Math 152: Linear Systems – Winter 2005

Section 3: Matrices and Determinants

Matrix operations

A matrix is a rectangular array of numbers. Here is an example of an $m \times n$ matrix.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

This is sometimes abbreviated $A = [a_{i,j}]$. An $m \times 1$ matrix is called a column vector and a $1 \times n$ matrix is called a row vector. (The convention is that $m \times n$ means m rows and n columns.

Addition and scalar multiplication are defined for matrices exactly as for vectors. If s is a number

$$s \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} = \begin{bmatrix} sa_{1,1} & sa_{1,2} & \cdots & sa_{1,n} \\ sa_{2,1} & sa_{2,2} & \cdots & sa_{2,n} \\ \vdots & \vdots & & \vdots \\ sa_{m,1} & sa_{m,2} & \cdots & sa_{m,n} \end{bmatrix},$$

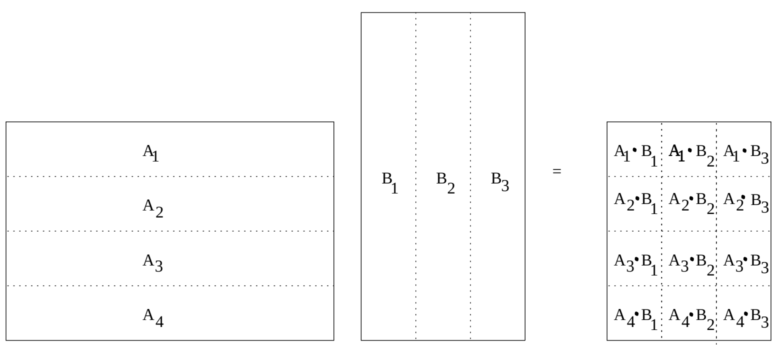
and

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix} = \begin{bmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \cdots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \cdots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & \cdots & a_{m,n}+b_{m,n} \end{bmatrix}$$

The product of an $m \times n$ matrix $A = [a_{i,j}]$ with a $n \times p$ matrix $B = [b_{i,j}]$ is a $m \times p$ matrix $C = [c_{i.j}]$ whose entries are defined by

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

An easy way to remember this is to chop the matrix A into m row vectors of length n and to chop B into p column vectors also of length n, as in the following diagram. The i,jth entry of the product is then the dot product $A_i \cdot B_j$.



Notice that the matrix product AB only makes sense if the the number of columns of A equals the number of rows of B. So $A^2 = AA$ only makes sense for a square matrix.

Here is an example

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1\times1+0\times3+1\times5+2\times7 & 1\times2+0\times4+1\times6+2\times8 \\ 1\times1+1\times3+1\times5+4\times7 & 1\times2+1\times4+1\times6+4\times8 \\ 0\times1+0\times3+1\times5+1\times7 & 0\times2+0\times4+1\times6+1\times8 \end{bmatrix} = \begin{bmatrix} 20 & 24 \\ 37 & 44 \\ 12 & 14 \end{bmatrix}$$

Notice that if A is an $m \times n$ matrix, and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

and

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

Then the equation

$$A\mathbf{x} = \mathbf{b}$$

is a short way of writing the system of linear equations corresponding to the augmented matrix $[A|\mathbf{b}]$.

We will see shortly why matrix multiplication is defined the way it is. For now, you should be aware of some important properties that *don't* hold for matrix multiplication, even though they are true for multiplication of numbers. First of all, in general, AB is *not* equal to BA, even when both products are defined and have the same size. For example, if

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

and

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

then

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

but

$$BA = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

This example also shows that two non-zero matrices can be multiplied together to give the zero matrix.

Here is a list of properties that do hold for matrix multiplication.

1. $A + B = B + A$
2. $A + (B + C) = (A + B) + C$
3. $s(A + B) = sA + sB$
4. $(s + t)A = sA + tA$
5. $(st)A = s(tA)$
6. $1A = A$
7. $A + \mathbf{0} = A$ (here $\mathbf{0}$ is the matrix with all entries zero)
8. $A - A = A + (-1)A = \mathbf{0}$
9. $A(B + C) = AB + AC$
10. $(A + B)C = AC + BC$
11. $A(BC) = (AB)C$
12. $s(AB) = (sA)B = A(sB)$

Problem 3.1: Define

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 1 \end{bmatrix} \; B = \begin{bmatrix} -1 & 2 \\ -3 & 1 \\ -2 & 1 \end{bmatrix} \; C = \begin{bmatrix} 2 & -2 & 0 \end{bmatrix} \; D = \begin{bmatrix} 2 \\ -11 \\ 2 \end{bmatrix}$$

Compute all products of two of these (i.e., AB, AC, etc.) that are defined.

Problem 3.2: Compute $A^2 = AA$ and $A^3 = AAA$ for

$$A = \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$A = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Problem 3.3: Let

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

a) Find A^2, A^3 and A^4.

b) Find A^k for all positive integers k.

c) Find e^{tA} (part of the problem is to invent a reasonable definition!)

d) Find a square root of A (i.e., a matrix B with $B^2 = A$).

e) Find all square roots of A.

Problem 3.4: Compute A^k for $k = 2, 3, 4$ when

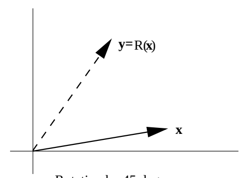
$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Linear Transformations and Matrices

Recall that a function f is a rule that takes an input value x and produces an output value $y = f(x)$. Functions are sometimes called transformations or maps (since they transform, or map, the input value to the output value). In calculus, you have mostly dealt with functions whose input values and output values are real numbers. However, it is also useful to consider functions whose input values and output values are vectors.

We have already encountered this idea when we discussed quadratic functions. A quadratic function such as $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ can be considered as a transformation (or map) whose input is the vector $\mathbf{x} = [x_1, x_2, x_3]$ and whose output is the number $y = x_1^2 + x_2^2 + x_3^2$. In this case we could write $f(x_1, x_2, x_3)$ as $f(\mathbf{x})$.

An example of a transformation whose inputs and outputs are both vectors in two dimensions is rotation by some angle, say 45°. If $\mathbf{x}$ is the input vector, then the output vector $R(\mathbf{x})$ is the vector you get by rotating $\mathbf{x}$ by 45° in the counter-clockwise direction.



Rotation by 45 degrees

A transformation T is called *linear* if for any two input vectors $\mathbf{x}$ and $\mathbf{y}$ and any two numbers s and t,

$$T(s\mathbf{x} + t\mathbf{y}) = sT(\mathbf{x}) + tT(\mathbf{y}) \tag{3.1}$$

This condition is saying that when we scalar multiply and add two vectors, it doesn't matter whether we (i) do scalar multiplication and addition first and then apply a linear transformation, or (ii) do a linear transformation first and then do scalar multiplication and addition. In both cases we get the same answer. The linearity condition (3.1) is equivalent to the following two conditions:

(i) For any two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}).$$

(ii) For any vector $\mathbf{x}$ and any scalar s,

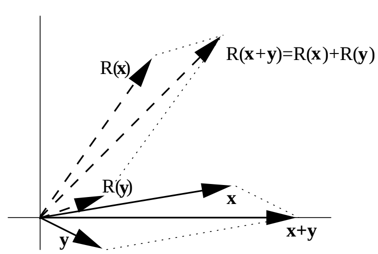
$$T(s\mathbf{x}) = sT(\mathbf{x})$$

Notice that the quadratic function f above is not a linear transformation, since

$$f(2\mathbf{x}) = (2x_1)^2 + (2x_2)^2 + (2x_3)^2 = 4(x_1^2 + x_2^2 + x_3^2) = 4f(\mathbf{x}).$$

So $f(2\mathbf{x})$ is not equal to $2f(\mathbf{x})$ as would need to be true if f were linear.

However, rotation by 45° is a linear transformation. The following picture demonstrates that condition (i) holds.



Rotation by 45 degrees

Problem 3.5: Let $\mathbf{a}$ be a fixed vector. Show that the transformation $T(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ is not a linear transformation.

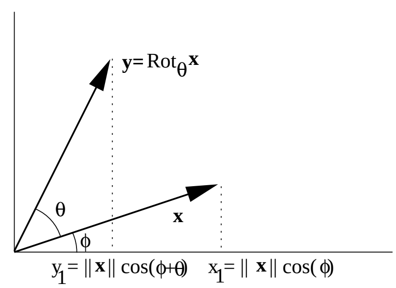
Problem 3.6: Let $\mathbf{a}$ be a fixed vector. Show that the transformation $T(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ is a linear transformation (whose output values are numbers).

The most important example of a linear transformation is multiplication by a matrix. If we regard vectors as column vectors, then multiplying an n dimensional vector $\mathbf{x}$ with an $m \times n$ matrix A results in an m dimensional vector $\mathbf{y} = A\mathbf{x}$. The linearity property (3.1) is a consequence of properties (9) and (12) of matrix multiplication. We will see that in fact every linear transformation is of this form.

Rotations in two dimensions

Let us obtain a formula for the transformation that rotates a vector in two dimensions counterclockwise by θ degrees. Let $\mathbf{x}$ be an arbitrary vector. Denote by $\mathrm{Rot}_\theta\mathbf{x}$ the vector obtained by rotating $\mathbf{x}$ counterclockwise by θ degrees. If the angle between $\mathbf{x}$ and the x axis is ϕ, then the components of $\mathbf{x}$ can be written $\mathbf{x} = [x_1, x_2]$ with $x_1 = \|x\|\cos(\phi)$ and $x_2 = \|x\|\sin(\phi)$.

To obtain the vector that has been rotated by θ degrees, we simply need to add θ to ϕ in this representation. Thus $\mathbf{y} = \mathrm{Rot}_\theta\mathbf{x} = [y_1, y_2]$, where $y_1 = \|x\|\cos(\phi + \theta)$ and $y_2 = \|x\|\sin(\phi + \theta)$.



To simplify this we can use the addition formulas for sin and cos. Recall that

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$$
$$\sin(a+b) = \cos(a)\sin(b) + \sin(a)\cos(b)$$

Thus

$$\begin{aligned} y_1 &= \|x\|\cos(\phi+\theta) \\ &= \|x\|(\cos(\phi)\cos(\theta) - \sin(\phi)\sin(\theta)) \\ &= \cos(\theta)x_1 - \sin(\theta)x_2 \end{aligned}$$

$$\begin{aligned} y_2 &= \|x\|\sin(\phi+\theta) \\ &= \|x\|(\sin(\phi)\cos(\theta) + \cos(\phi)\sin(\theta)) \\ &= \sin(\theta)x_1 + \cos(\theta)x_2 \end{aligned}$$

Notice that this can be written as a matrix product.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The matrix $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, also denoted Rot_θ, is called a rotation matrix. What this formula is saying is that the linear transformation of rotation by θ degrees in the same as the linear transformation of multiplication by the matrix. In other words, if we want to know the co-ordinates of the vector obtained by rotating $\mathbf{x}$ by θ degrees, we simply calculate $\mathrm{Rot}_\theta \mathbf{x}$.

Projections in two dimensions

Now we consider the transformation which projects a vector $\mathbf{x}$ in the direction of another vector $\mathbf{a}$. We already have a formula for this transformation. In the special case that $\mathbf{a}$ has unit length, the formula is

$$\mathrm{Proj}_{\mathbf{a}}\mathbf{x} = (\mathbf{x} \cdot \mathbf{a})\mathbf{a}.$$

It follows from the properties of the dot product that

$$\begin{aligned} \mathrm{Proj}_{\mathbf{a}}(s\mathbf{x} + t\mathbf{y}) &= ((s\mathbf{x} + t\mathbf{y}) \cdot \mathbf{a})\mathbf{a} \\ &= ((s\mathbf{x} \cdot \mathbf{a} + t\mathbf{y} \cdot \mathbf{a})\mathbf{a} \\ &= s((\mathbf{x} \cdot \mathbf{a})\mathbf{a}) + t((\mathbf{y} \cdot \mathbf{a})\mathbf{a}) \\ &= s\mathrm{Proj}_{\mathbf{a}}\mathbf{x} + t\mathrm{Proj}_{\mathbf{a}}\mathbf{y} \end{aligned}$$

Thus $\mathrm{Proj}_{\mathbf{a}}$ is a linear transformation. Let us now see that $\mathrm{Proj}_{\mathbf{a}}$ is also given by multiplication by a matrix. If $\mathbf{a} = [a_1, a_2]$, then

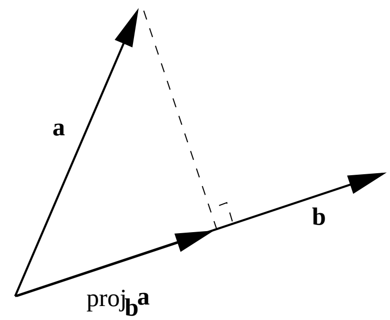
$$\begin{aligned} \mathrm{Proj}_{\mathbf{a}}\mathbf{x} &= \begin{bmatrix} (x_1 a_1 + x_2 a_2) a_1 \\ (x_1 a_1 + x_2 a_2) a_2 \end{bmatrix} \\ &= \begin{bmatrix} a_1^2 x_1 + a_1 a_2 x_2 \\ a_2 a_1 x_1 + a_2^2 x_2 \end{bmatrix} \\ &= \begin{bmatrix} a_1^2 & a_1 a_2 \\ a_2 a_1 & a_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

If $\mathbf{a}$ is the unit vector making an angle of θ with the x axis, then $a_1 = \cos(\theta)$ and $a_2 = \sin(\theta)$. Using half angle formulas, we have

$$\begin{aligned} a_1^2 = \cos^2(\theta) &= \frac{1 + \cos(2\theta)}{2} \\ a_2^2 = \sin^2(\theta) &= \frac{1 - \cos(2\theta)}{2} \\ a_1 a_2 = \cos(\theta)\sin(\theta) &= \frac{\sin(2\theta)}{2} \end{aligned}$$

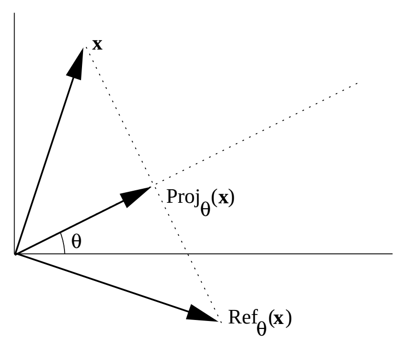
Thus the matrix which when multiplied by $\mathbf{x}$ produces the projection of $\mathbf{x}$ onto the line making an angle of θ with the x axis is given by

$$\mathrm{Proj}_\theta = \frac{1}{2} \begin{bmatrix} 1 + \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & 1 - \cos(2\theta) \end{bmatrix}$$



Reflections in two dimensions

A third example of a geometric linear transformation is reflection across a line. The following figure illustrates reflection across a line making an angle θ with the x axis. Let $\mathrm{Ref}_\theta \mathbf{x}$ denote the reflected vector.



We can obtain the matrix for reflection from the following observation.

The vector with tail at $\mathbf{x}$ and head at $\mathrm{Proj}_\theta \mathbf{x}$ is $\mathrm{Proj}_\theta \mathbf{x} - \mathbf{x}$. If we add twice this vector to $\mathbf{x}$, we arrive at $\mathrm{Ref}_\theta \mathbf{x}$. Therefore

$$\begin{aligned}\mathrm{Ref}_\theta \mathbf{x} &= \mathbf{x} + 2(\mathrm{Proj}_\theta \mathbf{x} - \mathbf{x}) \\ &= 2\mathrm{Proj}_\theta \mathbf{x} - \mathbf{x}\end{aligned}$$

Now if $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $I\mathbf{x} = \mathbf{x}$ for any vector $\mathbf{x}$, since

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1x_1 + 0x_2 \\ 0x_1 + 1x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

I is called the identity matrix.

Now we can write

$$\mathrm{Ref}_\theta \mathbf{x} = 2\mathrm{Proj}_\theta \mathbf{x} - I\mathbf{x} = (2\mathrm{Proj}_\theta - I)\mathbf{x}.$$

This means that the matrix for reflections is $2\mathrm{Proj}_\theta - I$. Explicitly

$$\begin{aligned}\mathrm{Ref}_\theta &= \begin{bmatrix} 1+\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & 1-\cos(2\theta) \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}\end{aligned}$$

Problem 3.7: Find the matrices which project on the lines

(a) $x_1 = x_2$ (b) $3x_1 + 4x_2 = 0$.

Problem 3.8: Find the matrices that reflect about the lines (a) $x_1 = x_2$ (b) $3x_1 + 4x_2 = 0$.

Problem 3.9: Find the matrices which rotate about the origin in two dimensions by (a) $\pi/4$, (b) $\pi/2$ (c) π

Problem 3.10: Find the matrix which first reflects about the line making an angle of ϕ with the x axis, and then reflects about the line making an angle of θ with the x axis.

Every linear transformation is multiplication by a matrix

We have just seen three examples of linear transformations whose action on a vector is given by multiplication by a matrix. Now we will see that for *any* linear transformation $T(\mathbf{x})$ there is a matrix T such that $T(\mathbf{x})$ is the matrix product $T\mathbf{x}$.

To illustrate this suppose that T is a linear transformation that takes three dimensional vectors as input.

Let $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ be the standard basis vectors in three dimensions, that is

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Then any vector can be written

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

Now, using the linearity property of the linear transformation T, we obtain

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + x_3T(\mathbf{e}_3)$$

Now take the three vectors $T(\mathbf{e}_1)$, $T(\mathbf{e}_2)$ and $T(\mathbf{e}_3)$ and put them in the columns of a matrix which we'll also call T. Then

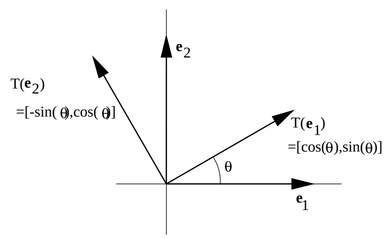
$$T\mathbf{x} = \left[\begin{array}{c|c|c} T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = T(\mathbf{e}_1)x_1 + T(\mathbf{e}_2)x_2 + T(\mathbf{e}_3)x_3 = T(\mathbf{x})$$

In other words, the action of the transformation T on a vector $\mathbf{x}$ is the same as multiplying x by the matrix $T = \left[\begin{array}{c|c|c} T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \end{array}\right]$

The same idea works in any dimension. To find the matrix of a linear transformation $T(\mathbf{x})$ we take the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ (where $\mathbf{e}_k$ has zeros everywhere except for a 1 in the kth spot) and calculate the action of the linear transformation on each one. We then take the transformed vectors

$T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)$ and put them into the columns of a matrix $T = \left[T(\mathbf{e}_1) \Big| T(\mathbf{e}_2) \Big| \cdots \Big| T(\mathbf{e}_n) \right]$. This matrix T then reproduces the action of the linear transformation, that is, $T(\mathbf{x}) = T\mathbf{x}$.

To see how this works in practice, lets recalculate the matrix for rotations in two dimensions. Under a rotation angle of θ, the vector $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ gets transformed to $T(\mathbf{e}_1) = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$ while the vector $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ gets transformed to $T(\mathbf{e}_2) = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}$



According to our prescription, we must now put these two vectors into the columns of a matrix. This gives the matrix

$$T = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

which is exactly the same as Rot_θ.

Composition of linear transformations and matrix product

Suppose we apply one linear transformation T and then follow it by another linear transformation S. For example, think of first rotating a vector and then reflecting it across a line. Then $S(T(\mathbf{x}))$ is again a linear transformation, since

$$S(T(s\mathbf{x} + t\mathbf{y})) = S(sT(\mathbf{x}) + tT(\mathbf{y})) = sS(T(\mathbf{x})) + tS(T(\mathbf{y})).$$

What is the matrix for the composition $S(T(\mathbf{x}))$? We know that there are matrices S and T that reproduce the action of $S(\mathbf{x})$ and $T(\mathbf{x})$. So $T(\mathbf{x})$ is the matrix product $T\mathbf{x}$ and $S(T\mathbf{x}))$ is the matrix product $S(T\mathbf{x})$ (here the parenthesis just indicate in which order we are doing the matrix product) But matrix multiplication is associative. So $S(T\mathbf{x}) = (ST)\mathbf{x}$. In other words the matrix for the composition of $S(\mathbf{x})$ and $T(\mathbf{x})$ is simply the matrix product of the corresponding matrices S and T.

For example, to compute the matrix for the transformation of rotation by 45° followed by reflection about the line making an angle of 30° with the x axis we simply compute the product

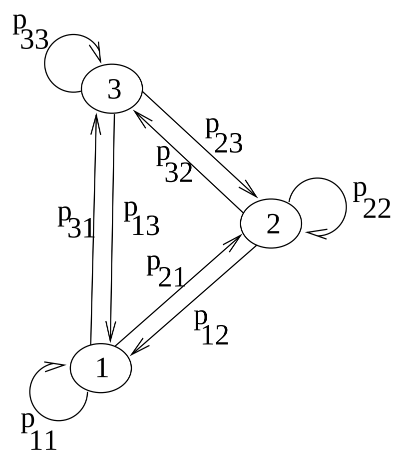
$$\begin{aligned}
\mathrm{Ref}_{30^\circ}\mathrm{Rot}_{45^\circ} &= \begin{bmatrix} \cos(60^\circ) & \sin(60^\circ) \\ \sin(60^\circ) & -\cos(60^\circ) \end{bmatrix} \begin{bmatrix} \cos(45^\circ) & -\sin(45^\circ) \\ \sin(45^\circ) & \cos(45^\circ) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sqrt{2}+\sqrt{6}}{4} & \frac{-\sqrt{2}+\sqrt{6}}{4} \\ \frac{\sqrt{6}-\sqrt{2}}{4} & \frac{-\sqrt{6}-\sqrt{2}}{4} \end{bmatrix}
\end{aligned}$$

Problem 3.11: Find the matrix which first reflects about the line making an angle of ϕ with the x axis and then reflects about the line making an angle of θ with the x axis. Give another geometric interpretation of this matrix.

Problem 3.12: Find the matrix that rotates about the z axis by and angle of θ in three dimensions.

Application of matrix product: random walk

Consider a system with three states, labeled 1, 2 and 3 in the diagram below.



To make the problem more vivid, one can imagine these as being actual locations. A random walker starts off at some location, say location 1 at time 0. Then at a sequence of times, $1, 2, \ldots, n \ldots$, the walker either stays where he is, or moves to one of the other locations. The next location is chosen randomly, but according to the transition probabilities $p_{i,j}$. These are numbers between 0 and 1 that measure how likely it is that, starting from location j, the walker will move to location i. If $p_{i,j} = 0$, then there is no chance that the walker will move from j to i, and if $p_{i,j} = 1$, then the walker will move for sure from j to i.

Since the walker must either move from j to another site or stay put, the sum of these probabilities must equal one:

$$p_{1,j} + p_{2,j} + p_{3,j} = \sum_{i=1}^{3} p_{i,j} = 1$$

At each time n there is a vector $\mathbf{x}_n = \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ x_{n,3} \end{bmatrix}$ that gives the probabilities that the walker is in location 1, 2 or 3 at time n. Let us compute the vector $\mathbf{x}_{n+1}$, given $\mathbf{x}_n$. To start, we must compute $x_{n+1,1}$, the probability that the walker is at location 1 at time $n+1$. There are three ways the walker can end up at location 1. The walker might have been at location 1 at time n and have stayed there. The probability of this is $p_{1,1}x_{n,1}$. Or he might have been at location 2 and have moved to 1. The probability of this is $p_{1,2}x_{n,2}$. Finally, he might have been at location 3 and have moved to 1. The probability of this is $p_{1,3}x_{n,3}$. Thus the total probability that the walker is at location 1 at time $n+1$ is the sum

$$x_{n+1,1} = p_{1,1}x_{n,1} + p_{1,2}x_{n,2} + p_{1,3}x_{n,3}$$

Similarly, for all i

$$x_{n+1,i} = p_{i,1}x_{n,1} + p_{i,2}x_{n,2} + p_{i,3}x_{n,3}$$

But this is exactly the formula for matrix multiplication. So

$$\mathbf{x}_{n+1} = P\mathbf{x}_n$$

where P is the matrix with entries $[p_{ij}]$.

Now suppose the initial probabilities are given by some vector $\mathbf{x}_0$. For example, if the walker starts off at location 1, then

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Then after one time step, we have

$$\mathbf{x}_1 = P\mathbf{x}_0$$

Then,

$$\mathbf{x}_2 = P\mathbf{x}_1 = P^2\mathbf{x}_0$$

and so on, so after n time steps

$$\mathbf{x}_n = P^n\mathbf{x}_0$$

Notice that the sum $x_{n,1} + x_{n,2} + x_{n,3} = \sum_{i=1}^{3} x_{n,i}$ should equal one, since the total probability of being in one of the three locations must add up to one.

If the initial vector has this property, then it is preserved for all later times. To see this, suppose that $\sum_{j=1}^{3} x_j = 1$ Then, since $\sum_{i=1}^{3} p_{ij} = 1$

$$\begin{aligned}\sum_{i=1}^{3}(Px)_i &= \sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij}x_j \\ &= \sum_{j=1}^{3}\left(\sum_{i=1}^{3} p_{ij}\right) x_j \\ &= \sum_{j=1}^{3} x_j \\ &= 1\end{aligned}$$

In other words, the vector for the next time step also has components summing to one.

Of course, one can generalize this to a system with an arbitrary number of states or locations.

Later in the course we will see how to use eigenvalues and eigenvectors to compute the limit as n tends to infinity of this expression.

Problem 3.13: Consider a random walk with 3 states, where the probability of staying in the same location is zero. Suppose

the probability of moving from location 1 to location 2 is $1/2$

the probability of moving from location 2 to location 1 is $1/3$

the probability of moving from location 3 to location 1 is $1/4$

Write down the matrix P. What is the probability that a walker starting in location 1 is in location 2 after two time steps?

Problem 3.14: Consider a random walk with 3 states, where all the probabilities $p_{i,j}$ are all equal to $1/3$. What is P, P^n. Compute the probabilities $P^n\mathbf{x}_0$ when $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ (ie the walker starts in location 1), $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ (ie the walker starts in location 2), and $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (ie the walker starts in location 3)

The transpose

If A is an $m \times n$ matrix, then its transpose A^T is the matrix obtained by flipping A about its diagonal. So the columns of A^T are the rows of A (in the same order) and vice versa. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

then

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Another way of saying this is that the i,jth entry of A^T is the same as the j,ith entry of A, that is,

$$a^T_{i,j} = a_{j,i}.$$

There are two important formulas to remember for the transpose of a matrix. The first gives a relation between the transpose and the dot product. If A is an $n \times m$ matrix, then for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ we have

$$\mathbf{y} \cdot (A\mathbf{x}) = (A^T\mathbf{y}) \cdot \mathbf{x} \tag{3.2}$$

Problem 3.15: Verify formula (3.2) for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

The proof of this formula is a simple calculation.

$$\begin{aligned}
\mathbf{y} \cdot (A\mathbf{x}) &= \sum_{i=1}^{m} y_i \left(\sum_{j=1}^{n} a_{i,j} x_j \right) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} y_i a_{i,j} x_j \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} a^T_{j,i} y_i x_j \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a^T_{j,i} y_i \right) x_j \\
&= (A^T\mathbf{y}) \cdot \mathbf{x}
\end{aligned}$$

In fact, the formula (3.2) could be used to define the transpose. Given A, there is exactly one matrix A^T for which (3.2) is true for every $\mathbf{x}$ and $\mathbf{y}$, and this matrix is the transpose.

The second important formula relates the transpose of a product of matrices to the transposes of each one. For two matrices A and B such that AB is defined the formula reads

$$(AB)^T = B^T A^T \tag{3.3}$$

Notice that the order of the factors is reversed on the right side. To see why (3.3) is true, notice that on the one hand

$$\mathbf{y} \cdot (AB\mathbf{x}) = ((AB)^T\mathbf{y}) \cdot \mathbf{x}$$

while on the other hand

$$\mathbf{y} \cdot (AB\mathbf{x}) = \mathbf{y} \cdot (A(B\mathbf{x})) = (A^T\mathbf{y}) \cdot (B\mathbf{x}) = (B^T A^T\mathbf{y}) \cdot \mathbf{x}$$

Thus $((AB)^T\mathbf{y}) \cdot \mathbf{x} = (B^T A^T\mathbf{y}) \cdot \mathbf{x}$ for every $\mathbf{x}$ and $\mathbf{y}$. This can only be true if (3.3) holds.

Problem 3.16: What is $(A^T)^T$?

Problem 3.17: Verify (3.3) for $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

Problem 3.18: Show that if A and B are both $m \times n$ matrices such that $\mathbf{y} \cdot (A\mathbf{x}) = \mathbf{y} \cdot (B\mathbf{x})$ for every $\mathbf{y} \in \mathbb{R}^m$ and every $\mathbf{x} \in \mathbb{R}^n$, then $A = B$.

Problem 3.19: Show that if you think of (column) vectors in $\mathbb{R}^n$ as $n \times 1$ matrices then

$$\mathbf{x}\dot{\mathbf{y}} = \mathbf{x}^T\mathbf{y}$$

Now use this formula and (3.3) to derive (3.2).

Quadratic functions revisited

Let us restate our results on minimization of quadratic functions using matrix notation. A quadratic function of $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ can be written in matrix form as

$$f(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{b} \cdot \mathbf{x} + c$$

where A is an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$ and c is a number. The vector containing the partial derivatives can be computed to be

$$\begin{bmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} = (A + A^T)\mathbf{x} + \mathbf{b}$$

Recall that we made the assumption that $a_{ij} = a_{ji}$ when we considered this problem before. This property can be stated in compact form as

$$A = A^T.$$

If this is true then $(A + A^T) = 2A$ so

$$\begin{bmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} = 2A\mathbf{x} + \mathbf{b}$$

To find the minimum value of f (if it exists) we need to find the value of $\mathbf{x}$ for which the vector above is zero. In other words, $\mathbf{x}$ solves the equation

$$2A\mathbf{x} = -\mathbf{b}.$$

This is the same equation that we derived before.

Least squares solutions

Lets take another look the situation where a system of linear equations, which we now can write

$$B\mathbf{x} = \mathbf{c},$$

has no solution. Typically this will be the case if there are more equations than variables, that is, B is an matrix with more rows than columns. In this case there is no value of $\mathbf{x}$ that makes the left side equal the right side. However, we may try to find the value of $\mathbf{x}$ for which the right side $B\mathbf{x}$ is *closest* to the left side $\mathbf{c}$.

One way to go about this is to try to minimize distance between the left and right sides. It is more convenient to minimize the square of the distance. This quantity can be written

$$\begin{aligned} \|B\mathbf{x} - \mathbf{c}\|^2 &= (B\mathbf{x} - \mathbf{c}) \cdot (B\mathbf{x} - \mathbf{c}) \\ &= (B\mathbf{x}) \cdot (B\mathbf{x}) - (B\mathbf{x}) \cdot \mathbf{c} - \mathbf{c} \cdot (B\mathbf{x}) + \mathbf{c} \cdot \mathbf{c} \\ &= \mathbf{x} \cdot (B^T B\mathbf{x}) - 2(B^T\mathbf{c}) \cdot \mathbf{x} + \mathbf{c} \cdot \mathbf{c} \end{aligned}$$

This is a quadratic function, written in matrix form. We want to use the formula of the previous section with $A = B^T B$ and $\mathbf{b} = B^T\mathbf{c}$. Before we can do so, we must verify that $A = B^T B$ satisfies $A^T = A$. This is true because

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

Thus the formula of the previous section implies that the minimum occurs at the value of $\mathbf{x}$ that solves the linear equation

$$B^T B\mathbf{x} = B^T\mathbf{c}$$

Here we have canceled a factor of 2 on each side.

Now lets derive the same result in another way. Think of all the values of $B\mathbf{x}$, as $\mathbf{x}$ ranges through all possible values in $\mathbb{R}^n$ as forming a (high dimensional) plane in $\mathbb{R}^m$. Our goal is to find the value of $\mathbf{x}$ so that the corresponding value of $B\mathbf{x}$ on the plane is closest to $\mathbf{c}$. Using the analogy to the geometric picture in three dimensions, we see that the minimum will occur when $B\mathbf{x} - \mathbf{c}$ is orthogonal to the plane. This means that the dot product of $B\mathbf{x} - \mathbf{c}$ with every vector in the plane, that is, every vector of the form $B\mathbf{y}$, should be zero. Thus we have

$$(B\mathbf{y}) \cdot (B\mathbf{x} - \mathbf{c}) = \mathbf{0}$$

for every $\mathbf{y} \in \mathbb{R}^n$. This is the same as

$$\mathbf{y} \cdot (B^T(B\mathbf{x} - \mathbf{c})) = \mathbf{y} \cdot (B^T B\mathbf{x} - B^T\mathbf{c}) = \mathbf{0}$$

for every $\mathbf{y} \in \mathbb{R}^n$. This can happen only if

$$B^T B\mathbf{x} = B^T\mathbf{c}$$

which is the same result we obtained before.

Problem 3.20: Find the least squares solution to

$$\begin{aligned} x_1 \quad +x_2 \quad &= 1 \\ x_1 \qquad\qquad &= 1 \\ x_1 \quad +x_2 \quad &= 0 \end{aligned}$$

Compare $B\mathbf{x}$ and $\mathbf{b}$.

Problem 3.21: Refer back to the least squares fit example, where we tried to find the best straight line going through a collection of points (x_i, y_i). Another way of formulating this problem is this. The line $y = ax + b$ passes through the point (x_i, y_i) if

$$ax_i + b = y_i \tag{3.4}$$

So, saying that the straight line passes through all n points is the same as saying that a and b solve the system of n linear equations given by (3.4) for $i = 1, \ldots, n$. Of course, unless the points all actually lie on the same line, this system of equations has no solutions. Show that the least squares solution to this problem is the same as we obtained before. (You may simplify the problem by assuming there are only three points (x_1, y_1), (x_2, y_2) and (x_3, y_3).)

Matrix Inverses

To solve the numerical equation

$$ax = b$$

for x we simply multiply both sides by $a^{-1} = \frac{1}{a}$. Then, since $a^{-1}a = 1$, we find

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

Of course, if $a = 0$ this doesn't work, since we cannot divide by zero. In fact, if $a = 0$ the equation $ax = b$ either has no solutions, if $b \neq 0$, or infinitely many solutions (every value of x), if $b = 0$.

We have seen that a system of linear equations can be rewritten

$$A\mathbf{x} = \mathbf{b}$$

where is A is a known matrix $\mathbf{x}$ is the unknown vector to be solved for, and $\mathbf{b}$ is a known vector. Suppose we could find an inverse matrix B (analogous to a^{-1}) with the property that $BA = I$ (recall that I denotes the identity matrix). Then we could matrix multiply both sides of the equation by B yielding

$$BA\mathbf{x} = B\mathbf{b}$$

But $BA\mathbf{x} = I\mathbf{x} = \mathbf{x}$, so $\mathbf{x} = B\mathbf{b}$. Thus there is a unique solution and we have a formula for it.

Just as in the numerical case, where a could be zero, we can't expect to find an inverse matrix in all cases. After all, we know that there are linear systems of equations with no solutions and with infinitely many solutions. In these situations there can be no inverse matrix.

When considering matrix inverses, we will always assume that we are dealing with square (i.e., $n \times n$) matrices.

We now make a definition.

Definition: If A is an $n \times n$ matrix, then B is called the *inverse* of A, and denoted $B = A^{-1}$, if

$$BA = I,$$

where I is the $n \times n$ identity matrix (with each diagonal entry equal to 1 and all other entries 0).

Here is an example. Suppose $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$ then the inverse matrix is $B = A^{-1} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$, since

$$\begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix} = \begin{bmatrix} 6-5 & 3-3 \\ 10-10 & -5+6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This means that to solve the linear equation

$$\begin{array}{lll} 2x_1 & +x_2 & = 2 \\ 5x_1 & +3x_2 & = 4 \end{array}$$

we write it as a matrix equation

$$\begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

and then multiply both sides by the inverse to obtain

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Here is an example of a matrix that doesn't have an inverse. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. To see that A doesn't have an inverse, notice that the homogeneous equations $A\mathbf{x} = \mathbf{0}$ has a non-zero solution $\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If A had an inverse B, then we could multiply both sides of the equation $A\mathbf{x} = \mathbf{0}$ by B to obtain $\mathbf{x} = B\mathbf{0} = \mathbf{0}$. But this is false. Therefore there cannot be an inverse for A.

Clearly, having an inverse is somehow connected to whether or not there are any non-zero solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Recall that $A\mathbf{x} = \mathbf{0}$ has only the zero solution precisely when $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$.

Let A be an $n \times n$ matrix. The following conditions are equivalent:

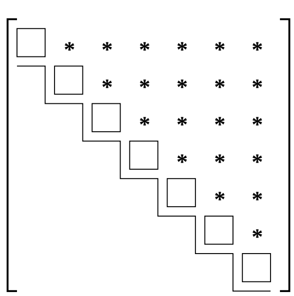
(1) A is invertible.

(2) The equation $A\mathbf{x} = \mathbf{b}$ always has a unique solution.

(3) The equation $A\mathbf{x} = \mathbf{0}$ has as the only solution $\mathbf{x} = \mathbf{0}$.

(4) The rank of A is n.

(5) The reduced form of A looks like



We already know that the conditions (2), (3), (4) and (5) are all equivalent.

We also just have seen that if A is invertible with inverse B, then the solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = B\mathbf{b}$ so it exists and since we have a formula for it, it is unique.

So we just have to show that if $A\mathbf{x} = \mathbf{b}$ always has a unique solution, then A has an inverse. Consider the transformation that takes a vector $\mathbf{b}$ to the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$, ie, $T\mathbf{b} = \mathbf{x}$. It is easy to check that this is a linear transformation, since if $T\mathbf{b}_1 == \mathbf{x}_1$, i.e., $A\mathbf{x}_1 = \mathbf{b}_1$ and $T\mathbf{b}_2 = \mathbf{x}_2$, i.e., $A\mathbf{x}_2 = \mathbf{b}_2$, then $A(t_1\mathbf{x}_1 + t_2\mathbf{x}_2) = t_1A\mathbf{x}_1 + t_2A\mathbf{x}_2 = t_1\mathbf{b}_1 + t_2\mathbf{b}_2$, so that $T(t_1\mathbf{b}_1 + t_2\mathbf{b}_2) = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 = t_1T\mathbf{b}_1 + t_2T\mathbf{b}_1$

Since T is a linear transformation, it is given by some matrix B, and since $T(A\mathbf{x}) = \mathbf{x}$, we must have $BA\mathbf{x} = \mathbf{x}$ which implies that BA is the identity matrix.

Going back to our first example, notice that not only is $BA = I$, but $AB = I$ too, since

$$\begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} = \begin{bmatrix} 6-5 & -2+2 \\ 15-15 & -5+6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For a general choice of A and B, BA need not be equal to AB. But if B is the inverse of A, then it *is* always true that $AB = BA = I$.

To see this, suppose that A is invertible with $BA = I$ but we don't know yet whether $AB = I$. So what we need to show is that for every vector $\mathbf{x}$, $AB\mathbf{x} = \mathbf{x}$. First notice that if A is invertible, then any vector $\mathbf{x}$ can be written in the form $A\mathbf{y}$ for some $\mathbf{y}$, since this is just the same as saying that the equation $A\mathbf{y} = \mathbf{x}$ has a solution $\mathbf{y}$. Thus $AB\mathbf{x} = ABA\mathbf{y} = AI\mathbf{y} = A\mathbf{y} = \mathbf{x}$.

Problem 3.22: Which of the following matrices are invertible?

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

(b) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 4 \\ 0 & 1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 4 \\ 0 & 1 & 2 \end{bmatrix}$

Problem 3.23: Find the inverse for $\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$

Computing the inverse

How can we compute the inverse of an $n \times n$ matrix A? Suppose that A is invertible and B is the inverse. Then $AB = I$. We can rewrite this equation as follows. Think of the columns of B as being column vectors so that

$$B = \left[\mathbf{b}_1 \middle| \mathbf{b}_2 \middle| \cdots \middle| \mathbf{b}_n\right]$$

Then the rules of matrix multiplication imply that

$$AB = \left[A\mathbf{b}_1 \middle| A\mathbf{b}_2 \middle| \cdots \middle| A\mathbf{b}_n\right]$$

Now the identity matrix can also be written as a matrix of column vectors. In this case the kth column is simply the matrix with zeros everywhere except for a 1 in the kth place, in other words the vector $\mathbf{e}_k$. Thus

$$I = \left[\begin{array}{c|c|c|c} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{array}\right]$$

So if $AB = I$ then the n equations

$$\begin{aligned} A\mathbf{b}_1 &= \mathbf{e}_1 \\ A\mathbf{b}_2 &= \mathbf{e}_2 \\ &\vdots \\ A\mathbf{b}_n &= \mathbf{e}_n \end{aligned}$$

hold. If we solve each of these equations for $\mathbf{b}_1$, $\mathbf{b}_2$, ..., $\mathbf{b}_n$, then we have found the inverse B.

Here is a simple example. Suppose we want to find the inverse for $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$. According to our discussion, we must solve $A\mathbf{b}_1 = \mathbf{e}_1$ and $A\mathbf{b}_2 = \mathbf{e}_2$. The augmented matrix for $A\mathbf{b}_1 = \mathbf{e}_1$ is

$$\left[\begin{array}{cc|c} 2 & 1 & 1 \\ 5 & 3 & 0 \end{array}\right]$$

We now perform a sequence of row operations. First divide the first row by 2. This gives

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

Now subtract 5 times the first row from the second row. This gives

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

Now subtract the second row from the first row. This gives

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

Finally, multiply the second row by 2. This gives

$$\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & -5 \end{array}\right].$$

Therefore $\mathbf{b}_1 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$ The augmented matrix for $A\mathbf{b}_2 = \mathbf{e}_2$ is

$$\left[\begin{array}{cc|c} 2 & 1 & 0 \\ 5 & 3 & 1 \end{array}\right]$$

We now perform a sequence of row operations. First divide the first row by 2. This gives

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

Now subtract 5 times the first row from the second row. This gives

$$\left[\begin{array}{cc|c} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{array}\right].$$

Now subtract the second row from the first row. This gives

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1/2 & 1 \end{array}\right].$$

Finally, multiply the second row by 2. This gives

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right].$$

Therefore $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. So

$$B = \left[\mathbf{b}_1 \middle| \mathbf{b}_2\right] = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$$

Notice that we performed *exactly the same sequence of row operations* in finding $\mathbf{b}_1$ and $\mathbf{b}_2$. This is because the row operations only depend on the left side of the augmented matrix, in other words, the matrix A. If we used this procedure to find the inverse of an $n \times n$ matrix, we would end up doing exactly the same row operations n times. Clearly this is a big waste of effort! We can save a lot of work by solving all the equations at the same time. To do this we make a super-augmented matrix with both right sides.

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 5 & 3 & 0 & 1 \end{array}\right]$$

Now we only have to go through the sequence of row operations once, keeping track of both right sides simultaneously. Going through the same sequence, we obtain

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

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

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

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

Notice that the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are automatically arranged as columns on the right side, so the matrix on the right is the inverse B.

The same procedure works for any size of square matrix. To find the inverse of A form the super-augmented matrix $[A|I]$. Then do a sequence of row operations to reduce A to the identity. If the resulting matrix is $[I|B]$ then B is the inverse matrix.

What happens if A doesn't have an inverse? In this case it will be impossible to reduce A to the identity matrix, since the rank of A is less than n. So the procedure will fail, as it must.

As another example, let us now compute the inverse of an arbitrary invertible 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We will see that A is invertible precisely when its determinant $\Delta = ad - bc$ is non-zero. So lets assume this is the case, and do the computation. To start with, lets assume that $ac \neq 0$. Then neither a or c are zero. Here is the sequence of row transformations.

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{cc|cc} ac & bc & c & 0 \\ ac & ad & 0 & a \end{array}\right] \begin{array}{l} c(1) \\ a(2) \end{array}$$

Notice that multiplication by a and by c would not be legal row transformations if either a or c were zero.

$$\left[\begin{array}{cc|cc} ac & bc & c & 0 \\ 0 & ad-bc & -c & a \end{array}\right] \begin{array}{l} \\ (2)-(1) \end{array}$$

$$\left[\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ 0 & 1 & -c/\Delta & a/\Delta \end{array}\right] \begin{array}{l} (1/ac)(1) \\ (1/\Delta)(2) \end{array}$$

$$\left[\begin{array}{cc|cc} 1 & 0 & d/\Delta & -b/\Delta \\ 0 & 1 & -c/\Delta & a/\Delta \end{array}\right] \begin{array}{c} (1)-(b/a)(2) \\ (1/\Delta)(2) \end{array}$$

Thus the inverse matrix is

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

This was derived under the additional assumption that $ac \neq 0$. However one can check directly that the same formula works, so long as $\Delta = ad - bc \neq 0$.

Problem 3.24: Determine which of these matrices are invertible, and find the inverse for the invertible ones.

(a) $\begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & 3 \\ -1 & -1 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -1 \\ 2 & -1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 1 & 4 \\ 3 & 2 & 5 \\ 0 & -1 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$

Row operations and elementary matrices

Recall that there are three row operation that are used in Gaussian elimination: (1) multiplication of a row by a non-zero number, (2) add a multiple of one row to another row and (3) exchanging two rows.

It turns out that each elementary row operation can be implemented by left multiplication by a matrix. In other words, for each elementary row operation there is a matrix Q such that QA is what you get by doing that row operation to the matrix A.

Here is an example. Suppose

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

and suppose that the row operation is multiplying the first row by 2. Then the matrix you get by doing that row operation to the matrix A is

$$A' = \begin{bmatrix} 2 & 0 & 4 & 2 \\ 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

In this case the matrix Q turns out to be

$$Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 & 2 \\ 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

i.e., $QA = A'$.

Now suppose that the elementary row operation is subtracting twice the first row from the second row. Then the matrix you get by doing that row operation to the matrix A is

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

In this case the matrix Q turns out to be

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

Since

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

i.e., again, $QA = A'$.

Finally, suppose that the elementary row operation is exchanging the second and the third rows. Then the matrix you get by doing that row operation to the matrix A is

$$A' = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 0 & 1 \end{bmatrix}$$

In this case the matrix Q turns out to be

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Since

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 0 & 1 \end{bmatrix}$$

i.e., again, $QA = A'$.

How can we find the matrices Q (called elementary matrices)? Here is the procedure. Start with the identity matrix I and do the row transformation to it. The resulting matrix Q is the matrix that implements that row transformation by multiplication from the left. Notice that this is true in the examples above. In the first example, the row transformation was multiplying the first row by 2. If you multiply the first row of $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ by two you get $Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

In the second example, the row transformation was subtracting twice the first row from the second row. If you subtract twice the second row from the first row of I by two you get $Q = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

In the third example, the row transformation was exchanging the second and third rows. If you exchange the second and third rows of I, you get $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Problem 3.25: Each elementary matrix is invertible, and the inverse is also an elementary matrix. Find the inverses of the three examples of elementary matrices above. Notice that the inverse elementary matrix is the matrix for the row transformation that undoes the original row transformation.

Elementary matrices are useful in theoretical studies of the Gaussian elimination process. We will use them briefly when studying determinants.

Suppose A is a matrix, and R is its reduced form. Then we can obtain R from A via a sequence of elementary row operations. Suppose that the corresponding elementary matrices are Q_1, Q_2, ..., Q_k. Then, starting with A, the matrix after the first elementary row operation is $Q_1 A$, then after the second elementary row operation is $Q_2 Q_1 A$, and so on, until we have

$$Q_k Q_{k-1} \cdots Q_2 Q_1 A = R.$$

Now let us apply the inverse matrices, starting with Q_k^{-1}. This gives

$$Q_k^{-1} Q_k Q_{k-1} \cdots Q_2 Q_1 A = Q_{k-1} \cdots Q_2 Q_1 A = Q_k^{-1} R.$$

Continuing in this way we see that

$$A = Q_1^{-1} Q_2^{-1} \cdots Q_k^{-1} R$$

In the special case that A is an $n \times n$ invertible matrix, A can be reduced to the identity matrix. In other words, we can take $R = I$. In this case A can be written as a product of elementary matrices.

$$A = Q_1^{-1}Q_2^{-1}\cdots Q_k^{-1}I = Q_1^{-1}Q_2^{-1}\cdots Q_k^{-1}$$

Notice that in this case

$$A^{-1} = Q_kQ_{k-1}\cdots Q_2Q_1.$$

As an example, let us write the matrix $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$ as a product of elementary matrices. The sequence of row transformations that reduce A to the identity are:

1) (1/2)(R1)

2) (R2)-5(R1)

3) (R1)-(R2)

4) 2(R2)

The corresponding elementary matrices and their inverses are

$$Q_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \quad Q_1^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix} \quad Q_2^{-1} = \begin{bmatrix} 1 & 0 \\ 5 & 1 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \quad Q_3^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$Q_4 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad Q_4^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

Therefore

$$A = Q_1^{-1}Q_2^{-1}Q_3^{-1}Q_4^{-1}$$

or

$$\begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

Problem 3.26: Write the matrix $\begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & 3 \\ -1 & -1 & 1 \end{bmatrix}$ as a product of elementary matrices.

Determinants: definition

We have already encountered determinants for 2×2 and 3×3 matrices. For 2×2 matrices

$$\det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

For 3×3 matrices we can define the determinant by expanding along the top row:

$$\det \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = a_{1,1} \det \begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix} - a_{1,2} \det \begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} + a_{1,3} \det \begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$$

If we multiply out the 2×2 determinants in this definition we arrive at the expression

$$\det \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{2,3}a_{3,2} + a_{1,2}a_{2,3}a_{3,1} - a_{1,2}a_{2,1}a_{3,3} + a_{1,3}a_{2,1}a_{1,2} - a_{1,3}a_{2,2}a_{3,1}$$

We now make a similar definition for an $n \times n$ matrix. Let A be an $n \times n$ matrix. Define $M_{i,j}$ to be the $(n-1) \times (n-1)$ matrix obtained by crossing out the ith row and the jth column. So, for example, if

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 0 & 1 & 2 \\ 3 & 4 & 5 & 6 \end{bmatrix}$$

then

$$M_{1,2} = \begin{bmatrix} \times & \times & \times & \times \\ 5 & \times & 7 & 8 \\ 9 & \times & 1 & 2 \\ 3 & \times & 5 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 8 \\ 9 & 1 & 2 \\ 3 & 5 & 6 \end{bmatrix}$$

We now define the determinant of an $n \times n$ matrix A to be

$$\det(A) = a_{1,1} \det(M_{1,1}) - a_{1,2} \det(M_{1,2}) + \cdots \pm a_{1,n} \det(M_{1,n}) = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det(M_{1,j}).$$

Of course, this formula still contains determinants on the right hand side. However, they are determinants of $(n-1) \times (n-1)$ matrices. If we apply this definition to those determinants we get a more complicated formula involving $(n-2) \times (n-2)$ matrices, and so on, until we arrive at an extremely long expression (with $n!$ terms) involving only numbers.

Calculating an expression with $n!$ is completely impossible, even with the fastest computers, when n gets reasonable large. For example $100! =$ 9332621544 3944152681 6992388562 6670049071 5968264381 6214685929 6389521759 9993229915 6089414639 7615651828 6253697920 8272237582 5118521091 6864000000 0000000000 00000000 Yet, your computer at home can probably compute the determinant of a 100×100 matrix in a few seconds. The secret, of course, is to compute the determinant in a different way. We start by computing the determinant of triangular matrices.

Triangular matrices

Recall that triangular matrices are matrices whose entries above or below the diagonal are all zero. For 2×2 matrices

$$\det\begin{bmatrix} a_{1,1} & a_{1,2} \\ 0 & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}0 = a_{1,1}a_{2,2}$$

and

$$\det\begin{bmatrix} a_{1,1} & 0 \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - 0a_{2,1} = a_{1,1}a_{2,2}$$

so the determinant is the product of the diagonal elements. For 3×3 matrices

$$\begin{aligned}\det\begin{bmatrix} a_{1,1} & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} &= a_{1,1}\det\begin{bmatrix} a_{2,2} & 0 \\ a_{3,2} & a_{3,3} \end{bmatrix} - 0 + 0 \\ &= a_{1,1}a_{2,2}a_{3,3}\end{aligned}$$

A similar expansion shows that the determinant of an $n \times n$ lower triangular matrix is the product of the diagonal elements. For upper triangular matrices we have

$$\det\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ 0 & 0 & a_{3,3} \end{bmatrix} = a_{1,1}\det\begin{bmatrix} a_{2,2} & a_{2,3} \\ 0 & a_{3,3} \end{bmatrix} - a_{1,2}\det\begin{bmatrix} 0 & a_{2,3} \\ 0 & a_{3,3} \end{bmatrix} + a_{1,3}\det\begin{bmatrix} 0 & a_{2,2} \\ 0 & 0 \end{bmatrix}$$

Since we already know that the determinant of a 2×2 triangular matrix is the product of the diagonals, we can see easily that the last two terms in this expression are zero. Thus we get

$$\begin{aligned}\det\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ 0 & 0 & a_{3,3} \end{bmatrix} &= a_{1,1}\det\begin{bmatrix} a_{2,2} & a_{2,3} \\ 0 & a_{3,3} \end{bmatrix} \\ &= a_{1,1}a_{2,2}a_{3,3}\end{aligned}$$

Once we know that the determinant of a 3×3 upper triangular matrix is the product of the diagonal elements, we can do a similar calculation to the one above to conclude that determinant of a 4×4 upper triangular matrix is the product of the diagonal elements, and so on.

Thus, the determinant of any (upper or lower) triangular $n\times n$ matrix is the product of the diagonal elements.

We know that an arbitrary $n \times n$ matrix can be reduced to an upper (or lower) triangular matrix by a sequence of row operations. This is the key to computing the determinant efficiently. We need to determine how the determinant of a matrix changes when we do an elementary row operation on it.

Exchanging two rows changes the sign of the determinant

We start with the elementary row operation of exchanging two rows. For 2×2 determinants,

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

while

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

so exchanging two rows changes the sign of the determinant.

We can do a similar calculation for 3×3 matrices. Its a a bit messier, but still manageable. Again, we find that exchanging two rows changes the sign of the determinant.

How about the $n \times n$ case? We will assume that we have already proved the result for the $(n-1) \times (n-1)$ case, and show how we can use this to show the result for an $n \times n$ matrix. Thus knowing the result for 2×2 matrices, implies it for 3×3, which in turn implies it for 4×4 matrices, and so on. We consider three cases, depending on which rows we are exchanging. Suppose A is the original matrix and A' is the matrix with two rows exchanged.

(1) Exchanging two row other than the first row: In this case we cross out the first row and any column from A' we obtain $M'_{1,j}$ which is the same as the matrix $M_{1,j}$ (corresponding to A) except with two of its rows exchanged. Since the size of $M_{1,j}$ is $n-1$ we know that $\det(M'_{1,j}) = -\det(M_{1,j})$ so

$$\begin{aligned}\det(A') &= \sum_{j=1}^{n}(-1)^{j+1}a_{1,j}\det(M'_{1,j})\\ &= -\sum_{j=1}^{n}(-1)^{j+1}a_{1,j}\det(M_{1,j})\\ &= -\det(A)\end{aligned}$$

(2) Exchanging the first and second row. Do see that this changes the sign of the determinant we have to expand the expansion. The following is a bit sketchy. I'll probably skip it in class, but give the argument here for completeness. If we expand $M_{1,j}$ we get

$$\det(M_{1,j}) = \sum_{k=1}^{j-1}(-1)^{k+1}a_{2,k}\det(M_{1,2,j,k}) + \sum_{k=j+1}^{n}(-1)^{k}a_{2,k}\det(M_{1,2,j,k})$$

where $M_{1,2,j,k}$ is the matrix obtained from A by deleting the first and second rows, and the jth and kth columns. Inserting this into the expansion for A gives

$$\det(A) = \sum_{j=1}^{n}\sum_{k=1}^{j-1}(-1)^{j+k}a_{1,j}a_{2,k}\det(M_{1,2,j,k}) - \sum_{j=1}^{n}\sum_{k=j+1}^{n}(-1)^{j+k}a_{1,j}a_{2,k}\det(M_{1,2,j,k})$$

The sum splits into two parts. Flipping the first two rows of A just exchanges the two sums. In other words $S - R$ becomes $R - S$ which is $-(S-R)$. So exchanging the first two rows also changes the sign of the determinant.

(3) Exchanging the first row with the kth row. We can effect this exchange by first exchanging the kth and the second row, then exchanging the first and the second row, then exchanging the kth and the second row again. Each flip changes the determinant by a minus sign, and since there are three flips, the overall change is by a minus sign.

Thus we can say that for any $n \times n$ matrix, exchanging two rows changes the sign of the determinant.

One immediate consequence of this fact is that a matrix with two rows the same has determinant zero. This is because if exchange the two rows the determinant changes by a minus sign, but the matrix doesn't change. Thus $\det(A) = -\det(A)$ which is only possible if $\det(A) = 0$.

The determinant is linear in each row separately

To say that the determinant is linear in the jth row means that if we write a matrix as a matrix of row vectors,

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

then

$$\det(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ s\mathbf{b} + t\mathbf{c} \\ \vdots \\ \mathbf{a}_n \end{bmatrix}) = s\det(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{b} \\ \vdots \\ \mathbf{a}_n \end{bmatrix}) + t\det(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{c} \\ \vdots \\ \mathbf{a}_n \end{bmatrix})$$

It is easy to from the expansion formula that the determinant is linear in the first row. For a 3×3 example we have

$$\begin{aligned}
\det(\begin{bmatrix} sb_1 + tc_1 & sb_2 + tc_2 & sb_3 + tc_3 \\ a_{2,1} & a_{2,2} & a2,3 \\ a_{3,1} & a_{3,2} & a3,3 \end{bmatrix}) \\
&= (sb_1 + tc_1)\det(M_{1,1}) - (sb_2 + tc_2)\det(M_{1,2}) + (sb_3 + tc_3)\det(M_{1.3}) \\
&= s(b_1\det(M_{1,1}) - b_2\det(M_{1,2}) + b_3\det(M_{1.3})) \\
&\quad + t(c_1\det(M_{1,1}) - c_2\det(M_{1,2}) + c_3\det(M_{1.3})) \\
&= s\det(\begin{bmatrix} b_1 & b_2 & b_3 \\ a_{2,1} & a_{2,2} & a2,3 \\ a_{3,1} & a_{3,2} & a3,3 \end{bmatrix}) + t\det(\begin{bmatrix} c_1 & c_2 & c_3 \\ a_{2,1} & a_{2,2} & a2,3 \\ a_{3,1} & a_{3,2} & a3,3 \end{bmatrix})
\end{aligned}$$

A similiar calculation can be done for any $n \times n$ matrix to show linearity in the first row. To show linearity in some other row, we first swap that row and the first row, then use linearity in the first row, and then swap back again. So

$$
\begin{aligned}
\det\left(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ s\mathbf{b}+t\mathbf{c} \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right) &= -\det\left(\begin{bmatrix} s\mathbf{b}+t\mathbf{c} \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right) \\
&= -s\det\left(\begin{bmatrix} \mathbf{b} \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right) - t\det\left(\begin{bmatrix} \mathbf{c} \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right) \\
&= s\det\left(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{b} \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right) + t\det\left(\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{c} \\ \vdots \\ \mathbf{a}_n \end{bmatrix}\right)
\end{aligned}
$$

Notice that linearity in each row separately does **not** mean that $\det(A+B) = \det(A) + \det(B)$.

Multiplying a row by a constant mulitplies the determinant by the constant

This is a special case of linearity.

Adding a multiple of one row to another doesn't change the determinant

Now we will see that the most often used row operation—adding a multiple of one row to another—doesn't change the determinant at all. Let A be an $n \times n$ matrix. Write A as a matrix of rows.

$$
A = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}
$$

Adding s times the ith row to the jth row yields

$$A' = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j + s\mathbf{a}_i \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

So

$$\det(A') = \det(\begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j + s\mathbf{a}_i \\ \vdots \\ \mathbf{a}_n \end{bmatrix}) = \det(\begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}) + s\det(\begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_n \end{bmatrix}) = \det(A) + 0$$

Here we used linearity in a row and the fact that the determinant of a matrix with two rows the same is zero.

Calculation of determinant using row operations

We can now use elementary row operations to compute the determinant of

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

The sequence of row operations that transforms this matrix into an upper triangular one is (R2)-(R1), (R3)-2(R1), exchange (R2) and (R3). The determinant doesn't change under the first two transformations, and changes sign under the third. Thus

$$\begin{aligned} \det(\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 1 \\ 2 & 3 & 0 \end{bmatrix}) &= \det(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 2 & 3 & 0 \end{bmatrix}) \\ &= \det(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 0 & -1 & -6 \end{bmatrix}) \\ &= -\det(\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -6 \\ 0 & 0 & -2 \end{bmatrix}) \\ &= -(1)(-1)(-2) = -2 \end{aligned}$$

Problem 3.27: Find the determinant of

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix}$$

Problem 3.28: Find the determinant of
$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 2 & 4 & 8 \\ 1 & -2 & 4 & -8 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

The determinant of QA

To begin, we compute the determinants of the elementary matrices. Recall that if A' is the matrix obtained from A by an elementary row operation, then

(1) $\det(A') = -\det(A)$ if the row operation is swapping two rows

(2) $\det(A') = s\det(A)$ if the row operation is multiplying a row by s

(2) $\det(A') = \det(A)$ if the row operation is adding a multiple of one row to another

Recall that the elementary matrices are obtained from the identity matrix I by an elementary row operation. So we can take $A = I$ and $A' = Q$ in the formulas above to obtain

(1) $\det(Q) = -\det(I) = -1$ if the row operation is swapping two rows

(2) $\det(Q) = s\det(I) = s$ if the row operation is multiplying a row by s

(2) $\det(Q) = \det(I) = 1$ if the row operation is adding a multiple of one row to another

Going back to the first set of formulas, we have that in each case $A' = QA$. In each case the factor in front of $\det(A)$ is exactly $\det(Q)$ So we see that in each case

$$\det(QA) = \det(Q)\det(A).$$

This formula can be generalized. If Q_1, Q_2, ..., Q_k are elementary matrices then $\det(Q_1Q_2Q_3\cdots Q_kA) = \det(Q_1)\det(Q_2Q_3\cdots Q_kA) = \det(Q_1)\det(Q_2)\det(Q_3\cdots Q_kA)$ and so on, so we arrive at the formula

$$\det(Q_1Q_2Q_3\cdots Q_kA) = \det(Q_1)\det(Q_2)\cdots\det(Q_k)\det(A).$$

The determinant of A is zero exactly when A is not invertible

Recall that if R denotes the reduced form of A, obtained by performing the sequence of row reductions corresponding to Q_1, Q_2, ..., Q_k, then

$$A = Q_1^{-1}Q_2^{-1}\cdots Q_k^{-1}R$$

Each Q_i^{-1} is an elementary matrix, therefore

$$\det(A) = \det(Q_1^{-1})\det(Q_2^{-1})\cdots\det(Q_k^{-1})\det(R)$$

If A is not invertible, then R has a row of zeros along the bottom. Thus R is an upper triangular matrix with at least one zero on the diagonal. The determinant of R is the product of the diagonal elements so $\det(R) = 0$. Thus $\det(A) = 0$ too.

If A is invertible, then we can reduce A to to identity matrix. In other words, we can take $R = I$. Then $\det(R) = 1$. Each $\det(Q_i^{-1})$ is non-zero too, so $\det(A) \neq 0$.

Inverses of Products

If both A and B are invertible, then so is AB. The inverse of AB is given by $B^{-1}A^{-1}$. To check this, simply compute

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

If one of A or B is not invertible then AB is not invertible. To see this recall that a matrix C is not invertible exactly whenever there is a non-zero solution $\mathbf{x}$ to $C\mathbf{x} = \mathbf{0}$.

If B is not invertible, then there is a non-zero vector $\mathbf{x}$ with $B\mathbf{x} = \mathbf{0}$. Then $AB\mathbf{x} = A\mathbf{0} = \mathbf{0}$ so AB is not invertible too.

If B is invertible, but A is not, then there is a non-zero $\mathbf{x}$ with $A\mathbf{x} = 0$. Let $\mathbf{y} = B^{-1}\mathbf{x}$. Since B^{-1} is invertible, $\mathbf{y}$ cannot be zero. We have $AB\mathbf{y} = ABB^{-1}\mathbf{x} = A\mathbf{x} = \mathbf{0}$ so AB is not invertible.

The product formula: $\det(AB) = \det(A)\det(B)$

If either A or B is non-invertible, then AB is non-invertible too. Thus $\det(AB) = 0$ and one of $\det(A)$ or $\det(B)$ is zero, so $\det(A)\det(B) = 0$ too. Thus $\det(AB) = \det(A)\det(B)$.

If both A and B are invertible, then

$$A = Q_1^{-1}Q_2^{-1}\cdots Q_k^{-1}$$

so

$$\det(A) = \det(Q_1^{-1})\det(Q_2^{-1})\cdots\det(Q_k^{-1})$$

and

$$B = \tilde{Q}_1^{-1}\tilde{Q}_2^{-1}\cdots\tilde{Q}_j^{-1}$$

so

$$\det(B) = \det(\tilde{Q}_1^{-1})\det(\tilde{Q}_2^{-1})\cdots\det(\tilde{Q}_j^{-1})$$

Therefore

$$AB = Q_1^{-1}Q_2^{-1}\cdots Q_k^{-1}\tilde{Q}_1^{-1}\tilde{Q}_2^{-1}\cdots\tilde{Q}_j^{-1}$$

so

$$\det(AB) = \det(Q_1^{-1})\det(Q_2^{-1})\cdots\det(Q_k^{-1})\det(\tilde{Q}_1^{-1})\det(\tilde{Q}_2^{-1})\cdots\det(\tilde{Q}_j^{-1}) = \det(A)\det(B)$$

The determinant of the transpose

Recall that the transpose A^T of a matrix A is the matrix you get when you flip A about its diagonal. If A is an $n \times n$ matrix, so is A^T and we can ask what the relationship between the determinants of these two matrices is. It turns out that they are the same.

$$\det(A^T) = \det(A).$$

If A is an upper or lower triangular matrix, this follows from the fact that the determinant of a triangular matrix is the product of the diagonal entries. If A is an arbitrary $n \times n$ matrix then the formula follows from two facts.

(1) The transpose of a product of two matrices is given by $(AB)^T = B^T A^T$

This implies that $(A_1 A_2 \cdots A_n)^T = A_n^T \cdots A_2^T A_1^T$.

(2) For an elementary matrix Q we have $\det(Q^T) = \det(Q)$.

If you accept these two facts, then we may write

$$A = Q_1^{-1} Q_2^{-1} \cdots Q_k^{-1} R$$

where R is upper triangular. Thus

$$A^T = R^T (Q_k^{-1})^T \cdots (Q_2^{-1})^T (Q_1^{-1})^T$$

so

$$\begin{aligned}
\det(A^T) &= \det(R^T)\det((Q_k^{-1})^T) \cdots \det((Q_1^{-1})^T) \\
&= \det(R)\det(Q_k^{-1}) \cdots \det(Q_1^{-1}) \\
&= \det(Q_1^{-1}) \cdots \det(Q_k^{-1})\det(R) \\
&= \det(A)
\end{aligned}$$

More expansion formulas

We can use the properties of the determinant to derive alternative expansion formulas. Recall that we defined the determinant to be

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det(M_{1,j}).$$

In other words, we expanded along the top row. Now lets see that we can expand along other rows as well. Let A be the original matrix with rows $\mathbf{a}_1 = [a_{1,1}, a_{1,2}, \ldots, a_{1,n}]$, $\ldots$ $\mathbf{a}_n = [a_{n,1}, a_{n,2}, \ldots, a_{n,n}]$. For example, if A is a 5×5 matrix then

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \\ \mathbf{a}_5 \end{bmatrix}$$

Suppose we want to expand along the fourth row. Let A' be the matrix, where the fourth row of A has been moved to the first row, with all other rows still in the same order, i.e.,

$$A' = \begin{bmatrix} \mathbf{a}_4 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_5 \end{bmatrix}$$

How is the determinant of A' related to the determinant of A? We can change A to A' be a series of row flips as follows:

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \\ \mathbf{a}_5 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_4 \\ \mathbf{a}_3 \\ \mathbf{a}_5 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_4 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_5 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{a}_4 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_5 \end{bmatrix} = A'$$

We have performed 3 flips, so $\det(A') = (-1)^3 \det(A) = -\det(A)$.

In general, to move the ith row to the top in this way, we must perform $i-1$ flips, so $\det(A') = (-1)^{i-1} \det(A)$

Notice that A' is a matrix with the properties

(1) $a'_{1,j} = a_{i,j}$, since we have moved the ith row to the top

(2) $M'_{1,j} = M_{i,j}$, since we haven't changed the order of the other rows.

Therefore

$$\begin{aligned} \det(A) &= (-1)^{i-1} \det(A') \\ &= (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j+1} a'_{1,j} \det(M'_{1,j}) \\ &= (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j+1} a_{i,j} \det(M_{i,j}) \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(M_{i,j}) \end{aligned}$$

This is the formula for expansion along the ith row.

As an example lets compute the determinant of a 3×3 matrix by expanding along the second row.

$$\begin{aligned} \det \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix} &= -\det \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} + 3 \det \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} - \det \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \\ &= -2 + 6 + 3 - 9 - 2 + 2 = -2 \end{aligned}$$

The formula for expanding along the ith row is handy if the matrix happens to have a row with many zeros.

Using the fact that $\det(A) = \det(A^T)$ we can also write down expansion formulas along columns, since the columns of A are the rows of A^T.

We end up with the formula

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(M_{i,j})$$

As an example lets compute the determinant of a 3×3 matrix by expanding along the second column.

$$\begin{aligned} \det \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix} &= -2 \det \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 3 \det \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} - 2 \det \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \\ &= -2 + 2 + 3 - 9 - 2 + 6 = -2 \end{aligned}$$

The formula for expanding along the ith row is handy if the matrix happens to have a column with many zeros.

An impractical formula for the inverse

We can use the expansion formulas of the previous section to obtain a formula for the inverse of a matrix A. This formula is really only practical for 3×3 matrices, since for larger matrices, the work involved in computing the determinants appearing is prohibitive.

We begin with the expansion formula

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(M_{i,j})$$

If A is invertible, then $\det(A) \neq 0$ so we can divide by it to obtain

$$1 = \sum_{j=1}^{n} a_{i,j} \frac{(-1)^{i+j} \det(M_{i,j})}{\det(A)}$$

Now suppose we take the matrix A and replace the ith row by the kth row for some $k \neq i$. The resulting matrix A' has two rows the same, so its determinant is zero. Its expansion is the same as that for A, except that $a_{i,j}$ is replaced by $a_{k,j}$. Thus, if $k \neq i$

$$0 = \sum_{j=1}^{n} (-1)^{i+j} a_{k,j} \det(M_{i,j})$$

Dividing by $\det(A)$ yields

$$0 = \sum_{j=1}^{n} a_{k,j} \frac{(-1)^{i+j} \det(M_{i,j})}{\det(A)}$$

Now let B be the matrix with entries

$$b_{i,j} = \frac{(-1)^{i+j} \det(M_{j,i})}{\det(A)}$$

This turns out to be the inverse A^{-1}.

It gets a bit confusing with all the indices, but lets think about what we need to show. The k, ith entry of the product AB is given by

$$\begin{aligned} (AB)_{k,i} &= \sum_{j=1}^{n} a_{k,j} b_{j,i} \\ &= \sum_{j=1}^{n} a_{k,j} \frac{(-1)^{i+j} \det(M_{i,j})}{\det(A)} \end{aligned}$$

According to the formulas above, this sum is equal to 1 if $k = i$ and equal to 0 if $k \neq i$. In other words, AB is the identity matrix I. This shows that $B = A^{-1}$

Cramer's rule

Given an impractical way to compute the inverse, we can derive an impractical formula for the solution of a system of n equations in n unknowns, i.e., a matrix equation

$$A\mathbf{x} = \mathbf{b}$$

The solution $\mathbf{x}$ is equal to $A^{-1}\mathbf{b}$. So if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$, then using the formula of the previous section for the inverse, we get

$$x_i = \sum_{j=1}^{n} A^{-1}_{i,j} b_j = \sum_{j=1}^{n} \frac{(-1)^{i+j} \det(M_{j,i})}{\det(A)} b_j$$

but this is exactly $(1/\det(A))$ times the formula for expanding the determinant matrix obtained from A by replacing the ith column with $\mathbf{b}$. Thus

$$x_i = \frac{\det(\text{matrix obtained from } A \text{ by replacing the } i\text{th column with } \mathbf{b})}{\det(A)}$$

Problem 3.29: Compute

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

by expanding along the second row, and by expanding along the third column.

Problem 3.30: Find the inverse of $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 3 \\ 3 & 0 & 1 \end{bmatrix}$ using the impractical formula

Problem 3.31: Solve the equation

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 3 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

using Cramer's rule.
