Orthogonality: Sections 3.1 & 3.2

Recall, a basis is a set of independent vectors that span a space.

Geometrically, it is a set of coordinate axes.

"In choosing a basis, we usually choose an orthogonal basis."
(perpendicular)

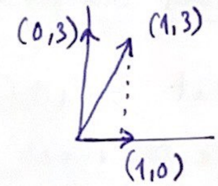
- Basis can convert geometric constructions into algebraic calculation.
- An orthogonal basis makes such calculations easier.
- Another specialization is the orthonormal basis, where each vector has length 1 & is orthogonal.

Length of a vector: $\|v\|$

In 2-D ~ hypotenuse of a right angled triangle.

$v = (v_1, v_2)$, $\|v\|^2 = v_1^2 + v_2^2$ eg $v = (1,3)$, $\|v\| = 1^2 + 3^2 = 10$

In 3-D, $v = (v_1, v_2, v_3)$ is the diagonal of a box ~~with~~, in a similar way.



$\|v\|^2 = v_1^2 + v_2^2 + v_3^2$.

Defn: The length of $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ is the positive square root of $v^T v$: $\|v\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$.

Orthogonal vectors:

Q. When are two vectors v_1 and v_2 perpendicular?

A: In the plane spanned by v_1 & v_2, the vectors are orthogonal provided they form a right angled triangle.

$\rightarrow \|v_1\|^2 + \|v_2\|^2 = \|v_1 - v_2\|^2$ where $v_1 = (v_{11}, v_{12}, \dots, v_{1n})$

$v_2 = (v_{21}, v_{22}, \dots, v_{2n})$

iff $(v_{11}^2 + v_{12}^2 + \dots + v_{1n}^2)$

$+ (v_{21}^2 + v_{22}^2 + \dots + v_{2n}^2) = (v_{11} - v_{21})^2 + \dots + (v_{1n} - v_{2n})^2$

iff $v_{11} \cdot v_{21} + v_{12} \cdot v_{22} + \dots + v_{1n} \cdot v_{2n} = 0$.

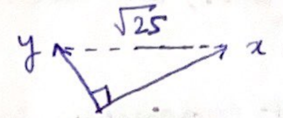
Defn: Two vectors $x = (x_1, \dots, x_n)$ & $y = (y_1, \dots, y_n)$ are orthogonal when $x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \underbrace{x^T y}_{} = 0$.

$\hookrightarrow$ Inner product / scalar product / dot prod.

Remark: If $x^T y > 0$ then the angle betⁿ x & y $< 90°$.
If $x^T y < 0$ - - - - - - - - . $> 90°$.

eg: $x = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$, $y = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, $x^T y = 4(-1) + 2 \times 2 = 0$.

$\|x\| = \sqrt{20}$, $\|y\| = \sqrt{5}$



- $x^T x = \|x\|^2$ for any vector $x \in \mathbb{R}^n$.
- The only vector orthogonal to itself is the zero vector.
- $x = 0$ is orthogonal to every vector in $\mathbb{R}^n$.

Theorem: If non-zero vectors $v_1, v_2, \ldots, v_k$ are mutually orthogonal (i.e every vector is perpendicular to every other) then those vectors are linearly independent.

Proof: Let $c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = 0$ where c_i's are scalars.
We take the inner product of both sides with v_1 to get,

$$v_1^T (c_1 v_1 + \cdots + c_k v_k) = c_1 v_1^T \cdot v_1 = 0$$

Since v_i's are non-zero so, $v_1^T v_1 \neq 0$ so, $c_1 = 0$.
Similarly, we can show all c_i's $= 0$ which implies independence. //

eg: The vectors $e_1, e_2, \ldots, e_n \in \mathbb{R}^n$ are orthogonal vectors.
They are also of unit length, so they form an orthonormal basis.
These vectors point along the axis: If we rotate them then we get a new orthonormal basis, a new system of mutually orthogonal unit vectors.

Let us now return to the concept of the inner product and lengths.

Theorem Let $u, v, w \in \mathbb{R}^n$ be vectors and α be a scalar. Then, we have.

(i) $u^T v = v^T u$, (ii) $(u+v)^T w = u^T w + v^T w$,

(iii) $(\alpha u)^T v = \alpha u^T v = u^T (\alpha v)$

(iv) $u^T u \geq 0$ and $u^T u = 0$ iff $u = 0$.

Theorem: Let $u \in \mathbb{R}^n$ and let α be a scalar. Then

$$\|\alpha u\| = |\alpha| \, \|u\|.$$

Proof: $\|\alpha u\| = \sqrt{(\alpha u)^T(\alpha u)} = \sqrt{\alpha^2(u^Tu)} = |\alpha|\sqrt{u^Tu} = |\alpha| \, \|u\|$. //

• This theorem actually says that any non-zero vector $u \in \mathbb{R}^n$ can be scaled to obtain a new unit vector in the same direction as u.

Let $u \neq 0 \Rightarrow \|u\| \neq 0$. Define $v = \frac{1}{\|u\|} u$, here $\frac{1}{\|u\|}$ is a scalar.

Using the previous theorem we have, $\|v\| = \frac{1}{\|u\|} \|u\| = 1$.

This process is called normalization of u.

Orthogonal Subspaces:

Def^n: Two subspaces V and W of the same space $\mathbb{R}^n$ are orthogonal if every vector $v \in V$ is orthogonal to every vector $w \in W$.

eg: The subspace $\{0\}$ is orthogonal to all subspaces.

eg: $V = \text{span}\{(1,0,0,0),(1,1,0,0)\}$, $W = \text{span}\{(0,0,4,5)\}$.

The line W is orthogonal to the plane V.

Remark: Two planes cannot be orthogonal to each other. In $\mathbb{R}^3$, there will be lines which are not perpendicular.

Fundamental theorem of orthogonality: The row space is orthogonal to the nullspace in $\mathbb{R}^n$. The column space is orthogonal to the left nullspace in $\mathbb{R}^m$.

Proof: If x is in the nullspace, then $Ax = 0$.

If v is in the row space then it is a combination of the rows, i.e. $v = A^Tz$ for some vector z.

Now, $v^Tx = (A^Tz)^Tx = z^TAx = z^T0 = 0$.

For the other part, let us look at $A^Ty = 0$ or, $y^TA = 0$.

$$y^T A = (y_1 \cdots y_m) \begin{pmatrix} \text{col } 1 & \text{col } 2 & \cdots & \text{col } n \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \end{pmatrix} = (0\ 0 \ldots 0)$$

i.e: The vector y is orthogonal to every column. So, y is orthogonal to every combination of the columns. i.e It is orthogonal to the col$^{\underline{n}}$ space. So, every vector $y \in N(A^T)$ is orthogonal to $C(A)$.

Thus, $N(A^T) \perp C(A)$. //.

Remark: Something more is true, $N(A)$ contains every vector orthogonal to the row space, $C(A^T)$.

Def$^{\underline{n}}$: Given a subspace V of $\mathbb{R}^n$, the space of all vectors which are orthogonal to V is called the orthogonal complement of V, denoted by $V^{\perp}$.

Theorem: $N(A) = (C(A^T))^{\perp}$ and $N(A^T) = (C(A))^{\perp}$.

Proof: Let z be a vector orthogonal to $N(A)$ but outside $C(A^T)$. This would mean that adding z as an extra row of A would enlarge $C(A^T)$ but this would violate the rank-nullity theorem. So, every vector orthogonal to the null space is in the row space.

Apply the same reasoning to A^T to get the other part. //.

Theorem: $Ax = b$ is solvable iff $y^T b = 0$ whenever $y^T A = 0$.

Proof: b is in $C(A)$, so b must be perpendicular to $N(A^T)$. //.

- b must be orthogonal to every vector that is orthogonal to the columns.

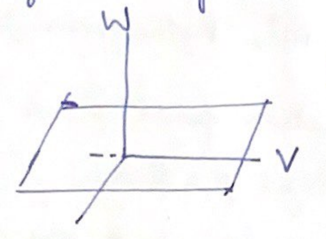
Remark: Two subspaces V & W can be orthogonal without being complements.

eg: $V = \text{span}\{(0,1,0)\}$, $W = \text{span}\{(0,0,1)\}$.

↳ $W^{\perp}$ is a 2-D plane.

Result: If $W = V^\perp$ then $V = W^\perp$ and $\dim V + \dim W = n$ where $V, W \subseteq \mathbb{R}^n$. In other words, $V^{\perp\perp} = V$.

- Splitting $\mathbb{R}^n$ into orthogonal parts, splits every $v \in \mathbb{R}^n$ as $v = x + y$, where x is the projection onto the subspace V, the orthogonal component ~~of~~ is the projection of v onto W.
- In $\mathbb{R}^3$: orthogonal complements ~ a plane and a line (NOT two lines).

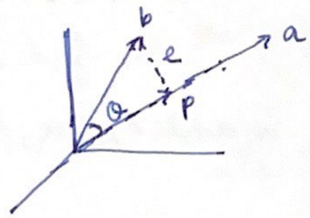


Line $W \perp$ plane V.

$V = W^\perp$.

We also want to understand inner products which are non-zero and angles which are not-right angled.

Geometrically, finding the projection p in 3-D is as follows:



, $e = b - p$

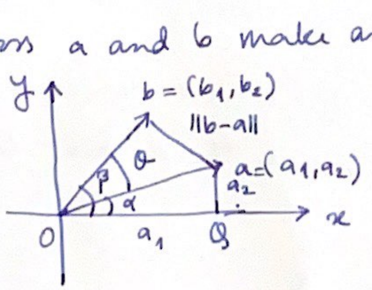
p = projection of b onto line through ~~b~~ a.

- Distance from a point b to line in the dir^n of a. $= e$.
- The line connecting b to p is $\perp$ to a.

For a plane (or any subspace say, S) the problem is to find the point p on that subspace that is closest to b. This is <u>the projection of b onto the subspace</u>.

<u>Inner products and Angles:</u> Let us examine the 2-D case first.

Suppose the vectors a and b make angles α and β with the x-axis.



$\|a\|$ = hyp. of $\Delta\, OQa$.

$\Rightarrow \sin\alpha = \dfrac{a_2}{\|a\|}, \cos\alpha = \dfrac{a_1}{\|a\|}$.

$\text{Ily}, \sin\beta = \dfrac{b_2}{\|b\|}, \cos\beta = \dfrac{b_1}{\|b\|}$.

Now, $\cos\theta = \cos(\beta - \alpha) = \cos\beta\cos\alpha + \sin\beta\sin\alpha$

$$= \frac{a_1 b_1 + a_2 b_2}{\|a\|\,\|b\|} = \frac{a^T b}{\|a\|\,\|b\|}.$$

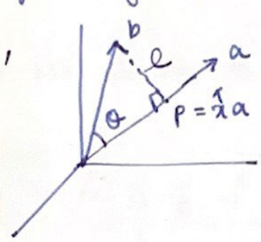
~~Defn~~: Theorem: The cosine of the angle betn any non-zero vectors a and b is, $\cos\theta = \dfrac{a^T b}{\|a\|\,\|b\|}$.

Law of cosines: $\|b-a\|^2 = \|b\|^2 + \|a\|^2 - 2\|b\|\,\|a\|\cos\theta$.

(When $\theta = 90°$, $\|b-a\|^2 = \|b\|^2 + \|a\|^2$, Pythagoras.)

Projection onto a line: The projection point p is actually some multiple of the given vector a, say, $p = \hat{x} a$.

Recall,



The line from b to the closest point $p = \hat{x} a$ is $\perp$ to the vector a.

$\Rightarrow (b-p) \perp a \Rightarrow a^T(b-p) = 0$

$\Rightarrow a^T b - \hat{x} a^T a = 0$

$\Rightarrow \hat{x} = \dfrac{a^T b}{a^T a}$. //

Schwarz Inequality: For vectors a and b, we have,

$$|a^T b| \le \|a\|\,\|b\|.$$

Proof: We have, $\|e\|^2 = \|b-p\|^2 \geq 0 \Leftrightarrow \left\|b - \dfrac{a^T b}{a^T a} a\right\|^2 \geq 0$

iff ~~then~~ $\left(b - \dfrac{a^T b}{a^T a} a\right)^T \left(b - \dfrac{a^T b}{a^T a} a\right) \geq 0$

iff $b^T b - 2\dfrac{(a^T b)^2}{a^T a} + \left(\dfrac{a^T b}{a^T a}\right)^2 a^T a \geq 0$

iff $(b^T b)(a^T a) - (a^T b)^2 \geq 0$

Then just take square roots. //

Remark: Here equality holds iff b is a multiple of a. (i.e. b is identical with its projection p, and the distance betn b and p is 0.)

Ex: Project $b=(1,2,3)$ onto the line through $a=(1,1,1)$ to get $\hat{x}$ & p.

Sol^n: $\hat{x} = \frac{a^Tb}{a^Ta} = 2$, $p = \hat{x}a = (2,2,2)$.

Transposes from Inner Products: The transpose A^T can be defined by the following property: The inner product of Ax with y equals the inner product of x with A^Ty.

i.e. $(Ax)^T y = x^T(A^Ty)$.

Theorem: $(AB)^T = B^TA^T$.

Proof: $(ABx)^Ty = (Bx)^T(A^Ty) = x^T(B^TA^Ty)$.

Corollary: $(A^{-1})^T = (A^T)^{-1}$. (Use $(AB)^{-1} = B^{-1}A^{-1}$.)

Projection Matrices: The projection matrix P is the matrix that multiplies b and produces p.

So, $p = a\frac{a^Tb}{a^Ta}$ so the proj. matrix $P = \frac{aa^T}{a^Ta}$.

eg: Let $a=(1,1,1)$, then $P = \frac{aa^T}{a^Ta} = \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}(1\,1\,1) = \begin{pmatrix}1/3 & 1/3 & 1/3\\ 1/3 & 1/3 & 1/3\\ 1/3 & 1/3 & 1/3\end{pmatrix}$.

Properties: (i) P is symmetric.

(ii) $P^2 = P$. (P^2b is the projection of Pb which is already on the line. So, $P^2b = Pb$.)

To project b onto a subspace, the ~~problem~~ projection p is now,

$p = A\hat{x} = A(A^TA)^{-1}A^Tb$. (Like before!!)

Here the projection matrix is $P = A(A^TA)^{-1}A^T$.

(We are constructing a $\perp^r$ line from b to $C(A)$.)

Theorem: (i) $P^2 = P$, (ii) $P^T = P$.

Conversely, any symm. matrix with $P^2 = P$ represents a projection.

Proof: (i) is easy.

(ii) $P^T = (A^T)^T ((A^TA)^{-1})^T A^T = A(A^TA)^{-1}A^T = P$.

For the converse, we have to show, if $P^2 = P$ and $P^T = P$ then, Pb is the projection of b onto the col^{m} space of P.

The vector $(b - Pb)$ is $\perp$ to the space.

Let Pc be any vector in the space, then,

$$(b-Pb)^T Pc = b^T(I-P)^T Pc = b^T(P-P^2)c = 0.$$

i.e. $b - Pb \perp$ to the space and Pb is the projection onto the col^{m} space. //