From: column-wise multiplication
To: element-wise multiplication

① Axiom 1: matrix-vector multiplication $A\vec{b}$ is defined as linear combination of columns of A with coefficients from $\vec{b}$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} =$$

$$b_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + b_3 \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

(this also shows why the number of columns of A must equal the number of entries of $\vec{b}$)

② Axiom 2: definition of matrix-matrix multiplication AB:

If $B = \left[\, \vec{b}_1 \mid \dots \mid \vec{b}_p \,\right]$,

$$AB = \left[\, A\vec{b}_1 \mid \dots \mid A\vec{b}_p \,\right]$$

Note that the axiom 2 is designed to produce the following result for compatible matrices A & B and vector $\vec{v}$:
$(AB)\vec{v} = A(B\vec{v})$, which is covered later in this chapter

In other words: multiplying A and B first and then applying the result to $\vec{v}$ must give the same outcome as first applying B to $\vec{v}$ and then applying A to the result

The concept is illustrated for a 2×3 A and a 3×2 B,
but is true for any m×n A and any n×p B

Suppose $C = AB = \left[\, A\vec{b}_1 \mid \dots \mid A\vec{b}_p \,\right] = \left[\, \vec{c}_1 \mid \dots \mid \vec{c}_p \,\right]$

Expanding this equality gives us

$$C = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}, \text{ where}$$

- $\vec{c}_1 = b_{11}\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{21}\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + b_{31}\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$

- $\vec{c}_2 = b_{12}\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{22}\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + b_{32}\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$

From here follows

$$C = \begin{bmatrix} a_{11}\times b_{11}+a_{12}\times b_{21}+a_{13}\times b_{31} & a_{11}\times b_{12}+a_{12}\times b_{22}+a_{13}\times b_{32} \\ a_{21}\times b_{11}+a_{22}\times b_{21}+a_{23}\times b_{31} & a_{21}\times b_{12}+a_{22}\times b_{22}+a_{23}\times b_{32} \end{bmatrix}$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Equivalently, each entry c_{ij} is the dot product of row i of A with column j of B

The highlighted row and column below show how one entry of C is formed:

This is known as the entry-wise multiplication rule,
often taken as a definition and universally used for computations

From: element-wise multiplication
To: row-wise multiplication

If any step of this page is not obvious, you can verify it by
expanding entries with the entry-wise multiplication rule in the reverse direction

Will start with entry-wise multiplication rule
that was derived earlier from column-wise multiplication:

$$C = AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11} \times b_{11} + a_{12} \times b_{21} + a_{13} \times b_{31} & a_{11} \times b_{12} + a_{12} \times b_{22} + a_{13} \times b_{32} \\ a_{21} \times b_{11} + a_{22} \times b_{21} + a_{23} \times b_{31} & a_{21} \times b_{12} + a_{22} \times b_{22} + a_{23} \times b_{32} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}\begin{bmatrix} b_{11} & b_{12} \end{bmatrix} + a_{12}\begin{bmatrix} b_{21} & b_{22} \end{bmatrix} + a_{13}\begin{bmatrix} b_{31} & b_{32} \end{bmatrix} \\ \\ a_{21}\begin{bmatrix} b_{11} & b_{12} \end{bmatrix} + a_{22}\begin{bmatrix} b_{21} & b_{22} \end{bmatrix} + a_{23}\begin{bmatrix} b_{31} & b_{32} \end{bmatrix} \end{bmatrix}$$

The above equation

- shows each row of C as a linear combination of rows of B with coefficients from A
- can be rewritten as

$$\begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \\ \\ \begin{bmatrix} a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \end{bmatrix}$$

or, equivalently:

$$\begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \times B \\ \\ \begin{bmatrix} a_{21} & a_{22} & a_{23} \end{bmatrix} \times B \end{bmatrix}$$

$$\Leftrightarrow$$

$$\text{row}_i(C) = \text{row}_i(A) \times B$$

From: element-wise multiplication
To: sum of outer products

Will start with entry-wise multiplication rule
that was derived earlier from column-wise multiplication:

$$C = AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}\times b_{11}+a_{12}\times b_{21}+a_{13}\times b_{31} & a_{11}\times b_{12}+a_{12}\times b_{22}+a_{13}\times b_{32} \\ a_{21}\times b_{11}+a_{22}\times b_{21}+a_{23}\times b_{31} & a_{21}\times b_{12}+a_{22}\times b_{22}+a_{23}\times b_{32} \end{bmatrix} =$$

Above matrix can be presented as a sum of n=3 matrices:
(n = column count of A and row count of B)

$$\begin{bmatrix} a_{11}\times b_{11} & a_{11}\times b_{12} \\ a_{21}\times b_{11} & a_{21}\times b_{12} \end{bmatrix} + \begin{bmatrix} a_{12}\times b_{21} & a_{12}\times b_{22} \\ a_{22}\times b_{21} & a_{22}\times b_{22} \end{bmatrix} + \begin{bmatrix} a_{13}\times b_{31} & a_{13}\times b_{32} \\ a_{23}\times b_{31} & a_{23}\times b_{32} \end{bmatrix}$$

Each of the matrices represents a product of column k of A and row k of B
also called outer product:

$$\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \end{bmatrix} + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \begin{bmatrix} b_{21} & b_{22} \end{bmatrix} + \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix} \begin{bmatrix} b_{31} & b_{32} \end{bmatrix}$$

Each outer product is formed from a column vector and a row vector,
and the result is a rank-1 matrix

If the last step is not obvious, you can verify it by
expanding entries with the entry-wise multiplication rule in the reverse direction

---

The 4 ways of matrix-matrix multiplication

① Column-wise multiplication:

This is the fundamental definition and key to understanding linear transformations
Each column of C is a linear combination of the columns of A,
using the entries of the corresponding column of B
as the coefficients

Geometrically, this represents how each vector of B is directed by A

② Entry-wise multiplication:

This is the rule used for actual computations.
It is derived directly from the column-wise definition
by focusing on a single scalar result at a time.

To find the entry in row i and column j of C,
we take the dot product of row i from A
and column j from B.

Mathematically, this is expressed as:

$$c_{ij} = \sum_{k=1}^{n} \left(a_{ik} b_{kj} \right)$$

③ Row-wise multiplication:

Each row of the result matrix C is a linear combination of the rows of B, using the entries of the corresponding row of A as the coefficients

Geometrically, rows of A represent linear constraints: lines, planes or hyperplanes, depending on dimension
Rows of C represent projection of rows of B on those constraints

④ Sum of outer products:

Matrix multiplication can be decomposed as a sum of rank-1 matrices:

$$C = AB = \vec{a}_1\vec{b}_1^{\mathsf{T}} + \vec{a}_2\vec{b}_2^{\mathsf{T}} + \dots + \vec{a}_n\vec{b}_n^{\mathsf{T}}$$

where $\vec{a}_k$ is column k of A and $\vec{b}_k^{\mathsf{T}}$ is row k of B
(n is the column count of A and the row count of B)

Each term $\vec{a}_k\vec{b}_k^{\mathsf{T}}$ is an outer product:

- it is a rank-1 matrix (unless one factor is zero)
- it maps all vectors into a single direction in the output space (rank 1)

Why this viewpoint matters:

- It shows that C is built by adding simple rank-1 layers
- If only a few such terms dominate, C is well-approximated by a low-rank matrix
- This idea leads directly to Singular Value Decomposition (SVD), which orders rank-1 contributions by importance

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:

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

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

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. Distributivity over matrix addition:

For conformable matrices A, B, C:

$$A(B + C) = AB + AC$$
$$(A + B)C = AC + BC$$

Justification:
Matrix–matrix multiplication is defined column-wise, and matrix–vector multiplication distributes over vector addition.

2. Distributivity over vector addition:

For any matrix A and vectors $\vec{c}_1$, $\vec{c}_2$:

$$A(\vec{c}_1 + \vec{c}_2) = A\,\vec{c}_1 + A\,\vec{c}_2$$

Justification:
This follows directly from the definition of matrix–vector multiplication as a linear combination of columns.

3. Compatibility with scalar multiplication:

For any scalar c_1, matrix A, and vector $\vec{c}_1$:

$$A(c_1\,\vec{c}_1) = c_1\,(A\,\vec{c}_1)$$
$$(c_1\,A)\,\vec{c}_1 = c_1\,(A\,\vec{c}_1)$$

Justification:
Scalar–vector and scalar–matrix multiplication are defined so that scalar factors pull out of linear combinations.

4. Associativity:

For conformable matrices A, B, C:
$(AB)C = A(BC)$

Justification:
Previously proven using only column-wise definitions.

5. Identity matrix:

There exists a matrix I such that:
$AI = A$ and $IA = A$

Justification:
The identity matrix leaves each column unchanged under matrix–vector multiplication.

6. Non-commutativity:

In general:
$AB \neq BA$

Justification:
Matrix multiplication represents composition of linear transformations; composition is not commutative in general.

7. Zero matrix:

For the zero matrix 0:
$A0 = 0$ and $0A = 0$

Justification:
All columns of the zero matrix are zero vectors, and linear combinations

of zero vectors remain zero.

All properties above follow directly from the column-wise definitions of matrix–vector and matrix–matrix multiplication.

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:

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

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

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 $AB \neq BA$ in general

Size importance: A acts on the output of B:

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

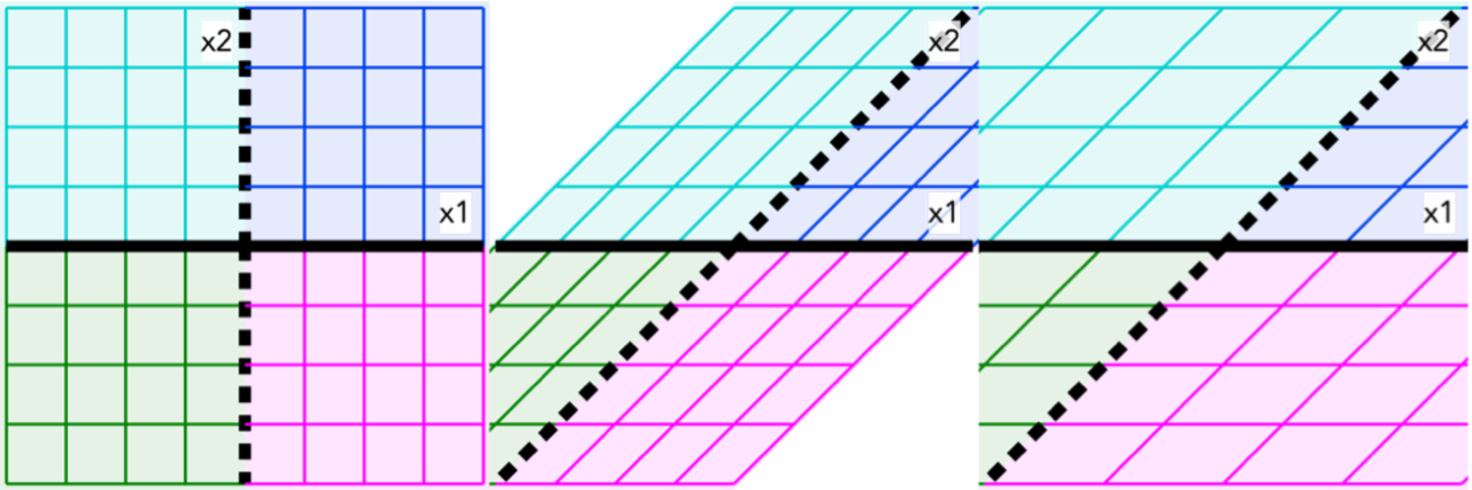
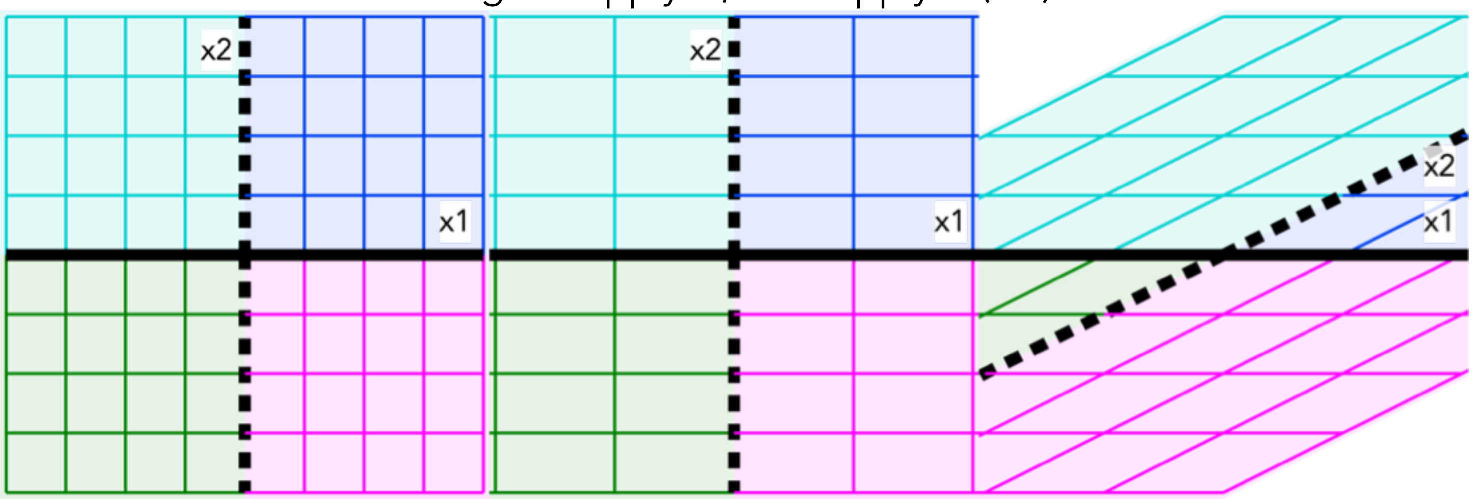


Image 2: apply A, then apply B (BA)



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

Composition of linear transformations vs composition of algebraic functions

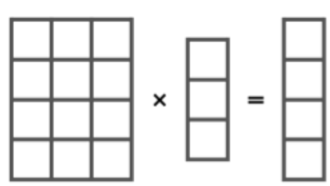
Algebraic functions	Matrix transformations
Start with a value x	Start with a vector $\vec{x}$
Apply g first	Apply B first
Result: g(x)	Result: $B\vec{x}$
Apply f next	Apply A next
Result: f(g(x))	Result: $A(B\vec{x})$
Written as composition $f \circ g$	Written as product AB
Order is right-to-left	Order 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

---

Matrix product shapes

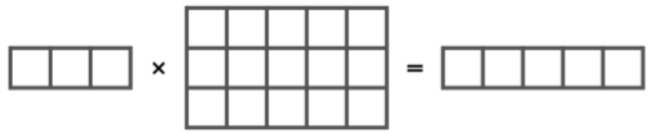
① Matrix × vector

$A\vec{x}$



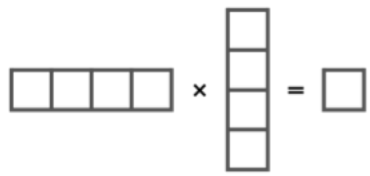
② Row vector × matrix

$\vec{y}^{\mathsf{T}}A$



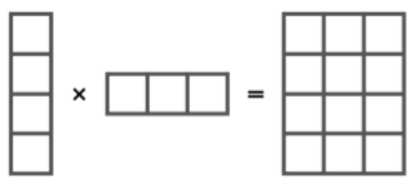
③ Dot product as a matrix product

$\vec{x}^{\mathsf{T}}\vec{y}$



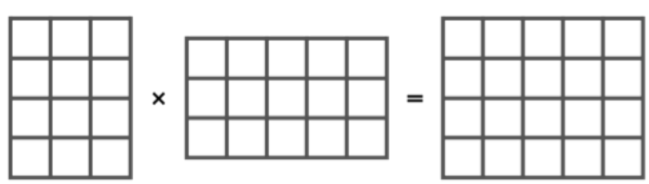
④ Outer product or rank-1 matrix

$\vec{x}\vec{y}^{\mathsf{T}}$



⑤ Matrix × matrix

AB



Block multiplication, 2×2 form

Any matrix A can be partitioned into blocks as shown below:

$$A = \left[\begin{array}{c|c} A11 & A12 \\ \hline A21 & A22 \end{array}\right]$$

If matrix B is partitioned in a similar way and with compatible block sizes,

$$AB = \left[\begin{array}{c|c} A11\times B11+A12\times B21 & A11\times B12+A12\times B22 \\ \hline A21\times B11+A22\times B21 & A21\times B12+A22\times B22 \end{array}\right]$$

Motivation:

Block matrix multiplication lets us treat a large matrix as a collection of smaller matrices

This is especially useful when some blocks have special structure:
- Identity blocks I, which leave other blocks unchanged
- Diagonal blocks D, which scale rows or columns
- Zero blocks, which eliminate entire terms

In these cases, multiplication is performed without expanding all entries

Block multiplication is also useful in proofs & decompositions

Partitioning rules:

1. Just as matrix A can have any number of rows,
it can be freely partitioned in the horizontal direction

2. Just as matrix B can have any number of columns,
it can be freely partitioned in the vertical direction

3. Partitioning is compatible if the corresponding block dimensions match,
so that each block product is defined
(analogous to full matrices A and B: their inner dimensions before partitioning must also match)

Partition mode (block sizes):

Assume we split A into 2×2 blocks by cutting its rows into $m1$ and $m2$,
and its columns into $n1$ and $n2$:
A is $(m1+m2) \times (n1+n2)$.

Similarly, split B into 2×2 blocks by cutting its rows into $n1$ and $n2$,
and its columns into $p1$ and $p2$:
B is $(n1+n2) \times (p1+p2)$.

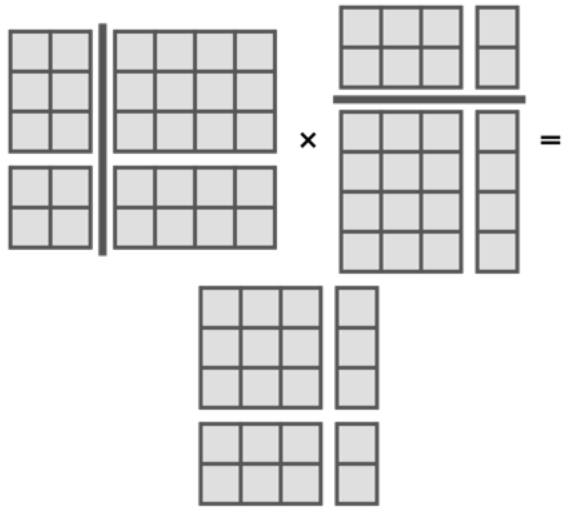
The inner dimensions of the full matrices match: $(n1+n2) = (n1+n2)$,
and the block cuts use the same $n1$ and $n2$, so each block product is defined.

Block dimensions:
$A11$ is $m1 \times n1$, $A12$ is $m1 \times n2$
$A21$ is $m2 \times n1$, $A22$ is $m2 \times n2$
$B11$ is $n1 \times p1$, $B12$ is $n1 \times p2$
$B21$ is $n2 \times p1$, $B22$ is $n2 \times p2$

Therefore, addition is valid for all pairs:

(A11)(B11) is m1×p1 and (A12)(B21) is m1×p1
(A11)(B12) is m1×p2 and (A12)(B22) is m1×p2
(A21)(B11) is m2×p1 and (A22)(B21) is m2×p1
(A21)(B12) is m2×p2 and (A22)(B22) is m2×p2

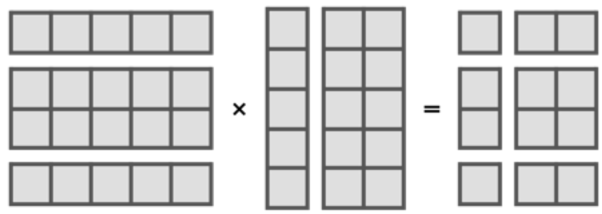
Below is one example of appropriate partition of compatible matrices A & B



Other valid forms of block multiplication

Block partitioning is not restricted to the 2×2 form
Any partitioning is valid, as long as every required block product is defined

Consider example of a valid multiplication (A4×5) × (B5×3):
- A is divided onto an arbitrary number of vertically stacked blocks
- B is divided onto an arbitrary number of horizontally stacked blocks
- Block sizes are also arbitrary



• C partition is shown & every block of C is a result of valid multiplication:

‣ C11 =A1×B1

‣ C12 =A1×B2

‣ C21 =A2×B1

‣ C22 =A2×B2

‣ C31 =A3×B1

‣ C32 =A3×B2

This completes Step 1 of the partition:
an arbitrary division of A into vertically stacked blocks
and B into horizontally stacked blocks

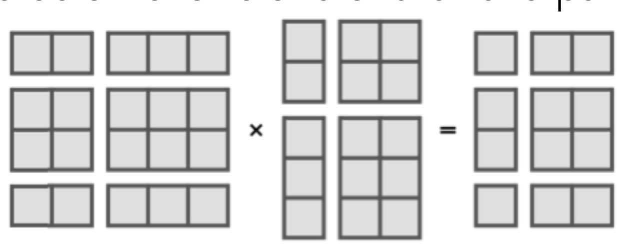
Step 2 introduces an additional partition:

• A is further divided into horizontally stacked blocks
• B is further divided into vertically stacked blocks

To carry out this step in a valid fashion,

• Note that the inner size of A matches the outer size of B
• Choose an arbitrary integer k
• Divide inner dimension of A and outer dimension of B
into k segments in such way that
they are not required to be equal, but match between A and B

Image below shows one of the valid partitions



- C partition is shown & every block of C again is a result of valid multiplication:
    - $C11 = A11B11 + A12B21$ (1×1)
    - $C12 = A11B12 + A12B22$ (1×2)
    - $C21 = A21B11 + A22B21$ (2×1)
    - $C22 = A21B12 + A22B22$ (2×2)
    - $C31 = A31B11 + A32B21$ (1×1)
    - $C32 = A31B12 + A32B22$ (1×2)

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:

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

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

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

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

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

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

---

① Product AD (D = diagonal matrix)

Consider example of 3×3 matrix A and 3×3 diagonal matrix D

$$A = \left[\vec{a}_1 \;\middle|\; \vec{a}_2 \;\middle|\; \vec{a}_3 \right]$$

$$D = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix} = \left[\vec{d}_1 \;\middle|\; \vec{d}_2 \;\middle|\; \vec{d}_3 \right]$$

By definition of matrix-matrix multiplication,

$$AD = \left[A\vec{d}_1 \;\middle|\; A\vec{d}_2 \;\middle|\; A\vec{d}_3 \right] =$$

$$\left[\begin{array}{c:c:c} \left[\vec{a}_1 \mid \vec{a}_2 \mid \vec{a}_3\right]\begin{bmatrix} d_{11} \\ 0 \\ 0 \end{bmatrix} & \left[\vec{a}_1 \mid \vec{a}_2 \mid \vec{a}_3\right]\begin{bmatrix} 0 \\ d_{22} \\ 0 \end{bmatrix} & \left[\vec{a}_1 \mid \vec{a}_2 \mid \vec{a}_3\right]\begin{bmatrix} 0 \\ 0 \\ d_{33} \end{bmatrix} \end{array}\right]$$

After applying matrix-vector multiplication rule to each column of AD,

$$AD = \left[\, d_{11}\vec{a}_1 \mid d_{22}\vec{a}_2 \mid d_{33}\vec{a}_3 \,\right]$$

Conclusion:
right multiplication by a diagonal matrix scales the columns of A

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

② Product DA (D = diagonal matrix)
Algebraic corollary (using transpose):

Suppose $B = A^T$ & $A = B^T$
so
$DA = DB^T$

Recall $(XY)^T = Y^TX^T$ and $D^T = D$,
therefore:

$$DB^T = \left((B^T)^T D^T\right)^T = (BD)^T$$

This, in effect,
• transposes A to obtain B
• scales columns of B, same as rows of A

• transposes back

$\Downarrow$

Left-multiplication of A by D scales rows of A

😊

here, we use this somewhat non-standard proof
to avoid repeating computations

Conclusion:
left multiplication by a diagonal matrix scales the rows of A

③ Identity matrix

The identity matrix I is a diagonal matrix with all diagonal entries equal to 1
From (1) and (2), diagonal multiplication scales columns (right) and rows (left)
With all diagonal entries equal to 1, nothing changes

Conclusion: $IA = AI = A$

Product of orthonormal matrices PQ

Recall definition of orthonormal matrix Q
introduced in 'Determinant of orthonormal matrix':
$Q^TQ = I$

Suppose $P^TP = I$ & $Q^TQ = I$

In order to determine if PQ satisfies definition of orthonormal matrix, compute product $(PQ)^T(PQ)$

We showed earlier that $(PQ)^T = Q^TP^T$

$$\Leftrightarrow$$

$$(PQ)^T(PQ) = Q^TP^TPQ = Q^TQ = I$$

$$\Downarrow$$

PQ does satisfy definition of orthonormal matrix

Product of upper triangular matrices TU

Suppose U is an invertible upper triangular matrix:

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}$$

In order to transform U into reduced echelon form
we need two types of elementary matrices E

1. Row replacement matrices E(r) to force above-diagonal entries to 0

Recall that E(r) are obtained by adding a multiple of row i to row j of the identity matrix

Since we only need to force above-diagonal entries, $j < i$

$$\Downarrow$$

Every E(r) is upper triangular; one possible example is shown below:

$$E(r) = \begin{bmatrix} 1 & 0 & 0 & s \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \left(E(r)\right)^{-1} = \begin{bmatrix} 1 & 0 & 0 & -s \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2. Row scaling matrices E(s) to force every diagonal entry to 1

Recall that every E(s) is obtained by scaling a row of I
and is therefore upper triangular

$\left(E(s)\right)^{-1}$ is also upper triangular

Because U is already upper triangular, no row swaps are required,
and there are no entries below the diagonal to eliminate
So U can be row reduced to I using only upper triangular elementary matrices

Now suppose matrix T is not upper triangular:

$$T = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ s & 0 & 0 & u_{44} \end{bmatrix}$$

Since T has a nonzero entry below the diagonal,
it cannot be row reduced using only upper triangular elementary row operations
At least one non-upper-triangular elementary matrix is required

Therefore:
a matrix is upper triangular if and only if it can be obtained using only upper triangular elementary row operations

Let U be an invertible upper triangular matrix
U can be row reduced to I using only upper triangular elementary matrices $E1$, $E2$, ..., Ek
Therefore
$U = E1^{-1} E2^{-1} \dots Ek^{-1}$

Each inverse Ei^{-1} is upper triangular

Now let V be another invertible upper triangular matrix
By the same reasoning,
$V = F1^{-1} F2^{-1} \dots Fm^{-1}$,
where each Fj is an upper triangular elementary matrix

Consider now the product
$W = UV$

Substituting the factorizations of U and V gives
$W = (E1^{-1} \dots Ek^{-1})(F1^{-1} \dots Fm^{-1})$

Thus W is obtained using only upper triangular elementary row operations
By the necessary-and-sufficient condition proved above,
W must itself be upper triangular

This well-known result is usually demonstrated using element-wise computations 🙂
Our goal here was to give a more structural proof based on elementary row operations

We fully acknowledge that this argument does not extend to non-invertible matrices

If you wish, you can also re-prove the result using an entry-wise computation based on the row–column rule for matrix multiplication

Learning to compare and compose different proof strategies is part of the goal here

$$\text{rank}(AB) \le \min\big(\text{rank}(A), \text{rank}(B)\big)$$

① Proof of rank(AB) ≤ rank(B)

Let $\vec{x} \in \text{null}(B)$

Then $B\vec{x} = \vec{0}$

Multiply by A: $AB\vec{x} = A(B\vec{x}) = A\vec{0} = \vec{0}$

So $\vec{x} \in \text{null}(AB)$

Therefore $\text{null}(B) \subseteq \text{null}(AB)$

② Proof of rank(AB) ≤ rank(A)

Let $\vec{y} \in \text{null}(A^T)$

Then $A^T\vec{y} = \vec{0}$

Multiply by B^T: $B^TA^T\vec{y} = B^T\vec{0} = \vec{0}$

But $B^TA^T = (AB)^T$, so $(AB)^T\vec{y} = \vec{0}$

So $\vec{y} \in \text{null}\big((AB)^T\big)$

Therefore $\text{null}(A^T) \subseteq \text{null}\big((AB)^T\big)$

Both A^T and $(AB)^T$ map from $\mathbb{R}^m$
By rank–nullity:
$\text{rank}(A^T) = m - \dim \text{null}(A^T)$

$$\text{rank}\left((AB)^T\right) = m - \dim \text{null}\left((AB)^T\right)$$

Since $\text{null}(A^T) \subseteq \text{null}\left((AB)^T\right)$, we have $\dim \text{null}(A^T) \leq \dim \text{null}\left((AB)^T\right)$

Therefore $\text{rank}\left((AB)^T\right) \leq \text{rank}(A^T)$

Recall $\text{rank}(M) = \text{rank}(M^T)$, so $\text{rank}(AB) \leq \text{rank}(A)$

Conclusion

$$\text{rank}(AB) \leq \min\left(\text{rank}(A), \text{rank}(B)\right)$$

Note This proof uses null space containment twice (for B and for A^T) to keep the logic symmetric

---

$\text{rank}(A^TA) = \text{rank}(A)$

If A is an $m \times n$ matrix, then A^TA is an $n \times n$ matrix

Claim
$\text{null}(A^TA) = \text{null}(A)$

Proof
Let $\vec{x} \in \mathbb{R}^n$

‣ If $\vec{x} \in \text{null}(A)$, then $A\vec{x} = \vec{0}$

Multiply by A^T: $A^TA\vec{x} = A^T\vec{0} = \vec{0}$

So $\vec{x} \in \text{null}(A^TA)$

‣ If $\vec{x} \in \text{null}(A^TA)$, then $A^TA\vec{x} = \vec{0}$

Multiply on the left by $\vec{x}^T$:

$\vec{x}^TA^TA\vec{x} = 0$

But $\vec{x}^TA^TA\vec{x} = (A\vec{x})^T(A\vec{x})$

So $(A\vec{x})^T(A\vec{x}) = 0$

This is a sum of squares, so it can be zero only when $A\vec{x} = \vec{0}$

Therefore $\vec{x} \in \text{null}(A)$

Thus $\text{null}(A^TA) = \text{null}(A)$

Rank consequence

Both A and A^TA map from $\mathbb{R}^n$

By rank–nullity:

$\text{rank}(A) = n - \dim \text{null}(A)$

$\text{rank}(A^TA) = n - \dim \text{null}(A^TA)$

Since the null spaces are equal, $\text{rank}(A^TA) = \text{rank}(A)$

Corollary

A has full column rank $\Leftrightarrow \text{null}(A) = \{\vec{0}\} \Leftrightarrow A^TA$ is invertible

Note that this allows $(A^TA)^{-1}$ to be used for Least Squares solutions