Associativity of matrix multiplication:
$(AB)C = A(BC)$
for any conformable matrices A, B & C

Step 1. Will prove the key vector identity:

For every vector $\vec{c}$, $(AB)\vec{c} = A(B\vec{c})$

- let $B = [\ \vec{b}1 \mid \vec{b}2 \mid \vec{b}3\ ]$

- let $\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$

Using matrix-vector multiplication definition:
- $B\vec{c} = \vec{b}1c_1 + \vec{b}2c_2 + \vec{b}3c_3$
- $A(B\vec{c}) = A\vec{b}1c_1 + A\vec{b}2c_2 + A\vec{b}3c_3$

Using matrix-matrix multiplication definition:
- $AB = [A\vec{b}1 \mid A\vec{b}2 \mid A\vec{b}3]$

Using matrix–vector multiplication definition:

$$(AB)\vec{c} = \left[\begin{array}{c|c|c} A\vec{b}1 & A\vec{b}2 & A\vec{b}3 \end{array}\right] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} =$$

$$= A\vec{b}1c_1 + A\vec{b}2c_2 + A\vec{b}3c_3 = A(B\vec{c})$$

Step 2. Extend from a single vector $\vec{c}$ to matrix C

Write C by its columns:

$$C = [\ \vec{c}1 \mid \vec{c}2 \mid \vec{c}3\ ]$$

Using column-wise matrix-matrix multiplication definition:

$$(AB)C = \left[\ (AB)\vec{c}1 \mid (AB)\vec{c}2 \mid (AB)\vec{c}3\ \right]$$

$$BC = [\ B\vec{c}1 \mid B\vec{c}2 \mid B\vec{c}3\ ]$$

$$A(BC) = \left[\ A(B\vec{c}1) \mid A(B\vec{c}2) \mid A(B\vec{c}3)\ \right]$$

Now apply Step 1 to each column $\vec{c}k$ ($k = 1,2,3$):

$$(AB)\vec{c}1 = A(B\vec{c}1)$$

$$(AB)\vec{c}2 = A(B\vec{c}2)$$

$$(AB)\vec{c}3 = A(B\vec{c}3)$$

So the corresponding columns in (AB)C and A(BC) are equal

Therefore:

$$(AB)C = A(BC)$$

This proof is written with three columns for simplicity;
the same logic applies to any number of columns

Other algebraic properties of matrix multiplication

① Identity

$$A(\vec{b} + \vec{c}) = A\,\vec{b} + A\,\vec{c}$$

was demonstrated in 'First steps' chapter

From this one, we can derive
② $A(B + C) = AB + AC$:

If $B = [\, \vec{b}_1 \mid \vec{b}_2 \,]$ and $C = [\, \vec{c}_1 \mid \vec{c}_2 \,]$,
- $(B + C) = [\, \vec{b}_1 + \vec{c}_1 \mid \vec{b}_2 + \vec{c}_2 \,]$

- $A\,(B + C) = A\,[\, \vec{b}_1 + \vec{c}_1 \mid \vec{b}_2 + \vec{c}_2 \,]$

$$= \left[\, A\,(\vec{b}_1 + \vec{c}_1) \mid A\,(\vec{b}_2 + \vec{c}_2) \,\right]$$
$$= [\, A\,\vec{b}_1 + A\,\vec{c}_1 \mid A\,\vec{b}_2 + A\,\vec{c}_2 \,]$$
$$= [\, A\,\vec{b}_1 \mid A\,\vec{b}_2 \,] + [\, A\,\vec{c}_1 \mid A\,\vec{c}_2 \,]$$
$$= AB + AC$$

③ Scaling identity for a scalar s
$$A\,(s\vec{b}) = s\,(A\,\vec{b})$$
was also demonstrated on the same page
This also can be generalized to
④ $A(sB) = s(AB)$
If $B = [\, \vec{b}_1 \mid \vec{b}_2 \,]$
- $sB = [\, s\,\vec{b}_1 \mid s\,\vec{b}_2 \,]$

- $A(sB) = A[\, s\,\vec{b}_1 \mid s\,\vec{b}_2 \,]$

$$= \left[\, A(s\,\vec{b}_1) \mid A(s\,\vec{b}_2) \,\right]$$
$$= \left[\, s(A\,\vec{b}_1) \mid s(A\,\vec{b}_2) \,\right]$$
$$= s[\, A\,\vec{b}_1 \mid A\,\vec{b}_2 \,]$$
$$= s(AB)$$

Recall that AB was defined by applying A to each column of B

Writing B as
$$B = [\, \vec{b}_1 \mid \vec{b}_2 \mid \dots \,]$$
applies the same definition and is used as a standard proof technique

⑤ Identity matrix

Recall the canonical basis vectors

$\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$

which span $\mathbb{R}^n$

We define the (always square) identity matrix as:

$I_n = [\ \vec{e}_1 \mid \vec{e}_2 \mid \ldots \mid \vec{e}_n\ ]$

or: ordered set of canonical basis vectorsIf A is m×n, then

$A\, I_n = A\, [\ \vec{e}_1 \mid \ldots \mid \vec{e}_n\ ]$

$= [\ A\,\vec{e}_1 \mid \ldots \mid A\,\vec{e}_n\ ]$

$= A$

- I_n acts on the domain of A:
  $A\, I_n = A$ (right-identity)
- I_m acts on the codomain of A:
  $I_m\, A = A$ (left-identity)

Matrix multiplication as
composition of linear transformations

Let B be an n×p matrix and A be an m×n matrix
Define two linear transformations:

$T\_B : \mathbb{R}^p \rightarrow \mathbb{R}^n$ by $T\_B(\vec{c}) = B\vec{c}$

$T\_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ by $T\_A(\vec{u}) = A\vec{u}$

Then applying B first and then A gives the composed transformation:

$$(T\_A \circ T\_B)(\vec{c}) = T\_A\big(T\_B(\vec{c})\big) = A(B\vec{c})$$

Now recall the key identity (proved in the associativity section):

$(AB)\vec{c} = A(B\vec{c})$ for every vector $\vec{c}$

Therefore, for every $\vec{c} \in \mathbb{R}^p$:

$(AB)\vec{c} = (T\_A \circ T\_B)(\vec{c})$

Conclusion:

The product matrix AB is exactly the single matrix that represents the composition "apply B, then apply A"

Order importance:

AB corresponds to "B then A", while BA corresponds to "A then B"
These are different compositions in general, so AB ≠ BA in general

Size importance: A acts on the output of B:

- Transformation by an $n \times p$ matrix B inputs p-vectors & outputs n-vectors
- Transformation by an $m \times n$ matrix A inputs n-vectors & outputs m-vectors

$\Downarrow$

For AB to be defined, domain of A must equal codomain of B
(equivalently: A must have n columns)

- The product C = AB inputs p-vectors and outputs m-vectors

$\Downarrow$

Outer sizes p and m propagate, inner size n vanishes

Just as multiplication AB proceeds right-to-left, input and output sizes are read right-to-left

Geometric interpretation (in 2D or 3D):

Applying B transforms the grid first; applying A transforms the result
The single matrix AB produces the same final transformed grid

Consider example:

$$\text{Matrix A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{(shear)}$$

$$\text{Matrix B} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{(scaling)}$$

Image 1: apply B, then apply A (AB)

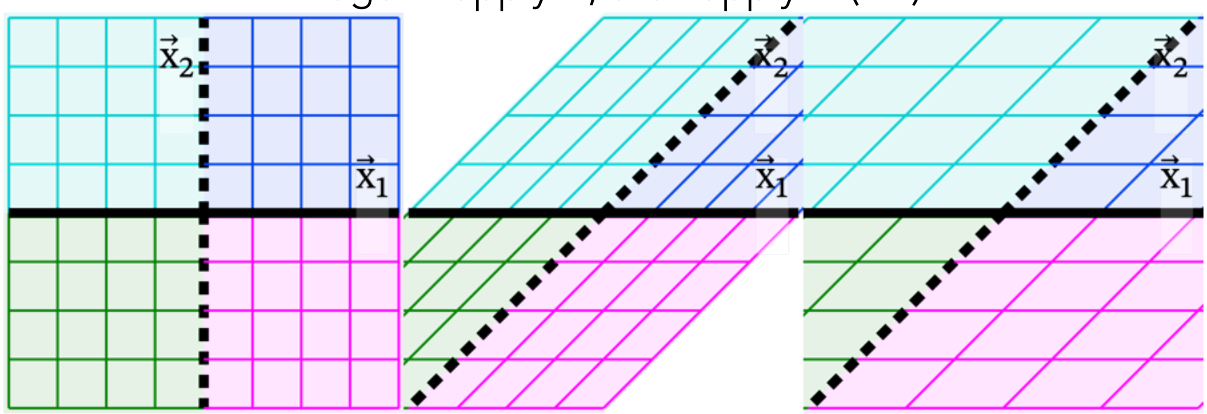
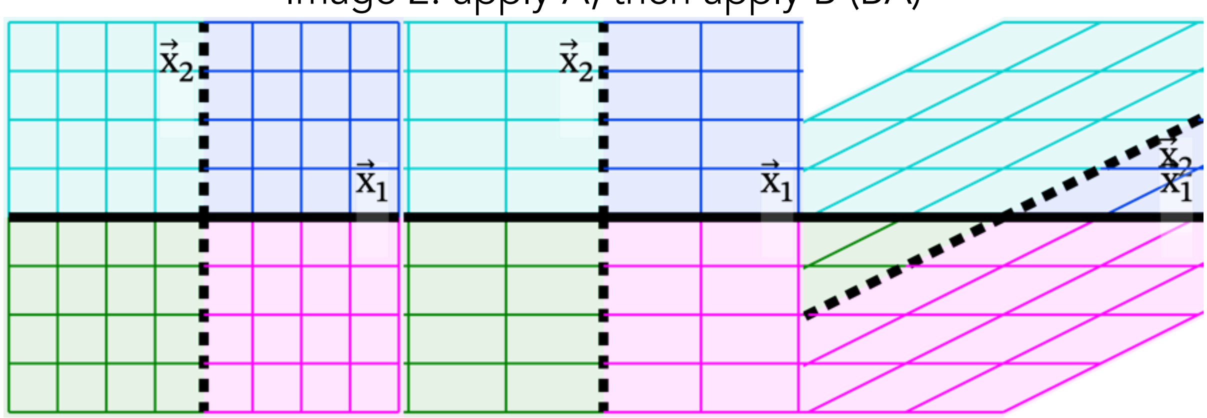


Image 2: apply A, then apply B (BA)



Matrix multiplication represents composition of linear transformations:
AB means “apply B, then apply A”

Composition of linear transformations vs
composition of algebraic functions

Algebraic functions	Matrix transformations
Start with a value x	Start with a vector $\vec{x}$
Apply g first	Apply B first
Result: g(x)	Result: $B\vec{x}$
Apply f next	Apply A next
Result: f(g(x))	Result: $A(B\vec{x})$
Written as composition $f \circ g$	Written as product AB
Order is right-to-left	Order is right-to-left
In general $f \circ g \neq g \circ f$	In general $AB \neq BA$
Domain of f must match range of g	Columns of A must match rows of B
Single function represents the composition	Single matrix represents the composition

---

Transpose of matrix product $(AB)^T$

$$(AB)^T = B^T A^T$$

Proof:

Recall the row–column rule for matrix multiplication:

each entry c_{ij} of $C = AB$ is the dot product of row i of A with column j of B:

$$(AB)_{ij} = (\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } B)$$

$$\Leftrightarrow$$

$$(AB)_{ij} = (\text{column } i \text{ of } A^T) \cdot (\text{row } j \text{ of } B^T)$$

$$\Leftrightarrow$$

$$(AB)^T_{ij} = (\text{column } j \text{ of } A^T) \cdot (\text{row } i \text{ of } B^T)$$

clarification: $(AB)^T_{ij} = (AB)_{ji}$

$$\Leftrightarrow$$

$$(AB)^T_{ij} = (\text{row } i \text{ of } B^T) \cdot (\text{column } j \text{ of } A^T)$$

$$= B^T \times A^T$$

Consider the following example of compatible A & B:

$$A = \left[\begin{array}{c|c|c} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{array}\right] = \begin{bmatrix} (\vec{a}_1)^T \\ (\vec{a}_2)^T \\ (\vec{a}_3)^T \end{bmatrix}$$

$$B = \left[\begin{array}{c|c|c} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{array}\right] = \begin{bmatrix} (\vec{b}_1)^\top \\ (\vec{b}_2)^\top \\ (\vec{b}_3)^\top \end{bmatrix}$$

$$A^\top = \left[\begin{array}{c|c|c} (\vec{a}_1)^\top & (\vec{a}_2)^\top & (\vec{a}_3)^\top \end{array}\right] \quad B^\top = \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vec{b}_3 \end{bmatrix}$$

$$A \times B = \begin{bmatrix} (\vec{a}_1)^\top \\ (\vec{a}_2)^\top \\ (\vec{a}_3)^\top \end{bmatrix} \left[\begin{array}{c|c|c} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{array}\right]$$

$$B^\top \times A^\top = \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vec{b}_3 \end{bmatrix} \left[\begin{array}{c|c|c} (\vec{a}_1)^\top & (\vec{a}_2)^\top & (\vec{a}_3)^\top \end{array}\right]$$

Applying the entry-wise multiplication rule, we see that each entry of $(AB)^T$ equals the corresponding entry of B^TA^T Thus, the two products satisfy identical equations and

$(AB)^T = B^TA^T$

Transpose and dot product

Adjoint identity

$(A\,\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A^T\,\vec{w})$

is derived from transpose multiplication rule

$(AB)^T = B^T A^T$

If A is m×n, $\vec{v} \in \mathbb{R}^n$, $\vec{w} \in \mathbb{R}^m$

Dot product can be written using transpose:

$\vec{u} \cdot \vec{w} = (\vec{u})^T\,\vec{w}$

Then

$(A\,\vec{v}) \cdot \vec{w}$

$= (A\,\vec{v})^T\,\vec{w}$

$= (\vec{v})^T\,A^T\,\vec{w}$

$= \vec{v} \cdot (A^T\,\vec{w})$

This can be intuitively explained as

- we compute $(A\,\vec{v})$ and measure its component along $\vec{w}$
- we compute $(A^T\,\vec{w})$ and measure its component along $\vec{v}$
- we get identical scalars

Will illustrate this with an $\mathbb{R}^2$ example:

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Image 1 shows transformation of $\vec{v}$ by A

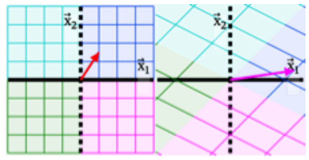
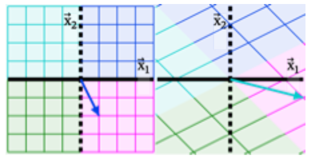


Image 2 shows transformation of $\vec{w}$ by A^T

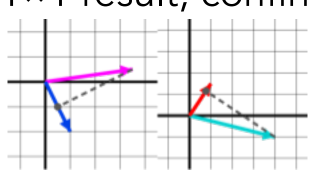


$$A\,\vec{v} = \begin{bmatrix} 3.5 \\ 0.5 \end{bmatrix} \quad A^T\,\vec{w} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$(A\,\vec{v}) \cdot \vec{w} = (A\,\vec{v})^T\,\vec{w} \begin{bmatrix} 3.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2.5 \end{bmatrix}$$

$$\vec{v} \cdot (A^T\,\vec{w}) = \vec{v}^T\,(A^T\,\vec{w}) \begin{bmatrix} 1 & 1.5 \end{bmatrix} \begin{bmatrix} 4 \\ -1 \end{bmatrix} = \begin{bmatrix} 2.5 \end{bmatrix}$$

Compare $(A\,\vec{v})^T\,\vec{w}$ and $\vec{v}^T\,(A^T\,\vec{w})$:
both give the same 1×1 result, confirming adjoint identity



- $(A\,\vec{v})$: magenta
- $\vec{w}$: blue
- $(A^T\,\vec{w})$: cyan

• $\vec{v}$: red

The scalar 2.5 is not a geometric length
It is a signed directional measurement
For nonzero $\vec{b}$, the dot product $\vec{a} \cdot \vec{b}$ equals
the projection length of $\vec{a}$ onto the direction of $\vec{b}$, multiplied by $|\vec{b}|$
Only when $|\vec{b}| = 1$ does the dot product equal the projection length itself

---

Matrix multiplication: what was defined and what followed

In this chapter, we treated matrix multiplication as a structure built from matrix–vector multiplication

Definitions used

• $A\vec{b}$ is defined as a linear combination of columns of A
• AB is defined by applying A to each column of B

From these definitions we derived

• the entry-wise rule: $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$
• the row-wise viewpoint
• the outer-product decomposition
• associativity: (AB)C = A(BC)
• distributivity, identity matrix, non-commutativity

We also used these viewpoints to interpret

- AB as composition of linear transformations
- size compatibility of products
- block multiplication and transpose of a product
- structured examples (diagonal, orthonormal, upper triangular)

Dimensional consequences of multiplication
(rank and null spaces) are discussed in the Subspaces chapter