Inverses and Determinants

I. Inverses.

Recall how to invert the linear filter

$$y_1 = ax_1 + bx_2,$$
$$y_2 = cx_1 + dx_2.$$

We form the equation $ae_2 - ce_1$, which is

$$ay_2 - cy_1 = (ad - bc)x_2.$$

If $ad - bc = 0$, then invertibility fails. If $ad - bc \neq 0$, then

$$x_2 = \frac{ay_2 - cy_1}{ad - bc},$$

and we can back substitute, or just similarly write $by_2 - dy_1 = (bc - ad)x_1$, to get

$$x_1 = \frac{dy_1 - by_2}{ad - bc}.$$

This shows that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Let's now go through the analogous process for 3×3 matrices. Consider the linear filter

$$ax_1 + bx_2 + cx_3 = y_1,$$
$$dx_1 + ex_2 + fx_3 = y_2,$$
$$gx_1 + hx_2 + ix_3 = y_3.$$

To invert it, we must solve for x_1, x_2, x_3, writing

$$x_1 = jy_1 + ky_2 + \ell y_3,$$
$$x_2 = my_1 + ny_2 + py_3,$$
$$x_3 = qy_1 + ry_2 + sy_3,$$

with the inverse existing if and only if this is possible.

In terms of augmented matrices, this amounts to putting

$$\left[\begin{array}{ccc|ccc} a & b & c & 1 & 0 & 0 \\ d & e & f & 0 & 1 & 0 \\ g & h & i & 0 & 0 & 1 \end{array}\right]$$

into the reduced row echelon form

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & j & k & \ell \\ 0 & 1 & 0 & m & n & p \\ 0 & 0 & 1 & q & r & s \end{array}\right],$$

Kiril Datchev, March 7, 2025. Questions, comments, and corrections are gratefully received at kdatchev@purdue.edu.

with the inverse existing if and only if this is the correct reduced row echelon form, i.e. if and only if each row has a pivot. In this situation,

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \begin{bmatrix} j & k & \ell \\ m & n & p \\ q & r & s \end{bmatrix}.$$

The same method works for $n \times n$ matrices for any n, but when $n \geq 3$ the analog of the formula $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ is much more complicated so it is generally better to use row operations to get to row echelon form.

Example 1.

Let's find the values of α for which

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & \alpha \end{bmatrix}$$

is invertible, and invert it for one of those values. Clearing the first column by applying $R_2 \to R_2 - R_1$ and $R_3 \to R_3 - R_1$ to

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & -2 & 0 & 0 & 1 & 0 \\ 1 & -1 & \alpha & 0 & 0 & 1 \end{array}\right]$$

yields

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -3 & -1 & -1 & 1 & 0 \\ 0 & -2 & \alpha - 1 & -1 & 0 & 1 \end{array}\right].$$

Clearing the second column by applying $R_1 \to 3R_1 + R_2$ and $R_3 \to 3R_3 - 2R_2$ gives

$$\left[\begin{array}{ccc|ccc} 3 & 0 & 2 & 2 & 1 & 0 \\ 0 & -3 & -1 & -1 & 1 & 0 \\ 0 & 0 & 3\alpha - 1 & -1 & -2 & 3 \end{array}\right].$$

We now see that A is invertible if and only if $\alpha \neq 1/3$.

Let's choose $\alpha = 2/3$ for simplicity. Clearing the third column by applying $R_1 \to R_1 - 2R_3$ and $R_2 \to R_2 + R_3$ gives

$$\left[\begin{array}{ccc|ccc} 3 & 0 & 0 & 4 & 5 & -6 \\ 0 & -3 & 0 & -2 & -1 & 3 \\ 0 & 0 & 1 & -1 & -2 & 3 \end{array}\right].$$

Finally, divide the first row by 3 and the second by -3 to reach reduced row echelon form:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4/3 & 5/3 & -2 \\ 0 & 1 & 0 & 2/3 & 1/3 & 1 \\ 0 & 0 & 1 & -1 & -2 & 3 \end{array}\right].$$

That gives

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

Let's check:

$$AA^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & 2/3 \end{bmatrix} \begin{bmatrix} 4/3 & 5/3 & -2 \\ 2/3 & 1/3 & 1 \\ -1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Exercise 1.

Invert

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

using this method, and check that this gives the same answer as using the formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Exercise 2.

Invert

$$\begin{bmatrix} 1 & -1 & 2 \\ -2 & 3 & -2 \\ 4 & -3 & 3 \end{bmatrix}.$$

II. Determinants.

The determinant of a 2×2 matrix is defined by

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Each of the three fundamental row operations has a simple effect on the determinant:

(1) Multiplying a row by a constant multiplies the determinant by the same constant:

$$\det \begin{bmatrix} \lambda a & \lambda b \\ c & d \end{bmatrix} = \lambda ad - \lambda bc = \lambda(ad - bc) = \lambda \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{1}$$

(2) Adding a constant mutiple of one row to another results in no change:

$$\det \begin{bmatrix} a + \lambda c & b + \lambda d \\ c & d \end{bmatrix} = (a + \lambda c)d - (b + \lambda d)c = ad - bc = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{2}$$

(3) Switching the rows flips the sign:

$$\det \begin{bmatrix} c & d \\ a & b \end{bmatrix} = cb - da = -(ad - bc) = -\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

(4) Moreover, the determinant of a matrix in reduced row echelon form is 1 if the matrix is the identity and 0 otherwise; i.e.

$$\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1, \quad \det \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \det \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \det \begin{bmatrix} 1 & b \\ 0 & 0 \end{bmatrix} = 0.$$

We define the determinant of a more general $n \times n$ matrix in such a way as to preserve these properties. Thus, one can compute the determinant of any square matrix by bringing the matrix to reduced row echelon form, keeping track of the row operations used along the way

Example 2. Let's check that if

$$A = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix},$$

then $\det A = adf$.

First, if d and f are both different from zero, then we can clear the off diagonal entries by taking $R_1 \to R_1 - gR_2 - hR_3$ and $R_2 \to R_2 - iR_3$, for some constants g, h, i. Since those operations don't affect the determinant (as in equation (2) above) we get

$$\det A = \det \begin{bmatrix} a & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & f \end{bmatrix}.$$

Next, constants multiplied to entire rows come out, as in equation (1), giving

$$\det \begin{bmatrix} a & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & f \end{bmatrix} = a \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & f \end{bmatrix} = ad \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & f \end{bmatrix} = adf \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = adf.$$

On the other hand, if either $d = 0$ or $f = 0$, then the reduced row echelon form of A is not the identity, so $\det A = 0 = adf$.

Example 3. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & \alpha \end{bmatrix}.$$

Applying row operations as in Example 1 yields

$$\det A = \det \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & \alpha \end{bmatrix} = \det \begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & -2 & \alpha - 1 \end{bmatrix}.$$

To keep the determinant unchanged, we take $R_3 \to R_3 - (2/3)R_2$ (rather than $R_3 \to 3R_3 - 2R_2$ as we did above):

$$\det A = \det \begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & \alpha - 1/3 \end{bmatrix}.$$

This brings us to the situation of Example 2, so

$$\det A = -3(\alpha - 1/3) = 1 - 3\alpha.$$

Example 4. A certain amount of work shows that

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

Exercise 3. Compute $\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ using row operations to get to reduced row echelon form, and confirm that this gives the same answer as using the formula $ad - bc$.

Exercise 4. Compute

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & \alpha \end{bmatrix}$$

as a function of α, using a combination of row operations and any of the results of the examples above that make the work easier.

III. Geometric meaning of determinants. The value of a 2×2 determinant is the signed area of the parallelogram spanned by the column vectors. The sign matches the sign of the angle from the first column to the second.

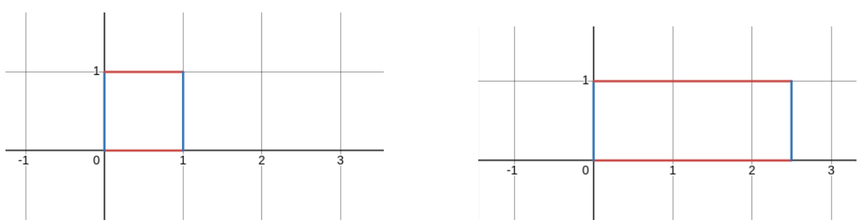
To check this, we check that the signed area obeys the same rules that the determinant does:

(1) Multiplying a row by a constant multiplies the signed area by the same constant because this corresponds to scaling by that constant in one direction.

This is easiest to see for a rectangle. For example, taking

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} 2.5 & 0 \\ 0 & 1 \end{bmatrix}$$

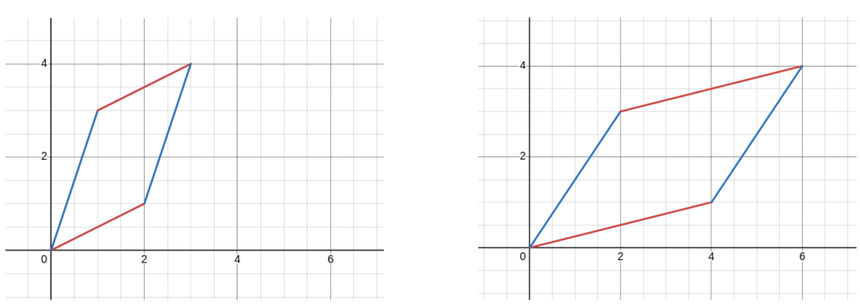
multiplies the area by 2.5:



The same thing works for a parallelogram. For example, taking

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

multiplies the area by 2:



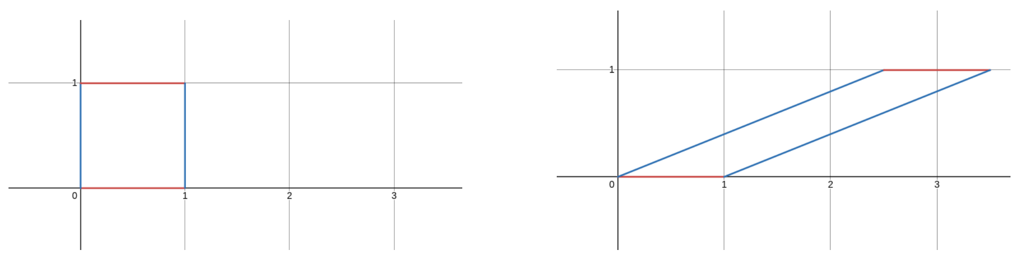
This can be checked by integration. If we compute the area by integrating first in x and then in y, in both cases the y integral has the same bounds (0 to 4 in the above pair of matrices) while for each y, the corresponding x interval is twice as long (for example, for $y = 2$ the range is 2/3 to 7/3 for the first matrix and 4/3 to 14/3 for the second matrix). For the figures, see here and here.

(2) Adding a constant mutiple of one row to another results in no change because this corresponds to shearing.

This is easiest to see for a rectangle. For example, taking

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} 1 & 2.5 \\ 0 & 1 \end{bmatrix}$$

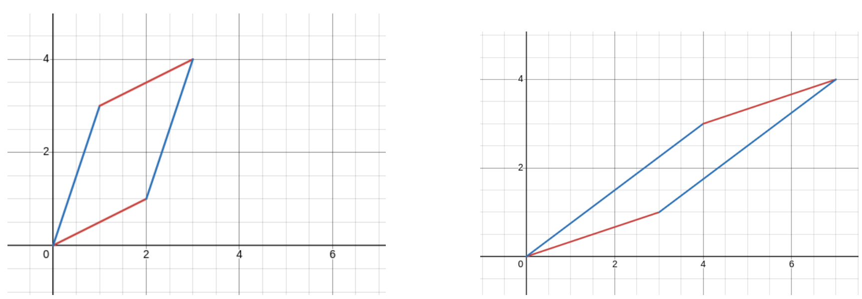
converts the square to a parallelogram having the same base and height:



It also works for parallelograms. For example, taking

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} 3 & 4 \\ 1 & 3 \end{bmatrix}$$

does not change the area:



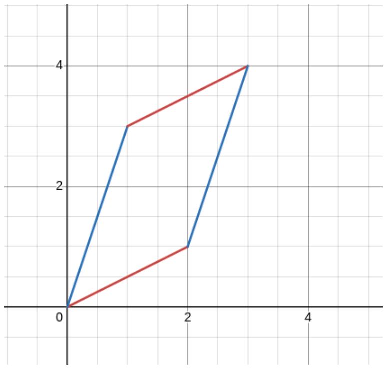
Once again, this can be checked by integration. If we compute the area by integrating first in x and then in y, in both cases the y integral has the same bounds (0 to 4 in the above pair of matrices) while for each y, the corresponding x interval is the same length (for example, for $y = 2$ the range is 2/3 to 7/3 for the first matrix and 7/3 to 13/3 for the second matrix). For the figures, see here and here.

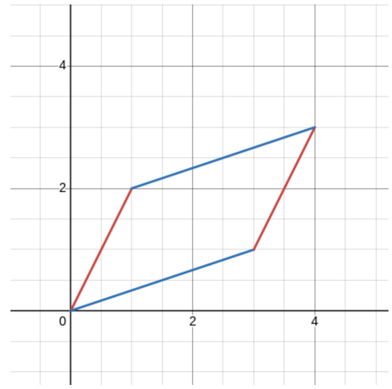
(3) Switching the rows flips the sign because the corresponding transformation is a reflection.

For example, taking

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$$

corresponds to





The parallelograms are congruent, but the angle from red to blue is positive in the first case and negative in the second. See here for the figures.

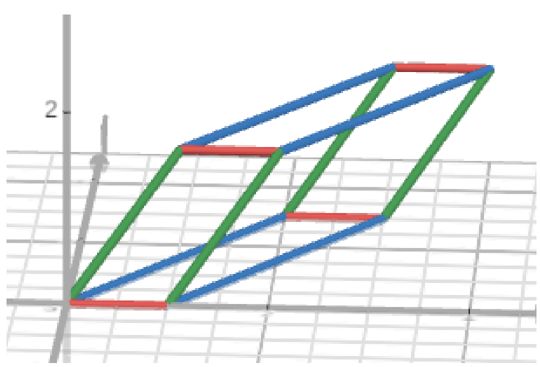
(4) Finally the signed area of the vectors spanned by the identity matrix is 1 because this is the unit square. For all other matrices in reduced row echelon form the area is zero because the y range is zero.

Since signed area of a parallelogram obeys the same four defining rules as the 2×2 determinant, it must give the same value.

For 3×3, we replaced signed area of a parallelogram by signed volume of a parallelepiped. Thus,

$$\det \begin{bmatrix} 1 & 2 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

gives the volume of the parallelepiped



because the red vector is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, the blue is $\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$, and the green is $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. See here for the figure.

The same principle applies to $n \times n$ matrices.

IV. Geometric criteria for invertibility/change of coordinates. The following are equivalent:

(1) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible

(2) $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \neq 0.$

(3) $\text{rref}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$

(4) We can solve the change of coordinates equation

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s_1 \begin{bmatrix} a \\ c \end{bmatrix} + s_2 \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} s_2 \\ s_2 \end{bmatrix}$$

uniquely for s_1 and s_2, given any x_1 and x_2.

(5) The vectors $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\begin{bmatrix} b \\ d \end{bmatrix}$ do not lie in a single line.

Condition (4) is another way of saying that $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\begin{bmatrix} b \\ d \end{bmatrix}$ span a nondegenerate parallelogram.

In three dimensions, the following analogous statements are equivalent:

(1) $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is invertible.

(2) $\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \neq 0.$

(3) $\text{rref}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

(4) We can solve the change of coordinates equation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = s_1 \begin{bmatrix} a \\ d \\ g \end{bmatrix} + s_2 \begin{bmatrix} b \\ e \\ h \end{bmatrix} + s_3 \begin{bmatrix} c \\ f \\ i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

uniquely for s_1, s_2, s_3, given any x_1, x_2, x_3.

(5) The vectors $\begin{bmatrix} a \\ d \\ g \end{bmatrix}$, $\begin{bmatrix} b \\ e \\ h \end{bmatrix}$, and $\begin{bmatrix} c \\ f \\ i \end{bmatrix}$ do not lie in a single plane.

Analogous equivalences hold in n dimensions.