§2.1: Matrix Operations

We have several ways to combine functions to make new functions:

- Addition: if f, g have the same domains and codomains, then we can set $(f+g)\mathbf{x} = f(\mathbf{x}) + g(\mathbf{x})$,
- Composition: if the codomain of f is the domain of g, then we can set $(g \circ f)\mathbf{x} = g(f(\mathbf{x}))$,
- Inverse (§2.2): if f is one-to-one and onto, then we can set $f^{-1}(\mathbf{y})$ to be the unique solution to $f(\mathbf{x}) = \mathbf{y}$.

It turns out that the sum, composition and inverse of linear transformations are also linear (exercise: prove it!), and we can ask how the standard matrix of the new function is related to the standard matrices of the old functions.

Notation:

The (i, j)-entry of a matrix A is the entry in row i, column j, and is written a_{ij} or $(A)_{ij}$.

The diagonal entries of A are the entries $a_{11}, a_{22}, \ldots$.

e.g. $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}$

A square matrix has the same number of rows as columns. The associated linear transformation has the same domain and codomain.

A diagonal matrix is a square matrix whose nondiagonal entries are 0.

e.g. $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

The identity matrix I_n is the $n \times n$ matrix whose diagonal entries are 1 and whose nondiagonal entries are 0.
It is the standard matrix for the identity transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{x}) = \mathbf{x}$.

e.g. $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Addition:

If A, B are the standard matrices for some linear transformations $S, T : \mathbb{R}^n \to \mathbb{R}^m$, then $(S+T)\mathbf{x} = S(\mathbf{x}) + T(\mathbf{x})$ is a linear transformation. What is its standard matrix $A + B$?

Proceed column by column:

First column of the standard matrix of $S + T$

$= (S+T)(\mathbf{e_1})$ definition of standard matrix of $S + T$

Addition:

If A, B are the standard matrices for some linear transformations $S, T : \mathbb{R}^n \to \mathbb{R}^m$, then $(S+T)\mathbf{x} = S(\mathbf{x}) + T(\mathbf{x})$ is a linear transformation. What is its standard matrix $A+B$?

Proceed column by column:

First column of the standard matrix of $S+T$

$= (S+T)(\mathbf{e_1})$ definition of standard matrix of $S+T$

$= S(\mathbf{e_1}) + T(\mathbf{e_1})$ definition of $S+T$

Addition:

If A, B are the standard matrices for some linear transformations $S, T : \mathbb{R}^n \to \mathbb{R}^m$, then $(S+T)\mathbf{x} = S(\mathbf{x}) + T(\mathbf{x})$ is a linear transformation. What is its standard matrix $A+B$?

Proceed column by column:

First column of the standard matrix of $S+T$

$= (S+T)(\mathbf{e_1})$ definition of standard matrix of $S+T$

$= S(\mathbf{e_1}) + T(\mathbf{e_1})$ definition of $S+T$

$=$ first column of A $+$ first column of B. definition of standard matrix of S and of T

Addition:

If A, B are the standard matrices for some linear transformations $S, T : \mathbb{R}^n \to \mathbb{R}^m$, then $(S+T)\mathbf{x} = S(\mathbf{x}) + T(\mathbf{x})$ is a linear transformation. What is its standard matrix $A+B$?

Proceed column by column:

First column of the standard matrix of $S+T$

$= (S+T)(\mathbf{e_1})$ definition of standard matrix of $S+T$

$= S(\mathbf{e_1}) + T(\mathbf{e_1})$ definition of $S+T$

$=$ first column of A + first column of B. definition of standard matrix of S and of T

i.e. $(i,1)$-entry of $A+B$=$a_{i1} + b_{i1}$.

The same is true of all the other columns, so $(A+B)_{ij}$=$a_{ij} + b_{ij}$.

Example: $A = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix}, \quad A+B = \begin{bmatrix} 5 & 1 & 6 \\ 2 & 8 & 9 \end{bmatrix}.$

Scalar multiplication:

If A is the standard matrix for a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^m$, and c is a scalar, then $(cS)\mathbf{x} = c(S\mathbf{x})$ is a linear transformation. What is its standard matrix cA?

Proceed column by column:

First column of the standard matrix of cS

$= (cS)(\mathbf{e_1})$	definition of standard matrix of cS
$= c(S\mathbf{e_1})$	definition of cS
$=$ first column of A multiplied by c.	definition of standard matrix of S

i.e. $(i,1)$-entry of cA=ca_{i1}.

The same is true of all the other columns, so $(cA)_{ij} = ca_{ij}$.

Example: $A = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix}, \quad c = -3, \quad cA = \begin{bmatrix} -12 & 0 & -15 \\ 3 & -9 & -6 \end{bmatrix}.$

Addition and scalar multiplication satisfy some familiar rules of arithmetic:

Let A, B, and C be matrices of the same size, and let r and s be scalars. Then

a. $A + B = B + A$

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

c. $A + 0 = A$

d. $r(A + B) = rA + rB$

e. $(r + s)A = rA + sA$

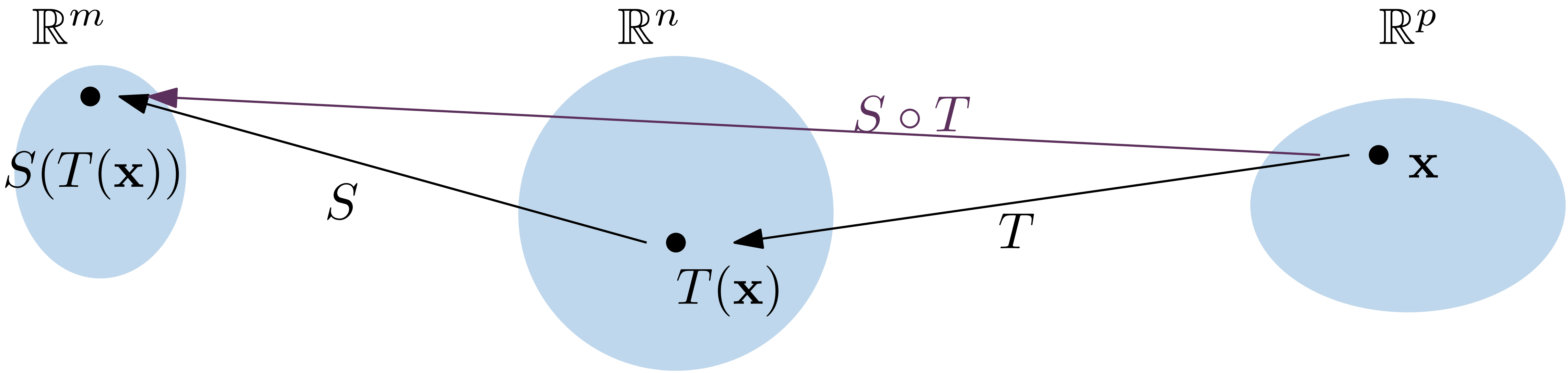
f. $r(sA) = (rs)A$

0 denotes the zero matrix:

$$0 = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

Composition:

If A is the standard matrix for a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^m$
and B is the standard matrix for a linear transformation $T : \mathbb{R}^p \to \mathbb{R}^n$
then the composition $S \circ T$ (T first, then S) is linear.
What is its standard matrix AB?



A is a $m \times n$ matrix,
B is a $n \times p$ matrix,
AB is a $m \times p$ matrix - so the (i, j)-entry of AB cannot simply be $a_{ij}b_{ij}$.

Composition:

Proceed column by column:

First column of the standard matrix of $S \circ T$

$= (S \circ T)(\mathbf{e_1})$ definition of standard matrix of $S \circ T$

$= S(T(\mathbf{e_1}))$ definition of $S \circ T$

$= S(\mathbf{b_1})$ definition of standard matrix of T (writing $\mathbf{b}_j$ for column j of B)

$= A\mathbf{b_1}$, and similarly for the other columns.

So
$$AB = A \begin{bmatrix} | & & | \\ \mathbf{b}_1 & \dots & \mathbf{b}_p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\mathbf{b}_1 & \dots & A\mathbf{b}_p \\ | & & | \end{bmatrix}.$$

The jth column of AB is a linear combination of the columns of A using weights from the jth column of B.

Composition:

Proceed column by column:

First column of the standard matrix of $S \circ T$
$= (S \circ T)(\mathbf{e_1})$ definition of standard matrix of $S \circ T$
$= S(T(\mathbf{e_1}))$ definition of $S \circ T$
$= S(\mathbf{b_1})$ definition of standard matrix of T (writing $\mathbf{b}_j$ for column j of B)
$= A\mathbf{b_1}$, and similarly for the other columns.

So
$$AB = A\begin{bmatrix} | & & | \\ \mathbf{b}_1 & \dots & \mathbf{b}_p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\mathbf{b}_1 & \dots & A\mathbf{b}_p \\ | & & | \end{bmatrix}.$$

The jth column of AB is a linear combination of the columns of A using weights from the jth column of B.

Another view is the row-column method: the (i,j)-entry of AB is $a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

Some familiar rules of arithmetic hold for matrix multiplication...

Let A be $m \times n$ and let B and C have sizes for which the indicated sums and products are defined (different sizes for each statement).

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left - distributive law)

c. $(B + C)A = BA + CA$ (right-distributive law)

d. $r(AB) = (rA)B = A(rB)$

for any scalar r

e. $I_m A = A = AI_n$ (identity for matrix multiplication)

Some familiar rules of arithmetic hold for matrix multiplication...

Let A be $m \times n$ and let B and C have sizes for which the indicated sums and products are defined (different sizes for each statement).

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B+C) = AB + AC$ (left - distributive law)

c. $(B+C)A = BA + CA$ (right-distributive law)

d. $r(AB) = (rA)B = A(rB)$
for any scalar r

e. $I_m A = A = AI_n$ (identity for matrix multiplication)

... but not all of them:

- Usually, $AB \neq BA$ (because order matters for function composition: $S \circ T \neq T \circ S$);
- It is possible for $AB = 0$ even if $A \neq 0$ and $B \neq 0$ - so you cannot solve matrix equations by 'factorising'.

A fun application of matrix multiplication:

Consider rotations counterclockwise about the origin.
Rotation through $(\theta + \varphi)$ = (rotation through θ) $\circ$ (rotation through φ).

$$\begin{bmatrix} \cos(\theta+\varphi) & -\sin(\theta+\varphi) \\ \sin(\theta+\varphi) & \cos(\theta+\varphi) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$
$$= \begin{bmatrix} \cos\theta\cos\varphi - \sin\theta\sin\varphi & -\cos\theta\sin\varphi - \sin\theta\cos\varphi \\ \sin\theta\cos\varphi + \cos\theta\sin\varphi & -\sin\theta\sin\varphi + \cos\theta\cos\varphi \end{bmatrix}.$$

So, equating the entries in the first column:
$\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi$
$\sin(\theta+\varphi) = \cos\theta\sin\varphi + \sin\theta\cos\varphi$

Powers:

For a square matrix A, the kth power of A is $A^k = \underbrace{A \dots A}_{k \text{ times}}$.

If A is the standard matrix for a linear transformation T, then A^k is the standard matrix for T^k, the function that "applies T k times".

Examples:

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}^3 = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \left(\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \right) = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 13 & 14 \\ 21 & 6 \end{bmatrix}.$$

$$\begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}^3 = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \left(\begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \right) = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 27 & 0 \\ 0 & -8 \end{bmatrix} = \begin{bmatrix} 3^3 & 0 \\ 0 & (-2)^3 \end{bmatrix}.$$

Exercise: show that $\begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}^k = \begin{bmatrix} a_{11}^k & 0 \\ 0 & a_{22}^k \end{bmatrix}$, and similarly for larger diagonal matrices.

We can consider polynomials involving square matrices:

Example: Let $p(x) = x^3 - 2x^2 + x - 2$, and $A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$ as on the previous page. Then

use the identity matrix instead of constants

$$p(A) = A^3 - 2A^2 + A - 2I_2 = \begin{bmatrix} 13 & 14 \\ 21 & 6 \end{bmatrix} - \begin{bmatrix} 14 & 4 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 12 \\ 18 & -8 \end{bmatrix}.$$

$$p(D) = D^3 - 2D^2 + D - 2I_2 = \begin{bmatrix} 27 & 0 \\ 0 & -8 \end{bmatrix} - \begin{bmatrix} 18 & 0 \\ 0 & 8 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & -20 \end{bmatrix} = \begin{bmatrix} p(3) & 0 \\ 0 & p(-2) \end{bmatrix}.$$

For a polynomial involving a single matrix, we can factorise and expand as usual:

Example: $x^3 - 2x^2 + x - 2 = (x^2 + 1)(x - 2)$, and

$$(A^2 + I_2)(A - 2I_2) = \begin{bmatrix} 8 & 2 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} -2 & 12 \\ 18 & -8 \end{bmatrix}.$$

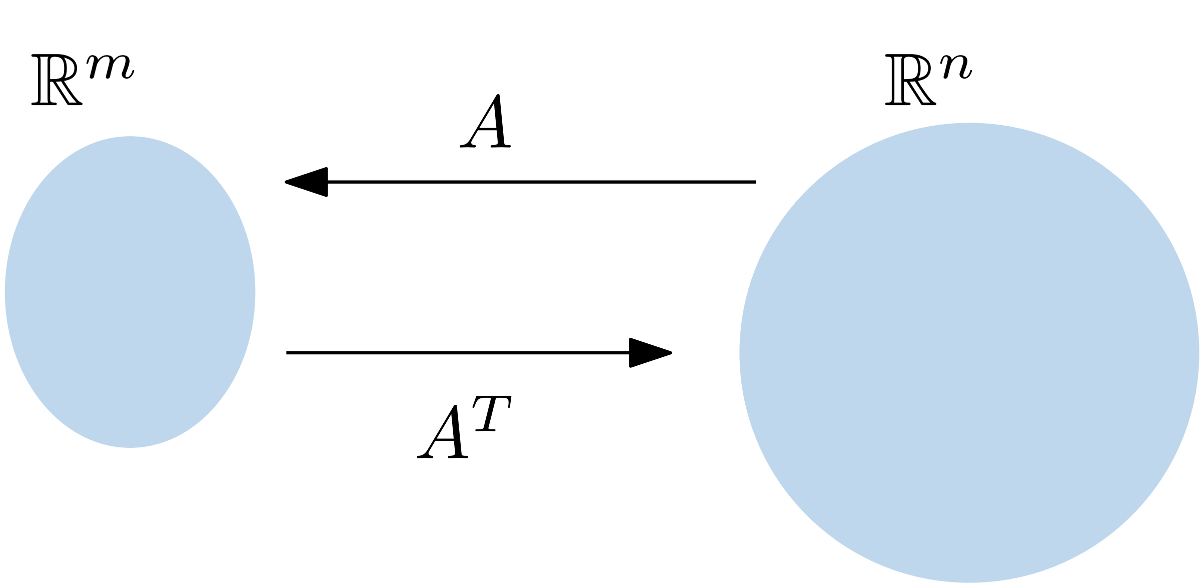
But be careful with the order when there are two or more matrices:

Example: $x^2 - y^2 = (x + y)(x - y)$, but

$$(A + D)(A - D) = A^2 - AD + DA - D^2 \neq A^2 - D^2.$$

Transpose:

The transpose of A is the matrix A^T whose (i, j)-entry is a_{ji}.
i.e. we obtain A^T by "flipping A through the main diagonal".



As a linear transformation, it "goes in the opposite direction", but it is NOT the inverse function.

Example: $A = \begin{bmatrix} 4 & 2 \\ 3 & -5 \\ 0 & 1 \end{bmatrix}, \quad A^T = \begin{bmatrix} 4 & 3 & 0 \\ 2 & -5 & 1 \end{bmatrix}.$

We will be interested in square matrices A such that
$A = A^T$ (symmetric matrix, self-adjoint linear transformation, §7.1), or
$A^{-1} = A^T$ (orthogonal matrix, or isometric linear transformation, §6.2).

Properties of the transpose:

Let A and B denote matrices whose sizes are appropriate for the following sums and products (different sizes for each statement).

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

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

c. For any scalar r, $(rA)^T =$

d. $(AB)^T =$

Example: $A = \begin{bmatrix} 4 & 2 \\ 3 & -5 \\ 0 & 1 \end{bmatrix}, \quad A^T = \begin{bmatrix} 4 & 3 & 0 \\ 2 & -5 & 1 \end{bmatrix}.$

Properties of the transpose:

Let A and B denote matrices whose sizes are appropriate for the following sums and products (different sizes for each statement).

a. $(A^T)^T = A$ (I.e., the transpose of A^T is A)

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

c. For any scalar r, $(rA)^T =$

d. $(AB)^T =$

Example: $A = \begin{bmatrix} 4 & 2 \\ 3 & -5 \\ 0 & 1 \end{bmatrix}, \quad A^T = \begin{bmatrix} 4 & 3 & 0 \\ 2 & -5 & 1 \end{bmatrix}.$

Properties of the transpose:

Let A and B denote matrices whose sizes are appropriate for the following sums and products (different sizes for each statement).

a. $(A^T)^T = A$ (I.e., the transpose of A^T is A)

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

c. For any scalar r, $(rA)^T = rA^T$

d. $(AB)^T =$

An example to explain d.
this is NOT a proof

$$\overset{A}{\begin{bmatrix} 4 & -2 \\ 3 & -5 \\ 0 & 1 \end{bmatrix}} \overset{B}{\begin{bmatrix} 2 & -3 \\ 6 & -7 \end{bmatrix}} = \begin{bmatrix} -4 & 2 \\ -24 & 26 \\ 6 & -7 \end{bmatrix}$$

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

Properties of the transpose:

Let A and B denote matrices whose sizes are appropriate for the following sums and products (different sizes for each statement).

a. $(A^T)^T = A$ (I.e., the transpose of A^T is A)

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

c. For any scalar r, $(rA)^T = rA^T$

d. $(AB)^T = B^T A^T$ (I.e. the transpose of a product of matrices equals the product of their transposes in reverse order.)

An example to explain d.
this is NOT a proof

$$\begin{array}{cc} A & B \\ \begin{bmatrix} 4 & -2 \\ 3 & -5 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 2 & -3 \\ 6 & -7 \end{bmatrix} \end{array} = \begin{bmatrix} -4 & 2 \\ -24 & 26 \\ 6 & -7 \end{bmatrix}$$

Guess: $\cancel{\begin{array}{cc} A^T & B^T \\ \begin{bmatrix} 4 & 3 & 0 \\ -2 & -5 & 1 \end{bmatrix} & \begin{bmatrix} 2 & 6 \\ -3 & -7 \end{bmatrix} \end{array}}$

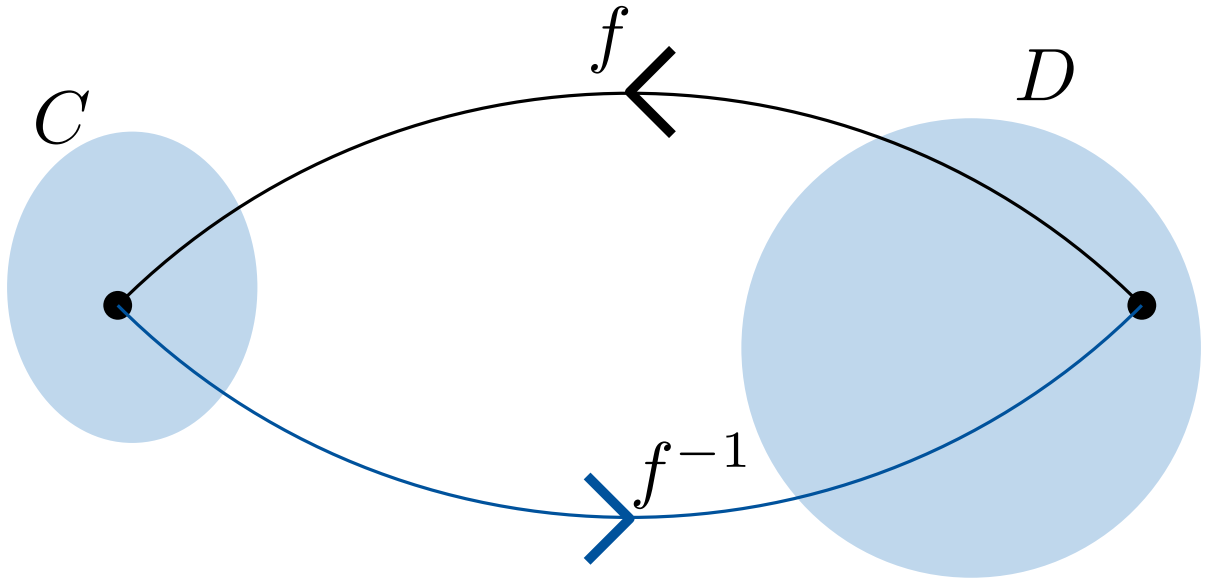
Cannot multiply these matrices!

Guess again: $\begin{array}{cc} B^T & A^T \\ \begin{bmatrix} 2 & 6 \\ -3 & -7 \end{bmatrix} & \begin{bmatrix} 4 & 3 & 0 \\ -2 & -5 & 1 \end{bmatrix} \end{array} = \begin{bmatrix} -4 & -24 & 6 \\ 2 & 26 & -7 \end{bmatrix}$

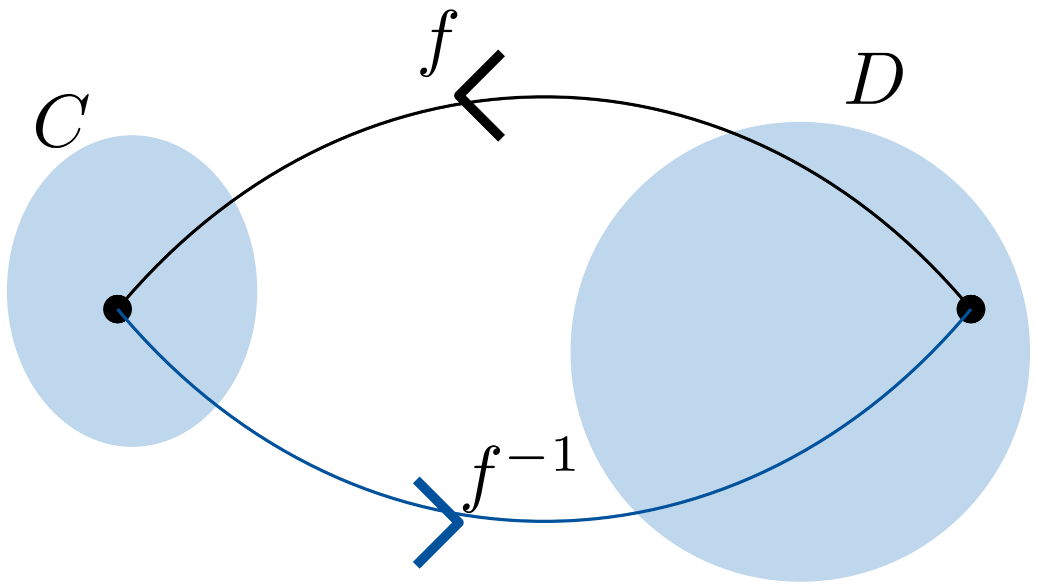
§2.2: The Inverse of a Matrix

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.
So f^{-1} exists if and only if f is one-to-one and onto. Then we say f is invertible.



§2.2: The Inverse of a Matrix



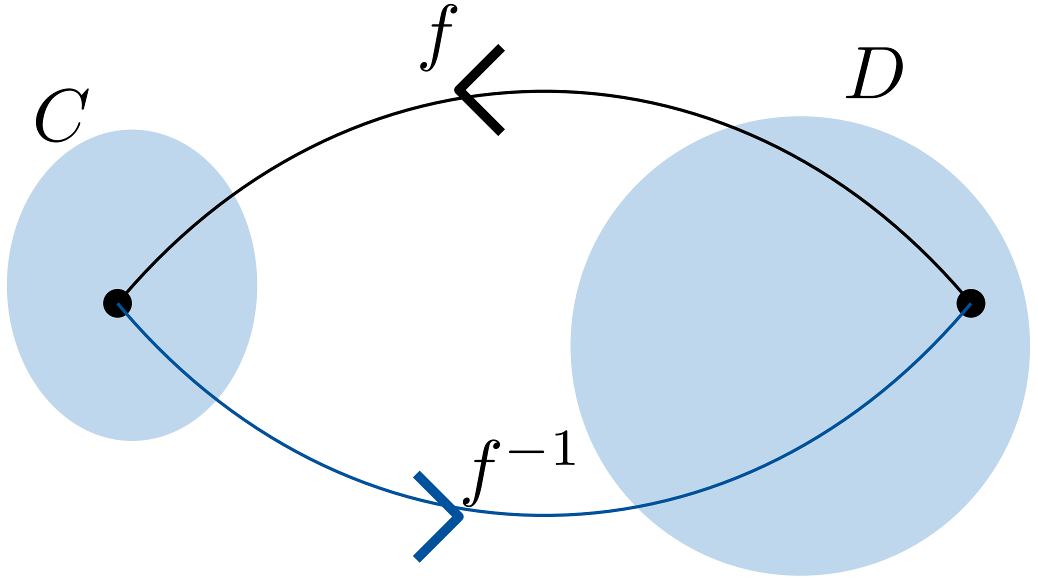
Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.
So f^{-1} exists if and only if f is one-to-one and onto. Then we say f is invertible.

Let T be a linear transformation whose standard matrix is A. From last week:

- T is one-to-one if and only if rref(A) has a pivot in every
- T is onto if and only if rref(A) has a pivot in every

§2.2: The Inverse of a Matrix



Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.
So f^{-1} exists if and only if f is one-to-one and onto. Then we say f is invertible.

Let T be a linear transformation whose standard matrix is A. From last week:

- T is one-to-one if and only if rref(A) has a pivot in every column.
- T is onto if and only if rref(A) has a pivot in every row.

So if T is invertible, then A must be a square matrix.

Warning: not all square matrices come from invertible linear transformations, e.g. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Definition: A $n \times n$ matrix A is *invertible* if there is a $n \times n$ matrix C satisfying $CA = AC = I_n$.

Fact: A matrix C with this property is unique:
if $BA = AC = I_n$, then $BAC = BI_n = B$ and $BAC = I_nC = C$ so $B = C$.

The matrix C is called the inverse of A, and is written A^{-1}. So

$$A^{-1}A = AA^{-1} = I_n.$$

A matrix that is not invertible is sometimes called singular.

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

In particular, if A is an invertible $n \times n$ matrix, then rref$(A) =$?

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof:

1. We show $A^{-1}\mathbf{b}$ is a solution

2. We show this is the unique solution:

In particular, if A is an invertible $n \times n$ matrix, then rref$(A) =$?

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof:

1. We show $A^{-1}\mathbf{b}$ is a solution (i.e. $A(A^{-1}\mathbf{b}) = \mathbf{b}$).

$A(A^{-1}\mathbf{b}) =$

2. We show this is the unique solution:

In particular, if A is an invertible $n \times n$ matrix, then rref$(A) =$?

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof:

1. We show $A^{-1}\mathbf{b}$ is a solution (i.e. $A(A^{-1}\mathbf{b}) = \mathbf{b}$).
$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, so $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$:

2. We show this is the unique solution:

In particular, if A is an invertible $n \times n$ matrix, then rref$(A) = ?$

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof:

1. We show $A^{-1}\mathbf{b}$ is a solution (i.e. $A(A^{-1}\mathbf{b}) = \mathbf{b}$).

$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, so $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$:

2. We show this is the unique solution:

Let $\mathbf{u}$ be any solution to $A\mathbf{x} = \mathbf{b}$, so: $$A\mathbf{u} = \mathbf{b}$$

Multiply both sides by A^{-1} on the left: $$A^{-1}(A\mathbf{u}) = A^{-1}\mathbf{b}$$
$$\mathbf{u} = A^{-1}\mathbf{b}.$$

In particular, if A is an invertible $n \times n$ matrix, then $\text{rref}(A) = ?$

Remember from calculus that the inverse of a function $f : D \to C$ is the function $f^{-1} : C \to D$ such that $f^{-1} \circ f =$ identity function on D and $f \circ f^{-1} =$ identity function on C.

Equivalently, $f^{-1}(y)$ is the unique solution to $f(x) = y$.

Theorem 5: Solving linear systems with the inverse: If A is an invertible $n \times n$ matrix, then, for each $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof:

1. We show $A^{-1}\mathbf{b}$ is a solution (i.e. $A(A^{-1}\mathbf{b}) = \mathbf{b}$).

$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, so $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$:

2. We show this is the unique solution:

Let $\mathbf{u}$ be any solution to $A\mathbf{x} = \mathbf{b}$, so: $\quad A\mathbf{u} = \mathbf{b}$

Multiply both sides by A^{-1} on the left: $\quad A^{-1}(A\mathbf{u}) = A^{-1}\mathbf{b}$

$$\mathbf{u} = A^{-1}\mathbf{b}.$$

In particular, if A is an invertible $n \times n$ matrix, then $\text{rref}(A) = I_n$.

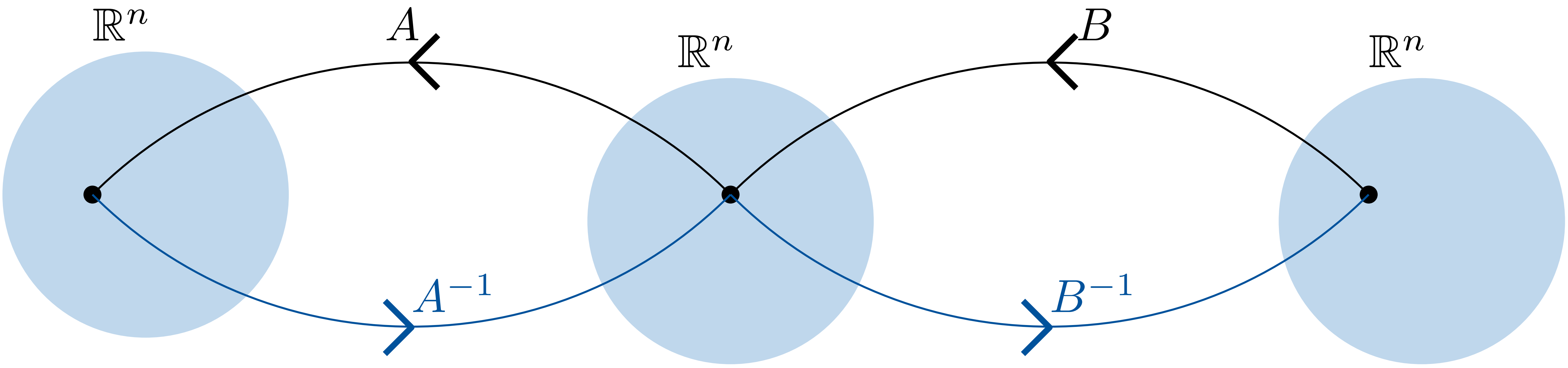
Properties of the inverse:

Suppose A and B are invertible. Then the following results hold:

a. A^{-1} is invertible and $(A^{-1})^{-1} = A$ (i.e. A is the inverse of A^{-1}).

b. AB is invertible and $(AB)^{-1} = ?$ (think about composition of functions, see diagram below)

c. A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$



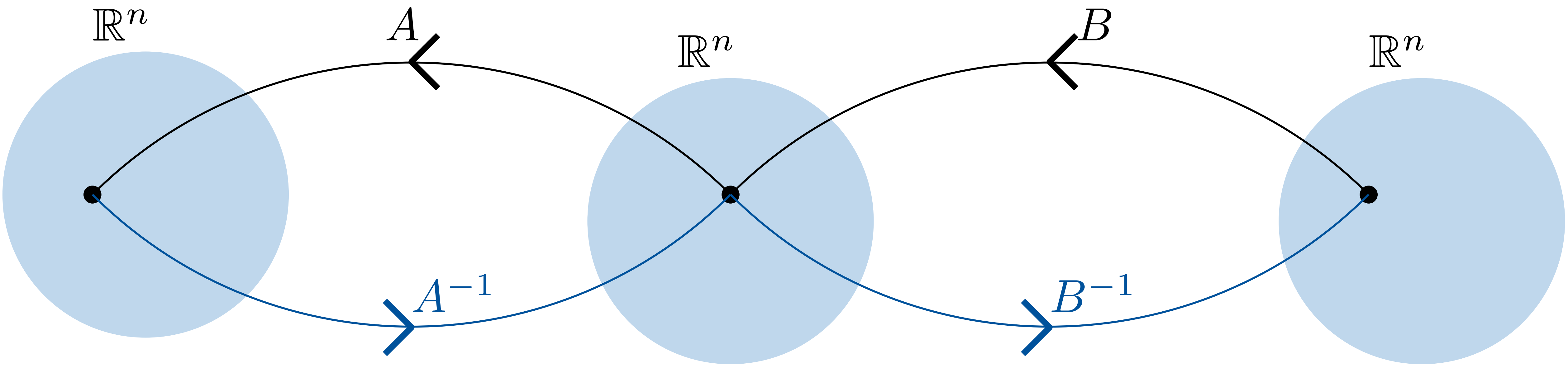
Properties of the inverse:

Suppose A and B are invertible. Then the following results hold:

a. A^{-1} is invertible and $(A^{-1})^{-1} = A$ (i.e. A is the inverse of A^{-1}).

b. AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ (think about composition of functions, see diagram below)

c. A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$



Exercise: prove these properties.

(Hint: to show X is the inverse of Y, i.e. $Y^{-1} = X$, you should check ???

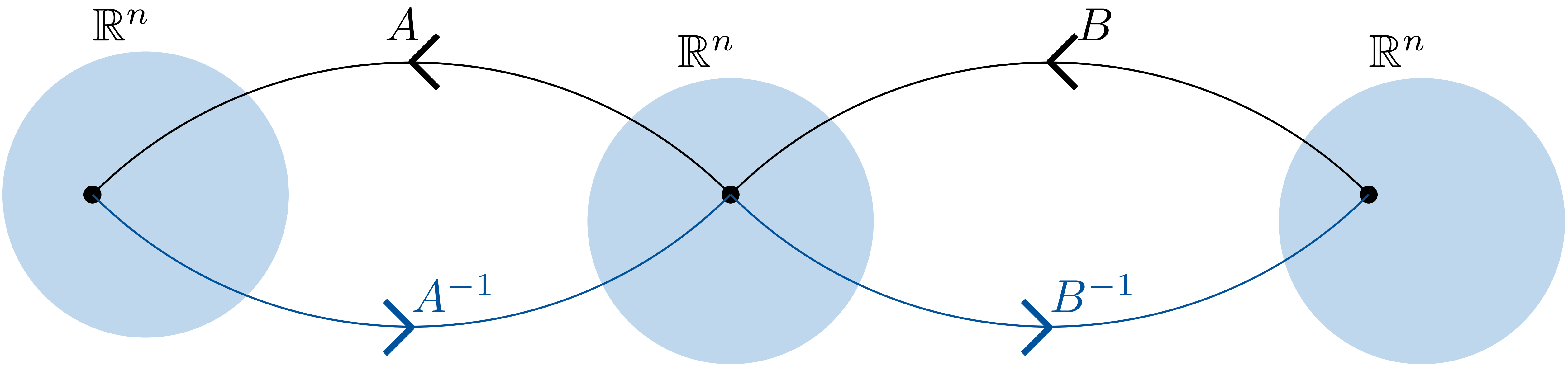
Properties of the inverse:

Suppose A and B are invertible. Then the following results hold:

a. A^{-1} is invertible and $(A^{-1})^{-1} = A$ (i.e. A is the inverse of A^{-1}).

b. AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ (think about composition of functions, see diagram below)

c. A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$



Exercise: prove these properties.
(Hint: to show X is the inverse of Y, i.e. $Y^{-1} = X$, you should check $XY = YX = I$.)

Inverse of a 2×2 matrix:

Fact: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

i) if $ad - bc \neq 0$, then A is invertible and $A^{-1} = \dfrac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$,

ii) if $ad - bc = 0$, then A is not invertible.

Inverse of a 2×2 matrix:

Fact: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

i) if $ad - bc \neq 0$, then A is invertible and $A^{-1} = \dfrac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$,

ii) if $ad - bc = 0$, then A is not invertible.

Proof of i):

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \left(\frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right) = \frac{1}{ad-bc} \begin{bmatrix} ad-bc & -ab+ba \\ cd-dc & -cb+da \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\left(\frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right) \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} da-bc & db-bd \\ -ca+ac & -cb+ad \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Proof of ii): next week.

Inverse of a 2×2 matrix:

Example: Let $A = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$, the standard matrix of rotation about the origin through an angle φ counterclockwise.

Example: Let $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, the standard matrix of projection to the x_1-axis.

Inverse of a 2×2 matrix:

Example: Let $A = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$, the standard matrix of rotation about the origin through an angle φ counterclockwise.

$\cos \varphi \cos \varphi - (-\sin \varphi) \sin \varphi = \cos^2 \varphi + \sin^2 \varphi = 1 \neq 0$ so A is invertible, and

$A^{-1} = \frac{1}{1} \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} = \begin{bmatrix} \cos(-\varphi) & -\sin(-\varphi) \\ \sin(-\varphi) & \cos(-\varphi) \end{bmatrix}$, the standard matrix of rotation about the origin through an angle φ clockwise.

Example: Let $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, the standard matrix of projection to the x_1-axis.

Inverse of a 2×2 matrix:

Example: Let $A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$, the standard matrix of rotation about the origin through an angle φ counterclockwise.

$\cos\varphi\cos\varphi - (-\sin\varphi)\sin\varphi = \cos^2\varphi + \sin^2\varphi = 1 \neq 0$ so A is invertible, and $A^{-1} = \frac{1}{1}\begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} = \begin{bmatrix} \cos(-\varphi) & -\sin(-\varphi) \\ \sin(-\varphi) & \cos(-\varphi) \end{bmatrix}$, the standard matrix of rotation about the origin through an angle φ clockwise.

Example: Let $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, the standard matrix of projection to the x_1-axis.

$1 \cdot 0 - 0 \cdot 0 = 0$ so B is not invertible.

Exercise: choose a matrix C that is the standard matrix of a reflection, and check that C is invertible and $C^{-1} = C$.

Inverse of a $n \times n$ matrix:

If A is the standard matrix of an invertible linear transformation T, then A^{-1} is the standard matrix of T^{-1}. So

$$A^{-1} = \begin{bmatrix} | & & | \\ T^{-1}(\mathbf{e}_1) & \dots & T^{-1}(\mathbf{e}_n) \\ | & & | \end{bmatrix}.$$

$T^{-1}(\mathbf{e}_i)$ is the unique solution to the equation $T(\mathbf{x}) = \mathbf{e_i}$, or equivalently $A\mathbf{x} = \mathbf{e_i}$. So if we row-reduce the augmented matrix $[A|\mathbf{e_i}]$, we should get $[I_n|T^{-1}(\mathbf{e}_i)]$. (Remember rref$(A) = I_n$.)

We carry out this row-reduction for all $\mathbf{e_i}$ at the same time:

$$[A|I_n] = \left[\begin{array}{c|ccc} & | & & | \\ A & \mathbf{e}_1 & \dots & \mathbf{e}_n \\ & | & & | \end{array}\right] \xrightarrow{\text{row reduction}} \left[\begin{array}{c|ccc} & | & & | \\ I_n & T^{-1}(\mathbf{e}_1) & \dots & T^{-1}(\mathbf{e}_n) \\ & | & & | \end{array}\right] = [I_n|A^{-1}].$$

We showed that, if A is invertible, then $[A|I_n]$ row-reduces to $[I_n|A^{-1}]$.
In other words, if we already knew that A was invertible, then we can find its inverse by row-reducing $[A|I_n]$.
It would be useful if we could apply this without first knowing that A is invertible.

We showed that, if A is invertible, then $[A|I_n]$ row-reduces to $[I_n|A^{-1}]$.
In other words, if we already knew that A was invertible, then we can find its inverse by row-reducing $[A|I_n]$.
It would be useful if we could apply this without first knowing that A is invertible.

Indeed, we can:
Fact: If $[A|I_n]$ row-reduces to $[I_n|C]$, then A is invertible and $C = A^{-1}$.

Proof: (different from textbook, not too important)
If $[A|I_n]$ row-reduces to $[I_n|C]$, then $\mathbf{c}_i$ is the unique solution to $A\mathbf{x} = \mathbf{e}_i$, so $AC\mathbf{e}_i = A\mathbf{c}_i = \mathbf{e}_i$ for all i, so $AC = I_n$.
Also, by switching the left and right sides, and reading the process backwards, $[C|I_n]$ row-reduces to $[I_n|A]$. So $\mathbf{a}_i$ is the unique solution to $C\mathbf{x} = \mathbf{e}_i$, so $CA\mathbf{e}_i = C\mathbf{a}_i = \mathbf{e}_i$ for all i, so $CA = I_n$.

We showed that, if A is invertible, then $[A|I_n]$ row-reduces to $[I_n|A^{-1}]$.
In other words, if we already knew that A was invertible, then we can find its inverse by row-reducing $[A|I_n]$.
It would be useful if we could apply this without first knowing that A is invertible.

Indeed, we can:

Fact: If $[A|I_n]$ row-reduces to $[I_n|C]$, then A is invertible and $C = A^{-1}$.

Proof: (different from textbook, not too important)

If $[A|I_n]$ row-reduces to $[I_n|C]$, then $\mathbf{c}_i$ is the unique solution to $A\mathbf{x} = \mathbf{e}_i$, so $AC\mathbf{e}_i = A\mathbf{c}_i = \mathbf{e}_i$ for all i, so $AC = I_n$.

Also, by switching the left and right sides, and reading the process backwards, $[C|I_n]$ row-reduces to $[I_n|A]$. So $\mathbf{a}_i$ is the unique solution to $C\mathbf{x} = \mathbf{e}_i$, so $CA\mathbf{e}_i = C\mathbf{a}_i = \mathbf{e}_i$ for all i, so $CA = I_n$.

In particular: an $n \times n$ matrix A is invertible if and only if $\text{rref}(A) = I_n$.
Also equivalent: $\text{rref}(A)$ has a pivot in every row and column.
For a square matrix, having a pivot in each row is the same as having a pivot in each column.

§2.3: Characterisations of Invertible Matrices

As observed at the end of the previous page: for a square $n \times n$ matrix A, the following are equivalent:

- A is invertible.
- $\text{rref}(A) = I_n$.
- $\text{rref}(A)$ has a pivot in every row.
- $\text{rref}(A)$ has a pivot in every column.

This means that, in the very special case when A is a square matrix, all the statements in the Existence of Solutions Theorem ("green theorem") and all the statements in the Uniqueness of Solutions Thoerem ("red theorem") are all equivalent, so we can put the two lists together to make a giant list of equivalent statements, on the next page. (The third list, in blue, comes from combining the corresponding green and red statements.) (Re the last line: you proved on ex. sheet #9 Q2c,d that it implies the higher lines; exercise: prove that the higher lines imply it.)

Theorem 8: Invertible Matrix Theorem (IMT): For a square $n \times n$ matrix A, the following are equivalent:

rref(A) has a pivot in every row.	rref(A) has a pivot in every column.	rref$(A) = I_n$.
$A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.	$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	$A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b}$ in $\mathbb{R}^n$.
The columns of A span $\mathbb{R}^n$.	The columns of A are linearly independent.	
The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.	The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is an invertible function.
There is a matrix D such that $AD = I_n$.	There is a matrix C such that $CA = I_n$.	A is invertible.

We will add more statements to the Invertible Matrix Theorem throughout the class.

Important consequences:

- line 3: A set of n vectors in $\mathbb{R}^n$ span $\mathbb{R}^n$ if and only if the set is linearly independent.
- line 4: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ (i.e. same domain and codomain) is one-to-one if and only if it is onto.

Students' main difficulty with IMT (or other theorems from later in the class) is when to use them, i.e. which theorems will help with which proof questions. Some tips:

- Each theorem connects two ideas, e.g. IMT connects existence and uniqueness. When the given information is about one idea, and the conclusion you want is about the other idea, then the theorem may be useful.
- If the situation of the question fits the conditions of the theorem, then that theorem may be useful. E.g. if you see a square matrix, consider IMT.

Theorem 8: Invertible Matrix Theorem continued: A is invertible if and only if A^T is invertible. (Proof: from p18 $(A^T)^{-1} = (A^{-1})^T$.)

This means that the statements in the Invertible Matrix Theorem are equivalent to the corresponding statements with "row" instead of "column", for example:

- The columns of an $n \times n$ matrix are linearly independent if and only if its rows span $\mathbb{R}^n$. (This is in fact also true for rectangular matrices.)
- If A is a square matrix and $A\mathbf{x} = \mathbf{b}$ has a unique solution for some $\mathbf{b}$, then the rows of A are linearly independent.

Theorem 8: Invertible Matrix Theorem continued: A is invertible if and only if A^T is invertible. (Proof: from p18 $(A^T)^{-1} = (A^{-1})^T$.)

This means that the statements in the Invertible Matrix Theorem are equivalent to the corresponding statements with "row" instead of "column", for example:

- The columns of an $n \times n$ matrix are linearly independent if and only if its rows span $\mathbb{R}^n$. (This is in fact also true for rectangular matrices.)
- If A is a square matrix and $A\mathbf{x} = \mathbf{b}$ has a unique solution for some $\mathbf{b}$, then the rows of A are linearly independent.

Advanced application (important for probability):

Let A be a square matrix. If the entries in each column of A sum to 1, then there is a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{v}$.

Hint: $(A - I)^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{0}$.