Strang

Malcolm

Started 5th June 2025

Contents

Chapter 1

Vectors and Matrices

1.1 Intuition for Dot product, Cosine formula, Schwarz and Triangle inequalities

Intuition for dot product
The unit vectors $\boldsymbol{v} = (\cos\alpha, \sin\alpha)$ and $\boldsymbol{w} = (\cos\beta, \sin\beta)$ are plotted as follows

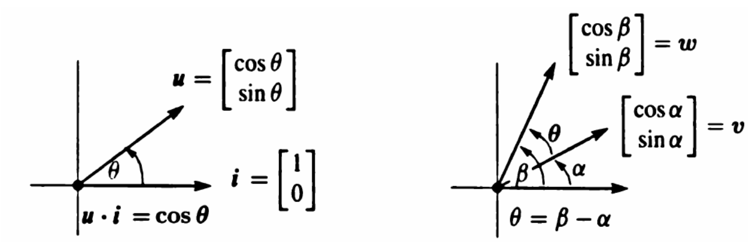


Figure 1.5: Unit vectors: $\boldsymbol{u} \cdot \boldsymbol{i} = \cos\theta$. The angle between the vectors is θ.

See first that when fixed in this form, the magnitude of both vectors is 1, with an angle $\beta - \alpha$ between them. These unit vectors have dot product

$$\boldsymbol{v} \cdot \boldsymbol{w} = \cos\alpha \cos\beta + \sin\alpha \sin\beta = \cos(\beta - \alpha)$$

We have θ as the angle between the two vectors; see that the sign of $\boldsymbol{v} \cdot \boldsymbol{w}$ tells us whether θ is below or above a right angle (due to the cosine function being negative for its argument $> \pi/2$ and positive for $< \pi/2$):

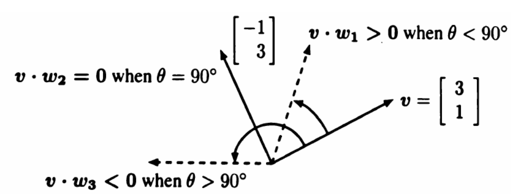


Figure 1.6: Small angle $\boldsymbol{v} \cdot \boldsymbol{w_1} > 0$. Right angle $\boldsymbol{v} \cdot \boldsymbol{w_2} = 0$. Large angle $\boldsymbol{v} \cdot \boldsymbol{w_3} < 0$.

(next page)

Cont.

The idea here is that the dot product reveals the exact angle θ; for unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$, the dot product $\boldsymbol{u} \cdot \boldsymbol{U}$ is the cosine of θ. Ths remains true in n dimensions (not shown).

See that any $\boldsymbol{u}$ and $\boldsymbol{v}$ can be fixed in the above form by normalising their lengths to get $\boldsymbol{u} = \boldsymbol{v}/||\boldsymbol{v}||$ and $\boldsymbol{U} = \boldsymbol{w}/||\boldsymbol{w}||$. After which their dot product would give $\cos\theta$. This leads us to the *cosine formula*:

$$\textbf{Cosine formula}: \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{||\boldsymbol{v}||\,||\boldsymbol{w}||} = \cos\theta \quad \text{if } \boldsymbol{v} \text{ and } \boldsymbol{w} \text{ are nonzero vectors}$$

Perpendicular vectors

See that when the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$ is $90°$, its cosine is 0; this gives us a way to test this. Also see that for perpendicular vectors:

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2$$

because

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = (\boldsymbol{v} + \boldsymbol{w}) \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w}$$

where $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.

Schwarz and Triangle inequalities

First, see from the cosine formula that the dot product of $\boldsymbol{v}/||\boldsymbol{v}||$ and $\boldsymbol{w}/||\boldsymbol{w}||$ never exceeds one (since $\cos\theta$ never exceeds one). This is the the *Schwarz inequality*:

$$\textbf{Schwarz inequality}: |\boldsymbol{v} \cdot \boldsymbol{w}| \leq ||\boldsymbol{v}||\,||\boldsymbol{w}||$$

The *Triangle inequality* comes directly from the Schwarz inequality:

$$\textbf{Triangle inequality}: ||\boldsymbol{v} + \boldsymbol{w}|| \leq ||\boldsymbol{v}|| + ||\boldsymbol{w}||$$

This can be seen from

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w} \leq ||\boldsymbol{v}||^2 + 2||\boldsymbol{v}||\,||\boldsymbol{w}|| + ||\boldsymbol{w}||^2$$

The square root gives us the triangle equality (side 3 cannot exceed side 1 + side 2).

1.2 Intuition for column rank being equal to row rank

If all columns are in the same direction, why does it happen that all the rows are the same direction?

Consider the matrix, see that column 2 is m times column 1:

$$\boldsymbol{A} = \begin{bmatrix} a & ma \\ b & mb \end{bmatrix}$$

See that the second row is just b/a times the first row—if the column rank is 1, then the row rank is 1. See that transposing the matrix, we have

$$\boldsymbol{A} = \begin{bmatrix} a(1) & b(1) \\ a(m) & b(m) \end{bmatrix}$$

which still has one independent column. Now consider the 3x3 case:

$$\boldsymbol{A} = \begin{bmatrix} a & ma & pa \\ b & mb & pb \\ c & mc & pc \end{bmatrix}$$

See that a similar deduction can also be made in this case, where the row rank of A is equal to its column rank.
(next page)

An informal proof
Consider any matrix $\boldsymbol{A}$, suppose we go from left to right, looking for independent columns of $\boldsymbol{A}$ using the following procedure:

1. If column 1 of $\boldsymbol{A}$ is not zero, put it in matrix $\boldsymbol{C}$
2. If column 2 of $\boldsymbol{A}$ is not a multiple of column 1, put it in into $\boldsymbol{C}$
3. If column 3 of $\boldsymbol{A}$ is not a combination of columns 1 and 2, put it into C. *continue*

See that at the end $\boldsymbol{C}$ will have r columns taken from $\boldsymbol{A}$, where r is the rank of $\boldsymbol{A}$ and $\boldsymbol{C}$. While the n columns of $\boldsymbol{A}$ are dependent, the r columns of $\boldsymbol{C}$ will surely be independent.

For instance consider $\boldsymbol{A}$ with rank 2

$$\boldsymbol{A} = \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} \quad \text{leads to} \quad \boldsymbol{C} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix}$$

Now consider another matrix $\boldsymbol{R}$ to be multiplied by $\boldsymbol{C}$ such that $\boldsymbol{A} = \boldsymbol{CR}$. The first and third columns of $\boldsymbol{A}$ are already in $\boldsymbol{C}$, so those respective columns in $\boldsymbol{R}$ make up a *identity matrix*; the second column of $\boldsymbol{A}$ is a multiple of the first, so we have

$$\boldsymbol{A} = \boldsymbol{CR} \quad \text{is} \quad \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(See that the ith row of $\boldsymbol{A}$ can be seen as a linear combination of the rows of $\boldsymbol{R}$ specified the ith row of $\boldsymbol{C}$. (or just consider $\boldsymbol{A}^T = \boldsymbol{R}^T\boldsymbol{C}^T$). We know that

1. $\boldsymbol{C}$ contains the full set of r independent columns of $\boldsymbol{A}$.
2. $\boldsymbol{R} = [\boldsymbol{I}\ \boldsymbol{F}]$ contains the identity matrix $\boldsymbol{I}$ in the same r columns that held $\boldsymbol{C}$.
3. The dependent columns of $\boldsymbol{A}$ are combinations of $\boldsymbol{CF}$ of the independent columns in $\boldsymbol{C}$.

Where the matrix $\boldsymbol{F}$ goes into the other $n-r$ columns of $\boldsymbol{R} = [\boldsymbol{I}\ \boldsymbol{F}]$. ($\boldsymbol{A} = \boldsymbol{CR}$ becomes $\boldsymbol{A} = \boldsymbol{C}[\boldsymbol{I}, \boldsymbol{F}] = [\boldsymbol{C}, \boldsymbol{CF}] = [\text{indep cols of } \boldsymbol{A},\ \text{dep cols of } \boldsymbol{A}]$ (in correct order).

See that $\boldsymbol{C}$ has the same column space as $\boldsymbol{A}$, and $\boldsymbol{R}$ *has the same row space* as $\boldsymbol{A}$ (every row of $\boldsymbol{A}$ is a combination of the rows of $\boldsymbol{R}$).
(next page)

Cont.
We had the example

$$\boldsymbol{A} = \boldsymbol{CR} \quad \text{is} \quad \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here is an informal proof that the row rank of $\boldsymbol{A}$ *equals the column rank of* $\boldsymbol{A}$ (based from facts we already know)

1. The r columns of $\boldsymbol{C}$ are independent (chosen that way from $\boldsymbol{A}$)
2. Every column of $\boldsymbol{A}$ is a combination of those r columns of $\boldsymbol{C}$ (since $\boldsymbol{A} = \boldsymbol{CR}$)
3. The r rows of $\boldsymbol{R}$ are independent (they contain the r by r matrix $\boldsymbol{I}$)
4. Every row of $\boldsymbol{A}$ is a combination of the $\boldsymbol{r}$ rows of $\boldsymbol{R}$

See that for every column of $\boldsymbol{A}$ that goes into $\boldsymbol{C}$, a column of $\boldsymbol{I}$ goes into $\boldsymbol{R}$, where each column of $\boldsymbol{I}$ in $\boldsymbol{R}$ adds an independent row.

This means that the column rank of $\boldsymbol{C}$ (column space of $\boldsymbol{A}$) is always equal to the row rank of $\boldsymbol{R}$ (row space of $\boldsymbol{A}$)—the column rank of $\boldsymbol{A}$ is equal to the row rank of $\boldsymbol{A}$.

More examples
Rank 2:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Rank 2:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Rank 1:

$$\begin{bmatrix} 1 & 2 & 10 & 100 \\ 3 & 6 & 30 & 300 \\ 2 & 4 & 20 & 200 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 10 & 100 \end{bmatrix}$$

1.3 Ways to multiply $AB = C$

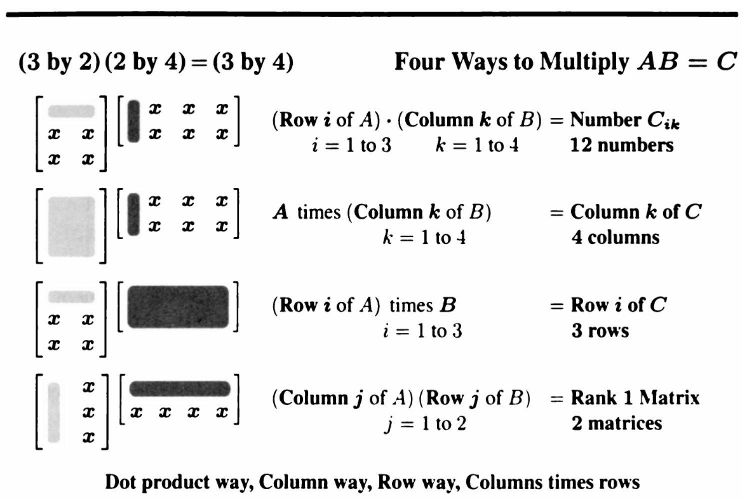
Multiplcation by columns of A and rows of B
A lesser known way to multiply $\boldsymbol{AB}$ is through considering the columns of $\boldsymbol{A}$ and the rows of $\boldsymbol{B}$ (contrary to the usual ideas where each entry of the result is a dot product of a row of $\boldsymbol{A}$ and column of $\boldsymbol{B}$):

$$\boxed{\begin{array}{c} \boldsymbol{AB} = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \boldsymbol{b}_1^* & - \\ & \vdots & \\ - & \boldsymbol{b}_n^* & - \end{bmatrix} = \boldsymbol{a}_1\boldsymbol{b}_1^* + \boldsymbol{a}_2\boldsymbol{b}_2^* + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^*. \\ \textbf{columns } \boldsymbol{a}_k \qquad \textbf{rows } \boldsymbol{b}_k^* \qquad \textbf{Add columns } \boldsymbol{a}_k \textbf{ times rows } \boldsymbol{b}_k^* \end{array}}$$

We multiply each column of $\boldsymbol{A}$ by each row of $\boldsymbol{B}$; this gives us n rank 1 matrices, which we then sum together; these matrices are called *outer products*

(usually we see the ith column of the result as a linear combination of the columns of $\boldsymbol{A}$ specified by the ith column of $\boldsymbol{B}$. However in this case the ith outer product is a matrix of the same size as the result, that contains all the contributions of the ith column of $\boldsymbol{A}$ to the final product. By summing this over all n columns of $\boldsymbol{A}$ we get the result.)

Summary of methods



(A nice way to intuit the the third method is to consider $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$; where the columns of $(\boldsymbol{AB})^T$ are the rows of $\boldsymbol{AB}$)

1.4 Solutions to $\boldsymbol{Ax} = \boldsymbol{b}$

Given a nxn matrix $\boldsymbol{A}$ and an nx1 column vector b, there are three outcomes for the vector $\boldsymbol{x}$ that solves $\boldsymbol{Ax} = \boldsymbol{b}$.

First there may be *no vector* $\boldsymbol{x}$ that solves $\boldsymbol{Ax} = \boldsymbol{b}$, or there may be exactly *one* solution, or there may be *infinitely many* solution vectors $\boldsymbol{x}$. Here are the possibilities:

1. **Exactly one solution** to $\boldsymbol{Ax} = \boldsymbol{b}$ means that $\boldsymbol{A}$ has independent columns (only one particular linear combination of the columns of $\boldsymbol{A}$ leads to $\boldsymbol{b}$. That combination is specified by $\boldsymbol{x}$). $\boldsymbol{A}$ is full rank and the only solution to $\boldsymbol{Ax} = \boldsymbol{0}$ is $\boldsymbol{x} = \boldsymbol{0}$. $\boldsymbol{A}$ *has an inverse matrix* $\boldsymbol{A}^{-1}$ (given $\boldsymbol{b}$, we can work backward to get $\boldsymbol{x}$ since only one $\boldsymbol{x}$ leads to $\boldsymbol{b}$).

2. **No solution** to $\boldsymbol{Ax} = \boldsymbol{b}$ means that $\boldsymbol{b}$ is not in the column space of $\boldsymbol{A}$, so $\boldsymbol{A}$ is not full rank.

3. **Infinitely many solutions**. See that when the columns of $\boldsymbol{A}$ are not independent (not full rank), then there are infinitely many ways to produce the zero vector $\boldsymbol{b} = \boldsymbol{0}$ (this is the meaning of dependent columns), and so there are infinitely many solutions to $\boldsymbol{AX} = \boldsymbol{0}$.

Also see that if $\boldsymbol{A}$ is not full rank it means that its column space is some subspace, where solutions only exist for $\boldsymbol{b}$ within that subspace.

As such, if there so happens to be a solution to $\boldsymbol{Ax} = \boldsymbol{b}$ then we can add any solution to $\boldsymbol{AX} = \boldsymbol{0}$:

$$\boldsymbol{A}(\boldsymbol{x} + \alpha\boldsymbol{X}) = \boldsymbol{Ax} + \alpha\boldsymbol{AX} = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}$$

For some constant α, which gives us $\boldsymbol{b}$ again—we have infinitely many solutions.

1.5 Elimination by elimination matrices