{\boldsymbol{v}}_1 + \frac{(\underline{\boldsymbol{v}}_2 \bullet \underline{\boldsymbol{x}})}{\|\underline{\boldsymbol{v}}_2\|^2}\underline{\boldsymbol{v}}_2 + \ldots + \frac{(\underline{\boldsymbol{v}}_p \bullet \underline{\boldsymbol{x}})}{\|\underline{\boldsymbol{v}}_p\|^2}\underline{\boldsymbol{v}}_p .$$

You can see how that would have played out in the previous exercise.

Fri Apr 6

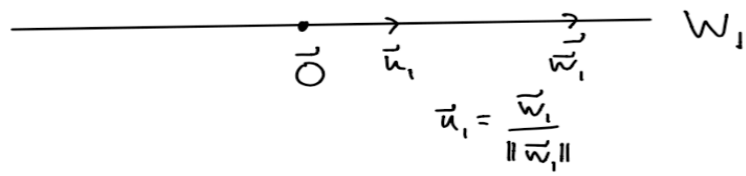
- 6.3-6.4 Gram-Schmidt process for constructing ortho-normal (or orthogonal) bases. The $A = Q\,R$ matrix factorization. (I'll bring notes to class for the second topic, if it looks like we'll have time on Friday. Otherwise we'll discuss it on Monday.)

<u>Announcements:</u>

<u>Warm-up Exercise</u>:

Start with a non-orthogonal basis $B = \{\underline{w}_1, \underline{w}_2, \ldots \underline{w}_p\}$ for a subspace W of $\mathbb{R}^n$. How can you convert it into an orthonormal basis? Here's how! The inductive process is called *Gram-Schmidt orthogonalization*.

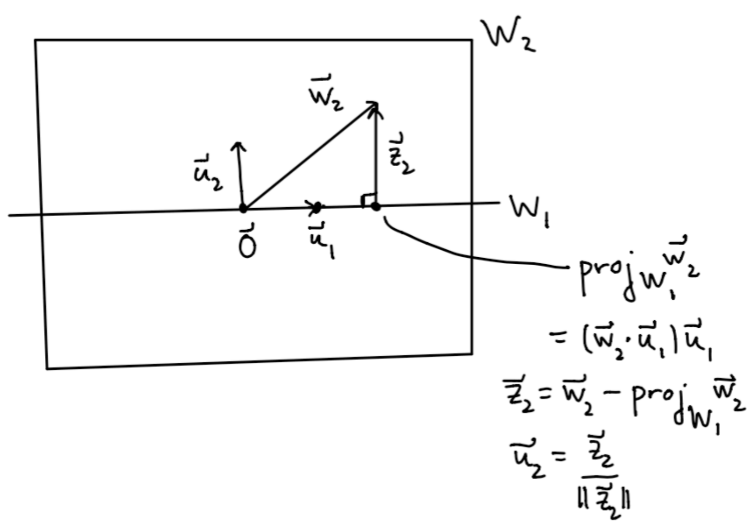
Let $W_1 = span\{\underline{w}_1\}$. Define $\underline{u}_1 = \dfrac{\underline{w}_1}{\|\underline{w}_1\|}$. Then $\{\underline{u}_1\}$ is an orthonormal basis for W_1.



Let $W_2 = span\{\underline{w}_1, \underline{w}_2\}$.

Let $\underline{z}_2 = \underline{w}_2 - proj_{W_1}\underline{w}_2$, so $\underline{z}_2 \perp \underline{u}_1$.

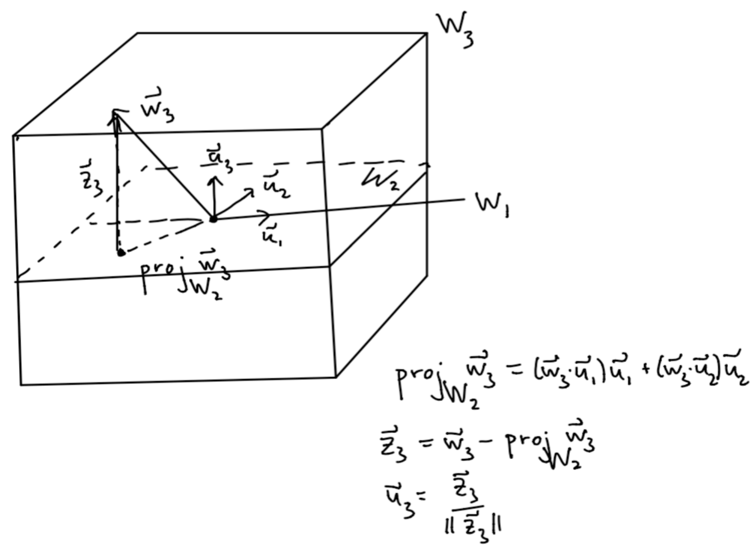
Define $\underline{u}_2 = \dfrac{\underline{z}_2}{\|\underline{z}_2\|}$. So $\{\underline{u}_1, \underline{u}_2\}$ is an orthonormal basis for W_2.



Let $W_3 = span\{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \underline{\boldsymbol{w}}_3\}$.

Let $\underline{\boldsymbol{z}}_3 = \underline{\boldsymbol{w}}_3 - proj_{W_2}\underline{\boldsymbol{w}}_3$, so $\underline{\boldsymbol{z}}_3 \perp W_2$.

Define $\underline{\boldsymbol{u}}_3 = \dfrac{\underline{\boldsymbol{z}}_3}{\|\underline{\boldsymbol{z}}_3\|}$. Then $\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \underline{\boldsymbol{u}}_3\}$ is an orthonormal basis for W_3.



Inductively,

Let $W_j = span\{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_j\} = span\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_{j-1}, \underline{\boldsymbol{w}}_j\}$.

Let $\underline{\boldsymbol{z}}_j = \underline{\boldsymbol{w}}_j - proj_{W_{j-1}}\underline{\boldsymbol{w}}_j = \underline{\boldsymbol{w}}_j - (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 - (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2 - \ldots (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_{j-1})\underline{\boldsymbol{u}}_{j-1}$.

Define $\underline{\boldsymbol{u}}_j = \dfrac{\underline{\boldsymbol{z}}_j}{\|\underline{\boldsymbol{z}}_j\|}$. Then $\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_j\}$ is an orthonormal basis for W_j .

Continue up to $j = p$.

Exercise 1 Perform Gram-Schmidt orthogonalization on the basis

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\}.$$

Sketch what you're doing, as you do it.

Exercise 2 Perform Gram-Schmidt on the basis

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \right\}.$$

This will proceed as in Exercise 1 until the third step, i.e.

$$\underline{\boldsymbol{u}}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{\boldsymbol{u}}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

The $A = Q\,R$ matrix decomposition:

We're denoting the original basis for W by $B = \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_p\}$. Denote the orthonormal basis we've constructed with Gram-Schmidt by $O = \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_p\}$. Because O is orthonormal it's easy to express these two bases in terms of each other. Notice

$$W_j = span\ \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_j\} = span\ \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_j\} \qquad \text{for each } 1 \le j \le p.$$

So,

$$\underline{\boldsymbol{w}}_1 = (\underline{\boldsymbol{w}}_1 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1$$

$$\underline{\boldsymbol{w}}_2 = (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_j = (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2 + \ldots.\ (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)\underline{\boldsymbol{u}}_j$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_p = \sum_{l=1}^{p} (\underline{\boldsymbol{w}}_l \cdot \underline{\boldsymbol{u}}_l)\underline{\boldsymbol{u}}_l\ .$$

Notice that the coefficients of the last terms in the sums above, namely $(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)$ can be computed as

$$(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j) = \underline{\boldsymbol{z}}_j \cdot \frac{\underline{\boldsymbol{z}}_j}{\|\underline{\boldsymbol{z}}_j\|} = \|\,\underline{\boldsymbol{z}}_j\,\|.$$

In matrix form (column by column) we have

$$* \quad \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \\ \vdots & \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \\ | & | & & | \end{bmatrix} \begin{bmatrix} \vec{w}_1\cdot\vec{u}_1 & \vec{w}_2\cdot\vec{u}_1 & \vec{w}_3\cdot\vec{u}_1 & \cdots & \vec{w}_p\cdot\vec{u}_1 \\ 0 & \vec{w}_2\cdot\vec{u}_2 & \vec{w}_3\cdot\vec{u}_2 & & \vec{w}_p\cdot\vec{u}_2 \\ 0 & 0 & \vec{w}_3\cdot\vec{u}_3 & \ddots & \vdots \\ \vdots & \vdots & 0 & & \\ 0 & 0 & 0 & & \vec{w}_p\cdot\vec{u}_p \end{bmatrix}$$

"A" — columns are original basis for W — $A_{n\times p}$

"Q" — columns are orthonormal — $Q_{n\times p}$

"R" — upper Δ'ular, with diagonal entries $\vec{w}_j\cdot\vec{u}_j = \|\vec{z}_j\|$ — $R_{p\times p}$

R also known as $P_{B\leftarrow O}$ ☺

Thus any matrix with linearly independent columns may be written in factored form as above, ($W = Col\,A$),

$$A_{n\times p} = Q_{n\times p}\,R_{p\times p}.$$

This factorization contains geometric information and can simplify the computational work needed to solve matrix equations $A\,\underline{\boldsymbol{x}} = \underline{\boldsymbol{b}}$.

From previous page...

$$* \qquad A_{n \times p} = Q_{n \times p} R_{p \times p}$$

<u>shortcut</u> (or what to do if you forgot the formulas for the entries of R) If you just know Q you can recover R by multiplying both sides of the $*$ equation on the previous page by the transpose Q^T of the Q matrix:

$$\begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_p^T \end{bmatrix} \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \end{bmatrix} = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_p^T \end{bmatrix} \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \end{bmatrix} R = IR = R!$$

$$A = Q\,R$$
$$Q^T A = Q^T Q\, R = I\, R = R.$$

<u>Example)</u> From <u>Exercise 1</u>

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\}, \quad O = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}.$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{w}}_1 \cdot \underline{\boldsymbol{u}}_1 & \underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_1 \\ 0 & \underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} = Q\,R.$$

<u>Exercise 3</u>) Verify that R could have been recovered via the formula

$$Q^T A = R$$

From previous page ...

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix}.$$

Exercise 4) Verify that the $A = Q\,R$ factorization in this example may be further factored as

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

- So, the transformation $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$ is a composition of (1) an area-preserving shear, followed by (2) a diagonal scaling that increases area by a factor of $\sqrt{2}\cdot 2\sqrt{2} = 4$, followed by a rotation of $\frac{\pi}{4}$, which does not effect area. Since determinants of products matrices are the products of determinants (we checked this back when we studied determinants), and area expansion factors of compositions are also the products of the area expansion factors, the generalization of this example gives another explanation of why the determinant of A (or its absolute value in general) coincides with the area expansion factor, in the 2×2 case. You show in your homework that the only possible Q matrices in the 2×2 case are rotations as above, or reflections across lines through the origin. In the latter case, the determinant of Q is -1, and the determinant of A is negative.

To be continued....