This tutorial teaches you about 4×4 matrices. It begins with transforming vectors and composing transformations. It also describes how 4×4 matrices can represent rotations and translations of three-dimensional space.

1 Transforming And Composing

A 4×4 matrix can transform four-dimensional vectors, in a way that's very similar to the two- and three-dimensional cases that we've already studied. If M is 4×4 and $\vec{v}$ is four-dimensional, then

$$M\vec{v} = \begin{bmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} M_{00}v_0 + M_{01}v_1 + M_{02}v_2 + M_{03}v_3 \\ M_{10}v_0 + M_{11}v_1 + M_{12}v_2 + M_{13}v_3 \\ M_{20}v_0 + M_{21}v_1 + M_{22}v_2 + M_{23}v_3 \\ M_{30}v_0 + M_{31}v_1 + M_{32}v_2 + M_{33}v_3 \end{bmatrix}.$$

If N is also a 4×4 matrix, then MN is a 4×4 matrix whose jth column is M times the jth column of N. In other words,

$$(MN)_{ij} = \sum_{k=0}^{3} M_{ik}N_{kj} = M_{i0}N_{0j} + M_{i1}N_{1j} + M_{i2}N_{2j} + M_{i3}N_{3j}.$$

Geometrically, MN is the composite transformation resulting from N followed in time by M. As always, matrix multiplication is associative and not commutative. The 4×4 identity matrix

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

satisfies $I\vec{v} = \vec{v}$ and $IM = M = MI$ for all M and $\vec{v}$.

2 Homogeneous Coordinates

In an earlier tutorial, we homogenized two dimensions into three dimensions, so that we could express translation using matrices. Now we homogenize three dimensions into four, again for the sake of translation.

In homogenization, any three-dimensional point $\vec{p}$ gets a 1 appended, and any 3×3 matrix M gets a row and column of 0s and 1s appended:

$$\vec{p} = \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ 1 \end{bmatrix}, \quad M = \begin{bmatrix} M_{00} & M_{01} & M_{02} & 0 \\ M_{10} & M_{11} & M_{12} & 0 \\ M_{20} & M_{21} & M_{22} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is not difficult to check that the new $M\vec{p}$ is the homogeneous version of the original $M\vec{p}$.

Translation by a three-dimensional vector $\vec{t}$ manifests as the 4×4 matrix

$$T = \begin{bmatrix} 1 & 0 & 0 & t_0 \\ 0 & 1 & 0 & t_1 \\ 0 & 0 & 1 & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inverse operation should intuitively be translation by $-\vec{t}$. So one might guess that

$$T^{-1} = \begin{bmatrix} 1 & 0 & 0 & -t_0 \\ 0 & 1 & 0 & -t_1 \\ 0 & 0 & 1 & -t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In fact, that guess is correct, as you can check by computing TT^{-1} and $T^{-1}T$.

3 Rotations Followed By Translations

The composite transformation resulting from M followed by T is

$$TM = \begin{bmatrix} 1 & 0 & 0 & t_0 \\ 0 & 1 & 0 & t_1 \\ 0 & 0 & 1 & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_{00} & M_{01} & M_{02} & 0 \\ M_{10} & M_{11} & M_{12} & 0 \\ M_{20} & M_{21} & M_{22} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} M_{00} & M_{01} & M_{02} & t_0 \\ M_{10} & M_{11} & M_{12} & t_1 \\ M_{20} & M_{21} & M_{22} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In computer graphics, we usually use this framework to express rotation followed by translation. So M (before it is homogenized) is a 3×3 rotation matrix R. The preceding matrix tutorial gave two useful ways of making such an R.

Sometimes, when we have a rotation-them-translation transformation TR, we need to know the inverse transformation. Intuitively, if TR means rotation by R then translation by $\vec{t}$, then the inverse transformation should be translation by $-\vec{t}$ then rotation by $R^{-1} = R^{\top}$. So it should be

$$
\begin{aligned}
(TR)^{-1} &= \begin{bmatrix} R_{00} & R_{10} & R_{20} & 0 \\ R_{01} & R_{11} & R_{21} & 0 \\ R_{02} & R_{12} & R_{22} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -t_0 \\ 0 & 1 & 0 & -t_1 \\ 0 & 0 & 1 & -t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} R_{00} & R_{10} & R_{20} & -R_{00}t_0 - R_{10}t_1 - R_{20}t_2 \\ R_{01} & R_{11} & R_{21} & -R_{01}t_0 - R_{11}t_1 - R_{21}t_2 \\ R_{02} & R_{12} & R_{22} & -R_{02}t_0 - R_{12}t_1 - R_{22}t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} R_{00} & R_{10} & R_{20} & \left(-R^\top \vec{t}\right)_0 \\ R_{01} & R_{11} & R_{21} & \left(-R^\top \vec{t}\right)_1 \\ R_{02} & R_{12} & R_{22} & \left(-R^\top \vec{t}\right)_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 & \left(-R^\top \vec{t}\right)_0 \\ 0 & 1 & 0 & \left(-R^\top \vec{t}\right)_1 \\ 0 & 0 & 1 & \left(-R^\top \vec{t}\right)_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{00} & R_{10} & R_{20} & 0 \\ R_{01} & R_{11} & R_{21} & 0 \\ R_{02} & R_{12} & R_{22} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\end{aligned}
$$

In other words, the inverse of TR is rotation by R^{-1} followed by translation by $-R^{-1}\vec{t}$. That's not obvious, but that's how it turns out. And it's worth emphasizing that this $(TR)^{-1}$ is *not* generally equal to $T^{-1}R^{-1}$, which would be

$$
T^{-1}R^{-1} = \begin{bmatrix} R_{00} & R_{10} & R_{20} & -t_0 \\ R_{01} & R_{11} & R_{21} & -t_1 \\ R_{02} & R_{12} & R_{22} & -t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$