Matrix

This section explores the concept behind the Matrix rather than how it works (which is assumed knowledge). It will be more in English than Mathematics.

Linear Transformation: (linear mapping, linear function or linear operator)

A function $f : V \to W$ that has this property: $\boxed{f(\mathbf{v_1}+\mathbf{v_2}) = f(\mathbf{v_1}) + f(\mathbf{v_2})}$ and $\boxed{f(a\mathbf{v}) = af(\mathbf{v})}$ for scalar $a \in K$.

It is easy to prove that $\boxed{f(\mathbf{0}) = \mathbf{0}}$ (by letting $\mathbf{v_2} = -\mathbf{v_1}$ or $a = 0$).

e.g. $f(x) = 5x$ is linear as $f(x+y) = 5(x+y) = 5x+5y,\ f(x)+f(y) = 5x+5y$ and $f(ax) = 5ax,\ af(x) = a \cdot 5x = 5ax$.
$f(x) = 5+x$ is non-linear as $f(x+y) = 5+(x+y),$ but $f(x)+f(y) = (5+x)+(5+y)$.

i.e. "Adding/scaling before linear mapping" and "linear mapping before adding/scaling" yield the same result.

Linear Transformations can be taken as converting a shape through scaling, reflecting, projecting, shearing or rotating. Here is an analogy: Photography is converting a 3-dimensional view into a 2-dimensional photo, which is a mapping from R^3 to R^2. Combining 3-D shapes then taking a photo is the same as taking their photos first then combining them on (a computer with proper software and skills). Shapes are like vectors. So adding the vectors first then mapping them through $f()$, is the same as mapping them first then adding them. (This is why matrix mathematics are used in 3-D animation software, together with some non-linear adjustments for perspective effects.)

$f(\mathbf{v_1}+\mathbf{v_2}) = f(\mathbf{v_1}) + f(\mathbf{v_2})$ is like photograph(shape1+shape2)=photograph(shape1)+photograph(shape2),
and $f(a\mathbf{v}) = af(\mathbf{v})$ is like photograph(enlarge(shape))=enlarge(photograph(shape)).

Matrix as a linear mapping:

An $m \times n$ matrix is a linear mapping $R^n \to R^m$. It takes an n-dimensional vector and turns it into an m-dimensional vector, and you can prove that matrix multiplication follows the rules of linear functions. Therefore multiplied by a matrix is a linear transformation. Conversely, any linear mappings (between finite dimensional vector spaces) can be represented as matrices. The inverse of a matrix undoes its tranformation, as $A^{-1}(A\mathbf{v}) = (A^{-1}A)\mathbf{v} = \mathbf{v}$.

In the following discussion, when a sample unit square or unit cube is referred to, it means that with its sides on the positive axises and $(0,0) \to (1,1)$ or $(0,0,0) \to (1,1,1)$ as its diagonal. (Note: These are not unit vectors).

Scaling: e.g. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$ where $k > 0, k \neq 1$

Replacing a 1 on the identity matrix by a positive value ($\neq 1$) results in a matrix which scales the corresponding axis. For example, if a_{22} of the identity matrix is replaced by 2, the resulting matrix enlarges the y-axis by a factor of 2, so the unit square becomes a tall rectangle. A factor between 0 and 1 would shrink the vector.

Reflecting: e.g. $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Replacing a 1 on the identity matrix by -1 results in a matrix which reflects the shape along the corresponding axis. For example, if a_{22} of the identity matrix is replaced by -1, the resulting matrix flips the shape long the y-axis (i.e. over the x-axis), so the unit square ended up in the fourth quadrant. This can be combined with scaling, so a value of -0.5, for example, will flip and shrink.

Projecting: e.g. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Replacing a 1 on the identity matrix by 0 results in a matrix which projects the shape along the corresponding axis to 0. For example, if a_{22} of I_3 is replaced by 0, the cube will be "flattened" along the y direction onto the xz-plane.

Shearing: e.g. $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

A non-zero value replacing a non-diagonal element of the identity matrix will "shift" a non-axis side of the shape along the direction of another axis. For example, if a_{12} of I_2 becomes 2, the top side of the unit square will shift to the right by an angle of $\tan^{-1}(2)$ (like a gradient of 2 on the y-axis).

Rotating: e.g. $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \tan\theta & \frac{1}{\cos\theta} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\tan\frac{\theta}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sin\theta & 1 \end{bmatrix} \begin{bmatrix} 1 & -\tan\frac{\theta}{2} \\ 0 & 1 \end{bmatrix}$

It can be proven that a rotation is a combination of shearing, reflecting and scaling. There are many ways to do that. For example, the last combination in the above example is (reading from the right) a shearing to the left by $\tan\frac{\theta}{2}$, then shearing up by $\sin\theta$ followed by shearing to the left again by $\tan\frac{\theta}{2}$. The net result is rotating by θ.

Please note that a series of conversions must be achieved by matrix multiplication. You cannot simply replace values on the same matrix. For example, flipping then shearing would be $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 0 & -1 \end{bmatrix}$ $\left(\text{not } \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}.\right)$

Determinant: An $n \times n$ square matrix can be taken as a collection of n n-D vectors, which represent the rows. Its determinant represents the size of the shape "half bound" by these vectors (i.e. a parallelogram in 2-D, a parallelepiped in 3-D, or an equivalent n-dimension shape).

Determinants can be positive or negative. If you move through the vectors from row 1 to row n and it follows the orientation of the axises, then the determinant is positive. Otherwise, it is negative. For example, in a 2×2 matrix, if the first vector rotates a smaller-than-π angle anti-clockwise and coincide with the second vector in direction, then the determinant is positive, as the x-axis and y-axis are in an anti-clockwise orientation. Similarly, in a 3×3 matrix, if the three vectors follow the right-hand rule, same as the orientation of the axises, the determinant will be positive.

Given the above framework, the following properties can be conceptualised:

If one row is zero, or two rows in proportion (parallel), the determinant is zero (as the parallelogram has no size).

Swapping two rows changes the sign of the determinant, as the orientation has been reversed.

Applying two matrices to a shape enlarge its size by the product of their determinants. $\det(AB) = \det(A) \cdot \det(B)$.

Multiplying one row by a constant causes the determinant to increase by the same factor. (e.g. doubling one side of a parallelogram doubles its size.)

Adding a multiple of one row to another does not change the determinant. (e.g. moving one side of a parallelogram parallel to the other side does not change its size.)