- The under-determined linear equations $A^T y = (1, 2)$ has (different) solutions

  $$B^T(1,2) = (1/3, 2/3, 38/9), \qquad C^T(1,2) = (0, 1/2, -1).$$

  (Recall that B^T and C^T are both right inverses of A^T.) We can find a solution of $A^T y = b$ for any vector b.

Left and right inverse of matrix product. Suppose A and D are compatible for the matrix product AD (*i.e.*, the number of columns in A is equal to the number of rows in D.) If A has a right inverse B and D has a right inverse E, then EB is a right inverse of AD. This follows from

$$(AD)(EB) = A(DE)B = A(IB) = AB = I.$$

If A has a left inverse C and D has a left inverse F, then FC is a left inverse of AD. This follows from

$$(FC)(AD) = F(CA)D = FD = I.$$

11.2 Inverse

If a matrix is left- *and* right-invertible, then the left and right inverses are unique and equal. To see this, suppose that $AX = I$ and $YA = I$, *i.e.*, X is any right inverse and Y is any left inverse of A. Then we have

$$X = (YA)X = Y(AX) = Y,$$

i.e., any left inverse of A is equal to any right inverse of A. This implies that the left inverse is unique: If we have $A\tilde{X} = I$, then the argument above tells us that $\tilde{X} = Y$, so we have $\tilde{X} = X$, *i.e.*, there is only one right inverse of A. A similar argument shows that Y (which is the same as X) is the only left inverse of A.

When a matrix A has both a left inverse Y and a right inverse X, we call the matrix $X = Y$ simply the *inverse* of A, and denote it as A^{-1}. We say that A is *invertible* or *nonsingular*. A square matrix that is not invertible is called *singular*.

Dimensions of invertible matrices. Invertible matrices must be square, since tall matrices are not right-invertible, while wide matrices are not left-invertible. A matrix A and its inverse (if it exists) satisfy

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

If A has inverse A^{-1}, then the inverse of A^{-1} is A; in other words, we have $(A^{-1})^{-1} = A$. For this reason we say that A and A^{-1} are inverses (of each other).

Solving linear equations with the inverse. Consider the square system of n linear equations with n variables, $Ax = b$. If A is invertible, then for any n-vector b,

$$x = A^{-1}b \tag{11.1}$$

is a solution of the equations. (This follows since A^{-1} is a right inverse of A.) Moreover, it is the *only* solution of $Ax = b$. (This follows since A^{-1} is a left inverse of A.) We summarize this very important result as

> *The square system of linear equations $Ax = b$, with A invertible, has the unique solution $x = A^{-1}b$, for any n-vector b.*

One immediate conclusion we can draw from the formula (11.1) is that the solution of a square set of linear equations is a linear function of the right-hand side vector b.

Invertibility conditions. For square matrices, left-invertibility, right-invertibility, and invertibility are equivalent: If a matrix is square and left-invertible, then it is also right-invertible (and therefore invertible) and vice-versa.

To see this, suppose A is an $n \times n$ matrix and left-invertible. This implies that the n columns of A are linearly independent. Therefore they form a basis and so any n-vector can be expressed as a linear combination of the columns of A. In particular, each of the n unit vectors e_i can be expressed as $e_i = Ab_i$ for some n-vector b_i. The matrix $B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$ satisfies

$$AB = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_n \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} = I.$$

So B is a right inverse of A.

We have just shown that for a square matrix A,

$$\text{left-invertibility} \quad \Longrightarrow \quad \text{column independence} \quad \Longrightarrow \quad \text{right-invertibility}.$$

(The symbol $\Longrightarrow$ means that the left-hand condition implies the right-hand condition.) Applying the same result to the transpose of A allows us to also conclude that

$$\text{right-invertibility} \quad \Longrightarrow \quad \text{row independence} \quad \Longrightarrow \quad \text{left-invertibility}.$$

So all six of these conditions are equivalent; if any one of them holds, so do the other five.

In summary, for a square matrix A, the following are equivalent.

- A is invertible.
- The columns of A are linearly independent.
- The rows of A are linearly independent.
- A has a left inverse.
- A has a right inverse.

Examples.

- The identity matrix I is invertible, with inverse $I^{-1} = I$, since $II = I$.
- A diagonal matrix A is invertible if and only if its diagonal entries are nonzero. The inverse of an $n \times n$ diagonal matrix A with nonzero diagonal entries is
$$A^{-1} = \begin{bmatrix} 1/A_{11} & 0 & \cdots & 0 \\ 0 & 1/A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/A_{nn} \end{bmatrix},$$
since
$$AA^{-1} = \begin{bmatrix} A_{11}/A_{11} & 0 & \cdots & 0 \\ 0 & A_{22}/A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn}/A_{nn} \end{bmatrix} = I.$$
In compact notation, we have
$$\mathbf{diag}(A_{11}, \ldots, A_{nn})^{-1} = \mathbf{diag}(A_{11}^{-1}, \ldots, A_{nn}^{-1}).$$
Note that the inverse on the left-hand side of this equation is the matrix inverse, while the inverses appearing on the right-hand side are scalar inverses.
- As a non-obvious example, the matrix
$$A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 2 \\ -3 & -4 & -4 \end{bmatrix}$$
is invertible, with inverse
$$A^{-1} = \frac{1}{30} \begin{bmatrix} 0 & -20 & -10 \\ -6 & 5 & -2 \\ 6 & 10 & 2 \end{bmatrix}.$$
This can be verified by checking that $AA^{-1} = I$ (or that $A^{-1}A = I$, since either of these implies the other).
- 2×2 *matrices.* A 2×2 matrix A is invertible if and only if $A_{11}A_{22} \neq A_{12}A_{21}$, with inverse
$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$
(There are similar formulas for the inverse of a matrix of any size, but they grow very quickly in complexity and so are not very useful in most applications.)
- *Orthogonal matrix.* If A is square with orthonormal columns, we have $A^TA = I$, so A is invertible with inverse $A^{-1} = A^T$.

Inverse of matrix transpose. If A is invertible, its transpose A^T is also invertible and its inverse is $(A^{-1})^T$:

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

Since the order of the transpose and inverse operations does not matter, this matrix is sometimes written as A^{-T}.

Inverse of matrix product. If A and B are invertible (hence, square) and of the same size, then AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}. \qquad (11.2)$$

The inverse of a product is the product of the inverses, in reverse order.

Dual basis. Suppose that A is invertible with inverse $B = A^{-1}$. Let $a_1, \ldots, a_n$ be the columns of A, and $b_1^T, \ldots, b_n^T$ denote the *rows* of B, *i.e.*, the columns of B^T:

$$A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}, \qquad B = \begin{bmatrix} b_1^T \\ \vdots \\ b_n^T \end{bmatrix}.$$

We know that $a_1, \ldots, a_n$ form a basis, since the columns of A are linearly independent. The vectors $b_1, \ldots, b_n$ also form a basis, since the rows of B are linearly independent. They are called the *dual basis* of $a_1, \ldots, a_n$. (The dual basis of $b_1, \ldots, b_n$ is $a_1, \ldots, a_n$, so they called *dual bases*.)

Now suppose that x is any n-vector. It can be expressed as a linear combination of the basis vectors $a_1, \ldots, a_n$:

$$x = \beta_1 a_1 + \cdots + \beta_n a_n.$$

The dual basis gives us a simple way to find the coefficients $\beta_1, \ldots, \beta_n$.

We start with $AB = I$, and multiply by x to get

$$x = ABx = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1^T \\ \vdots \\ b_n^T \end{bmatrix} x = (b_1^T x)a_1 + \cdots + (b_n^T x)a_n.$$

This means (since the vectors $a_1, \ldots, a_n$ are linearly independent) that $\beta_i = b_i^T x$. In words: The coefficients in the expansion of a vector in a basis are given by the inner products with the dual basis vectors. Using matrix notation, we can say that $\beta = B^T x = (A^{-1})^T x$ is the vector of coefficients of x in the basis given by the columns of A.

As a simple numerical example, consider the basis

$$a_1 = (1, 1), \qquad a_2 = (1, -1).$$

The dual basis consists of the rows of $\begin{bmatrix} a_1 & a_2 \end{bmatrix}^{-1}$, which are

$$b_1^T = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}, \qquad b_2^T = \begin{bmatrix} 1/2 & -1/2 \end{bmatrix}.$$

To express the vector $x = (-5, 1)$ as a linear combination of a_1 and a_2, we have

$$x = (b_1^T x)a_1 + (b_2^T x)a_2 = (-2)a_1 + (-3)a_2,$$

which can be directly verified.

Negative matrix powers. We can now give a meaning to matrix powers with negative integer exponents. Suppose A is a square invertible matrix and k is a positive integer. Then by repeatedly applying property (11.2), we get

$$(A^k)^{-1} = (A^{-1})^k.$$

We denote this matrix as A^{-k}. For example, if A is square and invertible, then $A^{-2} = A^{-1}A^{-1} = (AA)^{-1}$. With A^0 defined as $A^0 = I$, the identity $A^{k+l} = A^k A^l$ holds for all integers k and l.

Triangular matrix. A triangular matrix with nonzero diagonal elements is invertible. We first discuss this for a lower triangular matrix. Let L be $n \times n$ and lower triangular with nonzero diagonal elements. We show that the columns are linearly independent, *i.e.*, $Lx = 0$ is only possible if $x = 0$. Expanding the matrix-vector product, we can write $Lx = 0$ as

$$\begin{array}{rcl}
L_{11}x_1 & = & 0 \\
L_{21}x_1 + L_{22}x_2 & = & 0 \\
L_{31}x_1 + L_{32}x_2 + L_{33}x_3 & = & 0 \\
& \vdots & \\
L_{n1}x_1 + L_{n2}x_2 + \cdots + L_{n,n-1}x_{n-1} + L_{nn}x_n & = & 0.
\end{array}$$

Since $L_{11} \neq 0$, the first equation implies $x_1 = 0$. Using $x_1 = 0$, the second equation reduces to $L_{22}x_2 = 0$. Since $L_{22} \neq 0$, we conclude that $x_2 = 0$. Using $x_1 = x_2 = 0$, the third equation now reduces to $L_{33}x_3 = 0$, and since L_{33} is assumed to be nonzero, we have $x_3 = 0$. Continuing this argument, we find that all entries of x are zero, and this shows that the columns of L are linearly independent. It follows that L is invertible.

A similar argument can be followed to show that an upper triangular matrix with nonzero diagonal elements is invertible. One can also simply note that if R is upper triangular, then $L = R^T$ is lower triangular with the same diagonal, and use the formula $(L^T)^{-1} = (L^{-1})^T$ for the inverse of the transpose.

Inverse via QR factorization. The QR factorization gives a simple expression for the inverse of an invertible matrix. If A is square and invertible, its columns are linearly independent, so it has a QR factorization $A = QR$. The matrix Q is orthogonal and R is upper triangular with positive diagonal entries. Hence Q and R are invertible, and the formula for the inverse product gives

$$A^{-1} = (QR)^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T. \tag{11.3}$$

In the following section we give an algorithm for computing R^{-1}, or more directly, the product $R^{-1}Q^T$. This gives us a method to compute the matrix inverse.

11.3 Solving linear equations

Back substitution. We start with an algorithm for solving a set of linear equations, $Rx = b$, where the $n \times n$ matrix R is upper triangular with nonzero diagonal entries (hence, invertible). We write out the equations as

$$\begin{aligned} R_{11}x_1 + R_{12}x_2 + \cdots + R_{1,n-1}x_{n-1} + R_{1n}x_n &= b_1 \\ &\vdots \\ R_{n-2,n-2}x_{n-2} + R_{n-2,n-1}x_{n-1} + R_{n-2,n}x_n &= b_{n-2} \\ R_{n-1,n-1}x_{n-1} + R_{n-1,n}x_n &= b_{n-1} \\ R_{nn}x_n &= b_n. \end{aligned}$$

From the last equation, we find that $x_n = b_n/R_{nn}$. Now that we know x_n, we substitute it into the second to last equation, which gives us

$$x_{n-1} = (b_{n-1} - R_{n-1,n}x_n)/R_{n-1,n-1}.$$

We can continue this way to find $x_{n-2}, x_{n-3}, \ldots, x_1$. This algorithm is known as *back substitution*, since the variables are found one at a time, starting from x_n, and we substitute the ones that are known into the remaining equations.

Algorithm 11.1 BACK SUBSTITUTION

given an $n \times n$ upper triangular matrix R with nonzero diagonal entries, and an n-vector b.

For $i = n, \ldots, 1$,
    $x_i = (b_i - R_{i,i+1}x_{i+1} - \cdots - R_{i,n}x_n)/R_{ii}$.

(In the first step, with $i = n$, we have $x_n = b_n/R_{nn}$.) The back substitution algorithm computes the solution of $Rx = b$, *i.e.*, $x = R^{-1}b$. It cannot fail since the divisions in each step are by the diagonal entries of R, which are assumed to be nonzero.

Lower triangular matrices with nonzero diagonal elements are also invertible; we can solve equations with lower triangular invertible matrices using *forward substitution*, the obvious analog of the algorithm given above. In forward substitution, we find x_1 first, then x_2, and so on.

Complexity of back substitution. The first step requires 1 flop (division by R_{nn}). The next step requires one multiply, one subtraction, and one division, for a total of 3 flops. The kth step requires $k-1$ multiplies, $k-1$ subtractions, and one division, for a total of $2k-1$ flops. The total number of flops for back substitution is then

$$1 + 3 + 5 + \cdots + (2n-1) = n^2$$

flops.

This formula can be obtained from the formula (5.7), or directly derived using a similar argument. Here is the argument for the case when n is even; a similar