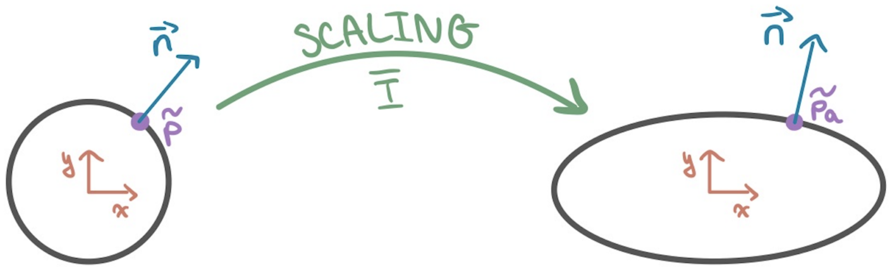
1.11 Normals

Remember: *Normals aren't normal.*

For example, take a point $\tilde{\boldsymbol{p}}$ (with coordinates $\overline{\mathsf{p}}$) on a sphere. The normal at this point is denoted by $\vec{\boldsymbol{n}}$ (with coordinates $\overline{\mathsf{n}}$). Then, we scale the sphere along the x-axis using the matrix $\underline{\overline{\mathsf{T}}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and we obtain an ellipse. We will denote the newly transformed point as $\tilde{\boldsymbol{p}}_a$ (with coordinates $\overline{\mathsf{p}}_a$) and the normal as $\vec{\boldsymbol{n}}_a$ (with coordinates $\overline{\mathsf{n}}_a$), as shown in the diagram below.



To transform the coordinates of the original point to the new point, we can directly apply the transformation matrix $\underline{\overline{\mathsf{T}}}$:

$$\overline{\mathsf{p}}_a = \underline{\overline{\mathsf{T}}}\,\overline{\mathsf{p}}$$

However, we cannot directly apply $\underline{\overline{\mathsf{T}}}$ to transform $\overline{\mathsf{n}}$ into $\overline{\mathsf{n}}_a$. Why?

Intuitively, for a circle with its origin at its center, we can think of the coordinates $\overline{\mathsf{p}}$ as being a multiple of $\overline{\mathsf{n}}$ times its radius r:

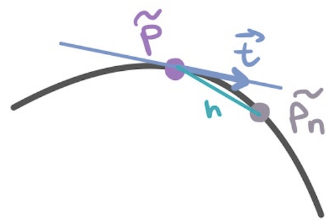
$$\overline{\mathsf{p}} = r\overline{\mathsf{n}} = \begin{pmatrix} rn_1 \\ rn_2 \\ rn_3 \end{pmatrix}$$

If we look at the diagram above, the point's x-coordinate will grow larger when we apply scaling in the x-direction using $\underline{\overline{\mathsf{T}}}$. If we do the same operation to the normal, we can expect the normal's x-coordinate to be scaled in the same way. However, we can see from the diagram that as the sphere becomes an ellipse, the normal's x-coordinate should actually become smaller as the surface stretches out. The normal's coordinates is no longer an integer multiple of the point's coordinates (not colinear anymore). Thus, we can conclude that normals do not transform the same way as points do.

To transform normals, we must first look at how tangents transform:

Take a point $\tilde{\boldsymbol{p}}$ with tangent $\vec{\boldsymbol{t}}$ and a nearby point $\tilde{\boldsymbol{p}}_n$. We can draw a line between them with length h. We can define the tangent as:

$$\vec{t} = \lim_{h \to 0} \frac{\widetilde{\boldsymbol{p}}_h - \widetilde{\boldsymbol{p}}}{h}$$



Since we know that points can be transformed by $\underline{\overline{\mathsf{T}}}$, we can say that the difference between two points can also be transformed by $\underline{\overline{\mathsf{T}}}$:

$$\underline{\overline{\mathsf{T}}}\ \vec{t} = \lim_{h \to 0} \frac{\underline{\overline{\mathsf{T}}}\ \widetilde{\boldsymbol{p}}_h - \underline{\overline{\mathsf{T}}}\ \widetilde{\boldsymbol{p}}}{h}$$

Therefore, we can conclude that tangents *do* transform the same way as points.

Going back to normals, we can define $\vec{\boldsymbol{n}}$ as the vector that satisfies:

$$\vec{\boldsymbol{n}} \cdot \vec{\boldsymbol{t}} = 0$$

$$\overline{\mathsf{n}}^T \overline{\mathsf{t}} = 0$$

After applying T, we want $\overline{\mathsf{n}}_a$ to satisfy:

$$\overline{\mathsf{n}}_a^T \overline{\mathsf{t}}_a = 0$$

...where $\overline{\mathsf{t}}_a = \underline{\overline{\mathsf{T}}}\overline{\mathsf{t}}$. We can write the equation above as:

$$(\overline{\mathsf{n}}_a^T \underline{\overline{\mathsf{T}}})\ \overline{\mathsf{t}} = 0$$

This tells us that:

$$\overline{\mathsf{n}}_a^T \underline{\overline{\mathsf{T}}} = \overline{\mathsf{n}}^T$$

$$\underline{\overline{\mathsf{T}}}^T \overline{\mathsf{n}}_a = \overline{\mathsf{n}}$$

$$\boxed{\overline{\mathsf{n}}_a = \underline{\overline{\mathsf{T}}}^{-T}\ \overline{\mathsf{n}}}$$

More specifically, since normals are vectors, the inverse transpose operation is only done on the upper left 3x3 matrix of $\underline{\overline{\mathsf{T}}}$:

$$T = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & 1 \end{array}\right)$$

$$\boxed{\overline{\mathsf{n}}_a = \left(\begin{array}{c|c} A^{-T} & 0 \\ \hline 0 & 1 \end{array}\right) \overline{\mathsf{n}}}$$

If $\underline{\overline{\mathsf{T}}}$ were the modelViewMatrix, then A^{-T} is the normalMatrix (supplied by Three.js).