Introduction to Matrices

Tom Davis
tomrdavis@earthlink.net
http://www.geometer.org/mathcircles
October 25, 2000

1 Definitions

A matrix (plural: matrices) is simply a rectangular array of "things". For now, we'll assume the "things" are numbers, but as you go on in mathematics, you'll find that matrices can be arrays of very general objects. Pretty much all that's required is that you be able to add, subtract, and multiply the "things".

Here are some examples of matrices. Notice that it is sometimes useful to have variables as entries, as long as the variables represent the same sorts of "things" as appear in the other slots. In our examples, we'll always assume that all the slots are filled with numbers. All our examples contain only real numbers, but matrices of complex numbers are very common.

$$\begin{pmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 4 & x & 17 \\ 2 & x+y & 7 & -19 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix}, \quad \begin{pmatrix} x & y & z & w \end{pmatrix}$$

The first example is a square 3×3 matrix; the next is a 2×4 matrix (2 rows and 4 columns—if we talk about a matrix that is "$m \times n$" we mean it has m rows and n columns). The final two examples consist of a single column matrix, and a single row matrix. These final two examples are often called "vectors"—the first is called a "column vector" and the second, a "row vector". We'll use only column vectors in this introduction.

Often we are interested in representing a general $m \times n$ matrix with variables in every location, and that is usually done as follows:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

The number in row i and column j is represented by a_{ij}, where $1 \leq i \leq m$ and $1 \leq j \leq n$. Sometimes when there is no question about the dimensions of a matrix, the entire matrix can simply be referred to as:

$$\left(a_{ij}\right).$$

1.1 Addition and Subtraction of Matrices

As long as you can add and subtract the "things" in your matrices, you can add and subtract the matrices themselves. The addition and subtraction occurs in the obvious way—element by element. Here are a couple of examples:

$$\begin{pmatrix} 1 & 3 & 7 \\ 2 & 6 & -4 \\ 2 & 15 & \pi \end{pmatrix} + \begin{pmatrix} 3 & 2 & 1 \\ 5.5 & 3 & -e \\ 2 & 5 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 4 & 5 & 8 \\ 7.5 & 9 & -4-e \\ 4 & 20 & \pi+\sqrt{2} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 & 7 \\ 2 & 6 & -4 \\ 2 & 15 & \pi \end{pmatrix} - \begin{pmatrix} 3 & 2 & 1 \\ 5.5 & 3 & -e \\ 2 & 5 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} -2 & 1 & 6 \\ -3.5 & 3 & e-4 \\ 0 & 10 & \pi - \sqrt{2} \end{pmatrix}$$

To find what goes in row i and column j of the sum or difference, just add or subtract the entries in row i and column j of the matrices being added or subtracted.

In order to make sense, both of the matrices in the sum or difference must have the same number of rows and columns. It makes no sense, for example, to add a 2×4 matrix to a 3×4 matrix.

1.2 Multiplication of Matrices

When you add or subtract matrices, the two matrices that you add or subtract must have the same number of rows and the same number of columns. In other words, both must have the same shape.

For matrix multiplication, all that is required is that the number of columns of the first matrix be the same as the number of rows of the second matrix. In other words, you can multiply an $m \times k$ matrix by a $k \times n$ matrix, with the $m \times k$ matrix on the left and the $k \times n$ matrix on the right. The example on the left below represents a legal multiplication since there are three columns in the left multiplicand and three rows in the right one; the example on the right doesn't make sense—the left matrix has three columns, but the right one has only 2 rows. If the matrices on the right were written in the reverse order with the 2×3 matrix on the left, it would represent a valid matrix multiplication.

$$\begin{pmatrix} 1 & 3 & 5 \\ 4 & 7 & 2 \\ 9 & 1 & 6 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 7 \\ 3 & 7 \\ 7 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}$$

So now we know what shapes of matrices it is legal to multiply, but how do we do the actual multiplication? Here is the method:

If we are multiplying an $m \times k$ matrix by a $k \times n$ matrix, the result will be an $m \times n$ matrix. The element in the product in row i and column j is gotten by multiplying term-wise all the elements in row i of the matrix on the left by all the elements in column j of the matrix on the right and adding them together.

Here is an example:

$$\begin{pmatrix} 1 & 3 & 2 \\ 5 & 0 & 7 \\ 6 & 9 & 8 \end{pmatrix} \begin{pmatrix} 4 & 11 \\ 6 & 10 \\ 5 & 9 \end{pmatrix} = \begin{pmatrix} 32 & 59 \\ 55 & 118 \\ 118 & 228 \end{pmatrix}$$

To find what goes in the first row and first column of the product, take the number from the first row of the matrix on the left: $(1, 3, 2)$, and multiply them, in order, by the numbers in the first column of the matrix on the right: $(4, 6, 5)$. Add the results: $1 \cdot 4 + 3 \cdot 6 + 2 \cdot 5 = 4 + 18 + 10 = 32$. To get the 228 in the third row and second column of the product, the use the numbers in the third row of the left matrix: $(6, 9, 8)$ and the numbers in the second column of the right matrix: $(11, 10, 9)$ to get $6 \cdot 11 + 9 \cdot 10 + 8 \cdot 9 = 66 + 90 + 72 = 228$.

Check your understanding by verifying that the other elements in the product matrix are correct.

In general, if we multiply a general $m \times k$ matrix by a general $k \times n$ matrix to get an $m \times n$ matrix as follows:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}$$

Then we can write c_{ij} (the number in row i, column j) as:

$$c_{ij} = \sum_{p=1}^{k} a_{ip}b_{pj}.$$

1.3 Square Matrices and Column Vectors

Although everything above has been stated in terms of general rectangular matrices, for the rest of this tutorial, we'll consider only two kinds of matrices (but of any dimension): square matrices, where the number of rows is equal to the number of columns, and column matrices, where there is only one column. These column matrices are often called "vectors", and there are many applications where they correspond exactly to what you commonly use as sets of coordinates for points in space. In the two-dimensional x-y plane, the coordinates $(1, 3)$ represent a point that is one unit to the right of the origin (in the direction of the x-axis), and three units above the origin (in the direction of the y-axis). That same point can be written as the following column vector:

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

If you wish to work in three dimensions, you'll need three coordinates to locate a point relative to the (three-dimensional) origin—an x-coordinate, a y-coordinate, and a z-coordinate. So the point you'd normally write as (x, y, z) can be represented by the column vector:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Quite often we will work with a combination of square matrices and column matrices, and in that case, if the square matrix has dimensions $n \times n$, the column vectors will have dimension $n \times 1$ (n rows and 1 column)[1].

1.4 Properties of Matrix Arithmetic

Matrix arithmetic (matrix addition, subtraction, and multiplication) satisfies many, *but not all* of the properties of normal arithmetic that you are used to. All of the properties below can be formally proved, and it's not too difficult, but we will not do so here. In what follows, we'll assume that different matrices are represented by upper-case letters: $M, N, P, \ldots$, and that column vectors are represented by lower-case letters: $v, w, \ldots$.

We will further assume that all the matrices are square matrices or column vectors, and that all are the same size, either $n \times n$ or $n \times 1$. Further, we'll assume that the matrices contain numbers (real or complex). Most of the properties listed below apply equally well to non-square matrices, assuming that the dimensions make the various multiplications and addtions/subtractions valid.

Perhaps the first thing to notice is that we can always multiply two $n \times n$ matrices, and we can multiply an $n \times n$ matrix by a column vector, but we cannot multiply a column vector by the matrix, nor a column vector by another. In other words, of the three matrix multiplications below, only the first one makes sense. Be sure you understand why.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} \qquad \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \qquad \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$$

[1] We could equally well use row vectors to correspond to coordinates, and this convention is used in many places. However, the use of column matrices for vectors is more common

Finally, an extremely useful matrix is called the “identity matrix”, and it is a square matrix that is filled with zeroes except for ones in the diagonal elements (having the same row and column number). Here, for example, is the 4 × 4 identity matrix:

$$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}.$$

The identity matrix is usually called “I” for any size square matrix. Usually you can tell the dimensions of the identity matrix from the surrounding context.

- Associative laws:

$$\begin{array}{ll}(MN)P = M(NP) & (MN)v = M(Nv)\\ (M+N)+P = M+(N+P) & (u+v)+w = u+(v+w)\end{array}$$

- Commutative laws for addition:

$$\begin{array}{ll}M+N = N+M & v+w = w+v\end{array}$$

- Distributive laws:

$$\begin{array}{ll}M(N\pm P) = MN\pm MP & (M\pm N)P = MP\pm NP\\ M(v\pm w) = Mv\pm Mw & (M\pm N)v = Mv\pm Nv\end{array}$$

- The identity matrix:

$$\begin{array}{ll}NI = IN = N & Iv = v\end{array}$$

Probably the most important thing to notice about the laws above is one that’s missing—multiplication of matrices is not in general commutative. It is easy to find examples of matrices M and N where $MN \neq NM$. In fact, matrices almost never commute under multiplication. Here’s an example of a pair that don’t:

$$\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}1&0\\1&1\end{pmatrix}=\begin{pmatrix}2&1\\1&1\end{pmatrix}; \qquad \begin{pmatrix}1&0\\1&1\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}=\begin{pmatrix}1&1\\1&2\end{pmatrix}.$$

So the order of multiplication is very important; that’s why you may have noticed the care that has been taken so far in describing multiplication of matrices in terms of “the matrix on the left”, and “the matrix on the right”.

The associative laws above are extremely useful, and to take one simple example, consider computer graphics. As we’ll see later, operations like rotation, translation, scaling, perspective, and so on, can all be represented by a matrix multiplication. Suppose you wish to rotate all the vectors in your drawing and then to translate the results. Suppose R and T are the rotation and translation matrices that do these jobs. If your picture has a million points in it, you can take each of those million points v and rotate them, calculating Rv for each vector v. Then, the result of that rotation can be translated: $T(Rv)$, so in total, there are two million matrix multiplications to make your picture. But the associative law tells us we can just multiply T by R once to get the matrix TR, and then multiply all million points by TR to get $(TR)v$, so all in all, there are only 1,000,001 matrix multiplications— one to produce TR and a million multiplications of TR by the individual vectors. That’s quite a savings of time.

The other thing to notice is that the identity matrix behaves just like 1 under multiplication—if you multiply any number by 1, it is unchanged; if you multiply any matrix by the identity matrix, it is unchanged.

2 Applications of Matrices

This section illustrates a tiny number of applications of matrices to real-world problems. Some details in the solutions have been omitted, but that's because entire books are written on some of the techniques. Our goal is to make it clear how a matrix formulation may simplify the solution.

2.1 Systems of Linear Equations

Let's start by taking a look at a problem that may seem a bit boring, but in terms of practical applications is perhaps the most common use of matrices: the solution of systems of linear equations. Following is a typical problem (although real-world problems may have hundreds of variables).

Solve the following system of equations:

$$\begin{aligned} x + 4y + 3z &= 7 \\ 2x + 5y + 4z &= 11 \\ x - 3y - 2z &= 5. \end{aligned}$$

The key observation is this: the problem above can be converted to matrix notation as follows:

$$\begin{pmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \\ 5 \end{pmatrix}. \tag{1}$$

The numbers in the square matrix are just the coefficients of x, y, and z in the system of equations. Check to see that the two forms—the matrix form and the system of equations form—represent exactly the same problem.

Ignoring all the difficult details, here is how such systems can be solved. Let M be the 3×3 square matrix in equation (1) above, so the equation looks like this:

$$M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \\ 5 \end{pmatrix}. \tag{2}$$

Suppose we can somehow find another matrix N such that $NM = I$. If we can, we can multiply both sides of equation (2) by N to obtain:

$$NM \begin{pmatrix} x \\ y \\ z \end{pmatrix} = I \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = N \begin{pmatrix} 7 \\ 11 \\ 5 \end{pmatrix},$$

so we can simply multiply our matrix N by the column matrix containing the numbers 7, 11, and 5 to get our solution.

Without explaining where we got it, the matrix on the left below is just such a matrix N. Check that the multiplication below does yield the identity matrix:

$$\begin{pmatrix} 2 & -1 & 1 \\ 8 & -5 & 2 \\ -11 & 7 & -3 \end{pmatrix} \begin{pmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So we just need to multiply that matrix N by the column vector containing 7, 11, and 5 to get our solution:

$$\begin{pmatrix} 2 & -1 & 1 \\ 8 & -5 & 2 \\ -11 & 7 & -3 \end{pmatrix} \begin{pmatrix} 7 \\ 11 \\ 5 \end{pmatrix} = \begin{pmatrix} 8 \\ 11 \\ -15 \end{pmatrix} = \begin{pmatrix} x \\ y \\ y \end{pmatrix}.$$

From this last equation we conclude that $x = 8$, $y = 11$, and $z = -15$ is a solution to the original system of equations. You can plug them in to check that they do indeed form a solution.

Although it doesn't happen all that often, sometimes the same system of equations needs to be solved for a variety of column vectors on the right—not just one. In that case, the solution to every one can be obtained by a single multiplication by the matrix N.

The matrix N is usually written as M^{-1}, called "M-inverse". It is a multiplicative inverse in just the same way that $1/3$ is the inverse of 3: $3 \cdot (1/3) = 1$, and 1 is the multiplicative identity, just as I is in matrix multiplication. Entire books are written that describe methods of finding the inverse of a matrix, so we won't go into that here.

Remember that for numbers, zero has no inverse; for matrices, it is much worse—many, many matrices do not have an inverse. Matrices without inverses are called "singular". Those with an inverse are called "non-singular".

Just as an example, the matrix on the left of the multiplication below can't possibly have an inverse, as we can see from the matrix on the right. No matter what the values are of $a, b, \ldots, i$, it is impossible to get anything but zeroes in certain spots in the diagonal, and we need ones in all the diagonal spots:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

If the set of linear equations has no solution, then it will be impossible to invert the associated matrix. For example, the following system of equations cannot possibly have a solution, since $x + y + z$ cannot possibly add to two different numbers (7 and 11) as would be required by the first two equations:

$$\begin{aligned} x + y + z &= 7 \\ x + y + z &= 11 \\ x - 3y - 2z &= 5. \end{aligned}$$

So obviously the associated matrix below cannot be inverted:

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

2.2 Computer Graphics

Some computers can only draw straight lines on the screen, but complicated drawings can be made with a long series of line-drawing instructions. For example, the letter "F" could be drawn in its normal orientation at the origin with the following set of instructions:

1. Draw a line from $(0, 0)$ to $(0, 5)$.
2. Draw a line from $(0, 5)$ to $(3, 5)$.
3. Draw a line from $(0, 3)$ to $(2, 3)$.

Imagine that you have a drawing that's far more complicated than the "F" above consisting of thousands of instructions similar to those above. Let's take a look at the following sorts of problems:

1. How would you convert the coordinates so that the drawing would be twice as big? How about stretched twice as high (y-direction) and three times as wide (x-direction)?
2. Could you draw the mirror image through the y-axis?

3. How would you shift the drawing 4 units to the right and 5 units down?

4. Could you rotate it $90°$ counter-clockwise about the origin? Could you rotate it by an angle θ counter-clockwise about the origin?

5. Could you rotate it by an angle θ about the point $(7, -3)$?

6. Ignoring the problem of making the drawing on the screen, what if your "drawing" were in three dimensions? Could you solve problems similar to those above to find the new (3-dimensional) coordinates after your object has been translated, scaled, rotated, et cetera?

It turns out that the answers to all of the problems above can be achieved by multiplying your vectors by a matrix. Of course a different matrix will solve each one. Here are the solutions:

Graphics Solution 1:

To scale in the x-direction by a factor of 2, we need to multiply all the x coordinates by 2. To scale in the y-direction, we similarly need to multiply the y coordinates by the same scale factor of 2. The solution to scale any drawing by a factor s_x in the x-direction and s_y in the y-direction is to multiply all the input vectors by a general scaling matrix as follows:

$$\begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_x x \\ s_y y \end{pmatrix}.$$

To uniformly scale everything to twice as big, let $s_x = s_y = 2$. To scale by a factor of 2 in the x-direction and 3 in the y-direction, let $s_x = 2$ and $s_y = 3$.

We'll illustrate the general procedure with the drawing instructions for the "F" that appeared earlier. The drawing commands are described in terms of a few points: $(0, 0)$, $(0, 5)$, $(3, 5)$, $(0, 3)$, and $(2, 3)$. If we write all five of those points as column vectors and multiply all five by the same matrix (either of the two above), we'll get five new sets of coordinates for the points. For example, in the case of the second example where the scaling is 2 times in x and 3 times in y, the five points will be converted by matrix multiplication to: $(0, 0)$, $(0, 15)$, $(6, 15)$, $(0, 9)$ and $(4, 9)$. If we rewrite the drawing instructions using these transformed points, we get:

1. Draw a line from $(0, 0)$ to $(0, 15)$.

2. Draw a line from $(0, 15)$ to $(6, 15)$.

3. Draw a line from $(0, 9)$ to $(6, 9)$.

Follow the instructions above and see that you draw an appropriately stretched "F". In fact, do the same thing for each of the matrix solutions in this set to verify that the drawing is transformed appropriately. Notice that if s_x or s_y is smaller than 1, the drawing will be shrunk—not expanded.

Graphics Solution 2:

A mirror image is just a scaling by -1. To mirror through the y-axis means that each x-coordinate will be replaced with its negative. Here's a matrix multiplication that will do the job:

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_x x \\ s_y y \end{pmatrix}.$$

Graphics Solution 3:

To translate points 4 to the right and 5 units down, you essentially need to add 4 to every x coordinate and to subtract 5 from every y coordinate. If you try to solve this exactly as in the examples above, you'll find it is impossible. To convince yourself it's impossible with *any* 2×2 matrix, consider what will happen to the origin: $(0, 0)$. You want to move it to $(4, -5)$, but look what happens if you multiply it by *any* 2×2 matrix (a, b, c, and d can be any numbers):

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

In other words, no matter what a, b, c, and d are, the matrix will map the origin back to the origin, so translation using this scheme is impossible.

But there's a great trick[2]. For every one of your two-dimensional vectors, add an artificial third component of 1. So the point $(3, 6)$ will become $(3, 6, 1)$, the origin will become $(0, 0, 1)$, et cetera. The column vectors will now have three rows, so the transformation matrix will need to be 3×3. To translate by t_x in the x-direction and t_y in the y-direction, multiply the artificially-enlarged vector by a matrix as follows:

$$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + t_x \\ y + t_y \\ 1 \end{pmatrix}.$$

The resulting vector is just what you want, and it also has the artificial 1 on the end that you can just throw away. To get the particular solution to the problem proposed above, simply let $t_x = 4$ and $t_y = -5$.

But now you're probably thinking, "That's a neat trick, but what happens to the matrices we had for scaling? What a pain to have to convert to the artificial 3-dimensional form and back if we need to mix scaling and translation." The nice thing is that we can *always* use the artificially extended form. Just use a slightly different form of the scaling matrix:

$$\begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} s_x x \\ s_y y \\ 1 \end{pmatrix}.$$

In the solutions that follow, we'll always add a 1 as the artificial third component[3].

Graphics Solution 4:

Convince yourself (by drawing a few examples, if necessary) that to rotate a point counter-clockwise by 90° about the origin, you will basically make the original x coordinate into a y coordinate, and vice-versa. But not quite. Anything that had a positive y coordinate will, after rotation by 90°, have a negative x coordinate and vice-versa. In other words, the new y coordinate is the old x coordinate, and the new x coordinate is the negative of the old y coordinate. Convince yourself that the following matrix does the trick (and notice that we've added the 1 as an artificial third component):

$$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} -y \\ x \\ 1 \end{pmatrix}.$$

The general solution for a rotation counter-clockwise by an angle θ is given by the following matrix multiplication:

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \\ 1 \end{pmatrix}.$$

If you've never seen anything like this before, you might consider trying it for a couple of simple angles, like $\theta = 45°$ or $\theta = 30°$ and put in the drawing coordinates for the letter "F" given earlier to see that it's transformed properly.

[2] In fact, it's a lot more than a trick—it is really part of projective geometry.

[3] But in the world of computer graphics or projective geometry, it is often useful to allow values other than 1—in perspective transformations, for example

Graphics Solution 5:

Here is where the power of matrices really comes through. Rather than solve the problem from scratch as we have above, let's just solve it using the information we already have. Why not translate the point $(7, -3)$ to the origin, then do a rotation about the origin, and finally, translate the result back to $(7, -3)$? Each of those operations can be achieved by a matrix multiplication. Here is the final solution:

$$\begin{pmatrix}1 & 0 & 7\\ 0 & 1 & -3\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & -7\\ 0 & 1 & 3\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\\ 1\end{pmatrix}.$$

Notice carefully the order of the matrix multiplication. The matrix closest to the $(x, y, 1)$ column vector is the first one that's applied to it—it should move $(7, -3)$ to the origin. To do that, we need to translate x coordinates by -7 and y coordinates by 3. The next operation to be done is the rotation by an arbitrary angle θ, using the matrix form from the previous problem. Finally, to translate back to $(7, -3)$ we have to translate in the opposite direction from what we did originally, and the matrix on the far left above does just that.

Remember that for any particular value of θ, $\sin\theta$ and $\cos\theta$ are just numbers, so if you knew the exact rotation angle, you could just plug the numbers in to the middle 3×3 matrix and multiply together the three matrices on the left. Then to transform any particular point, there would be only one matrix multiplication involved.

To convince yourself that we've got the right answer, why not put in a particular (simple) rotation of $90°$ into the matrices and work it out? $\cos 90° = 0$ and $\sin 90° = 1$, so the product of the three matrices on the left is:

$$\begin{pmatrix}1 & 0 & 7\\ 0 & 1 & -3\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & -7\\ 0 & 1 & 3\\ 0 & 0 & 1\end{pmatrix} = \begin{pmatrix}0 & -1 & 4\\ 1 & 0 & -10\\ 0 & 0 & 1\end{pmatrix}.$$

Try multiplying all the vectors from the "F" example by the single matrix on the right above and convince yourself that you've succeeded in rotating the "F" by $90°$ counter-clockwise about the point $(7, -3)$.

Graphics Solution 6:

The answer is yes. Of course you'll have to add an artificial fourth dimension which is always 1 to your three-dimensional coordinates, but the form of the matrices will be similar.

On the left below is the mechanism for scaling by s_x, s_y, and s_z in the x-, y-, and z-directions; on the right is a multiplication that translates by t_x, t_y, and t_z in the three directions.

$$\begin{pmatrix}s_x & 0 & 0 & 0\\ 0 & s_y & 0 & 0\\ 0 & 0 & s_z & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\\ 1\end{pmatrix} \qquad \begin{pmatrix}1 & 0 & 0 & t_x\\ 0 & 1 & 0 & t_y\\ 0 & 0 & 1 & t_z\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\\ 1\end{pmatrix}$$

Finally, to rotate by an angle of θ counter-clockwise about the three axes, multiply your vector on the left by the appropriate one of the following three matrices (left, middle, and right correspond to rotation about the x-axis, y-axis, and z-axis:

$$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & \cos\theta & -\sin\theta & 0\\ 0 & \sin\theta & \cos\theta & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \qquad \begin{pmatrix}\cos\theta & 0 & \sin\theta & 0\\ 0 & 1 & 0 & 0\\ -\sin\theta & 0 & \cos\theta & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \qquad \begin{pmatrix}\cos\theta & -\sin\theta & 0 & 0\\ \sin\theta & \cos\theta & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

2.3 Gambler's Ruin Problem

Consider the following questions:

1. Suppose you have \$20 and need \$50 for a bus ride home. Your only chance to get more money is by gambling in a casino. There is only one possible bet you can make at the casino—you must bet \$10, and then a fair coin is flipped. If "heads" results, you win \$10 (in other words, you get your original \$10 back plus an additional \$10), and if it's "tails", you lose the \$10 bet. The coin has exactly the same 50% probability of coming up heads or tails. What is the chance that you will get the \$50 you need? What if you had begun with \$10? What if you had begun with \$30?
2. Same as above, except this time you start with \$40 and need \$100 for the bus ticket. But this time, you can bet either \$10 or \$20 on each flip. What's your best betting strategy?
3. Real casinos don't give you a fair bet. Suppose the problem is the same as above, except that you have to make a \$10 bet on "red" or "black" on a roulette wheel. A roulette wheel in the United States has 18 red numbers, 18 black numbers and 2 green numbers. Winning a red or black bet doubles your money if you win (and you lose it all if you lose), but obviously you now have a slightly smaller chance of winning. A red or black bet has a probability of $18/38 = 9/19$ of winning on each spin of the wheel. Answer all the questions above if your bets must be made on a roulette wheel.

All of these problems are known as the "Gambler's Ruin Problem"—a gambler keeps betting until he goes broke, or until he reaches a certain goal, and there are no other possibilities. The problem has many elegant solutions which we will ignore. Let's just assume we're stupid and lazy, but we have a computer available to simulate the problem. Here is a nice matrix-based approach for stupid lazy people.

At any stage in the process, there are six possible states, depending on your "fortune". You have either \$0 (and you've lost), or you have \$10, \$20, \$30, \$40 (and you're still playing), or you have \$50 (and you've won). You know that when you begin playing, you are in a certain state (having \$20 in the case of the very first problem). After you've played a while, lots of different things could have happened, so depending on how long you've been going, you have various probabilities of being in the various states. Call the state with \$0 "state 0", and so on, up to "state 5" that represents having \$50.

At any particular time, let's let p_0 represent the probability of having \$0, p_1 the probability of having \$10, and so on, up to p_5 of having won with \$50.

We can give an entire probabilistic description of your state as a column vector like this:

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix}.$$

Now look at the individual situations. If you're in state 0, or in state 5, you will remain there for sure. If you're in any other state, there's a 50% chance of moving up a state and a 50% chance of moving down a state. Look at the following matrix multiplication:

$$\begin{pmatrix} 1 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \begin{pmatrix} p_0 + .5p_1 \\ .5p_2 \\ .5p_1 + .5p_3 \\ .5p_2 + .5p_4 \\ .5p_3 \\ .5p_4 + p_5 \end{pmatrix}. \tag{3}$$

Clearly, multiplication by the matrix above represents the change in probabilities of being in the various states, given an initial probabilistic distribution. The chance of being in state 0 after a coin flip is the chance you were there before, plus half of the chance that you were in state 1. Check that the others make sense as well.

So if the matrix on the left in equation (3) is called P, each time you multiply the vector corresponding to your initial situation by P, you'll find the probabilities of being in the various states. So after 1000 coin-flips, your state will be represented by $P^{1000}v$, where v is the vector representing your original situation ($p_0 = p_1 = p_3 = p_4 = p_5 = 0$ and $p_2 = 1$).

But multiplying P by itself 1000 times is a pain, even with a computer. Here's a nice trick: If you multiply P by itself, you get P^2. But now, you can multiply P^2 by itself to get P^4. Then multiply P^4 by itself to get P^8, and so on. With just 10 iterations of this technique, we can work out P^{1024}, which could even be done by hand, if you were desperate enough.

Here's what the computer says we get when we calculate P^{1024} to ten digits of accuracy:

$$\begin{pmatrix} 1 & .8 & .6 & .4 & .2 & 0 \\ 0 & 1.549 \times 10^{-95} & 0 & 2.506 \times 10^{-95} & 0 & 0 \\ 0 & 0 & 4.05 \times 10^{-95} & 0 & 2.506^{-95} & 0 \\ 0 & 2.506 \times 10^{-95} & 0 & 4.05 \times 10^{-95} & 0 & \\ 0 & 0 & 2.506 \times 10^{-95} & 0 & 1.549 \times 10^{-95} & 0 \\ 0 & .2 & .4 & .6 & .8 & 1 \end{pmatrix}.$$

For all practical purposes, we have:

$$P^{1024} = \begin{pmatrix} 1 & .8 & .6 & .4 & .2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .2 & .4 & .6 & .8 & 1 \end{pmatrix}.$$

Basically, if you start with \$0, \$10, . . . , \$50, you have probabilities of 0, .2, .4, .6, .8, and 1.0 of winning (reaching a goal of \$50), and the complementary probabilities of losing. As we said above, this can be proven rigorously, but just multiplying the transition matrix by itself repeatedly strongly implies the result.

To solve the second problem, simply do the same calculation with either 6 states (a 6×6 matrix) or 11 states (an 11×11 matrix) and find a high power of the matrix. You'll find it doesn't make any difference which strategy you use.

On the final problem, you can use exactly the same technique, but this time the original matrix P will not have entries of .5, but rather of $9/19$ and $10/19$. But after you know what P is, the process of finding a high power of P is exactly the same. In general, if you have a probability p of winning each bet and a probability of $q = 1 - p$ of losing, here's the transition matrix calculation:

$$\begin{pmatrix} 1 & q & 0 & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 & 0 \\ 0 & p & 0 & q & 0 & 0 \\ 0 & 0 & p & 0 & q & 0 \\ 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & p & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \begin{pmatrix} p_0 + qp_1 \\ qp_2 \\ pp_1 + qp_3 \\ pp_2 + qp_4 \\ pp_3 \\ pp_4 + p_5 \end{pmatrix}.$$

2.4 Board Game Design

Imagine that you are designing a board game for little children and you want to make sure that the game doesn't take too long to finish or the children will get bored. How many moves will it take, on average, for a child to complete the following games:

1. The game has 100 squares, and your marker begins on square 1. You roll a single die which turns up a number 1, 2, 3, 4, 5, or 6, each with probability 1/6. You advance your marker that many squares. When you reach square 100, the game is over. You do not have to hit 100 exactly—for example if you are on square 98 and throw a 3, the game is over.
2. Same as above, but now you must hit square 100 exactly. Is there any limit to how long this game could take?
3. Same as above, but certain of the squares have special markings:
   - Square 7 says "advance to square 55".
   - Square 99 says "go back to square 11".
   - Square 58 says "go back to square 45".
   - Squares 33 and 72 say, "lose a turn".
   - Square 50 says "lose two turns".

In a sense, this is exactly the same problem as "gambler's ruin" above, but not quite so uniform, at least in the third example. For the first problem, there are 100 states, representing the square you are currently upon. So a column vector 100 items long can represent the probability of being in any of the 100 states, and a transition matrix of size 100 × 100 can represent the transition probabilities. Rather than write out the entire 100 × 100 matrix for the game initially specified, let's write out the matrix for a game that's only 8 squares long. You begin on square 1, and win if you reach square 8. let p_1 be the probability of being on square 1, et cetera. Here's the transition matrix:

$$\begin{pmatrix}0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0\\ 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6\\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 2/6\\ 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 3/6\\ 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 4/6\\ 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 5/6\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}p_1\\p_2\\p_3\\p_4\\p_5\\p_6\\p_7\\p_8\end{pmatrix}.$$

If you insist on landing exactly on square 8 to win, the transformation matrix changes to this:

$$\begin{pmatrix}0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0\\ 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6\\ 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6\\ 0 & 0 & 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6\\ 0 & 0 & 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6\\ 0 & 0 & 0 & 0 & 0 & 4/6 & 1/6 & 1/6\\ 0 & 0 & 0 & 0 & 0 & 0 & 5/6 & 1/6\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}p_1\\p_2\\p_3\\p_4\\p_5\\p_6\\p_7\\p_8\end{pmatrix}.$$

In the situation where there are various "go to" squares, there is no chance of landing on them. As the problem is originally stated, it is impossible to land on squares 7, 99, and 58, so there appear to be only 97 possible places to stop. But there are really two states for each of squares 33 and 72—you either just landed there, or you landed there and have waited a turn. There are 3 states for square 50—just landed there, waited one turn, and waited two turns. Thus, the transition matrix contains 101 terms: $100 - 3 + 1 + 1 + 2 = 101$.

2.5 Graph Routes

Imagine a situation where there are 7 possible locations: 1, 2, . . . , 7. There are one-way streets connecting various pairs. For example, if you can get from location 3 to location 7, there will be a street labeled $(3, 7)$. Here is a complete list of the streets:

$(1,2)$, $(1,7)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,6)$, $(4,5)$, $(4,6)$, $(5,6)$, $(5,7)$, and $(6,7)$, $(7,1)$.

If you begin on location 1, and take 16 steps, how many different routes are there that put you on each of the locations? (The discussion that follows will be much easier to follow if you draw a picture. Make a circle of 7 dots, labeled 1 through 7, and for each of the pairs above, draw an arrow from the first dot of the pair to the second.)

This time, we can represent the number of paths to each location by a column vector:

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \end{pmatrix},$$

where p_i is the number of paths to location i. The initial vector for time=0 has $p_1 = 1$ and $p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = 0$. In other words, after zero steps, there is exactly one path to location 1, and no paths to other locations. You can see that if you multiply that initial vector by the matrix once, it will show exactly one path to 2 and one path to 7 and no other paths.

The transition matrix looks like this:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \end{pmatrix}. \tag{4}$$

which will generate the count of the number of paths in $n+1$ steps if the input is the number of paths at step n.

If the matrix on the left of equation (4) is called P, then after 16 steps, the counts will be given by:

$$P^{16} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \end{pmatrix}.$$

From a computer, here is P^{16}:

$$\begin{pmatrix} 481 & 440 & 104 & 280 & 292 & 188 & 288 \\ 288 & 293 & 72 & 192 & 176 & 104 & 188 \\ 188 & 184 & 48 & 125 & 120 & 72 & 104 \\ 188 & 184 & 48 & 125 & 120 & 72 & 104 \\ 292 & 300 & 77 & 203 & 197 & 120 & 176 \\ 384 & 404 & 107 & 281 & 280 & 173 & 264 \\ 728 & 676 & 188 & 480 & 476 & 288 & 481 \end{pmatrix}.$$

Since initially $p_1 = 1$ and $p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = 0$, we have:

$$P^{16}\begin{pmatrix}p_1\\p_2\\p_3\\p_4\\p_5\\p_6\\p_7\end{pmatrix} = \begin{pmatrix}481 & 440 & 104 & 280 & 292 & 188 & 288\\ 288 & 293 & 72 & 192 & 176 & 104 & 188\\ 188 & 184 & 48 & 125 & 120 & 72 & 104\\ 188 & 184 & 48 & 125 & 120 & 72 & 104\\ 292 & 300 & 77 & 203 & 197 & 120 & 176\\ 384 & 404 & 107 & 281 & 280 & 173 & 264\\ 728 & 676 & 188 & 480 & 476 & 288 & 481\end{pmatrix}\begin{pmatrix}1\\0\\0\\0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}481\\288\\188\\188\\292\\384\\728\end{pmatrix},$$

so there are 481 routes to location 1, 288 routes to location 2, 188 routes to location 3, and so on.

This is, of course, very difficult to check. Why don't you check the results for P^3—three step routes. The corresponding equation for only 3 steps is:

$$P^{16}\begin{pmatrix}p_1\\p_2\\p_3\\p_4\\p_5\\p_6\\p_7\end{pmatrix} = \begin{pmatrix}0 & 1 & 1 & 2 & 1 & 0 & 1\\ 1 & 0 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0 & 0 & 0 & 1\\ 3 & 1 & 0 & 0 & 0 & 0 & 0\\ 2 & 4 & 0 & 1 & 1 & 1 & 0\end{pmatrix}\begin{pmatrix}1\\0\\0\\0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\1\\0\\0\\1\\3\\2\end{pmatrix}.$$

So there is no way to get to location 1 in three steps, one way to get to location 2: $(1 \to 7 \to 1 \to 2)$, no ways to get to locations 3 or 4, one way to get to location 5: $(1 \to 2 \to 4 \to 5)$, three ways to get to location 6: $(1 \to 2 \to 3 \to 6)$, $(1 \to 2 \to 4 \to 6)$, and $(1 \to 2 \to 5 \to 6)$, and finally, two ways to get to location 7: $(1 \to 2 \to 5 \to 7)$ and $(1 \to 7 \to 1 \to 7)$.

Obviously, there is nothing special about the set of paths and number of locations in the problem above. For any given setup, you just need to work out the associated transition matrix and you're in business. Even if there are "loops"—paths that lead from a location to itself—there is no problem. A loop will simply generate an entry of 1 in the diagonal of the transition matrix.