Rigid Geometric Transformations and the Pinhole Camera Model

Carlo Tomasi

March 16, 2022

This note starts with a quick refresher of the geometry of rigid transformations in three-dimensional space, expressed in Cartesian coordinates. It then introduces a simple model for a camera, which relates coordinates of points in the world to coordinates of the perspective projections of these points on the image plane. This relation will later let us develop a method for reconstructing the three-dimensional geometry of a scene from two images of it.

1 Rigid Geometric Transformations

A rigid geometric transformation is a change between orthogonal Cartesian reference systems. We will attach one such system to each camera, or to a single camera as it moves around, and we need to be able to transform coordinates of world points between the reference systems. Since the reference systems are orthogonal, the next few subsections recall the main concepts of coordinates, orthogonality, orthogonal projection[1], and cross and triple products of vectors. This will hopefully just be a refresher for you. If you are rusty on the concepts please also look at the proofs in the Appendix. While appendices are optional reading, understanding some of the proofs (they are all easy) may make it easier for you to remember concepts.

1.1 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[2]: 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

[1] Please note that we talk about orthogonal projection here, and about perspective projection when we model cameras. These are two different projections.

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

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 and units of measure by specifying a point on each axis and at unit distance from the origin in the direction chosen as positive. These points are called the *unit points* of the system. 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 are often 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 unit point of the i-th reference axis. It follows that the Cartesian coordinates of the origin are $\mathbf{t} = (0,0,0)^T$, those of the unit points are the vectors $\mathbf{e}_x = (1,0,0)^T$, $\mathbf{e}_y = (0,1,0)^T$, and $\mathbf{e}_z = (0,0,1)^T$, and the vector $\mathbf{p} = (x,y,z)^T$ of coordinates of an arbitrary point in space can also be written as follows:

$$\mathbf{p} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z \ .$$

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

$$\mathbf{t} \, , \; x\mathbf{e}_x \, , \; x\mathbf{e}_x + y\mathbf{e}_y \, , \; \mathbf{p} \ .$$

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:

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

From the definition of norm of a vector we see that

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

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.

1.2 Orthogonality

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

Theorem 1.1.

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

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 Appendix 2 for a proof.

Setting $\theta = \pi/2$ in the result above, that is, making $\mathbf{a}$ and $\mathbf{b}$ perpendicular, yields another important corollary:

Corollary 1.2. *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. Note that “perpendicular” is a geometric property of two vectors, while “orthogonal” is an algebraic relationship involving their Cartesian coordinates.

1.3 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 on the line through $\mathbf{b}$ that is nearest to the endpoint of $\mathbf{a}$. See Figure 1.

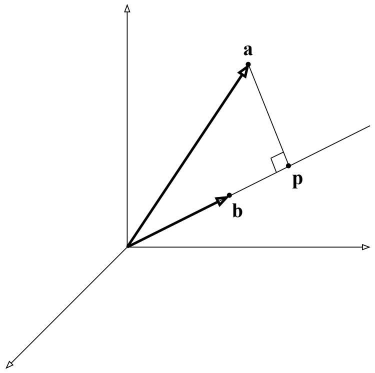


Figure 1: The vector from the origin to point $\mathbf{p}$ is the orthogonal projection of $\mathbf{a}$ onto $\mathbf{b}$. The line from the endpoint of $\mathbf{a}$ to $\mathbf{p}$ is orthogonal to $\mathbf{b}$.

Theorem 1.3. *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, rank-1 matrix:*

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

The signed magnitude *of the orthogonal projection is*

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

(A proof is given in the Appendix.) From the definition of orthogonal projection it follows that the line between a point $\mathbf{a}$ and its projection $\mathbf{p}$ onto the first reference axis is orthogonal to the first reference axis and, therefore, parallel to the first reference plane. Therefore, $\mathbf{a}$ and $\mathbf{p}$ are at the same distance from the first reference plane and on the same side of it. Thus, the first Cartesian coordinate of $\mathbf{a}$, which is defined as the signed distance of $\mathbf{a}$ from the first reference plane, is also the signed magnitude of $\mathbf{p}$. This reasoning can be applied to any of the three coordinates of $\mathbf{a}$, and we can conclude as follows.

Corollary 1.4. *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 $\mathbf{e}_x = (1,0,0)^T$, $\mathbf{e}_y = (0,1,0)^T$, $\mathbf{e}_z = (0,0,1)^T$. If $\mathbf{p} = (x,y,z)^T$, then obviously

$$\mathbf{e}_x^T\mathbf{p} = x \quad , \quad \mathbf{e}_y^T\mathbf{p} = y \quad , \quad \mathbf{e}_z^T\mathbf{p} = z \; .$$

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

1.4 The Cross Product

The *cross product* of two 3-dimensional vectors $\mathbf{a} = (a_x, a_y, a_z)^T$ and $\mathbf{b} = (b_x, b_y, b_z)^T$ is the 3-dimensional vector

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y \; , \; a_z b_x - a_x b_z \; , \; a_x b_y - a_y b_x)^T \; .$$

The following geometric interpretation is proven in the Appendix:

Theorem 1.5. *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 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 parallelogram with sides $\mathbf{a}$ and $\mathbf{b}$.

It is immediate to verify that the cross product of two vectors is a linear transformation of either vector (but not both):

$$\mathbf{a} \times (\mathbf{b}_1 + \mathbf{b}_2) = \mathbf{a} \times \mathbf{b}_1 + \mathbf{a} \times \mathbf{b}_2 \quad \text{and similarly} \quad (\mathbf{a}_1 + \mathbf{a}_2) \times \mathbf{b} = \mathbf{a}_1 \times \mathbf{b} + \mathbf{a}_2 \times \mathbf{b} \; .$$

So there must be a 3×3 matrix $[\mathbf{a}]_\times$ such that

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b} \; .$$

This matrix is convenient for repeatedly computing cross products of the form $\mathbf{a} \times \mathbf{p}$ where $\mathbf{a}$ is a fixed vector but $\mathbf{p}$ changes. Spelling out the definition of the cross product yields the following matrix:

$$[\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \; .$$

This matrix is skew-symmetric:

$$[\mathbf{a}]_\times^T = -[\mathbf{a}]_\times \; .$$

Of course, similar considerations hold for $\mathbf{b}$: Since

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} \; ,$$

we have

$$\mathbf{a} \times \mathbf{b} = -[\mathbf{b}]_\times \mathbf{a} = [\mathbf{b}]_\times^T \mathbf{a} \; .$$

1.5 The Triple Product

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

$$\mathbf{a}^T(\mathbf{b}\times\mathbf{c}) = a_x(b_y c_z - b_z c_y) - a_y(b_x c_z - b_z c_x) + a_z(b_x c_y - b_y c_x) \; .$$

It is immediate to verify that

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

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 and only if the three vectors form a right-handed triple. See Figure 2.

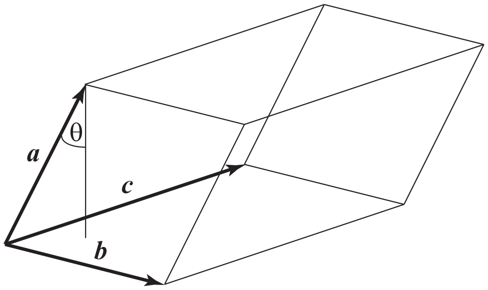


Figure 2: Up to a sign, the triple product of the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ is the volume of the parallelepiped with edges $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$.

1.6 Multiple Reference Systems

When two or more reference systems are involved, notation must be introduced to avoid possible ambiguities as to which reference system coordinates refer to. For instance, we may want to express the coordinates of the origin of one system with respect to the other system.

Reference systems will be identified with natural numbers, and the number zero is reserved for a privileged system called the *world reference system*. A left superscript is used to identify the reference system that a vector or a transformation is written in. A left superscript of zero can be optionally omitted.

Thus, ${}^2\mathbf{p}$ is the vector of the coordinates of point $\mathbf{p}$ in reference frame 2. The vector of world coordinates of the same point can be written as either ${}^0\mathbf{p}$ or just $\mathbf{p}$. The origin $\mathbf{t}$ of a reference system has always zero coordinates in that reference system, so

$${}^i\mathbf{t}_i = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

for all natural numbers i, and therefore also

$$\mathbf{t}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

since a zero left superscript is implied. Similarly, if **i**, **j**, **k** are the unit points of a reference system, we have

$$\left[\begin{array}{ccc} {}^i\mathbf{i}_i & {}^i\mathbf{j}_i & {}^i\mathbf{k}_i \end{array}\right] = \left[\begin{array}{ccc} \mathbf{i}_0 & \mathbf{j}_0 & \mathbf{k}_0 \end{array}\right] = I \; ,$$

the 3 × 3 identity matrix.

1.7 Rotation

A *rotation* is a transformation between two Cartesian references systems of equal origin and handedness and with unit points of equal magnitude. Let the two systems be S_0 (the world reference system) and S_1. Then, the common origin is

$$\mathbf{t}_0 = \mathbf{t}_1 = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right]$$

where the left superscripts were omitted because coordinates refer to the world reference system. The unit points of S_1 are $\mathbf{i}_1$, $\mathbf{j}_1$, $\mathbf{k}_1$ when their coordinates are expressed in S_0 (implied left superscript 0). Then a point with coordinates $\mathbf{p} = (x, y, z)^T$ in S_0 can be reached from the origin $\mathbf{t}_1$ common to the two systems by a polygonal path with the following four vertices:

$$\mathbf{t}_1 \quad , \quad \mathbf{a} = {}^1\!x\, \mathbf{i}_1 \quad , \quad \mathbf{b} = {}^1\!x\, \mathbf{i}_1 + {}^1\!y\, \mathbf{j}_1 \quad , \quad \mathbf{p} = {}^1\!x\, \mathbf{i}_1 + {}^1\!y\, \mathbf{j}_1 + {}^1\!z\, \mathbf{k}_1 \; .$$

The steps of this path are along the axes of S_1. The numbers ${}^1x$, ${}^1y$, ${}^1z$ are the signed magnitudes of the steps, and also the coordinates of the point in S_1. These step sizes are the signed magnitudes of the orthogonal projections of the point onto $\mathbf{i}_1$, $\mathbf{j}_1$, $\mathbf{k}_1$, and from Theorem 1.3 we see that

$${}^1\!x = \mathbf{i}_1^T \mathbf{p} \quad , \quad {}^1\!y = \mathbf{j}_1^T \mathbf{p} \quad , \quad {}^1\!z = \mathbf{k}_1^T \mathbf{p}$$

because the vectors $\mathbf{i}_1$, $\mathbf{j}_1$, $\mathbf{k}_1$ have unit norm. These three equations can be packaged into a single matrix equation that expresses the vector ${}^1\mathbf{p} = ({}^1x, {}^1y, {}^1z)^T$ as a function of $\mathbf{p}$:

$${}^1\mathbf{p} \; = \; R_1\, \mathbf{p} \quad \text{where} \quad R_1 = {}^0\!R_1 = \left[\begin{array}{c} \mathbf{i}_1^T \\ \mathbf{j}_1^T \\ \mathbf{k}_1^T \end{array}\right]$$

where the 3 × 3 matrix ${}^0R_1$ is called a *rotation* matrix.

This result was obtained without using the privileged status of the world reference system S_0 (except to omit some of the left superscripts). Therefore, the result must be general: Given any two Cartesian reference systems S_a and S_b with a common origin,

$${}^b\mathbf{p} = {}^a\!R_b\, {}^a\mathbf{p} \quad \text{where} \quad {}^a\!R_b = \left[\begin{array}{c} {}^a\mathbf{i}_b^T \\ {}^a\mathbf{j}_b^T \\ {}^a\mathbf{k}_b^T \end{array}\right] \; .$$

A rotation is a reversible transformation, and therefore the matrix ${}^aR_b$ must have an *inverse*, another matrix ${}^bR_a$ that transforms back from S_b to S_a:

$${}^a\mathbf{p} = {}^b\!R_a\, {}^b\mathbf{p} \; .$$

The proof of the following fact is given in the Appendix.

Theorem 1.6. *The inverse R^{-1} of a rotation matrix R is its transpose:*

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

Equivalently, if ${}^aR_b$ is the rotation whose rows are the unit points of reference systems b expressed in reference system a, then

$${}^bR_a = {}^aR_b^T \; .$$

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 $R = {}^aR_b$ are the unit vectors of S_b expressed in S_a, so its columns (the rows of the inverse transformation $R^T = {}^bR_a$) are the unit vectors of S_a expressed in S_b.

The equations in Theorem 1.6 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

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix}$$

where s_x, s_y s_z 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, the directions of both the x and the y axis can be flipped simultaneously by a 180-degree rotation around the z axis. No rotation can flip the directions 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 in turn 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 $R^T R = R R^T = I$ and $\det(R) = 1$.
> A diagonal matrix S is an inversion if and only if $S^T S = S S^T = I$ and $\det(S) = -1$.

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 relatively lengthy sequence of algebraic steps. This Section reviews orthogonality of rotation matrices from a geometric point of view, and derives the result above by conceptually simpler means. The rows of the rotation matrix

$$^aR_b = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} {}^a\mathbf{i}_b^T \\ {}^a\mathbf{j}_b^T \\ {}^a\mathbf{k}_b^T \end{bmatrix}$$

are by definition the unit vectors of the reference system S_b, expressed in the reference system S_a. This means that its entry r_{mn} is the signed magnitude of the orthogonal projection of the m-th unit vector in S_b onto the n-th unit vector in S_a. For instance,

$$r_{12} = {}^a\mathbf{i}_b^T \, {}^a\mathbf{j}_a \quad \text{and} \quad r_{31} = {}^a\mathbf{k}_b^T \, {}^a\mathbf{i}_a \ .$$

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 of the new system and the j-th axis of 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 ${}^a\mathbf{i}_b$ onto ${}^a\mathbf{j}_a$ is the same as the signed magnitude of the orthogonal projection of ${}^a\mathbf{j}_a$ onto ${}^a\mathbf{i}_b$. Since angles do not depend on reference system, the projection is the same when expressed in S_b:

$$r_{12} = {}^a\mathbf{i}_b^T \, {}^a\mathbf{j}_a = {}^a\mathbf{j}_a^T \, {}^a\mathbf{i}_b = {}^b\mathbf{j}_a^T \, {}^b\mathbf{i}_b$$

where in the second equality we merely switched the two vectors with each other and in the third we changed the reference system (*i.e.*, both left superscripts) from S_a to S_b.

This symmetry is the deep reason for orthogonality: When we want to go from the "new" system S_b back to the "old" system S_a through the inverse matrix $R^{-1} = {}^bR_a$, we seek to express the unit vectors of S_a in the system S_b, that is, we seek the signed magnitudes of the orthogonal projections of each unit vector of the "old" system S_a onto each of the unit vectors of the "new" system S_b . 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*:

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

The Cross Product is Covariant with Rotations. We saw that the cross product of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a third vector $\mathbf{c}$ that is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$, such that the triple $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is right-handed, and such as the signed magnitude of $\mathbf{c}$ is the product $\|\mathbf{a}\| \, \|\mathbf{b}\|$ times the magnitude of the sine of the angle θ between $\mathbf{a}$ and $\mathbf{b}$. If $\mathbf{a}$ and $\mathbf{b}$ are simultaneously rotated by the same rotation R to produce the new vectors $\mathbf{a}'$ and $\mathbf{b}'$, then the line orthogonal to $\mathbf{a}$ and $\mathbf{b}$ rotates the same way, because rotating $\mathbf{a}$ and $\mathbf{b}$ is the same as rotating the reference system in the opposite direction. Thus, the direction of $\mathbf{c}' = \mathbf{a}' \times \mathbf{b}'$ is that of $R\mathbf{c}$. In addition, a rotation does not change the magnitudes of vectors it is applied to, nor does it change the angle between any pair of vectors, or their handedness. Therefore, $\mathbf{c}' = R\mathbf{c}$. To summarize, this argument shows that the cross product is covariant with rotations:

$$(R\mathbf{a}) \times (R\mathbf{b}) = R\,(\mathbf{a} \times \mathbf{b}) \; .$$

In words, if you rotate the inputs to a cross product, the output rotates the same way.

1.8 Coordinate Transformation

A right-handed Cartesian system of reference S_1 can differ from the world reference system S_0 by a translation of the origin from $\mathbf{t}_0 = (0,0,0)^T$ to $\mathbf{t}_1$ and a rotation of the axes from unit points $\mathbf{i}_0 = \mathbf{e}_x$, $\mathbf{j}_0 = \mathbf{e}_y$, $\mathbf{k}_0 = \mathbf{e}_z$ to unit points $\mathbf{i}_1, \mathbf{j}_1, \mathbf{k}_1$. Suppose that the origin of frame S_0 is first translated to point $\mathbf{t}_1$ and *then* the resulting frame is rotated by R_1 (see Figure 3). Given a point with coordinates $\mathbf{p} = (x, y, z)^T$ in S_0, the coordinates ${}^1\mathbf{p} = ({}^1x, {}^1y, {}^1z)^T$ of the same point in S_1 are then

$$ {}^1\mathbf{p} = R_1(\mathbf{p} - \mathbf{t}_1) \; . \tag{1}$$

The translation is applied first, to yield the new coordinates $\mathbf{p} - \mathbf{t}_1$ in an intermediate frame. This translation does not change the directions of the coordinate axes, so the rotation from the intermediate frame to S_1 is the same rotation R_1 as from S_0 to S_1, which is applied thereafter.

The inverse transformation applies the inverse operations in reverse order:

$$\mathbf{p} = R_1^T\,{}^1\mathbf{p} + \mathbf{t}_1 \; . \tag{2}$$

This can also be verified algebraically from equation (1): Multiplying both sides by R_1^T from the left yields

$$R_1^T\,{}^1\mathbf{p} = R_1^T R_1(\mathbf{p} - \mathbf{t}_1) = \mathbf{p} - \mathbf{t}_1$$

and adding $\mathbf{t}_1$ to both sides yields equation (2). Thus,

$${}^1R_0 = {}^0R_1^T \quad \text{and} \quad {}^1\mathbf{t}_0 = -{}^0R_1\,{}^0\mathbf{t}_1$$

since

$$\mathbf{p} = R_1^T\,{}^1\mathbf{p} + \mathbf{t}_1 = R_1^T({}^1\mathbf{p} - (-R_1\mathbf{t}_1)) \; .$$

More generally, if

$${}^b\mathbf{p} = {}^aR_b({}^a\mathbf{p} - {}^a\mathbf{t}_b) \tag{3}$$

then

$${}^a\mathbf{p} = {}^bR_a({}^b\mathbf{p} - {}^b\mathbf{t}_a) \quad \text{where} \quad {}^bR_a = {}^aR_b^T \quad \text{and} \quad {}^b\mathbf{t}_a = -{}^aR_b\,{}^a\mathbf{t}_b \; . \tag{4}$$

The transformations (3) and (4) are said to be *rigid*, in that they preserve both distances and handedness. 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 the rotation matrix is 1, rather than 1 just in magnitude.

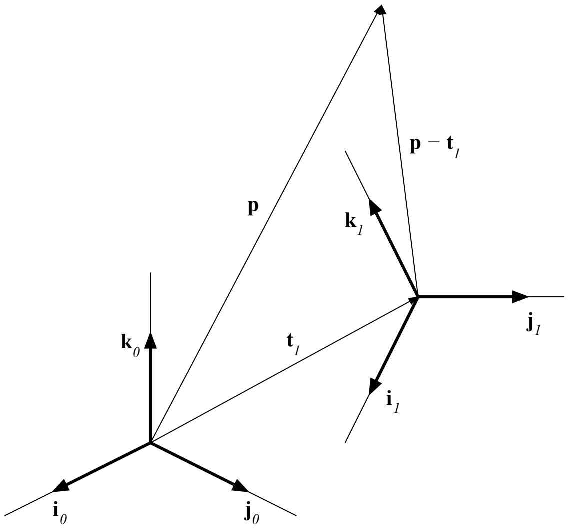


Figure 3: Transformation between two reference systems.

Chaining Rigid Transformations If two right-handed Cartesian systems S_a and S_b are given, say, in the world reference system S_0, then the transformation from S_a to S_b can be obtained in two steps, just as we did for rotations: First transform from S_a to S_0, then from S_0 to S_b:

$$\mathbf{p} = {}^0\mathbf{p} = {}^aR_0({}^a\mathbf{p} - {}^a\mathbf{t}_0) = {}^0R_a^T({}^a\mathbf{p} + {}^0R_a\,{}^0\mathbf{t}_a) = R_a^T\,{}^a\mathbf{p} + \mathbf{t}_a$$

and

$${}^b\mathbf{p} = {}^0R_b({}^0\mathbf{p} - {}^0\mathbf{t}_b) = R_b(\mathbf{p} - \mathbf{t}_b)$$

so that

$${}^b\mathbf{p} = R_b(R_a^T\,{}^a\mathbf{p} + \mathbf{t}_a - \mathbf{t}_b) = R_bR_a^T[{}^a\mathbf{p} + R_a(\mathbf{t}_a - \mathbf{t}_b)]\ ,$$

that is,

$${}^b\mathbf{p} = {}^aR_b({}^a\mathbf{p} - {}^a\mathbf{t}_b) \quad \text{where} \quad {}^aR_b = R_bR_a^T \quad \text{and} \quad {}^a\mathbf{t}_b = R_a(\mathbf{t}_b - \mathbf{t}_a)\ .$$

The following box summarizes these results:

If

$$ {}^b\mathbf{p} = R_b(\mathbf{p} - \mathbf{t}_b) \quad \text{and} \quad {}^a\mathbf{p} = R_a(\mathbf{p} - \mathbf{t}_a) $$

are the transformations between world coordinates and reference frames S_a and S_b, then the transformation from S_a to S_b is

$$ {}^b\mathbf{p} = {}^aR_b({}^a\mathbf{p} - {}^a\mathbf{t}_b) \quad \text{where} \quad {}^aR_b = R_b R_a^T \quad \text{and} \quad {}^a\mathbf{t}_b = R_a(\mathbf{t}_b - \mathbf{t}_a) $$

and the reverse transformation, from S_b to S_a, is

$$ {}^a\mathbf{p} = {}^bR_a({}^b\mathbf{p} - {}^b\mathbf{t}_a) \quad \text{where} \quad {}^bR_a = R_a R_b^T \quad \text{and} \quad {}^b\mathbf{t}_a = R_b(\mathbf{t}_a - \mathbf{t}_b) \; . $$

Consistently with these relations,

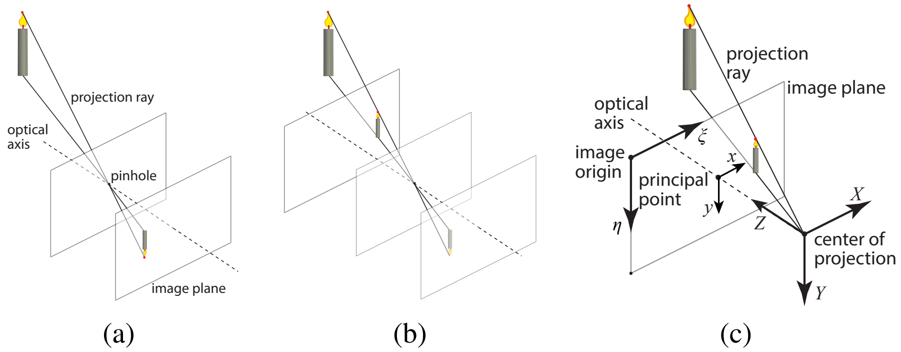
$$ {}^bR_a = {}^aR_b^T \quad \text{and} \quad {}^b\mathbf{t}_a = -{}^aR_b\,{}^a\mathbf{t}_b \; . $$

2 The Pinhole Camera Model

The images we process in computer vision are formed by light bouncing off surfaces in the world and into the lens of the camera. The light then hits an array of sensors, called *pixels*, inside the camera. Each sensor produces electric charges that are read by an electronic circuit and converted to voltages. These are in turn sampled by a device called a digitizer (or analog-to-digital converter) to produce the numbers that computers eventually process, called pixel values. Thus, the pixel values are a rather indirect encoding of the physical properties of visible surfaces. Is it not amazing that all those numbers in an image file carry information on how the properties of a packet of photons were changed by bouncing off a surface in the world? Even more amazing is that from this information we can perceive shapes and colors.

The study of what happens to the light that leaves surfaces in the world and makes it to the camera sensor is often encapsulated into what computer vision calls the *pinhole camera model*, a very much simplified description of camera optics that encapsulates the geometry of *perspective projection*. This section introduces this model. A later note, devoted to camera calibration, explains the key differences between the pinhole camera model and real lenses, and also describes what happens at the pixel level.

A pinhole camera is a box with five opaque faces and a translucent one. A very small hole is punched in the face of the box opposite to the translucent face. If you consider a single point in the world, such as the tip of the candle flame in Figure 4(a), only a thin beam from that point enters the pinhole and hits the translucent screen. Thus, the pinhole acts as a selector of light rays: without the pinhole and the box, any point on the screen would be illuminated from a whole hemisphere of directions, yielding a uniform coloring. With the pinhole, on the other hand, an inverted image of the visible world is formed on the screen. Since the screen is translucent, the image can be seen from outside the box. When the pinhole is reduced to a single point, this image is formed where the plane of the screen intersects the star of rays through the pinhole. Of course, a pinhole reduced to a point is an idealization: no power would pass through such a pinhole, and the image would be infinitely dim (black).



(a) (b) (c)

Figure 4: (a) Projection geometry for a pinhole camera. (b) If a screen could be placed in front of the pinhole, rather than behind, without blocking the projection rays, then the image would be upside-up. (c) What is left is the so-called *pinhole camera model*. The camera coordinate frame (X, Y, Z) is right-handed.

The fact that the image on the screen is inverted is mathematically inconvenient. It is therefore customary to consider instead the intersection of the star of rays through the pinhole with a plane parallel to the screen and *in front* of the pinhole as shown in Figure 4(b). This is of course an even greater idealization, since a screen in this position would block the light rays. So this model is a mathematical artifact, but a convenient one to use when doing mathematics. The new image is isomorphic to the old one, produced by the real pinhole camera, but upside-up.

In this model, the pinhole is called more appropriately the *center of projection* (Figure 4(c)). The front screen is the *image plane*. The distance between center of projection and image plane is the *focal distance*, and is denoted with f. The *optical axis* is the line through the center of projection that is perpendicular to the image plane. The point where the optical axis pierces the sensor plane is the *principal point*.

The origin of the *pixel image coordinate system* (ξ, η) is placed in the top left corner of the image. The *camera reference system* (X, Y, Z) axes are respectively parallel to ξ, η, and the optical axis, and the Z axis points towards the scene. With the choice in Figure 4(c), the camera reference system is right-handed. The Z coordinate of a point in the world is called the point's *depth*. The *canonical image coordinate system* (x, y) is oriented like the pixel image coordinate system, but its origin is at the principal point. Therefore, for a point in the image plane

$$x = X \quad \text{and} \quad y = Y \; .$$

The units used to measure point coordinates in the camera reference system (X, Y, Z) and in the canonical image reference system (x, y) are different from those used in the pixel image reference system (ξ, η). Typically, metric units (meters, centimeters, millimeters) are used in the first two systems system and pixels in the pixel image system. Pixels are the individual, rectangular elements on a digital camera's sensing array. Since pixels are not necessarily square, there may be a different number of pixels in a millimeter measured horizontally on the array than in a millimeter measured vertically, so two separate conversion units are needed to convert pixels to millimeters (or *vice versa*) in the two directions.

Every point on the image plane has a Z coordinate equal to f in the camera reference system. Both image reference systems, on the other hand, are two-dimensional, so the third coordinate is undefined in these systems, which differ from each other by a translation and two separate unit conversions:

Let ξ_0 and η_0 be the coordinates in pixels of the principal point $\boldsymbol{\pi}_0$ of the image in the pixel image reference system (ξ, η). Then an image point $\mathbf{p}$ with coordinates (x, y) in millimeters in the canonical image reference system has pixel image coordinates (in pixels)

$$\xi = s_x x + \xi_0 \quad \text{and} \quad \eta = s_y y + \eta_0 \tag{5}$$

where s_x and s_y are scaling constants expressed in pixels per millimeter.

The *projection equations* relate the camera-system coordinates $\mathbf{P} = (X, Y, Z)$ of a point in space to the canonical image coordinates $\mathbf{p} = (x, y)$ of the projection of $\mathbf{P}$ onto the image plane and then, in turn, to the pixel image coordinates $\boldsymbol{\pi} = (\xi, \eta)$ of the projection. These equations can be easily derived for the x coordinate from the top view in Figure 5. From this Figure we see that the triangle with orthogonal sides of length X and Z is similar to that with orthogonal sides of length x and f (the focal distance), so that

$X/Z = x/f$. Similarly, for the Y coordinate, one gets $Y/Z = y/f$. In conclusion,

Under perspective projection, the world point $\mathbf{P}$ with coordinates (X, Y, Z) projects to the image point with coordinates

$$\begin{aligned} x &= f\frac{X}{Z} \\ y &= f\frac{Y}{Z} \,. \end{aligned} \tag{6}$$

One way to make units of measure consistent in these projection equations is to measure all quantities in the same unit, say, millimeters. In this case, the two constants s_x and s_y in equation (5) have the dimension of pixels per millimeter. However, it is sometimes more convenient to express x, y, and f in pixels (image dimensions) and X, Y, Z in millimeters (world dimensions). The ratios x/f, y/f, X/Z, and Y/Z are then dimensionless, so the equations (6) are dimensionally consistent with this choice as well. In this case, the two constants s_x and s_y in equation (5) are dimensionless as well.

An even simpler choice for the projection equations (6) is to express x, y, and f in units of focal distance, so that $f = 1$. In that case, of course, s_x and s_y have the dimensions of pixels per focal distance.

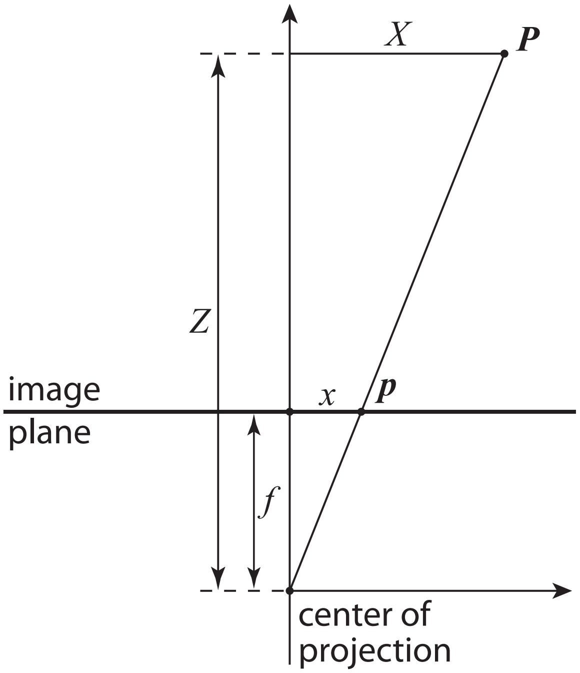


Figure 5: A top view of figure 4 (c).

Equations (5) and (6) can be written somewhat more compactly as follows:

$$\mathbf{p} = K_f \frac{\mathbf{P}}{Z} \quad \text{and} \quad \boldsymbol{\xi} = K_s \mathbf{p} + \boldsymbol{\pi}_0$$

where

$$K_f = \left[\begin{array}{ccc} f & 0 & 0 \\ 0 & f & 0 \end{array} \right] \quad \text{and} \quad K_s = \left[\begin{array}{cc} s_x & 0 \\ 0 & s_y \end{array} \right] .$$

Even more compact notation could be attained by using homogeneous coordinates. However, the additional compactness does not justify the cost of introducing this representation given the scope of these notes.

Of course, if the world points $\mathbf{P}$ are in a frame of reference different from the camera's, coordinates are to be transformed by an appropriate rigid transformation into the camera's reference frame, before applying the projection equations.

Appendix: Proofs

Theorem 1.1

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

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 1.3

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, rank-1 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[3] $P_{\mathbf{b}}$ 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 1.5

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 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:

$$\begin{array}{rcl}
\mathbf{c}^T\mathbf{a} &=& (a_y b_z - a_z b_y)a_x \,+\, (a_z b_x - a_x b_z)a_y \,+\, (a_x b_y - a_y b_x)a_z = 0 \\
\mathbf{c}^T\mathbf{b} &=& (a_y b_z - a_z b_y)b_x \,+\, (a_z b_x - a_x b_z)b_y \,+\, (a_x b_y - a_y b_x)b_z = 0
\end{array}$$

(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:

$$\begin{array}{rcl}
(\mathbf{a}^T\mathbf{b})^2 &=& (a_x b_x + a_y b_y + a_z b_z)\,(a_x b_x + a_y b_y + a_z b_z) \\
&=& \;\;a_x^2 b_x^2 + a_x b_x a_y b_y + a_x b_x a_z b_z \\
&& +a_y^2 b_y^2 + a_x b_x a_y b_y + a_y b_y a_z b_z \\
&& +a_z^2 b_z^2 + a_x b_x a_z b_z + a_y b_y a_z b_z \\
&=& a_x^2 b_x^2 + a_y^2 b_y^2 + a_z^2 b_z^2 + 2a_x b_x a_y b_y + 2a_y b_y a_z b_z + 2a_x b_x a_z b_z
\end{array}$$

and

$$\begin{array}{rcl}
\|\mathbf{a}\times\mathbf{b}\|^2 &=& (a_y b_z - a_z b_y)^2 \,+\, (a_z b_x - a_x b_z)^2 \,+\, (a_x b_y - a_y b_x)^2 \\
&=& \;\;a_y^2 b_z^2 + a_z^2 b_y^2 - 2a_y b_y a_z b_z \\
&& +a_x^2 b_z^2 + a_z^2 b_x^2 - 2a_x b_x a_z b_z \\
&& +a_x^2 b_y^2 + a_y^2 b_x^2 - 2a_x b_x a_y b_y \\
&=& \;\;a_x^2 b_y^2 + a_y^2 b_x^2 + a_y^2 b_z^2 + a_z^2 b_y^2 + a_x^2 b_z^2 + a_z^2 b_x^2 \\
&& -2a_x b_x a_y b_y - 2a_y b_z a_y b_y - 2a_x b_x a_z b_z
\end{array}$$

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

so that

$$(\mathbf{a}^T\mathbf{b})^2 \;+\; \|\mathbf{a}\times\mathbf{b}\|^2 \;=\; a_x^2b_x^2 + a_x^2b_y^2 + a_x^2b_z^2 + a_y^2b_x^2 + a_y^2b_y^2 + a_y^2b_z^2 + a_z^2b_x^2 + a_z^2b_y^2 + a_z^2b_z^2$$

but also

$$\|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \;=\; a_x^2b_x^2 + a_x^2b_y^2 + a_x^2b_z^2 + a_y^2b_x^2 + a_y^2b_y^2 + a_y^2b_z^2 + a_z^2b_x^2 + a_z^2b_y^2 + a_z^2b_z^2$$

so that

$$(\mathbf{a}^T\mathbf{b})^2 \;+\; \|\mathbf{a}\times\mathbf{b}\|^2 \;=\; \|\mathbf{a}\|^2\,\|\mathbf{b}\|^2 \tag{7}$$

as desired. The result on the magnitude is a consequence of equation (7). 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

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

Theorem 1.6

The inverse R^{-1} of a rotation matrix R is its transpose:

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

Equivalently, if ${}^aR_b$ is the rotation whose rows are the unit points of reference systems b expressed in reference system a, then

$${}^bR_a = {}^aR_b^T\;.$$

Proof. Assume that $a = 0$ and $b = 1$, so that left superscripts equal to 0 can be omitted. This assumption can be made without loss of generality, because the following proof makes no use of the privileged nature of system S_0, other than for simplifying away left superscripts. Also, for further brevity, let

$$R = {}^0R_1\;.$$

When we rotate point $\mathbf{p}$ through R we obtain a vector ${}^1\mathbf{p}$ of coordinates in S_1. We then look for a new matrix $R^{-1} = {}^1R_0$ that applied to ${}^1\mathbf{p}$ gives back the original vector $\mathbf{p}$:

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

that is, by combining these two equations,

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

Since this equality 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 R^{-1} is called 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 of this fact is a single line: suppose that B is the right inverse of A and 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 **i**, **j**, **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.