Matrices and Matrix Operations

This week we will learn about:

- Matrices,
- Matrix addition, scalar multiplication, and matrix multiplication,
- The transpose, and
- Matrix powers and adjacency matrices of graphs.

Extra reading and watching:

- Section 1.3 in the textbook
- Lecture videos 8, 9, 10, 11, and 12 on YouTube
- Matrix multiplication at Wikipedia
- Transpose at Wikipedia

Extra textbook problems:

⋆ 1.3.1, 1.3.2, 1.3.4, 1.3.12

⋆⋆ 1.3.3, 1.3.5–1.3.7, 1.3.9, 1.3.11, 1.3.13–1.3.15

⋆⋆⋆ 1.3.8

☠ none this week

Matrices

Previously, we introduced vectors, which can be thought of as 1D lists of numbers. Now we start working with matrices, which are 2D arrays of numbers:

> **Definition 3.1 — Matrices**
>
> A **matrix** is a rectangular array of numbers. Those numbers are called the **entries** or **elements** of the matrix.

Example. Examples of matrices.

$A = \begin{bmatrix} 1 & 3 \\ -4 & 0 \end{bmatrix}$, $\quad B = \begin{bmatrix} 7 & -8 & \pi \\ 2.7 & 0 & 3/5 \end{bmatrix}$, $\quad C = \begin{bmatrix} 3 & & 1 & & 7 \\ & 4 & & -1 & \end{bmatrix}$

✓ ✓ ✗

The **size** of a matrix is a description of the number of rows and columns that it has. A matrix with m rows and n columns has size $m \times n$.

A is 2×2, B is 2×3

Rows first, then columns.
Mnemonics: "RC cola", "RC car"

A $1 \times n$ matrix is called a **row matrix** or **row vector**. An $m \times 1$ matrix is called a **column matrix** or **column vector**. An $n \times n$ matrix is called **square**.

A is square, $D = \begin{bmatrix} 3 & 0 & 4 \end{bmatrix}$ is a row vector,

$E = \begin{bmatrix} 8 \\ -3 \\ 4.2 \end{bmatrix}$ is a column vector.

We use double subscripts to specify individual entries of a matrix: the entry of the matrix A in row i and column j is denoted by $a_{i,j}$. For example, if

$$A = \begin{bmatrix} 2 & 8 & -6 \\ 0 & -4 & 3 \end{bmatrix} \qquad (2 \times 3)$$

then $a_{1,3} = -6$ and $a_{2,2} = -4$.

Similarly, when we say "the (i,j)-entry of A", we mean $a_{i,j}$. Another notation for this is $[A]_{i,j}$, and we will see some examples shortly where this notation is advantageous.

With this notation in mind, a general $m \times n$ matrix A has the following form:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

Two matrices are **equal** if they have the same size and all of their entries (in the same positions) are equal to each other.

Example. Some (un)equal matrices.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

$A \neq B$ (2 and 3 swapped), $B \neq C$, $B \neq D$,
$A \neq C$ (C has extras zeros), $C \neq D$.
But $A = D$. ✓

We use $\mathcal{M}_{m,n}$ to denote the set of $m \times n$ matrices, and the shorthand $\mathcal{M}_n$ for the set of $n \times n$ matrices.

Just like we could add vectors or multiply vectors by a scalar, we can also add matrices and multiply matrices by scalars, and their definitions are exactly what you would expect:

Definition 3.2 — Matrix Addition and Scalar Multiplication

Suppose A and B are $m \times n$ matrices, and $c \in \mathbb{R}$ is a scalar. Then their **sum** $A+B$ is the matrix whose (i,j)-entry is $a_{i,j} + b_{i,j}$, and the **scalar multiplication** cA is the matrix whose (i,j)-entry is $ca_{i,j}$.

In other words, these operations are just performed entry-wise, as you might expect. The definition of matrix addition only makes sense when A and B have the same size.

Example. Matrix addition and scalar multiplication.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -4 \\ 2 & \pi \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 0 & 1 \\ -1 & 2 & 4 \end{bmatrix}.$$

$$A + B = \begin{bmatrix} 1+3 & 2-4 \\ 3+2 & 4+\pi \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 5 & 4+\pi \end{bmatrix}.$$

$$2C = \begin{bmatrix} 6 & 0 & 2 \\ -2 & 4 & 8 \end{bmatrix}$$

$A + C$? × Does not exist!

Matrix subtraction is defined analogously:

$A - B = A + (-1)B$ is the matrix whose (i,j)-entry is $a_{i,j} - b_{i,j}$. For example,

$$A - B = \begin{bmatrix} 1-3 & 2-(-4) \\ 3-2 & 4-\pi \end{bmatrix} = \begin{bmatrix} -2 & 6 \\ 1 & 4-\pi \end{bmatrix}.$$

Matrix addition, subtraction, and scalar multiplication satisfy all of the "natural" properties you might expect (e.g., $A+B=B+A$). We state these properties as a theorem:

Theorem 3.1 — Properties of Matrix Operations

Let $A, B, C \in \mathcal{M}_{m,n}$ and let $c, d \in \mathbb{R}$ be scalars. Then

a) $A+B=B+A$ (commutativity)

b) $(A+B)+C=A+(B+C)$ (associativity)

c) $c(A+B)=cA+cB$ (distributivity)

d) $(c+d)A=cA+dA$ (distributivity)

e) $c(dA)=(cd)A$

Proof. We will only prove part (c) of the theorem. The remaining parts of the theorem can be proved similarly: just use the definition of matrix addition and use the fact that all of these properties hold for addition of real numbers.

Let $1 \leq i \leq m$ and $1 \leq j \leq n$. Then

$[c(A+B)]_{i,j} = c[A+B]_{i,j} = c(a_{i,j}+b_{i,j}) = ca_{i,j} + cb_{i,j}$. Also,

the same!

$[cA+cB]_{i,j} = [cA]_{i,j} + [cB]_{i,j} = ca_{i,j} + cb_{i,j}$.

$\therefore [c(A+B)]_{i,j} = [cA+cB]_{i,j}$ for all i,j, so $c(A+B) = cA+cB$.

This completes the proof. ∎

Matrix Multiplication

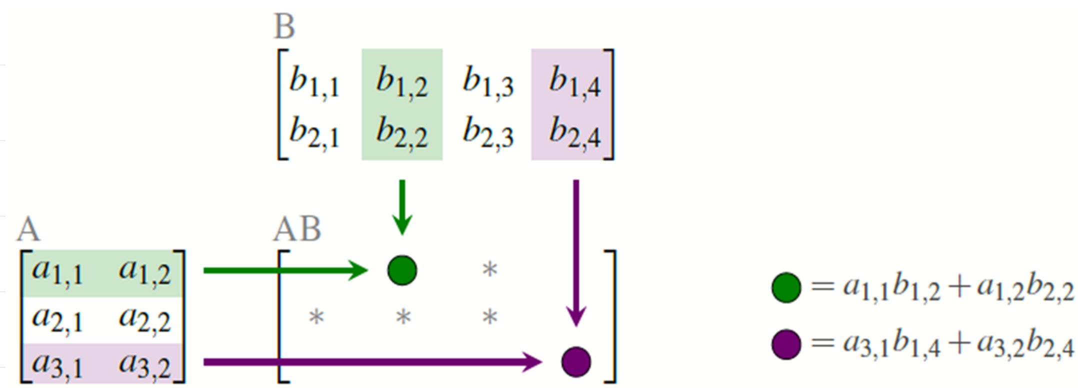
What about matrix multiplication? Recall that multiplication was a bit tricky with vectors: we only saw the dot product, which "multiplied" two vectors to give us a number. Matrix multiplication is a bit different than this, and looks quite messy and ugly at first glance. So hold onto your hats...

Definition 3.3 — Matrix Multiplication

If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then their **product** AB is the $m \times p$ matrix whose (i,j)-entry is:

$$[AB]_{i,j} \stackrel{\text{def}}{=} a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}.$$

In other words, the product AB is the matrix whose entries are all of the possible dot products of the rows of A with the columns of B.



We emphasize that the matrix product AB only makes sense if A has the same number of *columns* as B has *rows*. For example, it does not make sense to multiply a 2×3 matrix by another 2×3 matrix, but it does make sense to multiply a 2×3 matrix by a 3×7 matrix.

Example. Compute the product of two matrices.

$$A = \begin{bmatrix} 1 & 3 \\ -2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 1 \\ 3 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 2 \\ 3 & -1 & 4 \end{bmatrix}.$$

$$AB = \begin{bmatrix} 1(-2)+3(3) & 1(1)+3(-1) \\ (-2)(-2)+1(3) & (-2)(1)+1(-1) \end{bmatrix} = \begin{bmatrix} 7 & -2 \\ 7 & -3 \end{bmatrix}$$

$$AC = \begin{bmatrix} 1(1)+3(3) & 1(0)+3(-1) & 1(2)+3(4) \\ -2(1)+1(3) & -2(0)+1(-1) & -2(2)+1(4) \end{bmatrix} = \begin{bmatrix} 10 & -3 & 14 \\ 1 & -1 & 0 \end{bmatrix}$$

CA? × Does not exist!

When performing matrix multiplication, double-check that the sizes of your matrices actually make sense. In particular, the inner dimensions of the matrices must be equal, and the outer dimensions of the matrices will be the dimensions of the matrix product:

$$\underset{m\times n}{A} \qquad \underset{n\times p}{B} \qquad \underset{m\times p}{AB}$$

must match!

As always, we have defined a new operation (matrix multiplication), so we want to know what properties it satisfies.

Theorem 3.2 — Properties of Matrix Multiplication

Let A, B, and C be matrices (with sizes such that all of the multiplications below make sense) and let $c \in \mathbb{R}$ be a scalar. Then

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

b) $A(B + C) = AB + AC$ (left distributivity)

c) $(A + B)C = AC + BC$ (right distributivity)

d) $c(AB) = (cA)B = A(cB)$

Proof. We will only prove part (b) of the theorem. The remaining parts of the theorem can be proved similarly: just use the definition of matrix multiplication.

Suppose $A \in M_{m,n}$ and $B, C \in M_{n,p}$ and let $1 \le i \le m$ and $1 \le j \le p$. Then

$$[A(B+C)]_{i,j} = a_{i,1}(b_{1,j} + c_{1,j}) + a_{i,2}(b_{2,j} + c_{2,j}) + \cdots + a_{i,n}(b_{n,j} + c_{n,j})$$
$$= a_{i,1}b_{1,j} + a_{i,1}c_{1,j} + a_{i,2}b_{2,j} + a_{i,2}c_{2,j} + \cdots + a_{i,n}b_{n,j} + a_{i,n}c_{n,j}$$

$$[AB + AC]_{i,j} = [AB]_{i,j} + [AC]_{i,j} = (a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j})$$
$$+ (a_{i,1}c_{1,j} + a_{i,2}c_{2,j} + \cdots + a_{i,n}c_{n,j})$$

These are the same! $\therefore$ $A(B+C) = AB + AC$. ∎

Notice that we did *not* say anything about commutativity (i.e., we did not claim that $AB = BA$). Why not?

Example. Commutativity of matrix multiplication?

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 1 \end{bmatrix}$

$$AB = \begin{bmatrix} 1(2)+2(0) & 1(-1)+2(3) \\ 3(2)+4(0) & 3(-1)+4(3) \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 6 & 9 \end{bmatrix}$$

not the same!

$$BA = \begin{bmatrix} 2(1)+(-1)(3) & 2(2)+(-1)(4) \\ 0(1)+3(3) & 0(2)+3(4) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 9 & 12 \end{bmatrix}$$

AC exists, but CA does not

CD is 2×2, but DC is 3×3.

Example. FOILing matrices.

We have to be really careful of non-commutativity of matrix multiplication!

Does $(A+B)^2 = A^2 + 2AB + B^2$? No!

$$(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2$$

maybe different

One particularly important square matrix is the one that consists entirely of 0 entries, except with 1s on its diagonal. This is called the **identity matrix**, and it is denoted by I (or sometimes I_n if we want to emphasize it is $n \times n$).

Similarly, the **zero matrix** is the one with all entries equal to 0. We denote it by O (or $O_{m,n}$ if we care that it is $m \times n$).

Example. The identity matrix and zero matrix.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{etc.} \qquad O_{2,3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

If $A = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}$ then $AI_2 = \begin{bmatrix} 2(1)+3(0) & 2(0)+3(1) \\ (-1)(1)+4(0) & (-1)(0)+4(1) \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} = A$

and $I_2 A = \begin{bmatrix} 1(2)+0(-1) & 1(3)+0(4) \\ 0(2)+1(-1) & 0(3)+1(4) \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} = A.$

The previous example suggests the following general result, which is indeed true:

Theorem 3.3 — Multiplication by Identity or Zero

If $A \in \mathcal{M}_{m,n}$ then $AI_n = A = I_m A$ and $AO_n = O_{m,n} = O_m A$.

We won't prove the above theorem, but hopefully it seems believable enough.

Example. Diagonal matrices.

$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 6 \end{bmatrix}$ are diagonal.

$$AB = \begin{bmatrix} 2(4)+0(0) & 2(0)+0(-2) \\ 0(4)+3(0) & 0(0)+3(-2) \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & -6 \end{bmatrix}.$$

equals entrywise product!

In general, the product of two diagonal matrices is just the entry-wise product of the two matrices:

$$\begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} b_{1,1} & 0 & \cdots & 0 \\ 0 & b_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{n,n} \end{bmatrix} = \begin{bmatrix} a_{1,1}b_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2}b_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n}b_{n,n} \end{bmatrix}$$

The Transpose of a Matrix

We now introduce an operation on matrices that changes the shape of a matrix, but not its contents. Specifically, it swaps the role of the rows and columns of a matrix:

Definition 3.4 — The Transpose

Suppose $A \in \mathcal{M}_{m,n}$ is an $m \times n$ matrix. Then its **transpose**, which we denote by A^T, is the $n \times m$ matrix whose (i, j)-entry is $a_{j,i}$.

Intuitively, the transpose of a matrix is obtained by mirroring it across its main diagonal.

Example. Let's compute a transpose or two.

$$A = \begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 4 & 0 & 7 \\ 2 & -3 & 6 & 5 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \qquad B^T = \begin{bmatrix} 1 & 2 \\ 4 & -3 \\ 0 & 6 \\ 7 & 5 \end{bmatrix}$$

Let's now think about some basic properties that the transpose satisfies:

Theorem 3.4 — Properties of the Transpose

Let A and B be matrices (with sizes such that the operations below make sense) and let $c \in \mathbb{R}$ be a scalar. Then

a) $(A^T)^T = A$

b) $(A + B)^T = A^T + B^T$

c) $(AB)^T = B^T A^T$ $\longleftarrow$ not a typo! not equal to A^TB^T!

d) $(cA)^T = cA^T$

Proof. Parts (a), (b), and (d) of the theorem are intuitive enough, so we will only prove part (c):

Suppose $A \in M_{m,n}$ and $B \in M_{n,p}$, and let $1 \leq i \leq m$ and $1 \leq j \leq p$. Then

$$[(AB)^T]_{i,j} = [AB]_{j,i} = a_{j,1}b_{1,i} + a_{j,2}b_{2,i} + \cdots + a_{j,n}b_{n,i}$$

the same!

and

$$[B^TA^T]_{i,j} = [B^T]_{i,1}[A^T]_{1,j} + [B^T]_{i,2}[A^T]_{2,j} + \cdots + [B^T]_{i,n}[A^T]_{n,j}$$
$$= b_{1,i}a_{j,1} + b_{2,i}a_{j,2} + \cdots + b_{n,i}a_{j,n}$$

These are the same!

$\therefore \; (AB)^T = B^TA^T$. ∎

As a bit of a side note: would you have initially guessed that $(AB)^T = A^TB^T$? Situations like this are why we prove things rather than just guessing based on what "looks believable".

Example. Let's compute some more transposes.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 0 \\ -2 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad B^T = \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & 1 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}.$$

$$AB = \begin{bmatrix} -1 & 2 \\ 1 & 4 \end{bmatrix}, \quad A^TB^T = \begin{bmatrix} 3 & 1 \\ 6 & 0 \end{bmatrix}, \quad B^TA^T = \begin{bmatrix} -1 & 1 \\ 2 & 4 \end{bmatrix}$$

transpose of each other

$$AC = \begin{bmatrix} 3 & 3 & 2 \\ 7 & 5 & 4 \end{bmatrix}, \quad C^TA^T = \begin{bmatrix} 3 & 7 \\ 3 & 5 \\ 2 & 4 \end{bmatrix}, \quad A^TC^T?$$

× DNE!

Example. Transpose of the product of many matrices.

$(ABC)^T = ((AB)C)^T = C^T(AB)^T = C^TB^TA^T$

Order reverses, no matter how many matrices there are: $(ABCD)^T = D^TC^TB^TA^T$, etc.

The transpose has the useful property that it converts a column vector into the corresponding row vector, and vice-versa. Furthermore, if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are column vectors, then we can use our usual matrix multiplication rule to see that

$$\mathbf{v}^T\mathbf{w} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = v_1w_1 + v_2w_2 + \cdots + v_nw_n = \vec{v} \cdot \vec{w}.$$

In other words, we can use matrix multiplication to recover the dot product.

Matrix Powers

Matrix multiplication also lets us define *powers* of square matrices. For an integer $k \geq 1$, we define

$$A^k = AA\cdots A \qquad (k \text{ copies of } A)$$

and we also define $A^0 = I$ (analogously to how we define $a^0 = 1$ whenever a is a non-zero real number). The next theorem follows almost immediately from this definition:

Theorem 3.5 — Properties of Matrix Powers

If A is square and k and r are nonnegative integers, then

- $A^kA^r = A^{k+r}$, and
- $(A^k)^r = A^{kr}$.

$(xy)^k = x^ky^k$ if $x, y \in \mathbb{R}$ ✓

$(AB)^k = A^kB^k$ if $A, B \in M_n$ ✗

Example. Compute some matrix powers.

$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$, $A^2 = \begin{bmatrix} -1 & -2 \\ 4 & -1 \end{bmatrix}$, $A^3 = \begin{bmatrix} -5 & -1 \\ 2 & -5 \end{bmatrix}$

$A^4 = \begin{bmatrix} -7 & 4 \\ -8 & -7 \end{bmatrix}$ $\left(= A^3A = AA^3 = (A^2)^2\right)$

$$A^8 = (A^4)^2 = \begin{bmatrix} (-7)(-7)+4(-8) & (-7)(4)+(4)(-7) \\ (-8)(-7)+(-7)(-8) & (-8)(4)+(-7)(-7) \end{bmatrix} = \begin{bmatrix} 17 & -56 \\ 112 & 17 \end{bmatrix}.$$

Large powers can be computed quickly via exponentiation-by-squaring.

Later in the course, we will see how to define things like A^{-3} and $A^{\sqrt{2}}$.

Block Matrices

Oftentimes, there are clear "patterns" in the entries of a large matrix, and it might be useful to break that matrix down into smaller chunks based on some partition of its rows and columns. For example:

$$A = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 & -2 & 3 \end{array}\right] \quad \text{and} \quad B = \left[\begin{array}{cc|cc} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right]$$

$$A = \begin{bmatrix} I_3 & I_3 \\ O_{2,3} & C \end{bmatrix}, \text{ where } C = \begin{bmatrix} 2 & 1 & -1 \\ 0 & -2 & 3 \end{bmatrix} \text{ and}$$

$$B = \begin{bmatrix} D & O_{3,2} \\ O_{3,2} & D \end{bmatrix}, \text{ where } D = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 1 \end{bmatrix}.$$

When A and B are written in this way, as matrices whose elements are themselves matrices, they are called **block matrices**. Viewing matrices in this way often simplifies calculations and reveals structure, especially when the matrix has a lot of zeroes.

Remarkably, multiplication of block matrices works exactly as it does for regular matrices:

$$AB = \begin{bmatrix} I & I \\ O & C \end{bmatrix} \begin{bmatrix} D & O \\ O & D \end{bmatrix} = \begin{bmatrix} ID + IO & IO + ID \\ OD + CO & OO + CD \end{bmatrix} = \begin{bmatrix} D & D \\ O & CD \end{bmatrix}$$

$$\text{Well, } CD = \begin{bmatrix} 2 & 1 & -1 \\ 0 & -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ -7 & 1 \end{bmatrix}, \text{ so}$$

$$AB = \begin{bmatrix} D & D \\ O & CD \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ -1 & 1 & -1 & 1 \\ \hline 0 & 0 & 5 & 4 \\ 0 & 0 & -7 & 1 \end{array}\right].$$

And indeed, this is the exact same answer we would have gotten if we computed AB the "long way".

We have to be careful when performing block matrix multiplication: it is only valid if we choose the sizes of the blocks so that each and every matrix multiplication being performed makes sense.

Example. Suppose

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 3 & -1 \\ -1 & 0 & 2 \end{bmatrix}.$$

Which of the following block matrix multiplications make sense?

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}^2 \times$$

Partitions do not line up!

$$\left[\begin{array}{cc|ccc} 1 & 2 & 1 & 3 & -1 \\ -1 & 0 & -1 & 0 & 2 \\ \hline 1 & 3 & -1 & 1 & 2 \\ -1 & 0 & 2 & -1 & 0 \end{array}\right] ?$$

Either way, we can only take powers of square matrices.

$$\underbrace{\begin{bmatrix} A & B \\ O & I_3 \end{bmatrix}}_{2\times 2} \underbrace{\begin{bmatrix} A & A \\ O & A \\ I_2 & O \end{bmatrix}}_{3\times 2} \times$$

inner dimensions do not match

$$\begin{bmatrix} A & B \\ O & I_3 \end{bmatrix} \begin{bmatrix} A & A \\ O & A \end{bmatrix} = \begin{bmatrix} A^2 & A^2 + BA \\ O & A \end{bmatrix} \times$$

BA does not make sense (nor does $I_3 A$ at bottom-right).

$$\begin{bmatrix} A & B \\ O & I_3 \end{bmatrix} \begin{bmatrix} B & O \\ I_3 & I_3 \end{bmatrix} = \begin{bmatrix} AB + B & B \\ I_3 & I_3 \end{bmatrix}$$

$$= \left[\begin{array}{ccc|ccc} 0 & 6 & 2 & 1 & 3 & -1 \\ -2 & -3 & 3 & -1 & 0 & 2 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$AB = \begin{bmatrix} -1 & 3 & 3 \\ -1 & -3 & 1 \end{bmatrix}$$

$$AB + B = \begin{bmatrix} 0 & 6 & 2 \\ -2 & -3 & 3 \end{bmatrix}$$

Partitioning matrices in different ways can lead to new insights about how matrix multiplication works.

Theorem 3.6 — Matrix–Vector Multiplication

Suppose $A \in \mathcal{M}_{m,n}$ has columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and $\mathbf{v} \in \mathbb{R}^n$ is a column vector. Then

$$A\mathbf{v} = v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + \cdots + v_n\mathbf{a}_n.$$

Proof. We simply perform block matrix multiplication:

Write A and $\vec{v}$ as block matrices:

$$A = \left[\vec{a}_1 \,\vdots\, \vec{a}_2 \,\vdots\, \cdots \,\vdots\, \vec{a}_n\right] \quad \text{and} \quad \vec{v} = \begin{bmatrix} v_1 \\ \hdashline v_2 \\ \hdashline \vdots \\ \hdashline v_n \end{bmatrix}. \quad \text{Then}$$

$(1 \times n)$ $(n \times 1)$

$$A\vec{v} = \vec{a}_1 v_1 + \vec{a}_2 v_2 + \cdots + \vec{a}_n v_n = v_1\vec{a}_1 + v_2\vec{a}_2 + \cdots + v_n\vec{a}_n.$$ ■

We of course can compute $A\mathbf{v}$ directly from the definition, but it's nice to have multiple ways to think about things.

Theorem 3.7 — Matrix Multiplication is Column-Wise

Suppose $A \in \mathcal{M}_{m,n}$ and $B \in \mathcal{M}_{n,p}$ are matrices. If $\mathbf{b}_j$ is the j-th column of B, then

$$AB = \left[\, A\mathbf{b}_1 \mid A\mathbf{b}_2 \mid \cdots \mid A\mathbf{b}_p \,\right].$$

Proof. Again, we perform block matrix multiplication:

$A = [A]$ (1×1 block matrix)

$B = \left[\vec{b}_1 \,\vdots\, \vec{b}_2 \,\vdots\, \cdots \,\vdots\, \vec{b}_p\right]$ (1×p block matrix)

Then $AB = \left[A\vec{b}_1 \,\vdots\, A\vec{b}_2 \,\vdots\, \cdots \,\vdots\, A\vec{b}_p\right]$.

That's it! ■