# 6

# Robotics and Automatic Geometric Theorem Proving

In this chapter we will consider two recent applications of concepts and techniques from algebraic geometry in areas of computer science. First, continuing a theme introduced in several examples in Chapter 1, we will develop a systematic approach that uses algebraic varieties to describe the space of possible configurations of mechanical linkages such as robot "arms." We will use this approach to solve the forward and inverse kinematic problems of robotics for certain types of robots.

Second, we will apply the algorithms developed in earlier chapters to the study of automatic geometric theorem proving, an area that has been of interest to researchers in artificial intelligence. When the hypotheses of a geometric theorem can be expressed as polynomial equations relating the cartesian coordinates of points in the Euclidean plane, the geometrical propositions deducible from the hypotheses will include all the statements that can be expressed as polynomials in the ideal generated by the hypotheses.

## §1 Geometric Description of Robots

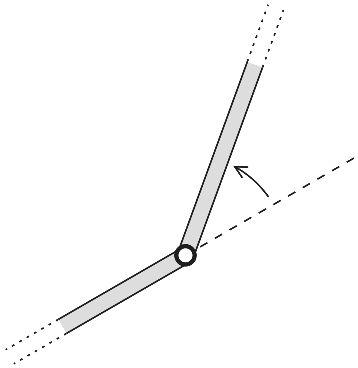
To treat the space of configurations of a robot geometrically, we need to make some simplifying assumptions about the components of our robots and their mechanical properties. We will not try to address many important issues in the engineering of actual robots (such as what types of motors and mechanical linkages would be used to achieve what motions, and how those motions would be controlled). Thus, we will restrict ourselves to highly idealized robots. However, within this framework, we will be able to indicate the types of problems that actually arise in robot motion description and planning.

We will always consider robots constructed from rigid links or segments, connected by joints of various types. For simplicity, we will consider only robots in which the segments are connected *in series*, as in a human limb. One end of our robot "arm" will usually be fixed in position. At the other end will be the "hand" or "effector," which will sometimes be considered as a final segment of the robot. In actual robots, this "hand" might be provided with mechanisms for grasping objects or with tools for performing some task. Thus, one of the major goals is to be able to describe and specify the position and orientation of the "hand."

Since the segments of our robots are rigid, the possible motions of the entire robot assembly are determined by the motions of the joints. Many actual robots are constructed using

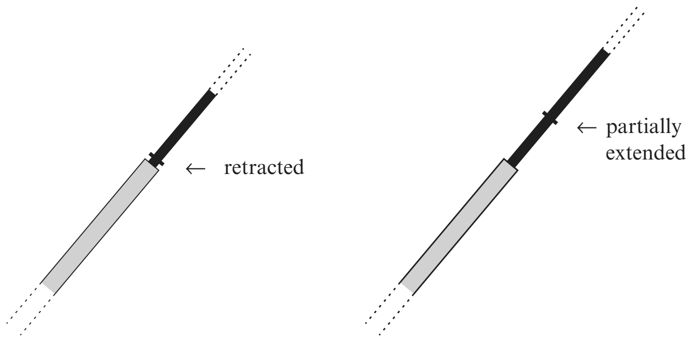
- planar revolute joints, and
- prismatic joints.

A planar revolute joint permits a *rotation* of one segment relative to another. We will assume that both of the segments in question lie in one plane and all motions of the joint will leave the two segments in that plane. (This is the same as saying that the axis of rotation is perpendicular to the plane in question.)



a revolute joint

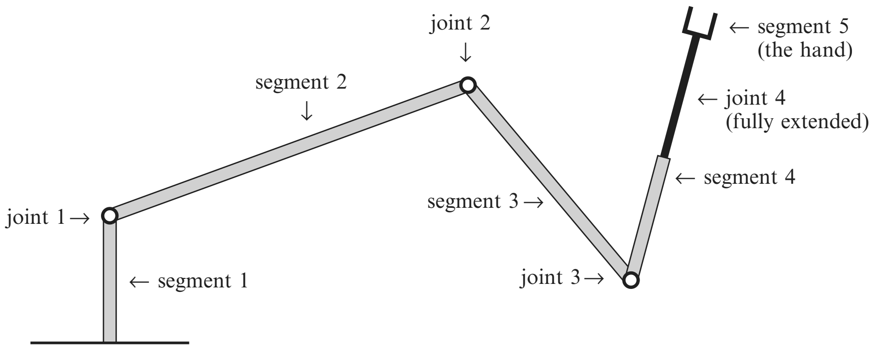
A prismatic joint permits one segment of a robot to move by sliding or *translation* along an axis. The following sketch shows a schematic view of a prismatic joint between two segments of a robot lying in a plane. Such a joint permits translational motion along a line in the plane.




a prismatic joint

If there are several joints in a robot, we will assume for simplicity that the joints all lie in the same plane, that the axes of rotation of all revolute joints are perpendicular to that plane, and, in addition, that the translation axes for the prismatic joints all lie in the plane of the joints. Thus, all motion will take place in one plane. Of course, this leads to a very restricted class of robots. Real robots must usually be capable of 3-dimensional motion. To achieve this, other types and combinations of joints are used. These include "ball" joints allowing rotation about any axis passing through some point in $\mathbb{R}^3$ and helical or "screw" joints combining rotation and translation along the axis of rotation in $\mathbb{R}^3$. It would also be possible to connect several segments of a robot with planar revolute joints, but with *nonparallel* axes of rotation. All of these possible configurations can be treated by methods similar to the ones we will present, but we will not consider them in detail. Our purpose here is to illustrate how affine varieties can be used to describe the geometry of robots, not to present a treatise on practical robotics. The planar robots provide a class of relatively uncomplicated but illustrative examples for us to consider.

**Example 1.** Consider the following planar robot "arm" with three revolute joints and one prismatic joint. All motions of the robot take place in the plane of the paper.




For easy reference, we number the segments and joints of a robot in increasing order out from the fixed end to the hand. Thus, in the above figure, segment 2 connects joints 1 and 2, and so on. Joint 4 is prismatic, and we will regard segment 4 as having variable length, depending on the setting of the prismatic joint. In this robot, the hand of the robot comprises segment 5.

In general, the position or setting of a revolute joint between segments $i$ and $i+1$ can be described by measuring the angle $\theta$ (counterclockwise) from segment $i$ to segment $i + 1$. Thus, the totality of settings of such a joint can be parametrized by a *circle* $S^1$ or by the interval $[0, 2\pi]$ with the endpoints identified. (In some cases, a revolute joint may not be free to rotate through a full circle, and then we would parametrize the possible settings by a subset of $S^1$.)

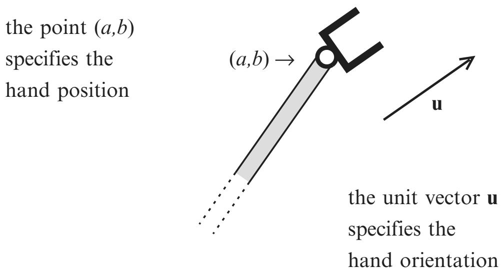
Similarly, the setting of a prismatic joint can be specified by giving the distance the joint is extended or, as in Example 1, by the total length of the segment (i.e., the distance between the end of the joint and the previous joint). Either way, the settings of a prismatic joint can be parametrized by a finite interval of real numbers.

If the joint settings of our robot can be specified independently, then the possible settings of the whole collection of joints in a planar robot with $r$ revolute joints and $p$ prismatic joints can be parametrized by the Cartesian product

$$\mathcal{J} = S^1 \times \cdots \times S^1 \times I_1 \times \cdots \times I_p,$$

where there is one $S^1$ factor for each revolute joint, and $I_j$ gives the settings of the $j$th prismatic joint. We will call $\mathcal{J}$ the *joint space* of the robot.

We can describe the space of possible configurations of the "hand" of a planar robot as follows. Fixing a Cartesian coordinate system in the plane, we can represent the possible positions of the "hand" by the points $(a, b)$ of a region $U \subset \mathbb{R}^2$. Similarly, we can represent the orientation of the "hand" by giving a unit vector aligned with some specific feature of the hand. Thus, the possible hand orientations are parametrized by vectors $\mathbf{u}$ in $V = S^1$. For example, if the "hand" is attached to a revolute joint, then we have the following picture of the hand configuration:




We will call $C = U \times V$ the *configuration space* or *operational space* of the robot's "hand."

Since the robot's segments are assumed to be rigid, each collection of joint settings will place the "hand" in a uniquely determined location, with a uniquely determined orientation. Thus, we have a function or mapping

$$f : \mathcal{J} \longrightarrow \mathcal{C}$$

which encodes how the different possible joint settings yield different hand configurations.

The two basic problems we will consider can be described succinctly in terms of the mapping $f : \mathcal{J} \longrightarrow \mathcal{C}$ described above:

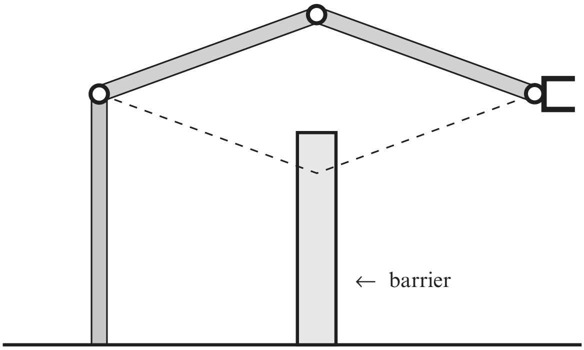
- (Forward Kinematic Problem) Can we give an explicit description or formula for $f$ in terms of the joint settings (our coordinates on $\mathcal{J}$) and the dimensions of the segments of the robot "arm"?
- (Inverse Kinematic Problem) Given $c \in \mathcal{C}$, can we determine one or all the $j \in \mathcal{J}$ such that $f(j) = c$?

In §2, we will see that the forward problem is relatively easily solved. Determining the position and orientation of the "hand" from the "arm" joint settings is mainly a matter of being systematic in describing the relative positions of the segments on either side of a joint. Thus, the forward problem is of interest mainly as a preliminary to the inverse problem. We will show that the mapping $f : \mathcal{J} \longrightarrow \mathcal{C}$ giving the "hand" configuration as a function of the joint settings may be written as a polynomial mapping as in Chapter 5, §1.

The inverse problem is somewhat more subtle since our explicit formulas will not be linear if revolute joints are present. Thus, we will need to use the general results on systems of polynomial equations to solve the equation

$$f(j) = c. \tag{1}$$

One feature of nonlinear systems of equations is that there can be several different solutions, even when the entire set of solutions is finite. We will see in §3 that this is true for a planar robot arm with three (or more) revolute joints. As a practical matter, the potential nonuniqueness of the solutions of the systems (1) is sometimes very desirable. For instance, if our real world robot is to work in a space containing physical obstacles or barriers to movement in certain directions, it may be the case that some of the solutions of (1) for a given $c \in \mathcal{C}$ correspond to positions that are not physically reachable:



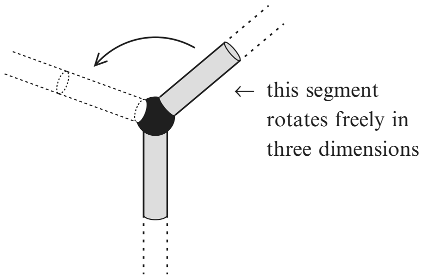

To determine whether it is possible to reach a given position, we might need to determine *all* solutions of (1), then see which one(s) are feasible given the constraints of the environment in which our robot is to work.

**EXERCISES FOR §1**

1. Give descriptions of the joint space $\mathcal{J}$ and the configuration space $\mathcal{C}$ for the planar robot picture in Example 1 in the text. For your description of $\mathcal{C}$, determine a bounded subset of $U \subset \mathbb{R}^2$ containing all possible hand positions. Hint: The description of $U$ will depend on the lengths of the segments.
2. Consider the mapping $f : \mathcal{J} \to \mathcal{C}$ for the robot pictured in Example 1 in the text. On geometric grounds, do you expect $f$ to be a *one-to-one* mapping? Can you find two different ways to put the hand in some particular position with a given orientation? Are there more than two such positions?

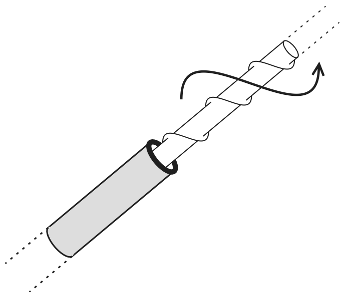
The text discussed the joint space $\mathcal{J}$ and the configuration space $\mathcal{C}$ for planar robots. In the following problems, we consider what $\mathcal{J}$ and $\mathcal{C}$ look like for robots capable of motion in three dimensions.

3. What would the configuration space $\mathcal{C}$ look like for a 3-dimensional robot? In particular, how can we describe the possible hand orientations?
4. A "ball" joint at point $B$ allows segment 2 in the robot pictured below to rotate by any angle about any axis in $\mathbb{R}^3$ passing through $B$. (Note: The motions of this joint are similar to those of the "joystick" found in some computer games.)




a ball joint

   a. Describe the set of possible joint settings for this joint mathematically. Hint: The distinct joint settings correspond to the possible direction vectors of segment 2.
   b. Construct a one-to-one correspondence between your set of joint settings in part (a) and the unit sphere $S^2 \subset \mathbb{R}^3$. Hint: One simple way to do this is to use the spherical angular coordinates $\phi, \theta$ on $S^2$.
5. A helical or "screw" joint at point $H$ allows segment 2 of the robot pictured below to extend out from $H$ along the the line $L$ in the direction of segment 1, while rotating about the axis $L$.



a helical or "screw" joint

The rotation angle $\theta$ (measured from the original, unextended position of segment 2) is given by $\theta = l \cdot \alpha$, where $l \in [0, m]$ gives the distance from $H$ to the other end of segment 2 and

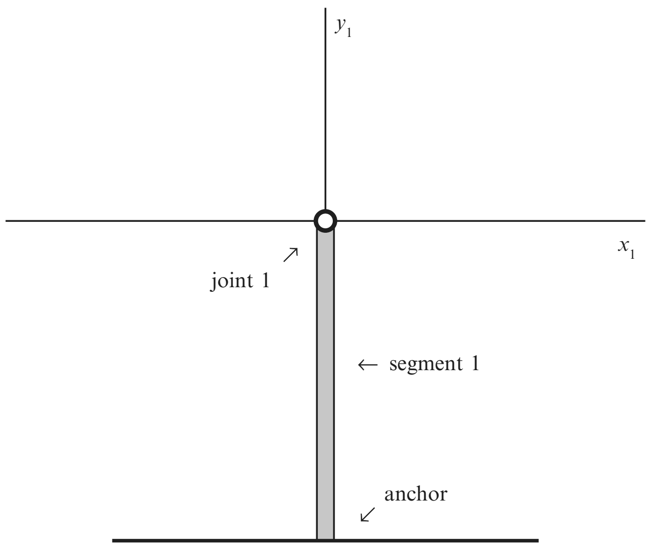
$\alpha$ is a constant angle. Give a mathematical description of the space of joint settings for this joint.

6. Give a mathematical description of the joint space $\mathcal{J}$ for a 3-dimensional robot with two "ball" joints and one helical joint.

## §2 The Forward Kinematic Problem

In this section, we will present a standard method for solving the forward kinematic problem for a given robot "arm." As in §1, we will only consider robots in $\mathbb{R}^2$, which means that the "hand" will be constrained to lie in the plane. Other cases will be studied in the exercises.

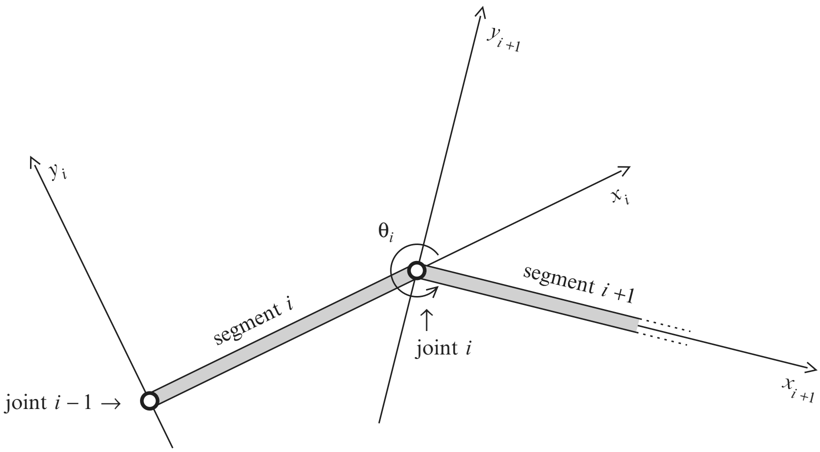
All of our robots will have a first segment that is anchored, or fixed in position. In other words, there is no movable joint at the initial endpoint of segment 1. With this convention, we will use a standard rectangular coordinate system in the plane to describe the position and orientation of the "hand." The origin of this coordinate system is placed at joint 1 of the robot arm, which is also fixed in position since all of segment 1 is. For example:




The Global $(x_1, y_1)$ Coordinate System

In addition to the global $(x_1, y_1)$ coordinate system, we introduce a local rectangular coordinate system at each of the *revolute joints* to describe the relative positions of the segments meeting at that joint. Naturally, these coordinate systems will *change* as the position of the "arm" varies.

At a revolute joint $i$, we introduce an $(x_{i+1}, y_{i+1})$ coordinate system in the following way. The origin is placed at joint $i$. We take the positive $x_{i+1}$-axis to lie along the direction of segment $i+1$ (in the robot's current position). Then the positive $y_{i+1}$-axis is determined to form a normal right-handed rectangular coordinate system. Note that for each $i \geq 2$, the $(x_i, y_i)$ coordinates of joint $i$ are $(l_i, 0)$, where $l_i$ is the length of segment $i$.




Our first goal is to relate the $(x_{i+1}, y_{i+1})$ coordinates of a point with the $(x_i, y_i)$ coordinates of that point. Let $\theta_i$ be the counterclockwise angle from the $x_i$-axis to the $x_{i+1}$-axis. This is the same as the joint setting angle $\theta_i$ described in §1. From the diagram above, we see that if a point $q$ has $(x_{i+1}, y_{i+1})$ coordinates

$$q = (a_{i+1}, b_{i+1}),$$

then to obtain the $(x_i, y_i)$ coordinates of $q$, say

$$q = (a_i, b_i),$$

we rotate by the angle $\theta_i$ (to align the $x_i$- and $x_{i+1}$-axes), and then translate by the vector $(l_i, 0)$ (to make the origins of the coordinate systems coincide). In the exercises, you will show that rotation by $\theta_i$ is accomplished by multiplying by the rotation matrix

$$\begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}.$$

It is also easy to check that translation is accomplished by adding the vector $(l_i, 0)$. Thus, we get the following relation between the $(x_i, y_i)$ and $(x_{i+1}, y_{i+1})$ coordinates of $q$:

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix} \cdot \begin{pmatrix} a_{i+1} \\ b_{i+1} \end{pmatrix} + \begin{pmatrix} l_i \\ 0 \end{pmatrix}.$$

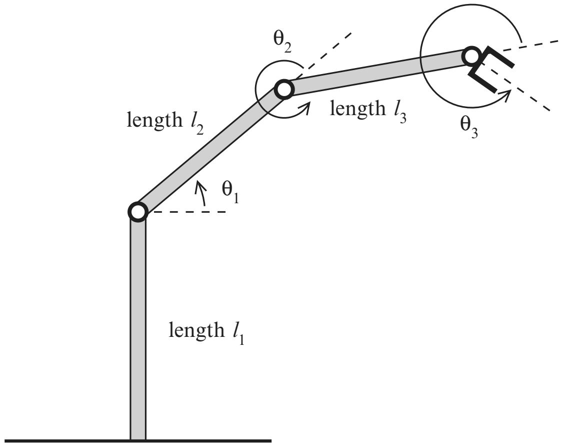
This coordinate transformation is also commonly written in a shorthand form using a

$3 \times 3$ matrix and 3-component vectors:

$$(1) \qquad \begin{pmatrix} a_i \\ b_i \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i & l_i \\ \sin\theta_i & \cos\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{i+1} \\ b_{i+1} \\ 1 \end{pmatrix} = A_i \cdot \begin{pmatrix} a_{i+1} \\ b_{i+1} \\ 1 \end{pmatrix}.$$

This allows us to combine the rotation by $\theta_i$ with the translation along segment $i$ into a single $3 \times 3$ matrix $A_i$.

**Example 1.** With this notation in hand, let us next consider a general plane robot "arm" with three revolute joints:




We will think of the hand as segment 4, which is attached via the revolute joint 3 to segment 3. As before, $l_i$ will denote the length of segment $i$. We have

$$A_1 = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

since the origin of the $(x_2, y_2)$ coordinate system is also placed at joint 1. We also have matrices $A_2$ and $A_3$ as in formula (1). The key observation is that the global coordinates of any point can be obtained by starting in the $(x_4, y_4)$ coordinate system and working our way back to the global $(x_1, y_1)$ system one joint at a time. That is, we multiply the $(x_4, y_4)$ coordinate vector of the point $A_3, A_2, A_1$ in turn:

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = A_1 A_2 A_3 \begin{pmatrix} x_4 \\ y_4 \\ 1 \end{pmatrix}.$$

Using the trigonometric addition formulas, this equation can be written as

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos(\theta_1+\theta_2+\theta_3) & -\sin(\theta_1+\theta_2+\theta_3) & l_3\cos(\theta_1+\theta_2) + l_2\cos\theta_1 \\ \sin(\theta_1+\theta_2+\theta_3) & \cos(\theta_1+\theta_2+\theta_3) & l_3\sin(\theta_1+\theta_2) + l_2\sin\theta_1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_4 \\ y_4 \\ 1 \end{pmatrix}.$$

Since the $(x_4, y_4)$ coordinates of the hand are $(0, 0)$ (because the hand is attached directly to joint 3), we obtain the $(x_1, y_1)$ coordinates of the hand by setting $x_4 = y_4 = 0$ and computing the matrix product above. The result is

$$(2) \qquad \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} l_3 \cos(\theta_1 + \theta_2) + l_2 \cos\theta_1 \\ l_3 \sin(\theta_1 + \theta_2) + l_2 \sin\theta_1 \\ 1 \end{pmatrix}.$$

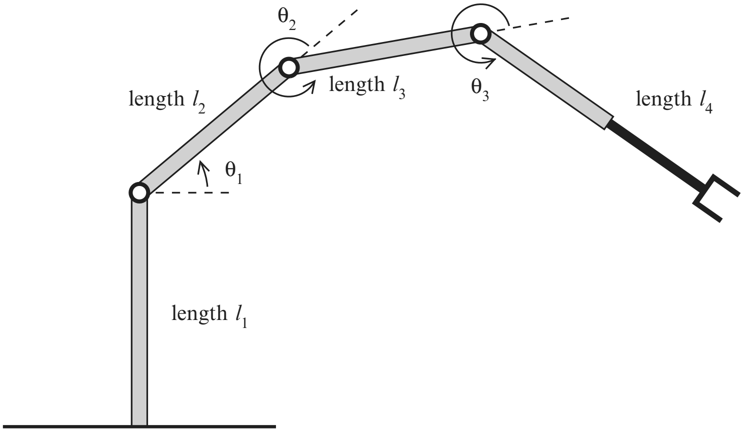
The hand orientation is determined if we know the angle between the $x_4$-axis and the direction of any particular feature of interest to us on the hand. For instance, we might simply want to use the direction of the $x_4$-axis to specify this orientation. From our computations, we know that the angle between the $x_1$-axis and the $x_4$-axis is simply $\theta_1 + \theta_2 + \theta_3$. Knowing the $\theta_i$ allows us to also compute this angle.

If we combine this fact about the hand orientation with the formula (2) for the hand position, we get an explicit description of the mapping $f : \mathcal{J} \to \mathcal{C}$ introduced in §1. As a function of the joint angles $\theta_i$, the configuration of the hand is given by

$$(3) \qquad f(\theta_1 + \theta_2 + \theta_3) = \begin{pmatrix} l_3 \cos(\theta_1 + \theta_2) + l_2 \cos\theta_1 \\ l_3 \sin(\theta_1 + \theta_2) + l_2 \sin\theta_1 \\ \theta_1 + \theta_2 + \theta_3 \end{pmatrix}.$$

The same ideas will apply when any number of planar revolute joints are present. You will study the explict form of the function $f$ in these cases in Exercise 7.

**Example 2.** Prismatic joints can also be handled within this framework. For instance, let us consider a planar robot whose first three segments and joints are the same as those of the robot in Example 1, but which has an additional prismatic joint between segment 4 and the hand. Thus, segment 4 will have variable length and segment 5 will be the hand.



The translation axis of the prismatic joint lies along the direction of segment 4. We can describe such a robot as follows. The three revolute joints allow us exactly the same freedom in placing joint 3 as in the robot studied in Example 1. However, the prismatic joint allows us to change the length of segment 4 to any value between $l_4 = m_1$ (when retracted) and $l_4 = m_2$ (when fully extended). By the reasoning given in Example 1, if the setting $l_4$ of the prismatic joint is known, then the position of the hand will be given by multiplying the product matrix $A_1 A_2 A_3$ times the $(x_4, y_4)$ coordinate vector of the hand, namely $(l_4, 0)$. It follows that the configuration of the hand is given by

$$(4) \qquad g(\theta_1, \theta_2, \theta_3, l_4) = \begin{pmatrix} l_4 \cos(\theta_1 + \theta_2 + \theta_3) + l_3 \cos(\theta_1 + \theta_2) + l_2 \cos\theta_1 \\ l_4 \sin(\theta_1 + \theta_2 + \theta_3) + l_3 \sin(\theta_1 + \theta_2) + l_2 \sin\theta_1 \\ \theta_1 + \theta_2 + \theta_3 \end{pmatrix}.$$

As before, $l_2$ and $l_3$ are constant, but $l_4 \in [m_1, m_2]$ is now another variable. The hand orientation will be given by $\theta_1 + \theta_2 + \theta_3$ as before since the setting of the prismatic joint will not affect the direction of the hand.

We will next discuss how formulas such as (3) and (4) may be converted into representations of $f$ and $g$ as polynomial or rational mappings in suitable variables. The joint variables for revolute and for prismatic joints are handled differently. For the revolute joints, the most direct way of converting to a polynomial set of equations is to use an idea we have seen several times before, for example, in Exercise 8 of Chapter 2, §8. Even though $\cos\theta$ and $\sin\theta$ are transcendental functions, they give a parametrization

$$x = \cos\theta,$$
$$y = \sin\theta$$

of the algebraic variety $\mathbf{V}(x^2 + y^2 - 1)$ in the plane. Thus, we can write the components of the right-hand side of (3) or, equivalently, the entries of the matrix $A_1 A_2 A_3$ in (2) as functions of

$$c_i = \cos\theta_i,$$
$$s_i = \sin\theta_i,$$

subject to the constraints

$$(5) \qquad c_i^2 + s_i^2 - 1 = 0$$

for $i = 1, 2, 3$. Note that the variety defined by these three equations in $\mathbb{R}^6$ is a concrete realization of the joint space $\mathcal{J}$ for this type of robot. Geometrically, this variety is just a Cartesian product of three copies of the circle.

Explicitly, we obtain from (3) an expression for the hand position as a function of the variables $c_1, s_1, c_2, s_2, c_3, s_3$. Using the trigonometric addition formulas, we can write

$$\cos(\theta_1 + \theta_2) = \cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2 = c_1 c_2 - s_1 s_2.$$

Similarly,

$$\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1 = s_1 c_2 + s_2 c_1.$$

Thus, the $(x_1, y_1)$ coordinates of the hand position are:

$$\begin{pmatrix} l_3(c_1c_2 - s_1s_2) + l_2c_1 \\ l_3(s_1c_2 + s_2c_1) + l_2s_1 \end{pmatrix}. \tag{6}$$

In the language of Chapter 5, we have defined a polynomial mapping from the variety $\mathcal{J} = \mathbf{V}(x_1^2 + y_1^2 - 1, x_2^2 + y_2^2 - 1, x_3^2 + y_3^2 - 1)$ to $\mathbb{R}^2$. Note that the hand *position* does not depend on $\theta_3$. That angle enters only in determining the hand orientation.

Since the hand orientation depends directly on the angles $\theta_i$ themselves, it is *not* possible to express the orientation itself as a polynomial in $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$. However, we can handle the orientation in a similar way. See Exercise 3.

Similarly, from the mapping $g$ in Example 2, we obtain the polynomial form

$$\begin{pmatrix} l_4(c_1(c_2c_3 - s_2s_3) - s_1(c_2s_3 + c_3s_2)) + l_3(c_1c_2 - s_1s_2) + l_2c_1 \\ l_4(s_1(c_2c_3 - s_2s_3) + c_1(c_2s_3 + c_3s_2)) + l_3(s_1c_2 + s_2c_1) + l_2s_1 \end{pmatrix} \tag{7}$$

for the $(x_1, y_1)$ coordinates of the hand position. Here $\mathcal{J}$ is the subset $V \times [m_1, m_2]$ of the variety $V \times \mathbb{R}$, where $V = \mathbf{V}(x_1^2 + y_1^2 - 1, x_2^2 + y_2^2 - 1, x_3^2 + y_3^2 - 1)$. The length $l_4$ is treated as another ordinary variable in (7), so our component functions are polynomials in $l_4$, and the $c_i$ and $s_i$.

A second way to write formulas (3) and (4) is based on the *rational* parametrization

$$\begin{aligned} x &= \frac{1 - t^2}{1 + t^2}, \\ y &= \frac{2t}{1 + t^2} \end{aligned} \tag{8}$$

of the circle introduced in §3 of Chapter 1. [In terms of the trigonometric parametrization, $t = \tan(\theta/2)$.] This allows us to express the mapping (3) in terms of three variables $t_i = \tan(\theta_i/2)$. We will leave it as an exercise for the reader to work out this alternate explicit form of the mapping $f : \mathcal{J} \to \mathcal{C}$ in Example 1. In the language of Chapter 5, the variety $\mathcal{J}$ for the robot in Example 1 is birationally equivalent to $\mathbb{R}^3$. We can construct a rational parametrization $\rho : \mathbb{R}^3 \to \mathcal{J}$ using three copies of the parametrization (8). Hence, we obtain a rational mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$, expressing the hand coordinates of the robot arm as functions of $t_1, t_2, t_3$ by taking the composition of $\rho$ with the hand coordinate mapping in the form (6).

Both of these forms have certain advantages and disadvantages for practical use. For the robot of Example 1, one immediately visible advantage of the rational mapping obtained from (8) is that it involves only three variables rather than the six variables $s_i, c_i, i = 1, 2, 3$, needed to describe the full mapping $f$ as in Exercise 3. In addition, we do not need the three extra constraint equations (5). However, the $t_i$ values corresponding to joint positions with $\theta_i$ close to $\pi$ are awkwardly large, and there is no $t_i$ value corresponding to $\theta_i = \pi$. We do not obtain every theoretically possible hand position in the image of the mapping $f$ when it is expressed in this form. Of course, this might not actually be a problem if our robot is constructed so that segment $i + 1$ is not free to fold back onto segment $i$ (that is, the joint setting $\theta_i = \pi$ is not possible).

The polynomial form (6) is more unwieldy, but since it comes from the trigonometric (unit-speed) parametrization of the circle, it does not suffer from the potential shortcomings of the rational form. It would be somewhat better suited for revolute joints that can freely rotate through a full circle.

**EXERCISES FOR §2**

1. Consider the plane $\mathbb{R}^2$ with an orthogonal right-handed coordinate system $(x_1, y_1)$. Now introduce a second coordinate system $(x_2, y_2)$ by rotating the first counterclockwise by an angle $\theta$. Suppose that a point $q$ has $(x_1, y_1)$ coordinates $(a_1, b_1)$ and $(x_2, y_2)$ coordinates $(a_2, b_2)$. We claim that

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}.$$

To prove this, first express the $(x_1, y_1)$ coordinates of $q$ in polar form as

$$q = (a_1, b_1) = (r\cos\alpha, r\sin\alpha).$$

a. Show that the $(x_2, y_2)$ coordinates of $q$ are given by

$$q = (a_2, b_2) = (r\cos(\alpha + \theta), r\sin(\alpha + \theta)).$$

b. Now use trigonometric identities to prove the desired formula.

2. In Examples 1 and 2, we used a $3 \times 3$ matrix $A$ to represent each of the changes of coordinates from one local system to another. Those changes of coordinates were rotations, followed by translations. These are special types of *affine transformations*.

a. Show that any affine transformation in the plane

$$\begin{aligned} x' &= ax + by + e, \\ y' &= cx + dy + f \end{aligned}$$

can be represented in a similar way:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

b. Give a similar representation for affine transformations of $\mathbb{R}^3$ using $4 \times 4$ matrices.

3. In this exercise, we will reconsider the hand orientation for the robots in Examples 1 and 2. Namely, let $\alpha = \theta_1 + \theta_2 + \theta_3$ be the angle giving the hand orientation in the $(x_1, y_1)$ coordinate system.

a. Using the trignomometric addition formulas, show that

$$c = \cos\alpha, \quad s = \sin\alpha$$

can be expressed as polynomials in $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$. Thus, the whole mapping $f$ can be expressed in polynomial form, at the cost of introducing an extra coordinate function for $\mathcal{C}$.

b. Express $c$ and $s$ using the rational parametrization (8) of the circle.

4. Consider a planar robot with a revolute joint 1, segment 2 of length $l_2$, a prismatic joint 2 with settings $l_3 \in [0, m_3]$, and a revolute joint 3, with segment 4 being the hand.

a. What are the joint and configuration spaces $\mathcal{J}$ and $\mathcal{C}$ for this robot?

b. Using the method of Examples 1 and 2, construct an explicit formula for the mapping $f : \mathcal{J} \to \mathcal{C}$ in terms of the trigonometric functions of the joint angles.
c. Convert the function $f$ into a polynomial mapping by introducing suitable new coordinates.

5. Rewrite the mappings $f$ and $g$ in Examples 1 and 2, respectively, using the rational parametrization (8) of the circle for each revolute joint. Show that in each case the hand position and orientation are given by rational mappings on $\mathbb{R}^n$. (The value of $n$ will be different in the two examples.)
6. Rewrite the mapping $f$ for the robot from Exercise 4, using the rational parametrization (8) of the circle for each revolute joint.
7. Consider a planar robot with a fixed segment 1 as in our examples in this section and with $n$ revolute joints linking segments of length $l_2, \ldots, l_n$. The hand is segment $n + 1$, attached to segment $n$ by joint $n$.
   a. What are the joint and configuration spaces for this robot?
   b. Show that the mapping $f : \mathcal{J} \to \mathcal{C}$ for this robot has the form

$$f(\theta_1, \ldots, \theta_n) = \begin{pmatrix} \sum_{i=1}^{n-1} l_{i+1} \cos\left(\sum_{j=1}^{i} \theta_j\right) \\ \sum_{i=1}^{n-1} l_{i+1} \sin\left(\sum_{j=1}^{i} \theta_j\right) \\ \sum_{i=1}^{n} \theta_i \end{pmatrix}.$$

   Hint: Argue by induction on $n$.
8. Another type of 3-dimensional joint is a "spin" or nonplanar revolute joint that allows one segment to rotate or spin in the plane perpendicular to the other segment. In this exercise, we will study the forward kinematic problem for a 3-dimensional robot containing two "spin" joints. As usual, segment 1 of the robot will be fixed, and we will pick a global coordinate system $(x_1, y_1, z_1)$ with the origin at joint 1 and segment 1 on the $z_1$-axis. Joint 1 is a "spin" joint with rotation axis along the $z_1$-axis, so that segment 2 rotates in the $(x_1, y_1)$-plane. Then segment 2 has length $l_2$ and joint 2 is a second "spin" joint connecting segment 2 to segment 3. The axis for joint 2 lies along segment 2, so that segment 3 always rotates in the plane perpendicular to segment 2.
   a. Construct a local right-handed orthogonal coordinate system $(x_2, y_2, z_2)$ with origin at joint 1, with the $x_2$-axis in the direction of segment 2 and the $y_2$-axis in the $(x_1, y_1)$-plane. Give an explicit formula for the $(x_1, y_1, z_1)$ coordinates of a general point, in terms of its $(x_2, y_2, z_2)$ coordinates and of the joint angle $\theta_1$.
   b. Express your formula from part (a) in matrix form, using the $4 \times 4$ matrix representation for affine space transformations given in part (b) of Exercise 2.
   c. Now, construct a local orthogonal coordinate system $(x_3, y_3, z_3)$ with origin at joint 2, the $x_3$-axis in the direction of segment 3, and the $z_3$-axis in the direction of segment 2. Give an explicit formula for the $(x_2, y_2, z_2)$ coordinates of a point in terms of its $(x_3, y_3, z_3)$ coordinates and the joint angle $\theta_2$.
   d. Express your formula from part (c) in matrix form.
   e. Give the transformation relating the $(x_3, y_3, z_3)$ coordinates of a general point to its $(x_1, y_1, z_1)$ coordinates in matrix form. Hint: This will involve suitably multiplying the matrices found in parts (b) and (d).
9. Consider the robot from Exercise 8.
   a. Using the result of part c of Exercise 8, give an explicit formula for the mapping $f : \mathcal{J} \to \mathcal{C}$ for this robot.
   b. Express the hand position for this robot as a polynomial function of the variables $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$.

c. The orientation of the hand (the end of segment 3) of this robot can be expressed by giving a unit vector in the direction of segment 3, expressed in the global coordinate system. Find an expression for the hand orientation.

## §3 The Inverse Kinematic Problem and Motion Planning

In this section, we will continue the discussion of the robot kinematic problems introduced in §1. To begin, we will consider the inverse kinematic problem for the planar robot arm with three revolute joints studied in Example 1 of §2. Given a point $(x_1, y_1) = (a, b) \in \mathbb{R}^2$ and an orientation, we wish to determine whether it is possible to place the hand of the robot at that point with that orientation. If it is possible, we wish to find all combinations of joint settings that will accomplish this. In other words, we want to determine the *image* of the mapping $f : \mathcal{J} \to \mathcal{C}$ for this robot; for each $c$ in the image of $f$, we want to determine the *inverse image* $f^{-1}(c)$.

It is quite easy to see geometrically that if $l_3 = l_2 = l$, the hand of our robot can be placed at any point of the closed disk of radius $2l$ centered at joint 1—the origin of the $(x_1, y_1)$ coordinate system. On the other hand, if $l_3 \neq l_2$, then the hand positions fill out a closed annulus centered at joint 1. (See, for example, the ideas used in Exercise 14 of Chapter 1, §2.) We will also be able to see this using the solution of the forward problem derived in equation (6) of §2. In addition, our solution will give *explicit formulas* for the joint settings necessary to produce a given hand position. Such formulas could be built into a control program for a robot of this kind.

For this robot, it is also easy to control the hand orientation. Since the setting of joint 3 is independent of the settings of joints 1 and 2, we see that, given any $\theta_1$ and $\theta_2$, it is possible to attain any desired orientation $\alpha = \theta_1 + \theta_2 + \theta_3$ by setting $\theta_3 = \alpha - (\theta_1 + \theta_2)$ accordingly.

To simplify our solution of the inverse kinematic problem, we will use the above observation to ignore the hand orientation. Thus, we will concentrate on the position of the hand, which is a function of $\theta_1$ and $\theta_2$ alone. From equation (6) of §2, we see that the possible ways to place the hand at a given point $(x_1, y_1) = (a, b)$ are described by the following system of polynomial equations:

$$
\begin{aligned}
a &= l_3(c_1c_2 - s_1s_2) + l_2c_1, \\
b &= l_3(c_1s_2 + c_2s_1) + l_2s_1, \\
0 &= c_1^2 + s_1^2 - 1, \\
0 &= c_2^2 + s_2^2 - 1
\end{aligned}
\tag{1}
$$

for $s_1, c_1, s_2, c_2$. To solve these equations, we compute a Groebner basis using lex order with the variables ordered

$$c_2 > s_2 > c_1 > s_1.$$

Our solutions will depend on the values of $a, b, l_2, l_3$, which appear as symbolic parameters in the coefficients of the Groebner basis:

$$
\begin{aligned}
& c_2 - \frac{a^2+b^2-l_2^2-l_3^2}{2l_2l_3}, \\
& s_2 + \frac{a^2+b^2}{al_3}s_1 - \frac{a^2b+b^3+b(l_2^2-l_3^2)}{2al_2l_3}, \\
(2) \qquad & c_1 + \frac{b}{a}s_1 - \frac{a^2+b^2+l_2^2-l_3^2}{2al_2}, \\
& s_1^2 - \frac{a^2b+b^3+b(l_2^2-l_3^2)}{l_2(a^2+b^2)}s_1 \\
& \quad + \frac{(a^2+b^2)^2+(l_2^2-l_3^2)^2-2a^2(l_2^2+l_3^2)+2b^2(l_2^2-l_3^2)}{4l_2^2(a^2+b^2)}.
\end{aligned}
$$

In algebraic terms, this is the reduced Groebner basis for the ideal $I$ generated by the polynomials in (1) in the ring $\mathbb{R}(a, b, l_2, l_3)[s_1, c_1, s_2, c_2]$. That is, we allow denominators that depend only on the parameters $a, b, l_2, l_3$.

This is the first time we have computed a Groebner basis over a field of rational functions and one has to be a bit careful about how to interpret (2). Working over $\mathbb{R}(a, b, l_2, l_3)$ means that $a, b, l_2, l_3$ are abstract variables over $\mathbb{R}$, and, in particular, they are algebraically independent [i.e., if $p$ is a polynomial with real coefficients such that $p(a, b, l_2, l_3) = 0$, then $p$ must be the zero polynomial]. Yet, in practice, we want $a, b, l_2, l_3$ to be certain specific real numbers. When we make such a substitution, the polynomials (1) generate an ideal $\overline{I} \subset \mathbb{R}[c_1, s_1, c_2, s_2]$ corresponding to a specific hand position of a robot with specific segment lengths. The key question is whether (2) remains a Groebner basis for $\overline{I}$ under this substitution. In general, the replacement of variables by specific values in a field is called *specialization*, and the question is how a Groebner basis behaves under specialization.

A first observation is that we expect problems when a specialization causes any of the denominators in (2) to vanish. This is typical of how specialization works: things usually behave nicely for most (but not all) values of the variables. In the exercises, you will prove that there is a proper subvariety $W \subset \mathbb{R}^4$ such that (2) specializes to a Groebner basis of $\overline{I}$ whenever $a, b, l_2, l_3$ take values in $\mathbb{R}^4 - W$. We also will see that there is an algorithm for finding $W$. The subtle point is that, in general, the vanishing of denominators is not the only thing that can go wrong (you will work out some examples in the exercises). Fortunately, in the example we are considering, it can be shown that $W$ is, in fact, defined by the vanishing of the denominators. This means that if we choose values $l_2 \neq 0$, $l_3 \neq 0$, $a \neq 0$, and $a^2 + b^2 \neq 0$, then (2) still gives a Groebner basis of (1). The details of the argument will be given in Exercise 9.

Given such a specialization, two observations follow immediately from the form of the leading terms of the Groebner basis (2). First, any zero $s_1$ of the last polynomial can be extended uniquely to a full solution of the system. Second, the set of solutions of (1) is a *finite* set for this choice of $a, b, l_2, l_3$. Indeed, since the last polynomial in (2) is quadratic in $s_1$, there can be at most *two* distinct solutions. It remains to see which $a, b$ yield *real* values for $s_1$ (the relevant solutions for the geometry of our robot).

To simplify the formulas somewhat, we will specialize to the case $l_2 = l_3 = 1$. In Exercise 1, you will show that by either substituting $l_2 = l_3 = 1$ directly into (2) or setting $l_2 = l_3 = 1$ in (1) and recomputing a Groebner basis in $\mathbb{R}(a, b)[s_1, c_1, s_2, c_2]$, we obtain the same result:

$$(3) \qquad \begin{aligned} &c_2 - \frac{a^2+b^2-2}{2}, \\ &s_2 + \frac{a^2+b^2}{a}s_1 - \frac{a^2b+b^3}{2a}, \\ &c_1 + \frac{b}{a}s_1 - \frac{a^2+b^2}{2a}, \\ &s_1^2 - bs_1 + \frac{(a^2+b^2)^2-4a^2}{4(a^2+b^2)}. \end{aligned}$$

Other choices for $l_2$ and $l_3$ will be studied in Exercise 4. [Although (2) remains a Groebner basis for any nonzero values of $l_2$ and $l_3$, the geometry of the situation changes rather dramatically if $l_2 \neq l_3$.]

It follows from our earlier remarks that (3) is a Groebner basis for (1) for all specializations of $a$ and $b$ where $a \neq 0$ and $a^2 + b^2 \neq 0$. Thus, the hand positions with $a = 0$ or $a = b = 0$ appear to have some special properties. We will consider the general case $a \neq 0$ first. Note that this implies $a^2 + b^2 \neq 0$ as well since $a, b \in \mathbb{R}$. Solving the last equation in (3) by the quadratic formula, we find that

$$s_1 = \frac{b}{2} \pm \frac{|a|\sqrt{4-(a^2+b^2)}}{2\sqrt{a^2+b^2}}.$$

Note that the solution(s) of this equation are real if and only if $0 < a^2 + b^2 \leq 4$, and when $a^2 + b^2 = 4$, we have a double root. From the geometry of the system, that is exactly what we expect. The distance from joint 1 to joint 3 is at most $l_2 + l_3 = 2$, and positions with $l_2 + l_3 = 2$ can be reached in only one way—by setting $\theta_2 = 0$ so that segment 3 and segment 2 are pointing in the same direction.

Given $s_1$, we may solve for $c_1, s_2, c_2$ using the other elements of the Groebner basis (3). Since $a \neq 0$, we obtain exactly one value for each of these variables for each possible $s_1$ value. (In fact, the value of $c_2$ does not depend on $s_1$—see Exercise 2.) Further, since $c_1^2 + s_1^2 - 1$ and $c_2^2 + s_2^2 - 1$ are in the ideal generated by (3), the values we get for $s_1, c_1, s_2, c_1$, uniquely determine the joint angles $\theta_1$ and $\theta_2$. Thus, the cases where $a \neq 0$ are easily handled.

We now take up the case of the possible values of $s_1, c_1, s_2, c_2$ when $a = b = 0$. Geometrically, this means that joint 3 is placed at the origin of the $(x_1, y_1)$ system—at the same point as joint 1. Most of the polynomials in our basis (2) are undefined when we try to substitute $a = b = 0$ in the coefficients. So this is a case where specialization fails. With a little thought, the geometric reason for this is visible. There are actually *infinitely many* different possible configurations that will place joint 3 at the origin since segments 2 and 3 have equal lengths. The angle $\theta_1$ can be specified arbitrarily, then setting $\theta_2 = \pi$ will fold segment 3 back along segment 2, placing joint 3 at (0, 0).

These are the only joint settings placing the hand at $(a, b) = (0, 0)$. You will derive the same results by a different method in Exercise 3.

Finally, we ask what happens if $a = 0$ but $b \neq 0$. From the geometry of the robot arm, we would guess that there should be nothing out of the ordinary about these points. Indeed, we could handle them simply by changing coordinates (rotating the $x_1$-, $y_1$-axes, for example) to make the first coordinate of the hand position any nonzero number. Nevertheless, there is an *algebraic* problem since some denominators in (2) vanish at $a = 0$. This is another case where specialization fails. In such a situation, we must substitute $a = 0$ (and $l_2 = l_3 = 1$) into (1) and then recompute the Groebner basis. We obtain

$$
\begin{aligned}
&c_2 - \frac{b^2 - 2}{2},\\
&s_2 - bc_1,\\
&c_1^2 + \frac{b^2 - 4}{4},\\
&s_1 - \frac{b}{2}.
\end{aligned}
\tag{4}
$$

Note that the *form* of the Groebner basis for the ideal is different under this specialization. One difference between this basis and the general form (2) is that the equation for $s_1$ now has degree 1. Also, the equation for $c_1$ (rather than the equation for $s_1$) has degree 2. Thus, we obtain two distinct real values for $c_1$ if $|b| < 2$ and one value for $c_1$ if $|b| = 2$. As in the case $a \neq 0$ above, there are at most two distinct solutions, and the solutions coincide when we are at a point on the boundary of the disk of radius 2. In Exercise 2, you will analyze the geometric meaning of the solutions with $a = 0$ and explain why there is only one distinct value for $s_1$ in this special case.

This completes the analysis of our robot arm. To summarize, given any $(a, b)$ in $(x_1, y_1)$ coordinates, to place joint 3 at $(a, b)$, there are

- infinitely many distinct settings of joint 1 when $a^2 + b^2 = 0$,
- two distinct settings of joint 1 when $a^2 + b^2 < 4$,
- one setting of joint 1 when $a^2 + b^2 = 4$,
- no possible settings of joint 1 when $a^2 + b^2 > 4$.

The cases $a^2 + b^2 = 0, 4$ (but *not* the special cases $a = 0, b \neq 0$) are examples of what are known as *kinematic singularities* for this robot. We will give a precise definition of this concept and discuss some of its meaning below.

In the exercises, you will consider the robot arm with three revolute joints and one prismatic joint introduced in Example 2 of §2. There are more restrictions here for the hand orientation. For example, if $l_4$ lies in the interval $[0, 1]$, then the hand can be placed in any position in the closed disk of radius 3 centered at $(x_1, y_1) = (0, 0)$. However, an interesting difference is that points on the boundary circle can only be reached with one hand orientation.

Before continuing our discussion of robotics, let us make some final comments about specialization. In the example given above, we assumed that we could compute Groebner bases over function fields. In practice, not all computer algebra systems can do this directly—some systems do not allow the coefficients to lie in a function field. The standard method for avoiding this difficulty will be explored in Exercise 10. Another question is how to determine which specializations are the bad ones. One way to attack this problem will be discussed in Exercise 8. Finally, we should mention that there is a special kind of Groebner basis, called a *comprehensive Groebner basis*, which has the property that it remains a Groebner basis under *all* specializations. Such Groebner bases are discussed in the appendix to BECKER and WEISPFENNING (1993).

We will conclude our discussion of the geometry of robots by studying kinematic singularities and some of the issues they raise in robot motion planning. The following discussion will use some ideas from advanced multivariable calculus that we have not encountered before.

Let $f : \mathcal{J} \to \mathcal{C}$ be the function expressing the hand configuration as a function of the joint settings. In the explicit parametrizations of the space $\mathcal{J}$ that we have used, each component of $f$ is a differentiable function of the variables $\theta_i$. For example, this is clearly true for the mapping $f$ for a planar robot with three revolute joints:

$$(5) \qquad f(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} l_3 \cos(\theta_1 + \theta_2) + l_2 \cos\theta_1 \\ l_3 \sin(\theta_1 + \theta_2) + l_2 \sin\theta_1 \\ \theta_1 + \theta_2 + \theta_3 \end{pmatrix}.$$

Hence, we can compute the Jacobian matrix (or matrix of partial derivatives) of $f$ with respect to the variables $\theta_1, \theta_2, \theta_3$. We write $f_i$ for the $i$-th component function of $f$. Then, by definition, the Jacobian matrix is

$$J_f(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} \dfrac{\partial f_1}{\partial \theta_1} & \dfrac{\partial f_1}{\partial \theta_2} & \dfrac{\partial f_1}{\partial \theta_3} \\ \dfrac{\partial f_2}{\partial \theta_1} & \dfrac{\partial f_2}{\partial \theta_2} & \dfrac{\partial f_2}{\partial \theta_3} \\ \dfrac{\partial f_3}{\partial \theta_1} & \dfrac{\partial f_3}{\partial \theta_2} & \dfrac{\partial f_3}{\partial \theta_3} \end{pmatrix}.$$

For example, the mapping $f$ in (5) has the Jacobian matrix

$$(6) \qquad J_f(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} -l_3 \sin(\theta_1 + \theta_2) - l_2 \sin\theta_1 & -l_3 \sin(\theta_1 + \theta_2) & 0 \\ l_3 \cos(\theta_1 + \theta_2) + l_2 \cos\theta_1 & l_3 \cos(\theta_1 + \theta_2) & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

From the matrix of functions $J_f$, we obtain matrices with constant entries by substituting particular values $j = (\theta_1, \theta_2, \theta_3)$. We will write $J_f(j)$ for the substituted matrix, which plays an important role in advanced multivariable calculus. Its key property is that $J_f(j)$ defines a linear mapping which is the *best linear approximation* of the function $f$ at $j \in \mathcal{J}$. This means that near $j$, the function $f$ and the linear function given by $J_f(j)$ have roughly the same behavior. In this sense, $J_f(j)$ represents the *derivative* of the mapping $f$ at $j \in \mathcal{J}$.

To define what is meant by a kinematic singularity, we need first to assign *dimensions* to the joint space $\mathcal{J}$ and the configuration space $\mathcal{C}$ for our robot, to be denoted by $\dim(\mathcal{J})$

and $\dim(\mathcal{C})$, respectively. We will do this in a very intuitive way. The dimension of $\mathcal{J}$, for example, will be simply the number of independent "degrees of freedom" we have in setting the joints. Each planar joint (revolute or prismatic) contributes 1 dimension to $\mathcal{J}$. Note that this yields a dimension of 3 for the joint space of the plane robot with three revolute joints. Similarly, $\dim(\mathcal{C})$ will be the number of independent degrees of freedom we have in the configuration (position and orientation) of the hand. For our planar robot, this dimension is also 3.

In general, suppose we have a robot with $\dim(\mathcal{J}) = m$ and $\dim(\mathcal{C}) = n$. Then differentiating $f$ as before, we will obtain an $n \times m$ Jacobian matrix of functions. If we substitute in $j \in \mathcal{J}$, we get the linear map $J_f(j) : \mathbb{R}^m \to \mathbb{R}^n$ that best approximates $f$ near $j$. An important invariant of a matrix is its *rank*, which is the maximal number of linearly independent columns (or rows). The exercises will review some of the properties of the rank. Since $J_f(j)$ is an $n \times m$ matrix, its rank will always be less than or equal to $\min(m, n)$. For instance, consider our planar robot with three revolute joints and $l_2 = l_3 = 1$. If we let $j = (0, \frac{\pi}{2}, \frac{\pi}{3})$, then formula (6) gives us

$$J_f\left(0, \frac{\pi}{2}, \frac{\pi}{3}\right) = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

This matrix has rank exactly 3 (the largest possible in this case).

We say that $J_f(j)$ has *maximal rank* if its rank is $\min(m,\ n)$ (the largest possible value), and, otherwise, $J_f(j)$ has *deficient rank*. When a matrix has deficient rank, its kernel is larger and image smaller than one would expect (see Exercise 14). Since $J_f(j)$ closely approximates $f$, $J_f(j)$ having deficient rank should indicate some special or "singular" behavior of $f$ itself near the point $j$. Hence, we introduce the following definition.

**Definition 1.** *A* **kinematic singularity** *for a robot is a point $j \in \mathcal{J}$ such that $J_f(j)$ has rank strictly less than* $\min(\dim(\mathcal{J}), \dim(\mathcal{C}))$.

For example, the kinematic singularities of the 3-revolute joint robot occur exactly when the matrix (6) has rank $\leq 2$. For square $n \times n$ matrices, having deficient rank is equivalent to the vanishing of the determinant. We have

$$\begin{aligned} 0 = \det(J_f) &= \sin(\theta_1 + \theta_2)\cos\theta_1 - \cos(\theta_1 + \theta_2)\sin\theta_1 \\ &= \sin\theta_2 \end{aligned}$$

if and only if $\theta_2 = 0$ or $\theta_2 = \pi$. Note that $\theta_2 = 0$ corresponds to a position in which segment 3 extends past segment 2 along the positive $x_2$-axis, whereas $\theta_2 = \pi$ corresponds to a position in which segment 3 is folded back along segment 2. These are exactly the two special configurations we found earlier in which there are not exactly two joint settings yielding a particular hand configuration.

Kinematic singularities are essentially unavoidable for planar robot arms with three or more revolute joints.

**Proposition 2.** *Let $f : \mathcal{J} \to \mathcal{C}$ be the configuration mapping for a planar robot with $n \geq 3$ revolute joints. Then there exist kinematic singularities $j \in \mathcal{J}$.*

**Proof.** By Exercise 7 of §2, we know that $f$ has the form

$$f(\theta_1, \ldots, \theta_n) = \begin{pmatrix} \sum_{i=1}^{n-1} l_{i+1} \cos\left(\sum_{j=1}^{i} \theta_j\right) \\ \sum_{i=1}^{n-1} l_{i+1} \sin\left(\sum_{j=1}^{i} \theta_j\right) \\ \sum_{i=1}^{n} \theta_i \end{pmatrix}.$$

Hence, the Jacobian matrix $J_f$ will be the $3 \times n$ matrix

$$\begin{pmatrix} -\sum_{i=1}^{n-1} l_{i+1} \sin\left(\sum_{j=1}^{i} \theta_j\right) & -\sum_{i=2}^{n-1} l_{i+1} \sin\left(\sum_{j=1}^{i} \theta_j\right) & \ldots & -l_n \sin(\theta_{n-1}) & 0 \\ \sum_{i=1}^{n-1} l_{i+1} \cos\left(\sum_{j=1}^{i} \theta_j\right) & \sum_{i=2}^{n-1} l_{i+1} \cos\left(\sum_{j=1}^{i} \theta_j\right) & \ldots & l_n \cos(\theta_{n-1}) & 0 \\ 1 & 1 & \ldots & 1 & 1 \end{pmatrix}.$$

Since we assume $n \geq 3$, by the definition, a kinematic singularity is a point where the rank of $J_f$ is $\leq 2$. If $j \in \mathcal{J}$ is a point where all $\theta_i \in \{0, \pi\}$, then every entry of the first row of $J_f(j)$ is zero. Hence, rank $J_f(j) \leq 2$ for those $j$. □

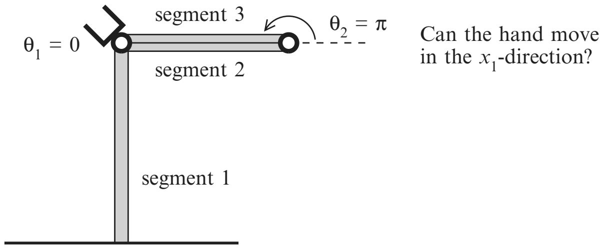
Descriptions of the possible motions of robots such as the ones we have developed are used in an essential way in *planning* the motions of the robot needed to accomplish the tasks that are set for it. The methods we have sketched are suitable (at least in theory) for implementation in programs to control robot motion automatically. The main goal of such a program would be to instruct the robot what joint setting changes to make in order to take the hand from one position to another. The basic problems to be solved here would be first, to find a *parametrized path* $c(t) \in \mathcal{C}$ starting at the initial hand configuration and ending at the new desired configuration, and second, to find a corresponding path $j(t) \in \mathcal{J}$ such that $f(j(t)) = c(t)$ for all $t$. In addition, we might want to impose extra constraints on the paths used such as the following:

1. If the configuration space path $c(t)$ is closed (i.e., if the starting and final configurations are the same), we might also want path $j(t)$ to be a closed path. This would be especially important for robots performing a *repetitive* task such as making a certain weld on an automobile body. Making certain the joint space path is closed means that the whole cycle of joint setting changes can simply be repeated to perform the task again.
2. In any real robot, we would want to limit the *joint speeds* necessary to perform the prescribed motion. Overly fast (or rough) motions could damage the mechanisms.
3. We would want to do as little total joint movement as possible to perform each motion.

Kinematic singularities have an important role to play in motion planning. To see the undesirable behavior that can occur, suppose we have a configuration space path $c(t)$ such that the corresponding joint space path $j(t)$ passes through or near a kinematic singularity. Using the multivariable chain rule, we can differentiate $c(t) = f(j(t))$ with respect to $t$ to obtain

$$c'(t) = J_f(j(t)) \cdot j'(t). \tag{7}$$

We can interpret $c'(t)$ as the velocity of our configuration space path, whereas $j'(t)$ is the corresponding *joint space velocity*. If at some time $t_0$ our joint space path passes *through* a kinematic singularity for our robot, then, because $J_f(j(t_0))$ is a matrix of deficient rank, equation (7) may have *no* solution for $j'(t_0)$, which means there may be *no* smooth joint paths $j(t)$ corresponding to configuration paths that move in certain directions. As an example, consider the kinematic singularities with $\theta_2 = \pi$ for our planar robot with three revolute joints. If $\theta_1 = 0$, then segments 2 and 3 point along the $x_1$-axis:



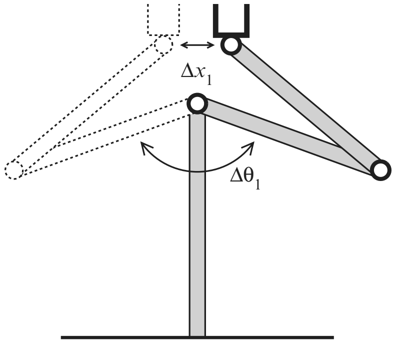

At a Kinematic Singularity

With segment 3 folded back along segment 2, there is no way to move the hand in the $x_1$-direction. More precisely, suppose that we have a configuration path such that $c'(t_0)$ is in the direction of the $x_1$-axis. Then, using formula (6) for $J_f$, equation (7) becomes

$$c'(t_0) = J_f(t_0) \cdot j'(t_0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \cdot j'(t_0).$$

Because the top row of $J_f(t_0)$ is identically zero, this equation has no solution for $j'(t_0)$ since we want the $x_1$ component of $c'(t_0)$ to be nonzero. Thus, $c(t)$ is a configuration path for which there is no corresponding smooth path in joint space. This is typical of what can go wrong at a kinematic singularity.

For $j(t_0)$ *near* a kinematic singularity, we may still have bad behavior since $J_f(j(t_0))$ may be close to a matrix of deficient rank. Using techniques from numerical linear algebra, it can be shown that in (7), if $J_f(j(t_0))$ is close to a matrix of deficient rank, very large joint space velocities may be needed to achieve a small configuration space velocity. For a simple example of this phenomenon, again consider the kinematic singularities of our planar robot with 3 revolute joints with $\theta_2 = \pi$ (where segment 3 is folded back along segment 2). As the diagram on the next page suggests, in order to move from position A to position B, both near the origin, a large change in $\theta_1$ will be needed to move the hand a short distance.




Near a Kinematic Singularity

To avoid undesirable situations such as this, care must be taken in specifying the desired configuration space path $c(t)$. The study of methods for doing this in a systematic way is an active field of current research in robotics, and unfortunately beyond the scope of this text. For readers who wish to pursue this topic further, a standard basic reference on robotics is the text by PAUL (1981). The survey by BUCHBERGER (1985) contains another discussion of Groebner basis methods for the inverse kinematic problem. A readable introduction to much of the more recent work on the inverse kinematic problem and motion control, with references to the original researach papers, is given in BAILLIEUL ET AL. (1990).

**EXERCISES FOR §3**

1. Consider the specialization of the Groebner basis (2) to the case $l_2 = l_3 = 1$.
   a. First, substitute $l_2 = l_3 = 1$ directly into (2) and simplify.
   b. Now, set $l_2 = l_3 = 1$ in (1) and compute a Groebner basis for the "specialized" ideal generated by (1), again using lex order with $c_2 > s_2 > c_1 > s_1$. Compare with your results from part (a) (Your results should be the same.)
2. This exercise studies the geometry of the planar robot with three revolute joints discussed in the text with the dimensions specialized to $l_2 = l_3 = 1$.
   a. Draw a diagram illustrating the two solutions of the inverse kinematic problem for the robot in the general case $a \neq 0, a^2 + b^2 \neq 4$. Why is $c_2$ independent of $s_1$ here? Hint: What kind of quadrilateral is formed by the segments of the robot in the two possible settings to place the hand at $(a, b)$? How are the two values of $\theta_2$ related?
   b. By drawing a diagram, or otherwise, explain the meaning of the two solutons of (4) in the case $a = 0$. In particular, explain why it is reasonable that $s_1$ has only one value. Hint: How are the two values of $\theta_1$ in your diagram related?
3. Consider the robot arm discussed in the text with $l_2 = l_3 = 1$. Set $a = b = 0$ in (1) and recompute a Groebner basis for the ideal. How is this basis different from the bases (3) and (4)? How does this difference explain the properties of the kinematic singularity at (0, 0)?
4. In this exercise, you will study the geometry of the robot discussed in the text when $l_2 \neq l_3$.
   a. Set $l_2 = 1, l_3 = 2$ and solve the system (2) for $s_1, c_1, s_2, c_2$. Interpret your results geometrically, identifying and explaining all special cases. How is this case different from the case $l_2 = l_3 = 1$ done in the text?
   b. Now, set $l_2 = 2, l_3 = 1$ and answer the same questions as in part (a).

As we know from the examples in the text, the form of a Groebner basis for an ideal can change if symbolic parameters appearing in the coefficients take certain special values. In Exercises 5–9, we will study some further examples of this phenomenon and prove some general results.

5. We begin with another example of how denominators in a Groebner basis can cause problems under specialization. Consider the ideal $I = \langle f, g\rangle$, where $f = x^2 - y$, $g = (y - tx)(y - t) = -txy + t^2x + y^2 - ty$, and $t$ is a symbolic parameter. We will use lex order with $x > y$.
   a. Compute a reduced Groebner basis for $I$ in $\mathbb{R}(t)[x, y]$. What polynomials in $t$ appear in the denominators in this basis?
   b. Now set $t = 0$ in $f, g$ and recompute a Groebner basis. How is this basis different from the one in part (a)? What if we clear denominators in the basis from part a and set $t = 0$?
   c. How do the points in the variety $\mathbf{V}(I) \subset \mathbb{R}^2$ depend on the choice of $t \in \mathbb{R}$. Is it reasonable that $t = 0$ is a special case?
   d. The first step of Buchberger's algorithm to compute a Groebner basis for $I$ would be to compute the S-polynomial $S(f, g)$. Compute this S-polynomial by hand in $\mathbb{R}(t)[x, y]$. Note that the special case $t = 0$ is already distinguished at this step.
6. This exercise will explore a more subtle example of what can go wrong during a specialization. Consider the ideal $I = \langle x + ty, x + y\rangle \subset \mathbb{R}(t)[x, y]$, where $t$ is a symbolic parameter. We will use lex order with $x > y$.
   a. Show that $\{x, y\}$ is a reduced Groebner basis of $I$. Note that neither the original basis nor the Groebner basis have any denominators.
   b. Let $t = 1$ and show that $\{x+y\}$ is a Groebner basis for the specialized ideal $\overline{I} \subset \mathbb{R}[x, y]$.
   c. To see why $t = 1$ is special, express the Groebner basis $\{x, y\}$ in terms of the original basis $\{x + ty, x + y\}$. What denominators do you see? In the next problem, we will explore the general case of what is happening here.
7. In this exercise, we will derive a condition under which the form of a Groebner basis *does not* change under specialization. Consider the ideal

$$I = \langle f_i(t_1, \dots, t_m, x_1, \dots, x_n) : 1 \le i \le s\rangle$$

   in $k(t_1, \dots, t_m)[x_1, \dots, x_n]$ and fix a monomial order. We think of $t_1, \dots, t_m$ as symbolic parameters appearing in the coefficients of $f_1, \dots, f_s$. By dividing each $f_i$ by its leading coefficient [which lies in $k(t_1, \dots, t_m)$], we may assume that the leading coefficients of the $f_i$ are all equal to 1. Then let $\{g_1, \dots, g_t\}$ be a reduced Groebner basis for $l$. Thus the leading coefficients of the $g_i$ are also 1. Finally, let $(t_1, \dots, t_m) \mapsto (a_1, \dots, a_m) \in k^m$ be a specialization of the parameters such that none of the denominators of the $f_i$ or $g_i$ vanish at $(a_1, \dots, a_m)$.
   a. If we use the division algorithm to find $A_{ij} \in k(t_1, \dots, t_m)[x_1, \dots, x_n]$ such that

$$f_i = \sum_{j=1}^{t} A_{ij} g_j,$$

   then show that none of the denominators of $A_{ij}$ vanish at $(a_1, \dots, a_m)$.
   b. We also know that $g_j$ can be written

$$g_j = \sum_{i=1}^{s} B_{ji} f_i,$$

   for some $B_{ij} \in k(t_1, \dots, t_m)[x_1, \dots, x_n]$. As Exercise 6 shows, the $B_{ji}$ may introduce new denominators. So assume, in addition, that none of the denominators of the $B_{ji}$

vanish under the specialization $(t_1, \ldots, t_m) \mapsto (a_1, \ldots, a_m)$. Let $\overline{I}$ denote the ideal in $k[x_1, \ldots, x_n]$ generated by the specialized $f_i$. Under these assumptions, prove that the specialized $g_j$ form a basis of $\overline{I}$.

c. Show that the specialized $g_j$ form a Groebner basis for $\overline{I}$. Hint: The monomial order used to compute $\overline{I}$ only deals with terms in the variables $x_j$. The parameters $t_j$ are "constants" as far as the ordering is concerned.

d. Let $d_1, \ldots, d_M \in k[t_1, \ldots, t_m]$ be all denominators that appear among $f_i$, $g_j$, and $B_{ji}$, and let $W = \mathbf{V}(d_1 \cdot d_2 \cdots d_M) \subset k^m$. Conclude that the $g_j$ remain a Groebner basis for the $f_i$ under all specializations $(t_1, \ldots, t_m) \mapsto (a_1, \ldots, a_m) \in k^m - W$.

8. We next describe an algorithm for finding which specializations preserve a Groebner basis. We will use the notation of Exercise 7. Thus, we want an algorithm for finding the denominators $d_1, \ldots, d_M$ appearing in the $f_i$, $g_j$, and $B_{ji}$. This is easy to do for the $f_i$ and $g_j$, but the $B_{ji}$ are more difficult. The problem is that since the $f_i$ are not a Groebner basis, we cannot use the division algorithm to find the $B_{ji}$. Fortunately, we only need the denominators. The idea is to work in the ring $k[t_1, \ldots, t_m, x_1, \ldots, x_n]$. If we multiply the $f_i$ and $g_j$ by suitable polynomials in $k[t_1, \ldots, t_m]$, we get

$$\tilde{f}_i, \tilde{g}_j \in k[t_1, \ldots, t_m, x_1, \ldots, x_n].$$

Let $\overline{I} \subset k[t_1, \ldots, t_m, x_1, \ldots, x_n]$ be the ideal generated by the $\tilde{f}_i$.

a. Suppose $g_j = \sum_{i=1}^{s} B_{ji} f_i$ in $k(t_1, \ldots, t_m)[x_1, \ldots, x_n]$ and let $d \in k[t_1, \ldots, t_m]$ be a polynomial that clears all denominators for the $g_j$, the $f_i$, and the $B_{ji}$. Then prove that

$$d \in (\tilde{I} : \tilde{g}_j) \cap k[t_1, \ldots, t_m],$$

where $\overline{I} : \tilde{g}_j$ is the ideal quotient as defined in §4 of Chapter 4.

b. Give an algorithm for computing $(\tilde{I} : \tilde{g}_j) \cap k[t_1, \ldots, t_m]$ and use this to describe an algorithm for finding the subset $W \subset k^m$ described in part (d) of Exercise 7.

9. The algorithm described in Exercise 8 can lead to lengthy calculations which may be too much for some computer algebra systems. Fortunately, quicker methods are available in some cases. Let $f_i, g_j \in k(t_1, \ldots, t_m)[x_1, \ldots, x_n]$ be as in Exercises 7 and 8, and suppose we suspect that the $g_j$ will remain a Groebner basis for the $f_i$ under all specializations where the denominators of the $f_i$ and $g_j$ do not vanish. How can we check this quickly?

a. Let $d \in k[t_1, \ldots, t_m]$ be the least common multiple of all denominators in the $f_i$ and $g_j$ and let $\tilde{f}_i, \tilde{g}_j \in k[t_1, \ldots, t_m, x_1, \ldots, x_n]$ be the polynomials we get by clearing denominators. Finally, let $\overline{I}$ be the ideal in $k[t_1, \ldots, t_m, x_1, \ldots, x_n]$ generated by the $\tilde{f}_i$. If $d\tilde{g}_j \in \tilde{I}$ for all $j$, then prove that specialization works for all $(t_1, \ldots, t_m) \mapsto (a_1, \ldots, a_m) \in k^m - \mathbf{V}(d)$.

b. Describe an algorithm for checking the criterion given in part a. For efficiency, what monomial order should be used?

c. Apply the algorithm of part (b) to equations (1) in the text. This will prove that (2) remains a Groebner basis for (1) under all specializations where $l_2 \neq 0, l_3 \neq 0, a \neq 0$, and $a^2 + b^2 \neq 0$.

10. In this exercise, we will learn how to compute a Groebner basis for an ideal in $k(t_1, \ldots, t_m)[x_1, \ldots, x_n]$ by working in the polynomial ring $k[t_1, \ldots, t_m, x_1, \ldots, x_n]$. This is useful when computing Groebner bases using computer algebra systems that won't allow the coefficients to lie in a function field. The first step is to fix a term order such that any monomial involving one of the $x_i$'s is greater than all monomials in $t_1, \ldots, t_m$ alone. For example, one could use a product order or lex order with $x_1 > \cdots > x_n > t_1 > \cdots > t_n$.

a. If $I$ is an ideal in $k(t_1, \ldots, t_m)[x_1, \ldots, x_n]$, show that $I$ can be written in the form

$$I = \langle f_i(t_1, \ldots, t_m, x_1, \ldots, x_n) : 1 \leq i \leq s \rangle,$$

where each $f_i \in k[t_1, \ldots, t_m, x_1, \ldots, x_n]$.

b. Now let $\tilde{I}$ be the ideal in $k[t_1, \ldots, t_m, x_1, \ldots, x_n]$. generated by $f_1, \ldots, f_s$, and let $g_1, \ldots, g_t$ be a reduced Groebner basis for $\tilde{I}$ with respect to the above term order. If any of the $g_i$ lie in $k[t_1, \ldots, t_n]$ show that $I = k(t_1, \ldots, t_m)[x_1, \ldots, x_n]$.

c. Let $g_1, \ldots, g_t$ be the Groebner basis of $\tilde{I}$ from part b, and assume that none of the $g_i$ lie in $k[t_1, \ldots, t_m]$. Then prove that $g_1, \ldots, g_t$ are a Groebner basis for $I$ (using the term order induced on monomials in $x_1, \ldots, x_n$).

11. Consider the planar robot with two revolute joints and one prismatic joint described in Exercise 4 of §2.
    a. Given a desired hand position and orientation, set up a system of equations as in (1) of this section whose solutions give the possible joint settings to reach that hand configuration. Take the length of segment 2 to be 1.
    b. Using a computer algebra system, solve your equations by computing a Groebner basis for the ideal generated by the equations from part (a) with respect to a suitable lex order. Note: Some experimentation may be necessary to find a reasonable variable order.
    c. What is the solution of the inverse kinematic problem for this robot. That is, which hand positions and orientations are possible? How many different joint settings yield a given hand configuration? (Do not forget that the setting of the prismatic joint is limited to a finite interval in $[0, m_3] \subset \mathbb{R}$.)
    d. Does this robot have any kinematic singularities according to Definition 1? If so, describe them.
12. Consider the planar robot with three joints and one prismatic joint that we studied in Example 2 of § 2.
    a. Given a desired hand position and orientation, set up a system of equations as in (1) of this section whose solutions give the possible joint settings to reach that hand configuration. Assume that segments 2 and 3 have length 1, and that segment 4 varies in length between 1 and 2. Note: Your system of equations for this robot *should* involve the hand orientation.
    b. Solve your equations by computing a Groebner basis for the ideal generated by your equations with respect to a suitable lex order. Note: Some experimentation may be necessary to find a reasonable variable order. The "wrong" variable order can lead to a completely intractable problem in this example.
    c. What is the solution of the inverse kinematic problem for this robot? That is, which hand positions and orientations are possible? How does the set of possible hand orientations vary with the position? (Do not forget that the setting $l_4$ of the prismatic joint is limited to the finite interval in $[1, 2] \subset \mathbb{R}$.)
    d. How many different joint settings yield a given hand configuration in general? Are these special cases?
    e. Does this robot have any kinematic singularities according to Definition 1? If so, describe the corresponding robot configurations and relate them to part (d).
13. Consider the 3-dimensional robot with two "spin" joints from Exercise 8 of §2.
    a. Given a desired hand position and orientation, set up a system of equations as in (1) of this section whose solutions give the possible joint settings to reach that hand configuration. Take the length of segment 2 to be 4, and the length of segment 3 to be 2, if you like.
    b. Solve your equations by computing a Groebner basis for the ideal generated by your equations with respect to a suitable lex order. Note: In this case there will be an element

of the Groebner basis that depends only on the hand position coordinates. What does this mean geometrically? Is your answer reasonable in terms of the geometry of this robot?

c. What is the solution of the inverse kinematic problem for this robot? That is, which hand positions and orientations are possible?

d. How many different joint settings yield a given hand configuration in general? Are these special cases?

e. Does this robot have any kinematic singularities according to Definition 1?

14. Let $A$ be an $m \times n$ matrix with real entries. We will study the rank of $A$, which is the maximal number of linearly independent columns (or rows) in $A$. Multiplication by $A$ gives a linear map $L_A : \mathbb{R}^n \to \mathbb{R}^m$, and from linear algebra, we know that the rank of $A$ is the dimension of the image of $L_A$. As in the text, $A$ has maximal rank if its rank is $\min(m, n)$. To understand what maximal rank means, there are three cases to consider.

a. If $m = n$, show that $A$ has maximal rank $\Leftrightarrow \det(A) \neq 0 \Leftrightarrow L_A$ is an isomorphism of vector spaces.

b. If $m < n$, show that $A$ has maximal rank $\Leftrightarrow$ the equation $A \cdot \mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^m \Leftrightarrow L_A$ is a surjective (onto) mapping.

c. If $m > n$, show that $A$ has maximal rank $\Leftrightarrow$ the equation $A \cdot \mathbf{x} = \mathbf{b}$ has at most one solution for all $\mathbf{b} \in \mathbb{R}^m \Leftrightarrow L_A$ is an injective (one-to-one) mapping.

15. A robot is said to be *kinematically redundant* if the dimension of its joint space $\mathcal{J}$ is *larger* than the dimension of its configuration space $\mathcal{C}$.

a. Which of the robots considered in this section (in the text and in Exercises 11–13 above) are kinematically redundant?

b. (This part requires knowledge of the Implicit Function Theorem.) Suppose we have a kinematically redundant robot and $j \in \mathcal{J}$ is not a kinematic singularity. What can be said about the inverse image $f^{-1}(f(j))$ in $\mathcal{J}$? In particular, how many different ways are there to put the robot in the configuration given by $f(j)$?

16. Verify the chain rule formula (7) explicitly for the planar robot with three revolute joints. Hint: Substitute $\theta_i = \theta_i(t)$ and compute the derivative of the configuration space path $f(\theta_1(t), \theta_2(t), \theta_3(t))$ with respect to $t$.

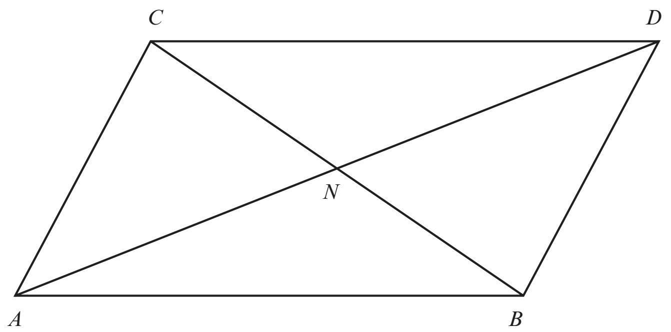
## §4 Automatic Geometric Theorem Proving

The geometric descriptions of robots and robot motion we studied in the first three sections of this chapter were designed to be used as tools by a control program to help plan the motions of the robot to accomplish a given task. In the process, the control program could be said to be "reasoning" about the geometric constraints given by the robot's design and its environment and to be "deducing" a feasible solution to the given motion problem. In this section and in the next, we will examine a second subject which has some of the same flavor—automated geometric reasoning in general. We will give two algorithmic methods for determining the validity of general statements in Euclidean geometry. Such methods are of interest to researchers both in artificial intelligence (AI) and in geometric modeling because they have been used in the design of programs that, in effect, can prove or disprove conjectured relationships between, or theorems about, plane geometric objects.

Few people would claim that such programs embody an understanding of the *meaning* of geometric statements comparable to that of a human geometer. Indeed, the whole question of whether a computer is capable of intelligent behavior is one that is still completely unresolved. However, it is interesting to note that a number of new (that is, apparently previously unknown) theorems have been verified by these methods. In a limited sense, these "theorem provers" are capable of "reasoning" about geometric configurations, an area often considered to be solely the domain of human intelligence.

The basic idea underlying the methods we will consider is that once we introduce Cartesian coordinates in the Euclidean plane, the hypotheses and the conclusions of a large class of geometric theorems can be expressed as *polynomial equations* between the coordinates of collections of points specified in the statements. Here is a simple but representative example.

**Example 1.** Let $A, B, C, D$ be the vertices of a parallelogram in the plane, as in the figure below.




It is a standard geometric theorem that the two diagonals $\overline{AD}$ and $\overline{BC}$ of any parallelogram intersect at a point ($N$ in the figure) which bisects both diagonals. In other words, $AN = DN$ and $BN = CN$, where, as usual, $XY$ denotes the length of the line segment $\overline{XY}$ joining the two points $X$ and $Y$. The usual proof from geometry is based on showing that the triangles $\Delta ANC$ and $\Delta BND$ are congruent. See Exercise 1.

To relate this theorem to algebraic geometry, we will show how the configuration of the parallelogram and its diagonals (the hypotheses of the theorem) and the statement that the point $N$ bisects the diagonals (the conclusion of the theorem) can be expressed in polynomial form.

The properties of parallelograms are unchanged under translations and rotations in the plane. Hence, we may begin by translating and rotating the parallelogram to place it in any position we like, or equivalently, by choosing our coordinates in any convenient fashion. The simplest way to proceed is as follows. We place the vertex $A$ at the origin and align the side $\overline{AB}$ with the horizontal coordinate axis. In other words, we can take $A = (0, 0)$ and $B = (u_1, 0)$ for some $u_1 \neq 0 \in \mathbb{R}$. In what follows we will think of $u_1$ as an indeterminate or variable whose value can be chosen arbitrarily in $\mathbb{R} - \{0\}$. The vertex $C$ of the parallelogram can be at any point $C = (u_2, u_3)$, where $u_2, u_3$ are new indeterminates independent of $u_1$, and $u_3 \neq 0$. The remaining vertex $D$ is now completely determined by the choice of $A, B, C$.

It will always be true that when constructing the geometric configuration described by a theorem, some of the coordinates of some points will be *arbitrary*, whereas the remaining coordinates of points will be *determined* (possibly up to a finite number of choices) by the arbitrary ones. To indicate arbitrary coordinates, we will consistently use variables $u_i$, whereas the other coordinates will be denoted $x_j$. It is important to note that this division of coordinates into two subsets is in no way uniquely specified by the hypotheses of the theorem. Different constructions of a figure, for example, may lead to different sets of arbitrary variables and to different translations of the hypotheses into polynomial equations.

Since $D$ is determined by $A$, $B$, and $C$, we will write $D = (x_1, x_2)$. One hypothesis of our theorem is that the quadrilateral $ABDC$ is a parallelogram or, equivalently, that the opposite pairs of sides are parallel and, hence, have the same slope. Using the slope formula for a line segment, we see that one translation of these statements is as follows:

$$\begin{aligned} \overline{AB} \parallel \overline{CD} &: 0 = \frac{x_2 - u_3}{x_1 - u_2}, \\ \overline{AC} \parallel \overline{BD} &: \frac{u_3}{u_2} = \frac{x_2}{x_1 - u_1}. \end{aligned}$$

Clearing denominators, we obtain the polynomial equations

$$\begin{aligned} h_1 &= x_2 - u_3 = 0, \\ h_2 &= (x_1 - u_1)u_3 - x_2u_2 = 0. \end{aligned} \tag{1}$$

(Below, we will discuss another way to get equations for $x_1$ and $x_2$.)

Next, we construct the intersection point of the diagonals of the parallelogram. Since the coordinates of the intersection point $N$ are determined by the other data, we write $N = (x_3, x_4)$. Saying that $N$ is the intersection of the diagonals is equivalent to saying that $N$ lies on both of the lines $\overline{AD}$ and $\overline{BC}$, or to saying that the triples $A, N, D$ and $B, N, C$ are *collinear*. The latter form of the statement leads to the simplest formulation of these hypotheses. Using the slope formula again, we have the following relations:

$$\begin{aligned} A, N, D \text{ collinear} &: \frac{x_4}{x_3} = \frac{u_3}{x_1}, \\ B, N, C \text{ collinear} &: \frac{x_4}{x_3 - u_1} = \frac{u_3}{u_2 - u_1}. \end{aligned}$$

Clearing denominators again, we have the polynomial equations

$$\begin{aligned} h_3 &= x_4x_1 - x_3u_3 = 0, \\ h_4 &= x_4(u_2 - u_1) - (x_3 - u_1)u_3 = 0. \end{aligned} \tag{2}$$

The system of four equations formed from (1) and (2) gives one translation of the hypotheses of our theorem.

The conclusions can be written in polynomial form by using the distance formula for two points in the plane (the Pythagorean Theorem) and squaring:

$$\begin{aligned} AN = ND &: x_3^2 + x_4^2 = (x_3 - x_1)^2 + (x_4 - x_2)^2, \\ BN = NC &: (x_3 - u_1)^2 + x_4^2 = (x_3 - u_2)^2 + (x_4 - u_3)^2. \end{aligned}$$

Cancelling like terms, the conclusions can be written as

$$
(3) \qquad \begin{aligned} g_1 &= x_1^2 - 2x_1x_3 - 2x_4x_2 + x_2^2 = 0, \\ g_2 &= 2x_3u_1 - 2x_3u_2 - 2x_4u_3 - u_1^2 + u_2^2 + u_3^2 = 0. \end{aligned}
$$

Our translation of the theorem states that the two equations in (3) should hold when the hypotheses in (1) and (2) hold.

As we noted earlier, different translations of the hypotheses and conclusions of a theorem are possible. For instance, see Exercise 2 for a different translation of this theorem based on a different construction of the parallelogram (that is, a different collection of arbitrary coordinates). There is also a great deal of freedom in the way that hypotheses can be translated. For example, the way we represented the hypothesis that *ABDC* is a parallelogram in (1) is typical of the way a *computer program* might translate these statements, based on a general method for handling the hypothesis $\overline{AB} \parallel \overline{CD}$. But there is an alternate translation based on the observation that, from the parallelogram law for vector addition, the coordinate vector of the point $D$ should simply be the *vector sum* of the coordinate vectors $B = (u_1, 0)$ and $C = (u_2, u_3)$. Writing $D = (x_1, x_2)$, this alternate translation would be

$$
(4) \qquad \begin{aligned} h_1' &= x_1 - u_1 - u_2 = 0, \\ h_2' &= x_2 - u_3 = 0. \end{aligned}
$$

These equations are much simpler than the ones in (1). If we wanted to design a geometric theorem-prover that could translate the hypothesis "*ABDC* is a parallelogram" *directly* (without reducing it to the equivalent form "$\overline{AB} \parallel \overline{CD}$ and $\overline{AC} \parallel \overline{BD}$"), the translation (4) would be preferable to (1).

Further, we could also use $h_2'$ to *eliminate* the variable $x_2$ from the hypotheses and conclusions, yielding an even simpler system of equations. In fact, with complicated geometric constructions, preparatory simplifications of this kind can sometimes be necessary. They often lead to much more tractable systems of equations.

The following proposition lists some of the most common geometric statements that can be translated into polynomial equations.

**Proposition 2.** *Let A, B, C, D, E, F be points in the plane. Each of the following geometric statements can be expressed by one or more polynomial equations:*

(i) *$\overline{AB}$ is parallel to $\overline{CD}$.*
(ii) *$\overline{AB}$ is perpendicular to $\overline{CD}$.*
(iii) *A, B, C are collinear.*
(iv) *The distance from A to B is equal to the distance from C to D: AB = CD.*
(v) *C lies on the circle with center A and radius AB.*
(vi) *C is the midpoint of $\overline{AB}$.*
(vii) *The acute angle ∠ABC is equal to the acute angle ∠DEF.*
(viii) *$\overline{BD}$ bisects the angle ∠ABC.*

**Proof.** General methods for translating statements (i), (iii), and (iv) were illustrated in Example 1; the general cases are exactly the same. Statement (v) is equivalent to $AC = AB$. Hence, it is a special case of (iv) and can be treated in the same way. Statement (vi) can be reduced to a conjunction of two statements: $A, C, B$ are collinear, *and* $AC = CB$. We, thus, obtain two equations from (iii) and (iv). Finally, (ii), (vii), and (viii) are left to the reader in Exercise 4. □

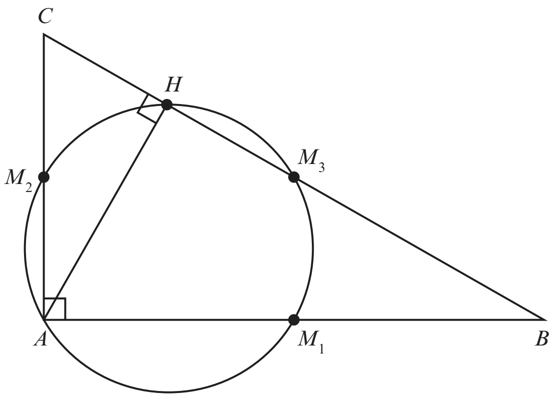
Exercise 3 gives several other types of statements that can be translated into polynomial equations. We will say that a geometric theorem is *admissible* if both its hypotheses and its conclusions admit translations into polynomial equations. There are always many different equivalent formulations of an admissible theorem; the translation will never be unique.

Correctly translating the hypotheses of a theorem into a system of polynomial equations can be accomplished most readily if we think of *constructing* a figure illustrating the configuration in question point by point. This is exactly the process used in Example 1 and in the following example.

**Example 3.** We will use Proposition 2 to translate the following beautiful result into polynomial equations.

**Theorem (The Circle Theorem of Apollonius).** *Let $\Delta ABC$ be a right triangle in the plane, with right angle at A. The midpoints of the three sides and the foot of the altitude drawn from A to $\overline{BC}$ all lie on one circle.*

The theorem is illustrated in the following figure:




In Exercise 1, you will give a conventional geometric proof of the Circle Theorem. Here we will make the translation to polynomial form, showing that the Circle Theorem is admissible. We begin by constructing the triangle. Placing $A$ at (0,0) and $B$ at $(u_1, 0)$, the hypothesis that $\angle CAB$ is a right angle says $C = (0, u_2)$. (Of course, we are taking a shortcut here; we could also make $C$ a general point and add the hypothesis $CA \perp AB$, but that would lead to more variables and more equations.)

Next, we construct the three midpoints of the sides. These points have coordinates $M_1 = (x_1, 0)$, $M_2 = (0, x_2)$, and $M_3 = (x_3, x_4)$. As in Example 1, we use the convention that $u_1, u_2$ are to be arbitrary, where as the $x_j$ are determined by the values of $u_1, u_2$. Using part (vi) of Proposition 2, we obtain the equations

$$
\begin{aligned}
h_1 &= 2x_1 - u_1 = 0,\\
h_2 &= 2x_2 - u_2 = 0,\\
h_3 &= 2x_3 - u_1 = 0,\\
h_4 &= 2x_4 - u_2 = 0.
\end{aligned} \tag{5}
$$

The next step is to construct the point $H = (x_5, x_6)$, the foot of the altitude drawn from $A$. We have two hypotheses here:

$$
\begin{aligned}
AH \perp BC &: h_5 = x_5u_1 - x_6u_2 = 0,\\
B, H, C \text{ collinear} &: h_6 = x_5u_2 + x_6u_1 - u_1u_2 = 0.
\end{aligned} \tag{6}
$$

Finally, we must consider the statement that $M_1, M_2, M_3, H$ lie on a circle. A general collection of four points in the plane lies on *no* single circle (this is why the statement of the Circle Theorem is interesting). But *three* noncollinear points always do lie on a circle (the *circumscribed circle* of the triangle they form). Thus, our conclusion can be restated as follows: if we construct the circle containing the noncollinear triple $M_1, M_2, M_3$, then $H$ must lie on *this* circle also. To apply part (v) of Proposition 2, we must know the center of the circle, so this is an additional point that must be constructed. We call the center $O = (x_7, x_8)$ and derive two additional hypotheses:

$$
\begin{aligned}
M_1O = M_2O &: h_7 = (x_1 - x_7)^2 + x_8^2 - x_7^2 - (x_8 - x_2)^2 = 0,\\
M_1O = M_3O &: h_8 = (x_1 - x_7)^2 + (0 - x_8)^2 - (x_3 - x_7)^2 - (x_4 - x_8)^2 = 0.
\end{aligned} \tag{7}
$$

Our conclusion is $HO = M_1O$, which takes the form

$$
g = (x_5 - x_7)^2 + (x_6 - x_8)^2 - (x_1 - x_7)^2 - x_8^2 = 0. \tag{8}
$$

We remark that both here and in Example 1, the number of hypotheses and the number of *dependent* variables $x_j$ are *the same*. This is typical of properly posed geometric hypotheses. We expect that given values for the $u_i$, there should be at most finitely many different combinations of $x_j$ satisfying the equations.

We now consider the typical form of an admissible geometric theorem. We will have some number of arbitrary coordinates, or independent variables in our construction, denoted by $u_1, \ldots, u_m$. In addition, there will be some collection of dependent variables $x_1, \ldots, x_n$. The hypotheses of the theorem will be represented by a collection of polynomial equations in the $u_i, x_j$. As we noted in Example 3, it is typical of a properly posed theorem that the number of hypotheses is equal to the number of dependent variables, so we will write the hypotheses as

$$
\begin{aligned}
h_1(u_1, \ldots, u_m, x_1, \ldots, x_n) &= 0,\\
&\ \vdots\\
h_n(u_1, \ldots, u_m, x_1, \ldots, x_n) &= 0.
\end{aligned}
\tag{9}
$$

The conclusions of the theorem will also be expressed as polynomials in the $u_i, x_j$. It suffices to consider the case of *one* conclusion since if there are more, we can simply treat them one at a time. Hence, we will write the conclusion as

$$g(u_1, \ldots, u_m, x_1, \ldots, x_n) = 0.$$

The question to be addressed is: how can the fact that $g$ follows from $h_1, \ldots, h_n$ be deduced *algebraically*? The basic idea is that we want $g$ to vanish whenever $h_1, \ldots, h_n$ do. We observe that the hypotheses (9) are equations that define a *variety*

$$V = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{R}^{m+n}.$$

This leads to the following definition.

**Definition 4.** *The conclusion $g$* **follows strictly** *from the hypotheses $h_1, \ldots, h_n$ if $g \in \mathbf{I}(V) \subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$, where $V = \mathbf{V}(h_1, \ldots, h_n)$.*

Although this definition seems reasonable, we will see later that it is too strict. Most geometric theorems have some "degenerate" cases that Definition 4 does not take into account. But for the time being, we will use the above notion of "follows strictly."

One drawback of Definition 4 is that because we are working over $\mathbb{R}$, we do not have an effective method for determining $\mathbf{I}(V)$. But we still have the following useful criterion.

**Proposition 5.** *If $g \in \sqrt{\langle h_1, \ldots, h_n\rangle}$, then $g$ follows strictly from $h_1, \ldots, h_n$.*

**Proof.** The hypothesis $g \in \sqrt{\langle h_1, \ldots, h_n\rangle}$ implies that $g^s \in \langle h_1, \ldots, h_n\rangle$ for some $s$. Thus, $g^s = \sum_{i=1}^n A_i h_i$, where $A_i \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$. Then $g^s$, and, hence, $g$ itself, must vanish whenever $h_1, \ldots, h_n$ do. □

Note that the converse of this proposition fails whenever $\sqrt{\langle h_1, \ldots, h_n\rangle} \subsetneq \mathbf{I}(V)$, which can easily happen when working over $\mathbb{R}$. Nevertheless, Proposition 5 is still useful because we can test whether $g \in \sqrt{\langle h_1, \ldots, h_n\rangle}$ using the radical membership algorithm from Chapter 4, §2. Let $\bar{I} = \langle h_1, \ldots, h_n, 1 - yg\rangle$ in the ring $\mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n, y]$. Then Proposition 8 of Chapter 4, §2 implies that

$$g \in \sqrt{\langle h_1, \ldots, h_n\rangle} \iff \{1\} \text{ is the reduced Groebner basis of } \bar{I}.$$

If this condition is satisfied, then $g$ follows strictly from $h_1, \ldots, h_n$.

If we work over $\mathbb{C}$, we can get a better sense of what $g \in \sqrt{\langle h_1, \ldots, h_n\rangle}$ means. By allowing solutions in $\mathbb{C}$, the hypotheses $h_1, \ldots, h_n$ define a variety $V_{\mathbb{C}} \subset \mathbb{C}^{m+n}$. Then, in Exercise 9, you will use the Strong Nullstellensatz to show that

$$\begin{aligned} g \in \sqrt{\langle h_1, \ldots, h_n \rangle} &\subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n] \\ \iff g \in \mathbf{I}(V_{\mathbb{C}}) &\subset \mathbb{C}[u_1, \ldots, u_m, x_1, \ldots, x_n]. \end{aligned}$$

Thus, $g \in \sqrt{\langle h_1, \ldots, h_n \rangle}$ means that $g$ "follows strictly over $\mathbb{C}$" from $h_1, \ldots, h_n$.

Let us apply these concepts to an example. This will reveal why Definition 4 is too strong.

**Example 1.** [continued] To see what can go wrong if we proceed as above, consider the theorem on the diagonals of a parallelogram from Example 1, taking as hypotheses the four polynomials from (1) and (2):

$$\begin{aligned} h_1 &= x_2 - u_3, \\ h_2 &= (x_1 - u_1)u_3 - u_2 x_2, \\ h_3 &= x_4 x_1 - x_3 u_3, \\ h_4 &= x_4(u_2 - u_1) - (x_3 - u_1)u_3. \end{aligned}$$

We will take as conclusion the *first* polynomial from (3):

$$g = x_1^2 - 2x_1x_3 - 2x_4x_2 + x_2^2.$$

To apply Proposition 5, we must compute a Groebner basis for

$$\bar{I} = \langle h_1, h_2, h_3, h_4, 1 - yg \rangle \subset \mathbb{R}[u_1, u_2, u_3, x_1, x_2, x_3, x_4, y].$$

Surprisingly enough, we do *not* find $\{1\}$. (You will use a computer algebra system in Exercise 10 to verify this.) Since the statement is a true geometric theorem, we must try to understand why our proposed method failed in this case.

The reason can be seen by computing a Groebner basis for $I = \langle h_1, h_2, h_3, h_4 \rangle$ in $\mathbb{R}[u_1, u_2, u_3, x_1, x_2, x_3, x_4]$, using lex order with $x_1 > x_2 > x_3 > x_4 > u_1 > u_2 > u_3$. The result is

$$\begin{aligned} f_1 &= x_1x_4 + x_4u_1 - x_4u_2 - u_1u_3, \\ f_2 &= x_1u_3 - u_1u_3 - u_2u_3, \\ f_3 &= x_2 - u_3, \\ f_4 &= x_3u_3 + x_4u_1 - x_4u_2 - u_1u_3, \\ f_5 &= x_4u_1^2 - x_4u_1u_2 - \frac{1}{2}u_1^2u_3 + \frac{1}{2}u_1u_2u_3, \\ f_6 &= x_4u_1u_3 - \frac{1}{2}u_1u_3^2. \end{aligned}$$

The variety $V = \mathbf{V}(h_1, h_2, h_3, h_4) = \mathbf{V}(f_1, \ldots, f_6)$ in $\mathbb{R}^7$ defined by the hypotheses is actually *reducible*. To see this, note that $f_2$ factors as $(x_1 - u_1 - u_2)u_3$, which implies that

$$V = \mathbf{V}(f_1, x_1 - u_1 - u_2, f_3, f_4, f_5, f_6) \cup \mathbf{V}(f_1, u_3, f_3, f_4, f_5, f_6).$$

Since $f_5$ and $f_6$ also factor, we can continue this decomposition process. Things simplify dramatically if we recompute the Groebner basis at each stage, and, in the exercises, you will show that this leads to the decomposition

$$V = V' \cup U_1 \cup U_2 \cup U_3$$

into irreducible varieties, where

$$\begin{aligned}
V' &= \mathbf{V}\left(x_1 - u_1 - u_2, x_2 - u_3, x_3 - \frac{u_1 + u_2}{2}, x_4 - \frac{u_3}{2}\right),\\
U_1 &= \mathbf{V}(x_2, x_4, u_3),\\
U_2 &= \mathbf{V}(x_1, x_2, u_1 - u_2, u_3),\\
U_3 &= \mathbf{V}(x_1 - u_2, x_2 - u_3, x_3u_3 - x_4u_2, u_1).
\end{aligned}$$

You will also show that none of these varieties are contained in the others, so that $V', U_1, U_2, U_3$ are the irreducible components of $V$.

The problem becomes apparent when we interpret the components $U_1, U_2, U_3 \subset V$ in terms of the parallelogram *ABDC*. On $U_1$ and $U_2$, we have $u_3 = 0$. This is troubling since $u_3$ was supposed to be arbitrary. Further, when $u_3 = 0$, the vertex $C$ of our paralleogram lies on $\overline{AB}$ and, hence we do not have a parallelogram at all. This is a *degenerate* case of our configuration, which we intended to *rule out* by the hypothesis that *ABDC* was an honest parallelogram in the plane. Similarly, we have $u_1 = 0$ on $U_3$, which again is a degenerate configuration.

You can also check that on $U_1 = \mathbf{V}(x_2, x_4, u_3)$, our conclusion $g$ becomes $g = x_1^2 - 2x_1x_3$, which is *not* zero since $x_1$ and $x_3$ are arbitrary on $U_1$. This explains why our first attempt failed to prove the theorem. Once we exclude the degenerate cases $U_1, U_2, U_3$, the above method easily shows that $g$ vanishes on $V'$. We leave the details as an exercise.

Our goal is to develop a general method that can be used to decide the validity of a theorem, taking into account any degenerate special cases that may need to be excluded. To begin, we use Theorem 2 of Chapter 4, §6 to write $V = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{R}^{m+n}$ as a finite union of irreducible varieties,

$$V = V_1 \cup \cdots \cup V_k. \tag{10}$$

As we saw in the continuation of Example 1, it may be the case that some polynomial equation involving *only* the $u_i$ holds on one or more of these irreducible components of $V$. Since our intent is that the $u_i$ should be essentially independent, we want to *exclude* these components from consideration if they are present. We introduce the following terminology.

**Definition 6.** *Let $W$ be an irreducible variety in the affine space $\mathbb{R}^{m+n}$ with coordinates $u_1, \ldots, u_m, x_1, \ldots, x_n$. We say that the functions $u_1, \ldots, u_m$ are* **algebraically independent on** *$W$ if no nonzero polynomial in the $u_i$ alone vanishes identically on $W$.*

Equivalently, Definition 6 states that $u_1, \ldots, u_m$ are algebraically independent on $W$ if $\mathbf{I}(W) \cap \mathbb{R}[u_1, \ldots, u_m] = \{0\}$.

Thus, in the decomposition of the variety $V$ given in (10), we can regroup the irreducible components in the following way:

$$V = W_1 \cup \cdots \cup W_p \cup U_1 \cup \cdots \cup U_q, \tag{11}$$

where $u_1, \ldots, u_m$ *are* algebraically independent on the components $W_i$ and *are not* algebraically independent on the components $U_j$. Thus, the $U_j$, represent "degenerate" cases of the hypotheses of our theorem. To ensure that the variables $u_i$ are actually arbitrary in the geometric configurations we study, we should consider *only* the subvariety

$$V' = W_1 \cup \cdots \cup W_p \subset V.$$

Given a conclusion $g \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ we want to prove, we are not interested in how $g$ behaves on the degenerate cases. This leads to the following definition.

**Definition 7.** *The conclusion $g$* **follows generically** *from the hypotheses $h_1, \ldots, h_n$ if $g \in \mathbf{I}(V') \subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$, where, as above, $V' \subset \mathbb{R}^{m+n}$ is the union of the components of the variety $V = \mathbf{V}(h_1, \ldots, h_n)$ on which the $u_i$ are algebraically independent.*

Saying a geometric theorem is "true" in the usual sense means precisely that its conclusion(s) follow generically from its hypotheses. The question becomes, given a conclusion $g$: Can we determine when $g \in \mathbf{I}(V')$? That is, can we develop a criterion that determines whether $g$ vanishes on every component of $V$ on which the $u_i$ *are algebraically independent*, ignoring what happens on the possible "degenerate" components?

Determining the decomposition of a variety into irreducible components is not always easy, so we would like a method to determine whether a conclusion follows generically from a set of hypotheses that *does not* require knowledge of the decomposition (11). Further, even if we could find $V'$, we would still have the problem of computing $\mathbf{I}(V')$.

Fortunately, it is possible to show that $g$ follows generically from $h_1, \ldots, h_n$ without knowing the decomposition of $V$ given in (11). We have the following result.

**Proposition 8.** *In the situation described above, $g$ follows generically from $h_1, \ldots, h_n$ whenever there is some nonzero polynomial $c(u_1, \ldots, u_m) \in \mathbb{R}[u_1, \ldots, u_m]$ such that*

$$c \cdot g \in \sqrt{H},$$

*where $H$ is the ideal generated by the hypotheses $h_i$ in $\mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$.*

**Proof.** Let $V_j$ be one of the irreducible components of $V'$. Since $c \cdot g \in \sqrt{H}$, we see that $c \cdot g$ vanishes on $V$ and, hence, on $V_j$. Thus, the product $c \cdot g$ is in $\mathbf{I}(V_j)$. But $V_j$ is irreducible, so that $\mathbf{I}(V_j)$ is a prime ideal by Proposition 3 of Chapter 4, §5. Thus,

$c \cdot g \in \mathbf{I}(V_j)$ implies either $c$ or $g$ is in $\mathbf{I}(V_j)$. We know $c \notin \mathbf{I}(V_j)$ since no nonzero polynomial in the $u_i$ alone vanishes on this component. Hence, $g \in \mathbf{I}(V_j)$, and since this is true for each component of $V'$, it follows that $g \in \mathbf{I}(V')$. □

For Proposition 8 to give a practical way of determining whether a conclusion follows generically from a set of hypotheses, we need a criterion for deciding when there is a nonzero polynomial $c$ with $c \cdot g \in \sqrt{H}$. This is actually quite easy to do. By the definition of the radical, we know that $c \cdot g \in \sqrt{H}$ if and only if

$$(c \cdot g)^s = \sum_{j=1}^{n} A_j h_j$$

for some $A_j \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ and $s \geq 1$. If we divide both sides of this equation by $c^s$, we obtain

$$g^s = \sum_{j=1}^{n} \frac{A_j}{c^s} h_j,$$

which shows that $g$ is in the radical of the ideal $\widetilde{H}$ generated by $h_1, \ldots, h_n$ over the ring $\mathbb{R}(u_1, \ldots, u_m)[x_1, \ldots, x_n]$ (in which we allow denominators depending only on the $u_i$). Conversely, if $g \in \sqrt{\widetilde{H}}$, then

$$g^s = \sum_{j=1}^{n} B_j h_j,$$

where the $B_j \in \mathbb{R}(u_1, \ldots, u_m)[x_1, \ldots, x_n]$. If we find a least common denominator $c$ for all terms in all the $B_j$ and multiply both sides by $c^s$ (clearing denominators in the process), we obtain

$$(c \cdot g)^s = \sum_{j=1}^{n} B'_j h_j,$$

where $B'_j \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ and $c$ depends only on the $u_i$. As a result, $c \cdot g \in \sqrt{H}$. These calculations and the radical membership algorithm from §2 of Chapter 4 establish the following corollary of Proposition 8.

**Corollary 9.** *In the situation of Proposition 8, the following are equivalent:*

(i) *There is a nonzero polynomial* $c \in \mathbb{R}[u_1, \ldots, u_m]$ *such that* $c \cdot g \in \sqrt{H}$.

(ii) $g \in \sqrt{\widetilde{H}}$, *where* $\widetilde{H}$ *is the ideal generated by the* $h_j$ *in* $\mathbb{R}(u_1, \ldots, u_m)[x_1, \ldots, x_n]$.

(iii) $\{1\}$ *is the reduced Groebner basis of the ideal*

$$\langle h_1, \ldots, h_n, 1 - yg \rangle \subset \mathbb{R}(u_1, \ldots, u_m)[x_1, \ldots, x_n, y].$$

If we combine part (iii) of this corollary with Proposition 8, we get an algorithmic method for proving that a conclusion follows generically from a set of hypotheses. We will call this the Groebner basis method in geometric theorem proving.

To illustrate the use of this method, we will consider the theorem on parallelograms from Example 1 once more. We compute a Groebner basis of $\langle h_1, h_2, h_3, h_4, 1 - yg\rangle$ in the ring $\mathbb{R}(u_1, u_2, u_3)[x_1, x_2, x_3, x_4, y]$. This computation *does yield* $\{1\}$ as we expect. Making $u_1, u_2, u_3$ invertible by passing to $\mathbb{R}(u_1, u_2, u_3)$ as our field of coefficients in effect *removes* the degenerate cases encountered above, and the conclusion *does* follow generically from the hypotheses. Moreover, in Exercise 12, you will see that $g$ itself (and not some higher power) actually lies in $\langle h_1, h_2, h_3, h_4\rangle \subset \mathbb{R}(u_1, u_2, u_2)[x_1, x_2, x_3, x_4]$.

Note that the Groebner basis method does not tell us what the degenerate cases are. The information about these cases is contained in the polynomial $c \in \mathbb{R}[u_1, \ldots, u_m]$, for $c \cdot g \in \sqrt{H}$ tells us that $g$ follows from $h_1, \ldots, h_n$ whenever $c$ does not vanish (this is because $c \cdot g$ vanishes on $V$). In Exercise 14, we will give an algorithm for finding $c$.

Over $\mathbb{C}$, we can think of Corollary 9 in terms of the variety $V_{\mathbb{C}} = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{C}^{m+n}$ as follows. Decomposing $V_{\mathbb{C}}$ as in (11), let $V'_{\mathbb{C}} \subset V_{\mathbb{C}}$ be the union of those components where the $u_i$ are algebraically independent. Then Exercise 15 will use the Nullstellensatz to prove that

$$\begin{aligned} \exists c \neq 0 \text{ in } \mathbb{R}[u_1, \ldots, u_m] \ \text{ with } c \cdot g \in \sqrt{\langle h_1, \ldots, h_n\rangle} \subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_u] \\ \iff g \in \mathbf{I}(V'_{\mathbb{C}}) \subset \mathbb{C}[u_1, \ldots, u_m, x_1, \ldots, x_n]. \end{aligned}$$

Thus, the conditions of Corollary 9 mean that $g$ "follows generically over $\mathbb{C}$" from the hypotheses $h_1, \ldots, h_n$.

This interpretation points out what is perhaps the main limitation of the Groebner basis method in geometric theorem proving: it can only prove theorems where the conclusions follow generically over $\mathbb{C}$, even though we are only interested in what happens over $\mathbb{R}$. In particular, there are theorems which are true over $\mathbb{R}$ but not over $\mathbb{C}$ [see STURMFELS (1989) for an example]. Our methods will fail for such theorems.

When using Corollary 9, it is often unnecessary to consider the radical of $\widetilde{H}$. In many cases, the first power of the conclusion is in $\widetilde{H}$ already. So most theorem proving programs in effect use an *ideal* membership algorithm first to test if $g \in \widetilde{H}$, and only go on to the radical membership test if that initial step fails.

To illustrate this, we continue with the Circle Theorem of Apollonius from Example 3. Our hypotheses are the eight polynomials $h_i$ from (5)–(7). We begin by computing a Groebner basis (using lex order) for the ideal $\widetilde{H}$, which yields

$$\begin{aligned} f_1 &= x_1 - u_1/2, \\ f_2 &= x_2 - u_2/2, \\ f_3 &= x_3 - u_1/2, \\ f_4 &= x_4 - u_2/2, \\ f_5 &= x_5 - \frac{u_1 u_2^2}{u_1^2 + u_2^2}, \\ f_6 &= x_6 - \frac{u_1^2 u_2}{u_1^2 + u_2^2}, \\ f_7 &= x_7 - u_1/4, \\ f_8 &= x_8 - u_2/4. \end{aligned} \tag{12}$$

We leave it as an exercise to show that the conclusion (8) reduces to zero on division by this Groebner basis. Thus, $g$ itself is in $\widetilde{H}$, which shows that $g$ follows generically from $h_1, \ldots, h_8$. Note that we must have either $u_1 \neq 0$ or $u_2 \neq 0$ in order to solve for $x_5$ and $x_6$. The equations $u_1 = 0$ and $u_2 = 0$ describe degenerate right "triangles" in which the three vertices are not distinct, so we certainly wish to rule these cases out. It is interesting to note, however, that if either $u_1$ or $u_2$ is nonzero, the conclusion is still true. For instance, if $u_1 \neq 0$ but $u_2 = 0$, then the vertices $C$ and $A$ coincide. From (5) and (6), the midpoints $M_1$ and $M_3$ coincide, $M_2$ coincides with $A$, and $H$ coincides with $A$ as well. As a result, there *is* a circle (infinitely many of them in fact) containing $M_1, M_2, M_3$, and $H$ in this degenerate case. In Exercise 16, you will study what happens when $u_1 = u_2 = 0$.

We conclude this section by noting that there is one further subtlety that can occur when we use this method to prove or verify a theorem. Namely, there are cases where the given statement of a geometric theorem conceals one or more unstated "extra" hypotheses. These may very well not be included when we make a direct translation to a system of polynomial equations. This often results in a situation where the variety $V'$ is *reducible* or, equivalently, where $p \geq 2$ in (11). In this case, it may be true that the intended conclusion is zero *only* on some of the reducible components of $V'$, so that any method based on Corollary 9 would fail. We will study an example of this type in Exercise 17. If this happens, we may need to reformulate our hypotheses to exclude the extraneous, unwanted components of $V'$.

**EXERCISES FOR §4**

1. This exercise asks you to give geometric proofs of the theorems that we studied in Examples 1 and 3.
   a. Give a standard Euclidean proof of the theorem of Example 1. Hint: Show $\Delta ANC \cong \Delta BND$.
   b. Give a standard Euclidean proof of the Circle Theorem of Apollonius from Example 3. Hint: First show that $\overline{AB}$ and $\overline{M_2M_3}$ are parallel.
2. This exercise shows that it is possible to give translations of a theorem based on different collections of arbitrary coordinates. Consider the parallelogram *ABDC* from Example 1 and begin by placing $A$ at the origin.
   a. Explain why it is also possible to consider both of the coordinates of $D$ as arbitrary variables: $D = (u_1, u_2)$.
   b. With this choice, explain why we can specify the coordinates of $B$ as $B = (u_3, x_1)$. That is, the $x$-coordinate of $B$ is arbitrary, but the $y$-coordinate is determined by the choices of $u_1, u_2, u_3$.
   c. Complete the translation of the theorem based on this choice of coordinates.
3. Let $A, B, C, D, E, F, G, H$ be points in the plane.
   a. Show that the statement $\overline{AB}$ is tangent to the circle through $A, C, D$ can be expressed by polynomial equations. Hint: Construct the center of the circle first. Then, what is true about the tangent and the radius of a circle at a given point?
   b. Show that the statement $AB \cdot CD = EF \cdot GH$ can be expressed by one or more polynomial equations.
   c. Show that the statement $\frac{AB}{CD} = \frac{EF}{GH}$ can be expressed by one or more polynomial equations.

d. The *cross ratio* of the ordered 4-tuple of distinct *collinear* points $(A, B, C, D)$ is defined to be the real number

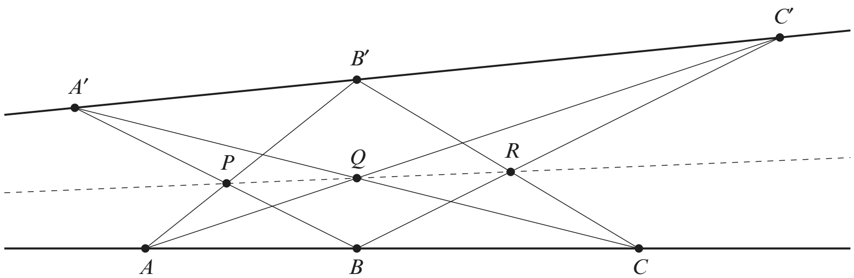
$$\frac{AC \cdot BD}{AD \cdot BC}.$$

Show that the statement "The cross ratio of $(A, B, C, D)$ is equal to $\rho \in \mathbb{R}$" can be expressed by one or more polynomial equations.

4. In this exercise, you will complete the proof of Proposition 2 in the text.
   a. Prove part (ii).
   b. Show that if $\alpha, \beta$ are acute angles, then $\alpha = \beta$ if and only if $\tan\alpha = \tan\beta$. Use this fact and part (c) of Exercise 3 to prove part (vii) of Proposition 2. Hint: To compute the tangent of an angle, you can construct an appropriate right triangle and compute a ratio of side lengths.
   c. Prove part (viii).
5. Let $\Delta ABC$ be a triangle in the plane. Recall that the *altitude* from $A$ is the line segment from $A$ meeting the opposite side $\overline{BC}$ at a right angle. (We may have to extend $\overline{BC}$ here to find the intersection point.) A standard geometric theorem asserts that the three altitudes of a triangle meet at a single point $H$, often called the *orthocenter* of the triangle. Give a translation of the hypotheses and conclusion of this theorem as a system of polynomial equations.
6. Let $\Delta ABC$ be a triangle in the plane. It is a standard theorem that if we let $M_1$ be the midpoint of $\overline{BC}$, $M_2$ be the midpoint of $\overline{AC}$ and $M_3$ be the midpoint of $\overline{AB}$, then the segments $\overline{AM_1}$, $\overline{BM_2}$ and $\overline{CM_3}$ meet at a single point $M$, often called the *centroid* of the triangle. Give a translation of the hypotheses and conclusion of this theorem as a system of polynomial equations.
7. Let $\Delta ABC$ be a triangle in the plane. It is a famous theorem of Euler that the *circumcenter* (the center of the circumscribed circle), the *orthocenter* (from Exercise 5), and the *centroid* (from Exercise 6) are always *collinear*. Translate the hypotheses and conclusion of this theorem into a system of polynomial equations. (The line containing the three "centers" of the triangle is called the *Euler line* of the triangle.)
8. A beautiful theorem ascribed to Pappus concerns two collinear triples of points $A, B, C$ and $A', B', C'$. Let

$$\begin{aligned} P &= \overline{AB'} \cap \overline{A'B}, \\ Q &= \overline{AC'} \cap \overline{A'C}, \\ R &= \overline{BC'} \cap \overline{B'C} \end{aligned}$$

be as in the figure:



Then it is always the case that $P, Q, R$ are collinear points. Give a translation of the hypotheses and conclusion of this theorem as a system of polynomial equations.

9. Given $h_1, \ldots, h_n \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$, let $V_{\mathbb{C}} = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{C}^{m+n}$. If $g \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$, the goal of this exercise is to prove that
$$\begin{aligned} g \in \sqrt{\langle h_1, \ldots, h_n \rangle} &\subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n] \\ \Longleftrightarrow g \in \mathbf{I}(V_{\mathbb{C}}) &\subset \mathbb{C}[u_1, \ldots, u_m, x_1, \ldots, x_n]. \end{aligned}$$
   a. Prove the $\Rightarrow$ implication.
   b. Use the Strong Nullstellensatz to show that if $g \in \mathbf{I}(V_{\mathbb{C}})$, then there are polynomials $A_j \in \mathbb{C}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ such that $g^s = \sum_{j=1}^n A_j h_j$ for some $s \geq 1$.
   c. Explain why $A_j$ can be written $A_j = A'_j + iA''_j$, where $A'_j, A''_j$ are polynomials with real coefficients. Use this to conclude that $g^s = \sum_{j=1}^n a'_j h_j$, which will complete the proof of the $\Leftarrow$ implication. Hint: $g$ and $h_1, \ldots, h_n$ have real coefficients.
10. Verify the claim made in Example 1 that $\{1\}$ is not the unique reduced Groebner basis for the ideal $\overline{I} = \langle h_1, h_2, h_3, h_4, 1 - yg \rangle$.
11. This exercise will study the decomposition into reducible components of the variety defined by the hypotheses of the theorem from Example 1.
   a. Verify the claim made in the continuation of Example 1 that
$$V = \mathbf{V}(f_1, x_1 - u_1 - u_2, f_3, \ldots, f_6) \cup \mathbf{V}(f_1, u_3, f_3, \ldots, f_6) = V_1 \cup V_2.$$
   b. Compute Groebner bases for the defining equations of $V_1$ and $V_2$. Some of the polynomials should factor and use this to decompose $V_1$ and $V_2$.
   c. By continuing this process, show that $V$ is the union of the varieties $V', U_1, U_2, U_3$ defined in the text.
   d. Prove that $V', U_1, U_2, U_3$ are irreducible and that none of them are contained in the union of the others. This shows that $V', U_1, U_2, U_3$ are the reducible components of $V$.
   e. On which irreducible component of $V$ is the conclusion of the theorem valid?
   f. Suppose we take as hypotheses the four polynomials in (4) and (2). Is $W = \mathbf{V}(h'_1, h'_2, h_3, h_4)$ reducible? How many components does it have?
12. Verify the claim made in Example 1 that the conclusion $g$ itself (and not some higher power) is in the ideal generated by $h_1, h_2, h_3, h_4$ in $\mathbb{R}(u_1, u_2, u_3)[x_1, x_2, x_3, x_4]$.
13. By applying part (iii) of Corollary 9, verify that $g$ follows generically from the $h_j$ for each of the following theorems. What is the lowest power of $g$ which is contained in the ideal $\widetilde{H}$ in each case?
   a. the theorem on the orthocenter of a triangle (Exercise 5),
   b. the theorem on the centroid of a triangle (Exercise 6),
   c. the theorem on the Euler line of a triangle (Exercise 7),
   d. Pappus's Theorem (Exercise 8).
14. In this exercise, we will give an algorithm for finding a nonzero $c \in \mathbb{R}[u_1, \ldots, u_m]$ such that $c \cdot g \in \sqrt{H}$, assuming that such a $c$ exists. We will work with the ideal
$$\overline{H} = \langle h_1, \ldots, h_n, 1 - yg \rangle \subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n, y].$$
   a. Show that the conditions of Corollary 9 are equivalent to $\overline{H} \cap \mathbb{R}[u_1, \ldots, u_m] \neq \{0\}$. Hint: Use condition (iii) of the corollary.
   b. If $c \in \overline{H} \cap \mathbb{R}[u_1, \ldots, u_m]$, prove that $c \cdot g \in \sqrt{H}$. Hint: Adapt the argument used in equations (2)–(4) in the proof of Hilbert's Nullstellensatz in Chapter 4, §1.
   c. Describe an algorithm for computing $\overline{H} \cap \mathbb{R}[u_1, \ldots, u_m]$. For maximum efficiency, what monomial order should you use?

Parts (a)–(c) give an algorithm which decides if there is a nonzero $c$ with $c \cdot g \in \sqrt{H}$ and *simultaneously* produces the required $c$. Parts (d) and (e) below give some interesting properties of the ideal $\overline{H} \cap \mathbb{R}[u_1, \ldots, u_m]$.

d. Show that if the conclusion $g$ fails to hold for some choice of $u_1, \ldots, u_m$, then $(u_1, \ldots, u_m) \in W = \mathbf{V}(\overline{H} \cap \mathbb{R}[u_1, \ldots, u_m]) \subset \mathbb{R}^m$. Thus, $W$ records the degenerate cases where $g$ fails.

e. Show that $\sqrt{\overline{H} \cap \mathbb{R}[u_1, \ldots, u_m]}$ gives *all* $c$'s for which $c \cdot g \in \sqrt{H}$. Hint: One direction follows from part (a). If $c \cdot g \in \sqrt{H}$, note the $\overline{H}$ contains $(c \cdot g)$'s and $1 - gy$. Now adapt the argument given in Proposition 8 of Chapter 4, §2 to show that $c^s \in \overline{H}$.

15. As in Exercise 9, suppose that we have $h_1, \ldots, h_n \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$. Then we get $V_{\mathbb{C}} = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{C}^{m+n}$. As we did with $V$, let $V'_{\mathbb{C}}$ be the union of the irreducible components of $V_{\mathbb{C}}$ where $u_1, \ldots, u_n$ are algebraically independent. Given $g \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$, we want to show that

$$\begin{aligned} &\exists c \neq 0 \text{ in } \mathbb{R}[u_1, \ldots, u_m] \text{ with } c \cdot g \in \sqrt{\langle h_1, \ldots, h_n \rangle} \subset \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n] \\ &\quad \Longleftrightarrow g \in \mathbf{I}(V'_{\mathbb{C}}) \subset \mathbb{C}[u_1, \ldots, u_m, x_1, \ldots, x_n]. \end{aligned}$$

a. Prove the $\Rightarrow$ implication. Hint: See the proof of Proposition 8.

b. Show that if $g \in \mathbf{I}(V'_{\mathbb{C}})$, then there is a nonzero polynomial $c \in \mathbb{C}[u_1, \ldots, u_m]$ such that $c \cdot g \in \mathbf{I}(V_{\mathbb{C}})$. Hint: Write $V_{\mathbb{C}} = V'_{\mathbb{C}} \cup U'_1 \cup \cdots \cup U'_q$, where $u_1, \ldots, u_m$ are algebraically dependent on each $U'_j$. This means there is a nonzero polynomial $c_j \in \mathbb{C}[u_1, \ldots, u_m]$ which vanishes on $U'_j$.

c. Show that the polynomial $c$ of part b can be chosen to have real coefficients. Hint: If $\overline{c}$ is the polynomial obtained from $c$ by taking the complex conjugates of the coefficients, show that $\overline{c}$ has real coefficients.

d. Once we have $c \in \mathbb{R}[u_1, \ldots, u_m]$ with $c \cdot g \in \mathbf{I}(V_{\mathbb{C}})$, use Exercise 9 to complete the proof of the $\Leftarrow$ implication.

16. This exercise deals with the Circle Theorem of Apollonius from Example 3.

a. Show that the conclusion (8) reduces to 0 on division by the Groebner basis (12) given in the text.

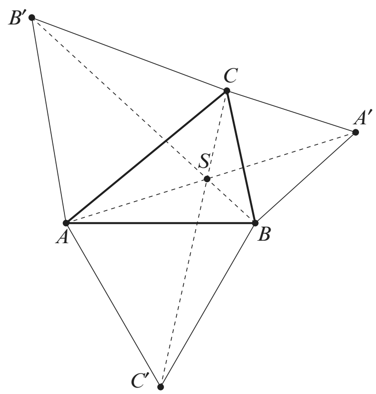
b. Discuss the case $u_1 = u_2 = 0$ in the Circle Theorem. Does the conclusion follow in this degenerate case?

c. Note that in the diagram in the text illustrating the Circle Theorem, the circle is shown passing through the vertex $A$ in addition to the three midpoints and the foot of the altitude drawn from $A$. Does this conclusion also follow from the hypotheses?

17. In this exercise, we will study a case where a direct translation of the hypotheses of a "true" theorem leads to extraneous components on which the conclusion is actually false. Let $\Delta ABC$ be a triangle in the plane. We construct three new points $A', B', C'$ such that the triangles $\Delta A'BC$, $\Delta AB'C$, $\Delta ABC'$ are *equilateral*. The intended construction is illustrated in the figure on the next page.

Our theorem is that the three line segments $\overline{AA'}, \overline{BB'}, \overline{CC'}$ all meet in a single point $S$. (We call $S$ *the Steiner point* or *Fermat point* of the triangle. If no angle of the original triangle was greater than $\frac{2\pi}{3}$, it can be shown that the three segments $\overline{AS}, \overline{BS}, \overline{CS}$ form the network of shortest total length connecting the points $A, B, C$.)

a. Give a conventional geometric proof of the theorem, assuming the construction is done as in the figure.




b. Now, translate the hypotheses and conclusion of this theorem directly into a set of polynomial equations.
c. Apply the test based on Corollary 9 to determine whether the conclusion follows generically from the hypotheses. The test should fail. Note: This computation may require a *great* deal of ingenuity to push through on some computer algebra systems. This is a complicated system of polynomials.
d. (The key point) Show that there are other ways to construct a figure which is consistent with the hypotheses *as stated*, but which do not agree with the figure above. Hint: Are the points $A'$, $B'$, $C'$ uniquely determined by the hypotheses as stated? Is the statement of the theorem valid for these alternate constructions of the figure? Use this to explain why part c did not yield the expected result. (These alternate constructions correspond to points in different components of the variety defined by the hypotheses.)
e. How can the hypotheses of the theorem be reformulated to exclude the extraneous components?

## §5 Wu's Method

In this section, we will study a second commonly used algorithmic method for proving theorems in Euclidean geometry based on systems of polynomial equations. This method, introduced by the Chinese mathematician Wu Wen-Tsün, was developed *before* the Groebner basis method given in §4. It is also more commonly used than the Groebner basis method in practice because it is usually more efficient.

Both the elementary version of Wu's method that we will present, and the more refined versions, use an interesting variant of the division algorithm for multivariable polynomials introduced in Chapter 2, §3. The idea here is to follow the one-variable polynomial division algorithm as closely as possible, and we obtain a result known as the *pseudodivision* algorithm. To describe the first step in the process, we consider two polynomials in the ring $k[x_1, \ldots, x_n, y]$, written in the form

$$
\begin{aligned}
f &= c_p y^p + \cdots + c_1 y + c_0,\\
g &= d_m y^m + \cdots + d_1 y + d_0,
\end{aligned}
\tag{1}
$$

where the coefficients $c_i, d_j$ are polynomials in $x_1, \ldots, x_n$. Assume that $m \leq p$. Proceeding as in the one-variable division algorithm for polynomials in $y$, we can attempt to remove the leading term $c_p y^p$ in $f$ by subtracting a multiple of $g$. However, this is not possible directly unless $d_m$ divides $c_p$ in $k[x_1, \ldots, x_n]$. In pseudodivision, we first *multiply* $f$ by $d_m$ to ensure that the leading coefficient is divisible by $d_m$, then proceed as in one-variable division. We can state the algorithm formally as follows.

**Proposition 1.** *Let $f, g \in k[x_1, \ldots, x_n, y]$ be as in* (1) *and assume $m \leq p$ and $g \neq 0$.*
(i) *There is an equation*
$$d_m^s f = qg + r,$$
*where $q, r \in k[x_1, \ldots, x_n, y]$, $s \geq 0$, and either $r = 0$ or the degree of $r$ in $y$ is less than $m$.*
(ii) *$r \in \langle f, g\rangle$ in the ring $k[x_1, \ldots, x_n, y]$.*

**Proof.** (i) Polynomials $q, r$ satisfying the conditions of the proposition can be constructed by the following algorithm, called *pseudodivision with respect to y*. We use the notation $\deg(h, y)$ for the degree of the polynomial $h$ in the variable $y$ and $LC(h, y)$ for the leading coefficient of $h$ as a polynomial in $y$—that is, the coefficient of $y^{\deg(h,y)}$ in $h$.

Input: $f, g$
Output: $q, r$

$r := f; q := 0$
While $r \neq 0$ AND $\deg(r, y) \geq m$ DO
  $r := d_m r - LC(r, y) g y^{\deg(r,y)-m}$
  $q := d_m q + LC(r, y) y^{\deg(r,y)-m}$

Note that if we follow this procedure, the body of the WHILE loop will be executed at most $p - m + 1$ times. Thus, the power $s$ in $d_m^s f = qg + r$ can be chosen so that $s \leq p - m + 1$. We leave the rest of the proof, and the consideration of whether $q, r$ are unique, to the reader as Exercise 1.

From $d_m^s f = qg + r$, it follows that $r = d_m^s f - qg \in \langle f, g\rangle$, which completes the proof of the proposition. □

The polynomials $q, r$ are known as a *pseudoquotient* and a *pseudoremainder* of $f$ on pseudodivision by $g$, with respect to the variable $y$. We will use the notation $\mathrm{Rem}(f, g, y)$ for the pseudoremainder produced by the algorithm given in the proof of Proposition 1. For example, if we pseudodivide $f = x^2y^3 - y$ by $g = x^3y - 2$ with respect to $y$ by the algorithm above, we obtain the equation
$$(x^3)^3 f = (x^8y^2 + 2x^5y + 4x^2 - x^6)g + 8x^2 - 2x^6.$$
In particular, the pseudoremainder is $\mathrm{Rem}(f, g, y) = 8x^2 - 2x^6$.

We note that there is a second, "slicker" way to understand what is happening in this algorithm. The same idea of *allowing denominators* that we exploited in §4 shows that pseudodivision is the same as

- ordinary one-variable polynomial division for polynomials in $y$, with coefficients in the rational function field $K = k(x_1, \ldots, x_n)$, followed by
- clearing denominators. You will establish this claim in Exercise 2, based on the observation that the only term that needs to be inverted in division of polynomials in $K[y]$($K$ any field) is the *leading coefficient* $d_m$ of the divisor $g$. Thus, the denominators introduced in the process of dividing $f$ by $g$ can all be cleared by multiplying by a suitable power $d_m^s$, and we get an equation of the form $d_m^s f = qg + r$.

In this second form, or directly, pseudodivision can be readily implemented in most computer algebra systems. Indeed, some systems include pseudodivision as one of the built-in operations on polynomials.

We recall the situation studied in §4, in which the hypotheses and conclusion of a theorem in Euclidean plane geometry are translated into a system of polynomials in variables $u_1, \ldots, u_m, x_1, \ldots, x_n$, with $h_1, \ldots, h_n$ representing the hypotheses and $g$ giving the conclusion. As in equation (11) of §4, we can group the irreducible components of the variety $V = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbb{R}^{m+n}$ as

$$V = V' \cup U,$$

where $V'$ is the union of the components on which the $u_i$ are algebraically independent. Our goal is to prove that $g$ vanishes on $V'$.

The elementary version of Wu's method that we will discuss is tailored for the case where $V'$ is *irreducible*. We note, however, that Wu's method can be extended to the more general reducible case also. The main algebraic tool needed (Ritt's decomposition algorithm based on *characteristic sets* for prime ideals) would lead us too far afield, though, so we will not discuss it. Note that, in practice, we usually do not know in advance whether $V'$ is irreducible or not. Thus, reliable "theorem-provers" based on Wu's method should include these more general techniques too.

Our simplified version of Wu's method uses the pseudodivision algorithm in *two* ways in the process of determining whether the equation $g = 0$ follows from $h_j = 0$.

- Step 1 of Wu's method uses pseudodivision to reduce the hypotheses to a system of polynomials $f_j$ that are in triangular form in the variables $x_1, \ldots, x_n$. That is, we seek

$$\begin{aligned} f_1 &= f_1(u_1, \ldots, u_m, x_1), \\ f_2 &= f_2(u_1, \ldots, u_m, x_1, x_2), \\ &\vdots \\ f_n &= f_n(u_1, \ldots, u_m, x_1, \ldots, x_n) \end{aligned} \tag{2}$$

  such that $\mathbf{V}(f_1, \ldots, f_n)$ again contains the irreducible variety $V'$, on which the $u_i$ are algebraically independent.

- Step 2 of Wu's method uses *successive* pseudodivision of the conclusion $g$ with respect to each of the variables $x_j$ to determine whether $g \in \mathbf{I}(V')$. We compute

$$
\begin{aligned}
R_{n-1} &= \mathrm{Rem}(g, f_n, x_n),\\
R_{n-2} &= \mathrm{Rem}(R_{n-1}, f_{n-1}, x_{n-1}),\\
&\vdots\\
R_1 &= \mathrm{Rem}(R_2, f_2, x_2),\\
R_0 &= \mathrm{Rem}(R_1, f_1, x_1).
\end{aligned}
\tag{3}
$$

- Then $R_0 = 0$ implies that $g$ follows from the hypotheses $h_j$ under an additional condition, to be made precise in Theorem 4.

To explain how Wu's method works, we need to explain each of these steps, beginning with the reduction to triangular form.

## *Step 1. Reduction to Triangular Form*

In practice, this reduction can almost always be accomplished using a procedure very similar to Gaussian elimination for systems of linear equations. We will not state any general theorems concerning our procedure, however, because there are some exceptional cases in which it might fail. (See the comments in 3 and 4 below.) A completely general procedure for accomplishing this kind of reduction may be found in Chou (1988).

The elementary version is performed as follows. We work one variable at a time, beginning with $x_n$.

1. Among the $h_j$, find all the polynomials containing the variable $x_n$. Call the set of such polynomials $S$. (If there are no such polynomials, the translation of our geometric theorem is most likely incorrect since it would allow $x_n$ to be arbitrary.)
2. If there is only one polynomial in $S$, then we can rename the polynomials, making that one polynomial $f_n'$, and our system of polynomials will have the form

$$
\begin{aligned}
f_1' &= f_1'(u_1, \ldots, u_m, x_1, \ldots, x_{n-1}),\\
&\vdots\\
f_{n-1}' &= f_{n-1}'(u_1, \ldots, u_m, x_1, \ldots, x_{n-1}),\\
f_n' &= f_n'(u_1, \ldots, u_m, x_1, \ldots, x_n).
\end{aligned}
\tag{4}
$$

3. If there is more than one polynomial in $S$, but some element of $S$ has degree 1 in $x_n$, then we can take $f_n'$ as that polynomial and replace all the other hypotheses in $S$ by their pseudoremainders on division by $f_n'$ with respect to $x_n$. [One of these pseudoremainders could conceivably be *zero*, but this would mean that $f_n'$ would divide $d^s h$, where $h$ is one of the other hypothesis polynomials and $d = LC(f_n', x_n)$. This is unlikely since $V'$ is assumed to be irreducible.] We obtain a system in the form (4) again. By part (ii) of Proposition 1, all the $f_j'$ are in the ideal generated by the $h_j$.

4. If there are several polynomials in $S$, but none has degree 1, then we repeat the steps:
   a. pick $a, b \in S$ where $0 < \deg(b, x_n) \le \deg(a, x_n)$;
   b. compute the pseudoremainder $r = \mathrm{Rem}(a, b, x_n)$;
   c. replace $S$ by $(S - \{a\}) \cup \{r\}$ (leaving the hypotheses not in $S$ unchanged),

   until eventually we are reduced to a system of polynomials of the form (4) again. Since the degree in $x_n$ are reduced each time we compute a pseudoremainder, we will eventually remove the $x_n$ terms from all but one of our polynomials. Moreover, by part (ii) of Proposition 1, each of the resulting polynomials is contained in the ideal generated by the $h_j$. Again, it is conceivable that we could obtain a zero pseudoremainder at some stage here. This would usually, but not always, imply reducibility, so it is unlikely. We then apply the same process to the polynomials $f'_1, \ldots, f'_{n-1}$ in (4) to remove the $x_{n-1}$ terms from all but one polynomial. Continuing in this way, we will eventually arrive at a system of equations in triangular form as in (2) above.

Once we have the triangular equations, we can relate them to the original hypotheses as follows.

**Proposition 2.** *Suppose that $f_1 = \cdots = f_n = 0$ are the triangular equations obtained from $h_1 = \cdots = h_n = 0$ by the above reduction algorithm. Then*

$$V' \subset V \subset \mathbf{V}(f_1, \ldots, f_n).$$

**Proof.** As we noted above, all the $f_j$ are contained in the ideal generated by the $h_j$. Thus, $\langle f_1, \ldots, f_n \rangle \subset \langle h_1, \ldots, h_n \rangle$ and hence, $V = \mathbf{V}(h_1, \ldots, h_n) \subset \mathbf{V}(f_1, \ldots, f_n)$ follows immediately. Since $V' \subset V$, we are done. □

**Example 3.** To illustrate the operation of this triangulation procedure, we will apply it to the hypotheses of the Circle Theorem of Apollonius from §4. Referring back to (5)–(7) of §4, we have

$$\begin{aligned}
h_1 &= 2x_1 - u_1,\\
h_2 &= 2x_2 - u_2,\\
h_3 &= 2x_3 - u_1,\\
h_4 &= 2x_4 - u_2,\\
h_5 &= u_2x_5 + u_1x_6 - u_1u_2,\\
h_6 &= u_1x_5 - u_2x_6,\\
h_7 &= x_1^2 - x_2^2 - 2x_1x_7 + 2x_2x_8,\\
h_8 &= x_1^2 - 2x_1x_7 - x_3^2 + 2x_3x_7 - x_4^2 + 2x_4x_8.
\end{aligned}$$

Note that this system is very nearly in triangular form in the $x_j$. In fact, this is often true, especially in the cases where each step of constructing the geometric configuration involves adding one new point.

At the first step of the triangulation procedure, we see that $h_7, h_8$ are the only polynomials in our set containing $x_8$. Even better, $h_8$ has degree 1 in $x_8$. Hence, we proceed as in step 3 of the triangulation procedure, making $f_8 = h_8$, and replacing $h_7$ by

$$\begin{aligned} f_7 &= \mathrm{Rem}(h_7, h_8, x_8) \\ &= (2x_1x_2 - 2x_2x_3 - 2x_1x_4)x_7 - x_1^2x_2 + x_2x_3^2 + x_1^2x_4 - x_2^2x_4 + x_2x_4^2. \end{aligned}$$

As this example indicates, we often ignore numerical constants when computing remainders. Only $f_7$ contains $x_7$, so nothing further needs to be done there. Both $h_6$ and $h_5$ contain $x_6$, but we are in the situation of step 3 in the procedure again. We make $f_6 = h_6$ and replace $h_5$ by

$$f_5 = \mathrm{Rem}(h_5, h_6, x_6) = (u_1^2 + u_2^2)x_5 - u_1u_2^2.$$

The remaining four polynomials are in triangular form already, so we take $f_i = h_i$ for $i = 1, 2, 3, 4$.

## *Step 2. Successive Pseudodivision*

The key step in Wu's method is the successive pseudodivsion operation given in equation (3) computing the *final remainder* $R_0$. The usefulness of this operation is indicated by the following theorem.

**Theorem 4.** *Consider the set of hypotheses and the conclusion for a geometric theorem. Let $R_0$ be the final remainder computed by the successive pseudodivision of g as in* (3)*, using the system of polynomials $f_1, \ldots, f_n$ in triangular form* (2)*. Let $d_j$ be the leading coefficient of $f_j$ as a polynomial in $x_j$ (so $d_j$ is a polynomial in $u_1, \ldots, u_m$ and $x_1, \ldots, x_{j-1}$). Then:*

(i) *There are nonnegative integers $s_1, \ldots, s_n$ and polynomials $A_1, \ldots, A_n$ in the ring $\mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ such that*

$$d_1^{s_1} \cdots d_n^{s_n} g = A_1 f_1 + \cdots + A_n f_n + R_0.$$

(ii) *If $R_0$ is the zero polynomial, then g is zero at every point of $V' - \mathbf{V}(d_1 d_2 \cdots d_n) \subset \mathbb{R}^{m+n}$.*

**Proof.** Part (i) follows by applying Proposition 1 repeatedly. Pseudodividing g by $f_n$ with respect to $x_n$, we have

$$R_{n-1} = d_n^{s_n} g - q_n f_n.$$

Hence, when we pseudodivide again with respect to $x_{n-1}$:

$$\begin{aligned} R_{n-2} &= d_{n-1}^{s_{n-1}}(d_n^{s_n} g - q_n f_n) - q_{n-1} f_{n-1} \\ &= d_{n-1}^{s_{n-1}} d_n^{s_n} g - q_{n-1} f_{n-1} - d_{n-1}^{s_{n-1}} q_n f_n. \end{aligned}$$

Continuing in the same way, we will eventually obtain an expression of the form

$$R_0 = d_1^{s_1} \cdots d_n^{s_n} g - (A_1 f_1 + \cdots + A_n f_n),$$

which is what we wanted to show.

(ii) By the result of part (i), if $R_0 = 0$, then at every point of the variety $W = \mathbf{V}(f_1, \ldots, f_n)$, either $g$ or one of the $d_j^{s_j}$ is zero. By Proposition 2, the variety $V'$ is contained in $W$, so the same is true on $V'$. The assertion follows. □

Even though they are not always polynomial relations in the $u_i$ alone, the equations $d_j = 0$, where $d_j$ is the leading coefficient of $f_j$, can often be interpreted as loci defining *degenerate special cases* of our geometric configuration.

**Example 3 (continued).** For instance, let us complete the application of Wu's method to the Circle Theorem of Apollonius. Our goal is to show that

$$g = (x_5 - x_7)^2 + (x_6 - x_8)^2 - (x_1 - x_7)^2 - x_8^2 = 0$$

is a consequence of the hypotheses $h_1 = \cdots = h_8 = 0$ (see (8) of §4). Using $f_1, \ldots, f_8$ computed above, we set $R_8 = g$ and compute the successive remainders

$$R_{i-1} = \operatorname{Rem}(R_i, f_i, x_i)$$

as $i$ decreases from 8 to 1. When computing these remainders, we always use the minimal exponent $s$ in Proposition 1, and in some cases, we ignore constant factors of the remainder. We obtain the following remainders.

$$\begin{aligned}
R_7 = {} & x_4x_5^2 - 2x_4x_5x_7 + x_4x_6^2 - x_4x_1^2 + 2x_4x_1x_7 + x_6x_1^2 - 2x_6x_1x_7 \\
& - x_6x_3^2 + 2x_6x_3x_7 - x_6x_4^2, \\
R_6 = {} & x_4^2x_1x_5^2 - x_4^2x_1^2x_5 - x_4x_1x_6x_3^2 + x_4^2x_1x_6^2 - x_4^3x_1x_6 + x_4^2x_2^2x_5 \\
& - x_4^2x_2^2x_1 - x_2x_4^3x_5 + x_2x_4^3x_1 - x_2x_1x_4x_5^2 - x_2x_1x_4x_6^2 \\
& + x_2x_3x_4x_5^2 + x_2x_3x_4x_6^2 - x_2x_3x_4x_1^2 + x_4x_1^2x_6x_3 + x_4x_2^2x_6x_1 \\
& - x_4x_2^2x_6x_3 + x_2x_1^2x_4x_5 - x_2x_3^2x_4x_5 + x_2x_3^2x_4x_1, \\
R_5 = {} & u_2^2x_4^2x_1x_5^2 - u_2^2x_4^2x_1^2x_5 + u_2^2x_4^2x_2^2x_5 - u_2^2x_4^2x_2^2x_1 - u_2^2x_2x_4^3x_5 \\
& + u_2^2x_2x_4^3x_1 - x_4u_2^2x_2x_1x_5^2 + x_4u_2^2x_2x_3x_5^2 - x_4u_2^2x_2x_3x_1^2 \\
& + x_4u_2^2x_2x_1^2x_5 - x_4u_2^2x_2x_3^2x_5 + x_4u_2^2x_2x_3^2x_1 - u_1x_5u_2x_4^3x_1 \\
& + x_4u_1x_5u_2x_2^2x_1 - x_4u_1x_5u_2x_1x_3^2 - x_4u_1x_5u_2x_2^2x_3 \\
& + x_4u_1x_5u_2x_1^2x_3 + u_1^2x_5^2x_4^2x_1 - x_4u_1^2x_5^2x_2x_1 + x_4u_1^2x_5^2x_2x_3, \\
R_4 = {} & -u_2^4x_4x_2x_3x_1^2 - u_2^4x_4^2x_2^2x_1 + u_2^4x_4x_2x_3^2x_1 + u_2^4x_4^3x_2x_1 \\
& - u_2^2x_4u_1^2x_2x_3x_1^2 - u_2^2x_4^2u_1^2x_2^2x_1 + u_2^2x_4u_1^2x_2x_3^2x_1 \\
& + u_2^2x_4^3u_1^2x_2x_1 - u_2^4x_4^3u_1x_2 - u_2^3x_4^3u_1^2x_1 + u_2^4x_4^2u_1x_2^2 \\
& - u_2^4x_4^2u_1x_1^2 + u_2^3x_4u_1^2x_2^2x_1 - u_2^3x_4u_1^2x_1x_3^2 - u_2^4x_4u_1x_2x_3^2 \\
& + u_2^4x_4u_1x_2x_1^2 - u_2^3x_4u_1^2x_2^2x_3 + u_2^3x_4u_1^2x_1^2x_3 + u_2^4x_4^2u_1^2x_1 \\
& - u_2^4x_4u_1^2x_2x_1 + u_2^4x_4u_1^2x_2x_3,
\end{aligned}$$

$$
\begin{aligned}
R_3 = {} & 4u_2^5x_2x_3^2x_1 - 4u_2^5u_1x_2x_3^2 + 4u_2^5u_1x_2x_1^2 - 4u_2^5x_2x_3x_1^2 \\
& - 3u_2^5u_1^2x_2x_1 + 4u_2^5u_1^2x_2x_3 - 4u_2^4u_1^2x_1x_3^2 - 4u_2^4u_1^2x_2^2x_3 \\
& + 2u_2^4u_1^2x_2^2x_1 + 4u_2^4u_1^2x_1^2x_3 - 4u_2^3u_1^2x_2x_3x_1^2 + 4u_2^3u_1^2x_2x_3^2x_1 \\
& - 2u_2^6x_2^2x_1 - 2u_2^6u_1x_1^2 + 2u_2^6u_1x_2^2 + u_2^6u_1^2x_1 + u_2^7x_2x_1 \\
& - u_2^7u_1x_2, \\
R_2 = {} & 2u_2^5u_1x_2x_1^2 - 2u_2^5u_1^2x_2x_1 + 2u_2^4u_1^2x_2^2x_1 - 2u_2^6x_2^2x_1 \\
& - 2u_2^6u_1x_1^2 + 2u_2^6u_1x_2^2 + u_2^6u_1^2x_1 + u_2^7x_2x_1 - u_2^7u_1x_2 \\
& + u_2^5u_1^3x_2 - 2u_2^4u_1^3x_2^2 + 2u_2^4u_1^3x_1^2 - 2u_2^3u_1^3x_2x_1^2 + u_2^3u_1^4x_2x_1 \\
& - u_2^4u_1^4x_1, \\
R_1 = {} & -2u_2^6u_1x_1^2 - u_2^4u_1^4x_1 + u_2^6u_1^2x_1 + 2u_2^4u_1^3x_1^2, \\
R_0 = {} & 0.
\end{aligned}
$$

By Theorem 4, Wu's method confirms that the Circle Theorem is valid when none of the leading coefficients of the $f_j$ is zero. The nontrivial conditions here are

$$
\begin{aligned}
d_5 &= u_1^2 + u_2^2 \neq 0, \\
d_6 &= u_2 \neq 0, \\
d_7 &= 2x_1x_2 - 2x_2x_3 - 2x_1x_4 \neq 0, \\
d_8 &= 2x_4 \neq 0.
\end{aligned}
$$

The second condition in this list is $u_2 \neq 0$, which says that the vertices $A$ and $C$ of the right triangle $\Delta ABC$ are distinct [recall we chose coordinates so that $A = (0, 0)$ and $C = (0, u_2)$ in Example 3 of §4]. This also implies the first condition since $u_1$ and $u_2$ are real. The condition $2x_4 \neq 0$ is equivalent to $u_2 \neq 0$ by the hypothesis $h_4 = 0$. Finally, $d_7 \neq 0$ says that the vertices of the triangle are distinct (see Exercise 5). From this analysis, we see that the Circle Theorem actually follows generically from its hypotheses as in §4.

The elementary version of Wu's method only gives $g = 0$ under the side conditions $d_j \neq 0$. In particular, note that in a case where $V'$ is reducible, it is entirely conceivable that one of the $d_j$ could vanish on an entire component of $V'$. If this happened, there would be no conclusion concerning the validity of the theorem for geometric configurations corresponding to points in that component.

Indeed, a much stronger version of Theorem 4 is known when the subvariety $V'$ for a given set of hypotheses is *irreducible*. With the extra algebraic tools we have omitted (Ritt's decomposition algorithm), it can be proved that there are special triangular form sets of $f_j$ (called characteristic sets) with the property that $R_0 = 0$ is a necessary and sufficient condition for $g$ to lie in $\mathbf{I}(V')$. In particular, it is never the case that one of the leading coefficients of the $f_j$ is identically zero on $V'$ so that $R_0 = 0$ implies that $g$ must vanish on all of $V'$. We refer the interested reader to CHOU (1988) for the details. Other treatments of characteristic sets and the Wu–Ritt algorithm can be found in MISHRA (1993) and WANG (1994b). There is also a Maple package called "charsets" which implements the method of characteristic sets [see WANG (1994a)].

Finally, we will briefly compare Wu's method with the method based on Groebner bases introduced in §4. These two methods apply to exactly the same class of geometric theorems and they usually yield equivalent results. Both make essential use of a division algorithm to determine whether a polynomial is in a given ideal or not. However, as we can guess from the triangulation procedure described above, the basic version of Wu's method at least is likely to be much quicker on a given problem. The reason is that simply triangulating a set of polynomials usually requires much less effort than computing a Groebner basis for the ideal they generate, or for the ideal $\widetilde{H} = \langle h_1, \ldots, h_n, 1 - yg\rangle$. This pattern is especially pronounced when the original polynomials themselves are nearly in triangular form, which is often the case for the hypotheses of a geometric theorem. In a sense, this superiority of Wu's method is only natural since Groebner bases contain much more information than triangular form sets. Note that we have not claimed anywhere that the triangular form set of polynomials even generates the same ideal as the hypotheses in either $\mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ or $\mathbb{R}(u_1, \ldots, u_m)[x_1, \ldots, x_n]$. In fact, this is not true in general (Exercise 4). Wu's method is an example of a technique tailored to solve a particular problem. Such techniques can often outperform general techniques (such as computing Groebner bases) that do many other things besides.

For the reader interested in pursuing this topic further, we recommend CHOU (1988), the second half of which is an annotated collection of 512 geometric theorems proved by Chou's program implementing Wu's method. Wu (1983) is a reprint of the original paper that introduced these ideas.

**EXERCISES FOR §5**

1. This problem completes the proof of Proposition 1 begun in the text.
   a. Complete the proof of (i) of the proposition.
   b. Show that $q, r$ in the equation $d_m^s f = qg + r$ in the proposition are definitely *not* unique if no condition is placed on the exponent $s$.
2. Establish the claim stated after Proposition 1 that pseudodivision is equivalent to ordinary polynomial division in the ring $K[y]$, where $K = k(x_1, \ldots, x_n)$.
3. Show that there is a unique minimal $s \le p - m + 1$ in Proposition 1 for which the equation $d_m^s f = qg + r$ exists, and that $q$ and $r$ are unique when $s$ is minimal. Hint: Use the uniqueness of the quotient and remainder for division in $k(x_1, \ldots, x_n)[y]$.
4. Show by example that applying the triangulation procedure described in this section to two polynomials $h_1, h_2 \in k[x_1, x_2]$ can yield polynomials $f_1, f_2$ that generate an ideal strictly smaller than $\langle h_1, h_2\rangle$. The same can be true for larger sets of polynomials as well.
5. Show that the nondegeneracy condition $d_7 \neq 0$ for the Circle Theorem is automatically satisfied if $u_1$ and $u_2$ are nonzero.
6. Use Wu's method to verify each of the following theorems. In each case, state the conditions $d_j \neq 0$ under which Theorem 4 implies that the conclusion follows from the hypotheses. If you also did the corresponding Exercises in §4, try to compare the time and/or effort involved with each method.
   a. The theorem on the diagonals of a parallelogram (Example 1 of §4).
   b. The theorem on the orthocenter of a triangle (Exercise 5 of §4).

c. The theorem on the centroid of a triangle (Exercise 6 of §4).
d. The theorem on the Euler line of a triangle (Exercise 7 of §4).
e. Pappus's Theorem (Exercise 8 of §4).

7. Consider the theorem from Exercise 17 of §4 (for which $V'$ is reducible according to a direct translation of the hypotheses).
   a. Apply Wu's method to this problem. (Your final remainder should be nonzero here.)
   b. Does Wu's method succeed for the reformulation from part e of Exercise 17 from §4?