Inverses

Learning Goal: Students see the role of the inverse in solving systems, and start to look at the theory and computation of inverses.

We have noted the matrix $I$ that has the property that for any matrix $A$, $AI = IA = A$. So $I$ is the identity for matrix multiplication. Now in solving a linear equation in one variable $ax = b$, we use the fact that $a$ has a multiplicative inverse, $a^{-1}$ and that $a^{-1}a = 1$: $a^{-1}ax = a^{-1}b$, or $x = a^{-1}b$. Wouldn’t it be nice if there was a similar concept for matrix multiplication? Then we could easily solve all of our linear equations!

Definition

If $A$ is a square matrix, and there is some matrix $B$ so that $AB = BA = I$, then $A$ is called invertible and $B$ is called $A^{-1}$.

As we shall see, the case for non-square matrices is somewhat more complex.

Now, just like not all real numbers have multiplicative inverses (0 being the exception), not all matrices have inverses (there are many exceptions). One of the goals of studying matrices is to determine when a matrix has an inverse.

Note that we have not proven that the matrix $B$ is unique. Or even that it might only work on one side of $A$, since matrix multiplication isn’t commutative. And we haven’t said anything about how to find it. So we have a lot of work ahead.

Inverse of a product

Let’s say that $A$ and $B$ are both invertible, with inverses $A^{-1}$ and $B^{-1}$ respectively. Can we find an inverse for combinations of $A$ and $B$? It turns out that usually $(A + B)^{-1}$ is out of reach, just like you don’t know whether $(a + b)$ has a reciprocal unless you know something about $a$ and $b$. But the reciprocal of $ab$ is easy to find. Similarly, $(AB)^{-1}$ is easy to find: it is $B^{-1}A^{-1}$.

Checking this is as easy as computing, using the associative rule for matrix multiplication: $(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = AA^{-1} = I$, and similarly for $B^{-1}A^{-1}(AB)$ (recall that we have to check on both sides because matrix multiplication is not commutative).

This is a common thing mathematically: if you compose a bunch of operations, the inverse of the composition is the inverses of the individual operations in reverse order.

Inverses of elementary matrices

We have already seen inverses of elementary matrices. Remember that to find the matrix that accomplishes one of our elementary row operations, perform that operation on $I$. The three types of elementary matrices and their inverses are:

- $E_{ij}$ is the matrix that adds a multiple, $c$, of row $j$ to row $i$ by taking $I$ and putting $c$ in the $ij$-position. It’s inverse subtracts $c$ times row $j$ from row $i$, or adds $-c$ times row $j$ to row $i$, so $E_{ij}^{-1}$ is just $I$ with $-c$ in position $ij$.
- $P_{ij}$ is the matrix that swaps row $i$ and row $j$, so it is $I$ with rows $i$ and $j$ swapped. Its inverse must swap them back. In other words, $P_{ij}$ is its own inverse.
- $D_{ij}$ is the matrix that multiplies a row by a non-zero constant, $c$. It is $I$ with the 1 in row $j$ replaced with $c$. To undo a multiplication, divide! So $D_{ij}^{-1}$ must by $I$ with $1/c$ in row $j$. 


**Left and Right Inverses**

Our definition of an inverse requires that it work on both sides of \( A \). Since matrix multiplication is not commutative, it is conceivable that some matrix may only have an inverse on one side or the other. If \( BA = I \) then \( B \) is a *left inverse* of \( A \) and \( A \) is a *right inverse* of \( B \). We will later show that for square matrices, the existence of any inverse on either side is equivalent to the existence of a unique two-sided inverse. For now, we content ourselves with:

**Theorem:** if \( A \) has any left inverse \( B \) and any right inverse \( C \), then \( B = C \) and in fact this is the *unique* two-sided inverse.

**Proof:** We have \( BA = I \) and \( AC = I \). So \( B = BI = B(AC) = (BA)C = IC = C \) so \( B = C \) as advertised. Thus \( B \) is a two-sided inverse. Now let \( D \) be an inverse on either side. If the left, then \( D = C \); if the right, then \( D = B \)—both of these are shown in the same way we proved \( B = C \) above. So any inverse equals \( C \) which equals \( B \)—that is, they’re all the same and work on either side.

**Finding an inverse**

One method of finding an inverse is to solve the systems \( Ax_1 = e_1, Ax_2 = e_2, \ldots, Ax_n = e_n \). (Recall that \( e_j = (0, 0, \ldots, 1, \ldots, 0) \) with the 1 in the \( j \)th position.) Why is this? Well, if we put the \( x \) vectors into a matrix, we find that \( A[x_1 \ x_2 \ \cdots \ x_n] = [e_1 \ e_2 \ \cdots \ e_n] = I \), so the matrix of \( x \)’s is a right inverse of \( A \).

We’ll work by example to invert the matrix \[
\begin{bmatrix}
1 & 2 & 3 \\
2 & 4 & 5 \\
-1 & 1 & 1
\end{bmatrix}
\]

augmented matrix: \[
\begin{bmatrix}
1 & 2 & 3 & 1 & 0 & 0 \\
2 & 4 & 5 & 0 & 1 & 0 \\
-1 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}
\]

Next, we subtract twice the first row from the second to yield \[
\begin{bmatrix}
1 & 2 & 3 & 1 & 0 & 0 \\
-2 & 1 & 0 & 2 & 4 & 5 \\
0 & 0 & 1 & -1 & 1 & 0
\end{bmatrix}
\]

We now have a zero in the pivot position, so we must do a row swap using \( P_{23} = \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 3 & 4
\end{bmatrix}
\]

to get \[
\begin{bmatrix}
1 & 2 & 3 & 1 & 0 & 0 \\
0 & 3 & 4 & 1 & 0 & 1 \\
0 & 0 & -1 & -2 & 1 & 0
\end{bmatrix}
\]. Under normal circumstances we would start the back-substitution here, but for reasons that will become clear in a second, let’s continue.
the elimination upwards. The next step will multiply by \( E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \) and then

\[
E_{13} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ to obtain } \begin{bmatrix} 1 & 2 & 0 & | & -5 & 3 & 0 \\ 0 & 3 & 0 & | & -7 & 4 & 1 \\ 0 & 0 & -1 & | & -2 & 1 & 0 \end{bmatrix} . \]

The next step is to multiply on the left by

\[
E_{12} = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ which gives us } \begin{bmatrix} 1 & 0 & 0 & | & -1/3 & 1/3 & -2/3 \\ 0 & 3 & 0 & | & -7 & 4 & 1 \\ 0 & 0 & -1 & | & -2 & 1 & 0 \end{bmatrix} . \]

Finally, I will multiply by matrices \( D_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \) and \( D_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \) to multiply each row by the reciprocal of the pivot to reach

\[
\begin{bmatrix} 0 & 1 & 0 & | & -7/3 & 4/3 & 1/3 \\ \hline 0 & 0 & 1 & | & 2 & -1 & 0 \end{bmatrix} . \]

Now let \( B \) be the product of all the \( D \)'s, \( P \)'s, and \( E \)'s in the order (right to left) in which they were used: \( B = D_3D_2E_{12}E_{13}E_{23}P_{23}E_{31}E_{21} \). Then \( B \) is a two sided inverse, and is the matrix which appears to the right of the dashed line in the final augmented matrix.

Why does this work? We can easily check the multiplication \( AB = BA = I \) for this case. But we seek to understand why the method works. And the explanation is simple.

Let \( A \) be any square matrix and form the augmented matrix \([A | I] \). We applied a bunch of row operations to \( A \) until we arrived at \( I \). These row operations could be collected above as \( B \).

Thus \( BA = I \) and \( B \) is a left inverse of \( A \). By block multiplication \( B[A | I] = [BA | BI] = [I | B] \).

So, as claimed, the matrix \( B \) appears to the right of the dashed line in the augmented matrix.

(Another way to think about it: apply the operations to \( I \), and we transform \( I \) to \( B \) as \( A \) is being transformed to \( I \).)

But why is it a two-sided inverse? Well, if we had stopped and done back-substitution, because \( A \) had all its pivots, we could get solutions to each of \( Ax = e_j \). We have already pointed out that if we put these \( x \)'s as the columns of a matrix, this matrix is a right inverse of \( A \). So \( A \) has both a left inverse and a right inverse, and as we showed above, that makes them equal and unique.

This continued elimination process is called Gauss-Jordan elimination. It can be employed to solve systems (because it really eliminates each variable from all equations but one, so you can just read off the solution—it is the “RREF” (row-reduced echelon form) technique), but it is not as efficient as back-substitution.

Invertibility

We state some facts about square invertible matrices.

- Square \( A \) has a two-sided inverse if and only if it has all of its pivots. For if \( A \) has all of its pivots, elimination proceeds smoothly, and we can solve \( Ax = b \) for any right-hand side—including \( b = e_j \)—so we can assemble these into a right inverse. But we can also
continue Gauss-Jordan elimination to find a left inverse, hence we have a two-sided inverse. On the other hand, if \( A \) is missing any pivot then elimination will eventually produce a zero row. If this was row \( i \), then there is no way we could solve \( Ax = e_i \), so there certainly can’t be a right inverse, so no two-sided inverse either!

- \( A \) has a two-sided inverse if and only if \( Ax = 0 \) has the unique solution \( x = 0 \). If \( A \) is invertible and \( Ax = 0 \), then \( A^{-1}Ax = A^{-1}0 \). But this simplifies to \( x = 0 \), so \( 0 \) is the unique solution. On the other hand, if you have a unique solution, you must have a complete set of pivots, so by the previous comment \( A \) has a two-sided inverse.

- \( A \) has a two-sided inverse if and only if \( Ax = b \) has a unique solution for each right-hand side \( b \). For if \( A \) is invertible, then \( x = A^{-1}Ax = A^{-1}b \) is the only possible solution, and it really is a solution because \( A(A^{-1}b) = Ib = b \). Conversely, if \( Ax = b \) has a unique solution for all \( b \), it has one for \( b = 0 \) and the previous comment applies to show \( A \) is invertible.

- If \( A \) has a one-sided inverse on either side, then \( A \) is invertible, and in fact this one-sided inverse is \( A \)'s unique two-sided inverse. For let \( A \) have right inverse \( C \), so the \( AC = I \). Now begin elimination on \( A \). Let \( M \) be the matrix that collects all the steps of the forward elimination (i.e., not the Jordan part where we eliminate upward). As we have seen, \( M \) is a product of elementary matrices, each of which is invertible, so the product \( M \) is also invertible. If \( A \) doesn’t have all of its pivots, elimination will produce a zero row. That is, \( MA \) will have a zero row. But then \( (MA)C \) will have this same row zero. This cause \( M(AC) = MI = M \) to have a zero row. But this can’t be because then \( MM^{-1} \) will have this same zero row, but \( MM^{-1} = I \)! So \( A \) must have all of its pivots, and thus be invertible—thus \( C \) is actually a two-sided inverse for \( A \). On the other hand, if \( B \) is a left inverse for \( A \) so that \( BA = I \), then \( A \) is a right inverse for \( B \), and the preceding argument shows that \( A \) is actually a two-sided inverse for \( B \), so \( AB = I \) and \( B \) is a right inverse of \( A \)!

To summarize, \( A \) is invertible if and only if it has a full set of pivots. If it has an inverse on either side, it is invertible with a unique two-sided inverse. Because it has a full set of pivots, the system \( Ax = b \) is always non-singular. Invertibility is equivalent to having a unique solution for any equation \( Ax = b \).

The problem with non-square matrices is that a matrix with more rows than columns is guaranteed to have zero rows after elimination (thus is inconsistent for some right-hand sides), and one with more columns than rows is guaranteed to have non-unique solutions when it has any at all. These matrices may have one-sided inverses, though. We will come to this later.

Note especially the example in the text. The inverse of a lower triangular matrix is itself lower triangular. Why? Because Gauss-Jordan elimination only needs to work downward. Similarly, in an upper triangular matrix we only need to work upward, so its inverse will be of the same form. Many properties of matrices are shared by their inverses.

Reading: 2.5
Problems: 2.5: 3, 4, 5, 7, 8, 9, 10 – 13, 18, 21, 23, 24, 25, 29, 32, 33, 35, 40, 42*