Section 2.2

The Inverse of a Matrix

The Definition of Inverse

Recall: The multiplicative inverse (or reciprocal) of a nonzero number a is the number b such that $ab = 1$. We define the inverse of a matrix in almost the same way.

Definition

Let A be an $n \times n$ square matrix. We say A is **invertible** (or **nonsingular**) if there is a matrix B of the same size, such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

identity matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

In this case, B is the **inverse** of A, and is written A^{-1}.

Example

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

I claim $B = A^{-1}$. Check:

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

✔

Do there exist two matrices A and B such that AB is the identity, but BA is not? If so, find an example. (Where both products make sense.)

Yes. Take $A = \begin{pmatrix} 1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. $AB = 0$ yet $BA = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$.

However, if A and B are *square* matrices, then $AB = I_n$ implies $BA = I_n$.

The 2×2 case

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The **determinant** of A is the number

$$\det(A) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Facts:

1. If $\det(A) \neq 0$, then A is invertible and $A^{-1} = \dfrac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.
2. If $\det(A) = 0$, then A is not invertible.

Why 1?

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

So we get the identity by dividing by $ad - bc$.

Example

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 \cdot 4 - 2 \cdot 3 = -2 \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = -\frac{1}{2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}.$$

Theorem

If A is invertible, then $Ax = b$ has exactly one solution for every b, namely:

$$x = A^{-1}b.$$

Why? "Divide by A"!

$$Ax = b \rightsquigarrow A^{-1}(Ax) = A^{-1}b \rightsquigarrow (A^{-1}A)x = A^{-1}b \rightsquigarrow I_n x = A^{-1}b \rightsquigarrow x = A^{-1}b.$$

$I_n x = x$ for every x

Important

If A is invertible and you know its inverse, then the easiest way to solve $Ax = b$ is by "dividing by A":

$$x = A^{-1}b.$$

Example

Solve the system

$$\begin{aligned} 2x + 3y + 2z &= 1 \\ x \quad\quad + 3z &= 1 \\ 2x + 2y + 3z &= 1 \end{aligned} \qquad \text{using} \qquad \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 3 \\ 2 & 2 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} -6 & -5 & 9 \\ 3 & 2 & -4 \\ 2 & 2 & -3 \end{pmatrix}.$$

Answer: $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 3 \\ 2 & 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -6 & -5 & 9 \\ 3 & 2 & -4 \\ 2 & 2 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$

The advantage of using inverses is it doesn't matter what's on the right-hand side of the $=$:

$$\begin{cases} 2x + 3y + 2z = b_1 \\ x \quad\quad + 3z = b_2 \\ 2x + 2y + 3z = b_3 \end{cases} \Longrightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 3 \\ 2 & 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Some Facts

Say A and B are invertible $n \times n$ matrices.

1. A^{-1} is invertible and its inverse is $(A^{-1})^{-1} = A$.
2. AB is invertible and its inverse is $(AB)^{-1} = \sout{A^{-1}B^{-1}}$ $B^{-1}A^{-1}$.

   Why? $(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$.
3. A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

   Why? $A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$.

Poll

If A, B, C are invertible $n \times n$ matrices, what is the inverse of ABC?

i. $A^{-1}B^{-1}C^{-1}$ ii. $B^{-1}A^{-1}C^{-1}$ iii. $C^{-1}B^{-1}A^{-1}$ iv. $C^{-1}A^{-1}B^{-1}$

It's (iii):

$$(ABC)(C^{-1}B^{-1}A^{-1}) = AB(CC^{-1})B^{-1}A^{-1} = A(BB^{-1})A^{-1}$$
$$= AA^{-1} = I_n.$$

In general, a product of invertible matrices is invertible, and the inverse is the product of the inverses, in the *reverse order*.

Computing A^{-1}

Let A be an $n \times n$ matrix. Here's how to compute A^{-1}.

1. Row reduce the augmented matrix $(\, A \mid I_n \,)$.
2. If the result has the form $(\, I_n \mid B \,)$, then A is invertible and $B = A^{-1}$.
3. Otherwise, A is not invertible.

Example

$$A = \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix}$$

[interactive]

Computing A^{-1}

Example

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_3 = R_3 + 3R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 3 & 1 \end{array}\right)$$

$$\xrightarrow{\substack{R_1 = R_1 - 2R_3 \\ R_2 = R_2 - R_3}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -6 & -2 \\ 0 & 1 & 0 & 0 & -2 & -1 \\ 0 & 0 & 2 & 0 & 3 & 1 \end{array}\right)$$

$$\xrightarrow{R_3 = R_3 \div 2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -6 & -2 \\ 0 & 1 & 0 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 & 3/2 & 1/2 \end{array}\right)$$

So $\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -6 & -2 \\ 0 & -2 & -1 \\ 0 & 3/2 & 1/2 \end{pmatrix}$.

Check: $\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix} \begin{pmatrix} 1 & -6 & -2 \\ 0 & -2 & -1 \\ 0 & 3/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ ✔

Why Does This Work?

First answer: We can think of the algorithm as simultaneously solving the equations

$$Ax_1 = e_1 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$Ax_2 = e_2 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$Ax_3 = e_3 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

Now note $A^{-1}e_i = A^{-1}(Ax_i) = x_i$, and x_i is the ith column in the augmented part. Also $A^{-1}e_i$ is the ith column of A^{-1}.

Second answer: Elementary matrices.

Elementary Matrices

Definition

An **elementary matrix** is a square matrix E which differs from I_n by one row operation.

There are three kinds, corresponding to the three elementary row operations:

scaling $(R_2 = 2R_2)$	row replacement $(R_2 = R_2 + 2R_1)$	swap $(R_1 \longleftrightarrow R_2)$
$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Fact: if E is the elementary matrix for a row operation, then EA differs from A by the same row operation.

Example:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 \\ 2 & 1 & 10 \\ 0 & -3 & -4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix} \xrightarrow{R_2 = R_2 + 2R_1} \begin{pmatrix} 1 & 0 & 4 \\ 2 & 1 & 10 \\ 0 & -3 & -4 \end{pmatrix}$$

Elementary Matrices

Continued

Fact: if E is the elementary matrix for a row operation, then EA differs from A by the same row operation.

Consequence

Elementary matrices are invertible, and the inverse is the elementary matrix which un-does the row operation.

$$\underset{R_2 = R_2 \times 2}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}} = \underset{R_2 = R_2 \div 2}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}} \qquad \underset{R_2 = R_2 + 2R_1}{\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}} = \underset{R_2 = R_2 - 2R_1}{\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}$$

$$\underset{R_1 \longleftrightarrow R_2}{\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}} = \underset{R_1 \longleftrightarrow R_2}{\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}}$$

Theorem

An $n \times n$ matrix A is invertible if and only if it is row equivalent to I_n. In this case, the sequence of row operations taking A to I_n also takes I_n to A^{-1}.

Why? Say the row operations taking A to I_n have elementary matrices $E_1, E_2, \ldots, E_k$. So

$$\text{note the order!} \longrightarrow E_k E_{k-1} \cdots E_2 E_1 A = I_n$$
$$\implies E_k E_{k-1} \cdots E_2 E_1 A A^{-1} = A^{-1}$$
$$\implies E_k E_{k-1} \cdots E_2 E_1 I_n = A^{-1}.$$

This means if you do these same row operations to A and to I_n, you'll end up with I_n and A^{-1}. This is what you do when you row reduce the augmented matrix:

$$\left(A \mid I_n \right) \rightsquigarrow \left(I_n \mid A^{-1} \right)$$