Chapter 2

Representation of a three dimensional moving scene

The study of the geometric relationship between a three-dimensional scene and its images taken from a moving camera boils down to the interplay between two fundamental transformations: the rigid body motion that models how the camera moves, and the perspective projection which describes the image formation process. Long before these two transformations were brought together, the theories for each had been developed independently.

The study of the principles of motion of a material body has a long history belonging to the foundations of physics. For our purpose, the more recent noteworthy insights to the understanding of the motion of rigid bodies came from Chasles and Poinsot in the early 1800s. Their findings led to current treatment of the subject which has since been widely adopted.

We start this chapter with an introduction to the three dimensional Euclidean space as well as to rigid body transformations. The next chapter will then focus on the perspective projection model of the camera.

2.1 Three-dimensional Euclidean space

We will use $\mathbb{E}^3$ to denote the familiar three-dimensional Euclidean space. In general, a Euclidean space is a set whose elements satisfy the five axioms of Euclid []. More practically, the three-dimensional Euclidean space can be represented by a (global) Cartesian coordinate frame: every point $p \in \mathbb{E}^3$ can be identified with a point in $\mathbb{R}^3$ by three coordinates: $\mathbf{X} \doteq [X_1, X_2, X_3]^T$. Sometime we will use $[X, Y, Z]^T$ to indicate individual coordinates instead of $[X_1, X_2, X_3]^T$. Through such an assignment of Cartesian frame, one establishes a one-to-one correspondence between $\mathbb{E}^3$ and $\mathbb{R}^3$, which allows us to safely talk about points and their coordinates as if they were the same thing.

Cartesian coordinates are the first step towards being able to measure distances and angles. In order to do so, $\mathbb{E}^3$ must be endowed with a *metric*. A precise definition of metric relies on the notion of *vector*. In the Euclidean space, a *vector* is identified by a pair of points $p, q \in \mathbb{E}^3$; that is, a vector v is defined as a directed arrow connecting p to q. The point p is usually called the base point of v. In coordinates, the vector v is represented by the triplet $[v_1, v_2, v_3]^T \in \mathbb{R}^3$, where each coordinate is the difference between the corresponding coordinates of the two points: if p has coordinates $\mathbf{X}$ and q has coordinates $\mathbf{Y}$, then v has coordinates[1]

$$v = \mathbf{Y} - \mathbf{X} \in \mathbb{R}^3.$$

One can also introduce the concept of *free vector*, a vector whose definition does not depend on its base point. If we have two pairs of points (p, q) and (p', q') with coordinates satisfying $\mathbf{Y} - \mathbf{X} = \mathbf{Y}' - \mathbf{X}'$, we say that they define the same free vector. Intuitively, this allows a vector v to be transported in parallel anywhere in $\mathbb{E}^3$. In particular, without loss of generality, one can assume that the base point is the origin of the Cartesian frame, so that $\mathbf{X} = 0$ and $v = \mathbf{Y}$. Note, however, that this notation is confusing: $\mathbf{Y}$ here denotes the coordinates of a vector, that happen to be the same as the coordinates of the point q just because we have chosen the point p to be the origin. Nevertheless, the reader should keep in mind that points and vectors are different geometric objects; for instance, as we will see shortly, a rigid body motion acts differently on points and vectors. So, keep the difference in mind!

The set of all (free) vectors form a *linear (vector) space*[2], where a linear combination of two vectors $v, u \in \mathbb{R}^3$ is given by:

$$\alpha v + \beta u = (\alpha v_1 + \beta u_1, \alpha v_2 + \beta u_2, \alpha v_3 + \beta u_3)^T \in \mathbb{R}^3, \quad \forall \alpha, \beta \in \mathbb{R}.$$

The *Euclidean metric* for $\mathbb{E}^3$ is then defined simply by an inner product on its vector space:

Definition 2.1 (Euclidean metric and inner product). *A bilinear function* $\langle \cdot, \cdot \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ *is an* inner product *if it is linear, symmetric and positive definite. That is,* $\forall u, v, w \in \mathbb{R}^3$

1. $\langle u, \alpha v + \beta w \rangle \; = \; \alpha \langle u, v \rangle + \beta \langle u, w \rangle, \quad \forall \; \alpha, \beta \in \mathbb{R},$
2. $\langle u, v \rangle \; = \; \langle v, u \rangle.$
3. $\langle v, v \rangle \; \geq \; 0$ *and* $\langle v, v \rangle \; = \; 0 \; \Leftrightarrow \; v = 0,$

The quantity $\|v\| = \sqrt{\langle v, v \rangle}$ is called the *Euclidean norm* (or 2-norm) of the vector v. It can be shown that, by a proper choice of the Cartesian frame, any inner product in $\mathbb{E}^3$ can be converted to the following familiar form:

$$\langle u, v \rangle \; = \; u^T v = u_1 v_1 + u_2 v_2 + u_3 v_3. \tag{2.1}$$

In most of this book (but not everywhere!) we will use the canonical inner product $\langle u, v \rangle \; = \; u^T v$ and, consequently, $\|v\| \; = \; \sqrt{v_1^2 + v_2^2 + v_3^2}$. When the inner product between two vectors is zero, $\langle u, v \rangle$, they are said to be *orthogonal*.

Finally, a Euclidean space $\mathbb{E}^3$ can then be formally described as a space which, with respect to a Cartesian frame, can be identified with $\mathbb{R}^3$ and has a metric (on its vector space) given by the above inner product. With such a metric, one can measure not only distances between points or angles between vectors, but also calculate the length of a curve, or the volume of a region[3]

While the inner product of two vectors returns a scalar, the so-called *cross product* returns a vector instead.

---

[1]Note that we use the same symbol v for a vector and its coordinates.

[2]Note that points do not.

[3]For example, if the trajectory of a moving particle p in $\mathbb{E}^3$ is described by a curve $\gamma(\cdot) : t \mapsto \mathbf{X}(t) \in \mathbb{R}^3, t \in [0, 1]$, then the total length of the curve is given by:

$$l(\gamma(\cdot)) = \int_0^1 \|\dot{\mathbf{X}}(t)\| \; dt.$$

where $\dot{\mathbf{X}}(t) = \frac{d}{dt}(\mathbf{X}(t)) \in \mathbb{R}^3$ is the so-called tangent vector to the curve.

Definition 2.2 (Cross product). *Given two vectors $u, v \in \mathbb{R}^3$, their* cross product *is a third vector with coordinates given by:*

$$u \times v = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix} \in \mathbb{R}^3.$$

It is immediate from this definition that the cross product of two vectors is linear: $u \times (\alpha v + \beta w) = \alpha u \times v + \beta u \times w \ \forall \ \alpha, \beta \in \mathbb{R}$. Furthermore, it is immediate to verify that

$$\langle u \times v, u \rangle \ = \ \langle u \times v, v \rangle \ = 0, \quad u \times v = -v \times u.$$

Therefore, the cross product of two vector is orthogonal to each of its factors, and the order of the factors defines an *orientation.*

If we fix u, the cross product can be interpreted as a map $v \mapsto u \times v$ between $\mathbb{R}^3$ and $\mathbb{R}^3$. Due to the linearity property, this map is in fact linear and, therefore, like all linear maps between vector spaces, it can be represented by a matrix. We call such a matrix $\widehat{u} \in \mathbb{R}^{3\times 3}$. It is immediate to verify by substitution that this matrix is given by

$$\widehat{u} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3}. \tag{2.2}$$

Hence, we can write $u \times v = \widehat{u}v$. Note that[4] $\widehat{u}$ is a 3×3 skew-symmetric matrix, i.e. $\widehat{u}^T = -\widehat{u}$. It is immediate to verify that for $e_1 \doteq [1,0,0]^T, e_2 \doteq [0,1,0]^T \in \mathbb{R}^3$, we have $e_1 \times e_2 = [0,0,1]^T \doteq e_3$. That is for, a standard Cartesian frame, the cross product of the principal axes X and Y gives the principal axis Z. The cross product therefore conforms with the *right-hand rule.*

The cross product allows us to define a map between a vector, u, and a skew-symmetric, 3×3 matrix $\widehat{u}$. Is the converse true? Can every 3×3 skew-symmetric matrix be associated with a three-dimensional vector u? The answer is yes, as it is easy to verify. Let $M \in \mathbb{R}^{3\times 3}$ be skew-symmetric, that is $M = -M^T$. By writing this equation in terms of the elements of the matrix, we conclude that $m_{11} = m_{22} = m_{23} = 0$ and $m_{ij} = -m_{ji}, \ i,j = 1,2,3$. This shows that a skew-symmetric matrix has only three degrees of freedom, for instance m_{21}, m_{13}, m_{32}. If we call $u_1 = m_{32}; \ u_2 = m_{13}; \ u_3 = m_{21}$, then $\widehat{u} = M$. Indeed, the vector space $\mathbb{R}^3$ and the space of skew-symmetric 3×3 matrices $so(3)$ can be considered as the same thing, with the cross product that maps one onto the other, $\times : \mathbb{R} \rightarrow so(3); \ u \mapsto \widehat{u}$, and the inverse map, called "vee", that extracts the components of the vector u from the skew-symmetric matrix $\widehat{u}$: $\vee : so(3) \rightarrow \mathbb{R}; \ M = -M^T \mapsto M^\vee = [m_{32}, m_{13}, m_{21}]^T$.

2.2 Rigid body motion

Consider an object moving in front of a camera. In order to describe its motion one should, in principle, specify the trajectory of every single point on the object, for instance by giving its coordinates as a function of time $\mathbf{X}(t)$. Fortunately, for rigid objects we do not need to specify the motion of every particle. As we will see shortly, it is sufficient to specify the motion of a point, and the motion of three coordinate axes attached to that point.

[4] In some computer vision literature, the matrix $\widehat{u}$ is also denoted as $u_\times$.

The condition that defines a rigid object is that the distance between any two points on it does not change over time as the object moves. So if $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are the coordinates of any two points p and q respectively, their distance between must satisfy:

$$\|\mathbf{X}(t) - \mathbf{Y}(t)\| = \text{constant}, \quad \forall t \in \mathbb{R}. \tag{2.3}$$

In other words, if v is the vector defined by the two points p and q, then the norm (or length) of v remains the same as the object moves: $\|v(t)\| = \text{constant}$. A *rigid body motion* is then a family of transformations that describe how the coordinates of every point on the object change as a function of time. We denote it by g:

$$\begin{array}{rrcl} g(t): & \mathbb{R}^3 & \rightarrow & \mathbb{R}^3 \\ & \mathbf{X} & \mapsto & g(t)(\mathbf{X}) \end{array}$$

If, instead of looking at the entire continuous moving path of the object, we concentrate on the transformation between its initial and final configuration, this transformation is usually called a *rigid body displacement* and is denoted by a single mapping:

$$\begin{array}{rrcl} g: & \mathbb{R}^3 & \rightarrow & \mathbb{R}^3 \\ & \mathbf{X} & \mapsto & g(\mathbf{X}) \end{array}$$

Besides transforming the coordinates of points, g also induces a transformation on vectors. Suppose v is a vector defined by two points p and q: $v = \mathbf{Y} - \mathbf{X}$; then, after the transformation g, we obtain a new vector:

$$g_*(v) \doteq g(\mathbf{Y}) - g(\mathbf{X}).$$

That g preserves the distance between any two points can be expressed in terms of vectors as $\|g_*(v)\| = \|v\|$ for $\forall v \in \mathbb{R}^3$.

However, preserving distances between points is not the only requirement that a rigid object moving in space satisfies. In fact, there are transformations that preserve distances, and yet they are not physically realizable. For instance, the mapping

$$f : [X_1, X_2, X_3]^T \mapsto [X_1, X_2, -X_3]^T$$

preserves distance but not orientation. It corresponds to a reflection of points about the XY plane as a double-sided mirror. To rule out this type of mapping, we require that any rigid body motion, besides preserving distance, preserves orientation as well. That is, in addition to preserving the norm of vectors, it must preserve their cross product. The coordinate transformation induced by a rigid body motion is called a *special Euclidean transformation*. The word "special" indicates the fact that it is orientation-preserving.

Definition 2.3 (Rigid body motion or special Euclidean transformation). *A mapping $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a* rigid body motion *or a* special Euclidean transformation *if it preserves the norm and the cross product of any two vectors:*

1. *Norm:* $\|g_*(v)\| = \|v\|, \ \forall v \in \mathbb{R}^3$.
2. *Cross product:* $g_*(u) \times g_*(v) = g_*(u \times v), \ \forall u, v \in \mathbb{R}^3$.

In the above definition of rigid body motion, it is explicitly required that the distance between points be preserved. Then how about angles between vectors? Although it is not explicitly stated in the definition, angles are indeed preserved by any rigid body motion since the inner product $\langle \cdot, \cdot \rangle$ can be expressed in terms of the norm $\| \cdot \|$ by the *polarization identity*:

$$u^T v = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2). \tag{2.4}$$

Hence, for any rigid body motion g, one can show that:

$$u^T v = g_*(u)^T g_*(v), \quad \forall u, v \in \mathbb{R}^3. \tag{2.5}$$

In other words, a rigid body motion can also be defined as one that preserves both inner product and cross product.

How do these properties help us describe rigid motion concisely? The fact that distances and orientations are preserved in a rigid motion means that individual points cannot translate relative to each other. However, they can rotate relative to each other, but they have to collectively, in order to not alter any mutual distance between points. Therefore, a rigid body motion can be described by the motion of any one point on the body, and the rotation of a coordinate frame attached to that point. In order to do this, we represent the *configuration* of a rigid body by attaching a Cartesian coordinate frame to some point on the rigid body and keeping track of the motion of this coordinate frame relative to a fixed frame.

To see this, consider a "fixed" (world) coordinate frame, given by three *orthonormal* vectors $e_1, e_2, e_3 \in \mathbb{R}^3$; that is, they satisfy

$$e_i^T e_j = \delta_{ij} \begin{cases} \delta_{ij} = 1 & \text{for} \quad i = j \\ \delta_{ij} = 0 & \text{for} \quad i \neq j \end{cases}. \tag{2.6}$$

Typically, the vectors are ordered so as to form a right-handed frame: $e_1 \times e_2 = e_3$. Then, after a rigid body motion g, we have:

$$g_*(e_i)^T g_*(e_j) = \delta_{ij}, \quad g_*(e_1) \times g_*(e_2) = g_*(e_3). \tag{2.7}$$

That is, the resulting three vectors still form a right-handed orthonormal frame. Therefore, a rigid object can always be associated with a right-handed, orthonormal frame, and its rigid body motion can be entirely specified by the motion of such a frame, which we call the object frame. In in Figure 2.1 we show an object (in this case a camera) moving relative to a fixed "world" coordinate frame W. In order to specify the configuration of the camera relative to the world frame W, one may pick a fixed point o on the camera and attach to it an orthonormal frame, the camera coordinate frame C. When the camera moves, the camera frame also moves as if it were fixed to the camera. The configuration of the camera is then determined by (1) the vector between the origin of the world frame o and the camera frame, $g(o)$, called the "translational" part and denoted as T, and (2) the relative orientation of the camera frame C, with coordinate axes $g_*(e_1), g_*(e_2), g_*(e_3)$, relative to the fixed world frame W with coordinate axes e_1, e_2, e_3,called the "rotational" part and denoted by R.

In the case of vision, there is no obvious choice of the origin o and the reference frame e_1, e_2, e_3. Therefore, we could choose the world frame to be attached to the camera and specify the translation and rotation of the scene (assuming it is rigid!), or we could attach the world frame to the scene and specify the motion of the camera. All that matters is the *relative* motion between the scene and the camera, and what choice of reference frame to make is, from the point of view of geometry[5], arbitrary.

Remark 2.1. *The set of rigid body motions, or special Euclidean transformations, is a (Lie) group, the so-called* special Euclidean group, *typically denoted as $SE(3)$. Algebraically, a* group *is a set G, with an operation of (binary) multiplication $\circ$ on elements of G which is:*

- closed*: If $g_1, g_2 \in G$ then also $g_1 \circ g_2 \in G$;*
- associative*: $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$, for all $g_1, g_2, g_3 \in G$;*
- unit element e*: $e \circ g = g \circ e = g$, for all $g \in G$;*
- invertible: *For every element $g \in G$, there exists an element $g^{-1} \in G$ such that $g \circ g^{-1} = g^{-1} \circ g = e$.*

In the next few sections, we will focus on studying in detail how to represent *the special Euclidean group $SE(3)$. More specifically, we will introduce a way to realize elements in the special Euclidean group $SE(3)$ as elements in a group of $n \times n$ non-singular (real) matrices whose multiplication is simply the matrix multiplication. Such a group of matrices is usually called a general linear group and denoted as $GL(n)$ and such a representation is called a* matrix representation. *A representation is a map*

$$\begin{array}{rccc} \mathcal{R}: & SE(3) & \rightarrow & GL(n) \\ & g & \mapsto & \mathcal{R}(g) \end{array}$$

which preserves the group structure of $SE(3)$.[6] *That is, the inverse of a rigid body motion and the composition of two rigid body motions are preserved by the map in the following way:*

$$\mathcal{R}(g^{-1}) = \mathcal{R}(g)^{-1}, \quad \mathcal{R}(g \circ h) = \mathcal{R}(g)\mathcal{R}(h), \quad \forall g, h \in SE(3). \tag{2.8}$$

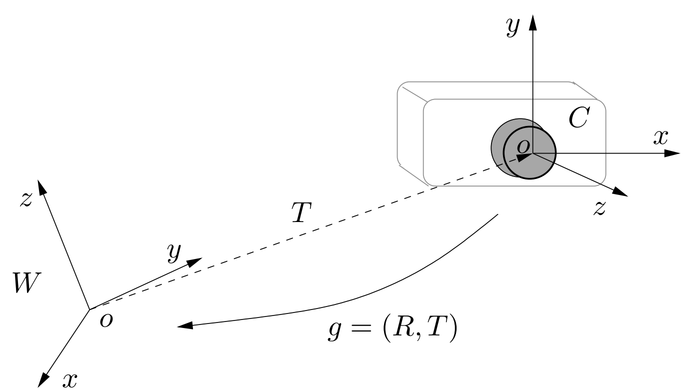

Figure 2.1: A rigid body motion which, in this instance, is between a camera and a world coordinate frame.

We start with the rotational component of motion.

[5]The neuroscience literature debates on whether the primate brain maintains a view-centered or an object-centered representation of the world. From the point of view of geometry, the two are equivalent, for they only differ by an arbitrary change of coordinates.

[6]Such a map is called *group homeomorphism* in algebra.

2.3 Rotational motion and its representations

Suppose we have a rigid object rotating about a fixed point $o \in \mathbb{E}^3$. How do we describe its orientation relative a chosen coordinate frame, say W? Without loss of generality, we may always assume that the origin of the world frame is the center of rotation o. If this is not the case, simply translate the origin to the point o. We now attach another coordinate frame, say C to the rotating object with origin also at o. The relation between these two coordinate frames is illustrated in Figure 2.2. Obviously, the configuration of the object is determined by the orientation of the frame

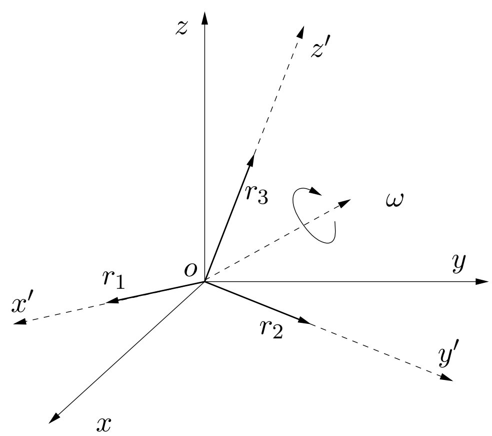

Figure 2.2: Rotation of a rigid body about a fixed point o. The solid coordinate frame W is fixed and the dashed coordinate frame C is attached to the rotating rigid body.

C. The orientation of the frame C relative to the frame W is determined by the coordinates of the three orthonormal vectors $r_1 = g_*(e_1), r_2 = g_*(e_2), r_3 = g_*(e_3) \in \mathbb{R}^3$ relative to the world frame W, as shown in Figure 2.2. The three vectors r_1, r_2, r_3 are simply the unit vectors along the three principal axes X', Y', Z' of the frame C respectively. The configuration of the rotating object is then completely determined by the following 3×3 matrix:

$$R_{wc} = [r_1, r_2, r_3] \in \mathbb{R}^{3\times 3}$$

with r_1, r_2, r_3 stacked in order as its three columns. Since r_1, r_2, r_3 form an orthonormal frame, it follows that:

$$r_i^T r_j = \delta_{ij} \begin{cases} \delta_{ij} = 1 & \text{for} \quad i = j \\ \delta_{ij} = 0 & \text{for} \quad i \neq j \end{cases} \qquad \forall i, j \in \{1, 2, 3\}.$$

This can be written in matrix form as:

$$R_{wc}^T R_{wc} = R_{wc} R_{wc}^T = I.$$

Any matrix which satisfies the above identity is called an *orthogonal matrix*. It follows from the definition that the inverse of an ortohgonal matrix is simply its transpose: $R_{wc}^{-1} = R_{wc}^T$. Since r_1, r_2, r_3 form a right-handed frame, we further have that the determinant of R_{wc} must be positive 1. This can be easily seen when looking at the determinant of the rotation matrix:

$$detR = r_1^T (r_2 \times r_3)$$

which is equal to 1 for right-handed coordinate systems. Hence R_{wc} is a *special orthogonal matrix* where, as before, the word "special" indicates orientation preserving. The space of all such special orthogonal matrices in $\mathbb{R}^{3\times 3}$ is usually denoted by:

$$\boxed{SO(3) = \{R \in \mathbb{R}^{3\times 3} \mid R^T R = I, \det(R) = +1\}}.$$

Traditionally, 3×3 special orthogonal matrices are called *rotation matrices* for obvious reasons. It is straightforward to verify that $SO(3)$ has a group structure. That is, it satisfies all four axioms of a group mentioned in the previous section. We leave the proof to the reader as an exercise. Therefore, the space $SO(3)$ is also referred to as the *special orthogonal group* of $\mathbb{R}^3$, or simply the *rotation group*. Directly from the definition of the rotation matrix, we can show that rotation indeed preserves both the inner product and the cross product of vectors. We also leave this as an exercise to the reader.

Going back to Figure 2.2, every rotation matrix $R_{wc} \in SO(3)$ represents a possible configuration of the object rotated about the point o. Besides this, R_{wc} takes another role as the matrix that represents the actual coordinates transformation from the frame C to the frame W. To see this, suppose that, for a given a point $p \in \mathbb{E}^3$, its coordinates with respect to the frame W are $\mathbf{X}_w = [X_{1w}, X_{2w}, X_{3w}]^T \in \mathbb{R}^3$. Since r_1, r_2, r_3 form a basis for $\mathbb{R}^3$, $\mathbf{X}_w$ can also be expressed as a linear combination of these three vectors, say $\mathbf{X}_w = X_{1c}r_1 + X_{2c}r_2 + X_{3c}r_3$ with $[X_{1c}, X_{2c}, X_{3c}]^T \in \mathbb{R}^3$. Obviously, $\mathbf{X}_c = [X_{1c}, X_{2c}, X_{3c}]^T$ are the coordinates of the same point p with respect to the frame C. Therefore, we have:

$$\mathbf{X}_w = X_{1c}r_1 + X_{2c}r_2 + X_{3c}r_3 = R_{wc}\mathbf{X}_c.$$

In this equation, the matrix R_{wc} transforms the coordinates $\mathbf{X}_c$ of a point p relative to the frame C to those $\mathbf{X}_w$ relative to the frame W. Since R_{wc} is a rotation matrix, its inverse is simply its transpose:

$$\mathbf{X}_c = R_{wc}^{-1}\mathbf{X}_w = R_{wc}^T\mathbf{X}_w.$$

That is, the inverse transformation is also a rotation; we call it R_{cw}, following an established convention, so that

$$R_{cw} = R_{wc}^{-1} = R_{wc}^T.$$

The configuration of the continuously rotating object can be then described as a trajectory $R(t) : t \mapsto SO(3)$ in the space $SO(3)$. For different times, the composition law of the rotation group then implies:

$$R(t_2, t_0) = R(t_2, t_1)R(t_1, t_0), \quad \forall t_0 < t_1 < t_2 \in \mathbb{R}.$$

Then, for a rotating camera, the world coordinates $\mathbf{X}_w$ of a fixed 3-D point p are transformed to its coordinates relative to the camera frame C by:

$$\mathbf{X}_c(t) = R_{cw}(t)\mathbf{X}_w.$$

Alternatively, if a point p fixed with respect to the camera frame with coordinates $\mathbf{X}_c$, its world coordinates $\mathbf{X}_w(t)$ as function of t are then given by:

$$\mathbf{X}_w(t) = R_{wc}(t)\mathbf{X}_c.$$

2.3.1 Canonical exponential coordinates

So far, we have shown that a rotational rigid body motion in $\mathbb{E}^3$ can be represented by a 3×3 rotation matrix $R \in SO(3)$. In the matrix representation that we have so far, each rotation matrix R is described and determined by its $3 \times 3 = 9$ entries. However, these 9 entries are not free parameters - they must satisfy the constraint $R^T R = I$. This actually imposes 6 independent constraints on the 9 entries. Hence the dimension of the rotation matrix space $SO(3)$ is only 3, and 6 parameters out of the 9 are in fact redundant. In this and the next section, we will introduce a few more representations (or parameterizations) for rotation matrix.

Given a curve $R(t) : \mathbb{R} \to SO(3)$ which describes a continuous rotational motion, the rotation must satisfy the following constraint:

$$R(t)R^T(t) = I.$$

Computing the derivative of the above equation with respect to time t, noticing that the right hand side is a constant matrix, we obtain:

$$\dot{R}(t)R^T(t) + R(t)\dot{R}^T(t) = 0 \quad \Rightarrow \quad \dot{R}(t)R^T(t) = -(\dot{R}(t)R^T(t))^T.$$

The resulting constraint which we obtain reflects the fact that the matrix $\dot{R}(t)R^T(t) \in \mathbb{R}^{3\times3}$ is a skew symmetric matrix (see Appendix **??**). Then, as we have seen, there must exist a vector, say $\omega(t) \in \mathbb{R}^3$ such that:

$$\widehat{\omega}(t) = \dot{R}(t)R^T(t).$$

Multiplying both sides by $R(t)$ to the right yields:

$$\dot{R}(t) = \widehat{\omega}(t)R(t). \tag{2.9}$$

Notice that, from the above equation, if $R(t_0) = I$ for $t = t_0$, we have $\dot{R}(t_0) = \widehat{\omega}(t_0)$. Hence, around the identity matrix I, a skew symmetric matrix gives a first-order approximation of rotation matrix:

$$R(t_0 + dt) \approx I + \widehat{\omega}(t_0)\ dt.$$

The space of all skew symmetric matrices is denoted as:

$$\boxed{so(3) = \{\widehat{\omega} \in \mathbb{R}^{3\times3} \mid \omega \in \mathbb{R}^3\}}$$

and thanks to the above observation it is also called the *tangent space* at the identity of the matrix group $SO(3)$[7]. If $R(t)$ is not at the identity, the tangent space at $R(t)$ is simply $so(3)$ transported to $R(t)$ by a multiplication of $R(t)$ to the right: $\dot{R}(t) = \widehat{\omega}(t)R(t)$. This also shows that elements of $SO(3)$ only depend upon three parameters.

Having understood its local approximation, we will now use this knowledge to obtain a representation for rotation matrices. Let us start by assuming that the matrix $\widehat{\omega}$ in (2.9) is constant:

$$\dot{R}(t) = \widehat{\omega}R(t). \tag{2.10}$$

In the above equation, $\widehat{\omega}$ can be interpreted as the *state transition matrix* for the following linear ordinary differential equation (ODE):

$$\dot{x}(t) = \widehat{\omega}x(t), \quad x(t) \in \mathbb{R}^3.$$

[7]Since $SO(3)$ is a Lie group, $so(3)$ is also called its Lie algebra.

It is then immediate to verify that the solution to the above ODE is given by:

$$x(t) = e^{\widehat{\omega} t} x(0) \tag{2.11}$$

where $e^{\widehat{\omega} t}$ is the matrix exponential:

$$e^{\widehat{\omega} t} = I + \widehat{\omega} t + \frac{(\widehat{\omega} t)^2}{2!} + \cdots + \frac{(\widehat{\omega} t)^n}{n!} + \cdots . \tag{2.12}$$

$e^{\widehat{\omega} t}$ is also denoted as $\exp(\widehat{\omega} t)$. Due to the uniqueness of the solution for the ODE (2.11), and assuming $R(0) = I$ as initial condition, we must have:

$$\boxed{R(t) = e^{\widehat{\omega} t}} \tag{2.13}$$

To confirm that the matrix $e^{\widehat{\omega} t}$ is indeed a rotation matrix, one can directly show from the definition of matrix exponential:

$$(e^{\widehat{\omega} t})^{-1} = e^{-\widehat{\omega} t} = e^{\widehat{\omega}^T t} = (e^{\widehat{\omega} t})^T.$$

Hence $(e^{\widehat{\omega} t})^T e^{\widehat{\omega} t} = I$. It remains to show that $\det(e^{\widehat{\omega} t}) = +1$ and we leave this fact to the reader as an exercise. A physical interpretation of the equation (2.13) is: if $\|\omega\| = 1$, then $R(t) = e^{\widehat{\omega} t}$ is simply a rotation around the axis $\omega \in \mathbb{R}^3$ by t radians. Therefore, the matrix exponential (2.12) indeed defines a map from the space $so(3)$ to $SO(3)$, the so-called *exponential map*:

$$\begin{aligned} \exp : so(3) &\rightarrow SO(3) \\ \widehat{\omega} \in so(3) &\mapsto e^{\widehat{\omega}} \in SO(3). \end{aligned}$$

Note that we obtained the expression (2.13) by assuming that the $\omega(t)$ in (2.9) is constant. This is however not always the case. So a question naturally arises here: Can every rotation matrix $R \in SO(3)$ be expressed in an exponential form as in (2.13)? The answer is yes and the fact is stated as the following theorem:

Theorem 2.1 (Surjectivity of the exponential map onto $SO(3)$). *For any $R \in SO(3)$, there exists a (not necessarily unique) $\omega \in \mathbb{R}^3$, $\|\omega\| = 1$ and $t \in \mathbb{R}$ such that $R = e^{\widehat{\omega} t}$.*

Proof. The proof of this theorem is by construction: if the rotation matrix R is given as:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix},$$

the corresponding t and ω are given by:

$$\boxed{t = \cos^{-1}\left(\frac{\text{trace}(R) - 1}{2}\right), \quad \omega = \frac{1}{2\sin(t)} \begin{bmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{bmatrix}}.$$

□

The significance of this theorem is that any rotation matrix can be realized by rotating around some fixed axis by a certain angle. However, the theorem only guarantees the surjectivity of the exponential map from $so(3)$ to $SO(3)$. Unfortunately, this map is not injective hence not one-to-one. This will become clear after we have introduced the so-called *Rodrigues' formula* for computing $R = e^{\widehat{\omega}t}$.

From the constructive proof for Theorem 2.1, we now know how to compute the exponential coordinates (ω, t) for a given rotation matrix $R \in SO(3)$. On the other hand, given (ω, t), how do we effectively compute the corresponding rotation matrix $R = e^{\widehat{\omega}t}$? One can certainly use the series (2.12) from the definition. The following theorem however simplifies the computation dramatically:

Theorem 2.2 (Rodrigues' formula for rotation matrix). *Given $\omega \in \mathbb{R}^3$ with $\|\omega\| = 1$ and $t \in \mathbb{R}$, the matrix exponential $R = e^{\widehat{\omega}t}$ is given by the following formula:*

$$\boxed{e^{\widehat{\omega}t} = I + \widehat{\omega}\sin(t) + \widehat{\omega}^2(1 - \cos(t))} \tag{2.14}$$

Proof. It is direct to verify that powers of $\widehat{\omega}$ can be reduced by the following two formulae:

$$\begin{aligned} \widehat{\omega}^2 &= \omega\omega^T - I, \\ \widehat{\omega}^3 &= -\widehat{\omega}. \end{aligned}$$

Hence the exponential series (2.12) can be simplified as:

$$e^{\widehat{\omega}t} = I + \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots\right)\widehat{\omega} + \left(\frac{t^2}{2!} - \frac{t^4}{4!} + \frac{t^6}{6!} - \cdots\right)\widehat{\omega}^2.$$

What in the brackets are exactly the series for $\sin(t)$ and $(1 - \cos(t))$. Hence we have $e^{\widehat{\omega}t} = I + \widehat{\omega}\sin(t) + \widehat{\omega}^2(1 - \cos(t))$. □

Using the Rodrigues' formula, it is immediate to see that if $t = 2k\pi, k \in \mathbb{Z}$, we have

$$e^{\widehat{\omega}2k\pi} = I$$

for all k. Hence, for a given rotation matrix $R \in SO(3)$ there are typically infinitely many exponential coordinates (ω, t) such that $e^{\widehat{\omega}t} = R$. The exponential map $\exp : so(3) \to SO(3)$ is therefore *not* one-to-one. It is also useful to know that the exponential map is not *commutative* either, i.e. for two $\widehat{\omega}_1, \widehat{\omega}_2 \in so(3)$, usually

$$e^{\widehat{\omega}_1}e^{\widehat{\omega}_2} \neq e^{\widehat{\omega}_2}e^{\widehat{\omega}_1} \neq e^{\widehat{\omega}_1 + \widehat{\omega}_2}$$

unless $\widehat{\omega}_1\widehat{\omega}_2 = \widehat{\omega}_2\widehat{\omega}_1$.

Remark 2.2. *In general, the difference between $\widehat{\omega}_1\widehat{\omega}_2$ and $\widehat{\omega}_2\widehat{\omega}_1$ is called the* Lie bracket *on $so(3)$, denoted as:*

$$[\widehat{\omega}_1, \widehat{\omega}_2] = \widehat{\omega}_1\widehat{\omega}_2 - \widehat{\omega}_2\widehat{\omega}_1, \quad \forall \widehat{\omega}_1, \widehat{\omega}_2 \in so(3).$$

Obviously, $[\widehat{\omega}_1, \widehat{\omega}_2]$ is also a skew symmetric matrix in $so(3)$. The linear structure of $so(3)$ together with the Lie bracket form the Lie algebra *of the (Lie) group $SO(3)$. For more details on the Lie group structure of $SO(3)$, the reader may refer to [21]. The set of all rotation matrices $e^{\widehat{\omega}t}, t \in \mathbb{R}$ is then called a* one parameter *subgroup of $SO(3)$ and the multiplication in such a subgroup is commutative, i.e. for the same $\omega \in \mathbb{R}^3$, we have:*

$$e^{\widehat{\omega}t_1}e^{\widehat{\omega}t_2} = e^{\widehat{\omega}t_2}e^{\widehat{\omega}t_1} = e^{\widehat{\omega}(t_1 + t_2)}, \quad \forall t_1, t_2 \in \mathbb{R}.$$

2.3.2 Quaternions and Lie-Cartan coordinates

Quaternions

We know that complex numbers $\mathbb{C}$ can be simply defined as $\mathbb{C} = \mathbb{R} + \mathbb{R}i$ with $i^2 = -1$. Quaternions are to generalize complex numbers in a similar fashion. The set of quaternions, denoted by $\mathbb{H}$, is defined as

$$\mathbb{H} = \mathbb{C} + \mathbb{C}j, \quad \text{with } j^2 = -1 \text{ and } i \cdot j = -j \cdot i. \tag{2.15}$$

So an element of $\mathbb{H}$ is of the form

$$q = q_0 + q_1 i + (q_2 + iq_3)j = q_0 + q_1 i + q_2 j + q_3 ij, \quad q_0, q_1, q_2, q_3 \in \mathbb{R}. \tag{2.16}$$

For simplicity of notation, in the literature ij is sometimes denoted as k. In general, the *multiplication* of any two quaternions is similar to the multiplication of two complex numbers, except that the multiplication of i and j is *anti-commutative*: $ij = -ji$. We can also similarly define the concept of *conjugation* for a quaternion

$$q = q_0 + q_1 i + q_2 j + q_3 ij \quad \Rightarrow \quad \bar{q} = q_0 - q_1 i - q_2 j - q_3 ij. \tag{2.17}$$

It is direct to check that

$$q\bar{q} = q_0^2 + q_1^2 + q_2^2 + q_3^2. \tag{2.18}$$

So $q\bar{q}$ is simply the square of the norm $\|q\|$ of q as a four dimensional vector in $\mathbb{R}^4$. For a non-zero $q \in \mathbb{H}$, i.e. $\|q\| \neq 0$, we can further define its *inverse* to be

$$q^{-1} = \bar{q}/\|q\|^2. \tag{2.19}$$

The multiplication and inverse rules defined above in fact endow the space $\mathbb{R}^4$ an algebraic structure of a *skew field*. $\mathbb{H}$ is in fact called a *Hamiltonian field*, besides another common name *quaternion field*.

One important usage of quaternion field $\mathbb{H}$ is that we can in fact embed the rotation group $SO(3)$ into it. To see this, let us focus on a special subgroup of $\mathbb{H}$, the so-called *unit quaternions*

$$\mathbb{S}^3 = \{q \in \mathbb{H} \mid \|q\|^2 = q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1\}. \tag{2.20}$$

It is obvious that the set of all unit quaternions is simply the unit sphere in $\mathbb{R}^4$. To show that $\mathbb{S}^3$ is indeed a group, we simply need to prove that it is closed under the multiplication and inverse of quaternions, i.e. the multiplication of two unit quaternions is still a unit quaternion and so is the inverse of a unit quaternion. We leave this simple fact as an exercise to the reader.

Given a rotation matrix $R = e^{\widehat{\omega} t}$ with $\omega = [\omega_1, \omega_2, \omega_3]^T \in \mathbb{R}^3$ and $t \in \mathbb{R}$, we can associate to it a unit quaternion as following

$$q(R) = \cos(t/2) + \sin(t/2)(\omega_1 i + \omega_2 j + \omega_3 ij) \in \mathbb{S}^3. \tag{2.21}$$

One may verify that this association preserves the group structure between $SO(3)$ and $\mathbb{S}^3$:

$$q(R^{-1}) = q^{-1}(R), \quad q(R_1 R_2) = q(R_1)q(R_2), \quad \forall R, R_1, R_2 \in SO(3). \tag{2.22}$$

Further study can show that this association is also *genuine*, i.e. for different rotation matrices, the associated unit quaternions are also different. In the opposite direction, given a unit quaternion

$q = q_0 + q_1 i + q_2 j + q_3 ij \in \mathbb{S}^3$, we can use the following formulae find the corresponding rotation matrix $R(q) = e^{\widehat{\omega} t}$

$$t = 2\arccos(q_0), \quad \omega_m = \begin{cases} q_m / \sin(t/2), & t \neq 0 \\ 0, & t = 0 \end{cases}, \quad m = 1, 2, 3. \tag{2.23}$$

However, one must notice that, according to the above formula, there are two unit quaternions correspond to the same rotation matrix: $R(q) = R(-q)$, as shown in Figure 2.3. Therefore,

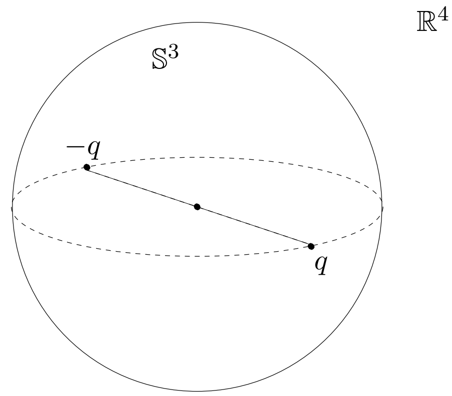

Figure 2.3: Antipodal unit quaternions q and $-q$ on the unit sphere $\mathbb{S}^3 \subset \mathbb{R}^4$ correspond to the same rotation matrix.

topologically, $\mathbb{S}^3$ is a double-covering of $SO(3)$. So $SO(3)$ is topologically the same as a three-dimensional projective plane $\mathbb{RP}^3$.

Compared to the exponential coordinates for rotation matrix that we studied in the previous section, using unit quaternions $\mathbb{S}^3$ to represent rotation matrices $SO(3)$, we have much less redundancy: there are only two unit quaternions correspond to the same rotation matrix while there are infinitely many for exponential coordinates. Furthermore, such a representation for rotation matrix is smooth and there is no singularity, as opposed to the *Lie-Cartan coordinates* representation which we will now introduce.

Lie-Cartan coordinates

Exponential coordinates and unit quaternions can both be viewed as ways to *globally* parameterize rotation matrices – the parameterization works for every rotation matrix practically the same way. On the other hand, the Lie-Cartan coordinates to be introduced below falls into the category of *local* parameterizations. That is, this kind of parameterizations are only good for a portion of $SO(3)$ but not for the entire space. The advantage of such local parameterizations is we usually need only three parameters to describe a rotation matrix, instead of four for both exponential coordinates: $(\omega, t) \in \mathbb{R}^4$ and unit quaternions: $q \in \mathbb{S}^3 \subset \mathbb{R}^4$.

In the space of skew symmetric matrices $so(3)$, pick a *basis* $(\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)$, i.e. the three vectors $\omega_1, \omega_2, \omega_3$ are linearly independent. Define a mapping (a parameterization) from $\mathbb{R}^3$ to $SO(3)$ as

$$\alpha : \ (\alpha_1, \alpha_2, \alpha_3) \quad \mapsto \quad \exp(\alpha \widehat{\omega}_1 + \alpha_2 \widehat{\omega}_2 + \alpha_3 \widehat{\omega}_3).$$

The coordinates $(\alpha_1, \alpha_2, \alpha_3)$ are called the *Lie-Cartan coordinates of the first kind* relative to the basis $(\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)$. Another way to parameterize the group $SO(3)$ using the same basis is to define another mapping from $\mathbb{R}^3$ to $SO(3)$ by

$$\beta : (\beta_1, \beta_2, \beta_3) \quad \mapsto \quad \exp(\beta_1 \widehat{\omega}_1) \exp(\beta_2 \widehat{\omega}_2) \exp(\beta_3 \widehat{\omega}_3).$$

The coordinates $(\beta_1, \beta_2, \beta_3)$ are then called the *Lie-Cartan coordinates of the second kind.*

In the special case when we choose $\omega_1, \omega_2, \omega_3$ to be the principal axes Z, Y, X, respectively, i.e.

$$\omega_1 = [0, 0, 1]^T \doteq \mathbf{z}, \quad \omega_2 = [0, 1, 0]^T \doteq \mathbf{y}, \quad \omega_3 = [1, 0, 0]^T \doteq \mathbf{x},$$

the Lie-Cartan coordinates of the second kind then coincide with the well-known ZYX *Euler angles* parametrization and $(\beta, \beta_2, \beta_3)$ are the corresponding Euler angles. The rotation matrix is then expressed by:

$$R(\beta_1, \beta_2, \beta_3) = \exp(\beta_1 \widehat{\mathbf{z}}) \exp(\beta_2 \widehat{\mathbf{y}}) \exp(\beta_3 \widehat{\mathbf{x}}). \tag{2.24}$$

Similarly we can define YZX Euler angles and ZYZ Euler angles. There are instances when this representation becomes singular and for certain rotation matrices, their corresponding Euler angles cannot be uniquelly determines. For example, the ZYX Euler angles become singular when $\beta_2 = -\pi/2$. The presence of such singularities is quite expected because of the topology of the space $SO(3)$. Globally $SO(3)$ is very much like a sphere in $\mathbb{R}^4$ as we have shown in the previous section, and it is well known that any attempt to find a global (three-dimensional) coordinate chart for it is doomed to fail.

2.4 Rigid body motion and its representations

In Section 2.3, we have studied extensively pure rotational rigid body motion and different representations for rotation matrix. In this section, we will study how to represent a rigid body motion in general - a motion with both rotation and translation.

Figure 2.4 illustrates a moving rigid object with a coordinate frame C attached to it. To describe

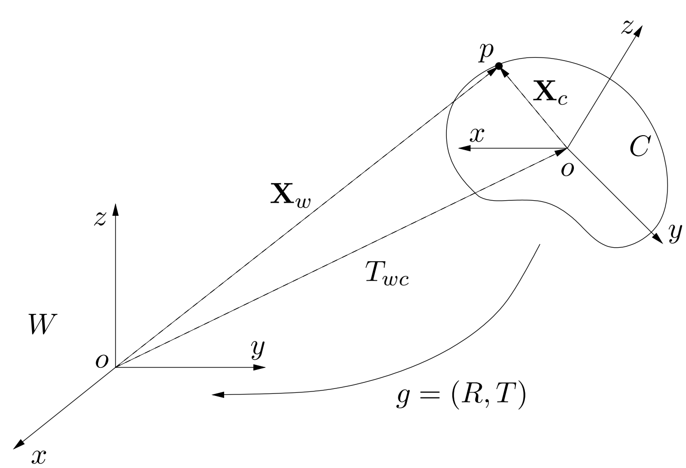

Figure 2.4: A rigid body motion between a moving frame C and a world frame W.

the coordinates of a point p on the object with respect to the world frame W, it is clear from the figure that the vector $\mathbf{X}_w$ is simply the sum of the translation $T_{wc} \in \mathbb{R}^3$ of the center of frame C relative to that of frame W and the vector $\mathbf{X}_c$ but expressed relative to frame W. Since $\mathbf{X}_c$ are the coordinates of the point p relative to the frame C, with respect to the world frame W, it becomes

$R_{wc}\mathbf{X}_c$ where $R_{wc} \in SO(3)$ is the relative rotation between the two frames. Hence the coordinates $\mathbf{X}_w$ are given by:

$$\mathbf{X}_w = R_{wc}\mathbf{X}_c + T_{wc}. \tag{2.25}$$

Usually, we denote the full rigid motion as $g_{wc} = (R_{wc}, T_{wc})$ or simply $g = (R, T)$ if the frames involved are clear from the context. Then g represents not only a description of the configuration of the rigid body object but a transformation of coordinates between the frames. In a compact form we may simply write:

$$\mathbf{X}_w = g_{wc}(\mathbf{X}_c).$$

The set of all possible configurations of rigid body can then be described as:

$$\boxed{SE(3) = \{g = (R, T) \mid R \in SO(3), T \in \mathbb{R}^3\} = SO(3) \times \mathbb{R}^3}$$

so called special Euclidean group $SE(3)$. Note that $g = (R, T)$ is not yet a matrix representation for the group $SE(3)$ as we defined in Section 2.2. To obtain such a representation, we must introduce the so-called homogeneous coordinates.

2.4.1 Homogeneous representation

One may have already noticed from equation (2.25) that, unlike the pure rotation case, the coordinate transformation for a full rigid body motion is not linear but *affine* instead.[8] Nonetheless, we may convert such an affine transformation to a linear one by using the so-called *homogeneous coordinates*: Appending 1 to the coordinates $\mathbf{X} = [X_1, X_2, X_3]^T \in \mathbb{R}^3$ of a point $p \in \mathbb{E}^3$ yields a vector in $\mathbb{R}^4$ denoted by $\bar{\mathbf{X}}$:

$$\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} \in \mathbb{R}^4.$$

Such an extension of coordinates, in effect, has embedded the Euclidean space $\mathbb{E}^3$ into a hyperplane in $\mathbb{R}^4$ instead of $\mathbb{R}^3$. Homogeneous coordinates of a vector $v = \mathbf{X}(q) - \mathbf{X}(p)$ are defined as the difference between homogeneous coordinates of the two points hence of the form:

$$\bar{v} = \begin{bmatrix} v \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}(q) \\ 1 \end{bmatrix} - \begin{bmatrix} \mathbf{X}(p) \\ 1 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ 0 \end{bmatrix} \in \mathbb{R}^4.$$

Notice that, in $\mathbb{R}^4$, vectors of the above form give rise to a subspace hence all linear structures of the original vectors $v \in \mathbb{R}^3$ are perfectly preserved by the new representation. Using the new notation, the transformation (2.25) can be re-written as:

$$\bar{\mathbf{X}}_w = \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix} = \begin{bmatrix} R_{wc} & T_{wc} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_c \\ 1 \end{bmatrix} =: \bar{g}_{wc}\bar{\mathbf{X}}_c$$

[8] We say two vectors u, v are related by a *linear* transformation if $u = Av$ for some matrix A, and by an *affine* transformation if $u = Av + b$ for some matrix A and vector b.

where the 4×4 matrix $\bar{g}_{wc} \in \mathbb{R}^{4\times 4}$ is called the *homogeneous representation* of the rigid motion $g_{wc} = (R_{wc}, T_{wc}) \in SE(3)$. In general, if $g = (R, T)$, then its homogeneous representation is:

$$\bar{g} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}. \tag{2.26}$$

Notice that, by introducing a little redundancy into the notation, we represent a rigid body transformation of coordinates by a linear matrix multiplication. The homogeneous representation of g in (2.26) gives rise to a natural matrix representation of the special Euclidean group $SE(3)$:

$$\boxed{SE(3) = \left\{ \bar{g} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \middle| R \in SO(3), T \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{4\times 4}}$$

It is then straightforward to verify that so-defined $SE(3)$ indeed satisfies all the requirements of a group. In particular, $\forall g_1, g_2$ and $g \in SE(3)$, we have:

$$\bar{g}_1\bar{g}_2 = \begin{bmatrix} R_1 & T_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_2 & T_2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_1R_2 & R_1T_2 + T_1 \\ 0 & 1 \end{bmatrix} \in SE(3)$$

and

$$\bar{g}^{-1} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} R^T & -R^TT \\ 0 & 1 \end{bmatrix} \in SE(3).$$

In homogeneous representation, the action of a rigid body transformation $g \in SE(3)$ on a vector $v = \mathbf{X}(q) - \mathbf{X}(p) \in \mathbb{R}^3$ becomes:

$$\bar{g}_*(\bar{v}) = \bar{g}\bar{\mathbf{X}}(q) - \bar{g}\bar{\mathbf{X}}(p) = \bar{g}\bar{v}.$$

That is, the action is simply represented by a matrix multiplication. The reader can verify that such an action preserves both inner product and cross product hence $\bar{g}$ indeed represents a rigid body transformation according to the definition we gave in Section 2.2.

2.4.2 Canonical exponential coordinates

In Section 2.3.1, we have studied exponential coordinates for rotation matrix $R \in SO(3)$. Similar coordination also exists for the homogeneous representation of a full rigid body motion $g \in SE(3)$. For the rest of this section, we demonstrate how to extend the results we have developed for rotational motion in Section 2.3.1 to a full rigid body motion. The results developed here will be extensively used throughout the entire book.

Consider that the motion of a continuously moving rigid body object is described by a curve from $\mathbb{R}$ to $SE(3)$: $g(t) = (R(t), T(t))$, or in homogeneous representation:

$$g(t) = \begin{bmatrix} R(t) & T(t) \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

Here, for simplicity of notation, we will remove the "bar" off from the symbol $\bar{g}$ for homogeneous representation and simply use g for the same matrix. We will use the same convention for point: $\mathbf{X}$ for $\bar{\mathbf{X}}$ and for vector: v for $\bar{v}$ whenever their correct dimension is clear from the context. Similar as in the pure rotation case, lets first look at the structure of the matrix $\dot{g}(t)g^{-1}(t)$:

$$\dot{g}(t)g^{-1}(t) = \begin{bmatrix} \dot{R}(t) & \dot{T}(t) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R^T(t) & -R^T(t)T(t) \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \dot{R}(t)R^T(t) & \dot{T}(t) - \dot{R}(t)R^T(t)T(t) \\ 0 & 0 \end{bmatrix}. \tag{2.27}$$

From our study of rotation matrix, we know $\dot{R}(t)R^T(t)$ is a skew symmetric matrix, i.e. there exists $\widehat{\omega}(t) \in so(3)$ such that $\widehat{\omega}(t) = \dot{R}(t)R^T(t)$. Define a vector $v(t) \in \mathbb{R}^3$ such that $v(t) = \dot{T}(t) - \widehat{\omega}(t)T(t)$. Then the above equation becomes:

$$\dot{g}(t)g^{-1}(t) = \begin{bmatrix} \widehat{\omega}(t) & v(t) \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

If we further define a matrix $\widehat{\xi} \in \mathbb{R}^{4\times 4}$ to be:

$$\widehat{\xi}(t) = \begin{bmatrix} \widehat{\omega}(t) & v(t) \\ 0 & 0 \end{bmatrix},$$

then we have:

$$\dot{g}(t) = (\dot{g}(t)g^{-1}(t))g(t) = \widehat{\xi}(t)g(t). \tag{2.28}$$

$\widehat{\xi}$ can be viewed as the "tangent vector" along the curve of $g(t)$ and used for approximate $g(t)$ locally:

$$g(t+dt) \approx g(t) + \widehat{\xi}(t)g(t)dt = \left(I + \widehat{\xi}(t)dt\right) g(t).$$

In robotics literature a 4×4 matrix of the form as $\widehat{\xi}$ is called a *twist*. The set of all twist is then denoted as:

$$\boxed{se(3) = \left\{ \widehat{\xi} = \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix} \middle| \, \widehat{\omega} \in so(3), v \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{4\times 4}}$$

$se(3)$ is called the tangent space (or Lie algebra) of the matrix group $SE(3)$. We also define two operators $\vee$ and $\wedge$ to convert between a twist $\widehat{\xi} \in se(3)$ and its *twist coordinates* $\xi \in \mathbb{R}^6$ as follows:

$$\begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix}^\vee \doteq \begin{bmatrix} v \\ \omega \end{bmatrix} \in \mathbb{R}^6, \qquad \begin{bmatrix} v \\ \omega \end{bmatrix}^\wedge \doteq \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

In the twist coordinates ξ, we will refer to v as the *linear velocity* and ω as the *angular velocity*, which indicates that they are related to either translational or rotational part of the full motion. Let us now consider a special case of the equation (2.28) when the twist $\widehat{\xi}$ is a constant matrix:

$$\dot{g}(t) = \widehat{\xi}g(t).$$

Hence we have again a time-invariant linear ordinary differential equation, which can be intergrated to give:

$$g(t) = e^{\widehat{\xi}t}g(0).$$

Assuming that the initial condition $g(0) = I$ we may conclude that:

$$\boxed{g(t) = e^{\widehat{\xi}t}}$$

where the twist exponential is:

$$e^{\widehat{\xi}t} = I + \widehat{\xi}t + \frac{(\widehat{\xi}t)^2}{2!} + \cdots + \frac{(\widehat{\xi}t)^n}{n!} + \cdots. \tag{2.29}$$

Using Rodrigues' formula introduced in the previous section, it is straightforward to obtain that:

$$\boxed{e^{\widehat{\xi}t} = \begin{bmatrix} e^{\widehat{\omega}t} & (I - e^{\widehat{\omega}t})\widehat{\omega}v + \omega\omega^T vt \\ 0 & 1 \end{bmatrix}} \tag{2.30}$$

It it clear from this expression that the exponential of $\widehat{\xi}t$ is indeed a rigid body transformation matrix in $SE(3)$. Therefore the exponential map defines a mapping from the space $se(3)$ to $SE(3)$:

$$\begin{aligned} \exp: \ se(3) &\rightarrow SE(3) \\ \widehat{\xi} \in se(3) &\mapsto e^{\widehat{\xi}} \in SE(3) \end{aligned}$$

and the twist $\widehat{\xi} \in se(3)$ is also called the *exponential coordinates* for $SE(3)$, as $\widehat{\omega} \in so(3)$ for $SO(3)$.

One question remains to answer: Can every rigid body motion $g \in SE(3)$ always be represented in such an exponential form? The answer is yes and is formulated in the following theorem:

Theorem 2.3 (Surjectivity of the exponential map onto $SE(3)$). *For any $g \in SE(3)$, there exist (not necessarily unique) twist coordinates $\xi = (v, \omega)$ and $t \in \mathbb{R}$ such that $g = e^{\widehat{\xi}t}$.*

Proof. The proof is constructive. Suppose $g = (R, T)$. For the rotation matrix $R \in SO(3)$ we can always find (ω, t) with $\|\omega\| = 1$ such that $e^{\widehat{\omega}t} = R$. If $t \neq 0$, from equation (2.30), we can solve for $v \in \mathbb{R}^3$ from the linear equation

$$(I - e^{\widehat{\omega}t})\widehat{\omega}v + \omega\omega^T vt = T \quad \Rightarrow \quad v = [(I - e^{\widehat{\omega}t})\widehat{\omega} + \omega\omega^T t]^{-1}T.$$

If $t = 0$, then $R = I$. We may simply choose $\omega = 0, v = T/\|T\|$ and $t = \|T\|$. □

Similar to the exponential coordinates for rotation matrix, the exponential map from $se(3)$ to $SE(3)$ is not injective hence not one-to-one. There are usually infinitely many exponential coordinates (or twists) that correspond to every $g \in SE(3)$. Similarly as in the pure rotation case, the linear structure of $se(3)$, together with the closure under the Lie bracket operation:

$$[\widehat{\xi}_1, \widehat{\xi}_2] = \widehat{\xi}_1\widehat{\xi}_2 - \widehat{\xi}_2\widehat{\xi}_1 = \begin{bmatrix} \widehat{\omega_1 \times \omega_2} & \omega_1 \times v_2 - \omega_2 \times v_1 \\ 0 & 0 \end{bmatrix} \in se(3).$$

makes $se(3)$ the Lie algebra for $SE(3)$. The two rigid body motions $g_1 = e^{\widehat{\xi}_1}$ and $g_2 = e^{\widehat{\xi}_2}$ commute with each other : $g_1 g_2 = g_2 g_1$, only if $[\widehat{\xi}_1, \widehat{\xi}_2] = 0$.

2.5 Coordinates and velocity transformation

In the above presentation of rigid body motion we described how 3-D points move relative to the camera frame. In computer vision we usually need to know how the coordinates of the points and their velocities change with respect to camera frames at different locations. This is mainly because that it is usually much more convenient and natural to choose the camera frame as the reference frame and to describe both camera motion and 3-D points relative to it. Since the camera may be moving, we need to know how to transform quantities such as coordinates and velocities from one camera frame to another. In particular, how to correctly express location and velocity of a (feature) point in terms of that of a moving camera. Here we introduce a few conventions that we will use for the rest of this book. The time $t \in \mathbb{R}$ will be always used as an index to register camera motion. Even in the discrete case when a few snapshots are given, we will order them by some time indexes as if th! ey had been taken in such order. We found that time is a good uniform index for both discrete case and continuous case, which will be treated in a unified way in this book. Therefore, we will use $g(t) = (R(t), T(t)) \in SE(3)$ or:

$$g(t) = \begin{bmatrix} R(t) & T(t) \\ 0 & 1 \end{bmatrix} \in SE(3)$$

to denote the relative displacement between some fixed world frame W and the camera frame C at time $t \in \mathbb{R}$. Here we will ignore the subscript cw from supposedly $g_{cw}(t)$ as long as the relativity is clear from the context. By default, we assume $g(0) = I$, i.e. at time $t = 0$ the camera frame coincides with the world frame. So if the coordinates of a point $p \in \mathbb{E}^3$ relative to the world frame are $\mathbf{X}_0 = \mathbf{X}(0)$, its coordinates relative to the camera at time t are then:

$$\boxed{\mathbf{X}(t) = R(t)\mathbf{X}_0 + T(t)} \tag{2.31}$$

or in homogeneous representation:

$$\boxed{\mathbf{X}(t) = g(t)\mathbf{X}_0}. \tag{2.32}$$

If the camera is at locations $g(t_1), \ldots, g(t_m)$ at time $t_1, \ldots, t_m$ respectively, then the coordinates of the same point p are given as $\mathbf{X}(t_i) = g(t_i)\mathbf{X}_0, i = 1, \ldots, m$ correspondingly. If it is only the position, not the time, that matters, we will very often use g_i as a shorthand for $g(t_i)$ and similarly $\mathbf{X}_i$ for $\mathbf{X}(t_i)$.

When the starting time is not $t = 0$, the relative motion between the camera at time t_2 relative to the camera at time t_1 will be denoted as $g(t_2, t_1) \in SE(3)$. Then we have the following relationship between coordinates of the same point p:

$$\mathbf{X}(t_2) = g(t_2, t_1)\mathbf{X}(t_1), \quad \forall t_2, t_1 \in \mathbb{R}.$$

Now consider a third position of the camera at $t = t_3 \in \mathbb{R}^3$, as shown in Figure 2.5. The relative motion between the camera at t_3 and t_2 is $g(t_3, t_2)$ and between t_3 and t_1 is $g(t_3, t_1)$. We then have

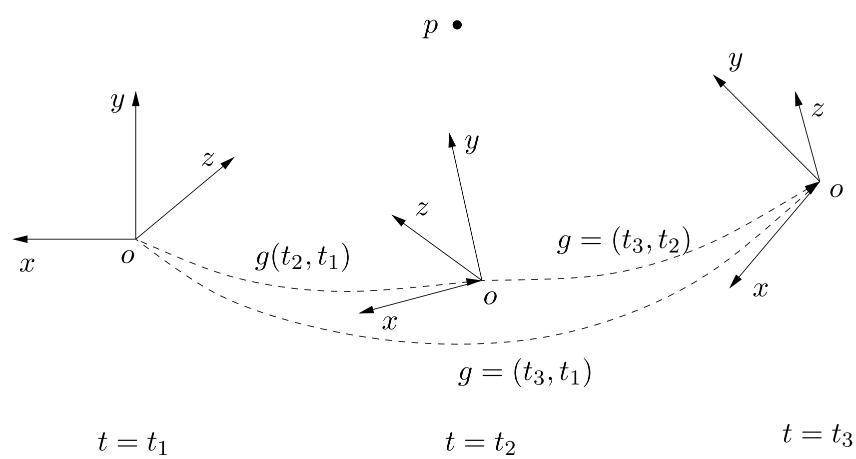

Figure 2.5: Composition of rigid body motions.

the following relationship among coordinates

$$\mathbf{X}(t_3) = g(t_3, t_2)\mathbf{X}(t_2) = g(t_3, t_2)g(t_2, t_1)\mathbf{X}(t_1).$$

Comparing with the direct relationship between coordinates at t_3 and t_1:

$$\mathbf{X}(t_3) = g(t_3, t_1)\mathbf{X}(t_1),$$

the following *composition rule* for consecutive motions must hold:

$$g(t_3, t_1) = g(t_3, t_2)g(t_2, t_1).$$

The composition rule describes the coordinates $\mathbf{X}$ of the point p relative to any camera position, if they are known with respect to a particular one. The same composition rule implies the *rule of inverse*

$$g^{-1}(t_2, t_1) = g(t_1, t_2)$$

since $g(t_2, t_1)g(t_1, t_2) = g(t_2, t_2) = I$. In case time is of no physical meaning to a particular problem, we might use g_{ij} as a shorthand for $g(t_i, t_j)$ with $t_i, t_j \in \mathbb{R}$.

Having understood transformation of coordinates, we now study what happens to velocity. We know that the coordinates $\mathbf{X}(t)$ of a point $p \in \mathbb{E}^3$ relative to a moving camera, are a function of time t:

$$\mathbf{X}(t) = g_{cw}(t)\mathbf{X}_0.$$

Then the velocity of the point p relative to the (instantaneous) camera frame is:

$$\dot{\mathbf{X}}(t) = \dot{g}_{cw}(t)\mathbf{X}_0.$$

In order express $\dot{\mathbf{X}}(t)$ in terms of quantities expressed in the moving frame we substitute for $\mathbf{X}_0 = g_{cw}^{-1}(t)\mathbf{X}_0$ and using the notion of twist define:

$$\widehat{V}_{cw}^c(t) = \dot{g}_{cw}(t)g_{cw}^{-1}(t) \in se(3) \tag{2.33}$$

where an expression for $\dot{g}_{cw}(t)g_{cw}^{-1}(t)$ can be found in (2.27). The above equation can be rewritten as:

$$\boxed{\dot{\mathbf{X}}(t) = \widehat{V}_{cw}^c(t)\mathbf{X}(t)} \tag{2.34}$$

Since $\widehat{V}_{cw}^c(t)$ is of the form:

$$\widehat{V}_{cw}^c(t) = \begin{bmatrix} \widehat{\omega}(t) & v(t) \\ 0 & 0 \end{bmatrix},$$

we can also write the velocity of the point in 3-D coordinates (instead of in the homogeneous coordinates) as:

$$\boxed{\dot{\mathbf{X}}(t) = \widehat{\omega}(t)\mathbf{X}(t) + v(t)}. \tag{2.35}$$

The physical interpretation of the symbol $\widehat{V}_{cw}^c$ is the velocity of the world frame moving relative to the camera frame, as viewed in the camera frame – the subscript and superscript of $\widehat{V}_{cw}^c$ indicate that. Usually, to clearly specify the physical meaning of a velocity, we need to specify: It is the velocity of which frame moving relative to which frame, and in which frame it is viewed. If we change where we view the velocity, the expression will change accordingly. For example suppose that a viewer is in another coordinate frame displaced relative to the camera frame by a rigid body transformation $g \in SE(3)$. Then the coordinates of the same point p relative to this frame are $\mathbf{Y}(t) = g\mathbf{X}(t)$. Compute the velocity in the new frame we have:

$$\dot{\mathbf{Y}}(t) = g\dot{g}_{cw}(t)g_{cw}^{-1}(t)g^{-1}\mathbf{Y}(t) = g\widehat{V}_{cw}^c g^{-1}\mathbf{Y}(t).$$

So the new velocity (or twist) is:

$$\widehat{V} = g\widehat{V}_{cw}^c g^{-1}.$$

This is the simply the same physical quantity but viewed from a different vantage point. We see that the two velocities are related through a mapping defined by the relative motion g, in particular:

$$\begin{array}{rccc} ad_g : & se(3) & \to & se(3) \\ & \widehat{\xi} & \mapsto & g\widehat{\xi}g^{-1}. \end{array}$$

This is the so-called *adjoint map* on the space $se(3)$. Using this notation in the previous example we have $\widehat{V} = ad_g(\widehat{V}^c_{cw})$. Clearly, the adjoint map transforms velocity from one frame to another. Using the fact that $g_{cw}(t)g_{wc}(t) = I$, it is straightforward to verify that:

$$\widehat{V}^c_{cw} = \dot{g}_{cw}g^{-1}_{cw} = -g^{-1}_{wc}\dot{g}_{wc} = -g_{cw}(\dot{g}_{wc}g^{-1}_{wc})g^{-1}_{cw} = ad_{g_{cw}}(-\widehat{V}^w_{wc}).$$

Hence $\widehat{V}^c_{cw}$ can also be interpreted as the *negated* velocity of the camera moving relative to the world frame, viewed in the (instantaneous) camera frame.

2.6 Summary

The rigid body motion introduced in this chapter is an element $g \in SE(3)$. The two most commonly used representation of elements of $g \in SE(3)$ are:

- Homogeneous representation:

$$\bar{g} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \text{ with } R \in SO(3) \text{ and } T \in \mathbb{R}^3.$$

- Twist representation:

$$g(t) = e^{\widehat{\xi}t} \text{ with the twist coordinates } \xi = (v, \omega) \in \mathbb{R}^6 \text{ and } t \in \mathbb{R}.$$

In the instantaneous case the velocity of a point with respect to the (instantaneous) camera frame is:

$$\dot{\mathbf{X}}(t) = \widehat{V}^c_{cw}(t)\mathbf{X}(t) \text{ where } \widehat{V}^c_{cw} = \dot{g}_{cw}g^{-1}_{cw}$$

and $g_{cw}(t)$ is the configuration of the camera with respect to the world frame. Using the actual 3D coordinates, the velocity of a 3D point yields the familiar relationship:

$$\dot{\mathbf{X}}(t) = \widehat{\omega}(t)\mathbf{X}(t) + v(t).$$

2.7 References

The presentation of the material in this chapter follows the development in [**?**]. More details on the abstract treatment of the material as well as further references can be also found there.

2.8 Exercises

1. **Linear vs. nonlinear maps**
   Suppose $A, B, C, X \in \mathbb{R}^{n\times n}$. Consider the following maps from $\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and determine if they are linear or not. Give a brief proof if true and a counterexample if false:

$$\begin{array}{llll} \text{(a)} & X & \mapsto & AX + XB \\ \text{(b)} & X & \mapsto & AX + BXC \\ \text{(c)} & X & \mapsto & AXA - B \\ \text{(d)} & X & \mapsto & AX + XBX \end{array}$$

   Note: A map $f : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto f(x)$ is called linear if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.

2. **Group structure of $SO(3)$**
   Prove that the space $SO(3)$ is a group, i.e. it satisfies all four axioms in the definition of group.

3. **Skew symmetric matrices**
   Given any vector $\omega = [\omega_1, \omega_2, \omega_3]^T \in \mathbb{R}^3$, define a 3×3 matrix associated to ω:

$$\widehat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \tag{2.36}$$

   According to the definition, $\widehat{\omega}$ is *skew symmetric*, i.e. $\widehat{\omega}^T = -\widehat{\omega}$. Now for any matrix $A \in \mathbb{R}^{3\times 3}$ with determinant $\det A = 1$, show that the following equation holds:

$$A^T \widehat{\omega} A = \widehat{A^{-1}\omega}. \tag{2.37}$$

   Then, in particular, if A is a rotation matrix, the above equation holds.
   Hint: Both $A^T\widehat{(\cdot)}A$ and $\widehat{A^{-1}(\cdot)}$ are linear maps with ω as the variable. What do you need to prove that two linear maps are the same?

4. **Rotation as rigid body motion**
   Given a rotation matrix $R \in SO(3)$, its action on a vector v is defined as Rv. Prove that any rotation matrix must preserve both the inner product and cross product of vectors. Hence, a rotation is indeed a rigid body motion.

5. **Range and null space**
   Recall that given a matrix $A \in \mathbb{R}^{m\times n}$, its *null space* is defined as a subspace of $\mathbb{R}^n$ consisting of all vectors $x \in \mathbb{R}^n$ such that $Ax = 0$. It is usually denoted as $Nu(A)$. The *range* of the matrix A is defined as a subspace of $\mathbb{R}^m$ consisting of all vectors $y \in \mathbb{R}^m$ such that there exists some $x \in \mathbb{R}^n$ such that $y = Ax$. It is denoted as $Ra(A)$. In mathematical terms,

$$Nu(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}, \quad Ra(A) = \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n, y = Ax\} \tag{2.38}$$

   (a) Recall that a set of vectors V is a subspace if for all vectors $x, y \in V$ and scalars $\alpha, \beta \in \mathbb{R}$, $\alpha x + \beta y$ is also a vector in V. Show that both $Nu(A)$ and $Ra(A)$ are indeed subspaces.

   (b) What are $Nu(\widehat{\omega})$ and $Ra(\widehat{\omega})$ for a non-zero vector $\omega \in \mathbb{R}^3$? Can you describe intuitively the geometric relationship between these two subspaces in $\mathbb{R}^3$? (A picture might help.)

6. **Properties of rotation matrices**
   Let $R \in SO(3)$ be a rotation matrix generated by rotating about a unit vector $\omega \in \mathbb{R}^3$ by θ radians. That is $R = e^{\widehat{\omega}\theta}$.

   (a) What are the eigenvalues and eigenvectors of $\widehat{\omega}$? You may use Matlab and try some examples first if you have little clue. If you happen to find a brutal-forth way to do it, can you instead use results in Exercise 3 to simplify the problem first?

   (b) Show that the eigenvalues of R are $1, e^{i\theta}, e^{-i\theta}$ where $i = \sqrt{-1}$ the imaginary unit. What is the eigenvector which corresponds to the eigenvalue 1? This actually gives another proof for $\det(e^{\widehat{\omega}\theta}) = 1 \cdot e^{i\theta} \cdot e^{-i\theta} = +1$ but not -1.

7. **Adjoint transformation on twist**
   Given a rigid body motion g and a twist $\widehat{\xi}$

$$g = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in SE(3), \quad \widehat{\xi} = \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix} \in se(3),$$

   show that $g\widehat{\xi}g^{-1}$ is still a twist. Notify what the corresponding ω and v terms have become in the new twist. The adjoint map is kind of a generalization of $R\widehat{\omega}R^T = \widehat{R\omega}$.