

Mar 16: Final theorems to recap all (Ch 17) 16)
Orthogonality

We have learned a lot about how to think of matrices: as collections of vectors, as coefficients of linear systems, as algebraic objects in their own right — and how all these aspects fit together.

Let's summarize the optimal cases in 3 theorems.

Thm 1: When A is $m \times n$ $\left[\right]$ ($m \geq n$)

Suppose A is an $m \times n$ matrix. Then the following are equivalent:

1) $\text{rank}(A) = n$

2) the columns of A are LI

3) the rows of A span \mathbb{R}^n

4) $\text{Row}(A) = \mathbb{R}^n$

5) $\text{Null}(A) = \{0\}$

6) The only solution to $A\vec{x} = \vec{0}$ is the trivial solution

7) If $A\vec{x} = \vec{b}$ is consistent, then it has a unique solution

8) The RREF of A is $\begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} I_n \\ O_{m-n, n} \end{bmatrix}$

9) $\text{rank}(A^T) = n$

If these hold, then $m \leq n$.

* This is NOT saying that EVERY $m \times n$ matrix has all these properties

Proof:

$\text{rank}(A) = n \iff n$ leading 1s in an RREF of A

\iff a leading 1 in every column of an RREF of A

\iff columns of A are LI (column space alg)

n leading 1s $\iff n$ nonzero rows in a RREF of A

$\iff \dim \text{Row}(A) = n$ (row space algorithm)

$\iff \text{Row}(A) = \mathbb{R}^n$ (since $\text{Row}(A) \subseteq \mathbb{R}^n$)

\iff rows of A span \mathbb{R}^n (definition of row space)

leading 1 in every column \Leftrightarrow RREF is $\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$ & 0s elsewhere

\Leftrightarrow no nonleading variables in the solution to $AX = \vec{0}$, so only trivial solution

\Leftrightarrow $\text{Null}(A) = \{\vec{0}\}$

\Leftrightarrow if $A\vec{x} = \vec{b}$ is consistent, no nonleading variables, so unique solution.

$\text{Row}(A) = \mathbb{R}^n \Leftrightarrow \text{Col}(A^T) = \mathbb{R}^n$ (because rows of A are columns of A^T)
 $\Leftrightarrow \text{rank}(A^T) = n$.

□

Thm 2: When A is $m \times n$ $\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$ ($m \leq n$)

Suppose A is an $m \times n$ matrix. Then the following are equivalent:

1) $\text{rank}(A) = m$

2) $\text{Col}(A) = \mathbb{R}^m$

3) the rows of A are LI

4) the columns of A span \mathbb{R}^m

5) for every choice of $\vec{b} \in \mathbb{R}^m$, the linear system $A\vec{x} = \vec{b}$ is consistent.

$$6) \text{rank}(A^T) = m.$$

If this is all true, then $m \leq n$.

* Again, this theorem is about the optimal case, it is not ALWAYS true.

Proof: $\text{rank}(A) = m \Leftrightarrow \dim \text{Col}(A) = m$
 $\Leftrightarrow \text{Col}(A) = \mathbb{R}^m$ since $\text{Col}(A) \subseteq \mathbb{R}^m$
 \Leftrightarrow Columns of A span \mathbb{R}^m .

$$\Leftrightarrow \dim \text{Row}(A) = m$$

\Leftrightarrow the m rows of A span an m -dimensional space

\Leftrightarrow the rows of A are LI (thm: max LI sets)

$\text{rank}(A) = m \Leftrightarrow$ a leading one in every row of A
 \Leftrightarrow no degenerate equations in an $R \times R$
of $[A|b]$, for any b .

$$\Leftrightarrow \text{rank}(A) = \text{rank}([A|b]) = m \text{ for all } b$$

$\Leftrightarrow Ax = b$ is consistent, for every choice of b .

$$\text{rank}(A) = m \Leftrightarrow \dim \text{Row } A = m$$

$$\Leftrightarrow \dim \text{Col}(A^T) = m \quad (\text{since columns of } A^T = \text{rows of } A)$$

$$\Leftrightarrow \text{rank}(A^T) = m$$

□

eg) A is a 300×200 matrix of rank 100.

so neither theorem holds: $Ax = \vec{0}$ has ∞ many solutions, neither the columns nor rows of A are LI, there are choices of b such that $Ax = b$ is inconsistent, etc.

Remark: $\text{Col}(A) = \{ A\vec{x} \mid \vec{x} \in \mathbb{R}^n \}$ is exactly the set of vectors \vec{b} for which $A\vec{x} = \vec{b}$ is consistent.

We can put all this together into one major thm, when A is $n \times n$ and both conditions hold.

Thm 3: When A is $n \times n$ (square)

Let A be an $n \times n$ matrix. Then the following conditions are equivalent:

1) $\text{rank } A = n$

2) $Ax = 0$ has only the trivial solution

3) $Ax = b$ is consistent \forall every $b \in \mathbb{R}^n$

4) Every linear system $Ax = b$ has a unique solution.

5) The RREF of A is I_n .

6) $\text{Null}(A) = \{ \vec{0} \}$

7) $\text{Col}(A) = \mathbb{R}^n$

8) $\text{Row}(A) = \mathbb{R}^n$

9) $\text{rank}(A^T) = n$

10) the columns of A are LI

11) the rows of A are LI

12) the columns of A span \mathbb{R}^n

13) the rows of A span \mathbb{R}^n

14) the columns of A form a basis for \mathbb{R}^n

15) the rows of A form a basis for \mathbb{R}^n

16) A is invertible

17) A^T is invertible

Recall: A is invertible means that there is a matrix, which we call A^{-1} , which has the property that $AA^{-1} = I$ & $A^{-1}A = I$.

Recall: A^T is the transpose

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad A^{-1} = \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \text{ but } A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

formula, 2×2 case

OK: so what we have gotten out of these chapters is a practical way to implement all our theory about basis and dimension: convert your vectors to vectors in \mathbb{R}^n using ambient coordinates (if necessary), do what you need to do with matrices, then convert back.

Next step: geometry in general vector spaces.

So one key feature of \mathbb{R}^n that we have ignored until now is geometry. In \mathbb{R}^n we have a dot product:

$$\vec{u} = (u_1, \dots, u_n)$$

$$\vec{v} = (v_1, \dots, v_n)$$

$$\vec{u} \cdot \vec{v} = u_1 v_1 + \dots + u_n v_n.$$

Sometimes, when we don't want to confuse the dot product with multiplication, we use the notation

$$\langle \vec{u}, \vec{v} \rangle = \vec{u} \cdot \vec{v}$$

↑ ↑
angle brackets

The dot product tells you the length of a vector and the angle between vectors:

$$\|\vec{u}\| = \sqrt{\langle \vec{u}, \vec{u} \rangle}$$

$$\text{angle } \theta \text{ between } \vec{u} \text{ \& } \vec{v} = \cos^{-1} \left(\frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|} \right)$$

Key point: \vec{u} and \vec{v} are orthogonal if $\langle \vec{u}, \vec{v} \rangle = 0$.

In fact, every vector space has a geometry (a notion of "dot product") but usually the formula has nothing to do with the dot product in \mathbb{R}^n .

For example: in the space of functions which are periodic, with period dividing 2π , the correct "inner product" is:

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x) dx$$

Then if $f(x) = \sin(x)$ you get $\langle f, g \rangle = 0$
 $g(x) = \cos(x)$

but $\langle f, f \rangle = \pi$ and $\langle g, g \rangle = \pi$.

So the trig functions

$\sin(x), \cos(x), \sin(2x), \cos(2x), \dots$

are all orthogonal with respect to this inner product. This is the point of Fourier series (2nd year) EECS.

Another example: when $A, B \in M_{m \times n}$ the inner product is $\langle A, B \rangle = \text{tr}(AB^T)$

\uparrow
"trace" = sum of the entries on the main diagonal.

Orthogonality: We will stick to the dot product here; later on (2nd year, 3rd year) you will do the same calculations with other inner products.

(All the results are true for all inner products!)

Defn: A set of vectors $\{\vec{v}_1, \dots, \vec{v}_m\}$ is called orthogonal if $\vec{v}_i \cdot \vec{v}_j = 0 \in \mathbb{R}$ for every $0 \leq i < j \leq m$ and $\vec{v}_i \neq \vec{0}$ for all $k \leq m$.

eg) $\{(1, 0), (0, 1)\}$ is an orthogonal set in \mathbb{R}^2 .

$\{(2, 0, 0), (0, 1, 1), (0, 1, -1)\}$ is an orthogonal set in \mathbb{R}^3

$\{(1, 1, 0), (1, 0, 1), (0, 0, 1)\}$ is not an orthogonal set in \mathbb{R}^3 because not all pairs of vectors are \perp .

So "orthogonal" feels like "linearly independent" in the sense that orthogonal vectors are definitely pointing in different directions.

Theorem (orthogonal sets are LI)

Let $\{\vec{v}_1, \dots, \vec{v}_k\}$ be an orthogonal set in \mathbb{R}^n .
Then $\{\vec{v}_1, \dots, \vec{v}_k\}$ is LI.

Pf. (This is a very cool trick:)

So suppose we had a dependence equation

$$a_1 \vec{v}_1 + \dots + a_k \vec{v}_k = \vec{0}.$$

Let's take the dot product with \vec{v}_1 :

$$\vec{v}_1 \cdot (a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_k \vec{v}_k) = \vec{v}_1 \cdot \vec{0}$$

$$\Rightarrow a_1 \vec{v}_1 \cdot \vec{v}_1 + a_2 \vec{v}_1 \cdot \vec{v}_2 + \dots + a_k \vec{v}_1 \cdot \vec{v}_k = 0$$

$$\Rightarrow a_1 \|\vec{v}_1\|^2 + a_2 \cdot 0 + \dots + a_k \cdot 0 = 0$$

$$\Rightarrow a_1 \|\vec{v}_1\|^2 = 0.$$

But $\vec{v}_1 \neq \vec{0}$ (orthogonal sets don't contain the zero vector)

$$\therefore a_1 = 0$$

We could do this with $\vec{v}_2, \vec{v}_3, \dots, \vec{v}_k$ and would conclude from each that $a_2 = 0, a_3 = 0, \dots, a_k = 0$.

\therefore the only solution is the trivial solution

$\Rightarrow \{\vec{v}_1, \dots, \vec{v}_k\}$ is the dependence equation

is LI

□

Very cool! Some immediate consequences:

- Every orthogonal set in \mathbb{R}^n has at most n elements

- A basis of \mathbb{R}^n which is an orthogonal set is called an orthogonal basis.
example: our standard basis is an orthogonal basis.

Orthogonal bases are the best:

Thm (expansion thm for coordinates relative to an orthogonal basis)

Suppose $B = \{\vec{v}_1, \dots, \vec{v}_n\}$ is an orthogonal basis for \mathbb{R}^n . Then for any vector \vec{v} in \mathbb{R}^n ,

we have:

$$\vec{v} = \frac{\vec{v} \cdot \vec{v}_1}{\|\vec{v}_1\|^2} \vec{v}_1 + \dots + \frac{\vec{v} \cdot \vec{v}_n}{\|\vec{v}_n\|^2} \vec{v}_n$$

that is, the coordinates of \vec{v} with respect to the basis B are

$$[\vec{v}]_B = \left(\frac{\vec{v} \cdot \vec{v}_1}{\|\vec{v}_1\|^2}, \frac{\vec{v} \cdot \vec{v}_2}{\|\vec{v}_2\|^2}, \dots, \frac{\vec{v} \cdot \vec{v}_n}{\|\vec{v}_n\|^2} \right).$$

In other words: to express \vec{v} as a linear combination of elements of an orthogonal basis, you don't have to do any row reduction — you just calculate some dot products.

* In function spaces for waveforms (periodic fcts), this theorem is the Fourier series expansion of your waveform