Chapter 4

The Configuration Space

Chapter 3 only covered how to model and transform a collection of bodies; however, for the purposes of planning it is important to define the state space. The state space for motion planning is a set of possible transformations that could be applied to the robot. This will be referred to as the *configuration space*, based on Lagrangian mechanics and the seminal work of Lozano-Pérez [656, 660, 657], who extensively utilized this notion in the context of planning (the idea was also used in early collision avoidance work by Udupa [947]). The motion planning literature was further unified around this concept by Latombe's book [588]. Once the configuration space is clearly understood, many motion planning problems that appear different in terms of geometry and kinematics can be solved by the same planning algorithms. This level of abstraction is therefore very important.

This chapter provides important foundational material that will be very useful in Chapters 5 to 8 and other places where planning over continuous state spaces occurs. Many concepts introduced in this chapter come directly from mathematics, particularly from topology. Therefore, Section 4.1 gives a basic overview of topological concepts. Section 4.2 uses the concepts from Chapter 3 to define the configuration space. After reading this, you should be able to precisely characterize the configuration space of a robot and understand its structure. In Section 4.3, obstacles in the world are transformed into obstacles in the configuration space, but it is important to understand that this transformation may not be explicitly constructed. The implicit representation of the state space is a recurring theme throughout planning. Section 4.4 covers the important case of kinematic chains that have loops, which was mentioned in Section 3.4. This case is so difficult that even the space of transformations usually cannot be explicitly characterized (i.e., parameterized).

4.1 Basic Topological Concepts

This section introduces basic topological concepts that are helpful in understanding configuration spaces. Topology is a challenging subject to understand in depth. The treatment given here provides only a brief overview and is designed to stim-

ulate further study (see the literature overview at the end of the chapter). To advance further in this chapter, it is not necessary to understand all of the material of this section; however, the more you understand, the deeper will be your understanding of motion planning in general.

4.1.1 Topological Spaces

Recall the concepts of open and closed intervals in the set of real numbers $\mathbb{R}$. The open interval $(0, 1)$ includes all real numbers between 0 and 1, *except* 0 and 1. However, for either endpoint, an infinite sequence may be defined that converges to it. For example, the sequence $1/2$, $1/4$, ..., $1/2^i$ converges to 0 as i tends to infinity. This means that we can choose a point in $(0, 1)$ within any small, positive distance from 0 or 1, but we cannot pick one exactly on the boundary of the interval. For a closed interval, such as $[0, 1]$, the boundary points are included.

The notion of an open set lies at the heart of topology. The open set definition that will appear here is a substantial generalization of the concept of an open interval. The concept applies to a very general collection of subsets of some larger space. It is general enough to easily include any kind of configuration space that may be encountered in planning.

A set X is called a *topological space* if there is a collection of subsets of X called *open sets* for which the following axioms hold:

1. The union of any number of open sets is an open set.

2. The intersection of a finite number of open sets is an open set.

3. Both X and $\emptyset$ are open sets.

Note that in the first axiom, the union of an infinite number of open sets may be taken, and the result must remain an open set. Intersecting an infinite number of open sets, however, does not necessarily lead to an open set.

For the special case of $X = \mathbb{R}$, the open sets include open intervals, as expected. Many sets that are not intervals are open sets because taking unions and intersections of open intervals yields other open sets. For example, the set

$$\bigcup_{i=1}^{\infty} \left(\frac{1}{3^i}, \frac{2}{3^i} \right), \tag{4.1}$$

which is an infinite union of pairwise-disjoint intervals, is an open set.

Closed sets   Open sets appear directly in the definition of a topological space. It next seems that closed sets are needed. Suppose X is a topological space. A subset $C \subset X$ is defined to be a *closed set* if and only if $X \setminus C$ is an open set. Thus, the complement of any open set is closed, and the complement of any closed set is open. Any closed interval, such as $[0, 1]$, is a closed set because its complement, $(-\infty, 0) \cup (1, \infty)$, is an open set. For another example, $(0, 1)$ is an open set;
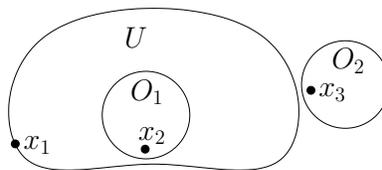
Figure 4.1: An illustration of the boundary definition. Suppose $X = \mathbb{R}^2$, and U is a subset as shown. Three kinds of points appear: 1) x_1 is a boundary point, 2) x_2 is an interior point, and 3) x_3 is an exterior point. Both x_1 and x_2 are limit points of U.

therefore, $\mathbb{R} \setminus (0,1) = (-\infty, 0] \cup [1, \infty)$ is a closed set. The use of "(" may seem wrong in the last expression, but "[" cannot be used because $-\infty$ and ∞ do not belong to $\mathbb{R}$. Thus, the use of "(" is just a notational quirk.

Are all subsets of X either closed or open? Although it appears that open sets and closed sets are opposites in some sense, the answer is *no*. For $X = \mathbb{R}$, the interval $[0, 2\pi)$ is neither open nor closed (consider its complement: $[2\pi, \infty)$ is closed, and $(-\infty, 0)$ is open). Note that for any topological space, X and $\emptyset$ are both open and closed!

Special points From the definitions and examples so far, it should seem that points on the "edge" or "border" of a set are important. There are several terms that capture where points are relative to the border. Let X be a topological space, and let U be any subset of X. Furthermore, let x be any point in X. The following terms capture the position of point x relative to U (see Figure 4.1):

- If there exists an open set O_1 such that $x \in O_1$ and $O_1 \subseteq U$, then x is called an *interior point* of U. The set of all interior points in U is called the *interior of U* and is denoted by int(U).

- If there exists an open set O_2 such that $x \in O_2$ and $O_2 \subseteq X \setminus U$, then x is called an *exterior point* with respect to U.

- If x is neither an interior point nor an exterior point, then it is called a *boundary point* of U. The set of all boundary points in X is called the *boundary of U* and is denoted by ∂U.

- All points in $x \in X$ must be one of the three above; however, another term is often used, even though it is redundant given the other three. If x is either an interior point or a boundary point, then it is called a *limit point* (or *accumulation point*) of U. The set of all limit points of U is a closed set called the *closure* of U, and it is denoted by cl(U). Note that cl$(U) = $ int$(U) \cup \partial U$.

For the case of $X = \mathbb{R}$, the boundary points are the endpoints of intervals. For example, 0 and 1 are boundary points of intervals, $(0, 1)$, $[0, 1]$, $[0, 1)$, and $(0, 1]$. Thus, U may or may not include its boundary points. All of the points in $(0, 1)$

are interior points, and all of the points in $[0, 1]$ are limit points. The motivation of the name "limit point" comes from the fact that such a point might be the limit of an infinite sequence of points in U. For example, 0 is the limit point of the sequence generated by $1/2^i$ for each $i \in \mathbb{N}$, the natural numbers.

There are several convenient consequences of the definitions. A closed set C contains the limit point of any sequence that is a subset of C. This implies that it contains all of its boundary points. The closure, cl, always results in a closed set because it adds all of the boundary points to the set. On the other hand, an open set contains none of its boundary points. These interpretations will come in handy when considering obstacles in the configuration space for motion planning.

Some examples   The definition of a topological space is so general that an incredible variety of topological spaces can be constructed.

Example 4.1 (The Topology of $\mathbb{R}^n$) We should expect that $X = \mathbb{R}^n$ for any integer n is a topological space. This requires characterizing the open sets. An *open ball* $B(x, \rho)$ is the set of points in the interior of a sphere of radius ρ, centered at x. Thus,

$$B(x, \rho) = \{x' \in \mathbb{R}^n \mid \|x' - x\| < \rho\}, \tag{4.2}$$

in which $\| \cdot \|$ denotes the Euclidean norm (or magnitude) of its argument. The open balls are open sets in $\mathbb{R}^n$. Furthermore, all other open sets can be expressed as a countable union of open balls.[1] For the case of $\mathbb{R}$, this reduces to representing any open set as a union of intervals, which was done so far.

Even though it is possible to express open sets of $\mathbb{R}^n$ as unions of balls, we prefer to use other representations, with the understanding that one could revert to open balls if necessary. The primitives of Section 3.1 can be used to generate many interesting open and closed sets. For example, any algebraic primitive expressed in the form $H = \{x \in \mathbb{R}^n \mid f(x) \leq 0\}$ produces a closed set. Taking finite unions and intersections of these primitives will produce more closed sets. Therefore, all of the models from Sections 3.1.1 and 3.1.2 produce an obstacle region $\mathcal{O}$ that is a closed set. As mentioned in Section 3.1.2, sets constructed only from primitives that use the $<$ relation are open. ∎

Example 4.2 (Subspace Topology) A new topological space can easily be constructed from a subset of a topological space. Let X be a topological space, and let $Y \subset X$ be a subset. The *subspace topology* on Y is obtained by defining the open sets to be every subset of Y that can be represented as $U \cap Y$ for some open set $U \subseteq X$. Thus, the open sets for Y are almost the same as for X, except that the points that do not lie in Y are trimmed away. New subspaces can be constructed by intersecting open sets of $\mathbb{R}^n$ with a complicated region defined by semi-algebraic models. This leads to many interesting topological spaces, some of

---

[1]Such a collection of balls is often referred to as a *basis*.

which will appear later in this chapter. ■

Example 4.3 (The Trivial Topology) For any set X, there is always one trivial example of a topological space that can be constructed from it. Declare that X and $\emptyset$ are the only open sets. Note that all of the axioms are satisfied. ■

Example 4.4 (A Strange Topology) It is important to keep in mind the almost absurd level of generality that is allowed by the definition of a topological space. A topological space can be defined for any set, as long as the declared open sets obey the axioms. Suppose a four-element set is defined as

$$X = \{\text{CAT}, \text{DOG}, \text{TREE}, \text{HOUSE}\}. \tag{4.3}$$

In addition to $\emptyset$ and X, suppose that $\{\text{CAT}\}$ and $\{\text{DOG}\}$ are open sets. Using the axioms, $\{\text{CAT}, \text{DOG}\}$ must also be an open set. Closed sets and boundary points can be derived for this topology once the open sets are defined. ■

After the last example, it seems that topological spaces are so general that not much can be said about them. Most spaces that are considered in topology and analysis satisfy more axioms. For $\mathbb{R}^n$ and any configuration spaces that arise in this book, the following is satisfied:

Hausdorff axiom: For any distinct $x_1, x_2 \in X$, there exist open sets O_1 and O_2 such that $x_1 \in O_1$, $x_2 \in O_2$, and $O_1 \cap O_2 = \emptyset$.

In other words, it is possible to separate x_1 and x_2 into nonoverlapping open sets. Think about how to do this for $\mathbb{R}^n$ by selecting small enough open balls. Any topological space X that satisfies the Hausdorff axiom is referred to as a *Hausdorff space*. Section 4.1.2 will introduce manifolds, which happen to be Hausdorff spaces and are general enough to capture the vast majority of configuration spaces that arise. We will have no need in this book to consider topological spaces that are not Hausdorff spaces.

Continuous functions A very simple definition of continuity exists for topological spaces. It nicely generalizes the definition from standard calculus. Let $f : X \to Y$ denote a function between topological spaces X and Y. For any set $B \subseteq Y$, let the *preimage* of B be denoted and defined by

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}. \tag{4.4}$$

Note that this definition does not require f to have an inverse.

The function f is called *continuous* if $f^{-1}(O)$ is an open set for every open set $O \subseteq Y$. Analysis is greatly simplified by this definition of continuity. For example, to show that any composition of continuous functions is continuous requires only a one-line argument that the preimage of the preimage of any open set always yields

an open set. Compare this to the cumbersome classical proof that requires a mess of δ's and ϵ's. The notion is also so general that continuous functions can even be defined on the absurd topological space from Example 4.4.

Homeomorphism: Making a donut into a coffee cup You might have heard the expression that to a topologist, a donut and a coffee cup appear the same. In many branches of mathematics, it is important to define when two basic objects are equivalent. In graph theory (and group theory), this equivalence relation is called an *isomorphism*. In topology, the most basic equivalence is a homeomorphism, which allows spaces that appear quite different in most other subjects to be declared equivalent in topology. The surfaces of a donut and a coffee cup (with one handle) are considered equivalent because both have a single hole. This notion needs to be made more precise!

Suppose $f : X \to Y$ is a bijective (one-to-one and onto) function between topological spaces X and Y. Since f is bijective, the inverse f^{-1} exists. If both f and f^{-1} are continuous, then f is called a *homeomorphism*. Two topological spaces X and Y are said to be *homeomorphic*, denoted by $X \cong Y$, if there exists a homeomorphism between them. This implies an equivalence relation on the set of topological spaces (verify that the reflexive, symmetric, and transitive properties are implied by the homeomorphism).

Example 4.5 (Interval Homeomorphisms) Any open interval of $\mathbb{R}$ is homeomorphic to any other open interval. For example, $(0, 1)$ can be mapped to $(0, 5)$ by the continuous mapping $x \mapsto 5x$. Note that $(0, 1)$ and $(0, 5)$ are each being interpreted here as topological subspaces of $\mathbb{R}$. This kind of homeomorphism can be generalized substantially using linear algebra. If a subset, $X \subset \mathbb{R}^n$, can be mapped to another, $Y \subset \mathbb{R}^n$, via a nonsingular linear transformation, then X and Y are homeomorphic. For example, the rigid-body transformations of the previous chapter were examples of homeomorphisms applied to the robot. Thus, the topology of the robot does not change when it is translated or rotated. (In this example, note that the robot itself is the topological space. This will not be the case for the rest of the chapter.)

Be careful when mixing closed and open sets. The space $[0, 1]$ is not homeomorphic to $(0, 1)$, and neither is homeomorphic to $[0, 1)$. The endpoints cause trouble when trying to make a bijective, continuous function. Surprisingly, a bounded and unbounded set may be homeomorphic. A subset X of $\mathbb{R}^n$ is called *bounded* if there exists a ball $B \subset \mathbb{R}^n$ such that $X \subset B$. The mapping $x \mapsto 1/x$ establishes that $(0, 1)$ and $(1, \infty)$ are homeomorphic. The mapping $x \mapsto 2\tan^{-1}(x)/\pi$ establishes that $(-1, 1)$ and all of $\mathbb{R}$ are homeomorphic!   ∎

Example 4.6 (Topological Graphs) Let X be a topological space. The previous example can be extended nicely to make homeomorphisms look like graph
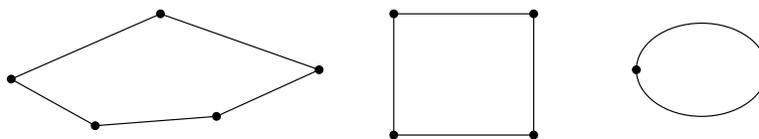
Figure 4.2: Even though the graphs are not isomorphic, the corresponding topological spaces may be homeomorphic due to useless vertices. The example graphs map into $\mathbb{R}^2$, and are all homeomorphic to a circle.
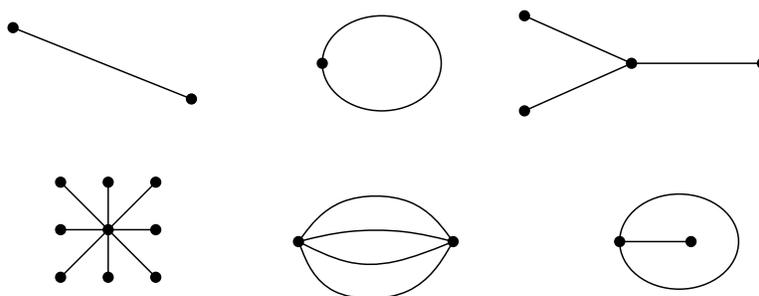
Figure 4.3: These topological graphs map into subsets of $\mathbb{R}^2$ that are not homeomorphic to each other.

isomorphisms. Let a *topological graph*[2] be a graph for which every vertex corresponds to a point in X and every edge corresponds to a continuous, injective (one-to-one) function, $\tau : [0, 1] \rightarrow X$. The image of τ connects the points in X that correspond to the endpoints (vertices) of the edge. The images of different edge functions are not allowed to intersect, except at vertices. Recall from graph theory that two graphs, $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$, are called *isomorphic* if there exists a bijective mapping, $f : V_1 \rightarrow V_2$ such that there is an edge between v_1 and v_1' in G_1, if and only if there exists an edge between $f(v_1)$ and $f(v_1')$ in G_2.

The bijective mapping used in the graph isomorphism can be extended to produce a homeomorphism. Each edge in E_1 is mapped continuously to its corresponding edge in E_2. The mappings nicely coincide at the vertices. Now you should see that two topological graphs are homeomorphic if they are isomorphic under the standard definition from graph theory.[3] What if the graphs are not isomorphic? There is still a chance that the topological graphs may be homeomorphic, as shown in Figure 4.2. The problem is that there appear to be "useless" vertices in the graph. By removing vertices of degree two that can be deleted without affecting the connectivity of the graph, the problem is fixed. In this case,

---

[2]In topology this is called a 1-complex [439].

[3]Technically, the images of the topological graphs, as subspaces of X, are homeomorphic, not the graphs themselves.

graphs that are not isomorphic produce topological graphs that are not homeomorphic. This allows many distinct, interesting topological spaces to be constructed. A few are shown in Figure 4.3.   ∎

4.1.2 Manifolds

In motion planning, efforts are made to ensure that the resulting configuration space has nice properties that reflect the true structure of the space of transformations. One important kind of topological space, which is general enough to include most of the configuration spaces considered in Part II, is called a manifold. Intuitively, a manifold can be considered as a "nice" topological space that behaves at every point like our intuitive notion of a surface.

Manifold definition   A topological space $M \subseteq \mathbb{R}^m$ is a *manifold*[4] if for every $x \in M$, an open set $O \subset M$ exists such that: 1) $x \in O$, 2) O is homeomorphic to $\mathbb{R}^n$, and 3) n is fixed for all $x \in M$. The fixed n is referred to as the *dimension* of the manifold, M. The second condition is the most important. It states that in the vicinity of any point, $x \in M$, the space behaves just like it would in the vicinity of any point $y \in \mathbb{R}^n$; intuitively, the set of directions that one can move appears the same in either case. Several simple examples that may or may not be manifolds are shown in Figure 4.4.

One natural consequence of the definitions is that $m \geq n$. According to Whitney's embedding theorem [449], $m \leq 2n+1$. In other words, $\mathbb{R}^{2n+1}$ is "big enough" to hold any n-dimensional manifold.[5] Technically, it is said that the n-dimensional manifold M is *embedded* in $\mathbb{R}^m$, which means that an injective mapping exists from M to $\mathbb{R}^m$ (if it is not injective, then the topology of M could change).

As it stands, it is impossible for a manifold to include its boundary points because they are not contained in open sets. A *manifold with boundary* can be defined requiring that the neighborhood of each boundary point of M is homeomorphic to a half-space of dimension n (which was defined for $n = 2$ and $n = 3$ in Section 3.1) and that the interior points must be homeomorphic to $\mathbb{R}^n$.

The presentation now turns to ways of constructing some manifolds that frequently appear in motion planning. It is important to keep in mind that two

---

[4]Manifolds that are not subsets of $\mathbb{R}^m$ may also be defined. This requires that M is a Hausdorff space and is second countable, which means that there is a countable number of open sets from which any other open set can be constructed by taking a union of some of them. These conditions are automatically satisfied when assuming $M \subseteq \mathbb{R}^m$; thus, it avoids these extra complications and is still general enough for our purposes. Some authors use the term *manifold* to refer to a *smooth manifold*. This requires the definition of a smooth structure, and the homeomorphism is replaced by diffeomorphism. This extra structure is not needed here but will be introduced when it is needed in Section 8.3.

[5]One variant of the theorem is that for smooth manifolds, $\mathbb{R}^{2n}$ is sufficient. This bound is tight because $\mathbb{RP}^n$ (n-dimensional projective space, which will be introduced later in this section), cannot be embedded in $\mathbb{R}^{2n-1}$.
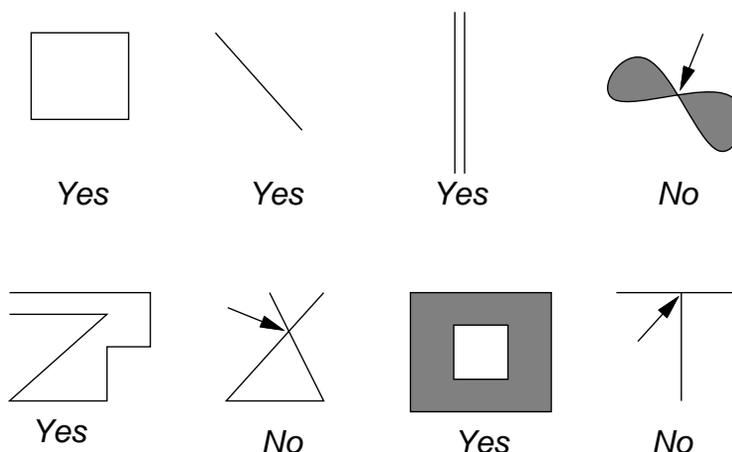
Figure 4.4: Some subsets of $\mathbb{R}^2$ that may or may not be manifolds. For the three that are not, the point that prevents them from being manifolds is indicated.

manifolds will be considered equivalent if they are homeomorphic (recall the donut and coffee cup).

Cartesian products There is a convenient way to construct new topological spaces from existing ones. Suppose that X and Y are topological spaces. The *Cartesian product*, $X \times Y$, defines a new topological space as follows. Every $x \in X$ and $y \in Y$ generates a point (x, y) in $X \times Y$. Each open set in $X \times Y$ is formed by taking the Cartesian product of one open set from X and one from Y. Exactly one open set exists in $X \times Y$ for every pair of open sets that can be formed by taking one from X and one from Y. Furthermore, these new open sets are used as a basis for forming the remaining open sets of $X \times Y$ by allowing any unions and finite intersections of them.

A familiar example of a Cartesian product is $\mathbb{R} \times \mathbb{R}$, which is equivalent to $\mathbb{R}^2$. In general, $\mathbb{R}^n$ is equivalent to $\mathbb{R} \times \mathbb{R}^{n-1}$. The Cartesian product can be taken over many spaces at once. For example, $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^n$. In the coming text, many important manifolds will be constructed via Cartesian products.

1D manifolds The set $\mathbb{R}$ of reals is the most obvious example of a 1D manifold because $\mathbb{R}$ certainly looks like (via homeomorphism) $\mathbb{R}$ in the vicinity of every point. The range can be restricted to the unit interval to yield the manifold $(0, 1)$ because they are homeomorphic (recall Example 4.5).

Another 1D manifold, which is not homeomorphic to $(0, 1)$, is a circle, $\mathbb{S}^1$. In this case $\mathbb{R}^m = \mathbb{R}^2$, and let

$$\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}. \tag{4.5}$$

If you are thinking like a topologist, it should appear that this particular circle is not important because there are numerous ways to define manifolds that are

homeomorphic to $\mathbb{S}^1$. For any manifold that is homeomorphic to $\mathbb{S}^1$, we will sometimes say that the manifold *is* $\mathbb{S}^1$, just represented in a different way. Also, $\mathbb{S}^1$ will be called a *circle*, but this is meant only in the topological sense; it only needs to be homeomorphic to the circle that we learned about in high school geometry. Also, when referring to $\mathbb{R}$, we might instead substitute $(0, 1)$ without any trouble. The alternative representations of a manifold can be considered as changing *parameterizations*, which are formally introduced in Section 8.3.2.

Identifications   A convenient way to represent $\mathbb{S}^1$ is obtained by *identification*, which is a general method of declaring that some points of a space are identical, even though they originally were distinct.[6] For a topological space X, let $X/\sim$ denote that X has been redefined through some form of identification. The open sets of X become redefined. Using identification, $\mathbb{S}^1$ can be defined as $[0, 1]/\sim$, in which the identification declares that 0 and 1 are equivalent, denoted as $0 \sim 1$. This has the effect of "gluing" the ends of the interval together, forming a closed loop. To see the homeomorphism that makes this possible, use polar coordinates to obtain $\theta \mapsto (\cos 2\pi\theta, \sin 2\pi\theta)$. You should already be familiar with 0 and 2π leading to the same point in polar coordinates; here they are just normalized to 0 and 1. Letting θ run from 0 up to 1, and then "wrapping around" to 0 is a convenient way to represent $\mathbb{S}^1$ because it does not need to be curved as in (4.5).

It might appear that identifications are cheating because the definition of a manifold requires it to be a subset of $\mathbb{R}^m$. This is not a problem because Whitney's theorem, as mentioned previously, states that any n-dimensional manifold can be embedded in $\mathbb{R}^{2n+1}$. The identifications just reduce the number of dimensions needed for visualization. They are also convenient in the implementation of motion planning algorithms.

2D manifolds   Many important, 2D manifolds can be defined by applying the Cartesian product to 1D manifolds. The 2D manifold $\mathbb{R}^2$ is formed by $\mathbb{R} \times \mathbb{R}$. The product $\mathbb{R} \times \mathbb{S}^1$ defines a manifold that is equivalent to an infinite cylinder. The product $\mathbb{S}^1 \times \mathbb{S}^1$ is a manifold that is equivalent to a torus (the surface of a donut).

Can any other 2D manifolds be defined? See Figure 4.5. The identification idea can be applied to generate several new manifolds. Start with an open square $M = (0, 1) \times (0, 1)$, which is homeomorphic to $\mathbb{R}^2$. Let (x, y) denote a point in the plane. A *flat cylinder* is obtained by making the identification $(0, y) \sim (1, y)$ for all $y \in (0, 1)$ and adding all of these points to M. The result is depicted in Figure 4.5 by drawing arrows where the identification occurs.

A *Möbius band* can be constructed by taking a strip of paper and connecting the ends after making a 180-degree twist. This result is not homeomorphic to the cylinder. The Möbius band can also be constructed by putting the twist into the identification, as $(0, y) \sim (1, 1 - y)$ for all $y \in (0, 1)$. In this case, the arrows are drawn in opposite directions. The Möbius band has the famous properties that

---

[6]This is usually defined more formally and called a *quotient topology*.

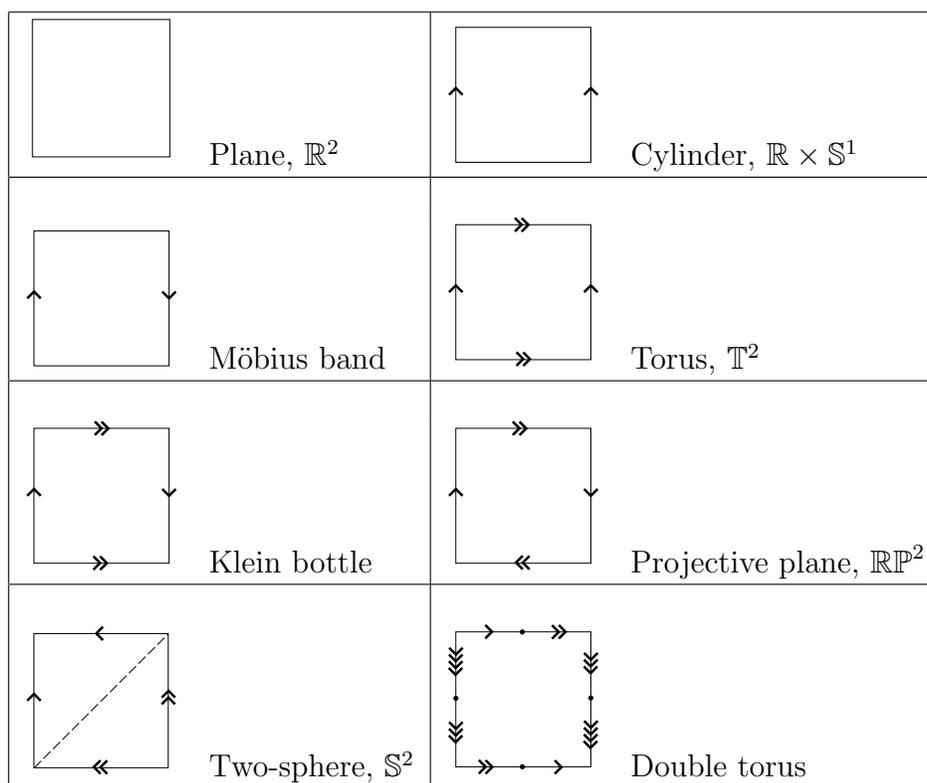|	
Plane, $\mathbb{R}^2$	Cylinder, $\mathbb{R} \times \mathbb{S}^1$
Möbius band	Torus, $\mathbb{T}^2$
Klein bottle	Projective plane, $\mathbb{RP}^2$
Two-sphere, $\mathbb{S}^2$	Double torus

Figure 4.5: Some 2D manifolds that can be obtained by identifying pairs of points along the boundary of a square region.

it has only one side (trace along the paper strip with a pencil, and you will visit both sides of the paper) and is nonorientable (if you try to draw it in the plane, without using identification tricks, it will always have a twist).

For all of the cases so far, there has been a boundary to the set. The next few manifolds will not even have a boundary, even though they may be bounded. If you were to live in one of them, it means that you could walk forever along any trajectory and never encounter the edge of your universe. It might seem like our physical universe is unbounded, but it would only be an illusion. Furthermore, there are several distinct possibilities for the universe that are not homeomorphic to each other. In higher dimensions, such possibilities are the subject of cosmology, which is a branch of astrophysics that uses topology to characterize the structure of our universe.

A *torus* can be constructed by performing identifications of the form $(0, y) \sim (1, y)$, which was done for the cylinder, and also $(x, 0) \sim (x, 1)$, which identifies the top and bottom. Note that the point $(0, 0)$ must be included and is identified with three other points. Double arrows are used in Figure 4.5 to indicate the top and bottom identification. All of the identification points must be added to M. Note that there are no twists. A funny interpretation of the resulting *flat torus* is as the universe appears for a spacecraft in some 1980s-style *Asteroids*-like video games.

The spaceship flies off of the screen in one direction and appears somewhere else, as prescribed by the identification.

Two interesting manifolds can be made by adding twists. Consider performing all of the identifications that were made for the torus, except put a twist in the side identification, as was done for the Möbius band. This yields a fascinating manifold called the *Klein bottle*, which can be embedded in $\mathbb{R}^4$ as a closed 2D surface in which the inside and the outside are the same! (This is in a sense similar to that of the Möbius band.) Now suppose there are twists in both the sides and the top and bottom. This results in the most bizarre manifold yet: the real projective plane, $\mathbb{RP}^2$. This space is equivalent to the set of all lines in $\mathbb{R}^3$ that pass through the origin. The 3D version, $\mathbb{RP}^3$, happens to be one of the most important manifolds for motion planning!

Let $\mathbb{S}^2$ denote the unit sphere, which is defined as

$$\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}. \tag{4.6}$$

Another way to represent $\mathbb{S}^2$ is by making the identifications shown in the last row of Figure 4.5. A dashed line is indicated where the equator might appear, if we wanted to make a distorted wall map of the earth. The poles would be at the upper left and lower right corners. The final example shown in Figure 4.5 is a *double torus*, which is the surface of a two-holed donut.

Higher dimensional manifolds   The construction techniques used for the 2D manifolds generalize nicely to higher dimensions. Of course, $\mathbb{R}^n$, is an n-dimensional manifold. An n-dimensional torus, $\mathbb{T}^n$, can be made by taking a Cartesian product of n copies of $\mathbb{S}^1$. Note that $\mathbb{S}^1 \times \mathbb{S}^1 \neq \mathbb{S}^2$. Therefore, the notation $\mathbb{T}^n$ is used for $(\mathbb{S}^1)^n$. Different kinds of n-dimensional cylinders can be made by forming a Cartesian product $\mathbb{R}^i \times \mathbb{T}^j$ for positive integers i and j such that $i + j = n$. Higher dimensional spheres are defined as

$$\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}, \tag{4.7}$$

in which $\|x\|$ denotes the Euclidean norm of x, and n is a positive integer. Many interesting spaces can be made by identifying faces of the cube $(0, 1)^n$ (or even faces of a polyhedron or polytope), especially if different kinds of twists are allowed. An n-dimensional projective space can be defined in this way, for example. *Lens spaces* are a family of manifolds that can be constructed by identification of polyhedral faces [834].

Due to its coming importance in motion planning, more details are given on projective spaces. The standard definition of an *n-dimensional real projective space* $\mathbb{RP}^n$ is the set of all lines in $\mathbb{R}^{n+1}$ that pass through the origin. Each line is considered as a point in $\mathbb{RP}^n$. Using the definition of $\mathbb{S}^n$ in (4.7), note that each of these lines in $\mathbb{R}^{n+1}$ intersects $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ in exactly two places. These intersection points are called *antipodal*, which means that they are as far from each other as possible on $\mathbb{S}^n$. The pair is also unique for each line. If we identify

all pairs of antipodal points of $\mathbb{S}^n$, a homeomorphism can be defined between each line through the origin of $\mathbb{R}^{n+1}$ and each antipodal pair on the sphere. This means that the resulting manifold, $\mathbb{S}^n / \sim$, is homeomorphic to $\mathbb{RP}^n$.

Another way to interpret the identification is that $\mathbb{RP}^n$ is just the upper half of $\mathbb{S}^n$, but with every equatorial point identified with its antipodal point. Thus, if you try to walk into the southern hemisphere, you will find yourself on the other side of the world walking north. It is helpful to visualize the special case of $\mathbb{RP}^2$ and the upper half of $\mathbb{S}^2$. Imagine warping the picture of $\mathbb{RP}^2$ from Figure 4.5 from a square into a circular disc, with opposite points identified. The result still represents $\mathbb{RP}^2$. The center of the disc can now be lifted out of the plane to form the upper half of $\mathbb{S}^2$.

4.1.3 Paths and Connectivity

Central to motion planning is determining whether one part of a space is reachable from another. In Chapter 2, one state was reached from another by applying a sequence of actions. For motion planning, the analog to this is connecting one point in the configuration space to another by a continuous path. Graph connectivity is important in the discrete planning case. An analog to this for topological spaces is presented in this section.

Paths   Let X be a topological space, which for our purposes will also be a manifold. A *path* is a continuous function, $\tau : [0, 1] \to X$. Alternatively, $\mathbb{R}$ may be used for the domain of τ. Keep in mind that a path is a function, not a set of points. Each point along the path is given by $\tau(s)$ for some $s \in [0, 1]$. This makes it appear as a nice generalization to the sequence of states visited when a plan from Chapter 2 is applied. Recall that there, a countable set of stages was defined, and the states visited could be represented as x_1, x_2, ... In the current setting $\tau(s)$ is used, in which s replaces the stage index. To make the connection clearer, we could use x instead of τ to obtain $x(s)$ for each $s \in [0, 1]$.

Connected vs. path connected   A topological space X is said to be *connected* if it cannot be represented as the union of two disjoint, nonempty, open sets. While this definition is rather elegant and general, if X is connected, it does not imply that a path exists between any pair of points in X thanks to crazy examples like the *topologist's sine curve*:

$$X = \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } y = \sin(1/x)\}. \tag{4.8}$$

Consider plotting X. The $\sin(1/x)$ part creates oscillations near the y-axis in which the frequency tends to infinity. After union is taken with the y-axis, this space is connected, but there is no path that reaches the y-axis from the sine curve.

How can we avoid such problems? The standard way to fix this is to use the path definition directly in the definition of connectedness. A topological space X is said to be *path connected* if for all $x_1, x_2 \in X$, there exists a path τ such that
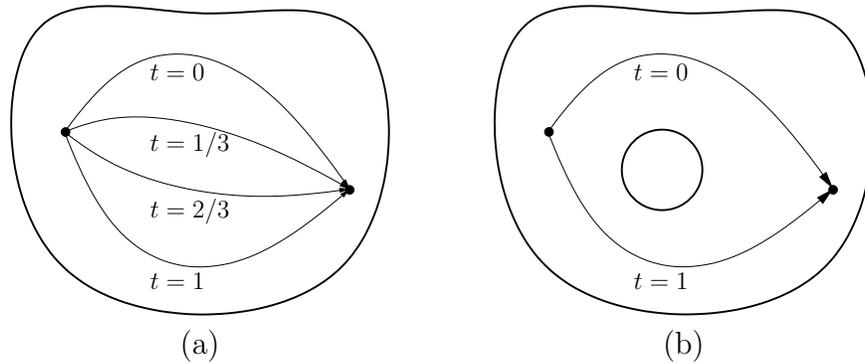
Figure 4.6: (a) Homotopy continuously warps one path into another. (b) The image of the path cannot be continuously warped over a hole in $\mathbb{R}^2$ because it causes a discontinuity. In this case, the two paths are not homotopic.

$\tau(0) = x_1$ and $\tau(1) = x_2$. It can be shown that if X is path connected, then it is also connected in the sense defined previously.

Another way to fix it is to make restrictions on the kinds of topological spaces that will be considered. This approach will be taken here by assuming that all topological spaces are manifolds. In this case, no strange things like (4.8) can happen,[7] and the definitions of connected and path connected coincide [451]. Therefore, we will just say a space is *connected*. However, it is important to remember that this definition of connected is sometimes inadequate, and one should really say that X is *path connected*.

Simply connected Now that the notion of connectedness has been established, the next step is to express different kinds of connectivity. This may be done by using the notion of homotopy, which can intuitively be considered as a way to continuously "warp" or "morph" one path into another, as depicted in Figure 4.6a.

Two paths τ_1 and τ_2 are called *homotopic* (with endpoints fixed) if there exists a continuous function $h : [0,1] \times [0,1] \to X$ for which the following four conditions are met:

1. (**Start with first path**) $h(s, 0) = \tau_1(s)$ for all $s \in [0, 1]$.

2. (**End with second path**) $h(s, 1) = \tau_2(s)$ for all $s \in [0, 1]$.

3. (**Hold starting point fixed**) $h(0, t) = h(0, 0)$ for all $t \in [0, 1]$.

4. (**Hold ending point fixed**) $h(1, t) = h(1, 0)$ for all $t \in [0, 1]$.

---

[7]The topologist's sine curve is not a manifold because all open sets that contain the point $(0, 0)$ contain some of the points from the sine curve. These open sets are not homeomorphic to $\mathbb{R}$.

The parameter t can be interpreted as a knob that is turned to gradually deform the path from τ_1 into τ_2. The first two conditions indicate that $t = 0$ yields τ_1 and $t = 1$ yields τ_2, respectively. The remaining two conditions indicate that the path endpoints are held fixed.

During the warping process, the path image cannot make a discontinuous jump. In $\mathbb{R}^2$, this prevents it from moving over holes, such as the one shown in Figure 4.6b. The key to preventing homotopy from jumping over some holes is that h must be continuous. In higher dimensions, however, there are many different kinds of holes. For the case of $\mathbb{R}^3$, for example, suppose the space is like a block of Swiss cheese that contains air bubbles. Homotopy can go around the air bubbles, but it cannot pass through a hole that is drilled through the entire block of cheese. Air bubbles and other kinds of holes that appear in higher dimensions can be characterized by generalizing homotopy to the warping of higher dimensional surfaces, as opposed to paths [439].

It is straightforward to show that homotopy defines an equivalence relation on the set of all paths from some $x_1 \in X$ to some $x_2 \in X$. The resulting notion of "equivalent paths" appears frequently in motion planning, control theory, and many other contexts. Suppose that X is path connected. If all paths fall into the same equivalence class, then X is called *simply connected*; otherwise, X is called *multiply connected*.

Groups The equivalence relation induced by homotopy starts to enter the realm of algebraic topology, which is a branch of mathematics that characterizes the structure of topological spaces in terms of algebraic objects, such as groups. These resulting groups have important implications for motion planning. Therefore, we give a brief overview. First, the notion of a group must be precisely defined. A *group* is a set, G, together with a binary operation, $\circ$, such that the following *group axioms* are satisfied:

1. (**Closure**) For any $a, b \in G$, the product $a \circ b \in G$.

2. (**Associativity**) For all $a, b, c \in G$, $(a \circ b) \circ c = a \circ (b \circ c)$. Hence, parentheses are not needed, and the product may be written as $a \circ b \circ c$.

3. (**Identity**) There is an element $e \in G$, called the *identity*, such that for all $a \in G$, $e \circ a = a$ and $a \circ e = a$.

4. (**Inverse**) For every element $a \in G$, there is an element a^{-1}, called the *inverse* of a, for which $a \circ a^{-1} = e$ and $a^{-1} \circ a = e$.

Here are some examples.

Example 4.7 (Simple Examples of Groups) The set of integers $\mathbb{Z}$ is a group with respect to addition. The identity is 0, and the inverse of each i is $-i$. The set $\mathbb{Q} \backslash 0$ of rational numbers with 0 removed is a group with respect to multiplication. The identity is 1, and the inverse of every element, q, is $1/q$ (0 was removed to

avoid division by zero).   ■

An important property, which only some groups possess, is *commutativity*: $a \circ b = b \circ a$ for any $a, b \in G$. The group in this case is called *commutative* or *Abelian*. We will encounter examples of both kinds of groups, both commutative and noncommutative. An example of a commutative group is vector addition over $\mathbb{R}^n$. The set of all 3D rotations is an example of a noncommutative group.

The fundamental group   Now an interesting group will be constructed from the space of paths and the equivalence relation obtained by homotopy. The *fundamental group*, $\pi_1(X)$ (or *first homotopy group*), is associated with any topological space, X. Let a (continuous) path for which $f(0) = f(1)$ be called a *loop*. Let some $x_b \in X$ be designated as a *base point*. For some arbitrary but fixed base point, x_b, consider the set of all loops such that $f(0) = f(1) = x_b$. This can be made into a group by defining the following binary operation. Let $\tau_1 : [0, 1] \to X$ and $\tau_2 : [0, 1] \to X$ be two loop paths with the same base point. Their product $\tau = \tau_1 \circ \tau_2$ is defined as

$$\tau(t) = \begin{cases} \tau_1(2t) & \text{if } t \in [0, 1/2) \\ \tau_2(2t - 1) & \text{if } t \in [1/2, 1]. \end{cases} \tag{4.9}$$

This results in a continuous loop path because τ_1 terminates at x_b, and τ_2 begins at x_b. In a sense, the two paths are concatenated end-to-end.

Suppose now that the equivalence relation induced by homotopy is applied to the set of all loop paths through a fixed point, x_b. It will no longer be important which particular path was chosen from a class; any representative may be used. The equivalence relation also applies when the set of loops is interpreted as a group. The group operation actually occurs over the set of equivalences of paths.

Consider what happens when two paths from different equivalence classes are concatenated using $\circ$. Is the resulting path homotopic to either of the first two? Is the resulting path homotopic if the original two are from the same homotopy class? The answers in general are *no* and *no*, respectively. The fundamental group describes how the equivalence classes of paths are related and characterizes the connectivity of X. Since fundamental groups are based on paths, there is a nice connection to motion planning.

Example 4.8 (A Simply Connected Space)   Suppose that a topological space X is simply connected. In this case, all loop paths from a base point x_b are homotopic, resulting in one equivalence class. The result is $\pi_1(X) = \mathbf{1_G}$, which is the group that consists of only the identity element.   ■

Example 4.9 (The Fundamental Group of $\mathbb{S}^1$)   Suppose $X = \mathbb{S}^1$. In this case, there is an equivalence class of paths for each $i \in \mathbb{Z}$, the set of integers.

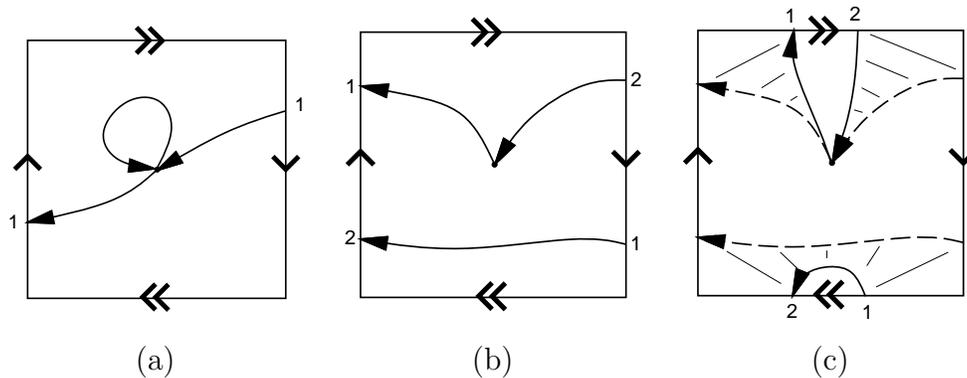(a)   (b)   (c)

Figure 4.7: An illustration of why $\pi_1(\mathbb{RP}^2) = \mathbb{Z}_2$. The integers 1 and 2 indicate precisely where a path continues when it reaches the boundary. (a) Two paths are shown that are not equivalent. (b) A path that winds around twice is shown. (c) This is homotopic to a loop path that does not wind around at all. Eventually, the part of the path that appears at the bottom is pulled through the top. It finally shrinks into an arbitrarily small loop.

If $i > 0$, then it means that the path winds i times around $\mathbb{S}^1$ in the counterclockwise direction and then returns to x_b. If $i < 0$, then the path winds around i times in the clockwise direction. If $i = 0$, then the path is equivalent to one that remains at x_b. The fundamental group is $\mathbb{Z}$, with respect to the operation of addition. If τ_1 travels i_1 times counterclockwise, and τ_2 travels i_2 times counterclockwise, then $\tau = \tau_1 \circ \tau_2$ belongs to the class of loops that travel around $i_1 + i_2$ times counterclockwise. Consider additive inverses. If a path travels seven times around $\mathbb{S}^1$, and it is combined with a path that travels seven times in the opposite direction, the result is homotopic to a path that remains at x_b. Thus, $\pi_1(\mathbb{S}^1) = \mathbb{Z}$.   ∎

Example 4.10 (The Fundamental Group of $\mathbb{T}^n$) For the torus, $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$, in which the ith component of $\mathbb{Z}^n$ corresponds to the number of times a loop path wraps around the ith component of $\mathbb{T}^n$. This makes intuitive sense because $\mathbb{T}^n$ is just the Cartesian product of n circles. The fundamental group $\mathbb{Z}^n$ is obtained by starting with a simply connected subset of the plane and drilling out n disjoint, bounded holes. This situation arises frequently when a mobile robot must avoid collision with n disjoint obstacles in the plane.   ∎

By now it seems that the fundamental group simply keeps track of how many times a path travels around holes. This next example yields some very bizarre behavior that helps to illustrate some of the interesting structure that arises in algebraic topology.

Example 4.11 (The Fundamental Group of $\mathbb{RP}^2$) Suppose $X = \mathbb{RP}^2$, the projective plane. In this case, there are only two equivalence classes on the space of loop paths. All paths that "wrap around" an even number of times are homotopic. Likewise, all paths that wrap around an odd number of times are homotopic. This strange behavior is illustrated in Figure 4.7. The resulting fundamental group therefore has only two elements: $\pi_1(\mathbb{RP}^2) = \mathbb{Z}_2$, the cyclic group of order 2, which corresponds to addition mod 2. This makes intuitive sense because the group keeps track of whether a sum of integers is odd or even, which in this application corresponds to the total number of traversals over the square representation of $\mathbb{RP}^2$. The fundamental group is the same for $\mathbb{RP}^3$, which arises in Section 4.2.2 because it is homeomorphic to the set of 3D rotations. Thus, there are surprisingly only two path classes for the set of 3D rotations. ∎

Unfortunately, two topological spaces may have the same fundamental group even if the spaces are not homeomorphic. For example, $\mathbb{Z}$ is the fundamental group of $\mathbb{S}^1$, the cylinder, $\mathbb{R} \times \mathbb{S}^1$, and the Möbius band. In the last case, the fundamental group does not indicate that there is a "twist" in the space. Another problem is that spaces with interesting connectivity may be declared as simply connected. The fundamental group of the sphere $\mathbb{S}^2$ is just $\mathbf{1_G}$, the same as for $\mathbb{R}^2$. Try envisioning loop paths on the sphere; it can be seen that they all fall into one equivalence class. Hence, $\mathbb{S}^2$ is simply connected. The fundamental group also neglects bubbles in $\mathbb{R}^3$ because the homotopy can warp paths around them. Some of these troubles can be fixed by defining second-order homotopy groups. For example, a continuous function, $[0,1] \times [0,1] \to X$, of two variables can be used instead of a path. The resulting homotopy generates a kind of sheet or surface that can be warped through the space, to yield a homotopy group $\pi_2(X)$ that wraps around bubbles in $\mathbb{R}^3$. This idea can be extended beyond two dimensions to detect many different kinds of holes in higher dimensional spaces. This leads to the *higher order homotopy groups*. A stronger concept than simply connected for a space is that its homotopy groups of all orders are equal to the identity group. This prevents all kinds of holes from occurring and implies that a space, X, is *contractible*, which means a kind of homotopy can be constructed that shrinks X to a point [439]. In the plane, the notions of *contractible* and *simply connected* are equivalent; however, in higher dimensional spaces, such as those arising in motion planning, the term *contractible* should be used to indicate that the space has no interior obstacles (holes).

An alternative to basing groups on homotopy is to derive them using *homology*, which is based on the structure of cell complexes instead of homotopy mappings. This subject is much more complicated to present, but it is more powerful for proving theorems in topology. See the literature overview at the end of the chapter for suggested further reading on algebraic topology.

4.2 Defining the Configuration Space

This section defines the manifolds that arise from the transformations of Chapter 3. If the robot has n degrees of freedom, the set of transformations is usually a manifold of dimension n. This manifold is called the *configuration space* of the robot, and its name is often shortened to *C-space*. In this book, the C-space may be considered as a special state space. To solve a motion planning problem, algorithms must conduct a search in the C-space. The C-space provides a powerful abstraction that converts the complicated models and transformations of Chapter 3 into the general problem of computing a path that traverses a manifold. By developing algorithms directly for this purpose, they apply to a wide variety of different kinds of robots and transformations. In Section 4.3 the problem will be complicated by bringing obstacles into the configuration space, but in Section 4.2 there will be no obstacles.

4.2.1 2D Rigid Bodies: $SE(2)$

Section 3.2.2 expressed how to transform a rigid body in $\mathbb{R}^2$ by a homogeneous transformation matrix, T, given by (3.35). The task in this chapter is to characterize the set of all possible rigid-body transformations. Which manifold will this be? Here is the answer and brief explanation. Since any $x_t, y_t \in \mathbb{R}$ can be selected for translation, this alone yields a manifold $M_1 = \mathbb{R}^2$. Independently, any rotation, $\theta \in [0, 2\pi)$, can be applied. Since 2π yields the same rotation as 0, they can be identified, which makes the set of 2D rotations into a manifold, $M_2 = \mathbb{S}^1$. To obtain the manifold that corresponds to all rigid-body motions, simply take $\mathcal{C} = M_1 \times M_2 = \mathbb{R}^2 \times \mathbb{S}^1$. The answer to the question is that the C-space is a kind of cylinder.

Now we give a more detailed technical argument. The main purpose is that such a simple, intuitive argument will not work for the 3D case. Our approach is to introduce some of the technical machinery here for the 2D case, which is easier to understand, and then extend it to the 3D case in Section 4.2.2.

Matrix groups The first step is to consider the set of transformations as a group, in addition to a topological space.[8] We now derive several important groups from sets of matrices, ultimately leading to $SO(n)$, the group of $n \times n$ rotation matrices, which is very important for motion planning. The set of all nonsingular $n \times n$ real-valued matrices is called the *general linear group*, denoted by $GL(n)$, with respect to matrix multiplication. Each matrix $A \in GL(n)$ has an inverse $A^{-1} \in GL(n)$, which when multiplied yields the identity matrix, $AA^{-1} = I$. The

---

[8]The groups considered in this section are actually Lie groups because they are smooth manifolds [63]. We will not use that name here, however, because the notion of a smooth structure has not yet been defined. Readers familiar with Lie groups, however, will recognize most of the coming concepts. Some details on Lie groups appear later in Sections 15.4.3 and 15.5.1.

matrices must be nonsingular for the same reason that 0 was removed from $\mathbb{Q}$. The analog of division by zero for matrix algebra is the inability to invert a singular matrix.

Many interesting groups can be formed from one group, G_1, by removing some elements to obtain a *subgroup*, G_2. To be a subgroup, G_2 must be a subset of G_1 and satisfy the group axioms. We will arrive at the set of rotation matrices by constructing subgroups. One important subgroup of $GL(n)$ is the *orthogonal group*, $O(n)$, which is the set of all matrices $A \in GL(n)$ for which $AA^T = I$, in which A^T denotes the matrix *transpose* of A. These matrices have orthogonal columns (the inner product of any pair is zero) and the determinant is always 1 or -1. Thus, note that AA^T takes the inner product of every pair of columns. If the columns are different, the result must be 0; if they are the same, the result is 1 because $AA^T = I$. The *special orthogonal group*, $SO(n)$, is the subgroup of $O(n)$ in which every matrix has determinant 1. Another name for $SO(n)$ is the *group of n-dimensional rotation matrices*.

A chain of groups, $SO(n) \leq O(n) \leq GL(n)$, has been described in which $\leq$ denotes "a subgroup of." Each group can also be considered as a topological space. The set of all $n \times n$ matrices (which is not a group with respect to multiplication) with real-valued entries is homeomorphic to $\mathbb{R}^{n^2}$ because n^2 entries in the matrix can be independently chosen. For $GL(n)$, singular matrices are removed, but an n^2-dimensional manifold is nevertheless obtained. For $O(n)$, the expression $AA^T = I$ corresponds to n^2 algebraic equations that have to be satisfied. This should substantially drop the dimension. Note, however, that many of the equations are redundant (pick your favorite value for n, multiply the matrices, and see what happens). There are only $\binom{n}{2}$ ways (pairwise combinations) to take the inner product of pairs of columns, and there are n equations that require the magnitude of each column to be 1. This yields a total of $n(n+1)/2$ independent equations. Each independent equation drops the manifold dimension by one, and the resulting dimension of $O(n)$ is $n^2 - n(n+1)/2 = n(n-1)/2$, which is easily remembered as $\binom{n}{2}$. To obtain $SO(n)$, the constraint $\det A = 1$ is added, which eliminates exactly half of the elements of $O(n)$ but keeps the dimension the same.

Example 4.12 (Matrix Subgroups) It is helpful to illustrate the concepts for $n = 2$. The set of all 2×2 matrices is

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R} \right\}, \tag{4.10}$$

which is homeomorphic to $\mathbb{R}^4$. The group $GL(2)$ is formed from the set of all nonsingular 2×2 matrices, which introduces the constraint that $ad - bc \neq 0$. The set of singular matrices forms a 3D manifold with boundary in $\mathbb{R}^4$, but all other elements of $\mathbb{R}^4$ are in $GL(2)$; therefore, $GL(2)$ is a 4D manifold.

Next, the constraint $AA^T = I$ is enforced to obtain $O(2)$. This becomes

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.11}$$

which directly yields four algebraic equations:

$$a^2 + b^2 = 1 \tag{4.12}$$
$$ac + bd = 0 \tag{4.13}$$
$$ca + db = 0 \tag{4.14}$$
$$c^2 + d^2 = 1. \tag{4.15}$$

Note that (4.14) is redundant. There are two kinds of equations. One equation, given by (4.13), forces the inner product of the columns to be 0. There is only one because $\binom{n}{2} = 1$ for $n = 2$. Two other constraints, (4.12) and (4.15), force the rows to be unit vectors. There are two because $n = 2$. The resulting dimension of the manifold is $\binom{n}{2} = 1$ because we started with $\mathbb{R}^4$ and lost three dimensions from (4.12), (4.13), and (4.15). What does this manifold look like? Imagine that there are two different two-dimensional unit vectors, (a, b) and (c, d). Any value can be chosen for (a, b) as long as $a^2 + b^2 = 1$. This looks like $\mathbb{S}^1$, but the inner product of (a, b) and (c, d) must also be 0. Therefore, for each value of (a, b), there are two choices for c and d: 1) $c = b$ and $d = -a$, or 2) $c = -b$ and $d = a$. It appears that there are two circles! The manifold is $\mathbb{S}^1 \sqcup \mathbb{S}^1$, in which $\sqcup$ denotes the union of disjoint sets. Note that this manifold is not connected because no path exists from one circle to the other.

The final step is to require that $\det A = ad - bc = 1$, to obtain $SO(2)$, the set of all 2D rotation matrices. Without this condition, there would be matrices that produce a rotated mirror image of the rigid body. The constraint simply forces the choice for c and d to be $c = -b$ and $a = d$. This throws away one of the circles from $O(2)$, to obtain a single circle for $SO(2)$. We have finally obtained what you already knew: $SO(2)$ is homeomorphic to $\mathbb{S}^1$. The circle can be parameterized using polar coordinates to obtain the standard 2D rotation matrix, (3.31), given in Section 3.2.2. ∎

Special Euclidean group Now that the group of rotations, $SO(n)$, is characterized, the next step is to allow both rotations and translations. This corresponds to the set of all $(n + 1) \times (n + 1)$ transformation matrices of the form

$$\left\{ \begin{pmatrix} R & v \\ 0 & 1 \end{pmatrix} \,\middle|\, R \in SO(n) \text{ and } v \in \mathbb{R}^n \right\}. \tag{4.16}$$

This should look like a generalization of (3.52) and (3.56), which were for $n = 2$ and $n = 3$, respectively. The R part of the matrix achieves rotation of an n-dimensional body in $\mathbb{R}^n$, and the v part achieves translation of the same body. The result is a group, $SE(n)$, which is called the *special Euclidean group*. As a topological space, $SE(n)$ is homeomorphic to $\mathbb{R}^n \times SO(n)$, because the rotation matrix and translation vectors may be chosen independently. In the case of $n = 2$, this means $SE(2)$ is homeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$, which verifies what was stated
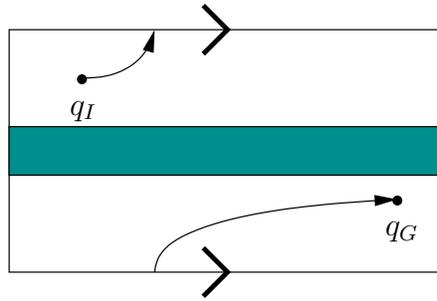
Figure 4.8: A planning algorithm may have to cross the identification boundary to find a solution path.

at the beginning of this section. Thus, the C-space of a 2D rigid body that can translate and rotate in the plane is

$$\mathcal{C} = \mathbb{R}^2 \times \mathbb{S}^1. \tag{4.17}$$

To be more precise, $\cong$ should be used in the place of $=$ to indicate that $\mathcal{C}$ could be any space homeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$; however, this notation will mostly be avoided.

Interpreting the C-space   It is important to consider the topological implications of $\mathcal{C}$. Since $\mathbb{S}^1$ is multiply connected, $\mathbb{R} \times \mathbb{S}^1$ and $\mathbb{R}^2 \times \mathbb{S}^1$ are multiply connected. It is difficult to visualize $\mathcal{C}$ because it is a 3D manifold; however, there is a nice interpretation using identification. Start with the open unit cube, $(0,1)^3 \subset \mathbb{R}^3$. Include the boundary points of the form $(x, y, 0)$ and $(x, y, 1)$, and make the identification $(x, y, 0) \sim (x, y, 1)$ for all $x, y \in (0, 1)$. This means that when traveling in the x and y directions, there is a "frontier" to the C-space; however, traveling in the z direction causes a wraparound.

It is very important for a motion planning algorithm to understand that this wraparound exists. For example, consider $\mathbb{R} \times \mathbb{S}^1$ because it is easier to visualize. Imagine a path planning problem for which $\mathcal{C} = \mathbb{R} \times \mathbb{S}^1$, as depicted in Figure 4.8. Suppose the top and bottom are identified to make a cylinder, and there is an obstacle across the middle. Suppose the task is to find a path from q_I to q_G. If the top and bottom were not identified, then it would not be possible to connect q_I to q_G; however, if the algorithm realizes it was given a cylinder, the task is straightforward. In general, it is very important to understand the topology of $\mathcal{C}$; otherwise, potential solutions will be lost.

The next section addresses $SE(n)$ for $n = 3$. The main difficulty is determining the topology of $SO(3)$. At least we do not have to consider $n > 3$ in this book.

4.2.2   3D Rigid Bodies: $SE(3)$

One might expect that defining $\mathcal{C}$ for a 3D rigid body is an obvious extension of the 2D case; however, 3D rotations are significantly more complicated. The resulting

C-space will be a six-dimensional manifold, $\mathcal{C} = \mathbb{R}^3 \times \mathbb{RP}^3$. Three dimensions come from translation and three more come from rotation.

The main quest in this section is to determine the topology of $SO(3)$. In Section 3.2.3, yaw, pitch, and roll were used to generate rotation matrices. These angles are convenient for visualization, performing transformations in software, and also for deriving the DH parameters. However, these were concerned with applying a single rotation, whereas the current problem is to characterize the set of all rotations. It is possible to use α, β, and γ to parameterize the set of rotations, but it causes serious troubles. There are some cases in which nonzero angles yield the identity rotation matrix, which is equivalent to $\alpha = \beta = \gamma = 0$. There are also cases in which a continuum of values for yaw, pitch, and roll angles yield the same rotation matrix. These problems destroy the topology, which causes both theoretical and practical difficulties in motion planning.

Consider applying the matrix group concepts from Section 4.2.1. The general linear group $GL(3)$ is homeomorphic to $\mathbb{R}^9$. The orthogonal group, $O(3)$, is determined by imposing the constraint $AA^T = I$. There are $\binom{3}{2} = 3$ independent equations that require distinct columns to be orthogonal, and three independent equations that force the magnitude of each column to be 1. This means that $O(3)$ has three dimensions, which matches our intuition since there were three rotation parameters in Section 3.2.3. To obtain $SO(3)$, the last constraint, $\det A = 1$, is added. Recall from Example 4.12 that $SO(2)$ consists of two circles, and the constraint $\det A = 1$ selects one of them. In the case of $O(3)$, there are two three-spheres, $\mathbb{S}^3 \sqcup \mathbb{S}^3$, and $\det A = 1$ selects one of them. However, there is one additional complication: Antipodal points on these spheres generate the same rotation matrix. This will be seen shortly when quaternions are used to parameterize $SO(3)$.

Using complex numbers to represent $SO(2)$ Before introducing quaternions to parameterize 3D rotations, consider using complex numbers to parameterize 2D rotations. Let the term *unit complex number* refer to any complex number, $a + bi$, for which $a^2 + b^2 = 1$.

The set of all unit complex numbers forms a group under multiplication. It will be seen that it is "the same" group as $SO(2)$. This idea needs to be made more precise. Two groups, G and H, are considered "the same" if they are *isomorphic*, which means that there exists a bijective function $f : G \to H$ such that for all $a, b \in G$, $f(a) \circ f(b) = f(a \circ b)$. This means that we can perform some calculations in G, map the result to H, perform more calculations, and map back to G without any trouble. The sets G and H are just two alternative ways to express the same group.

The unit complex numbers and $SO(2)$ are isomorphic. To see this clearly, recall that complex numbers can be represented in polar form as $re^{i\theta}$; a unit complex number is simply $e^{i\theta}$. A bijective mapping can be made between 2D rotation matrices and unit complex numbers by letting $e^{i\theta}$ correspond to the rotation matrix (3.31).

If complex numbers are used to represent rotations, it is important that they behave algebraically in the same way. If two rotations are combined, the matrices are multiplied. The equivalent operation is multiplication of complex numbers. Suppose that a 2D robot is rotated by θ_1, followed by θ_2. In polar form, the complex numbers are multiplied to yield $e^{i\theta_1}e^{i\theta_2} = e^{i(\theta_1+\theta_2)}$, which clearly represents a rotation of $\theta_1 + \theta_2$. If the unit complex number is represented in Cartesian form, then the rotations corresponding to $a_1 + b_1i$ and $a_2 + b_2i$ are combined to obtain $(a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)i$. Note that here we have not used complex numbers to express the solution to a polynomial equation, which is their more popular use; we simply borrowed their nice algebraic properties. At any time, a complex number $a + bi$ can be converted into the equivalent rotation matrix

$$R(a, b) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \tag{4.18}$$

Recall that only one independent parameter needs to be specified because $a^2 + b^2 = 1$. Hence, it appears that the set of unit complex numbers is the same manifold as $SO(2)$, which is the circle $\mathbb{S}^1$ (recall, that "same" means in the sense of homeomorphism).

Quaternions   The manner in which complex numbers were used to represent 2D rotations will now be adapted to using quaternions to represent 3D rotations. Let $\mathbb{H}$ represent the set of *quaternions*, in which each quaternion, $h \in \mathbb{H}$, is represented as $h = a + bi + cj + dk$, and $a, b, c, d \in \mathbb{R}$. A quaternion can be considered as a four-dimensional vector. The symbols i, j, and k are used to denote three "imaginary" components of the quaternion. The following relationships are defined: $i^2 = j^2 = k^2 = ijk = -1$, from which it follows that $ij = k$, $jk = i$, and $ki = j$. Using these, multiplication of two quaternions, $h_1 = a_1 + b_1i + c_1j + d_1k$ and $h_2 = a_2 + b_2i + c_2j + d_2k$, can be derived to obtain $h_1 \cdot h_2 = a_3 + b_3i + c_3j + d_3k$, in which

$$\begin{aligned} a_3 &= a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2 \\ b_3 &= a_1b_2 + a_2b_1 + c_1d_2 - c_2d_1 \\ c_3 &= a_1c_2 + a_2c_1 + b_2d_1 - b_1d_2 \\ d_3 &= a_1d_2 + a_2d_1 + b_1c_2 - b_2c_1. \end{aligned} \tag{4.19}$$

Using this operation, it can be shown that $\mathbb{H}$ is a group with respect to quaternion multiplication. Note, however, that the multiplication is not commutative! This is also true of 3D rotations; there must be a good reason.

For convenience, quaternion multiplication can be expressed in terms of vector multiplications, a dot product, and a cross product. Let $v = [b \ \ c \ \ d]$ be a three-dimensional vector that represents the final three quaternion components. The first component of $h_1 \cdot h_2$ is $a_1a_2 - v_1 \cdot v_2$. The final three components are given by the three-dimensional vector $a_1v_2 + a_2v_1 + v_1 \times v_2$.

In the same way that *unit* complex numbers were needed for $SO(2)$, *unit quaternions* are needed for $SO(3)$, which means that $\mathbb{H}$ is restricted to quaternions for
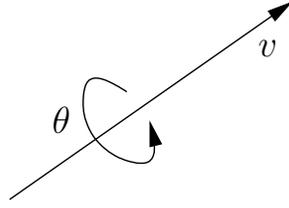
Figure 4.9: Any 3D rotation can be considered as a rotation by an angle θ about the axis given by the unit direction vector $v = [v_1 \; v_2 \; v_3]$.
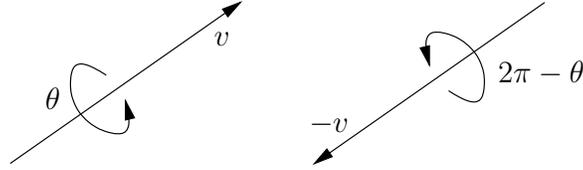

Figure 4.10: There are two ways to encode the same rotation.

which $a^2 + b^2 + c^2 + d^2 = 1$. Note that this forms a subgroup because the multiplication of unit quaternions yields a unit quaternion, and the other group axioms hold.

The next step is to describe a mapping from unit quaternions to $SO(3)$. Let the unit quaternion $h = a + bi + cj + dk$ map to the matrix

$$R(h) = \begin{pmatrix} 2(a^2 + b^2) - 1 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & 2(a^2 + c^2) - 1 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & 2(a^2 + d^2) - 1 \end{pmatrix}, \tag{4.20}$$

which can be verified as orthogonal and $\det R(h) = 1$. Therefore, it belongs to $SO(3)$. It is not shown here, but it conveniently turns out that h represents the rotation shown in Figure 4.9, by making the assignment

$$h = \cos\frac{\theta}{2} + \left(v_1 \sin\frac{\theta}{2}\right) i + \left(v_2 \sin\frac{\theta}{2}\right) j + \left(v_3 \sin\frac{\theta}{2}\right) k. \tag{4.21}$$

Unfortunately, this representation is not unique. It can be verified in (4.20) that $R(h) = R(-h)$. A nice geometric interpretation is given in Figure 4.10. The quaternions h and $-h$ represent the same rotation because a rotation of θ about the direction v is equivalent to a rotation of $2\pi - \theta$ about the direction $-v$. Consider the quaternion representation of the second expression of rotation with respect to the first. The real part is

$$\cos\left(\frac{2\pi - \theta}{2}\right) = \cos\left(\pi - \frac{\theta}{2}\right) = -\cos\left(\frac{\theta}{2}\right) = -a. \tag{4.22}$$

The i, j, and k components are

$$-v \sin\left(\frac{2\pi - \theta}{2}\right) = -v \sin\left(\pi - \frac{\theta}{2}\right) = -v \sin\left(\frac{\theta}{2}\right) = [-b \;\; -c \;\; -d]. \tag{4.23}$$

The quaternion $-h$ has been constructed. Thus, h and $-h$ represent the same rotation. Luckily, this is the only problem, and the mapping given by (4.20) is two-to-one from the set of unit quaternions to $SO(3)$.

This can be fixed by the identification trick. Note that the set of unit quaternions is homeomorphic to $\mathbb{S}^3$ because of the constraint $a^2 + b^2 + c^2 + d^2 = 1$. The algebraic properties of quaternions are not relevant at this point. Just imagine each h as an element of $\mathbb{R}^4$, and the constraint $a^2 + b^2 + c^2 + d^2 = 1$ forces the points to lie on $\mathbb{S}^3$. Using identification, declare $h \sim -h$ for all unit quaternions. This means that the antipodal points of $\mathbb{S}^3$ are identified. Recall from the end of Section 4.1.2 that when antipodal points are identified, $\mathbb{RP}^n \cong \mathbb{S}^n / \sim$. Hence, $SO(3) \cong \mathbb{RP}^3$, which can be considered as the set of all lines through the origin of $\mathbb{R}^4$, but this is hard to visualize. The representation of $\mathbb{RP}^2$ in Figure 4.5 can be extended to $\mathbb{RP}^3$. Start with $(0,1)^3 \subset \mathbb{R}^3$, and make three different kinds of identifications, one for each pair of opposite cube faces, and add all of the points to the manifold. For each kind of identification a twist needs to be made (without the twist, $\mathbb{T}^3$ would be obtained). For example, in the z direction, let $(x, y, 0) \sim (1 - x, 1 - y, 1)$ for all $x, y \in [0, 1]$.

One way to force uniqueness of rotations is to require staying in the "upper half" of $\mathbb{S}^3$. For example, require that $a \geq 0$, as long as the boundary case of $a = 0$ is handled properly because of antipodal points at the equator of $\mathbb{S}^3$. If $a = 0$, then require that $b \geq 0$. However, if $a = b = 0$, then require that $c \geq 0$ because points such as $(0, 0, -1, 0)$ and $(0, 0, 1, 0)$ are the same rotation. Finally, if $a = b = c = 0$, then only $d = 1$ is allowed. If such restrictions are made, it is important, however, to remember the connectivity of $\mathbb{RP}^3$. If a path travels across the equator of $\mathbb{S}^3$, it must be mapped to the appropriate place in the "northern hemisphere." At the instant it hits the equator, it must move to the antipodal point. These concepts are much easier to visualize if you remove a dimension and imagine them for $\mathbb{S}^2 \subset \mathbb{R}^3$, as described at the end of Section 4.1.2.

Using quaternion multiplication   The representation of rotations boiled down to picking points on $\mathbb{S}^3$ and respecting the fact that antipodal points give the same element of $SO(3)$. In a sense, this has nothing to do with the algebraic properties of quaternions. It merely means that $SO(3)$ can be parameterized by picking points in $\mathbb{S}^3$, just like $SO(2)$ was parameterized by picking points in $\mathbb{S}^1$ (ignoring the antipodal identification problem for $SO(3)$).

However, one important reason why the quaternion arithmetic was introduced is that the group of unit quaternions with h and $-h$ identified is also isomorphic to $SO(3)$. This means that a sequence of rotations can be multiplied together using quaternion multiplication instead of matrix multiplication. This is important because fewer operations are required for quaternion multiplication in comparison to matrix multiplication. At any point, (4.20) can be used to convert the result back into a matrix; however, this is not even necessary. It turns out that a point in the world, $(x, y, z) \in \mathbb{R}^3$, can be transformed by directly using quaternion arithmetic. An analog to the complex conjugate from complex numbers is needed.

For any $h = a + bi + cj + dk \in \mathbb{H}$, let $h^* = a - bi - cj - dk$ be its *conjugate*. For any point $(x, y, z) \in \mathbb{R}^3$, let $p \in \mathbb{H}$ be the quaternion $0 + xi + yj + zk$. It can be shown (with a lot of algebra) that the rotated point (x, y, z) is given by $h \cdot p \cdot h^*$. The i, j, k components of the resulting quaternion are new coordinates for the transformed point. It is equivalent to having transformed (x, y, z) with the matrix $R(h)$.

Finding quaternion parameters from a rotation matrix   Recall from Section 3.2.3 that given a rotation matrix (3.43), the yaw, pitch, and roll parameters could be directly determined using the atan2 function. It turns out that the quaternion representation can also be determined directly from the matrix. This is the inverse of the function in (4.20).[9]

For a given rotation matrix (3.43), the quaternion parameters $h = a + bi + cj + dk$ can be computed as follows [210]. The first component is

$$a = \tfrac{1}{2}\sqrt{r_{11} + r_{22} + r_{33} + 1}, \tag{4.24}$$

and if $a \neq 0$, then

$$b = \frac{r_{32} - r_{23}}{4a}, \tag{4.25}$$

$$c = \frac{r_{13} - r_{31}}{4a}, \tag{4.26}$$

and

$$d = \frac{r_{21} - r_{12}}{4a}. \tag{4.27}$$

If $a = 0$, then the previously mentioned equator problem occurs. In this case,

$$b = \frac{r_{13}r_{12}}{\sqrt{r_{12}^2 r_{13}^2 + r_{12}^2 r_{23}^2 + r_{13}^2 r_{23}^2}}, \tag{4.28}$$

$$c = \frac{r_{12}r_{23}}{\sqrt{r_{12}^2 r_{13}^2 + r_{12}^2 r_{23}^2 + r_{13}^2 r_{23}^2}}, \tag{4.29}$$

and

$$d = \frac{r_{13}r_{23}}{\sqrt{r_{12}^2 r_{13}^2 + r_{12}^2 r_{23}^2 + r_{13}^2 r_{23}^2}}. \tag{4.30}$$

This method fails if $r_{12} = r_{23} = 0$ or $r_{13} = r_{23} = 0$ or $r_{12} = r_{23} = 0$. These correspond precisely to the cases in which the rotation matrix is a yaw, (3.39), pitch, (3.40), or roll, (3.41), which can be detected in advance.

---

[9]Since that function was two-to-one, it is technically not an inverse until the quaternions are restricted to the upper hemisphere, as described previously.

Special Euclidean group   Now that the complicated part of representing $SO(3)$ has been handled, the representation of $SE(3)$ is straightforward. The general form of a matrix in $SE(3)$ is given by (4.16), in which $R \in SO(3)$ and $v \in \mathbb{R}^3$. Since $SO(3) \cong \mathbb{RP}^3$, and translations can be chosen independently, the resulting C-space for a rigid body that rotates and translates in $\mathbb{R}^3$ is

$$\mathcal{C} = \mathbb{R}^3 \times \mathbb{RP}^3, \tag{4.31}$$

which is a six-dimensional manifold. As expected, the dimension of $\mathcal{C}$ is exactly the number of degrees of freedom of a free-floating body in space.

4.2.3   Chains and Trees of Bodies

If there are multiple bodies that are allowed to move independently, then their C-spaces can be combined using Cartesian products. Let $\mathcal{C}_i$ denote the C-space of $\mathcal{A}_i$. If there are n free-floating bodies in $\mathcal{W} = \mathbb{R}^2$ or $\mathcal{W} = \mathbb{R}^3$, then

$$\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_n. \tag{4.32}$$

If the bodies are attached to form a kinematic chain or kinematic tree, then each C-space must be considered on a case-by-case basis. There is no general rule that simplifies the process. One thing to generally be careful about is that the full range of motion might not be possible for typical joints. For example, a revolute joint might not be able to swing all of the way around to enable any $\theta \in [0, 2\pi)$. If θ cannot wind around $\mathbb{S}^1$, then the C-space for this joint is homeomorphic to $\mathbb{R}$ instead of $\mathbb{S}^1$. A similar situation occurs for a spherical joint. A typical ball joint cannot achieve any orientation in $SO(3)$ due to mechanical obstructions. In this case, the C-space is not $\mathbb{RP}^3$ because part of $SO(3)$ is missing.

Another complication is that the DH parameterization of Section 3.3.2 is designed to facilitate the assignment of coordinate frames and computation of transformations, but it neglects considerations of topology. For example, a common approach to representing a spherical robot wrist is to make three zero-length links that each behave as a revolute joint. If the range of motion is limited, this might not cause problems, but in general the problems would be similar to using yaw, pitch, and roll to represent $SO(3)$. There may be multiple ways to express the same arm configuration.

Several examples are given below to help in determining C-spaces for chains and trees of bodies. Suppose $\mathcal{W} = \mathbb{R}^2$, and there is a chain of n bodies that are attached by revolute joints. Suppose that the first joint is capable of rotation only about a fixed point (e.g., it spins around a nail). If each joint has the full range of motion $\theta_i \in [0, 2\pi)$, the C-space is

$$\mathcal{C} = \mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 = \mathbb{T}^n. \tag{4.33}$$

However, if each joint is restricted to $\theta_i \in (-\pi/2, \pi/2)$, then $\mathcal{C} = \mathbb{R}^n$. If any transformation in $SE(2)$ can be applied to $\mathcal{A}_1$, then an additional $\mathbb{R}^2$ is needed.

In the case of restricted joint motions, this yields $\mathbb{R}^{n+2}$. If the joints can achieve any orientation, then $\mathcal{C} = \mathbb{R}^2 \times \mathbb{T}^n$. If there are prismatic joints, then each joint contributes $\mathbb{R}$ to the C-space.

Recall from Figure 3.12 that for $\mathcal{W} = \mathbb{R}^3$ there are six different kinds of joints. The cases of revolute and prismatic joints behave the same as for $\mathcal{W} = \mathbb{R}^2$. Each screw joint contributes $\mathbb{R}$. A cylindrical joint contributes $\mathbb{R} \times \mathbb{S}^1$, unless its rotational motion is restricted. A planar joint contributes $\mathbb{R}^2 \times \mathbb{S}^1$ because any transformation in $SE(2)$ is possible. If its rotational motions are restricted, then it contributes $\mathbb{R}^3$. Finally, a spherical joint can theoretically contribute $\mathbb{RP}^3$. In practice, however, this rarely occurs. It is more likely to contribute $\mathbb{R}^2 \times \mathbb{S}^1$ or $\mathbb{R}^3$ after restrictions are imposed. Note that if the first joint is a free-floating body, then it contributes $\mathbb{R}^3 \times \mathbb{RP}^3$.

Kinematic trees can be handled in the same way as kinematic chains. One issue that has not been mentioned is that there might be collisions between the links. This has been ignored up to this point, but obviously this imposes very complicated restrictions. The concepts from Section 4.3 can be applied to handle this case and the placement of additional obstacles in $\mathcal{W}$. Reasoning about these kinds of restrictions and the path connectivity of the resulting space is indeed the main point of motion planning.

4.3 Configuration Space Obstacles

Section 4.2 defined $\mathcal{C}$, the manifold of robot transformations in the absence of any collision constraints. The current section removes from $\mathcal{C}$ the configurations that either cause the robot to collide with obstacles or cause some specified links of the robot to collide with each other. The removed part of $\mathcal{C}$ is referred to as the obstacle region. The leftover space is precisely what a solution path must traverse. A motion planning algorithm must find a path in the leftover space from an initial configuration to a goal configuration. Finally, after the models of Chapter 3 and the previous sections of this chapter, the motion planning problem can be precisely described.

4.3.1 Definition of the Basic Motion Planning Problem

Obstacle region for a rigid body Suppose that the world, $\mathcal{W} = \mathbb{R}^2$ or $\mathcal{W} = \mathbb{R}^3$, contains an obstacle region, $\mathcal{O} \subset \mathcal{W}$. Assume here that a rigid robot, $\mathcal{A} \subset \mathcal{W}$, is defined; the case of multiple links will be handled shortly. Assume that both $\mathcal{A}$ and $\mathcal{O}$ are expressed as semi-algebraic models (which includes polygonal and polyhedral models) from Section 3.1. Let $q \in \mathcal{C}$ denote the *configuration* of $\mathcal{A}$, in which $q = (x_t, y_t, \theta)$ for $\mathcal{W} = \mathbb{R}^2$ and $q = (x_t, y_t, z_t, h)$ for $\mathcal{W} = \mathbb{R}^3$ (h represents the unit quaternion).

The *obstacle region*, $\mathcal{C}_{obs} \subseteq \mathcal{C}$, is defined as

$$\mathcal{C}_{obs} = \{q \in \mathcal{C} \mid \mathcal{A}(q) \cap \mathcal{O} \neq \emptyset\}, \tag{4.34}$$

which is the set of all configurations, q, at which $\mathcal{A}(q)$, the transformed robot, intersects the obstacle region, $\mathcal{O}$. Since $\mathcal{O}$ and $\mathcal{A}(q)$ are closed sets in $\mathcal{W}$, the obstacle region is a closed set in $\mathcal{C}$.

The leftover configurations are called the *free space*, which is defined and denoted as $\mathcal{C}_{free} = \mathcal{C} \setminus \mathcal{C}_{obs}$. Since $\mathcal{C}$ is a topological space and $\mathcal{C}_{obs}$ is closed, $\mathcal{C}_{free}$ must be an open set. This implies that the robot can come arbitrarily close to the obstacles while remaining in $\mathcal{C}_{free}$. If $\mathcal{A}$ "touches" $\mathcal{O}$,

$$\text{int}(\mathcal{O}) \cap \text{int}(\mathcal{A}(q)) = \emptyset \text{ and } \mathcal{O} \cap \mathcal{A}(q) \neq \emptyset, \qquad (4.35)$$

then $q \in \mathcal{C}_{obs}$ (recall that int means the interior). The condition above indicates that only their boundaries intersect.

The idea of getting arbitrarily close may be nonsense in practical robotics, but it makes a clean formulation of the motion planning problem. Since $\mathcal{C}_{free}$ is open, it becomes impossible to formulate some optimization problems, such as finding the shortest path. In this case, the closure, $\text{cl}(\mathcal{C}_{free})$, should instead be used, as described in Section 7.7.

Obstacle region for multiple bodies If the robot consists of multiple bodies, the situation is more complicated. The definition in (4.34) only implies that the robot does not collide with the obstacles; however, if the robot consists of multiple bodies, then it might also be appropriate to avoid collisions between different links of the robot. Let the robot be modeled as a collection, $\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\}$, of m links, which may or may not be attached together by joints. A single configuration vector q is given for the entire collection of links. We will write $\mathcal{A}_i(q)$ for each link, i, even though some of the parameters of q may be irrelevant for moving link $\mathcal{A}_i$. For example, in a kinematic chain, the configuration of the second body does not depend on the angle between the ninth and tenth bodies.

Let P denote the set of *collision pairs*, in which each collision pair, $(i, j) \in P$, represents a pair of link indices $i, j \in \{1, 2, \ldots, m\}$, such that $i \neq j$. If (i, j) appears in P, it means that A_i and A_j are not allowed to be in a configuration, q, for which $\mathcal{A}_i(q) \cap \mathcal{A}_j(q) \neq \emptyset$. Usually, P does not represent all pairs because consecutive links are in contact all of the time due to the joint that connects them. One common definition for P is that each link must avoid collisions with any links to which it is not attached by a joint. For m bodies, P is generally of size $O(m^2)$; however, in practice it is often possible to eliminate many pairs by some geometric analysis of the linkage. Collisions between some pairs of links may be impossible over all of $\mathcal{C}$, in which case they do not need to appear in P.

Using P, the consideration of robot self-collisions is added to the definition of $\mathcal{C}_{obs}$ to obtain

$$\mathcal{C}_{obs} = \left(\bigcup_{i=1}^{m} \{q \in \mathcal{C} \mid \mathcal{A}_i(q) \cap \mathcal{O} \neq \emptyset\} \right) \bigcup \left(\bigcup_{[i,j] \in P} \{q \in \mathcal{C} \mid \mathcal{A}_i(q) \cap \mathcal{A}_j(q) \neq \emptyset\} \right).$$
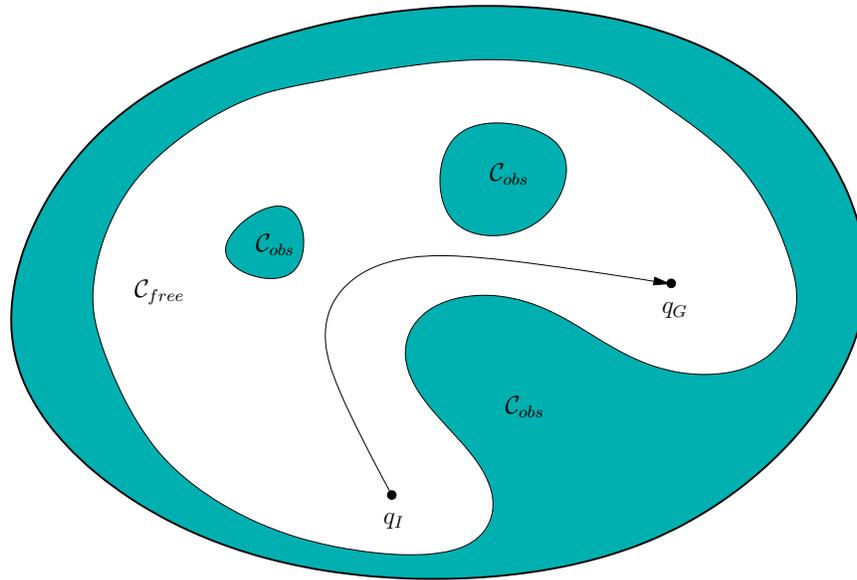$$(4.36)$$

Figure 4.11: The basic motion planning problem is conceptually very simple using C-space ideas. The task is to find a path from q_I to q_G in $\mathcal{C}_{free}$. The entire blob represents $\mathcal{C} = \mathcal{C}_{free} \cup \mathcal{C}_{obs}$.

Thus, a configuration $q \in \mathcal{C}$ is in $\mathcal{C}_{obs}$ if at least one link collides with $\mathcal{O}$ or a pair of links indicated by P collide with each other.

Definition of basic motion planning Finally, enough tools have been introduced to precisely define the motion planning problem. The problem is conceptually illustrated in Figure 4.11. The main difficulty is that it is neither straightforward nor efficient to construct an explicit boundary or solid representation of either $\mathcal{C}_{free}$ or $\mathcal{C}_{obs}$. The components are as follows:

Formulation 4.1 (The Piano Mover's Problem)

1. A *world* $\mathcal{W}$ in which either $\mathcal{W} = \mathbb{R}^2$ or $\mathcal{W} = \mathbb{R}^3$.

2. A semi-algebraic *obstacle region* $\mathcal{O} \subset \mathcal{W}$ in the world.

3. A semi-algebraic *robot* is defined in $\mathcal{W}$. It may be a rigid robot $\mathcal{A}$ or a collection of m links, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$.

4. The *configuration space* $\mathcal{C}$ determined by specifying the set of all possible transformations that may be applied to the robot. From this, $\mathcal{C}_{obs}$ and $\mathcal{C}_{free}$ are derived.

5. A configuration, $q_I \in \mathcal{C}_{free}$ designated as the *initial configuration*.

6. A configuration $q_G \in \mathcal{C}_{free}$ designated as the *goal configuration*. The initial and goal configurations together are often called a *query pair* (or *query*) and designated as (q_I, q_G).

7. A complete algorithm must compute a (continuous) *path*, $\tau : [0, 1] \to \mathcal{C}_{free}$, such that $\tau(0) = q_I$ and $\tau(1) = q_G$, or correctly report that such a path does not exist.

It was shown by Reif [817] that this problem is PSPACE-hard, which implies NP-hard. The main problem is that the dimension of $\mathcal{C}$ is unbounded.

4.3.2   Explicitly Modeling $\mathcal{C}_{obs}$: The Translational Case

It is important to understand how to construct a representation of $\mathcal{C}_{obs}$. In some algorithms, especially the combinatorial methods of Chapter 6, this represents an important first step to solving the problem. In other algorithms, especially the sampling-based planning algorithms of Chapter 5, it helps to understand why such constructions are avoided due to their complexity.

The simplest case for characterizing $\mathcal{C}_{obs}$ is when $\mathcal{C} = \mathbb{R}^n$ for $n = 1$, 2, and 3, and the robot is a rigid body that is restricted to translation only. Under these conditions, $\mathcal{C}_{obs}$ can be expressed as a type of convolution. For any two sets $X, Y \subset \mathbb{R}^n$, let their *Minkowski difference*[10] be defined as

$$X \ominus Y = \{x - y \in \mathbb{R}^n \mid x \in X \text{ and } y \in Y\}, \qquad (4.37)$$

in which $x - y$ is just vector subtraction on $\mathbb{R}^n$. The Minkowski difference between X and Y can also be considered as the Minkowski sum of X and $-Y$. The *Minkowski sum* $\oplus$ is obtained by simply adding elements of X and Y in (4.37), as opposed to subtracting them. The set $-Y$ is obtained by replacing each $y \in Y$ by $-y$.

In terms of the Minkowski difference, $\mathcal{C}_{obs} = \mathcal{O} \ominus \mathcal{A}(0)$. To see this, it is helpful to consider a one-dimensional example.

Example 4.13 (One-Dimensional C-Space Obstacle) In Figure 4.12, both the robot $\mathcal{A} = [-1, 2]$ and obstacle region $\mathcal{O} = [0, 4]$ are intervals in a one-dimensional world, $\mathcal{W} = \mathbb{R}$. The negation, $-\mathcal{A}$, of the robot is shown as the interval $[-2, 1]$. Finally, by applying the Minkowski sum to $\mathcal{O}$ and $-\mathcal{A}$, the C-space obstacle, $\mathcal{C}_{obs} = [-2, 5]$, is obtained.   ■

The Minkowski difference is often considered as a *convolution*. It can even be defined to appear the same as studied in differential equations and system theory.

---

[10]In some contexts, which include mathematics and image processing, the Minkowski difference or *Minkowski subtraction* is defined differently (instead, it is a kind of "erosion"). For this reason, some authors prefer to define all operations in terms of the Minkowski sum, $\oplus$, which is consistently defined in all contexts. Following this convention, we would define $X \oplus (-Y)$, which is equivalent to $X \ominus Y$.
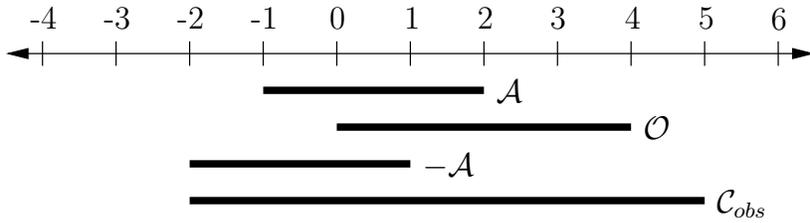
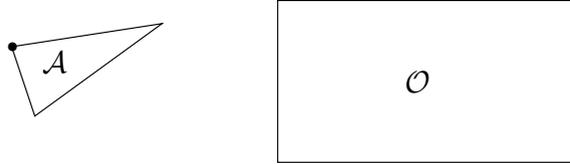Figure 4.12: A one-dimensional C-space obstacle.

Figure 4.13: A triangular robot and a rectangular obstacle.

For a one-dimensional example, let $f : \mathbb{R} \to \{0, 1\}$ be a function such that $f(x) = 1$ if and only if $x \in \mathcal{O}$. Similarly, let $g : \mathbb{R} \to \{0, 1\}$ be a function such that $g(x) = 1$ if and only if $x \in \mathcal{A}$. The convolution

$$h(x) = \int_{-\infty}^{\infty} f(\tau)g(x - \tau)d\tau, \tag{4.38}$$

yields a function h, for which $h(x) > 0$ if $x \in \text{int}(\mathcal{C}_{obs})$, and $h(x) = 0$ otherwise.

A polygonal C-space obstacle   A simple algorithm for computing $\mathcal{C}_{obs}$ exists in the case of a 2D world that contains a convex polygonal obstacle $\mathcal{O}$ and a convex polygonal robot $\mathcal{A}$ [657]. This is often called the *star algorithm*. For this problem, $\mathcal{C}_{obs}$ is also a convex polygon. Recall that nonconvex obstacles and robots can be modeled as the union of convex parts. The concepts discussed below can also be applied in the nonconvex case by considering $\mathcal{C}_{obs}$ as the union of convex components, each of which corresponds to a convex component of $\mathcal{A}$ colliding with a convex component of $\mathcal{O}$.

The method is based on sorting normals to the edges of the polygons on the basis of angles. The key observation is that every edge of $\mathcal{C}_{obs}$ is a translated edge from either $\mathcal{A}$ or $\mathcal{O}$. In fact, every edge from $\mathcal{O}$ and $\mathcal{A}$ is used exactly once in the construction of $\mathcal{C}_{obs}$. The only problem is to determine the ordering of these edges of $\mathcal{C}_{obs}$. Let α_1, α_2, ..., α_n denote the angles of the inward edge normals in counterclockwise order around $\mathcal{A}$. Let β_1, β_2, ..., β_n denote the outward edge normals to $\mathcal{O}$. After sorting both sets of angles in circular order around $\mathbb{S}^1$, $\mathcal{C}_{obs}$ can be constructed incrementally by using the edges that correspond to the sorted normals, in the order in which they are encountered.

Example 4.14 (A Triangular Robot and Rectangular Obstacle) To gain an understanding of the method, consider the case of a triangular robot and a rectangular obstacle, as shown in Figure 4.13. The black dot on $\mathcal{A}$ denotes the origin
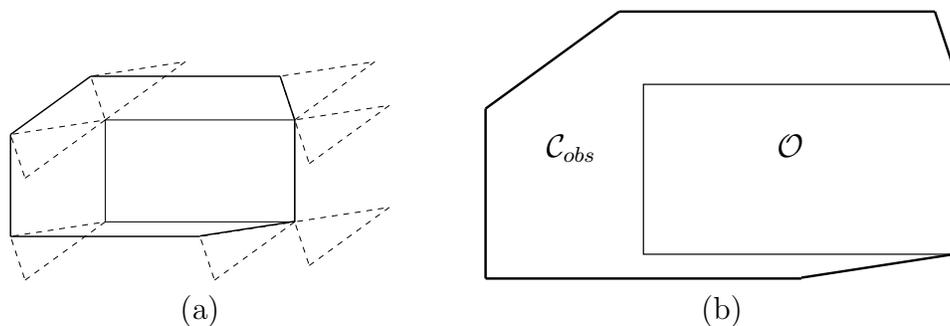
Figure 4.14: (a) Slide the robot around the obstacle while keeping them both in contact. (b) The edges traced out by the origin of $\mathcal{A}$ form $\mathcal{C}_{obs}$.
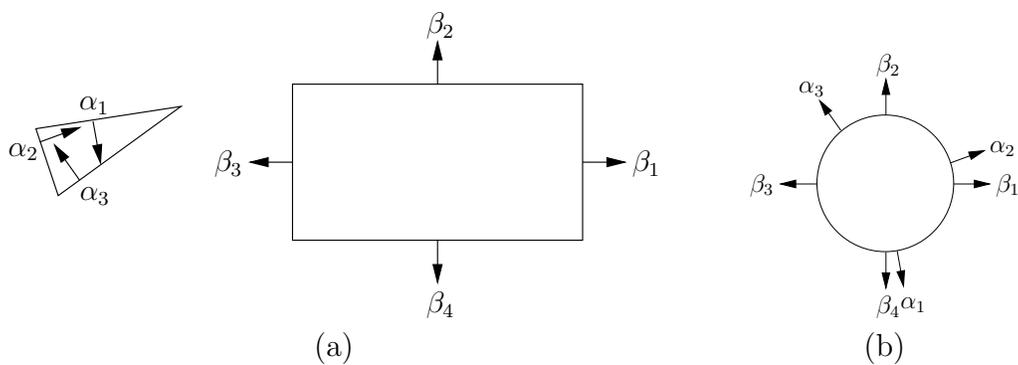

Figure 4.15: (a) Take the inward edge normals of $\mathcal{A}$ and the outward edge normals of $\mathcal{O}$. (b) Sort the edge normals around $\mathbb{S}^1$. This gives the order of edges in $\mathcal{C}_{obs}$.

of its body frame. Consider sliding the robot around the obstacle in such a way that they are always in contact, as shown in Figure 4.14a. This corresponds to the traversal of all of the configurations in $\partial \mathcal{C}_{obs}$ (the boundary of $\mathcal{C}_{obs}$). The origin of $\mathcal{A}$ traces out the edges of $\mathcal{C}_{obs}$, as shown in Figure 4.14b. There are seven edges, and each edge corresponds to either an edge of $\mathcal{A}$ or an edge of $\mathcal{O}$. The directions of the normals are defined as shown in Figure 4.15a. When sorted as shown in Figure 4.15b, the edges of $\mathcal{C}_{obs}$ can be incrementally constructed. ∎

The running time of the algorithm is $O(n + m)$, in which n is the number of edges defining $\mathcal{A}$, and m is the number of edges defining $\mathcal{O}$. Note that the angles can be sorted in linear time because they already appear in counterclockwise order around $\mathcal{A}$ and $\mathcal{O}$; they only need to be merged. If two edges are collinear, then they can be placed end-to-end as a single edge of $\mathcal{C}_{obs}$.

Computing the boundary of $\mathcal{C}_{obs}$ So far, the method quickly identifies each edge that contributes to $\mathcal{C}_{obs}$. It can also construct a solid representation of $\mathcal{C}_{obs}$ in terms of half-planes. This requires defining $n + m$ linear equations (assuming there are no collinear edges).

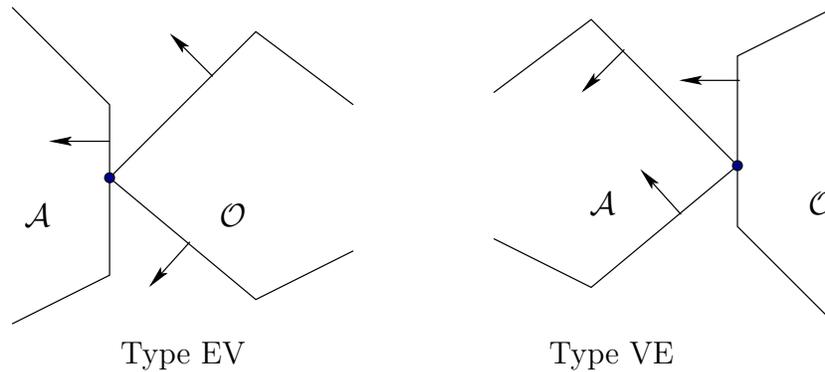Type EV   Type VE

Figure 4.16: Two different types of contact, each of which generates a different kind of $\mathcal{C}_{obs}$ edge [280, 657].

There are two different ways in which an edge of $\mathcal{C}_{obs}$ is generated, as shown in Figure 4.16 [282, 657]. *Type EV* contact refers to the case in which an edge of $\mathcal{A}$ is in contact with a vertex of $\mathcal{O}$. Type EV contacts contribute to n edges of $\mathcal{C}_{obs}$, once for each edge of $\mathcal{A}$. *Type VE* contact refers to the case in which a vertex of $\mathcal{A}$ is in contact with an edge of $\mathcal{O}$. This contributes to m edges of $\mathcal{C}_{obs}$. The relationships between the edge normals are also shown in Figure 4.16. For Type EV, the inward edge normal points between the outward edge normals of the obstacle edges that share the contact vertex. Likewise for Type VE, the outward edge normal of $\mathcal{O}$ points between the inward edge normals of $\mathcal{A}$.

Using the ordering shown in Figure 4.15b, Type EV contacts occur precisely when an edge normal of $\mathcal{A}$ is encountered, and Type VE contacts occur when an
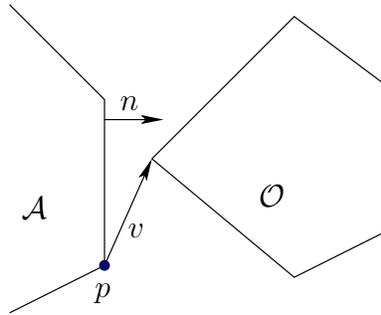
Figure 4.17: Contact occurs when n and v are perpendicular.

edge normal of $\mathcal{O}$ is encountered. The task is to determine the line equation for each occurrence. Consider the case of a Type EV contact; the Type VE contact can be handled in a similar manner. In addition to the constraint on the directions of the edge normals, the contact vertex of $\mathcal{O}$ must lie on the contact edge of $\mathcal{A}$. Recall that convex obstacles were constructed by the intersection of half-planes. Each edge of $\mathcal{C}_{obs}$ can be defined in terms of a supporting half-plane; hence, it is only necessary to determine whether the vertex of $\mathcal{O}$ lies on the line through the contact edge of $\mathcal{A}$. This condition occurs precisely as n and v are perpendicular, as shown in Figure 4.17, and yields the constraint $n \cdot v = 0$.

Note that the normal vector n does not depend on the configuration of $\mathcal{A}$ because the robot cannot rotate. The vector v, however, depends on the translation $q = (x_t, y_t)$ of the point p. Therefore, it is more appropriate to write the condition as $n \cdot v(x_t, y_t) = 0$. The transformation equations are linear for translation; hence, $n \cdot v(x_t, y_t) = 0$ is the equation of a line in $\mathcal{C}$. For example, if the coordinates of p are $(1, 2)$ for $\mathcal{A}(0, 0)$, then the expression for p at configuration (x_t, y_t) is $(1 + x_t, 2 + y_t)$. Let $f(x_t, y_t) = n \cdot v(x_t, y_t)$. Let $H = \{(x_t, y_t) \in \mathcal{C} \mid f(x_t, y_t) \leq 0\}$. Observe that any configurations not in H must lie in $\mathcal{C}_{free}$. The half-plane H is used to define one edge of $\mathcal{C}_{obs}$. The obstacle region $\mathcal{C}_{obs}$ can be completely characterized by intersecting the resulting half-planes for each of the Type EV and Type VE contacts. This yields a convex polygon in $\mathcal{C}$ that has $n + m$ sides, as expected.
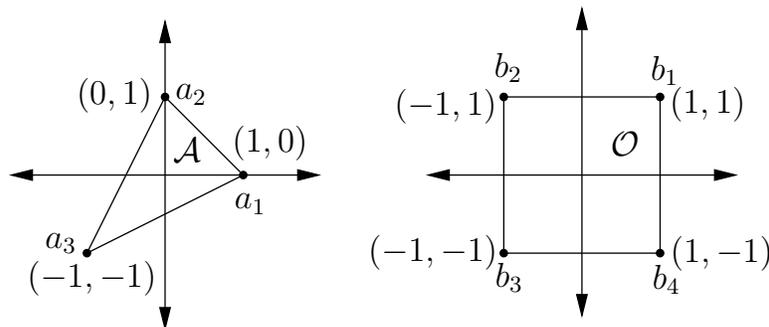

Figure 4.18: Consider constructing the obstacle region for this example.

Type	Vtx.	Edge	n	v	Half-Plane
VE	a_3	b_4-b_1	$[1,0]$	$[x_t - 2, y_t]$	$\{q \in \mathcal{C} \mid x_t - 2 \leq 0\}$
VE	a_3	b_1-b_2	$[0,1]$	$[x_t - 2, y_t - 2]$	$\{q \in \mathcal{C} \mid y_t - 2 \leq 0\}$
EV	b_2	a_3-a_1	$[1,-2]$	$[-x_t, 2 - y_t]$	$\{q \in \mathcal{C} \mid -x_t + 2y_t - 4 \leq 0\}$
VE	a_1	b_2-b_3	$[-1,0]$	$[2 + x_t, y_t - 1]$	$\{q \in \mathcal{C} \mid -x_t - 2 \leq 0\}$
EV	b_3	a_1-a_2	$[1,1]$	$[-1 - x_t, -y_t]$	$\{q \in \mathcal{C} \mid -x_t - y_t - 1 \leq 0\}$
VE	a_2	b_3-b_4	$[0,-1]$	$[x_t + 1, y_t + 2]$	$\{q \in \mathcal{C} \mid -y_t - 2 \leq 0\}$
EV	b_4	a_2-a_3	$[-2,1]$	$[2 - x_t, -y_t]$	$\{q \in \mathcal{C} \mid 2x_t - y_t - 4 \leq 0\}$

Figure 4.19: The various contact conditions are shown in the order as the edge normals appear around $\mathbb{S}^1$ (using inward normals for $\mathcal{A}$ and outward normals for $\mathcal{O}$).

Example 4.15 (The Boundary of $\mathcal{C}_{obs}$) Consider building a geometric model of $\mathcal{C}_{obs}$ for the robot and obstacle shown in Figure 4.18. Suppose that the orientation of $\mathcal{A}$ is fixed as shown, and $\mathcal{C} = \mathbb{R}^2$. In this case, $\mathcal{C}_{obs}$ will be a convex polygon with seven sides. The contact conditions that occur are shown in Figure 4.19. The ordering as the normals appear around $\mathbb{S}^1$ (using inward edge normals for $\mathcal{A}$ and outward edge normals for $\mathcal{O}$). The $\mathcal{C}_{obs}$ edges and their corresponding contact types are shown in Figure 4.19. ∎

A polyhedral C-space obstacle Most of the previous ideas generalize nicely for the case of a polyhedral robot that is capable of translation only in a 3D world that contains polyhedral obstacles. If $\mathcal{A}$ and $\mathcal{O}$ are convex polyhedra, the resulting $\mathcal{C}_{obs}$ is a convex polyhedron.

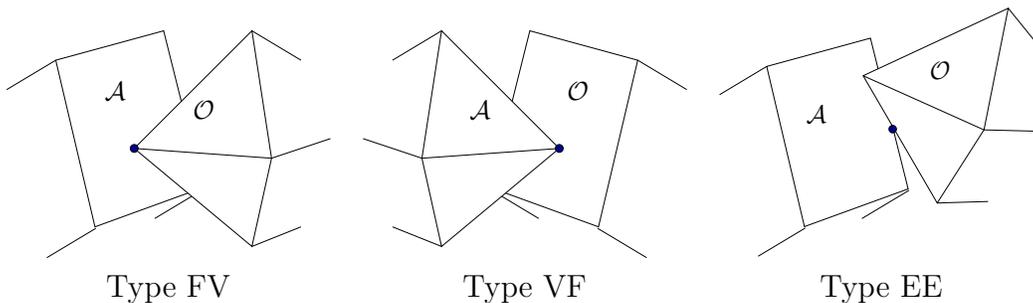Type FV   Type VF   Type EE

Figure 4.20: Three different types of contact, each of which generates a different kind of $\mathcal{C}_{obs}$ face.

There are three different kinds of contacts that each lead to half-spaces in $\mathcal{C}$:

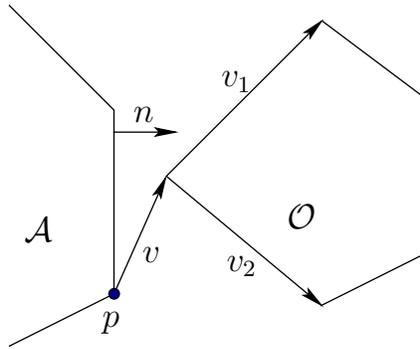1. **Type FV:** A face of $\mathcal{A}$ and a vertex of $\mathcal{O}$

Figure 4.21: An illustration to help in constructing $\mathcal{C}_{obs}$ when rotation is allowed.

2. **Type VF:** A vertex of $\mathcal{A}$ and a face of $\mathcal{O}$

3. **Type EE:** An edge of $\mathcal{A}$ and an edge of $\mathcal{O}$.

These are shown in Figure 4.20. Each half-space defines a face of the polyhedron, $\mathcal{C}_{obs}$. The representation of $\mathcal{C}_{obs}$ can be constructed in $O(n+m+k)$ time, in which n is the number of faces of $\mathcal{A}$, m is the number of faces of $\mathcal{O}$, and k is the number of faces of $\mathcal{C}_{obs}$, which is at most nm [411].

4.3.3   Explicitly Modeling $\mathcal{C}_{obs}$: The General Case

Unfortunately, the cases in which $\mathcal{C}_{obs}$ is polygonal or polyhedral are quite limited. Most problems yield extremely complicated C-space obstacles. One good point is that $\mathcal{C}_{obs}$ can be expressed using semi-algebraic models, for any robots and obstacles defined using semi-algebraic models, even after applying any of the transformations from Sections 3.2 to 3.4. It might not be true, however, for other kinds of transformations, such as warping a flexible material [32, 577].

Consider the case of a convex polygonal robot and a convex polygonal obstacle in a 2D world. Assume that any transformation in $SE(2)$ may be applied to $\mathcal{A}$; thus, $\mathcal{C} = \mathbb{R}^2 \times \mathbb{S}^1$ and $q = (x_t, y_t, \theta)$. The task is to define a set of algebraic primitives that can be combined to define $\mathcal{C}_{obs}$. Once again, it is important to distinguish between Type EV and Type VE contacts. Consider how to construct the algebraic primitives for the Type EV contacts; Type VE can be handled in a similar manner.

For the translation-only case, we were able to determine all of the Type EV contacts by sorting the edge normals. With rotation, the ordering of edge normals depends on θ. This implies that the applicability of a Type EV contact depends on θ, the robot orientation. Recall the constraint that the inward normal of $\mathcal{A}$ must point between the outward normals of the edges of $\mathcal{O}$ that contain the vertex of contact, as shown in Figure 4.21. This constraint can be expressed in terms of inner products using the vectors v_1 and v_2. The statement regarding the directions of the normals can equivalently be formulated as the statement that the angle between n and v_1, and between n and v_2, must each be less than $\pi/2$. Using inner products,

this implies that $n \cdot v_1 \geq 0$ and $n \cdot v_2 \geq 0$. As in the translation case, the condition $n \cdot v = 0$ is required for contact. Observe that n now depends on θ. For any $q \in \mathcal{C}$, if $n(\theta) \cdot v_1 \geq 0$, $n(\theta) \cdot v_2 \geq 0$, and $n(\theta) \cdot v(q) > 0$, then $q \in \mathcal{C}_{free}$. Let H_f denote the set of configurations that satisfy these conditions. These conditions imply that a point is in $\mathcal{C}_{free}$. Furthermore, any other Type EV and Type VE contacts could imply that more points are in $\mathcal{C}_{free}$. Ordinarily, $H_f \subset \mathcal{C}_{free}$, which implies that the complement, $\mathcal{C} \setminus H_f$, is a superset of $\mathcal{C}_{obs}$ (thus, $\mathcal{C}_{obs} \subset \mathcal{C} \setminus H_f$). Let $H_A = \mathcal{C} \setminus H_f$. Using the primitives

$$H_1 = \{q \in \mathcal{C} \mid n(\theta) \cdot v_1 \leq 0\}, \tag{4.39}$$

$$H_2 = \{q \in \mathcal{C} \mid n(\theta) \cdot v_2 \leq 0\}, \tag{4.40}$$

and

$$H_3 = \{q \in \mathcal{C} \mid n(\theta) \cdot v(q) \leq 0\}, \tag{4.41}$$

let $H_A = H_1 \cup H_2 \cup H_3$.

It is known that $\mathcal{C}_{obs} \subseteq H_A$, but H_A may contain points in $\mathcal{C}_{free}$. The situation is similar to what was explained in Section 3.1.1 for building a model of a convex polygon from half-planes. In the current setting, it is only known that any configuration outside of H_A must be in $\mathcal{C}_{free}$. If H_A is intersected with all other corresponding sets for each possible Type EV and Type VE contact, then the result is $\mathcal{C}_{obs}$. Each contact has the opportunity to remove a portion of $\mathcal{C}_{free}$ from consideration. Eventually, enough pieces of $\mathcal{C}_{free}$ are removed so that the only configurations remaining must lie in $\mathcal{C}_{obs}$. For any Type EV contact, $(H_1 \cup H_2) \setminus H_3 \subseteq \mathcal{C}_{free}$. A similar statement can be made for Type VE contacts. A logical predicate, similar to that defined in Section 3.1.1, can be constructed to determine whether $q \in \mathcal{C}_{obs}$ in time that is linear in the number of $\mathcal{C}_{obs}$ primitives.

One important issue remains. The expression $n(\theta)$ is not a polynomial because of the $\cos \theta$ and $\sin \theta$ terms in the rotation matrix of $SO(2)$. If polynomials could be substituted for these expressions, then everything would be fixed because the expression of the normal vector (not a unit normal) and the inner product are both linear functions, thereby transforming polynomials into polynomials. Such a substitution can be made using stereographic projection (see [588]); however, a simpler approach is to use complex numbers to represent rotation. Recall that when $a + bi$ is used to represent rotation, each rotation matrix in $SO(2)$ is represented as (4.18), and the 3×3 homogeneous transformation matrix becomes

$$T(a, b, x_t, y_t) = \begin{pmatrix} a & -b & x_t \\ b & a & y_t \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.42}$$

Using this matrix to transform a point $[x \ y \ 1]$ results in the point coordinates $(ax - by + x_t, bx + ay + y_t)$. Thus, any transformed point on $\mathcal{A}$ is a linear function of a, b, x_t, and y_t.

This was a simple trick to make a nice, linear function, but what was the cost? The dependency is now on a and b instead of θ. This appears to increase the

dimension of $\mathcal{C}$ from 3 to 4, and $\mathcal{C} = \mathbb{R}^4$. However, an algebraic primitive must be added that constrains a and b to lie on the unit circle.

By using complex numbers, primitives in $\mathbb{R}^4$ are obtained for each Type EV and Type VE contact. By defining $\mathcal{C} = \mathbb{R}^4$, the following algebraic primitives are obtained for a Type EV contact:

$$H_1 = \{(x_t, y_t, a, b) \in \mathcal{C} \mid n(x_t, y_t, a, b) \cdot v_1 \leq 0\}, \tag{4.43}$$

$$H_2 = \{(x_t, y_t, a, b) \in \mathcal{C} \mid n(x_t, y_t, a, b) \cdot v_2 \leq 0\}, \tag{4.44}$$

and

$$H_3 = \{(x_t, y_t, a, b) \in \mathcal{C} \mid n(x_t, y_t, a, b) \cdot v(x_t, y_t, a, b) \leq 0\}. \tag{4.45}$$

This yields $H_A = H_1 \cup H_2 \cup H_3$. To preserve the correct $\mathbb{R}^2 \times \mathbb{S}^1$ topology of $\mathcal{C}$, the set

$$H_s = \{(x_t, y_t, a, b) \in \mathcal{C} \mid a^2 + b^2 - 1 = 0\} \tag{4.46}$$

is intersected with H_A. The set H_s remains fixed over all Type EV and Type VE contacts; therefore, it only needs to be considered once.

Example 4.16 (A Nonlinear Boundary for $\mathcal{C}_{obs}$) Consider adding rotation to the model described in Example 4.15. In this case, all possible contacts between pairs of edges must be considered. For this example, there are 12 Type EV contacts and 12 Type VE contacts. Each contact produces 3 algebraic primitives. With the inclusion of H_s, this simple example produces 73 primitives! Rather than construct all of these, we derive the primitives for a single contact. Consider the Type VE contact between a_3 and b_4-b_1. The outward edge normal n remains fixed at $n = [1, 0]$. The vectors v_1 and v_2 are derived from the edges adjacent to a_3, which are a_3-a_2 and a_3-a_1. Note that each of a_1, a_2, and a_3 depend on the configuration. Using the 2D homogeneous transformation (3.35), a_1 at configuration (x_t, y_t, θ) is $(\cos\theta + x_t, \sin\theta + y_t)$. Using $a + bi$ to represent rotation, the expression of a_1 becomes $(a + x_t, b + y_t)$. The expressions of a_2 and a_3 are $(-b + x_t, a + y_t)$ and $(-a + b + x_t, -b - a + y_t)$, respectively. It follows that $v_1 = a_2 - a_3 = [a - 2b, 2a + b]$ and $v_2 = a_1 - a_3 = [2a - b, a + 2b]$. Note that v_1 and v_2 depend only on the orientation of $\mathcal{A}$, as expected. Assume that v is drawn from b_4 to a_3. This yields $v = a_3 - b_4 = [-a + b + x_t - 1, -a - b + y_t + 1]$. The inner products $v_1 \cdot n$, $v_2 \cdot n$, and $v \cdot n$ can easily be computed to form H_1, H_2, and H_3 as algebraic primitives.

One interesting observation can be made here. The only nonlinear primitive is $a^2 + b^2 = 1$. Therefore, $\mathcal{C}_{obs}$ can be considered as a linear polytope (like a polyhedron, but one dimension higher) in $\mathbb{R}^4$ that is intersected with a cylinder.   ∎

3D rigid bodies   For the case of a 3D rigid body to which any transformation in $SE(3)$ may be applied, the same general principles apply. The quaternion parameterization once again becomes the right way to represent $SO(3)$ because using (4.20) avoids all trigonometric functions in the same way that (4.18) avoided them for $SO(2)$. Unfortunately, (4.20) is not linear in the configuration variables,

as it was for (4.18), but it is at least polynomial. This enables semi-algebraic models to be formed for $\mathcal{C}_{obs}$. Type FV, VF, and EE contacts arise for the $SE(3)$ case. From all of the contact conditions, polynomials that correspond to each patch of $\mathcal{C}_{obs}$ can be made. These patches are polynomials in seven variables: x_t, y_t, z_t, a, b, c, and d. Once again, a special primitive must be intersected with all others; here, it enforces the constraint that unit quaternions are used. This reduces the dimension from 7 back down to 6. Also, constraints should be added to throw away half of $\mathbb{S}^3$, which is redundant because of the identification of antipodal points on $\mathbb{S}^3$.

Chains and trees of bodies For chains and trees of bodies, the ideas are conceptually the same, but the algebra becomes more cumbersome. Recall that the transformation for each link is obtained by a product of homogeneous transformation matrices, as given in (3.53) and (3.57) for the 2D and 3D cases, respectively. If the rotation part is parameterized using complex numbers for $SO(2)$ or quaternions for $SO(3)$, then each matrix consists of polynomial entries. After the matrix product is formed, polynomial expressions in terms of the configuration variables are obtained. Therefore, a semi-algebraic model can be constructed. For each link, all of the contact types need to be considered. Extrapolating from Examples 4.15 and 4.16, you can imagine that no human would ever want to do all of that by hand, but it can at least be automated. The ability to construct this representation automatically is also very important for the existence of theoretical algorithms that solve the motion planning problem combinatorially; see Section 6.4.

If the kinematic chains were formulated for $\mathcal{W} = \mathbb{R}^3$ using the DH parameterization, it may be inconvenient to convert to the quaternion representation. One way to avoid this is to use complex numbers to represent each of the θ_i and α_i variables that appear as configuration variables. This can be accomplished because only cos and sin functions appear in the transformation matrices. They can be replaced by the real and imaginary parts, respectively, of a complex number. The dimension will be increased, but this will be appropriately reduced after imposing the constraints that all complex numbers must have unit magnitude.

4.4 Closed Kinematic Chains

This section continues the discussion from Section 3.4. Suppose that a collection of links is arranged in a way that forms loops. In this case, the C-space becomes much more complicated because the joint angles must be chosen to ensure that the loops remain closed. This leads to constraints such as that shown in (3.80) and Figure 3.26, in which some links must maintain specified positions relative to each other. Consider the set of all configurations that satisfy such constraints. Is this a manifold? It turns out, unfortunately, that the answer is generally *no*. However, the C-space belongs to a nice family of spaces from algebraic geometry

called *varieties*. Algebraic geometry deals with characterizing the solution sets of polynomials. As seen so far in this chapter, all of the kinematics can be expressed as polynomials. Therefore, it may not be surprising that the resulting constraints are a system of polynomials whose solution set represents the C-space for closed kinematic linkages. Although the algebraic varieties considered here need not be manifolds, they can be decomposed into a finite collection of manifolds that fit together nicely.[11]

Unfortunately, a parameterization of the variety that arises from closed chains is available in only a few simple cases. Even the topology of the variety is extremely difficult to characterize. To make matters worse, it was proved in [489] that for every closed, bounded real algebraic variety that can be embedded in $\mathbb{R}^n$, there exists a linkage whose C-space is homeomorphic to it. These troubles imply that most of the time, motion planning algorithms need to work directly with implicit polynomials. For the algebraic methods of Section 6.4.2, this does not pose any conceptual difficulty because the methods already work directly with polynomials. Sampling-based methods usually rely on the ability to efficiently sample configurations, which cannot be easily adapted to a variety without a parameterization. Section 7.4 covers recent methods that extend sampling-based planning algorithms to work for varieties that arise from closed chains.

4.4.1 Mathematical Concepts

To understand varieties, it will be helpful to have definitions of polynomials and their solutions that are more formal than the presentation in Chapter 3.

Fields Polynomials are usually defined over a *field*, which is another object from algebra. A field is similar to a group, but it has more operations and axioms. The definition is given below, and while reading it, keep in mind several familiar examples of fields: the rationals, $\mathbb{Q}$; the reals, $\mathbb{R}$; and the complex plane, $\mathbb{C}$. You may verify that these fields satisfy the following six axioms.

A *field* is a set $\mathbb{F}$ that has two binary operations, $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ (called *multiplication*) and $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ (called *addition*), for which the following axioms are satisfied:

1. (**Associativity**) For all $a, b, c \in \mathbb{F}$, $(a+b)+c = a+(b+c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

2. (**Commutativity**) For all $a, b \in \mathbb{F}$, $a + b = b + a$ and $a \cdot b = b \cdot a$.

3. (**Distributivity**) For all $a, b, c \in \mathbb{F}$, $a \cdot (b + c) = a \cdot b + a \cdot c$.

4. (**Identities**) There exist $0, 1 \in \mathbb{F}$, such that $a + 0 = a \cdot 1 = a$ for all $a \in \mathbb{F}$.

5. (**Additive Inverses**) For every $a \in \mathbb{F}$, there exists some $b \in \mathbb{F}$ such that $a + b = 0$.

---

[11]This is called a Whitney stratification [173, 968].

6. (**Multiplicative Inverses**) For every $a \in F$, except $a = 0$, there exists some $c \in \mathbb{F}$ such that $a \cdot c = 1$.

Compare these axioms to the group definition from Section 4.2.1. Note that a field can be considered as two different kinds of groups, one with respect to multiplication and the other with respect to addition. Fields additionally require commutativity; hence, we cannot, for example, build a field from quaternions. The distributivity axiom appears because there is now an interaction between two different operations, which was not possible with groups.

Polynomials Suppose there are n variables, $x_1, x_2, \ldots, x_n$. A *monomial* over a field $\mathbb{F}$ is a product of the form

$$x_1^{d_1} \cdot x_2^{d_2} \cdots \cdot x_n^{d_n}, \tag{4.47}$$

in which all of the exponents d_1, d_2, ..., d_n are positive integers. The *total degree* of the monomial is $d_1 + \cdots + d_n$.

A *polynomial f* in variables $x_1, \ldots, x_n$ with coefficients in $\mathbb{F}$ is a finite linear combination of monomials that have coefficients in $\mathbb{F}$. A polynomial can be expressed as

$$\sum_{i=1}^{m} c_i m_i, \tag{4.48}$$

in which m_i is a monomial as shown in (4.47), and $c_i \in \mathbb{F}$ is a *coefficient*. If $c_i \neq 0$, then each $c_i m_i$ is called a *term*. Note that the exponents d_i may be different for every term of f. The *total degree of f* is the maximum total degree among the monomials of the terms of f. The set of all polynomials in $x_1, \ldots, x_n$ with coefficients in $\mathbb{F}$ is denoted by $\mathbb{F}[x_1, \ldots, x_n]$.

Example 4.17 (Polynomials) The definitions correspond exactly to our intuitive notion of a polynomial. For example, suppose $\mathbb{F} = \mathbb{Q}$. An example of a polynomial in $\mathbb{Q}[x_1, x_2, x_3]$ is

$$x_1^4 - \tfrac{1}{2} x_1 x_2 x_3^3 + x_1^2 x_2^2 + 4. \tag{4.49}$$

Note that 1 is a valid monomial; hence, any element of $\mathbb{F}$ may appear alone as a term, such as the $4 \in \mathbb{Q}$ in the polynomial above. The total degree of (4.49) is 5 due to the second term. An equivalent polynomial may be written using nicer variables. Using x, y, and z as variables yields

$$x^4 - \tfrac{1}{2} xyz^3 + x^2 y^2 + 4, \tag{4.50}$$

which belongs to $\mathbb{Q}[x, y, z]$. ∎

The set $\mathbb{F}[x_1, \ldots, x_n]$ of polynomials is actually a group with respect to addition; however, it is not a field. Even though polynomials can be multiplied, some polynomials do not have a multiplicative inverse. Therefore, the set $\mathbb{F}[x_1, \ldots, x_n]$ is often referred to as a *commutative ring* of polynomials. A commutative ring is a set with two operations for which every axiom for fields is satisfied except the last one, which would require a multiplicative inverse.

Varieties   For a given field $\mathbb{F}$ and positive integer n, the n-dimensional *affine space* over $\mathbb{F}$ is the set

$$\mathbb{F}^n = \{(c_1, \ldots, c_n) \mid c_1, \ldots, c_n \in \mathbb{F}\}. \tag{4.51}$$

For our purposes in this section, an affine space can be considered as a vector space (for an exact definition, see [438]). Thus, $\mathbb{F}^n$ is like a vector version of the scalar field $\mathbb{F}$. Familiar examples of this are $\mathbb{Q}^n$, $\mathbb{R}^n$, and $\mathbb{C}^n$.

A polynomial in $f \in \mathbb{F}[x_1, \ldots, x_n]$ can be converted into a function,

$$f : \mathbb{F}^n \to \mathbb{F}, \tag{4.52}$$

by substituting elements of $\mathbb{F}$ for each variable and evaluating the expression using the field operations. This can be written as $f(a_1, \ldots, a_n) \in \mathbb{F}$, in which each a_i denotes an element of $\mathbb{F}$ that is substituted for the variable x_i.

We now arrive at an interesting question. For a given f, what are the elements of $\mathbb{F}^n$ such that $f(a_1, \ldots, a_n) = 0$? We could also ask the question for some nonzero element, but notice that this is not necessary because the polynomial may be redefined to formulate the question using 0. For example, what are the elements of $\mathbb{R}^2$ such that $x^2 + y^2 = 1$? This familiar equation for $\mathbb{S}^1$ can be reformulated to yield: What are the elements of $\mathbb{R}^2$ such that $x^2 + y^2 - 1 = 0$?

Let $\mathbb{F}$ be a field and let $\{f_1, \ldots, f_k\}$ be a set of polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. The set

$$V(f_1, \ldots, f_k) = \{(a_1, \ldots, a_n) \in \mathbb{F} \mid f_i(a_1, \ldots, a_n) = 0 \text{ for all } 1 \leq i \leq k\} \tag{4.53}$$

is called the *(affine) variety* defined by $f_1, \ldots, f_k$. One interesting fact is that unions and intersections of varieties are varieties. Therefore, they behave like the semi-algebraic sets from Section 3.1.2, but for varieties only equality constraints are allowed. Consider the varieties $V(f_1, \ldots, f_k)$ and $V(g_1, \ldots, g_l)$. Their intersection is given by

$$V(f_1, \ldots, f_k) \cap V(g_1, \ldots, g_l) = V(f_1, \ldots, f_k, g_1, \ldots, g_l), \tag{4.54}$$

because each element of $\mathbb{F}^n$ must produce a 0 value for each of the polynomials in $\{f_1, \ldots, f_k, g_1, \ldots, g_l\}$.

To obtain unions, the polynomials simply need to be multiplied. For example, consider the varieties $V_1, V_2 \subset \mathbb{F}$ defined as

$$V_1 = \{(a_1, \ldots, a_n) \in \mathbb{F} \mid f_1(a_1, \ldots, a_n) = 0\} \tag{4.55}$$

and

$$V_2 = \{(a_1, \ldots, a_n) \in \mathbb{F} \mid f_2(a_1, \ldots, a_n) = 0\}. \tag{4.56}$$

The set $V_1 \cup V_2 \subset \mathbb{F}$ is obtained by forming the polynomial $f = f_1 f_2$. Note that $f(a_1, \ldots, a_n) = 0$ if either $f_1(a_1, \ldots, a_n) = 0$ or $f_2(a_1, \ldots, a_n) = 0$. Therefore, $V_1 \cup V_2$ is a variety. The varieties V_1 and V_2 were defined using a single polynomial, but the same idea applies to any variety. All pairs of the form $f_i g_j$ must appear in the argument of $V(\cdot)$ if there are multiple polynomials.

4.4.2 Kinematic Chains in $\mathbb{R}^2$

To illustrate the concepts it will be helpful to study a simple case in detail. Let $\mathcal{W} = \mathbb{R}^2$, and suppose there is a chain of links, $\mathcal{A}_1$, ..., $\mathcal{A}_n$, as considered in Example 3.3 for $n = 3$. Suppose that the first link is attached at the origin of $\mathcal{W}$ by a revolute joint, and every other link, $\mathcal{A}_i$ is attached to $\mathcal{A}_{i-1}$ by a revolute joint. This yields the C-space

$$\mathcal{C} = \mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 = \mathbb{T}^n, \tag{4.57}$$

which is the n-dimensional torus.

Two links If there are two links, $\mathcal{A}_1$ and $\mathcal{A}_2$, then the C-space can be nicely visualized as a square with opposite faces identified. Each coordinate, θ_1 and θ_2, ranges from 0 to 2π, for which $0 \sim 2\pi$. Suppose that each link has length 1. This yields $a_1 = 1$. A point $(x, y) \in \mathcal{A}_2$ is transformed as

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 & 1 \\ \sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{4.58}$$

To obtain polynomials, the technique from Section 4.2.2 is applied to replace the trigonometric functions using $a_i = \cos\theta_i$ and $b_i = \sin\theta_i$, subject to the constraint $a_i^2 + b_i^2 = 1$. This results in

$$\begin{pmatrix} a_1 & -b_1 & 0 \\ b_1 & a_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_2 & -b_2 & 1 \\ b_2 & a_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{4.59}$$

for which the constraints $a_i^2 + b_i^2 = 1$ for $i = 1, 2$ must be satisfied. This preserves the torus topology of $\mathcal{C}$, but now the C-space is embedded in $\mathbb{R}^4$. The coordinates of each point are $(a_1, b_1, a_2, b_2) \in \mathbb{R}^4$; however, there are only two degrees of freedom because each a_i, b_i pair must lie on a unit circle.

Multiplying the matrices in (4.59) yields the polynomials, $f_1, f_2 \in \mathbb{R}[a_1, b_1, a_2, b_2]$,

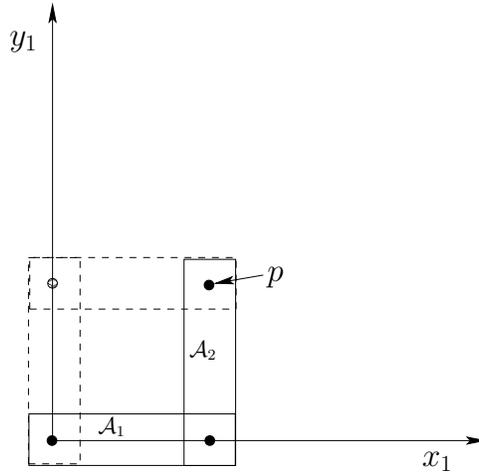$$f_1 = xa_1a_2 - ya_1b_2 - xb_1b_2 + ya_2b_1 + a_1 \tag{4.60}$$

Figure 4.22: Two configurations hold the point p at $(1, 1)$.

and

$$f_2 = -ya_1a_2 + xa_1b_2 + xa_2b_1 - yb_1b_2 + b_1, \qquad (4.61)$$

for the x and y coordinates, respectively. Note that the polynomial variables are configuration parameters; x and y are not polynomial variables. For a given point $(x, y) \in \mathcal{A}_2$, all coefficients are determined.

A zero-dimensional variety Now a kinematic closure constraint will be imposed. Fix the point $(1, 0)$ in the body frame of $\mathcal{A}_2$ at $(1, 1)$ in $\mathcal{W}$. This yields the constraints

$$f_1 = a_1a_2 - b_1b_2 + a_1 = 1 \qquad (4.62)$$

and

$$f_2 = a_1b_2 + a_2b_1 + b_1 = 1, \qquad (4.63)$$

by substituting $x = 1$ and $y = 0$ into (4.60) and (4.61). This yields the variety

$$V(a_1a_2 - b_1b_2 + a_1 - 1, a_1b_2 + a_2b_1 + b_1 - 1, a_1^2 + b_1^2 - 1, a_2^2 + b_2^2 - 1), \qquad (4.64)$$

which is a subset of $\mathbb{R}^4$. The polynomials are slightly modified because each constraint must be written in the form $f = 0$.

Although (4.64) represents the constrained configuration space for the chain of two links, it is not very explicit. Without an explicit characterization (i.e., a parameterization), it complicates motion planning. From Figure 4.22 it can be seen that there are only two solutions. These occur for $\theta_1 = 0$, $\theta_2 = \pi/2$ and $\theta_1 = \pi/2$, $\theta_2 = -\pi/2$. In terms of the polynomial variables, (a_1, b_1, a_2, b_2), the two solutions are $(1, 0, 0, 1)$ and $(0, 1, 0, -1)$. These may be substituted into each polynomial in (4.64) to verify that 0 is obtained. Thus, the variety represents two points in $\mathbb{R}^4$. This can also be interpreted as two points on the torus, $\mathbb{S}^1 \times \mathbb{S}^1$.

It might not be surprising that the set of solutions has dimension zero because there are four independent constraints, shown in (4.64), and four variables. Depending on the choices, the variety may be empty. For example, it is physically impossible to bring the point $(1,0) \in \mathcal{A}_2$ to $(1000,0) \in \mathcal{W}$.

A one-dimensional variety The most interesting and complicated situations occur when there is a continuum of solutions. For example, if one of the constraints is removed, then a one-dimensional set of solutions can be obtained. Suppose only one variable is constrained for the example in Figure 4.22. Intuitively, this should yield a one-dimensional variety. Set the x coordinate to 0, which yields

$$a_1 a_2 - b_1 b_2 + a_1 = 0, \tag{4.65}$$

and allow any possible value for y. As shown in Figure 4.23a, the point p must follow the y-axis. (This is equivalent to a three-bar linkage that can be constructed by making a third joint that is prismatic and forced to stay along the y-axis.) Figure 4.23b shows the resulting variety $V(a_1 a_2 - b_1 b_2 + a_1)$ but plotted in $\theta_1 - \theta_2$ coordinates to reduce the dimension from 4 to 2 for visualization purposes. To correctly interpret the figures in Figure 4.23, recall that the topology is $\mathbb{S}^1 \times \mathbb{S}^1$, which means that the top and bottom are identified, and also the sides are identified. The center of Figure 4.23b, which corresponds to $(\theta_1, \theta_2) = (\pi, \pi)$, prevents the variety from being a manifold. The resulting space is actually homeomorphic to two circles that touch at a point. Thus, even with such a simple example, the nice manifold structure may disappear. Observe that at (π, π) the links are completely overlapped, and the point p of $\mathcal{A}_2$ is placed at $(0,0)$ in $\mathcal{W}$. The horizontal line in Figure 4.23b corresponds to keeping the two links overlapping and swinging them around together by varying θ_1. The diagonal lines correspond to moving along configurations such as the one shown in Figure 4.23a. Note that the links and the y-axis always form an isosceles triangle, which can be used to show that the solution set is any pair of angles, θ_1, θ_2 for which $\theta_2 = \pi - \theta_1$. This is the reason why the diagonal curves in Figure 4.23b are linear. Figures 4.23c and 4.23d show the varieties for the constraints

$$a_1 a_2 - b_1 b_2 + a_1 = \tfrac{1}{8}, \tag{4.66}$$

and

$$a_1 a_2 - b_1 b_2 + a_1 = 1, \tag{4.67}$$

respectively. In these cases, the point $(0,1)$ in $\mathcal{A}_2$ must follow the $x = 1/8$ and $x = 1$ axes, respectively. The varieties are manifolds, which are homeomorphic to $\mathbb{S}^1$. The sequence from Figure 4.23b to 4.23d can be imagined as part of an animation in which the variety shrinks into a small circle. Eventually, it shrinks to a point for the case $a_1 a_2 - b_1 b_2 + a_1 = 2$, because the only solution is when $\theta_1 = \theta_2 = 0$. Beyond this, the variety is the empty set because there are no solutions. Thus, by allowing one constraint to vary, four different topologies are obtained: 1) two circles joined at a point, 2) a circle, 3) a point, and 4) the empty set.
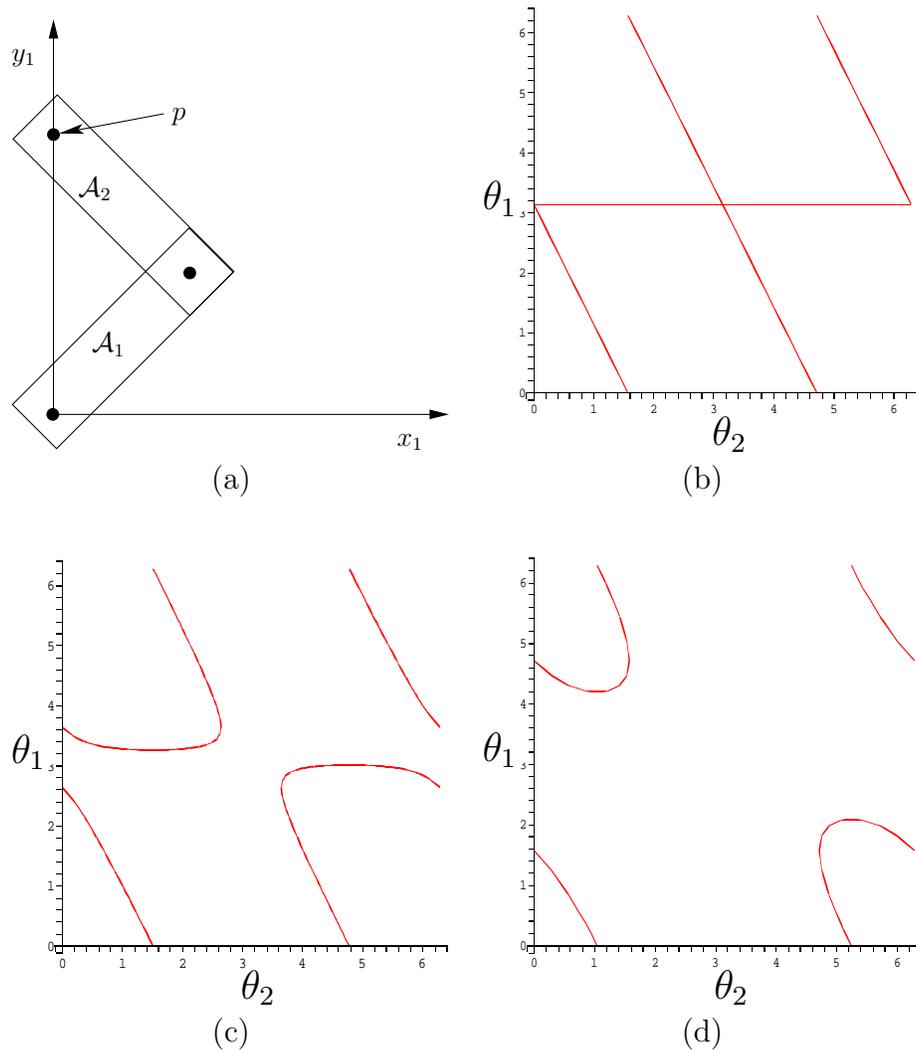
(a)

(b)

(c)

(d)

Figure 4.23: A single constraint was added to the point p on $\mathcal{A}_2$, as shown in (a). The curves in (b), (c), and (d) depict the variety for the cases of $f_1 = 0$, $f_1 = 1/8$, and $f_1 = 1$, respectively.

Three links Since visualization is still possible with one more dimension, suppose there are three links, $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$. The C-space can be visualized as a 3D cube with opposite faces identified. Each coordinate θ_i ranges from 0 to 2π, for which $0 \sim 2\pi$. Suppose that each link has length 1 to obtain $a_1 = a_2 = 1$. A point $(x, y) \in \mathcal{A}_3$ is transformed as

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 & 10 \\ \sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_3 & -\sin\theta_3 & 10 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$
(4.68)

To obtain polynomials, let $a_i = \cos\theta_i$ and $b_i = \sin\theta_i$, which results in

$$\begin{pmatrix} a_1 & -b_1 & 0 \\ b_1 & a_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_2 & -b_2 & 1 \\ b_2 & a_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_3 & -b_3 & 1 \\ b_3 & a_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$
(4.69)

for which the constraints $a_i^2 + b_i^2 = 1$ for $i = 1, 2, 3$ must also be satisfied. This preserves the torus topology of $\mathcal{C}$, but now it is embedded in $\mathbb{R}^6$. Multiplying the matrices yields the polynomials $f_1, f_2 \in \mathbb{R}[a_1, b_1, a_2, b_2, a_3, b_3]$, defined as

$$f_1 = 2a_1a_2a_3 - a_1b_2b_3 + a_1a_2 - 2b_1b_2a_3 - b_1a_2b_3 + a_1,$$
(4.70)

and

$$f_2 = 2b_1a_2a_3 - b_1b_2b_3 + b_1a_2 + 2a_1b_2a_3 + a_1a_2b_3,$$
(4.71)

for the x and y coordinates, respectively.

Again, consider imposing a single constraint,

$$2a_1a_2a_3 - a_1b_2b_3 + a_1a_2 - 2b_1b_2a_3 - b_1a_2b_3 + a_1 = 0,$$
(4.72)

which constrains the point $(1, 0) \in \mathcal{A}_3$ to traverse the y-axis. The resulting variety is an interesting manifold, depicted in Figure 4.24 (remember that the sides of the cube are identified).

Increasing the required f_1 value for the constraint on the final point causes the variety to shrink. Snapshots for $f_1 = 7/8$ and $f_1 = 2$ are shown in Figure 4.25. At $f_1 = 1$, the variety is not a manifold, but it then changes to $\mathbb{S}^2$. Eventually, this sphere is reduced to a point at $f_1 = 3$, and then for $f_1 > 3$ the variety is empty.

Instead of the constraint $f_1 = 0$, we could instead constrain the y coordinate of p to obtain $f_2 = 0$. This yields another 2D variety. If both constraints are enforced simultaneously, then the result is the intersection of the two original varieties. For example, suppose $f_1 = 1$ and $f_2 = 0$. This is equivalent to a kind of *four-bar mechanism* [310], in which the fourth link, $\mathcal{A}_4$, is fixed along the x-axis from 0 to 1. The resulting variety,

$$V(2a_1a_2a_3 - a_1b_2b_3 + a_1a_2 - 2b_1b_2a_3 - b_1a_2b_3 + a_1 - 1,$$
$$2b_1a_2a_3 - b_1b_2b_3 + b_1a_2 + 2a_1b_2a_3 + a_1a_2b_3),$$
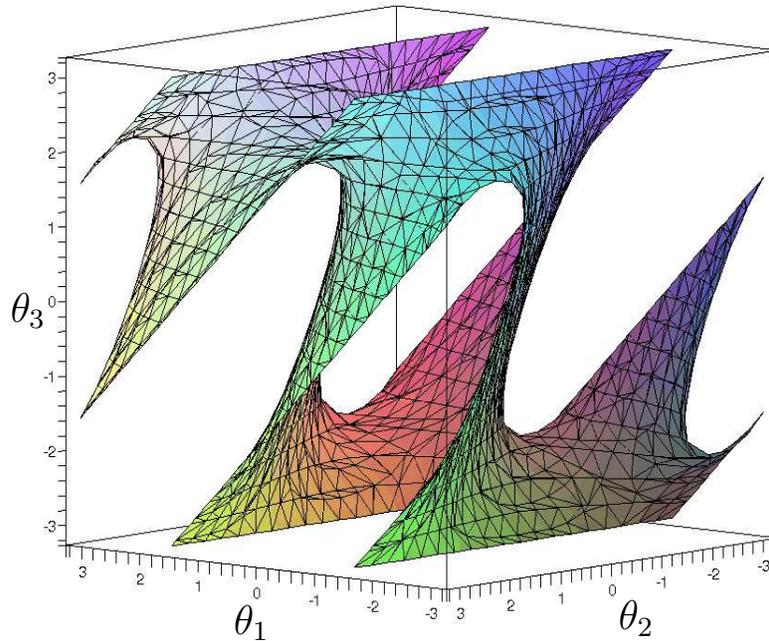(4.73)

Figure 4.24: The variety for the three-link chain with $f_1 = 0$ is a 2D manifold.

is depicted in Figure 4.26. Using the $\theta_1, \theta_2, \theta_3$ coordinates, the solution may be easily parameterized as a collection of line segments. For all $t \in [0, \pi]$, there exist solution points at $(0, 2t, \pi)$, $(t, 2\pi - t, \pi + t)$, $(2\pi - t, t, \pi - t)$, $(2\pi - t, \pi, \pi + t)$, and $(t, \pi, \pi - t)$. Note that once again the variety is not a manifold. A family of interesting varieties can be generated for the four-bar mechanism by selecting different lengths for the links. The topologies of these mechanisms have been determined for 2D and a 3D extension that uses spherical joints (see [698]).

4.4.3 Defining the Variety for General Linkages

We now describe a general methodology for defining the variety. Keeping the previous examples in mind will help in understanding the formulation. In the general case, each constraint can be thought of as a statement of the form:

> *The ith coordinate of a point $p \in \mathcal{A}_j$ needs to be held at the value x in the body frame of $\mathcal{A}_k$.*

For the variety in Figure 4.23b, the first coordinate of a point $p \in \mathcal{A}_2$ was held at the value 0 in $\mathcal{W}$ in the body frame of $\mathcal{A}_1$. The general form must also allow a point to be fixed with respect to the body frames of links other than $\mathcal{A}_1$; this did not occur for the example in Section 4.4.2

Suppose that n links, $\mathcal{A}_1, \ldots, \mathcal{A}_n$, move in $\mathcal{W} = \mathbb{R}^2$ or $\mathcal{W} = \mathbb{R}^3$. One link, $\mathcal{A}_1$ for convenience, is designated as the root as defined in Section 3.4. Some links

$$f_1 = 7/8 \qquad\qquad\qquad f_1 = 2$$

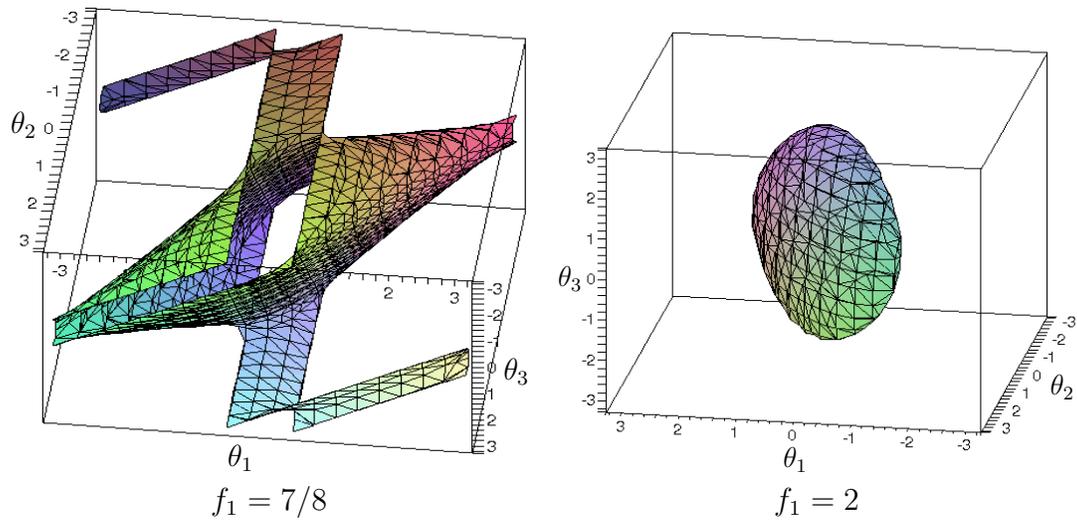Figure 4.25: If $f_1 > 0$, then the variety shrinks. If $1 < p < 3$, the variety is a sphere. At $f_1 = 0$ it is a point, and for $f_1 > 3$ it completely vanishes.

are attached in pairs to form joints. A *linkage graph*, $\mathcal{G}(V, E)$, is constructed from the links and joints. Each vertex of $\mathcal{G}$ represents a link in L. Each edge in $\mathcal{G}$ represents a joint. This definition may seem somewhat backward, especially in the plane because links often look like edges and joints look like vertices. This alternative assignment is also possible, but it is not easy to generalize to the case of a single link that has more than two joints. If more than two links are attached at the same point, each generates an edge of $\mathcal{G}$.

The steps to determine the polynomial constraints that express the variety are as follows:

1. Define the linkage graph $\mathcal{G}$ with one vertex per link and one edge per joint. If a joint connects more than two bodies, then one body must be designated as a junction. See Figures 4.27 and 4.28a. In Figure 4.28, links 4, 13, and 23 are designated as junctions in this way.

2. Designate one link as the root, $\mathcal{A}_1$. This link may either be fixed in $\mathcal{W}$, or transformations may be applied. In the latter case, the set of transformations could be $SE(2)$ or $SE(3)$, depending on the dimension of $\mathcal{W}$. This enables the entire linkage to move independently of its internal motions.

3. Eliminate the loops by constructing a spanning tree T of the linkage graph, $\mathcal{G}$. This implies that every vertex (or link) is reachable by a path from the root). Any spanning tree may be used. Figure 4.28b shows a resulting spanning tree after deleting the edges shown with dashed lines.

4. Apply the techniques of Section 3.4 to assign body frames and transformations to the resulting tree of links.
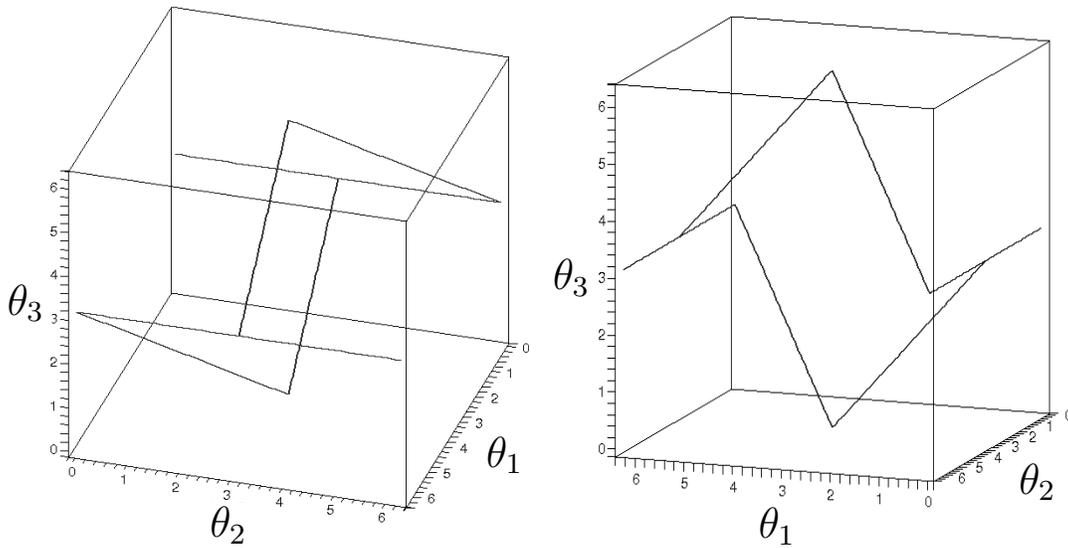
Figure 4.26: If two constraints, $f_1 = 1$ and $f_2 = 0$, are imposed, then the varieties are intersected to obtain a 1D set of solutions. The example is equivalent to a well-studied four-bar mechanism.

5. For each edge of $\mathcal{G}$ that does not appear in T, write a set of constraints between the two corresponding links. In Figure 4.28b, it can be seen that constraints are needed between four pairs of links: 14–15, 21–22, 23–24, and 19–23.

This is perhaps the trickiest part. For examples like the one shown in Figure 3.27, the constraint may be formulated as in (3.81). This is equivalent to what was done to obtain the example in Figure 4.26, which means that there are actually two constraints, one for each of the x and y coordinates. This will also work for the example shown in Figure 4.27 if all joints are revolute. Suppose instead that two bodies, $\mathcal{A}_j$ and $\mathcal{A}_k$, must be rigidly attached. This requires adding one more constraint that prevents mutual rotation. This could be achieved by selecting another point on $\mathcal{A}_j$ and ensuring that one of its coordinates is in the correct position in the body frame of $\mathcal{A}_k$. If four equations are added, two from each point, then one of them would be redundant because there are only three degrees of freedom possible for $\mathcal{A}_j$ relative to $\mathcal{A}_k$ (which comes from the dimension of $SE(2)$).

A similar but more complicated situation occurs for $\mathcal{W} = \mathbb{R}^3$. Holding a single point fixed produces three constraints. If a single point is held fixed, then $\mathcal{A}_j$ may achieve any rotation in $SO(3)$ with respect to $\mathcal{A}_k$. This implies that $\mathcal{A}_j$ and $\mathcal{A}_k$ are attached by a spherical joint. If they are attached by a revolute joint, then two more constraints are needed, which can be chosen from the coordinates of a second point. If $\mathcal{A}_j$ and $\mathcal{A}_k$ are rigidly attached, then one constraint from a third point is needed. In total, however, there can
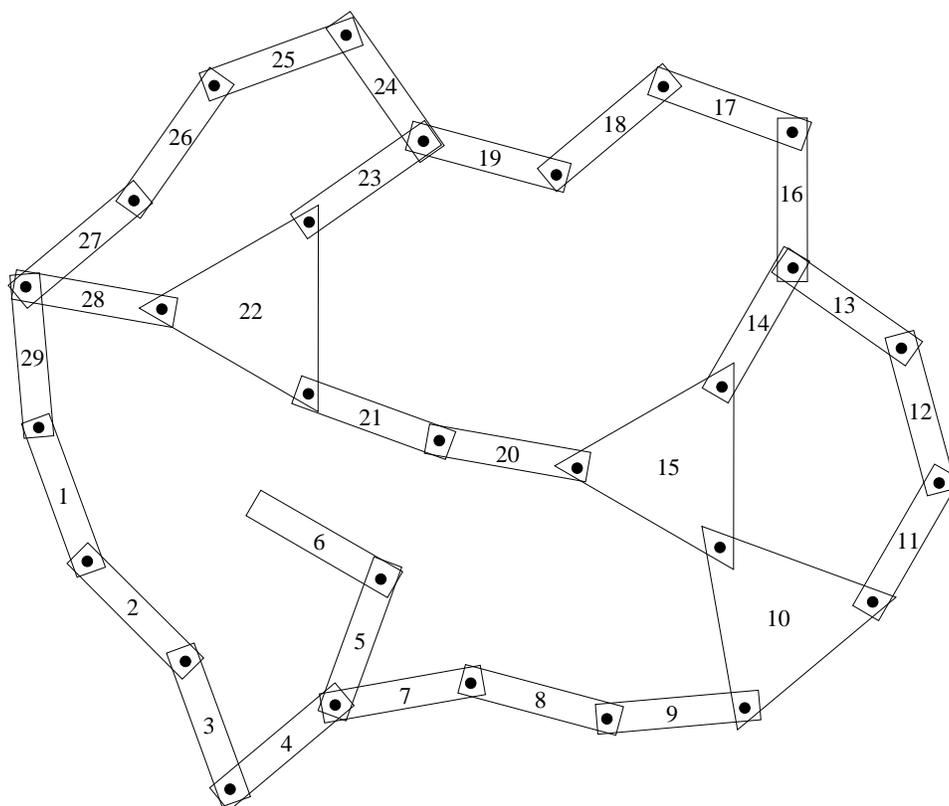
Figure 4.27: A complicated linkage that has 29 links, several loops, links with more than two bodies, and bodies with more than two links. Each integer i indicates link $\mathcal{A}_i$.

be no more than six independent constraints because this is the dimension of $SE(3)$.

6. Convert the trigonometric functions to polynomials. For any 2D transformation, the familiar substitution of complex numbers may be made. If the DH parameterization is used for the 3D case, then each of the $\cos\theta_i$, $\sin\theta_i$ expressions can be parameterized with one complex number, and each of the $\cos\alpha_i$, $\sin\alpha_i$ expressions can be parameterized with another. If the rotation matrix for $SO(3)$ is directly used in the parameterization, then the quaternion parameterization should be used. In all of these cases, polynomial expressions are obtained.

7. List the constraints as polynomial equations of the form $f = 0$. To write the description of the variety, all of the polynomials must be set equal to zero, as was done for the examples in Section 4.4.2.
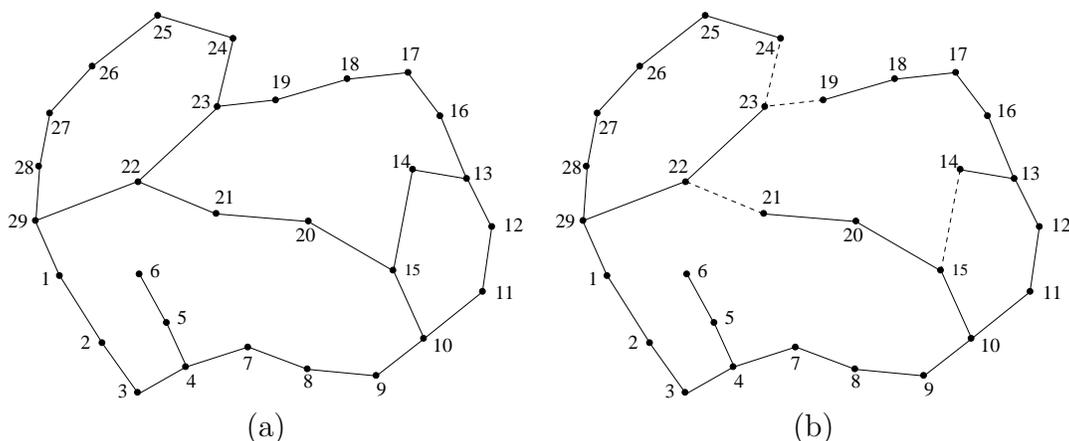
Figure 4.28: (a) One way to make the linkage graph that corresponds to the linkage in Figure 4.27. (b) A spanning tree is indicated by showing the removed edges with dashed lines.

Is it possible to determine the dimension of the variety from the number of independent constraints? The answer is generally *no*, which can be easily seen from the chains of links in Section 4.4.2; they produced varieties of various dimensions, depending on the particular equations. Techniques for computing the dimension exist but require much more machinery than is presented here (see the literature overview at the end of the chapter). However, there is a way to provide a simple upper bound on the number of degrees of freedom. Suppose the total degrees of freedom of the linkage in spanning tree form is m. Each independent constraint can remove at most one degree of freedom. Thus, if there are l independent constraints, then the variety can have no more than $m - l$ dimensions. One expression of this for a general class of mechanisms is the Kutzbach criterion; the planar version of this is called Grübler's formula [310].

One final concern is the obstacle region, $\mathcal{C}_{obs}$. Once the variety has been identified, the obstacle region and motion planning definitions in (4.34) and Formulation 4.1 do not need to be changed. The configuration space $\mathcal{C}$ must be redefined, however, to be the set of configurations that satisfy the closure constraints.

Further Reading

Section 4.1 introduced the basic definitions and concepts of topology. Further study of this fascinating subject can provide a much deeper understanding of configuration spaces. There are many books on topology, some of which may be intimidating, depending on your level of math training. For a heavily illustrated, gentle introduction to topology, see [535]. Another gentle introduction appears in [496]. An excellent text at the graduate level is available on-line: [439]. Other sources include [38, 451]. To understand the motivation for many technical definitions in topology, [911] is helpful. The manifold coverage in Section 4.1.2 was simpler than that found in most sources

because most sources introduce *smooth manifolds*, which are complicated by differentiability requirements (these were not needed in this chapter); see Section 8.3.2 for smooth manifolds. For the configuration spaces of points moving on a topological graph, see [5].

Section 4.2 provided basic C-space definitions. For further reading on matrix groups and their topological properties, see [63], which provides a transition into more advanced material on Lie group theory. For more about quaternions in engineering, see [210, 563]. The remainder of Section 4.2 and most of Section 4.3 were inspired by the coverage in [588]. C-spaces are also covered in [220]. For further reading on computing representations of $\mathcal{C}_{obs}$, see [513, 736] for bitmaps, and Chapter 6 and [865] for combinatorial approaches.

Much of the presentation in Section 4.4 was inspired by the nice introduction to algebraic varieties in [250], which even includes robotics examples; methods for determining the dimension of a variety are also covered. More algorithmic coverage appears in [704]. See [693] for detailed coverage of robots that are designed with closed kinematic chains.

Exercises

1. Consider the set $X = \{1, 2, 3, 4, 5\}$. Let X, $\emptyset$, $\{1, 3\}$, $\{1, 2\}$, $\{2, 3\}$, $\{1\}$, $\{2\}$, and $\{3\}$ be the collection of all subsets of X that are designated as *open sets*.

   (a) Is X a topological space?

   (b) Is it a topological space if $\{1, 2, 3\}$ is added to the collection of open sets? Explain.

   (c) What are the closed sets (assuming $\{1, 2, 3\}$ is included as an open set)?

   (d) Are any subsets of X neither open nor closed?

2. Continuous functions for the strange topology:

   (a) Give an example of a continuous function, $f : X \to X$, for the strange topology in Example 4.4.

   (b) Characterize the set of all possible continuous functions.

3. For the letters of the Russian alphabet, А, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я, determine which pairs are homeomorphic. Imagine each as a 1D subset of $\mathbb{R}^2$ and draw them accordingly before solving the problem.

4. Prove that homeomorphisms yield an equivalence relation on the collection of all topological spaces.

5. What is the dimension of the C-space for a cylindrical rod that can translate and rotate in $\mathbb{R}^3$? If the rod is rotated about its central axis, it is assumed that the rod's position and orientation are not changed in any detectable way. Express the C-space of the rod in terms of a Cartesian product of simpler spaces (such as $\mathbb{S}^1$, $\mathbb{S}^2$, $\mathbb{R}^n$, P^2, etc.). What is your reasoning?

6. Let $\tau_1 : [0, 1] \to \mathbb{R}^2$ be a loop path that traverses the unit circle in the plane, defined as $\tau_1(s) = (\cos(2\pi s), \sin(2\pi s))$. Let $\tau_2 : [0, 1] \to \mathbb{R}^2$ be another loop path: $\tau_1(s) = (-2 + 3\cos(2\pi s), \frac{1}{2}\sin(2\pi s))$. This path traverses an ellipse that is centered at $(-2, 0)$. Show that τ_1 and τ_2 are homotopic (by constructing a continuous function with an additional parameter that "morphs" τ_1 into τ_2).

7. Prove that homotopy yields an equivalence relation on the set of all paths from some $x_1 \in X$ to some $x_2 \in X$, in which x_1 and x_2 may be chosen arbitrarily.

8. Determine the C-space for a spacecraft that can translate and rotate in a 2D *Asteroids*-style video game. The sides of the screen are identified. The top and bottom are also identified. There are no "twists" in the identifications.

9. Repeat the derivation of H_A from Section 4.3.3, but instead consider Type VE contacts.

10. Determine the C-space for a car that drives around on a huge sphere (such as the earth with no mountains or oceans). Assume the sphere is big enough so that its curvature may be neglected (e.g., the car rests flatly on the earth without wobbling). [Hint: It is not $\mathbb{S}^2 \times \mathbb{S}^1$.]

11. Suppose that $\mathcal{A}$ and $\mathcal{O}$ are each defined as equilateral triangles, with coordinates $(0, 0)$, $(2, 0)$, and $(1, \sqrt{3})$. Determine the C-space obstacle. Specify the coordinates of all of its vertices and indicate the corresponding contact type for each edge.

12. Show that (4.20) is a valid rotation matrix for all unit quaternions.

13. Show that $\mathbb{F}[x_1, \ldots, x_n]$, the set of polynomials over a field $\mathbb{F}$ with variables $x_1, \ldots, x_n$, is a group with respect to addition.

14. Quaternions:

    (a) Define a unit quaternion h_1 that expresses a rotation of $-\frac{\pi}{2}$ around the axis given by the vector $[\frac{1}{\sqrt{3}} \quad \frac{1}{\sqrt{3}} \quad \frac{1}{\sqrt{3}}]$.

    (b) Define a unit quaternion h_2 that expresses a rotation of π around the axis given by the vector $[0 \quad 1 \quad 0]$.

    (c) Suppose the rotation represented by h_1 is performed, followed by the rotation represented by h_2. This combination of rotations can be represented as a single rotation around an axis given by a vector. Find this axis and the angle of rotation about this axis.

15. What topological space is contributed to the C-space by a spherical joint that achieves any orientation except the identity?

16. Suppose five polyhedral bodies float freely in a 3D world. They are each capable of rotating and translating. If these are treated as "one" composite robot, what is the topology of the resulting C-space (assume that the bodies are *not* attached to each other)? What is its dimension?
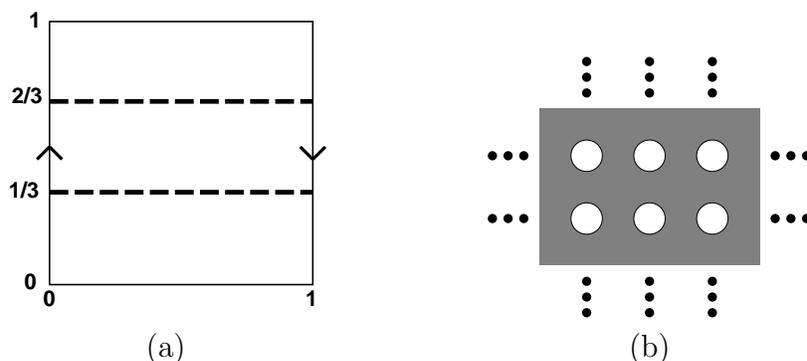
Figure 4.29: (a) What topological space is obtained after slicing the Möbius band? (b) Is a manifold obtained after tearing holes out of the plane?

17. Suppose a goal region $G \subseteq \mathcal{W}$ is defined in the C-space by requiring that the *entire* robot is contained in G. For example, a car may have to be parked entirely within a space in a parking lot.

    (a) Give a definition of $\mathcal{C}_{goal}$ that is similar to (4.34) but pertains to containment of $\mathcal{A}$ inside of G.

    (b) For the case in which $\mathcal{A}$ and G are convex and polygonal, develop an algorithm for efficiently computing $\mathcal{C}_{goal}$.

18. Figure 4.29a shows the Möbius band defined by identification of sides of the unit square. Imagine that scissors are used to cut the band along the two dashed lines. Describe the resulting topological space. Is it a manifold? Explain.

19. Consider Figure 4.29b, which shows the set of points in $\mathbb{R}^2$ that are remaining after a closed disc of radius $1/4$ with center (x, y) is removed for every value of (x, y) such that x and y are both integers.

    (a) Is the remaining set of points a manifold? Explain.

    (b) Now remove discs of radius $1/2$ instead of $1/4$. Is a manifold obtained?

    (c) Finally, remove disks of radius $2/3$. Is a manifold obtained?

20. Show that the solution curves shown in Figure 4.26 correctly illustrate the variety given in (4.73).

21. Find the number of faces of $\mathcal{C}_{obs}$ for a cube and regular tetrahedron, assuming $\mathcal{C}$ is $SE(3)$. How many faces of each contact type are obtained?

22. Following the analysis matrix subgroups from Section 4.2, determine the dimension of $SO(4)$, the group of 4×4 rotation matrices. Can you characterize this topological space?

23. Suppose that a kinematic chain of spherical joints is given. Show how to use (4.20) as the rotation part in each homogeneous transformation matrix, as opposed to

using the DH parameterization. Explain why using (4.20) would be preferable for motion planning applications.

24. Suppose that the constraint that c is held to position $(10, 10)$ is imposed on the mechanism shown in Figure 3.29. Using complex numbers to represent rotation, express this constraint using polynomial equations.

25. The Tangle toy is made of 18 pieces of macaroni-shaped joints that are attached together to form a loop. Each attachment between joints forms a revolute joint. Each link is a curved tube that extends around 1/4 of a circle. What is the dimension of the variety that results from maintaining the loop? What is its configuration space (accounting for internal degrees of freedom), assuming the toy can be placed anywhere in $\mathbb{R}^3$?

Implementations

26. Computing C-space obstacles:

    (a) Implement the algorithm from Section 4.3.2 to construct a convex, polygonal C-space obstacle.

    (b) Now allow the robot to rotate in the plane. For any convex robot and obstacle, compute the orientations at which the C-space obstacle fundamentally changes due to different Type EV and Type VE contacts becoming active.

    (c) Animate the changing C-space obstacle by using the robot orientation as the time axis in the animation.

27. Consider "straight-line" paths that start at the origin (lower left corner) of the manifolds shown in Figure 4.5 and leave at a particular angle, which is input to the program. The lines must respect identifications; thus, as the line hits the edge of the square, it may continue onward. Study the conditions under which the lines fill the entire space versus forming a finite pattern (i.e., a segment, stripes, or a tiling).