Rigid Geometric Transformations

Carlo Tomasi

This note is a quick refresher of the geometry of rigid transformations in three-dimensional space, expressed in Cartesian coordinates.

1 Cartesian Coordinates

Let us assume the notions of the *distance* between two points and the *angle* between lines to be known from geometry. The *law of cosines* is also stated without proof[1]: if a, b, c are the sides of a triangle and the angle between a and b is θ, then

$$c^2 = a^2 + b^2 - 2ab\cos\theta \; .$$

The special case for $\theta = \pi/2$ radians is known as Pythagoras' theorem.

The definitions that follow focus on three-dimensional space. Two-dimensional geometry can be derived as a special case when the third coordinate of every point is set to zero.

A *Cartesian reference system* for three-dimensional space is a point in space called the *origin* and three mutually perpendicular, directed lines though the origin called the *axes*. The order in which the axes are listed is fixed, and is part of the definition of the reference system. The plane that contains the second and third axis is the *first reference plane*. The plane that contains the third and first axis is the *second reference plane*. The plane that contains the first and second axis is the *third reference plane*.

It is customary to mark the axis directions by specifying a point on each axis and at unit distance from the origin. These points are called the *unit points* of the system, and the *positive direction* of an axis is from the origin towards the axis' unit point. A Cartesian reference system is *right-handed* if the smallest rotation that brings the first unit point to the second is counterclockwise when viewed from the third unit point. The system is *left-handed* otherwise.

The *Cartesian coordinates* of a point in three-dimensional space are the signed distances of the point from the first, second, and third reference plane, in this order, and collected into a vector. The sign for coordinate i is positive if the point is in the half-space (delimited by the i-th reference plane) that contains the positive half of the i-th reference axis. It follows that the Cartesian coordinates of the origin are $\mathbf{o} = (0,0,0)$, those of the unit points are the vectors $\mathbf{e}_1 = (1,0,0)$, $\mathbf{e}_2 = (0,1,0)$, and $\mathbf{e}_3 = (0,0,1)$, and the vector $\mathbf{p} = (x,y,z)$ of coordinates of an arbitrary point in space can also be written as follows:

$$\mathbf{p} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3 \; .$$

The point $\mathbf{p}$ can be reached from the origin by the following polygonal path:

$$\mathbf{o}\,,\; x\mathbf{e}_1\,,\; x\mathbf{e}_1 + y\mathbf{e}_2\,,\; \mathbf{p} \; .$$

[1] A proof based on trigonometry is straightforward but tedious, and a useful exercise.

Each segment of the path is followed by a right-angle turn, so Pythagoras' theorem can be applied twice to yield the distance of $\mathbf{p}$ from the origin:

$$d(\mathbf{o}, \mathbf{p}) = \sqrt{x^2 + y^2 + z^2} \ .$$

From the definition of norm of a vector we see that

$$d(\mathbf{o}, \mathbf{p}) = \|\mathbf{p}\| \ .$$

So the norm of the vector of coordinates of a point is the distance of the point from the origin. A vector is often drawn as an arrow pointing from the origin to the point whose coordinates are the components of the vector. Then, the result above shows that the *length* of that arrow is the norm of the vector. Because of this, the words "length" and "norm" are often used interchangeably.

2 Orthogonality

The law of cosines yields a geometric interpretation of the inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$:

Theorem 2.1.

$$\mathbf{a}^T\mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \, \cos\theta \tag{1}$$

where θ is the acute angle between the two arrows that represent $\mathbf{a}$ and $\mathbf{b}$ geometrically.

So the inner product of two vectors is the product of the lengths of the two arrows that represent them and of the cosine of the angle between them. See the appendix for a proof.

Setting $\theta = \pi/2$ in the result above yields another important corollary:

Corollary 2.2. *The arrows that represent two vectors $\mathbf{a}$ and $\mathbf{b}$ are mutually perpendicular if an only if the two vectors are orthogonal:*

$$\mathbf{a}^T\mathbf{b} = 0 \ .$$

Because of this result, the words "perpendicular" and "orthogonal" are often used interchangeably.

3 Orthogonal Projection

Given two vectors $\mathbf{a}$ and $\mathbf{b}$, the *orthogonal projection* of $\mathbf{a}$ onto $\mathbf{b}$ is the vector $\mathbf{p}$ that represents the point p on the line through $\mathbf{b}$ that is nearest to the endpoint of $\mathbf{a}$. See Figure 1.

Theorem 3.1. *The orthogonal projection of $\mathbf{a}$ onto $\mathbf{b}$ is the vector*

$$\mathbf{p} = P_{\mathbf{b}}\mathbf{a}$$

where $P_{\mathbf{b}}$ is the following square, symmetric matrix:

$$P_{\mathbf{b}} = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}} \ .$$

The signed magnitude *of the orthogonal projection is*

$$p = \frac{\mathbf{b}^T\mathbf{a}}{\|\mathbf{b}\|} = \|\mathbf{p}\| \ \mathrm{sign}(\mathbf{b}^T\mathbf{a}) \ .$$

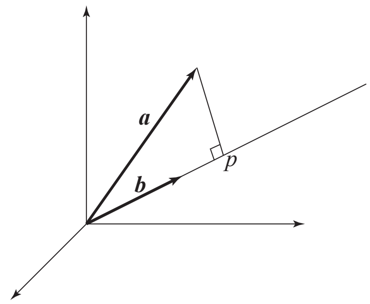

Figure 1: The vector from the origin to point p is the orthogonal projection of $\mathbf{a}$ onto $\mathbf{b}$. The line from the endpoint of $\mathbf{a}$ to p is orthogonal to $\mathbf{b}$.

From the definition of orthogonal projection we also see the following fact.

Corollary 3.2. *The coordinates of a point in space are the signed magnitudes of the orthogonal projections of the vector of coordinates of the point onto the three unit vectors that define the coordinate axes.*

This result is trivial in the basic Cartesian reference frame with unit points $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$. If $\mathbf{p} = (x, y, z)$, then obviously

$$\mathbf{e}_1\mathbf{p} = x \quad , \quad \mathbf{e}_2\mathbf{p} = y \quad , \quad \mathbf{e}_3\mathbf{p} = z \ .$$

The result becomes less trivial in Cartesian reference systems where the axes have different orientations, as we will see soon.

4 Cross Product

The *cross product* of two 3-dimensional vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ is the 3-dimensional vector

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2 \ , \ a_3b_1 - a_1b_3 \ , \ a_1b_2 - a_2b_1) \ .$$

The following geometric interpretation is proven in the Appendix:

Theorem 4.1. *The cross product of two three-dimensional vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is a vector* $\mathbf{c}$ *orthogonal to both* $\mathbf{a}$ *and* $\mathbf{b}$*, oriented so that the triple* $\mathbf{a}$*,* $\mathbf{b}$*,* $\mathbf{c}$ *is right-handed, and with magnitude*

$$\|\mathbf{c}\| \ = \ \|\mathbf{a} \times \mathbf{b}\| \ = \ \|\mathbf{a}\| \, \|\mathbf{b}\| \, \sin\theta$$

where θ *is the acute angle between* $\mathbf{a}$ *and* $\mathbf{b}$*.*

From its expression, we see that the magnitude of $\mathbf{a} \times \mathbf{b}$ is the area of a rectangle with sides $\mathbf{a}$ and $\mathbf{b}$.

Suppose that we need to compute cross products of the form $\mathbf{a} \times \mathbf{p}$ where $\mathbf{a}$ is a fixed vector but $\mathbf{p}$ changes. It is then convenient to write the cross product as the product of a matrix $\mathbf{a}_\times$ that depends on $\mathbf{a}$ and of $\mathbf{p}$. Spelling out the definition of the cross product yields the following anti-symmetric matrix:

$$\mathbf{a}_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \ .$$

The *triple product* of three-dimensional vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ is defined as follows:

$$\mathbf{a}^T(\mathbf{b}\times\mathbf{c}) = a_1(b_2c_3 - b_3c_2) + a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \; .$$

It is immediate to verify that

$$\mathbf{a}^T(\mathbf{b}\times\mathbf{c}) = \mathbf{b}^T(\mathbf{c}\times\mathbf{a}) = \mathbf{c}^T(\mathbf{a}\times\mathbf{b}) = -\mathbf{a}^T(\mathbf{c}\times\mathbf{b}) = -\mathbf{c}^T(\mathbf{b}\times\mathbf{a}) = -\mathbf{b}^T(\mathbf{a}\times\mathbf{c}) \; .$$

Again, from its expression, we see that the triple product of vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ is, up to a sign, the volume of a parallelepiped with edges $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$: the cross product $\mathbf{p} = \mathbf{b}\times\mathbf{c}$ is a vector orthogonal to the plane of $\mathbf{b}$ and $\mathbf{c}$, and with magnitude equal to the base area of the parallelepiped. The inner product of $\mathbf{p}$ and $\mathbf{a}$ is the magnitude of $\mathbf{p}$ times that of $\mathbf{a}$ times the cosine of the angle between them, that is, the base area of the parallelepiped times its height (or the negative of its height). This gives the volume of the solid, up to a sign. The sign is positive if the three vectors form a right-handed triple. See Figure 2.

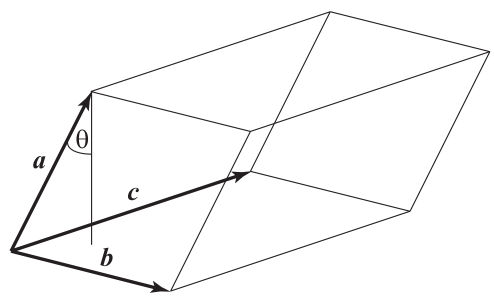

Figure 2: Up to a sign, the triple product of the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ is the volume of the parallelepiped with edges $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$.

5 Rotation

A *rotation* is a transformation between two Cartesian references systems C and C' of equal origin and handedness. Let $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ be the unit points of C, and $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ the unit points of C'. Then a point with coordinates $\mathbf{p} = (x, y, z)$ in C can be reached from the common origin $\mathbf{o}$ to the two systems by a polygonal path with the following four vertices:

$$\mathbf{o} \quad , \quad \mathbf{a} = x'\mathbf{i} \quad , \quad \mathbf{b} = x'\mathbf{i} + y'\mathbf{j} \quad , \quad \mathbf{p} = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} \; .$$

The steps of this path are along the axes of C'. The numbers x', y', z' are the magnitudes of the steps, and also the coordinates of the point in C'. These step sizes are the signed magnitudes of the orthogonal projections of the point onto $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, and from Theorem 3.1 we see that

$$x' = \mathbf{i}^T\mathbf{p} \quad , \quad y' = \mathbf{j}^T\mathbf{p} \quad , \quad z' = \mathbf{k}^T\mathbf{p}$$

because the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ have unit norm. These three equations can be packaged into a single matrix equation that expresses the vector $\mathbf{p}' = (x', y', z')$ as a function of $\mathbf{p}$:

$$\mathbf{p}' \; = \; R\,\mathbf{p} \quad \text{where} \quad R = \begin{bmatrix} \mathbf{i}^T \\ \mathbf{j}^T \\ \mathbf{k}^T \end{bmatrix}$$

where the 3×3 matrix R is called a *rotation* matrix.

A rotation is a reversible transformation, and therefore the matrix R must have an *inverse*, another matrix that transforms back from C' to C. The proof of the following fact is given in the Appendix.

Theorem 5.1. *The inverse of a rotation matrix is its transpose:*

$$R^T R = R R^T = I \; .$$

Note that R^T, being the inverse of R, is also a transformation between two Cartesian systems with the same origin and handedness, so R^T is a rotation matrix as well, and its rows must be mutually orthogonal unit vectors. Since the rows of R^T are the columns of R, we conclude that both the rows and columns of a rotation matrix are unit norm and orthogonal. This makes intuitive sense: just as the rows of R are the unit vectors of C' expressed in C, so its columns (the rows of the inverse transformation R^T) are the unit vectors of C expressed in C'.

The equations in Theorem 5.1 characterize combinations of rotations and possible inversions. An *inversion* (also known as a *mirror flip*) is a transformation that changes the direction of some of the axes. This is represented by a matrix of the form

$$S = \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix}$$

where s_1, s_2 s_3 are equal to either 1 or -1, and there is either one or three negative elements. It is easy to see that

$$S^T S = S S^T = I \; .$$

If there were zero or two negative elements, then S would be a rotation matrix, because the flip of two axes can be achieved by a rotation. For instance, flipping the x and y axes can be achieved by a 180-degree rotation around the z axis. No rotation can achieve the flip of an odd number of axes.

The *determinant* of a 3×3 matrix is the triple product of its rows. Direct manipulation shows that this is the same as the triple product of its columns. It is immediate to see that the determinant of a rotation matrix is 1:

$$\det(R) = \mathbf{i}^T(\mathbf{j} \times \mathbf{k}) = \mathbf{i}^T \mathbf{i} = 1$$

because

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} \quad , \quad \mathbf{j} \times \mathbf{k} = \mathbf{i} \quad , \quad \mathbf{k} \times \mathbf{i} = \mathbf{j} \; .$$

These equalities can be verified by the geometric interpretation of the cross product: each of the three vectors **i**, **j**, **k** is orthogonal to the other two, and its magnitude is equal to 1. The order of the vectors in the equalities above preserves handedness.

It is even easier to see that the determinant of an inversion matrix S is equal to -1. Thus, the following conclusion can be drawn.

A matrix R is a rotation if and only if $R^T R = R R^T = I \quad \text{and} \quad \det(R) = 1$.
A diagonal matrix S is an inversion if and only if $S^T S = S S^T = I \quad \text{and} \quad \det(S) = -1$.

Note that in particular the identity matrix I is a rotation, and $-I$ is an inversion.

Geometric Interpretation of Orthogonality The orthogonality result

$$R^{-1} = R^T$$

is very simple, and yet was derived in the Appendix through a comparatively lengthy sequence of algebraic steps. This Section reviews orthogonality of rotation matrices from a geometric point of view, and derives the result above by simpler means. The rows of the rotation matrix

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{i}^T \\ \mathbf{j}^T \\ \mathbf{k}^T \end{bmatrix} = \begin{bmatrix} i_1 & i_2 & i_3 \\ j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{bmatrix}$$

are the unit vectors of the rotated ("new") reference system, expressed in the original ("old") reference system. This means that its entry r_{ij} is the signed magnitude of the orthogonal projection of the i-th new unit vector onto the j-th old unit vector. For instance,

$$r_{12} = \mathbf{i}^T \mathbf{e}_2 \quad \text{and} \quad r_{31} = \mathbf{k}^T \mathbf{e}_1 \ .$$

However, the signed magnitude of the orthogonal projection of a unit vector onto another unit vector is simply the cosine of the angle between them:

$$r_{ij} = \cos \alpha_{ij}$$

where α_{ij} is the angle between the i-th axis in the new system and the j-th axis in the old.

Thus, the entries of a rotation matrix are *direction cosines*: they are all cosines of well-defined angles. This result also tells us that signed orthogonal projection magnitude is symmetric for unit vectors: For instance, the signed magnitude of the orthogonal projection of $\mathbf{i}$ onto $\mathbf{e}_2$ is the same as the signed magnitude of the orthogonal projection of $\mathbf{e}_2$ onto $\mathbf{i}$.

This symmetry is the deep reason for orthogonality: when we want to go from the "new" system $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ back to the "old" system $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ through the inverse matrix R^{-1}, we seek to express the latter unit vectors in the system of the former, that is, we seek the signed magnitudes of the orthogonal projections of each "old" unit vector onto each of the "new" unit vectors. Because of symmetry, these orthogonal projections are already available in the matrix R, just in a different arrangement: what we want in the rows of R^{-1} can be found in the columns of R. *Voilà*:

$$R^{-1} = R^T \ .$$

6 Coordinate Transformation

Two right-handed, Cartesian systems of reference C and C' can differ by a translation of the origin from $\mathbf{o}$ to $\mathbf{t}$ and a rotation of the axes from unit points $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ to unit points $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Suppose that the origin of frame C is first translated to point $\mathbf{t}$ (as expressed in C) and then the resulting frame is rotated by R (see Figure 3). Given a point with coordinates $\mathbf{p} = (x, y, z)$ in C, the coordinates $\mathbf{p}' = (x', y', z')$ of the same point in C' are then

$$\mathbf{p}' = R(\mathbf{p} - \mathbf{t}) \tag{2}$$

The translation is applied first, to yield the new coordinates $\mathbf{p} - \mathbf{t}$ in an intermediate frame C". This does not change the directions of the coordinate axes, so the rotation in C and in C" is expressed by the same rotation R, which is applied thereafter.

The inverse transformation applies the inverse operations in reverse order:

$$\mathbf{p} = R^T \mathbf{p}' + \mathbf{t} \ . \tag{3}$$

This can also be verified algebraically from equation (2): multiplying both sides by R^T from the left yields

$$R^T \mathbf{p}' = R^T R(\mathbf{p} - \mathbf{t}) = \mathbf{p} - \mathbf{t}$$

and adding $\mathbf{t}$ to both sides yields equation (3).

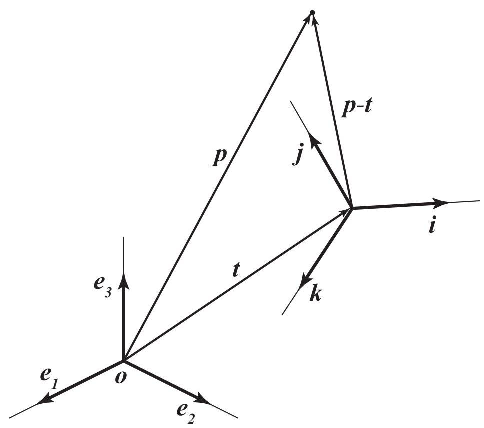

Figure 3: Transformation between two reference systems.

The transformations (2) and (3) are said to be *rigid*, in that they preserve distances. They are also sometimes referred to as *special Euclidean*, where the attribute "special" refers to the fact that mirror flips are not included—*i.e.*, the determinant of R is 1, rather than 1 in magnitude.

A Proofs

Theorem 2.1

$$\mathbf{a}^T\mathbf{b} = \|\mathbf{a}\|\,\|\mathbf{b}\|\,\cos\theta \tag{4}$$

where θ is the acute angle between the two arrows that represent $\mathbf{a}$ and $\mathbf{b}$ geometrically.

Proof. Consider a triangle with sides

$$a = \|\mathbf{a}\| \quad , \quad b = \|\mathbf{b}\| \quad , \quad c = \|\mathbf{b}-\mathbf{a}\|$$

and with an angle θ between a and b. Then the law of cosines yields

$$\|\mathbf{b}-\mathbf{a}\|^2 \;=\; \|\mathbf{a}\|^2 \,+\, \|\mathbf{b}\|^2 \,-\, 2\|\mathbf{a}\|\,\|\mathbf{b}\|\,\cos\theta\;.$$

From the definition of norm we then obtain

$$\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\mathbf{a}^T\mathbf{b} \;=\; \|\mathbf{a}\|^2 \,+\, \|\mathbf{b}\|^2 \,-\, 2\|\mathbf{a}\|\,\|\mathbf{b}\|\,\cos\theta\;.$$

Canceling equal terms and dividing by -2 yields the desired result.

Theorem 3.1

The orthogonal projection of $\mathbf{a}$ onto $\mathbf{b}$ is the vector

$$\mathbf{p} = P_{\mathbf{b}}\mathbf{a}$$

where $P_{\mathbf{b}}$ is the following square, symmetric matrix:

$$P_{\mathbf{b}} = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}}\;.$$

The signed magnitude *of the orthogonal projection is*

$$p = \frac{\mathbf{b}^T\mathbf{a}}{\|\mathbf{b}\|} = \|\mathbf{p}\|\,\mathrm{sign}(\mathbf{b}^T\mathbf{a})\;.$$

Proof. To prove this, observe that since by definition point p is on the line through $\mathbf{b}$, the orthogonal projection vector $\mathbf{p}$ has the form $\mathbf{p} = x\mathbf{b}$, where x is some real number. From elementary geometry, the line between p and the endpoint of $\mathbf{a}$ is shortest when it is perpendicular to $\mathbf{b}$:

$$\mathbf{b}^T(\mathbf{a} - x\mathbf{b}) = 0$$

which yields

$$x = \frac{\mathbf{b}^T\mathbf{a}}{\mathbf{b}^T\mathbf{b}}$$

so that

$$\mathbf{p} = x\mathbf{b} = \mathbf{b}\,x = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}}\,\mathbf{a}$$

as advertised. The magnitude of $\mathbf{p}$ can be computed as follows. First, observe that

$$P_{\mathbf{b}}^2 = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}} \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}} = \frac{\mathbf{b}\mathbf{b}^T\mathbf{b}\mathbf{b}^T}{(\mathbf{b}^T\mathbf{b})^2} = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}} = P_{\mathbf{b}}$$

so that the orthogonal-projection matrix[2] $P_{\mathbf{b}}$ is *idempotent*:

$$P_{\mathbf{b}}^2 = P_{\mathbf{b}} \ .$$

This means that applying the matrix once or multiple times has the same effect. Then,

$$\|\mathbf{p}\|^2 = \mathbf{p}^T\mathbf{p} = \mathbf{a}^T P_{\mathbf{b}}^T P_{\mathbf{b}}\mathbf{a} = \mathbf{a}^T P_{\mathbf{b}} P_{\mathbf{b}}\mathbf{a} = \mathbf{a}^T P_{\mathbf{b}}\mathbf{a} = \mathbf{a}^T \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}}\mathbf{a} = \frac{(\mathbf{b}^T\mathbf{a})^2}{\mathbf{b}^T\mathbf{b}}$$

which, once the sign of $\mathbf{b}^t\mathbf{a}$ is taken into account, yields the promised expression for the signed magnitude of $\mathbf{p}$.

Theorem 4.1

The cross product of two three-dimensional vectors $\mathbf{a}$ *and* $\mathbf{b}$ *is a vector* $\mathbf{c}$ *orthogonal to both* $\mathbf{a}$ *and* $\mathbf{b}$*, oriented so that the triple* $\mathbf{a}$*,* $\mathbf{b}$*,* $\mathbf{c}$ *is right-handed, and with magnitude*

$$\|\mathbf{c}\| \ = \ \|\mathbf{a}\times\mathbf{b}\| \ = \ \|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin\theta$$

where θ *is the acute angle between* $\mathbf{a}$ *and* $\mathbf{b}$*.*

Proof. That the cross product $\mathbf{c}$ of $\mathbf{a}$ and $\mathbf{b}$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$ can be checked directly:

$$\begin{aligned}
\mathbf{c}^T\mathbf{a} &= (a_2b_3 - a_3b_2)a_1 + (a_3b_1 - a_1b_3)a_2 + (a_1b_2 - a_2b_1)a_3 = 0 \\
\mathbf{c}^T\mathbf{b} &= (a_2b_3 - a_3b_2)b_1 + (a_3b_1 - a_1b_3)b_2 + (a_1b_2 - a_2b_1)b_3 = 0
\end{aligned}$$

(verify that all terms do indeed cancel). We also have

$$(\mathbf{a}^T\mathbf{b})^2 \ + \ \|\mathbf{a}\times\mathbf{b}\|^2 \ = \ \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2$$

as can be shown by straightforward manipulation:

$$\begin{aligned}
(\mathbf{a}^T\mathbf{b})^2 &= (a_1b_1 + a_2b_2 + a_3b_3)\,(a_1b_1 + a_2b_2 + a_3b_3) \\
&= \ \ a_1^2b_1^2 + a_1b_1a_2b_2 + a_1b_1a_3b_3 \\
&\quad + a_2^2b_2^2 + a_1b_1a_2b_2 + a_2b_2a_3b_3 \\
&\quad + a_3^2b_3^2 + a_1b_1a_3b_3 + a_2b_2a_3b_3 \\
&= a_1^2b_1^2 + a_2^2b_2^2 + a_3^2b_3^2 + 2a_1b_1a_2b_2 + 2a_2b_2a_3b_3 + 2a_1b_1a_3b_3
\end{aligned}$$

and

$$\begin{aligned}
\|\mathbf{a}\times\mathbf{b}\|^2 &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\
&= \ \ a_2^2b_3^2 + a_3^2b_2^2 - 2a_2b_2a_3b_3 \\
&\quad + a_1^2b_3^2 + a_3^2b_1^2 - 2a_1b_1a_3b_3 \\
&\quad + a_1^2b_2^2 + a_2^2b_1^2 - 2a_1b_1a_2b_2 \\
&= \ \ a_1^2b_2^2 + a_2^2b_1^2 + a_2^2b_3^2 + a_3^2b_2^2 + a_1^2b_3^2 + a_3^2b_1^2 \\
&\quad - 2a_1b_1a_2b_2 - 2a_2b_3a_2b_2 - 2a_1b_1a_3b_3
\end{aligned}$$

[2]The matrix that describes orthogonal projection is not an orthogonal matrix. It could not possibly be, since it is rank-deficient.

so that

$$(\mathbf{a}^T\mathbf{b})^2 \; + \; \|\mathbf{a}\times\mathbf{b}\|^2 \;\; = \;\; a_1^2b_1^2 + a_1^2b_2^2 + a_1^2b_3^2 + a_2^2b_1^2 + a_2^2b_2^2 + a_2^2b_3^2 + a_3^2b_1^2 + a_3^2b_2^2 + a_3^2b_3^2$$

but also

$$\|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \;\; = \;\; a_1^2b_1^2 + a_1^2b_2^2 + a_1^2b_3^2 + a_2^2b_1^2 + a_2^2b_2^2 + a_2^2b_3^2 + a_3^2b_1^2 + a_3^2b_2^2 + a_3^2b_3^2$$

so that

$$(\mathbf{a}^T\mathbf{b})^2 \; + \; \|\mathbf{a}\times\mathbf{b}\|^2 \;\; = \;\; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \tag{5}$$

as desired. The result on the magnitude is a consequence of equation (5). From this equation we obtain

$$\|\mathbf{a}\times\mathbf{b}\|^2 \;\; = \;\; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \; - \; (\mathbf{a}^T\mathbf{b})^2 \;\; = \;\; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \; - \; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2\,\cos^2\theta \;\; = \;\; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2\,\sin^2\theta$$

or

$$\|\mathbf{a}\times\mathbf{b}\| \;\; = \;\; \pm\|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin\theta \; .$$

Since the angle θ is acute (from equation (4)), all quantities in the last equation are nonnegative, so that the $-$ sign yields an impossible equation. This yields the promised result.

Theorem 5.1

The inverse of a rotation matrix is its transpose:

$$R^T R = RR^T = I \; .$$

Proof. When we rotated $\mathbf{p}$ through R we obtained a vector $\mathbf{p}'$ of coordinates in C'. We then look for a new matrix R^{-1} that applied to $\mathbf{p}'$ gives back the original vector $\mathbf{p}$:

$$\mathbf{p}' = R\mathbf{p} \quad \rightarrow \quad \mathbf{p} = R^{-1}\mathbf{p}'$$

that is,

$$\mathbf{p} = R^{-1}R\mathbf{p} \; .$$

Since this is to hold for *any* vector $\mathbf{p}$, we need to find R^{-1} such that

$$R^{-1}R = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \; .$$

The matrix I is called the *identity* matrix, and the matrix R^{-1} is the *left inverse* of R. However, even a *right inverse*, that is, a matrix Q such that

$$RQ = I$$

will do. This is because for *any* square matrix A, if the matrix B is the right inverse of A, that is, if $AB = I$, then B is also the left inverse:

$$BA = I \; .$$

The proof is a single line: suppose that the left inverse is a matrix C, so that $CA = I$. Then

$$C = CI = C(AB) = (CA)B = IB = B \; ,$$

which forces us to conclude that B and C are the same matrix. So we can drop "left" or "right" and merely say *inverse*.

The inverse R^{-1} of the rotation matrix R is more easily found by looking for a right inverse. The three vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ that make up the rows of R have unit norm,

$$\mathbf{i}^T\mathbf{i} = \mathbf{j}^T\mathbf{j} = \mathbf{k}^T\mathbf{k} = 1 \; ,$$

and are mutually orthogonal:

$$\mathbf{i}^T\mathbf{j} = \mathbf{j}^T\mathbf{k} = \mathbf{k}^T\mathbf{i} = 0 \; .$$

Because of this,

$$RR^T = \begin{bmatrix} \mathbf{i}^T \\ \mathbf{j}^T \\ \mathbf{k}^T \end{bmatrix} \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{i}^T\mathbf{i} & \mathbf{i}^T\mathbf{j} & \mathbf{i}^T\mathbf{k} \\ \mathbf{j}^T\mathbf{i} & \mathbf{j}^T\mathbf{j} & \mathbf{j}^T\mathbf{k} \\ \mathbf{k}^T\mathbf{i} & \mathbf{k}^T\mathbf{j} & \mathbf{k}^T\mathbf{k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

as promised.