Chapter 8

Polyhedral Regions and Polynomials

In this chapter we will consider a series of interrelated topics concerning polyhedral regions P in $\mathbb{R}^n$ and polynomials. In the first three sections we will see how some of the algebraic methods from earlier chapters give conceptual and computational tools for several classes of problems of intrinsic interest and practical importance. We will begin by considering Gröbner basis methods for integer optimization and combinatorial enumeration problems. We will also use module Gröbner bases to study piecewise polynomial, or *spline*, functions on polyhedral complexes.

The final two sections apply the same polyhedral geometry to furnish some further insight into the foundations of Gröbner basis theory. We will study the *Gröbner fan* of an ideal, a collection of polyhedral cones classifying the ideal's different reduced Gröbner bases, and use the Gröbner fan to develop a general basis conversion algorithm called the *Gröbner Walk*. The walk is applicable even when the ideal is not zero-dimensional, hence is more general than the FGLM algorithm from Chapter 2, §3.

Many of the topics in this chapter are also closely related to the material on polytopes and toric varieties from Chapter 7, but we have tried to make this chapter as independent as possible from Chapter 7 so that it can be read separately.

§1 Integer Programming

This section applies the theory of Gröbner bases to problems in integer programming. Most of the results depend only on the basic algebra of polynomial rings and facts about Gröbner bases for ideals. From Proposition (1.12) on, we will also need to use the language of Laurent polynomials, but the idea should be reasonably clear even if that concept is not familiar. The original reference for this topic is an article by Conti and Traverso, [CT], and another treatment may be found in [AL], Section 2.8. Further developments may be found in the articles [Tho1], [Tho2], [HT], and the book [Stu2]. For a general introduction to linear and integer programming, we recommend [Schri].

To begin, we will consider a very small, but in other ways typical, applied integer programming problem, and we will use this example to illustrate the key features of this class of problems. Suppose that a small local trucking firm has two customers, A and B, that generate shipments to the same location. Each shipment from A is a pallet weighing exactly 400 kilos and taking up 2 cubic meters of volume. Each pallet from B weighs 500 kilos and takes up 3 cubic meters. The shipping firm uses small trucks that can carry any load up to 3700 kilos, and up to 20 cubic meters. B's product is more perishable, though, and they are willing to pay a higher price for on-time delivery: \$ 15 per pallet versus \$ 11 per pallet from A. The question facing the manager of the trucking company is: How many pallets from each of the two companies should be included in each truckload to maximize the revenues generated?

Using A to represent the number of pallets from company A, and similarly B to represent the number of pallets from company B in a truckload, we want to maximize the revenue function $11A + 15B$ subject to the following constraints:

$$\begin{aligned} &4A + 5B \leq 37 && \text{(the weight limit, in 100's)} \\ &2A + 3B \leq 20 && \text{(the volume limit)} \\ &A, B \in \mathbb{Z}_{\geq 0}. && \end{aligned} \tag{1.1}$$

Note that both A, B must be integers. This is, as we will see, an important restriction, and the characteristic feature of *integer programming* problems.

Integer programming problems are generalizations of the mathematical translation of the question above. Namely, in an integer programming problem we seek the *maximum or minimum value* of some *linear* function

$$\ell(A_1, \ldots, A_n) = c_1A_1 + c_2A_2 + \cdots + c_nA_n$$

on the set of $(A_1, \ldots, A_n) \in \mathbb{Z}^n_{\geq 0}$ with $A_j \geq 0$ for all $1 \leq j \leq n$ satisfying a set of linear inequalities:

$$\begin{aligned} a_{11}A_1 + a_{12}A_2 + \cdots + a_{1n}A_n &\leq \ \text{(or } \geq) \ b_1 \\ a_{21}A_1 + a_{22}A_2 + \cdots + a_{2n}A_n &\leq \ \text{(or } \geq) \ b_2 \\ &\vdots \\ a_{m1}A_1 + a_{m2}A_2 + \cdots + a_{mn}A_n &\leq \ \text{(or } \geq) \ b_m. \end{aligned}$$

We assume in addition that the a_{ij}, and the b_i are all integers. Some of the coefficients c_j, a_{ij}, b_i may be negative, but we will always assume $A_j \geq 0$ for all j.

Integer programming problems occur in many contexts in engineering, computer science, operations research, and pure mathematics. With large numbers of variables and constraints, they can be *difficult* to solve. It is perhaps instructive to consider our small shipping problem (1.1) in detail. In geometric terms we are seeking a maximum for the function $11A + 15B$ on the integer points in the closed convex polygon P in $\mathbb{R}^2$ bounded above by portions of the lines $4A + 5B = 37$ (slope $-4/5$), $2A + 3B = 20$ (slope $-2/3$), and by the coordinate axes $A = 0$, and $B = 0$. See Fig. 8.1. The set of all points in $\mathbb{R}^2$ satisfying the inequalities from (1.1) is known as the feasible region.

(1.2) Definition. The *feasible region* of an integer programming problem is the set P of all $(A_1, \ldots, A_n) \in \mathbb{R}^n$ satisfying the inequalities in the statement of the problem.

The set of all points in $\mathbb{R}^n$ satisfying a single linear inequality of the form considered here is called a *closed half-space*. A *polyhedral region* or *polyhedron* in $\mathbb{R}^n$ is defined as the intersection of a finite number of closed half-spaces. Equation (1.4) in Chapter 7 shows that polytopes are bounded polyhedral regions. In fact, a polyhedral region is a polytope if and only if it is bounded in $\mathbb{R}^n$ (the other implication is shown for instance in [Ewa], Theorem 1.5). In this chapter we will consider both bounded and unbounded polyhedral regions.

It is possible for the feasible region of an integer programming problem to contain *no* integer points at all. There are no solutions of the optimization

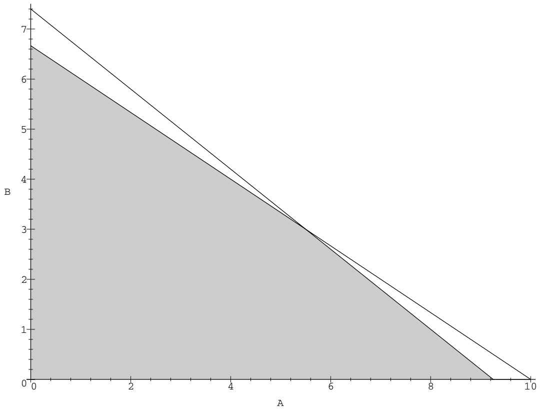

FIGURE 8.1. The feasible region P for (1.1)

problem in that case. For instance in $\mathbb{R}^2$ consider the region defined by

$$\begin{aligned} A + B &\leq 1 \\ 3A - B &\geq 1 \\ 2A - B &\leq 1, \end{aligned} \tag{1.3}$$

and $A, B \geq 0$.

Exercise 1. Verify directly (for example with a picture) that there are no integer points in the region defined by (1.3).

When n is small, it is often possible to analyze the feasible set of an integer programming problem geometrically and determine the integer points in it. However, even this can be complicated since any polyhedral region formed by intersecting half-spaces bounded by affine hyperplanes with equations defined over $\mathbb{Z}$ can occur. For example, consider the set P in $\mathbb{R}^3$ defined by inequalities:

$$\begin{array}{llll} 2A_1 + 2A_2 + 2A_3 & \leq 5 & -2A_1 + 2A_2 + 2A_3 & \leq 5 \\ 2A_1 + 2A_2 - 2A_3 & \leq 5 & -2A_1 + 2A_2 - 2A_3 & \leq 5 \\ 2A_1 - 2A_2 + 2A_3 & \leq 5 & -2A_1 - 2A_2 + 2A_3 & \leq 5 \\ 2A_1 - 2A_2 - 2A_3 & \leq 5 & -2A_1 - 2A_2 - 2A_3 & \leq 5. \end{array}$$

In Exercise 11, you will show that P is a solid regular octahedron, with 8 triangular faces, 12 edges, and 6 vertices.

Returning to the problem from (1.1), if we did not have the additional constraints $A, B \in \mathbb{Z}$, (if we were trying to solve a *linear programming* problem rather than an *integer programming* problem), the situation would be somewhat easier to analyze. For instance, to solve (1.1), we could apply the following simple geometric reasoning. The *level curves* of the revenue function $\ell(A, B) = 11A + 15B$ are lines of slope $-11/15$. The values of ℓ increase as we move out into the first quadrant. Since the slopes satisfy $-4/5 < -11/15 < -2/3$, it is clear that the revenue function attains its overall maximum on P at the vertex q in the interior of the first quadrant. Readers of Chapter 7 will recognize q as the face of P in the support line with normal vector $\nu = (-11, -15)$. See Fig. 8.2.

That point has rational, but not *integer* coordinates: $q = (11/2, 3)$. Hence q is *not* the solution of the integer programming problem! Instead, we need to consider only the integer points (A, B) in P. One *ad hoc* method that works here is to fix A, compute the largest B such that (A, B) lies in P, then compute the revenue function at those points and compare values for all possible A values. For instance, with $A = 4$, the largest B giving a point in P is $B = 4$, and we obtain $\ell(4, 4) = 104$. Similarly, with $A = 8$, the largest feasible B is $B = 1$, and we obtain $\ell(8, 1) = 103$. Note incidentally that *both* of these values are larger than the value of ℓ at the integer point closest to q in P—$(A, B) = (5, 3)$, where $\ell(5, 3) = 100$. This shows some

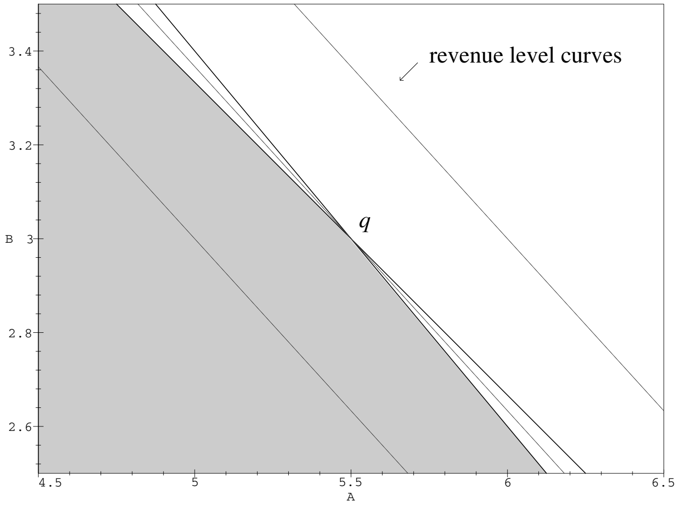

Figure 8.2. The linear programming maximum for (1.1)

of the potential subtlety of integer programming problems. Continuing in this way it can be shown that the maximum of ℓ occurs at $(A, B) = (4, 4)$.

Exercise 2. Verify directly (that is, by enumerating integer points as suggested above) that the solution of the shipping problem (1.1) is the point $(A, B) = (4, 4)$.

This sort of approach would be quite impractical for larger problems. Indeed, the general integer programming problem is known to be *NP-complete*, and so as Conti and Traverso remark, "even algorithms with theoretically bad worst case and average complexity can be useful ... , hence deserve investigation."

To discuss integer programming problems in general it will be helpful to standardize their statement to some extent. This can be done using the following observations.

1. We need only consider the problem of *minimizing* the linear function $\ell(A_1, \ldots, A_n) = c_1A_1 + c_2A_2 + \cdots + c_nA_n$, since maximizing a function ℓ on a set of integer n-tuples is the same as minimizing the function $-\ell$.

2. Similarly, by replacing an inequality
$$a_{i1}A_1 + a_{i2}A_2 + \cdots + a_{in}A_n \geq b_i$$
by the equivalent form
$$-a_{i1}A_1 - a_{i2}A_2 - \cdots - a_{in}A_n \leq -b_i,$$
we may consider only inequalities involving $\leq$.
3. Finally, by introducing additional variables, we can rewrite the linear constraint inequalities as *equalities*. The new variables are called "slack variables."

For example, using the idea in point 3 here the inequality
$$3A_1 - A_2 + 2A_3 \leq 9$$
can be replaced by
$$3A_1 - A_2 + 2A_3 + A_4 = 9$$
if $A_4 = 9 - (3A_1 - A_2 + 2A_3) \geq 0$ is introduced as a new variable to "take up the slack" in the original inequality. Slack variables will appear with coefficient zero in the function to be minimized.

Applying 1, 2, and 3 above, any integer programming problem can be put into the *standard form*:

$$\begin{aligned} &\text{Minimize: } c_1A_1 + \cdots + c_nA_n, \text{ subject to:}\\ &a_{11}A_1 + a_{12}A_2 + \cdots + a_{1n}A_n = b_1\\ &a_{21}A_1 + a_{22}A_2 + \cdots + a_{2n}A_n = b_2\\ &\qquad\qquad\vdots\\ &a_{m1}A_1 + a_{m2}A_2 + \cdots + a_{mn}A_n = b_m\\ &A_j \in \mathbb{Z}_{\geq 0},\ j = 1, \ldots n, \end{aligned} \tag{1.4}$$

where now n is the total number of variables (including slack variables). As before, we will call the set of all *real* n-tuples satisfying the constraint equations the *feasible region*. Note that this is a polyhedral region in $\mathbb{R}^n$ because the set of all $(A_1, \ldots, A_n) \in \mathbb{R}^n$ satisfying a linear equation $a_{j1}A_1 + \cdots + a_{jn}A_n = b_j$ is the intersection of the two half-spaces defined by $a_{j1}A_1 + \cdots + a_{jn}A_n \geq b_j$ and $a_{j1}A_1 + \cdots + a_{jn}A_n \leq b_j$.

For the rest of this section we will explore an alternative approach to integer programming problems, in which we translate such a problem into a question about polynomials. We will use the standard form (1.4) and first consider the case where all the coefficients are nonnegative: $a_{ij} \geq 0$, $b_i \geq 0$. The translation proceeds as follows. We introduce an indeterminate z_i for each of the equations in (1.4), and exponentiate to obtain an equality
$$z_i^{a_{i1}A_1 + a_{i2}A_2 + \cdots + a_{in}A_n} = z_i^{b_i}$$

for each $i = 1, \ldots, m$. Multiplying the left and right hand sides of these equations, and rearranging the exponents, we get another equality:

$$\prod_{j=1}^{n} (\prod_{i=1}^{m} z_i^{a_{ij}})^{A_j} = \prod_{i=1}^{m} z_i^{b_i}. \tag{1.5}$$

From (1.5) we get the following direct algebraic characterization of the integer n-tuples in the feasible region of the problem (1.4).

(1.6) Proposition. *Let k be a field, and define $\varphi : k[w_1, \ldots, w_n] \to k[z_1, \ldots, z_m]$ by setting*

$$\varphi(w_j) = \prod_{i=1}^{m} z_i^{a_{ij}}$$

for each $j = 1, \ldots, n$, and $\varphi(g(w_1, \ldots, w_n)) = g(\varphi(w_1), \ldots, \varphi(w_n))$ for a general polynomial $g \in k[w_1, \ldots, w_n]$. Then $(A_1, \ldots, A_n)$ is an integer point in the feasible region if and only if φ maps the monomial $w_1^{A_1} w_2^{A_2} \cdots w_n^{A_n}$ to the monomial $z_1^{b_1} \cdots z_m^{b_m}$.

Exercise 3. Prove Proposition (1.6).

For example, consider the standard form of our shipping problem (1.1), with slack variables C in the first equation and D in the second.

$$\begin{aligned} \varphi : k[w_1, w_2, w_3, w_4] &\to k[z_1, z_2] \\ w_1 &\mapsto z_1^4 z_2^2 \\ w_2 &\mapsto z_1^5 z_2^3 \\ w_3 &\mapsto z_1 \\ w_4 &\mapsto z_2. \end{aligned} \tag{1.7}$$

The integer points in the feasible region of this restatement of the problem are the (A, B, C, D) such that

$$\varphi(w_1^A w_2^B w_3^C w_4^D) = z_1^{37} z_2^{20}.$$

Exercise 4. Show that in this case *every* monomial in $k[z_1, \ldots, z_m]$ is the image of some monomial in $k[w_1, \ldots, w_n]$.

In other cases, φ may not be surjective, and the following test for membership in the image of a mapping is an important part of the translation of integer programming problems.

Since the image of φ in Proposition (1.6) is precisely the set of polynomials in $k[z_1, \ldots, z_m]$ that can be expressed as polynomials in the $f_j = \prod_{i=1}^{m} z_i^{a_{ij}}$, we can also write the image as $k[f_1, \ldots, f_n]$, the subring of $k[z_1, \ldots, z_m]$ generated by the f_j. The subring membership test given by parts a and b of the following Proposition is also used in studying rings of invariants for finite matrix groups (see [CLO], Chapter 7, §3).

(1.8) Proposition. *Suppose that $f_1, \ldots, f_n \in k[z_1, \ldots, z_m]$ are given. Fix a monomial order in $k[z_1, \ldots, z_m, w_1, \ldots, w_n]$ with the elimination property: any monomial containing one of the z_i is greater than any monomial containing only the w_j. Let $\mathcal{G}$ be a Gröbner basis for the ideal*

$$I = \langle f_1 - w_1, \ldots, f_n - w_n \rangle \subset k[z_1, \ldots, z_m, w_1, \ldots, w_n]$$

and for each $f \in k[z_1, \ldots, z_m]$, let $\overline{f}^{\mathcal{G}}$ be the remainder on division of f by $\mathcal{G}$. Then

a. *A polynomial f satisfies $f \in k[f_1, \ldots, f_n]$ if and only if $g = \overline{f}^{\mathcal{G}} \in k[w_1, \ldots, w_n]$.*
b. *If $f \in k[f_1, \ldots, f_n]$ and $g = \overline{f}^{\mathcal{G}} \in k[w_1, \ldots, w_n]$ as in part a, then $f = g(f_1, \ldots, f_n)$, giving an expression for f as a polynomial in the f_j.*
c. *If each f_j and f are monomials and $f \in k[f_1, \ldots, f_n]$, then g is also a monomial.*

In other words, part c says that in the situation of Proposition (1.6), if $z_1^{b_1} \cdots z_m^{b_m}$ is in the image of φ, then it is automatically the image of some *monomial* $w_1^{A_1} \cdots w_n^{A_n}$.

Proof. Parts a and b are proved in Proposition 7 of Chapter 7, §3 in [CLO], so we will not repeat them here.

To prove c, we note that each generator of I is a difference of two monomials. It follows that in the application of Buchberger's algorithm to compute $\mathcal{G}$, each S-polynomial considered and each nonzero S-polynomial remainder that goes into the Gröbner basis will be a difference of two monomials. This is true since in computing the S-polynomial, we are subtracting one difference of two monomials from another, and the leading terms cancel. Similarly, in the remainder calculation, at each step we subtract one difference of two monomials from another and cancellation occurs. It follows that every element of $\mathcal{G}$ will also be a difference of two monomials. When we divide a monomial by a Gröbner basis of this form, the remainder must be *a monomial*, since at each step we subtract a difference of two monomials from a single monomial and a cancellation occurs. Hence, if we are in the situation of parts a and b and the remainder is $g(w_1, \ldots, w_n) \in k[w_1, \ldots, w_n]$, then g must be a monomial. □

In the restatement of our example problem in (1.7), we would consider the ideal

$$I = \langle z_1^4 z_2^2 - w_1, z_1^5 z_2^3 - w_2, z_1 - w_3, z_2 - w_4 \rangle.$$

Using the *lex* order with the variables ordered

$$z_1 > z_2 > w_4 > w_3 > w_2 > w_1$$

(chosen to eliminate terms involving slack variables if possible), we obtain a Gröbner basis $\mathcal{G}$:

$$\begin{aligned}
g_1 &= z_1 - w_3,\\
g_2 &= z_2 - w_4,\\
g_3 &= w_4^2 w_3^4 - w_1\\
g_4 &= w_4 w_3^3 w_2 - w_1^2 \qquad (1.9)\\
g_5 &= w_4 w_3 w_1 - w_2\\
g_6 &= w_4 w_1^4 - w_3 w_2^3\\
g_7 &= w_3^2 w_2^2 - w_1^3.
\end{aligned}$$

(Note: An efficient implementation of Buchberger's algorithm is necessary for working out relatively large explicit examples using this approach, because of the large number of variables involved. We used `Singular` and *Macaulay 2* to compute the examples in this chapter.) So for instance, using g_1 and g_2 the monomial $f = z_1^{37} z_2^{20}$ reduces to $w_3^{37} w_4^{20}$. Hence f is in the image of φ from (1.7). But then further reductions are also possible, and the remainder on division is

$$\overline{f}^{\mathcal{G}} = w_2^4 w_1^4 w_3.$$

This monomial corresponds to the solution of the integer programming problem ($A = 4, B = 4$, and slack $C = 1$) that you verified in Exercise 2. In a sense, this is an accident, since the *lex* order that we used for the Gröbner basis and remainder computations did not take the revenue function ℓ explicitly into account.

To find the solution of an integer programming problem minimizing a given linear function $\ell(A_1, \ldots, A_n)$ we will usually need to use a monomial order specifically tailored to the problem at hand.

(1.10) Definition. A monomial order on $k[z_1, \ldots, z_m, w_1, \ldots, w_n]$ is said to be *adapted* to an integer programming problem (1.4) if it has the following two properties:

a. (Elimination) Any monomial containing one of the z_i is greater than any monomial containing only the w_j.
b. (Compatibility with ℓ) Let $A = (A_1, \ldots, A_n)$ and $A' = (A'_1, \ldots, A'_n)$. If the monomials $w^A, w^{A'}$ satisfy $\varphi(w^A) = \varphi(w^{A'})$ and $\ell(A_1, \ldots, A_n) > \ell(A'_1, \ldots, A'_n)$, then $w^A > w^{A'}$.

(1.11) Theorem. *Consider an integer programming problem in standard form (1.4). Assume all $a_{ij}, b_i \geq 0$ and let $f_j = \prod_{i=1}^m z_i^{a_{ij}}$ as before. Let $\mathcal{G}$*

be a Gröbner basis for

$$I = \langle f_1 - w_1, \ldots, f_n - w_n \rangle \subset k[z_1, \ldots, z_m, w_1, \ldots, w_n]$$

with respect to any adapted monomial order. Then if $f = z_1^{b_1} \cdots z_m^{b_m}$ *is in* $k[f_1, \ldots, f_n]$*, the remainder* $\overline{f}^{\mathcal{G}} \in k[w_1, \ldots, w_n]$ *will give a solution of (1.4) minimizing* ℓ*. (There are cases where the minimum is not unique and, if so, this method will only find one minimum.)*

Proof. Let $\mathcal{G}$ be a Gröbner basis for I with respect to an adapted monomial order. Suppose that $w^A = \overline{f}^{\mathcal{G}}$ so $\varphi(w^A) = f$, but that $A = (A_1, \ldots, A_n)$ is not a minimum of ℓ. That is, assume that there is some $A' = (A_1', \ldots, A_n') \neq A$ such that $\varphi(w^{A'}) = f$ and $\ell(A_1', \ldots, A_n') < \ell(A_1, \ldots, A_n)$. Consider the difference $h = w^A - w^{A'}$. We have $\varphi(h) = f - f = 0$. In Exercise 5 below, you will show that this implies $h \in I$. But then h must reduce to *zero* under the Gröbner basis $\mathcal{G}$ for I. However, because $>$ is an adapted order, the leading term of h must be w^A, and that monomial is reduced with respect to $\mathcal{G}$ since it is a remainder. This contradiction shows that A must give a minimum of ℓ. □

Exercise 5. Let $f_i \in k[z_1, \ldots, z_m]$, $i = 1, \ldots, n$, as above and define a mapping

$$\begin{aligned} \varphi : k[w_1, \ldots, w_n] &\rightarrow k[z_1, \ldots, z_m] \\ w_i &\mapsto f_i \end{aligned}$$

as in (1.6). Let $I = \langle f_1 - w_1, \ldots, f_n - w_n \rangle \subset k[z_1, \ldots, z_m, w_1, \ldots, w_n]$. Show that if $h \in k[w_1, \ldots, w_n]$ satisfies $\varphi(h) = 0$, then $h \in I \cap k[w_1, \ldots, w_n]$. Hint: See the proof of Proposition 3 from Chapter 7, §4 of [CLO].

Exercise 6. Why did the *lex* order used to compute the Gröbner basis in (1.9) correctly find the maximum value of $11A + 15B$ in our example problem (1.1)? Explain, using Theorem (1.11). (Recall, w_4 and w_3 corresponding to the slack variables were taken greater than w_2, w_1.)

Theorem (1.11) yields a Gröbner basis algorithm for solving integer programming problems with all $a_{ij}, b_i \geq 0$:

```
Input: $A, b$ from (1.4), an adapted monomial order $>$
Output: a solution of (1.4), if one exists


    $f_j := \prod_{i=1}^{m} z_i^{a_{ij}}$
    $I := \langle f_1 - w_1, \ldots, f_n - w_n \rangle$
```

```
$\mathcal{G} :=$ Gröbner basis of $I$ with respect to $>$
$f := \prod_{i=1}^{m} z_i^{b_i}$
$g := \overline{f}^{\mathcal{G}}$
IF $g \in k[w_1, \ldots, w_n]$ THEN
      its exponent vector gives a solution
ELSE
      there is no solution
```

Monomial orders satisfying both the elimination and compatibility properties from (1.10) can be specified in the following ways.

First, assume that all $c_j \geq 0$. Then it is possible to define a *weight order* $>_\ell$ on the w-variables using the linear function ℓ (see [CLO], Chapter 2, §4, Exercise 12). Namely order monomials in the w-variables alone first by ℓ-values:

$$w_1^{A_1} \cdots w_n^{A_n} >_\ell w_1^{A_1'} \cdots w_n^{A_n'}$$

if $\ell(A_1, \ldots, A_n) > \ell(A_1', \ldots, A_n')$ and break ties using any other fixed monomial order on $k[w_1, \ldots, w_n]$. Then incorporate this order into a product order on $k[z_1, \ldots, z_m, w_1, \ldots, w_n]$ with the z-variables greater than all the w-variables, to ensure that the elimination property from (1.10) holds.

If some $c_j < 0$, then the recipe above produces a total ordering on monomials in $k[z_1, \ldots, z_m, w_1, \ldots, w_n]$ that is compatible with multiplication and that satisfies the elimination property. But it will not be a well-ordering. So in order to apply the theory of Gröbner bases with respect to monomial orders, we will need to be more clever in this case. We begin with the following observation.

In $k[z_1, \ldots, z_m, w_1, \ldots, w_n]$, define a (non-standard) degree for each variable by setting $\deg(z_i) = 1$ for all $i = 1, \ldots, m$, and $\deg(w_j) = d_j = \sum_{i=1}^m a_{ij}$ for all $j = 1, \ldots, n$. Each d_j must be strictly positive, since otherwise the constraint equations would not depend on A_j. We say a polynomial $f \in k[z_1, \ldots, z_m, w_1, \ldots, w_n]$ is *homogeneous* with respect to these degrees if all the monomials $z^\alpha w^\beta$ appearing in f have the same (non-standard) total degree $|\alpha| + \sum_j d_j \beta_j$.

(1.12) Lemma. *With respect to the degrees d_j on w_j, the following statements hold.*

a. *The ideal $I = \langle f_1 - w_1, \ldots, f_n - w_n \rangle$ is homogeneous.*

b. *Every reduced Gröbner basis for the ideal I consists of homogeneous polynomials.*

PROOF. Part a follows since the given generators are homogeneous for these degrees—since $f_j = \prod_{i=1}^m z_i^{a_{ij}}$, the two terms in $f_j - w_j$ have the same degree.

Part b follows in the same way as for ideals that are homogeneous in the usual sense. The proof of Theorem 2 of Chapter 8, §3 of [CLO] goes over to non-standard assignments of degrees as well. □

For instance, in the *lex* Gröbner basis given in (1.9) above, it is easy to check that all the polynomials are homogeneous with respect to the degrees $\deg(z_i) = 1$, $\deg(w_1) = 6$, $\deg(w_2) = 8$, and $\deg(w_3) = \deg(w_4) = 1$.

Since $d_j > 0$ for all j, given the c_j from ℓ and $\mu > 0$ sufficiently large, all the entries of the vector

$$(c_1, \ldots, c_n) + \mu(d_1, \ldots, d_n)$$

will be positive. Let μ be any fixed number for which this is true. Consider the $(m+n)$-component weight vectors u_1, u_2:

$$\begin{aligned} u_1 &= (1, \ldots, 1, 0, \ldots, 0) \\ u_2 &= (0, \ldots, 0, c_1, \ldots, c_n) + \mu(0, \ldots, 0, d_1, \ldots, d_n). \end{aligned}$$

Then all entries of u_2 are nonnegative, and hence we can define a weight order $>_{u_1,u_2,\sigma}$ by comparing u_1-weights first, then comparing u_2-weights if the u_1-weights are equal, and finally breaking ties with any other monomial order $>_\sigma$.

Exercise 7. Consider an integer programming problem (1.4) in which $a_{ij}, b_i \geq 0$ for all i, j.

a. Show that the order $>_{u_1,u_2,\sigma}$ defined above satisfies the elimination condition from Definition (1.10).

b. Show that if $\varphi(w^A) = \varphi(w^{A'})$, then $w^A - w^{A'}$ is homogeneous with respect to the degrees $d_j = \deg(w_j)$.

c. Deduce that $>_{u_1,u_2,\sigma}$ is an adapted order.

For example, our shipping problem (in standard form) can be solved using the second method here. We take $u_1 = (1, 1, 0, 0, 0, 0)$, and letting $\mu = 2$, we see that

$$u_2 = (0, 0, -11, -15, 0, 0) + 2(0, 0, 6, 8, 1, 1) = (0, 0, 1, 1, 2, 2)$$

has all nonnegative entries. Finally, break ties with $>_\sigma$ = graded reverse lex on all the variables ordered $z_1 > z_2 > w_1 > w_2 > w_3 > w_4$. Here is a `Singular` session performing the Gröbner basis and remainder calculations. Note the definition of the monomial order $>_{u_1,u_2,\sigma}$ by means of weight vectors.

```
> ring R = 0,(z(1..2),w(1..4)),(a(1,1,0,0,0,0),
  a(0,0,1,1,2,2),dp);
```

```
> ideal I = z(1)^4*z(2)^2-w(1), z(1)^5*z(2)^3-w(2), z(1)-w(3),
  z(2)-w(4);
> ideal J = std(I);
> J;
J[1]=w(1)*w(3)*w(4)-1*w(2)
J[2]=w(2)^2*w(3)^2-1*w(1)^3
J[3]=w(1)^4*w(4)-1*w(2)^3*w(3)
J[4]=w(2)*w(3)^3*w(4)-1*w(1)^2
J[5]=w(3)^4*w(4)^2-1*w(1)
J[6]=z(2)-1*w(4)
J[7]=z(1)-1*w(3)
> poly f = z(1)^37*z(2)^20;
> reduce(f,J);
w(1)^4*w(2)^4*w(3)
```

We find

$$\overline{z_1^{37}z_2^{20}}^{\mathcal{G}} = w_1^4 w_2^4 w_3$$

as expected, giving the solution $A = 4, B = 4$, and $C = 1, D = 0$.

This computation could also be done using the Maple `Groebner` package or *Mathematica*, since the weight order $>_{u_1,u_1,grevlex}$ can be defined as one of the matrix orders $>_M$ explained in Chapter 1, §2. For example, we could use the 6×6 matrix with first row u_1, second row u_2, and next four rows coming from the matrix defining the *grevlex* order on $k[z_1, z_1, w_1, w_2, w_3, w_4]$ following the patterns from parts b and e of Exercise 6 in Chapter 1, §2:

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The other rows from the matrix defining the *grevlex* order are discarded because of linear dependences with previous rows in M.

Finally, we want to discuss general integer programming problems where some of the a_{ij} and b_i may be *negative*. There is no real conceptual difference in that case; the geometric interpretation of the integer programming problem is exactly the same, only the positions of the affine linear spaces bounding the feasible region change. But there is a difference in the algebraic translation. Namely, we cannot view the negative a_{ij} and b_i directly as exponents—that is not legal in an ordinary polynomial. One way to fix this problem is to consider what are called *Laurent polynomials* in the variables z_i instead—polynomial expressions in the z_i *and* z_i^{-1}, as defined in Chapter 7, §1 of this text. To deal with these more general objects without introducing a whole new set of m variables, we will use the *second* representation of the ring of Laurent polynomials, as presented in Exercise 15 of Chapter 7, §1:

$$k[z_1^{\pm 1},\ldots,z_m^{\pm 1}] \cong k[z_1,\ldots,z_m,t]/\langle tz_1\cdots z_m - 1\rangle.$$

In intuitive terms, this isomorphism works by introducing a single new variable t satisfying $tz_1\cdots z_m - 1 = 0$, so that formally t is the product of the inverses of the z_i: $t = z_1^{-1}\cdots z_m^{-1}$. Then each of the $\prod_{i=1}^m z_i^{a_{ij}}$ involved in the algebraic translation of the integer programming problem can be rewritten in the form $t^{e_j}\prod_{i=1}^m z_i^{a'_{ij}}$, where now all $a'_{ij} \geq 0$—we can just take $e_j \geq 0$ to be the negative of the smallest (most negative) a_{ij} that appears, and $a'_{ij} = e_j + a_{ij}$ for each i. Similarly, $\prod_{i=1}^m z_i^{b_i}$ can be rewritten in the form $t^e\prod_{i=1}^m z_i^{b'_i}$ with $e \geq 0$, and $b_i \geq 0$ for all i. It follows that the equation (1.5) becomes an equation between polynomial expressions in $t, z_1,\ldots,z_n$:

$$\prod_{j=1}^n \big(t^{e_j}\prod_{i=1}^m z_i^{a'_{ij}}\big)^{A_j} = t^e\prod_{i=1}^m z_i^{b'_i},$$

modulo the relation $tz_1\cdots z_m - 1 = 0$. We have a direct analogue of Proposition (1.6).

(1.13) Proposition. *Define a mapping*

$$\varphi : k[w_1,\ldots,w_n] \to k[z_1^{\pm 1},\ldots,z_m^{\pm 1}]$$

by setting

$$\varphi(w_j) = t^{e_j}\prod_{i=1}^m z_i^{a'_{ij}} \bmod \langle tz_1\cdots z_m - 1\rangle$$

for each $j = 1,\ldots,n$, and extending to general $g(w_1,\ldots,w_n) \in k[w_1,\ldots,w_n]$ as before. Then $(A_1,\ldots,A_n)$ is an integer point in the feasible region if and only if $\varphi(w_1^{A_1}w_2^{A_2}\cdots w_n^{A_n})$ and $t^e z_1^{b'_1}\cdots z_m^{b'_m}$ represent the same element in $k[z_1^{\pm 1},\ldots,z_m^{\pm 1}]$ (that is, their difference is divisible by $tz_1\cdots z_m - 1$).

Similarly, Proposition (1.8) goes over to this more general situation. We will write S for the image of φ in $k[z_1^{\pm 1},\ldots,z_m^{\pm 1}]$. Then we have the following version of the subring membership test.

(1.14) Proposition. *Suppose that $f_1,\ldots,f_n \in k[z_1,\ldots,z_m,t]$ are given. Fix a monomial order in $k[z_1,\ldots,z_m,t,w_1,\ldots,w_n]$ with the elimination property: any monomial containing one of the z_i or t is greater than any monomial containing only the w_j. Finally, let $\mathcal{G}$ be a Gröbner basis for the ideal*

$$J = \langle tz_1\cdots z_m - 1, f_1 - w_1,\ldots,f_n - w_n\rangle$$

in $k[z_1, \ldots, z_m, t, w_1, \ldots, w_n]$ *and for each* $f \in k[z_1, \ldots, z_m, t]$, *let* $\overline{f}^{\mathcal{G}}$ *be the remainder on division of* f *by* $\mathcal{G}$. *Then*

a. f *represents an element in* S *if and only if* $g = \overline{f}^{\mathcal{G}} \in k[w_1, \ldots, w_n]$.

b. *If* f *represents an element in* S *and* $g = \overline{f}^{\mathcal{G}} \in k[w_1, \ldots, w_n]$ *as in part a, then* $f = g(f_1, \ldots, f_n)$, *giving an expression for* f *as a polynomial in the* f_j.

c. *If each* f_j *and* f *are monomials and* f *represents an element in* S, *then* g *is also a monomial.*

The proof is essentially the same as the proof for Proposition (1.8) so we omit it.

We should also mention that there is a direct parallel of Theorem (1.11) saying that using monomial orders which have the elimination and compatibility properties will yield minimum solutions for integer programming problems and give an algorithm for their solution. For ℓ with only nonnegative coefficients, adapted orders may be constructed using product orders as above, making t and the z_i greater than any w_j. For a more general discussion of constructing monomial orders compatible with a given ℓ, we refer the reader to [CT].

We will conclude this section with an example illustrating the general case described in the previous paragraph. Consider the following problem in standard form:

$$\begin{aligned}
&\text{Minimize:}\\
&\qquad A + 1000B + C + 100D,\\
&\text{Subject to the constraints:}\\
&\qquad 3A - 2B + C = -1\\
&\qquad 4A + B - C - D = 5\\
&\qquad A, B, C, D \in \mathbb{Z}_{\geq 0}.
\end{aligned} \tag{1.15}$$

With the relation $tz_1z_2 - 1 = 0$, our ideal J in this case is

$$J = \langle tz_1z_2 - 1, z_1^3z_2^4 - w_1, t^2z_2^3 - w_2, tz_1^2 - w_3, tz_1 - w_4\rangle.$$

If we use an elimination order placing t, z_1, z_2 before the w-variables, and then the use a weight order compatible with ℓ on the w_j (breaking ties with graded reverse lex), then we obtain a Gröbner basis $\mathcal{G}$ for J consisting of the following polynomials:

$$\begin{aligned}
g_1 &= w_2w_3^2 - w_4\\
g_2 &= w_1w_4^7 - w_3^3\\
g_3 &= w_1w_2w_4^6 - w_3\\
g_4 &= w_1w_2^2w_3w_4^5 - 1
\end{aligned}$$

$$\begin{aligned}
g_5 &= z_2 - w_1w_2^2w_3w_4^4\\
g_6 &= z_1 - w_1w_2w_4^5\\
g_7 &= t - w_2w_3w_4.
\end{aligned}$$

From the right-hand sides of the equations, we consider $f = tz_2^6$. A remainder computation yields

$$\overline{f}^{\mathcal{G}} = w_1w_2^2w_4.$$

Since this is still a very small problem, it is easy to check by hand that the corresponding solution $(A = 1, B = 2, C = 0, D = 1)$ really does minimize $\ell(A, B, C, D) = A + 1000B + C + 100D$ subject to the constraints.

Exercise 8. Verify directly that the solution $(A, B, C, D) = (1, 2, 0, 1)$ of the integer programming problem (1.15) is correct. Hint: Show first that $B \geq 2$ in any solution of the constraint equations.

We should also remark that because of the special *binomial* form of the generators of the ideals in (1.11) and (1.13) and the simple polynomial remainder calculations involved here, there are a number of optimizations one could make in special-purpose Gröbner basis integer programming software. See [CT] for some preliminary results and [BLR] for additional developments. Algorithms described in the latter paper have been implemented in the `intprog` package distributed with the current version of CoCoA. The current version of `Singular` also contains an `intprog` library with procedures for integer programming.

Additional Exercises for §1

Exercise 9. What happens if you apply the Gröbner basis algorithm to any optimization problem on the polyhedral region in (1.3)?

Note: For the computational portions of the following problems, you will need to have access to a Gröbner basis package that allows you to specify mixed elimination-weight monomial orders as in the discussion following Theorem (1.11). One way to specify these orders is via suitable weight matrices as explained in Chapter 1, §2. See the example following Exercise 7 above.

Exercise 10. Apply the methods of the text to solve the following integer programming problems:

a.

$$\text{Minimize: } 2A + 3B + C + 5D, \text{ subject to:}$$

$$\begin{aligned} 3A + 2B + C + D &= 10 \\ 4A + B + C &= 5 \\ A, B, C, D &\in \mathbb{Z}_{\geq 0}. \end{aligned}$$

Verify that your solution is correct.

b. Same as a, but with the right-hand sides of the constraint equations changed to $20, 14$ respectively. How much of the computation needs to be redone?

c.

$$\begin{aligned} \text{Maximize: } 3A + 4B + 2C&, \text{ subject to:} \\ 3A + 2B + C &\leq 45 \\ A + 2B + 3C &\leq 21 \\ 2A + B + C &\leq 18 \\ A, B, C &\in \mathbb{Z}_{\geq 0}. \end{aligned}$$

Also, describe the feasible region for this problem geometrically, and use that information to verify your solution.

Exercise 11. Consider the set P in $\mathbb{R}^3$ defined by inequalities:

$$\begin{array}{llll} 2A_1 + 2A_2 + 2A_3 & \leq 5 & -2A_1 + 2A_2 + 2A_3 & \leq 5 \\ 2A_1 + 2A_2 - 2A_3 & \leq 5 & -2A_1 + 2A_2 - 2A_3 & \leq 5 \\ 2A_1 - 2A_2 + 2A_3 & \leq 5 & -2A_1 - 2A_2 + 2A_3 & \leq 5 \\ 2A_1 - 2A_2 - 2A_3 & \leq 5 & -2A_1 - 2A_2 - 2A_3 & \leq 5. \end{array}$$

Verify that P is a solid (regular) octahedron. (What are the vertices?)

Exercise 12.

a. Suppose we want to consider *all* the integer points in a polyhedral region $P \subset \mathbb{R}^n$ as feasible, not just those with non-negative coordinates. How could the methods developed in the text be adapted to this more general situation?

b. Apply your method from part a to find the minimum of $2A_1 - A_2 + A_3$ on the integer points in the solid octahedron from Exercise 11.

§2 Integer Programming and Combinatorics

In this section we will study a beautiful application of commutative algebra and the ideas developed in §1 to combinatorial enumeration problems. For those interested in exploring this rich subject farther, we recommend the marvelous book [Sta1] by Stanley. Our main example is discussed there and far-reaching generalizations are developed using more advanced algebraic tools. There are also connections between the techniques we will develop here, *invariant theory* (see especially [Stu1]), the theory of *toric varieties*

([Ful]), and the *geometry of polyhedra* (see [Stu2]). The prerequisites for this section are the theory of Gröbner bases for polynomial ideals, familiarity with quotient rings, and basic facts about Hilbert functions (see, e.g. Chapter 6, §4 of this book or Chapter 9, §3 of [CLO]).

Most of this section will be devoted to the consideration of the following classical counting problem. Recall that a *magic square* is an $n \times n$ integer matrix $M = (m_{ij})$ with the property that the sum of the entries in each row and each column is the same. A famous 4×4 magic square appears in the well-known engraving *Melancholia* by Albrecht Dürer:

$$\begin{array}{cccc}
16 & 3 & 2 & 13 \\
5 & 10 & 11 & 8 \\
9 & 6 & 7 & 12 \\
4 & 15 & 14 & 1
\end{array}$$

The row and column sums in this array all equal 34. Although the extra condition that the m_{ij} are the distinct integers $1, 2, \ldots, n^2$ (as in Dürer's magic square) is often included, we will *not* make that part of the definition. Also, many familiar examples of magic squares have diagonal sums equal to the row and column sum and other interesting properties; we will not require that either. Our problem is this:

(2.1) Problem. *Given positive integers s, n, how many different $n \times n$ magic squares with $m_{ij} \geq 0$ for all i, j and row and column sum s are there?*

There are related questions from statistics and the design of experiments of practical as well as purely mathematical interest. In some small cases, the answer to (2.1) is easily derived.

Exercise 1. Show that the number of 2×2 nonnegative integer magic squares with row and column sum s is precisely $s + 1$, for each $s \geq 0$. How are the squares with sum $s > 1$ related to those with $s = 1$?

Exercise 2. Show that there are exactly six 3×3 magic squares with nonnegative integer entries and $s = 1$, twenty-one with $s = 2$, and fifty-five with $s = 3$. How many are there in each case if we require that the two diagonal sums also equal s?

Our main goal in this section will be to develop a *general* way to attack this and similar counting problems where the objects to be counted can be identified with the integer points in a polyhedral region in $\mathbb{R}^N$ for some N, so that we are in the same setting as in the integer programming problems from §1. We will take a somewhat *ad hoc* approach though, and use only as much general machinery as we need to answer our question (2.1) for small values of n.

To see how (2.1) fits into this context, note that the entire set of $n \times n$ nonnegative integer magic squares M is the set of solutions in $\mathbb{Z}_{\geq 0}^{n \times n}$ of a system of linear equations with integer coefficients. For instance, in the 3×3 case, the conditions that all row and column sums are equal can be expressed as 5 independent equations on the entries of the matrix. Writing

$$\vec{m} = (m_{11}, m_{12}, m_{13}, m_{21}, m_{22}, m_{23}, m_{31}, m_{32}, m_{33})^T,$$

the matrix $M = (m_{ij})$ is a magic square if and only if

$$A_3 \vec{m} = 0, \tag{2.2}$$

where A_3 is the 5×9 integer matrix

$$A_3 = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 \end{pmatrix} \tag{2.3}$$

and $m_{ij} \geq 0$ for all i, j. Similarly, the $n \times n$ magic squares can be viewed as the solutions of a similar system $A_n \vec{m} = 0$ for an integer matrix A_n with n^2 columns.

Exercise 3.

a. Show that the 3×3 nonnegative integer magic squares are exactly the solutions of the system of linear equations (2.2) with matrix A_3 given in (2.3).

b. What is the minimal number of linear equations needed to define the corresponding space of $n \times n$ magic squares? Describe an explicit way to produce a matrix A_n as above.

As in the discussion following (1.4) of this chapter, the set $\{\vec{m} : A_3\vec{m} = 0\}$ is a polyhedral region in $\mathbb{R}^{3\times 3}$. However, there are three important differences between our situation here and the optimization problems considered in §1. First, there is no linear function to be optimized. Instead, we are mainly interested in understanding the structure of the entire set of integer points in a polyhedral region. Second, unlike the regions considered in the examples in §1, the region in this case is *unbounded*, and there are infinitely many integer points. Finally, we have a *homogeneous* system of equations rather than an inhomogeneous system, so the points of interest are elements of the kernel of the matrix A_n. In the following, we will write

$$K_n = \ker(A_n) \cap \mathbb{Z}_{\geq 0}^{n \times n}$$

for the set of all nonnegative integer $n \times n$ magic squares. We begin with a few simple observations.

(2.4) Proposition. *For each n,*

a. *K_n is closed under vector sums in $\mathbb{Z}^{n\times n}$, and contains the zero vector.*

b. *The set $\mathcal{C}_n$ of solutions of $A_n\vec{m} = 0$ satisfying $\vec{m} \in \mathbb{R}_{\geq 0}^{n\times n}$ forms a convex polyhedral cone in $\mathbb{R}^{n\times n}$, with vertex at the origin.*

Proof. Part a follows by linearity. For part b, recall that a *convex polyhedral cone with vertex at the origin* is the intersection of finitely many half-spaces containing the origin. Then $\mathcal{C}_n$ is polyhedral since the defining equations are the linear equations $A_n\vec{m} = 0$ and the linear inequalities $m_{ij} \geq 0 \in \mathbb{R}$. It is a cone since any positive real multiple of a point in $\mathcal{C}_n$ is also in $\mathcal{C}_n$. Finally, it is convex since if $\vec{m}$ and $\vec{m}'$ are two points in $\mathcal{C}_n$, any linear combination $x = r\vec{m} + (1-r)\vec{m}'$ with $r \in [0,1]$ also satisfies the equations $A_n x = 0$ and has nonnegative entries, hence lies in $\mathcal{C}_n$. □

A set M with a binary operation is said to be a *monoid* if the operation is associative and possesses an identity element in M. For example $\mathbb{Z}_{\geq 0}^{n\times n}$ is a monoid under vector addition. In this language, part a of the proposition says that K_n is a *submonoid* of $\mathbb{Z}_{\geq 0}^{n\times n}$.

To understand the structure of the submonoid K_n, we will seek to find a minimal set of additive generators to serve as building blocks for all the elements of K_n. The appropriate notion is given by the following definition.

(2.5) Definition. Let K be any submonoid of the additive monoid $\mathbb{Z}_{\geq 0}^{N}$. A finite subset $\mathcal{H} \subset K$ is said to be a *Hilbert basis* for K if it satisfies the following two conditions.

a. For every $k \in K$ there exist $h_i \in \mathcal{H}$ and nonnegative integers c_i such that $k = \sum_{i=1}^{q} c_i h_i$, and

b. $\mathcal{H}$ is minimal with respect to inclusion.

It is a general fact that Hilbert bases exist and are unique for all submonoids $K \subset \mathbb{Z}_{\geq 0}^{N}$. Instead of giving an existence proof, however, we will present a Gröbner basis *algorithm* for finding the Hilbert basis for the submonoid $K = \ker(A)$ in $\mathbb{Z}_{\geq 0}^{N}$ for any integer matrix with N columns. (This comes from [Stu1], §1.4.) As in §1, we translate our problem from the context of integer points to Laurent polynomials. Given an integer matrix $A = (a_{ij})$ with N columns and m rows say, we introduce an indeterminate z_i for each row, $i = 1, \ldots, m$, and consider the ring of Laurent polynomials:

$$k[z_1^{\pm 1}, \ldots, z_m^{\pm 1}] \cong k[z_1, \ldots, z_m, t]/\langle t z_1 \cdots z_m - 1\rangle.$$

(See §1 of this chapter and Exercise 15 of Chapter 7, §1.) Define a mapping

$$(2.6)\qquad \psi : k[v_1, \ldots, v_N, w_1, \ldots, w_N] \to k[z_1^{\pm 1}, \ldots, z_m^{\pm 1}][w_1, \ldots, w_N]$$

as follows. First take

$$(2.7)\qquad \psi(v_j) = w_j \cdot \prod_{i=1}^{m} z_i^{a_{ij}}$$

and $\psi(w_j) = w_j$ for each $j = 1, \ldots, N$, then extend to polynomials in $k[v_1, \ldots, v_N, w_1, \ldots, w_N]$ so as to make ψ a ring homomorphism.

The purpose of ψ is to detect elements of the kernel of A.

(2.8) Proposition. *A vector $\alpha^T \in \ker(A)$ if and only if $\psi(v^\alpha - w^\alpha) = 0$, that is if and only if $v^\alpha - w^\alpha$ is in the kernel of the homomorphism ψ.*

Exercise 4. Prove Proposition (2.8).

As in Exercise 5 of §1, we can write $J = \ker(\psi)$ as

$$J = I \cap k[v_1, \ldots, v_N, w_1, \ldots, w_N],$$

where

$$I = \langle w_j \cdot \prod_{i=1}^{m} z_i^{a_{ij}} - v_j : j = 1, \ldots N\rangle$$

in the ring $k[z_1^{\pm 1}, \ldots, z_m^{\pm 1}][v_1, \ldots, v_N, w_1, \ldots, w_N]$. The following theorem of Sturmfels (Algorithm 1.4.5 of [Stu1]) gives a way to find Hilbert bases.

(2.9) Theorem. *Let $\mathcal{G}$ be a Gröbner basis for I with respect to any elimination order $>$ for which all $z_i, t > v_j$, and all $v_j > w_k$. Let S be the subset of $\mathcal{G}$ consisting of elements of the form $v^\alpha - w^\alpha$ for some $\alpha \in \mathbb{Z}^N_{\geq 0}$. Then*

$$\mathcal{H} = \{\alpha : v^\alpha - w^\alpha \in S\}$$

is the Hilbert basis for K.

Proof. The idea of this proof is similar to that of Theorem (1.11) of this chapter. See [Stu1] for a complete exposition. □

Here is a first example to illustrate Theorem (2.9). Consider the submonoid of $\mathbb{Z}^4_{\geq 0}$ given as $K = \ker(A) \cap \mathbb{Z}^4_{\geq 0}$, for

$$A = \begin{pmatrix} 1 & 2 & -1 & 0 \\ 1 & 1 & -1 & -2 \end{pmatrix}.$$

To find a Hilbert basis for K, we consider the ideal I generated by

$$w_1 z_1 z_2 - v_1,\ w_2 z_1^2 z_2 - v_2,\ w_3 t - v_3,\ w_4 z_1^2 t^2 - v_4$$

and $z_1 z_2 t - 1$. Computing a Gröbner basis $\mathcal{G}$ with respect to an elimination order as in (2.9), we find only one is of the desired form:

$$v_1 v_3 - w_1 w_3$$

It follows that the Hilbert basis for K consists of a single element: $\mathcal{H} = \{(1, 0, 1, 0)\}$. It is not difficult to verify from the form of the matrix A that every element in K is an integer multiple of this vector. Note that the size of the Hilbert basis is not the same as the dimension of the kernel of the matrix A as a linear mapping on $\mathbb{R}^4$. In general, there is no connection between the size of the Hilbert basis for $K = \ker(A) \cap \mathbb{Z}^N_{\geq 0}$ and $\dim \ker(A)$; the number of elements in the Hilbert basis can be either larger than, equal to, or smaller than the dimension of the kernel, depending on A.

We will now use Theorem (2.9) to continue our work on the magic square enumeration problem. If we apply the method of the theorem to find the Hilbert basis for $\ker(A_3) \cap \mathbb{Z}^{3\times 3}_{\geq 0}$ (see equation (2.3) above) then we need to compute a Gröbner basis for the ideal I generated by

$$\begin{array}{ll} v_1 - w_1z_1z_2z_4z_5 & v_2 - w_2z_1^2z_2^2z_3^2z_5t \\ v_3 - w_3z_1^2z_2^2z_3^2z_4t & v_4 - w_4z_2z_4^2z_5^2t \\ v_5 - w_5z_2z_3z_5t & v_6 - w_6z_2z_3z_4t \\ v_7 - w_7z_1z_4^2z_5^2t & v_8 - w_8z_1z_3z_5t \\ \multicolumn{2}{c}{v_9 - w_9z_1z_3z_4t} \end{array}$$

and $z_1 \cdots z_5t - 1$ in the ring

$$k[z_1, \ldots, z_5, t, v_1, \ldots, v_9, w_1, \ldots, w_9].$$

Using an elimination order as described in Theorem (2.9) with the computer algebra system *Macaulay 2*, one obtains a very large Gröbner basis. (Because of the simple binomial form of the generators, however, the computation goes extremely quickly.) However, if we identify the subset S as in the theorem, there are only six polynomials corresponding to the Hilbert basis elements:

$$(2.10)\qquad \begin{array}{ll} v_3v_5v_7 - w_3w_5w_7 & v_3v_4v_8 - w_3w_4w_8 \\ v_2v_6v_7 - w_2w_6w_7 & v_2v_4v_9 - w_2w_4w_9 \\ v_1v_6v_8 - w_1w_6w_8 & v_1v_5v_9 - w_1w_5w_9. \end{array}$$

Expressing the corresponding 6-element Hilbert basis in matrix form, we see something quite interesting. The matrices we obtain are precisely the six 3×3 *permutation matrices*—the matrix representations of the permutations of the components of vectors in $\mathbb{R}^3$. (This should also agree with your results in the first part of Exercise 2.) For instance, the Hilbert basis element $(0, 0, 1, 0, 1, 0, 1, 0, 0)$ from the first polynomial in (2.10) corresponds to the matrix

$$T_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

which interchanges x_1, x_3, leaving x_2 fixed. Similarly, the other elements of the Gröbner basis give (in the order listed above)

$$S = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad S^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$T_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; T_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \; I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here S and S^2 are the cyclic permutations, T_{ij} interchanges x_i and x_j, and I is the identity.

Indeed, it is a well-known combinatorial theorem that the $n \times n$ permutation matrices form the Hilbert basis for the monoid K_n for all $n \geq 2$. See Exercise 9 below for a general proof.

This gives us some extremely valuable information to work with. By the definition of a Hilbert basis we have, for instance, that in the 3×3 case every element M of K_3 can be written as a linear combination

$$M = aI + bS + cS^2 + dT_{12} + eT_{13} + fT_{23},$$

where a, b, c, d, e, f are nonnegative integers. This is what we meant before by saying that we were looking for "building blocks" for the elements of our additive monoid of magic squares. The row and column sum of M is then given by

$$s = a + b + c + d + e + f.$$

It might appear at first glance that our problem is solved for 3×3 matrices. Namely for a given sum value s, it might seem that we just need to count the ways to write s as a sum of at most 6 nonnegative integers a, b, c, d, e, f. However, there is an added wrinkle here that makes the problem even more interesting: The 6 permutation matrices are not linearly independent. In fact, there is an obvious relation

$$(2.11) \qquad I + S + S^2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = T_{12} + T_{13} + T_{23}.$$

This means that for all $s \geq 3$ there are different combinations of coefficients that produce the same matrix sum. How can we take this (and other possible relations) into account and eliminate multiple counting?

First, we claim that in fact every equality

$$(2.12) \qquad \begin{aligned} &aI + bS + cS^2 + dT_{12} + eT_{13} + fT_{23} \\ &= a'I + b'S + c'S^2 + d'T_{12} + e'T_{13} + f'T_{23}, \end{aligned}$$

where $a, \ldots, f, a', \ldots, f'$ are nonnegative integers, is a consequence of the relation in (2.11), in the sense that if (2.12) is true, then the difference vector

$$(a, b, c, d, e, f) - (a', b', c', d', e', f')$$

is an integer multiple of the vector of coefficients $(1, 1, 1, -1, -1, -1)$ in the linear dependence relation

$$I + S + S^2 - T_{12} - T_{13} - T_{23} = 0,$$

which follows from (2.11).

This can be verified directly as follows.

Exercise 5.

a. Show that the six 3×3 permutation matrices span a 5-dimensional subspace of the vector space of 3×3 real matrices over $\mathbb{R}$.

b. Using part a, show that in every relation (2.12) with $a, \ldots, f' \in \mathbb{Z}_{\geq 0}$, $(a, b, c, d, e, f) - (a', b', c', d', e', f')$ is an integer multiple of the vector $(1, 1, 1, -1, -1, -1)$.

Given this, we can solve our problem in the 3×3 case by "retranslating" it into algebra. Namely we can identify the 6-tuples of coefficients $(a, b, c, d, e, f) \in \mathbb{Z}^6_{\geq 0}$ with monomials in 6 new indeterminates denoted $x_1, \ldots, x_6$:

$$\alpha = (a, b, c, d, e, f) \leftrightarrow x_1^a x_2^b x_3^c x_4^d x_5^e x_6^f.$$

By (2.11), though, we see that we want to think of $x_1x_2x_3$ and $x_4x_5x_6$ as being the same. This observation indicates that, in counting, we want to consider the element of the quotient ring

$$R = k[x_1, \ldots, x_6]/\langle x_1x_2x_3 - x_4x_5x_6\rangle$$

represented by the monomial x^α. Let $MS_3(s)$ be the number of distinct 3×3 integer magic squares with nonnegative entries, and row and column sum equal to s. Our next goal is to show that $MS_3(s)$ can be reinterpreted as the Hilbert function of the above ring R.

We recall from §4 of Chapter 6 that a homogeneous ideal $I \subset k[x_1, \ldots, x_n]$ gives a quotient ring $R = k[x_1, \ldots, x_n]/I$, and the Hilbert function $H_R(s)$ is defined by

$$(2.13) \quad H_R(s) = \dim_k k[x_1, \ldots, x_n]_s/I_s = \dim_k k[x_1, \ldots, x_n]_s - \dim_k I_s,$$

where $k[x_1, \ldots, x_n]_s$ is the vector space of homogeneous polynomials of total degree s, and I_s is the vector space of homogeneous polynomials of total degree s in I. In the notation of Chapter 9, §3 of [CLO], the Hilbert function of $R = k[x_1, \ldots, x_n]/I$ is written $HF_I(s)$. Since our focus here is on the ideal I, in what follows, we will call both $H_R(s)$ and $HF_I(s)$ the Hilbert function of I. It is a basic result that the Hilbert functions of I and $\langle \mathrm{LT}(I)\rangle$ (for any monomial order) are equal. Hence we can compute the Hilbert function by counting the number of standard monomials with respect to I for each total degree s—that is, monomials of total degree s in the complement of $\langle \text{LT}(I)\rangle$. For this and other information about Hilbert functions, the reader should consult [CLO], Chapter 9, §3 or Chapter 6, §4 of this book.

(2.14) Proposition. *The function $MS_3(s)$ equals the Hilbert function $H_R(s) = HF_I(s)$ of the homogeneous ideal $I = \langle x_1x_2x_3 - x_4x_5x_6\rangle$.*

Proof. The single element set $\{x_1x_2x_3 - x_4x_5x_6\}$ is a Gröbner basis for the ideal it generates with respect to any monomial order. Fix any order such that the leading term of the generator is $x_1x_2x_3$. Then the standard monomials of total degree s in $k[x_1, \ldots, x_6]$ are the monomials of total degree s that are not divisible by $x_1x_2x_3$.

Given any monomial $x^\alpha = x_1^a x_2^b x_3^c x_4^d x_5^e x_6^f$, let $A = \min(a, b, c)$, and construct

$$\alpha' = (a - A, b - A, c - A, d + A, e + A, f + A).$$

Since $x^{\alpha'}$ is not divisible by $x_1x_2x_3$, it is a standard monomial, and you will show in Exercise 6 below that it is the remainder on division of x^α by $x_1x_2x_3 - x_4x_5x_6$.

We need to show that the 3×3 magic squares with row and column sum s are in one-to-one correspondence with the standard monomials of degree s. Let M be a magic square, and consider any expression

$$M = aI + bS + cS^2 + dT_{12} + eT_{13} + fT_{23} \tag{2.15}$$

with $\alpha = (a, \ldots, f) \in \mathbb{Z}^6_{\geq 0}$. We associate to M the standard form in R of the monomial x^α, namely $x^{\alpha'}$ as above. In Exercise 7 you will show that this gives a well-defined mapping from the set of magic squares to the collection of standard monomials with respect to I, since by Exercise 5 any two expressions (2.15) for M yield the same standard monomial $x^{\alpha'}$. Moreover the row and column sum of M is the same as the total degree of the image monomial.

This mapping is clearly onto, since the exponent vector α' of any standard monomial can be used to give the coefficients in an expression (2.15). It is also one-to-one, since if M in (2.15) and

$$M_1 = a_1I + b_1S + c_1S^2 + d_1T_{12} + e_1T_{13} + f_1T_{23}$$

map to the same standard monomial α', then writing $A = \min(a, b, c)$, $A_1 = \min(a_1, b_1, c_1)$, we have

$$\begin{aligned}&(a - A, b - A, c - A, d + A, e + A, f + A)\\ &\quad = (a_1 - A_1, b_1 - A_1, c_1 - A_1, d_1 + A_1, e_1 + A_1, f_1 + A_1).\end{aligned}$$

It follows that $(a, \ldots, f)$ and $(a_1, \ldots, f_1)$ differ by the vector

$$(A - A_1)(1, 1, 1, -1, -1, -1).$$

Hence by (2.11), the magic squares M and M_1 are equal. □

For readers of Chapter 7, we would like to mention that there is also a much more conceptual way to understand the relationship between the monoid K_3 from our original problem and the ring R and the corresponding variety $\mathbf{V}(x_1x_2x_3 - x_4x_5x_6)$, using the theory of *toric varieties*. In particular, if $\mathcal{A} = \{\vec{m}_1, \ldots, \vec{m}_6\} \subset \mathbb{Z}^9$ is the set of integer vectors corresponding to the 3×3 permutation matrices as above (the Hilbert basis for K_3), and we define $\phi_{\mathcal{A}} : (\mathbb{C}^*)^9 \to \mathbb{P}^5$ by

$$\phi_{\mathcal{A}}(t) = (t^{\vec{m}_1}, \ldots, t^{\vec{m}_6})$$

as in §3 of Chapter 7, then it follows that the toric variety $X_{\mathcal{A}}$ (the Zariski closure of the image of $\phi_{\mathcal{A}}$) is the projective variety $\mathbf{V}(x_1x_2x_3 - x_4x_5x_6)$. The ideal $I_{\mathcal{A}} = \langle x_1x_2x_3 - x_4x_5x_6\rangle$ is called the *toric ideal* corresponding to $\mathcal{A}$. The defining homogeneous ideal of a toric variety is always generated by differences of monomials, as in this example. See the book [Stu2] for more details.

To conclude, Proposition (2.14) solves the 3×3 magic square counting problem as follows. By the proposition and (2.13), to find $MS_3(s)$, we simply subtract the number of nonstandard monomials of total degree s in 6 variables from the total number of monomials of total degree s in 6 variables. The nonstandard monomials are those divisible by $x_1x_2x_3$; removing that factor, we obtain an arbitrary monomial of total degree $s - 3$. Hence one expression is the following:

$$(2.16)\qquad \begin{aligned} MS_3(s) &= \binom{s+5}{5} - \binom{(s-3)+5}{5} \\ &= \binom{s+5}{5} - \binom{s+2}{5}. \end{aligned}$$

(Also see Exercise 8 below.) For example, $MS_3(1) = 6$ (binomial coefficients $\binom{m}{\ell}$ with $m < \ell$ are zero), $MS_3(2) = 21$, and $MS_3(3) = 56 - 1 = 55$. This is the first time the relation (2.11) comes into play.

For readers who have studied Chapter 6 of this book, we should also mention how free resolutions can be used to obtain (2.16). The key point is that the ideal $I = \langle x_1x_2x_3 - x_4x_5x_6\rangle$ is generated by a polynomial of degree 3, so that $I \cong k[x_1, \ldots, x_6](-3)$ as $k[x_1, \ldots, x_6]$-modules. Hence $R = k[x_1, \ldots, x_6]/I$ gives the exact sequence

$$0 \to k[x_1, \ldots, x_6](-3) \to k[x_1, \ldots, x_6] \to R \to 0.$$

Since $H_R(s) = HF_I(s) = MS_3(s)$ by Proposition (2.14), the formula (2.16) follows immediately by the methods of Chapter 6, §4.

These techniques and more sophisticated ideas from commutative algebra, including the theory of toric varieties, have also been applied to the $n \times n$ magic square problem and other related questions from statistics and the design of experiments. We will consider one aspect of the connection with statistics in Exercises 12 and 13 below. We refer the reader to [Sta1] and [Stu2] for a more complete discussion of this interesting connection between algebra and various other areas of the mathematical sciences.

Additional Exercises for §2

Exercise 6. Let R, α and α' be as in the proof of Proposition (2.14). Show that

$$x^\alpha = q(x_1,\dots,x_6)(x_1x_2x_3 - x_4x_5x_6) + x^{\alpha'},$$

where

$$q = \Big((x_1x_2x_3)^{A-1} + (x_1x_2x_3)^{A-2}(x_4x_5x_6) + \cdots + (x_4x_5x_6)^{A-1}\Big)\cdot$$
$$\cdot\, x_1^{a-A}x_2^{b-A}x_3^{c-A}x_4^dx_5^ex_6^f.$$

Deduce that $x^{\alpha'}$ is the standard form of x^α in R.

Exercise 7. Use Exercise 5 to show that if we have any two expressions as in (2.15) for a given M with coefficient vectors $\alpha = (a,\dots,f)$ and $\alpha_1 = (a_1,\dots,f_1)$, then the corresponding monomials x^α and x^{α_1} have the same standard form $x^{\alpha'}$ in $R = k[x_1,\dots,x_6]/\langle x_1x_2x_3 - x_4x_5x_6\rangle$.

Exercise 8. There is another formula, due to MacMahon, for the number of nonnegative integer magic squares of size 3 with a given sum s:

$$MS_3(s) = \binom{s+4}{4} + \binom{s+3}{4} + \binom{s+2}{4}.$$

Show that this formula and (2.16) are equivalent. Hint: This can be proved in several different ways by applying different binomial coefficient identities.

Exercise 9. Verifying that the Hilbert basis for $K_4 = \ker(A_4) \cap \mathbb{Z}^{4\times 4}_{\geq 0}$ consists of exactly 24 elements corresponding to the 4×4 permutation matrices is already a *large* calculation if you apply the Gröbner basis method of Theorem (2.9). For larger n, this approach quickly becomes infeasible because of the large number of variables needed to make the polynomial translation. Fortunately, there is also a non-computational proof that every $n \times n$ matrix M with nonnegative integer entries and row and column sums all equal to s is a linear combination of $n \times n$ permutation matrices with nonnegative integer coefficients. The proof is by induction on the number of nonzero entries in the matrix.

a. The base case of the induction is the case where exactly n of the entries are nonzero (why?). Show in this case that M is equal to sP for some permutation matrix P.
b. Now assume that the theorem has been proved for all M with k or fewer nonzero entries and consider an M with equal row and column sums and $k + 1$ nonzero entries. Using the transversal form of Hall's "marriage" theorem (see, for instance, [Bry]), show that there is some collection of n nonzero entries in M, one from each row and one from each column.
c. Continuing from b, let $d > 0$ be the smallest element in the collection of nonzero entries found in that part, let P be the permutation matrix corresponding to the locations of those nonzero entries, and apply the induction hypothesis to $M - dP$. Deduce the desired result on M.
d. A *doubly stochastic* matrix is an $n \times n$ matrix with nonnegative real entries, all of whose row and column sums equal 1. Adapt the proof sketched in parts a-c to show that the collection of doubly stochastic matrices is the convex hull of the set of $n \times n$ permutation matrices. (See Chapter 7, §1 for more details about convex hulls.)

Exercise 10.
a. How many 3×3 nonnegative integer magic squares with sum s are there if we add the condition that the two diagonal sums should *also* equal s?
b. What about the corresponding question for 4×4 matrices?

Exercise 11. Study the collections of *symmetric* 3×3 and 4×4 nonnegative integer magic squares. What are the Hilbert bases for the monoids of solutions of the corresponding equations? What relations are there? Find the number of squares with a given row and column sum s in each case.

Exercise 12. In this exercise, we will start to develop some ideas concerning *contingency tables* in statistics and see how they relate to the topics discussed in this section. A "two-way" contingency table is an $m \times n$ matrix C with rows labeled according to the m different possible values of some one characteristic of individuals in a population (e.g., political party affiliation, number of TV sets owned, etc. in a human population) and the columns are similarly labeled according to the n possible values of another different characteristic (e.g., response to an item on a questionnaire, age, etc.). The entries are nonnegative integers recording the numbers of individuals in a sample with each combination of values of the two characteristics. The *marginal distribution* of such a table is the collection of row and column sums, giving the total numbers of individuals having each characteristic. For example, if $m = n = 3$ and

$$C = \begin{pmatrix} 34 & 21 & 17 \\ 23 & 21 & 32 \\ 12 & 13 & 50 \end{pmatrix}$$

we have row sums $72, 76, 75$, and column sums $69, 55, 99$.

a. By following what we did for magic squares in the text, show that the collection of all $m \times n$ contingency tables with a given, *fixed* marginal distribution is the set of nonnegative integer solutions of a system of $m + n$ linear equations in mn variables. Give an explicit form for the matrix of your system.
b. Are your equations from part a independent? Why or why not?
c. Is the set of solutions of your system from part a a monoid in $\mathbb{Z}^{mn}_{\geq 0}$ in this case? Why or why not?

Exercise 13. This application comes originally from the article [DS] by Diaconis and Sturmfels. A typical question that statisticians seek to answer is: can we say two characteristics are correlated on the basis of data from a sample of the population? One way that has been proposed to study this sort of problem is to compare values of some statistical measure of correlation from a given sample contingency table and from the other tables with the *same marginal distribution*. In realistic situations it will usually be too difficult to list all the tables having the given marginal distribution (the number can be huge). So a sort of Monte Carlo approach will usually have to suffice. Some number of *randomly generated* tables having the same marginal distribution can be used instead of the whole set. The problem is then to find some efficient way to generate other elements of the collections of tables studied in Exercise 12, given any one element of that collection. Gröbner bases can be used here as follows.

a. Show that C and C' have the same marginal distribution if and only if the difference $T = C' - C$ is an element of the kernel of the matrix of the system of linear equations you found in Exercise 12, part a.
b. To find appropriate matrices T to generate random walks on the set of tables with a fixed marginal distribution, an idea similar to what we did in Theorem (1.11), Proposition (2.8), and Theorem (2.9) of this chapter can be used. Consider a set of "table entry variables" x_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$ (one for each entry in our tables), "row variables" y_i, $1 \leq i \leq m$, and "column variables" z_j, $1 \leq j \leq n$. Let I be the elimination ideal

$$I = \langle x_{ij} - y_i z_j : 1 \leq i \leq m, 1 \leq j \leq n \rangle \cap k[x_{ij} : 1 \leq i \leq m, 1 \leq j \leq n].$$

Show that any difference of monomials $x^\alpha - x^\beta$ contained in I gives a matrix T as in part a. Hint: Use the exponents from α as entries with positive signs, and the exponents from β as entries with negative signs.
c. Compute a Gröbner basis for I in the case $m = n = 3$ using a suitable lexicographic order. Interpret the matrices T you get this way.

§3 Multivariate Polynomial Splines

In this section we will discuss a recent application of the theory of Gröbner bases to the problem of constructing and analyzing the *piecewise polynomial* or *spline* functions with a specified degree of smoothness on polyhedral subdivisions of regions in $\mathbb{R}^n$. Two-variable functions of this sort are frequently used in computer-aided design to specify the shapes of curved surfaces, and the degree of smoothness attainable in some specified class of piecewise polynomial functions is an important design consideration. For an introductory treatment, see [Far]. Uni- and multivariate splines are also used to interpolate values or approximate other functions in numerical analysis, most notably in the *finite element method* for deriving approximate solutions to partial differential equations. The application of Gröbner bases to this subject appeared first in papers of L. Billera and L. Rose ([BR1], [BR2], [BR3], [Ros]). For more recent results, we refer the reader to [SS]. We will need to use the results on Gröbner bases for *modules* over polynomial rings from Chapter 5.

To introduce some of the key ideas, we will begin by considering the simplest case of one-variable spline functions. On the real line, consider the subdivision of an interval $[a, b]$ into two subintervals $[a, c] \cup [c, b]$ given by any c satisfying $a < c < b$. In rough terms, a piecewise polynomial function on this subdivided interval is any function of the form

$$f(x) = \begin{cases} f_1(x) & \text{if } x \in [a, c] \\ f_2(x) & \text{if } x \in [c, b], \end{cases} \tag{3.1}$$

where f_1 and f_2 are polynomials in $\mathbb{R}[x]$. Note that we can always make "trivial" spline functions using *the same* polynomial $f_1 = f_2$ on both subintervals, but those are less interesting because we do not have independent control over the shape of the graph of each piece. Hence we will usually be more interested in finding splines with $f_1 \neq f_2$. Of course, as stated, (3.1) gives us a well-defined function on $[a, b]$ if and only if $f_1(c) = f_2(c)$, and if this is true, then f is *continuous* as a function on $[a, b]$. For instance, taking $a = 0, c = 1, b = 2$, and

$$f(x) = \begin{cases} x + 1 & \text{if } x \in [0, 1] \\ x^2 - x + 2 & \text{if } x \in [1, 2], \end{cases}$$

we get a continuous polynomial spline function. See Fig. 8.3.

Since the polynomial functions f_1, f_2 are C^∞ functions (that is, they have derivatives of all orders) and their derivatives are also polynomials, we can consider the piecewise polynomial derivative functions

$$\begin{cases} f_1^{(r)}(x) & \text{if } x \in [a, c] \\ f_2^{(r)}(x) & \text{if } x \in [c, b] \end{cases}$$

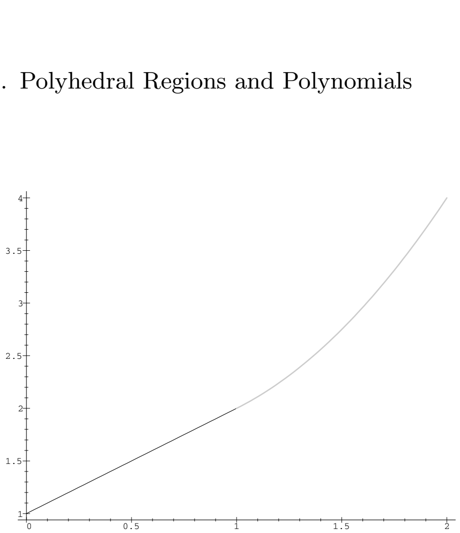

FIGURE 8.3. A continuous spline function

for any $r \geq 0$. As above, we see that f is a C^r function on $[a, b]$ (that is, f is r-times differentiable and its rth derivative, $f^{(r)}$, is continuous) if and only if $f_1^{(s)}(c) = f_2^{(s)}(c)$ for each s, $0 \leq s \leq r$. The following result gives a more algebraic version of this criterion.

(3.2) Proposition. *The piecewise polynomial function f in (3.1) defines a C^r function on $[a, b]$ if and only if the polynomial $f_1 - f_2$ is divisible by $(x - c)^{r+1}$ (that is, $f_1 - f_2 \in \langle (x - c)^{r+1} \rangle$ in $\mathbb{R}[x]$).*

For example, the spline function pictured in Fig. 8.3 is actually a C^1 function since $(x^2 - x + 2) - (x + 1) = (x - 1)^2$. We leave the proof of this proposition to the reader.

Exercise 1. Prove Proposition (3.2).

In practice, it is most common to consider classes of spline functions where the f_i are restricted to be polynomial functions of degree bounded by some fixed integer k. With $k = 2$ we get *quadratic* splines, with $k = 3$ we get *cubic* splines, and so forth.

We will work with two-component splines on a subdivided interval $[a, b] = [a, c] \cup [c, b]$ here. More general subdivisions are considered in Exercise 2 below. We can represent a spline function as in (3.1) by the ordered pair $(f_1, f_2) \in \mathbb{R}[x]^2$. From Proposition (3.2) it follows that the C^r splines form a vector subspace of $\mathbb{R}[x]^2$ under the usual componentwise addition and scalar multiplication. (Also see Proposition (3.10) below, which gives a stronger statement and which includes this one-variable situation as a special case.) Restricting the degree of each component as above, we get elements of the finite-dimensional vector subspace V_k of $\mathbb{R}[x]^2$ spanned by

$$(1,0),\ (x,0),\ \ldots,\ (x^k,0),\ (0,1),\ (0,x),\ \ldots,\ (0,x^k).$$

The C^r splines in V_k form a vector subspace $V_k^r \subset V_k$. We will focus on the following two questions concerning the V_k^r.

(3.3) Questions.

a. *What is the dimension of V_k^r?*

b. *Given k, what is the biggest r for which there exist C^r spline functions f in V_k^r for which $f_1 \neq f_2$?*

We can answer both of these questions easily in this simple setting. First note that any piecewise polynomial in V_k can be uniquely decomposed as the sum of a spline of the form (f,f), and a spline of the form $(0,g)$:

$$(f_1,f_2) = (f_1,f_1) + (0, f_2 - f_1).$$

Moreover, both terms on the right are again in V_k. Any spline function of the form (f,f) is automatically C^r for every $r \geq 0$. On the other hand, by Proposition (3.2), a spline of the form $(0,g)$ defines a C^r function if and only if $(x-c)^{r+1}$ divides g, and this is possible only if $r+1 \leq k$. If $r+1 \leq k$, any linear combination of $(0,(x-c)^{r+1}),\ldots,(0,(x-c)^k)$ gives an element of V_k^r, and these $k-r$ piecewise polynomial functions, together with the $(1,1),(x,x),\ldots,(x^k,x^k)$ give a basis for V_k^r. These observations yield the following answers to (3.3).

(3.4) Proposition. *For one-variable spline functions on a subdivided interval $[a,b] = [a,c] \cup [c,b]$, The dimension of the space V_k^r is*

$$\dim(V_k^r) = \begin{cases} k+1 & \textit{if}\ r+1 > k \\ 2k-r+1 & \textit{if}\ r+1 \leq k. \end{cases}$$

The space V_k^r contains spline functions not of the form (f,f) if and only if $r+1 \leq k$.

For instance, there are C^1 quadratic splines for which $f_1 \neq f_2$, but no C^2 quadratic splines except the ones of the form (f,f). Similarly there are C^2 cubic splines for which $f_1 \neq f_2$, but no C^3 cubic splines of this form. The vector space V_3^2 of C^2 cubic spline functions is 5-dimensional by (3.4). This means, for example, that there is a 2-dimensional space of C^2 cubic splines with any given values $f(a) = A$, $f(c) = C$, $f(b) = B$ at $x = a, b, c$. Because this freedom gives additional control over the shape of the graph of the spline function, one-variable cubic splines are used extensively as interpolating functions in numerical analysis.

The reader should have no difficulty extending all of the above to spline functions on any subdivided interval $[a,b]$, where the subdivision is specified by an arbitrary partition.

Exercise 2. Consider a partition

$$a = x_0 < x_1 < x_2 < \cdots < x_{m-1} < x_m = b$$

of the interval $[a, b]$ into m smaller intervals.

a. Let $(f_1, \ldots, f_m) \in \mathbb{R}[x]^m$ be an m-tuple of polynomials. Define f on $[a, b]$ by setting $f|_{[x_{i-1}, x_i]} = f_i$, Show that f is a C^r function on $[a, b]$ if and only if for each i, $1 \leq i \leq m - 1$, $f_{i+1} - f_i \in \langle (x - x_i)^{r+1} \rangle$.
b. What is the dimension of the space of C^r splines with $\deg f_i \leq k$ for all i? Find a basis. Hint: There exists a nice “triangular” basis generalizing what we did in the text for the case of two subintervals.
c. Show that there is a 2-dimensional space of C^2 cubic spline functions interpolating any specified values at the x_i, $i = 0, \ldots, n$.

We now turn to multivariate splines. Corresponding to subdivisions of intervals in $\mathbb{R}$, we will consider certain subdivisions of *polyhedral* regions in $\mathbb{R}^n$. As in Chapter 7, a *polytope* is the convex hull of a finite set in $\mathbb{R}^n$, and by (1.4) of that chapter, a polytope can be written as the intersection of a collection of affine half-spaces. In constructing partitions of intervals in $\mathbb{R}$, we allowed the subintervals to intersect only at common endpoints. Similarly, in $\mathbb{R}^n$ we will consider subdivisions of polyhedral regions into polytopes that intersect only along common faces.

The major new feature in $\mathbb{R}^n$, $n \geq 2$ is the much greater geometric freedom possible in constructing such subdivisions. We will use the following language to describe them.

(3.5) Definition.

a. A *polyhedral complex* $\Delta \subset \mathbb{R}^n$ is a finite collection of polytopes such that the faces of each element of Δ are elements of Δ, and the intersection of any two elements of Δ is an element of Δ. We will sometimes refer to the k-dimensional elements of a complex Δ as *k-cells*.
b. A polyhedral complex $\Delta \subset \mathbb{R}^n$ is said to be *pure* n-dimensional if every maximal element of Δ (with respect to inclusion) is an n-dimensional polyhedron.
c. Two n-dimensional polytopes in a complex Δ are said to be *adjacent* if they intersect along a common face of dimension $n - 1$.
d. Δ is said to be a *hereditary* complex if for every $\tau \in \Delta$ (including the empty set), any two n-dimensional polytopes σ, σ' of Δ that contain τ can be connected by a sequence $\sigma = \sigma_1, \sigma_2, \ldots, \sigma_m = \sigma'$ in Δ such that each σ_i is n-dimensional, each σ_i contains τ, and σ_i and σ_{i+1} are adjacent for each i.

The cells of a complex give a particularly well-structured subdivision of the polyhedral region $R = \cup_{\sigma \in \Delta} \sigma \subset \mathbb{R}^n$.

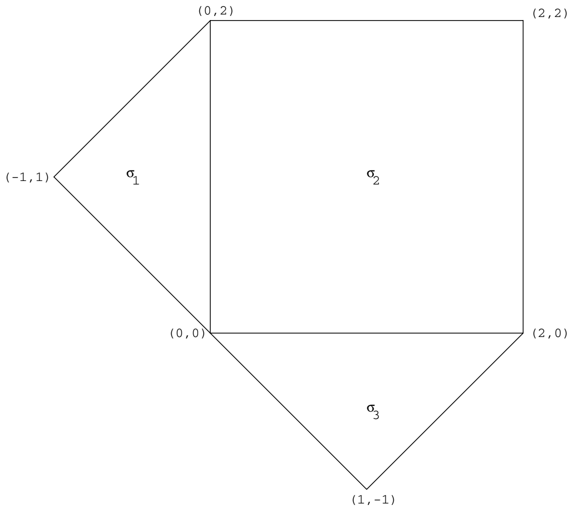

Figure 8.4. A polyhedral complex in $\mathbb{R}^2$

Here are some examples to illustrate the meaning of these conditions. For example, Fig. 8.4 is a picture of a polyhedral complex in $\mathbb{R}^2$ consisting of 18 polytopes in all—the three 2-dimensional polygons $\sigma_1, \sigma_2, \sigma_3$, eight 1-cells (the edges), six 0-cells (the vertices at the endpoints of edges), and the empty set, $\emptyset$.

The condition on intersections in the definition of a complex rules out collections of polyhedra such as the ones in Fig. 8.5. In the collection on

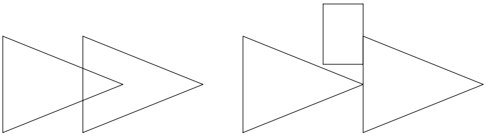

Figure 8.5. Collections of polygons that are not complexes

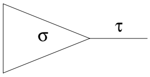

Figure 8.6. A non-pure complex

the left (which consists of *two triangles*, their six edges, their six vertices and the empty set), the intersection of the two 2-cells is not a cell of the complex. Similarly, in the collection on the right (which consists of two triangles and a rectangle, together with their edges and vertices, and the empty set) the 2-cells meet along subsets of their edges, but not along entire edges.

A complex such as the one in Fig. 8.6 is not pure, since τ is maximal and only 1-dimensional.

A complex is *not* hereditary if it is not connected, or if it has maximal elements meeting only along faces of codimension 2 or greater, with no other connection via n-cells, as is the case for the complex in Fig. 8.7. (Here, the cells are the two triangles, their edges and vertices, and finally the empty set.)

Let Δ be any pure n-dimensional polyhedral complex in $\mathbb{R}^n$, let $\sigma_1, \ldots, \sigma_m$ be a given, fixed, ordering of the n-cells in Δ, and let $R = \cup_{i=1}^m \sigma_i$. Generalizing our discussion of univariate splines above, we introduce the following collections of piecewise polynomial functions on R.

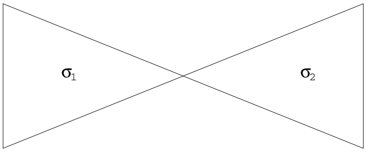

Figure 8.7. A non-hereditary complex

(3.6) Definition.

a. For each $r \geq 0$ we will denote by $C^r(\Delta)$ the collection of C^r functions f on R (that is, functions such that all rth order partial derivatives exist and are continuous on R) such that for every $\delta \in \Delta$ including those of dimension $< n$, the restriction $f|_\delta$ is a polynomial function $f_\delta \in \mathbb{R}[x_1, \ldots, x_n]$.

b. $C^r_k(\Delta)$ is the subset of $f \in C^r(\Delta)$ such that the restriction of f to each cell in Δ is a polynomial function of degree k or less.

Our goal is to study the analogues of Questions (3.3) for the $C^r_k(\Delta)$. Namely, we wish to compute the dimensions of these spaces over $\mathbb{R}$, and to determine when they contain nontrivial splines.

We will restrict our attention in the remainder of this section to complexes Δ that are both pure and hereditary. If σ_i, σ_j are adjacent n-cells of Δ, then they intersect along an interior $(n-1)$-cell $\sigma_{ij} \in \Delta$, a polyhedral subset of an affine hyperplane $\mathbf{V}(\ell_{ij})$, where $\ell_{ij} \in \mathbb{R}[x_1, \ldots, x_n]$ is a polynomial of total degree 1. Generalizing Proposition (3.2) above, we have the following algebraic characterization of the elements of $C^r(\Delta)$ in the case of a pure, hereditary complex.

(3.7) Proposition. *Let Δ be a pure, hereditary complex with m n-cells σ_i. Let $f \in C^r(\Delta)$, and for each i, $1 \leq i \leq m$, let $f_i = f|_{\sigma_i} \in \mathbb{R}[x_1, \ldots, x_n]$. Then for each adjacent pair σ_i, σ_j in Δ, $f_i - f_j \in \langle \ell_{ij}^{r+1} \rangle$. Conversely, any m-tuple of polynomials $(f_1, \ldots, f_m)$ satisfying $f_i - f_j \in \langle \ell_{ij}^{r+1} \rangle$ for each adjacent pair σ_i, σ_j of n-cells in Δ defines an element $f \in C^r(\Delta)$ when we set $f|_{\sigma_i} = f_i$.*

The meaning of Proposition (3.7) is that for pure n-dimensional complexes $\Delta \subset \mathbb{R}^n$, piecewise polynomial functions are determined by their restrictions to the n-cells $\sigma_1, \ldots, \sigma_m$ in Δ. In addition, for hereditary complexes, the C^r property for piecewise polynomial functions f may be checked by comparing *only* the restrictions $f_i = f|_{\sigma_i}$ and $f_j = f|_{\sigma_j}$ for *adjacent* pairs of n-cells.

Proof. If f is an element of $C^r(\Delta)$, then for each adjacent pair σ_i, σ_j of n-cells in Δ, $f_i - f_j$ and all its partial derivatives of order up to and including r must vanish on $\sigma_i \cap \sigma_j$. In Exercise 3 below you will show that this implies $f_i - f_j$ is an element of $\langle \ell_{ij}^{r+1} \rangle$.

Conversely, suppose we have $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_n]$ such that $f_i - f_j$ is an element of $\langle \ell_{ij}^{r+1} \rangle$ for each adjacent pair of n-cells in Δ. In Exercise 3 below, you will show that this implies that f_i and its partial derivatives of order up to and including r agree with f_j and its corresponding derivatives at each point of $\sigma_i \cap \sigma_j$. But the $f_1, \ldots, f_m$ define a C^r function on R if and only if for *every* $\delta \in \Delta$ and every pair of n-cells σ_p, σ_q containing δ (not only adjacent ones) f_p and its partial derivatives of order up to and including r agree with f_q and its corresponding derivatives at each point of δ. So let p, q be any pair of indices for which $\delta \subset \sigma_p \cap \sigma_q$. Since Δ is hereditary, there is a sequence of n-cells

$$\sigma_p = \sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_k} = \sigma_q,$$

each containing δ, such that σ_{i_j} and $\sigma_{i_{j+1}}$ are adjacent. By assumption, this implies that for each j, $f_{i_j} - f_{i_{j+1}}$ and all its partial derivatives of orders up to and including r vanish on $\sigma_{i_j} \cap \sigma_{i_{j+1}} \supset \delta$. But

$$f_p - f_q = (f_{i_1} - f_{i_2}) + (f_{i_2} - f_{i_3}) + \cdots + (f_{i_{k-1}} - f_{i_k})$$

and each term on the right and its partials up to and including order r vanish on δ. Hence $f_1, \dots, f_m$ define an element of $C^r(\Delta)$. □

Exercise 3. Let σ, σ' be two adjacent n-cells in a polyhedral complex Δ, and let $\sigma \cap \sigma' \subset \mathbf{V}(\ell)$ for a linear polynomial $\ell \in \mathbb{R}[x_1, \dots, x_n]$.

a. Show that if $f, f' \in \mathbb{R}[x_1, \dots, x_n]$ satisfy $f - f' \in \langle \ell^{r+1} \rangle$, then the partial derivatives of all orders $\leq r$ of f and f' agree at every point in $\sigma \cap \sigma'$.

b. Conversely if the partial derivatives of all orders $\leq r$ of f and f' agree at every point in $\sigma \cap \sigma'$, show that $f - f' \in \langle \ell^{r+1} \rangle$.

Fixing any one ordering on the n-cells σ_i in Δ, we will represent elements f of $C^r(\Delta)$ by ordered m-tuples $(f_1, \dots, f_m) \in \mathbb{R}[x_1, \dots, x_n]^m$, where $f_i = f|_{\sigma_i}$.

Consider the polyhedral complex Δ in $\mathbb{R}^2$ from Fig. 8.4, with the numbering of the 2-cells given there. It is easy to check that Δ is hereditary. The interior edges are given by $\sigma_1 \cap \sigma_2 \subset \mathbf{V}(x)$ and $\sigma_2 \cap \sigma_3 \subset \mathbf{V}(y)$. By the preceding proposition, an element $(f_1, f_2, f_3) \in \mathbb{R}[x, y]^3$ gives an element of $C^r(\Delta)$ if and only if

$$\begin{aligned} f_1 - f_2 &\in \langle x^{r+1} \rangle, \quad \text{and} \\ f_2 - f_3 &\in \langle y^{r+1} \rangle. \end{aligned}$$

To prepare for our next result, note that these inclusions can be rewritten in the form

$$\begin{aligned} f_1 - f_2 + x^{r+1} f_4 &= 0 \\ f_2 - f_3 + y^{r+1} f_5 &= 0 \end{aligned}$$

for some $f_4, f_5 \in \mathbb{R}[x, y]$. These equations can be rewritten again in vector-matrix form as

$$\begin{pmatrix} 1 & -1 & 0 & x^{r+1} & 0 \\ 0 & 1 & -1 & 0 & y^{r+1} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus, elements of $C^r(\Delta)$ are projections onto the first three components of elements of the kernel of the map $\mathbb{R}[x,y]^5 \to \mathbb{R}[x,y]^2$ defined by

$$M(\Delta, r) = \begin{pmatrix} 1 & -1 & 0 & x^{r+1} & 0 \\ 0 & 1 & -1 & 0 & y^{r+1} \end{pmatrix}. \tag{3.8}$$

By Proposition (1.10) and Exercise 9 of §3 of Chapter 5, it follows that $C^r(\Delta)$ has the structure of a *module* over the ring $\mathbb{R}[x,y]$. This observation allows us to apply the theory of Gröbner bases to study splines.

Our next result gives a corresponding statement for $C^r(\Delta)$ in general. We begin with some necessary notation. Let Δ be a pure, hereditary polyhedral complex in $\mathbb{R}^n$. Let m be the number of n-cells in Δ, and let e be the number of *interior* $(n-1)$-cells (the intersections $\sigma_i \cap \sigma_j$ for adjacent n-cells). Fix some ordering $\tau_1, \ldots, \tau_e$ for the interior $(n-1)$-cells and let ℓ_s be a linear polynomial defining the affine hyperplane containing τ_s. Consider the $e \times (m+e)$ matrix $M(\Delta, r)$ with the following block decomposition:

$$M(\Delta, r) = (\partial(\Delta) \mid D). \tag{3.9}$$

(Note: the orderings of the rows and columns are determined by the orderings of the indices of the n-cells and the interior $(n-1)$-cells, but any ordering can be used.) In (3.9), $\partial(\Delta)$ is the $e \times m$ matrix defined by this rule: In the sth row, if $\tau_s = \sigma_i \cap \sigma_j$ with $i < j$, then

$$\partial(\Delta)_{sk} = \begin{cases} +1 & \text{if } k = i \\ -1 & \text{if } k = j \\ 0 & \text{otherwise.} \end{cases}$$

In addition, D is the $e \times e$ diagonal matrix

$$D = \begin{pmatrix} \ell_1^{r+1} & 0 & \cdots & 0 \\ 0 & \ell_2^{r+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \ell_e^{r+1} \end{pmatrix}.$$

Then as in the example above we have the following statement.

(3.10) Proposition. *Let Δ be a pure, hereditary polyhedral complex in $\mathbb{R}^n$, and let $M(\Delta, r)$ be the matrix defined in (3.9) above.*

a. *An m-tuple $(f_1, \ldots, f_m)$ is in $C^r(\Delta)$ if and only if there exist $(f_{m+1}, \ldots, f_{m+e})$ such that $f = (f_1, \ldots, f_m, f_{m+1}, \ldots, f_{m+e})^T$ is an element of the kernel of the map $\mathbb{R}[x_1, \ldots, x_n]^{m+e} \to \mathbb{R}[x_1, \ldots, x_n]^e$ defined by the matrix $M(\Delta, r)$.*
b. *$C^r(\Delta)$ has the structure of a module over the ring $\mathbb{R}[x_1, \ldots, x_n]$. In the language of Chapter 5, it is the image of the projection homomorphism from $\mathbb{R}[x_1, \ldots, x_n]^{m+e}$ onto $\mathbb{R}[x_1, \ldots, x_n]^m$ (in the first m components) of the module of syzygies on the columns of $M(\Delta, r)$.*
c. *$C^r_k(\Delta)$ is a finite-dimensional vector subspace of $C^r(\Delta)$.*

Proof. Part a is essentially just a restatement of Proposition (3.7). For each interior $(n-1)$-cell $\tau_s = \sigma_i \cap \sigma_j$, $(i < j)$ we have an equation

$$f_i - f_j = -\ell_s^{r+1} f_{m+s}$$

for some $f_{m+s} \in \mathbb{R}[x_1, \ldots, x_n]$. This is the equation obtained by setting the sth component of the product $M(\Delta, r)f$ equal to zero.

Part b follows immediately from part a as in Chapter 5, Proposition (1.10) and Exercise 9 of Chapter 5, §3.

Part c follows by a direct proof, or more succinctly from part b, since $C_k^r(\Delta)$ is closed under sums and products by constant polynomials. □

The Gröbner basis algorithm based on Schreyer's Theorem (Chapter 5, Theorem (3.3)) may be applied to compute a Gröbner basis for the kernel of $M(\Delta, r)$ for each r, and from that information the dimensions of, and bases for, the $C_k^r(\Delta)$ may be determined.

As a first example, let us compute the $C^r(\Delta)$ for the complex $\Delta \subset \mathbb{R}^2$ from (3.8). We consider the matrix as in (3.8) with $r = 1$ first. Using any monomial order in $\mathbb{R}[x, y]^5$ with $e_5 > \cdots > e_1$, we compute a Gröbner basis for $\ker(M(\Delta, 1)$ (that is, the module of syzygies of the columns of $M(\Delta, 1)$) and we find three basis elements, the transposes of

$$\begin{aligned} g_1 &= (1, 1, 1, 0, 0) \\ g_2 &= (-x^2, 0, 0, 1, 0) \\ g_3 &= (-y^2, -y^2, 0, 0, 1). \end{aligned}$$

(In this simple case, it is easy to write down these syzygies by inspection. They must generate the module of syzygies because of the form of the matrix $M(\Delta, r)$—the last three components of the vector f are arbitary, and these determine the first two.) The elements of $C^1(\Delta)$ are given by projection on the first three components, so we see that the general element of $C^1(\Delta)$ will have the form

$$\begin{aligned} &f(1, 1, 1) + g(-x^2, 0, 0) + h(-y^2, -y^2, 0) \\ &\quad = (f - gx^2 - hy^2, f - hy^2, f), \end{aligned} \tag{3.11}$$

where $f, g, h \in \mathbb{R}[x, y]^2$ are arbitrary polynomials. Note that the triples with $g = h = 0$ are the "trivial" splines where we take the same polynomial on each σ_i, while the other generators contribute terms supported on only one or two of the 2-cells. The algebraic structure of $C^1(\Delta)$ as a module over $\mathbb{R}[x, y]$ is very simple—$C^1(\Delta)$ is a *free* module and the given generators form a module basis. (Billera and Rose show in Lemma 3.3 and Theorem 3.5 of [BR3] that the same is true for $C^r(\Delta)$ for *any* hereditary complex $\Delta \subset \mathbb{R}^2$ and all $r \geq 1$.) Using the decomposition it is also easy to count the dimension of $C_k^1(\Delta)$ for each k. For $k = 0, 1$, we have only the "trivial" splines, so $\dim C_0^1(\Delta) = 1$, and $\dim C_1^1(\Delta) = 3$ (a vector space basis is $\{(1,1,1), (x,x,x), (y,y,y)\}$). For $k \geq 2$, there are nontrivial splines as well, and we see by counting monomials of the appropriate degrees in f, g, h that

$$\dim C_k^1(\Delta) = \binom{k+2}{2} + 2\binom{(k-2)+2}{2} = \binom{k+2}{2} + 2\binom{k}{2}.$$

Also see Exercise 9 below for a more succinct way to package the information from the function $\dim C_k^1(\Delta)$.

For larger r, the situation is entirely analogous in this example. A Gröbner basis for the kernel of $M(\Delta, r)$ is given by

$$\begin{aligned} g_1 &= (1,1,1,0,0)^T \\ g_2 &= (-x^{r+1}, 0, 0, 1, 0)^T \\ g_3 &= (-y^{r+1}, -y^{r+1}, 0, 0, 1)^T, \end{aligned}$$

and we have that $C^r(\Delta)$ is a free module over $\mathbb{R}[x,y]$ for all $r \geq 0$. Thus

$$\dim C_k^r(\Delta) = \begin{cases} \binom{k+2}{2} & \text{if } k < r+1 \\ \binom{k+2}{2} + 2\binom{k-r+1}{2} & \text{if } k \geq r+1. \end{cases}$$

Our next examples, presented as exercises for the reader, indicate some of the subtleties that can occur for more complicated complexes. (Additional examples can be found in the exercises at the end of the section.)

Exercise 4. In $\mathbb{R}^2$, consider the convex quadrilateral

$$R = \mathrm{Conv}(\{(2,0), (0,1), (-1,1), (-1,-2)\})$$

(notation as in §1 of Chapter 7), and subdivide R into triangles by connecting each vertex to the origin by line segments. We obtain in this way a pure, hereditary polyhedral complex Δ containing four 2-cells, eight 1-cells (four interior ones), five 0-cells, and $\emptyset$. Number the 2-cells $\sigma_1, \ldots, \sigma_4$ proceeding counter-clockwise around the origin starting from the triangle $\sigma_1 = \mathrm{Conv}(\{(2,0), (0,0), (0,1)\})$. The interior 1-cells of Δ are then

$$\begin{aligned} &\sigma_1 \cap \sigma_2 \subset \mathbf{V}(x) \\ &\sigma_2 \cap \sigma_3 \subset \mathbf{V}(x+y) \\ &\sigma_3 \cap \sigma_4 \subset \mathbf{V}(2x-y) \\ &\sigma_1 \cap \sigma_4 \subset \mathbf{V}(y). \end{aligned}$$

a. Using this ordering on the interior 1-cells, show that we obtain

$$\begin{pmatrix} 1 & -1 & 0 & 0 & x^{r+1} & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & (x+y)^{r+1} & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & (2x-y)^{r+1} & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & y^{r+1} \end{pmatrix}$$

for the matrix $M(\Delta, r)$.

b. With $r = 1$, for instance, show that a Gröbner basis for the $\mathbb{R}[x, y]$-module of syzygies on the columns of $M(\Delta, 1)$ is given by the transposes of the following vectors

$$\begin{aligned} g_1 &= (1, 1, 1, 1, 0, 0, 0, 0) \\ g_2 &= (1/4)(3y^2, 6x^2 + 3y^2, 4x^2 - 4xy + y^2, 0, 6, -2, -1, -3) \\ g_3 &= (2xy^2 + y^3, 0, 0, y, -y, 0, -2x - y) \\ g_4 &= (-3xy^2 - 2y^3, x^3 - 3xy^2 - 2y^3, 0, 0, x, -x + 2y, 0, 3x + 2y) \\ g_5 &= (x^2y^2, 0, 0, 0, -y^2, 0, 0, -x^2). \end{aligned}$$

c. As before, the elements of $C^1(\Delta)$ are obtained by projection onto the first four components. From this, show that there are only "trivial" splines in $C^1_0(\Delta)$ and $C^1_1(\Delta)$, but g_2 and its multiples give nontrivial splines in all degrees $k \geq 2$, while g_3 and g_4 also contribute terms in degrees $k \geq 3$.

d. Show that the g_i form a basis for $C^1(\Delta)$, so it is a free module. Thus

$$\dim C^1_k(\Delta) = \begin{cases} 1 & \text{if } k = 0 \\ 3 & \text{if } k = 1 \\ 7 & \text{if } k = 2 \\ \binom{k+2}{2} + \binom{k}{2} + 2\binom{k-1}{2} & \text{if } k \geq 3. \end{cases}$$

We will next consider a second polyhedral complex Δ' in $\mathbb{R}^2$ which has the same combinatorial data as Δ in Exercise 4 (that is, the numbers of k-cells are the same for all k, the containment relations are the same, and so forth), but which is in special position.

Exercise 5. In $\mathbb{R}^2$, consider the convex quadrilateral

$$R = \mathrm{Conv}(\{(2, 0), (0, 1), (-1, 0), (0, -2)\}).$$

Subdivide R into triangles by connecting each vertex to the origin by line segments. This gives a pure, hereditary polyhedral complex Δ' with four 2-cells, eight 1-cells (four interior ones), five 0-cells, and $\emptyset$. Number the 2-cells $\sigma_1, \ldots, \sigma_4$ proceeding counter-clockwise around the origin starting from the triangle σ_1 with vertices $(2, 0)$, $(0, 0)$, $(0, 1)$. The interior 1-cells of

Δ are then

$$\begin{aligned}\sigma_1 \cap \sigma_2 &\subset \mathbf{V}(x)\\ \sigma_2 \cap \sigma_3 &\subset \mathbf{V}(y)\\ \sigma_3 \cap \sigma_4 &\subset \mathbf{V}(x)\\ \sigma_1 \cap \sigma_4 &\subset \mathbf{V}(y).\end{aligned}$$

This is what we meant before by saying that Δ' is in special position—the interior edges lie on only two distinct lines, rather than four of them.

a. Using this ordering on the interior 1-cells, show that we obtain

$$M(\Delta', r) = \begin{pmatrix} 1 & -1 & 0 & 0 & x^{r+1} & 0 & 0 & 0\\ 0 & 1 & -1 & 0 & 0 & y^{r+1} & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & 0 & x^{r+1} & 0\\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & y^{r+1}\end{pmatrix}.$$

b. With $r = 1$, for instance, show that a Gröbner basis for the $\mathbb{R}[x,y]$-module of syzygies on the columns of $M(\Delta', 1)$ is given by the transposes of

$$\begin{aligned}g_1' &= (1,1,1,1,0,0,0,0)\\ g_2' &= (0,x^2,x^2,0,1,0,-1,0)\\ g_3' &= (y^2,y^2,0,0,0,-1,0,-1)\\ g_4' &= (x^2y^2,0,0,0,-y^2,0,0,-x^2).\end{aligned}$$

Note that these generators have a different form (in particular, the components have different total degrees) than the generators for the syzygies on the columns of $M(\Delta, 1)$.

c. Check that the g_i' form a basis of $C^1(\Delta')$, and that

$$\dim C_k^1(\Delta') = \begin{cases} 1 & \text{if } k = 0\\ 3 & \text{if } k = 1\\ 8 & \text{if } k = 2\\ 16 & \text{if } k = 3\\ \binom{k+2}{2} + 2\binom{k}{2} + \binom{k-2}{2} & \text{if } k \geq 3.\end{cases}$$

Comparing Exercises 4 and 5, we see that the dimensions of $C_k^r(\Delta)$ can depend on more than just the combinatorial data of the polyhedral complex Δ—they can vary depending on the positions of the interior $(n-1)$-cells.

The recent paper [Ros] of Lauren Rose sheds some light on examples like these. To describe her results, it will be convenient to use the following notion.

(3.12) Definition. The *dual graph* G_Δ of a pure n-dimensional complex Δ is the graph with vertices corresponding to the n-cells in Δ, and edges corresponding to adjacent pairs of n-cells.

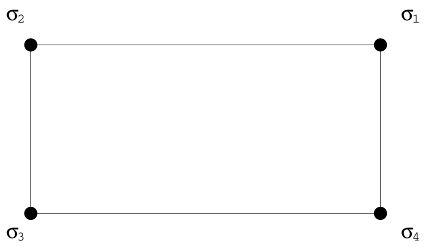

FIGURE 8.8. The dual graph

For instance, the dual graphs for the complexes in Exercises 4 and 5 are both equal to the graph in Fig. 8.8. By an easy translation of the definition in (3.5), the dual graph of a hereditary complex is *connected*.

As before, we will denote by e the number of interior $(n-1)$-cells and let $\delta_1, \ldots, \delta_e$ denote some ordering of them. Choose an ordering on the vertices of G_Δ (or equivalently on the n-cells of Δ), and consider the induced orientations of the edges. If $\delta = jk$ is the oriented edge from vertex j to vertex k in G_Δ, corresponding to the interior $(n-1)$-cell $\delta = \sigma_j \cap \sigma_k$, let ℓ_δ be the equation of the affine hyperplane containing δ. By convention, we take the *negative*, $-\ell_\delta$, as the defining equation for the affine hyperplane containing the edge kj with reversed orientation. For simplicity, we will also write ℓ_i for the linear polynomial ℓ_{δ_i}. Finally, let $\mathcal{C}$ denote the set of cycles in G_Δ. Then, following Rose, we consider a module $B^r(\Delta)$ built out of syzygies on the ℓ_i^{r+1}.

(3.13) Definition. $B^r(\Delta) \subset \mathbb{R}[x_1, \ldots, x_n]^e$ is the submodule defined by

$$B^r(\Delta) = \{(g_1, \ldots, g_e) \in \mathbb{R}[x_1, \ldots, x_n]^e : \text{ for all } c \in \mathcal{C}, \sum_{\delta \in c} g_\delta \ell_\delta^{r+1} = 0\}.$$

The following observation is originally due to Schumaker for the case of bivariate splines (see [Schu]). Our treatment follows Theorem 2.2 of [Ros].

(3.14) Theorem. *If Δ is hereditary, then $C^r(\Delta)$ is isomorphic to $B^r(\Delta) \oplus \mathbb{R}[x_1, \ldots, x_n]$ as an $\mathbb{R}[x_1, \ldots, x_n]$-module.*

PROOF. Consider the mapping

$$\varphi : C^r(\Delta) \to B^r(\Delta) \oplus \mathbb{R}[x_1, \ldots, x_n]$$

defined in the following way. By (3.7), for each $f = (f_1, \ldots, f_m)$ in $C^r(\Delta)$ and each interior $(n-1)$-cell $\delta_i = \sigma_j \cap \sigma_k$, we have $f_j - f_k = g_i \ell_i^{r+1}$ for some $g_i \in \mathbb{R}[x_1, \ldots, x_n]$. Let

$$\varphi(f) = \big((g_1, \ldots, g_e), f_1\big)$$

(the f_1 is the component in the $\mathbb{R}[x_1, \ldots, x_n]$ summand). For each cycle c in the dual graph, $\sum_{\delta \in c} g_\delta \ell_\delta^{r+1}$ equals a sum of the form $\sum (f_j - f_k)$, which cancels completely to 0 since c is a cycle. Hence, the e-tuple $(g_1, \ldots, g_e)$ is an element of $B^r(\Delta)$. It is easy to see that φ is a homomorphism of $\mathbb{R}[x_1, \ldots, x_n]$-modules.

To show that φ is an isomorphism, consider any

$$\big((g_1, \ldots, g_e), f\big) \in B^r(\Delta) \oplus \mathbb{R}[x_1, \ldots, x_n].$$

Let $f_1 = f$. For each i, $2 \le i \le m$, since G_Δ is connected, there is some path from vertex σ_1 to σ_i in G_Δ, using the edges in some set E. Let $f_i = f + \sum_{\delta \in E} g_\delta \ell_\delta^{r+1}$, where as above the g_δ are defined by $f_j - f_k = g_\delta \ell_\delta^{r+1}$ if δ is the oriented edge jk. Any two paths between these two vertices differ by a combination of cycles, so since $(g_1, \ldots, g_e) \in B^r(\Delta)$, f_i is a well-defined polynomial function on σ_i, and the m-tuple $(f_1, \ldots, f_m)$ gives a well-defined element of $C^r(\Delta)$ (why?). We obtain in this way a homomorphism

$$\psi : B^r(\Delta) \oplus \mathbb{R}[x_1, \ldots, x_n] \to C^r(\Delta),$$

and it is easy to check that ψ and φ are inverses. □

The algebraic reason for the special form of the generators of the module $C^1(\Delta)$ in Exercise 5 as compared to those in Exercise 4 can be read off easily from the alternate description of $C^1(\Delta)$ given by Theorem (3.14). For the dual graph shown in Fig. 8.8 on the previous page, there is exactly one cycle. In Exercise 4, numbering the edges counterclockwise, we have

$$\ell_1^2 = x^2,\ \ell_2^2 = (x+y)^2,\ \ell_3^2 = (2x-y)^2,\ \ell_4^2 = y^2.$$

It is easy to check that the dimension over $\mathbb{R}$ of the subspace of $B(\Delta)$ with g_i constant for all i is 1, so that applying the mapping ψ from the proof of Theorem (3.14), the quotient of the space $C_2^1(\Delta)$ of quadratic splines modulo the trivial quadratic splines is 1-dimensional. (The spline g_2 from part b of the exercise gives a basis.) On the other hand, in Exercise 5,

$$\ell_1^2 = x^2,\ \ell_2^2 = y^2,\ \ell_3^2 = x^2,\ \ell_4^2 = y^2,$$

so $B^1(\Delta)$ contains both $(1, 0, -1, 0)$ and $(0, 1, 0, -1)$. Under ψ, we obtain that the quotient of $C_2^1(\Delta)$ modulo the trivial quadratic splines is two-dimensional.

As an immediate corollary of Theorem (3.14), we note the following general sufficient condition for $C^r(\Delta)$ to be a free module.

(3.15) Corollary. *If Δ is hereditary and G_Δ is a tree (i.e., a connected graph with no cycles), then $C^r(\Delta)$ is free for all $r \geq 0$.*

PROOF. If there are no cycles, then $B^r(\Delta)$ is equal to the free module $\mathbb{R}[x_1, \ldots, x_n]^e$, and the corollary follows from Theorem (3.14). This result is Theorem 3.1 of [Ros]. □

Returning to bivariate splines, for *generic* pure 2-dimensional hereditary *simplicial* complexes Δ in $\mathbb{R}^2$ (that is, complexes where all 2-cells are *triangles* whose edges are in sufficiently general position) giving triangulations of 2-manifolds with boundary in the plane, there is a simple combinatorial formula for $\dim C^1_k(\Delta)$ first conjectured by Strang, and proved by Billera (see [Bil1]). The form of this dimension formula given in [BR1] is the following:

$$(3.16) \qquad \dim C^1_k(\Delta) = \binom{k+2}{2} + (h_1 - h_2)\binom{k}{2} + 2h_2\binom{k-1}{2}.$$

Here h_1 and h_2 are determined by purely combinatorial data from Δ:

$$(3.17) \qquad h_1 = V - 3 \quad \text{and } h_2 = 3 - 2V + E,$$

where V is the number of 0-cells, and E is the number of 1-cells in Δ. (Also see Exercise 12 below for Strang's original dimension formula, and its connection to (3.16).)

For example, the simplicial complex Δ in Exercise 4, in which the interior edges lie on four distinct lines (the generic situation) has $V = 5$ and $E = 8$, so $h_1 = 2$ and $h_2 = 1$. Hence (3.16) agrees with the formula from part d of the exercise. On the other hand, the complex Δ' from Exercise 5 is not generic as noted above, and (3.16) is not valid for Δ'.

Interestingly enough, there is no corresponding statement for $n \geq 3$. Moreover, the modules $C^r(\Delta)$ can fail to be free modules even in very simple cases (see part c of Exercise 10 below for instance). The paper [Sche] gives necessary and sufficient conditions for freeness of $C^r(\Delta)$ and shows that the first three terms of its Hilbert polynomial can be determined from the combinatorics and local geometry of Δ. The case $n = 3$, $r = 1$ is also studied in [ASW]. Nevertheless, this is still an area with many open questions.

ADDITIONAL EXERCISES FOR §3

Exercise 6. Investigate the modules $C^r(\Delta)$ and $C^r(\Delta')$, $r \geq 2$, for the complexes from Exercises 4 and 5. What are $\dim C^r_k(\Delta)$ and $\dim C^r_k(\Delta')$?

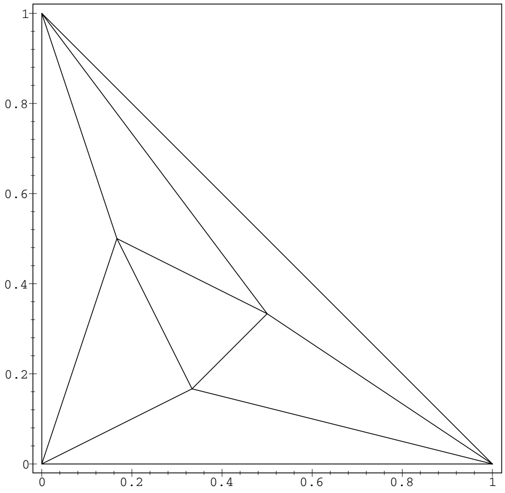

FIGURE 8.9. Figure for Exercise 7

Exercise 7. Let Δ be the simplicial complex in $\mathbb{R}^2$ given in Fig. 8.9. The three interior vertices are at $(1/3, 1/6), (1/2, 1/3)$, and $(1/6, 1/2)$.

a. Find the matrix $M(\Delta, r)$ for each $r \geq 0$.
b. Show that
$$\dim C^1_k(\Delta) = \binom{k+2}{2} + 6\binom{k-1}{2}$$
(where if $k < 3$, by convention, the second term is taken to be zero).
c. Verify that formula (3.16) is valid for this Δ.

Exercise 8. In the examples we presented in the text, the components of our Gröbner basis elements were all homogeneous polynomials. This will not be true in general. In particular, this may fail if some of the interior $(n-1)$-cells of our complex Δ lie on hyperplanes which do not contain the origin in $\mathbb{R}^n$. Nevertheless, there is a variant of *homogeneous coordinates* used to specify points in projective spaces—see [CLO] Chapter 8—that we can use if we want to work with homogeneous polynomials exclusively. Namely, think of a given pure, hereditary complex Δ as a subset of the hyperplane $x_{n+1} = 1$, a copy of $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$. By considering the *cone* $\overline{\sigma}$

over each k-cell $\sigma \in \Delta$ with vertex at $(0, \ldots, 0, 0)$ in $\mathbb{R}^{n+1}$, we get a new polyhedral complex $\overline{\Delta}$ in $\mathbb{R}^{n+1}$.

a. Show that n-cells σ, σ' from Δ are adjacent if and only the corresponding $\overline{\sigma}, \overline{\sigma'}$ are adjacent $(n+1)$-cells in $\overline{\Delta}$. Show that $\overline{\Delta}$ is hereditary.

b. What are the equations of the interior n-cells in $\overline{\Delta}$?

c. Given $f = (f_1 \ldots, f_m) \in C^r_k(\Delta)$, show that the component-wise homogenization with respect to x_{n+1}, $f^h = (f^h_1 \ldots, f^h_m)$, gives an element of $C^r_k(\overline{\Delta})$.

d. How are the matrices $M(\Delta, r)$ and $M(\Delta', r)$ related?

e. Describe the relation between $\dim C^r_k(\Delta)$ and $\dim C^r_k(\overline{\Delta})$.

Exercise 9. In this exercise we will assume that the construction of Exercise 8 has been applied, so that $C^r(\Delta)$ is a graded module over $\mathbb{R}[x_0, \ldots, x_n]$. Then the formal power series

$$H(C^r(\Delta), u) = \sum_{k=0}^{\infty} \dim C^r_k(\Delta) u^k$$

is the *Hilbert series* of the graded module $C^r(\Delta)$. This is the terminology of Exercise 24 of Chapter 6, §4, and that exercise showed that the Hilbert series can be written in the form

$$(3.18) \qquad H(C^r(\Delta), u) = P(u)/(1-u)^{n+1},$$

where $P(u)$ is a polynomial in u with coefficients in $\mathbb{Z}$. We obtain the series from (3.18) by using the formal geometric series expansion

$$1/(1-u) = \sum_{k=0}^{\infty} u^k.$$

a. Show that the Hilbert series for the module $C^1(\Delta)$ from (3.8) with $r = 1$ is given by

$$(1 + 2u^2)/(1-u)^3.$$

b. Show that the Hilbert series for the module $C^1(\Delta)$ from Exercise 4 is

$$(1 + u^2 + 2u^3)/(1-u)^3.$$

c. Show that the Hilbert series for the module $C^1(\Delta')$ from Exercise 5 is

$$(1 + 2u^2 + u^4)/(1-u)^3.$$

d. What is the Hilbert series for the module $C^1(\Delta)$ from Exercise 7 above?

Exercise 10. Consider the polyhedral complex Δ in $\mathbb{R}^3$ formed by subdividing the octahedron with vertices $\pm e_i$, $i = 1, 2, 3$ into 8 tetrahedra by adding an interior vertex at the origin.

a. Find the matrix $M(\Delta, r)$.
b. Find formulas for the dimensions of $C^1_k(\Delta)$ and $C^2_k(\Delta)$.
c. What happens if we move the vertex of the octahedron at e_3 to $(1, 1, 1)$ to form a new, combinatorially equivalent, subdivided octahedron Δ'? Using *Macaulay 2*'s `hilbertSeries` command, compute the Hilbert series of the graded module $\ker M(\Delta', 1)$ and from the result deduce that $C^1(\Delta')$ cannot be a free module. Hint: In the expression (3.19) for the dimension series of a free module, the coefficients in the numerator $P(t)$ must all be positive; do you see why?

Exercise 11. This exercise uses the language of exact sequences and some facts about graded modules from Chapter 6. The method used in the text to compute dimensions of $C^r_k(\Delta)$ requires the computation of a Gröbner basis for the module of syzygies on the columns of $M(\Delta, r)$, and it yields information leading to explicit bases of the spline spaces $C^r_k(\Delta)$. If bases for these spline spaces are not required, there is another method which can be used to compute the Hilbert series directly from $M(\Delta, r)$ without computing the syzygy module. We will assume that the construction of Exercise 8 has been applied, so that the last e columns of the matrix $M(\Delta, r)$ consist of homogeneous polynomials of degree $r + 1$. Write $R = \mathbb{R}[x_1, \ldots, x_n]$ and consider the exact sequence of graded R-modules

$$0 \to \ker M(\Delta, r) \to R^m \oplus R(-r-1)^e \to \operatorname{im} M(\Delta, r) \to 0.$$

a. Show that the Hilbert series of $R^m \oplus R(-r-1)^e$ is given by

$$(m + eu^{r+1})/(1-u)^{n+1}.$$

b. Show that the Hilbert series of the graded module $\ker M(\Delta, r)$ is the *difference* of the Hilbert series from part a and the Hilbert series of the image of $M(\Delta, r)$.

The Hilbert series of the image can be computed by applying Buchberger's algorithm to the module M generated by the columns of $M(\Delta, r)$, then applying the fact that M and $\langle \textsc{lt}(M)\rangle$ have the same Hilbert function.

Exercise 12. Strang's original conjectured formula for the dimension of $C^1_k(\Delta)$ for a simplicial complex in the plane with F triangles, E_0 *interior edges*, and V_0 *interior vertices* was

$$\dim C^1_k(\Delta) = \binom{k+2}{2} F - (2k+1)E_0 + 3V_0, \tag{3.19}$$

and this is the form proved in [Bil1]. In this exercise, you will show that this form is equivalent to (3.16), under the assumption that Δ gives a triangulation of a topological disk in the plane. Let E and V be the total numbers of edges and vertices respectively.

a. Show that $V - E + F = 1$ and $V_0 - E_0 + F = 1$ for such a triangulation. Hint: One approach is to use induction on the number of triangles. In topological terms, the first equation gives the usual Euler characteristic, and the second gives the Euler characteristic relative to the boundary.

b. Use part a and the edge-counting relation $3F = E + E_0$ to show that $E = 3 + 2E_0 - 3V_0$ and $V = 3 + E_0 - 2V_0$.

c. Show that if F is eliminated using part a, and the expressions for V and E from part b are substituted into (3.16), then (3.19) is obtained. Conversely, show that (3.19) implies (3.16).

Exercise 13. The methods introduced in this section work for some *algebraic*, but non-polyhedral, decompositions of regions in $\mathbb{R}^n$ as well. We will not essay a general development. Instead we will indicate the idea with a simple example. In $\mathbb{R}^2$ suppose we wanted to construct C^r piecewise polynomial functions on the union R of the regions σ_1, σ_2, σ_3 as in Fig. 8.10. The outer boundary is the circle of radius 1 centered at the origin, and the three interior edges are portions of the curves $y = x^2$, $x = -y^2$, and $y = x^3$, respectively.

We can think of this as a non-linear embedding of an abstract 2-dimensional polyhedral complex.

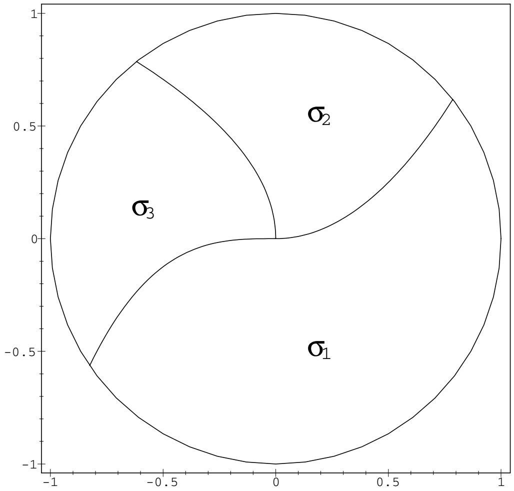

FIGURE 8.10. Figure for Exercise 13

a. Show that a triple $(f_1, f_2, f_3) \in \mathbb{R}[x, y]^3$ defines a C^r spline function on R if and only if

$$\begin{aligned} f_1 - f_2 &\in \langle (y - x^2)^{r+1} \rangle \\ f_2 - f_3 &\in \langle (x + y^2)^{r+1} \rangle \\ f_1 - f_3 &\in \langle (y - x^3)^{r+1} \rangle. \end{aligned}$$

b. Express the C^1 splines on this subdivided region as the kernel of an appropriate matrix with polynomial entries, and find the Hilbert function for the kernel.

Exercise 14. (The Courant functions and the face ring of a complex, see [Sta1]) Let Δ be a pure n-dimensional, hereditary complex in $\mathbb{R}^n$. Let $v_1, \ldots, v_q$ be the vertices of Δ (the 0-cells).

a. For each i, $1 \leq i \leq q$, show that there is a unique function $X_i \in C_1^0(\Delta)$ (that is, X_i is continuous, and restricts to a linear function on each n-cell) such that

$$X_i(v_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The X_i are called the *Courant functions* of Δ.

b. Show that

$$X_1 + \cdots + X_q = 1,$$

the constant function 1 on Δ.

c. Show that if $\{v_{i_1}, \ldots, v_{i_p}\}$ is any collection of vertices which do *not* form the vertices of any k-cell in Δ, then

$$X_{i_1} \cdot X_{i_2} \cdots X_{i_p} = 0,$$

the constant function 0 on Δ.

d. For a complex Δ with vertices $v_1, \ldots, v_q$, following Stanley and Reisner, we can define the *face ring* of Δ, denoted $\mathbb{R}[\Delta]$, as the quotient ring

$$\mathbb{R}[\Delta] = \mathbb{R}[x_1, \ldots, x_q]/I_\Delta,$$

where I_Δ is the ideal generated by the monomials $x_{i_1} x_{i_2} \cdots x_{i_p}$ corresponding to collections of vertices which are *not* the vertex set of any cell in Δ. Show using part c that there is a ring homomorphism from $\mathbb{R}[\Delta]$ to $\mathbb{R}[X_1, \ldots, X_q]$ (the subalgebra of $C^0(\Delta)$ generated over $\mathbb{R}$ by the Courant functions) obtained by mapping x_i to X_i for each i.

Billera has shown that in fact $C^0(\Delta)$ equals the algebra generated by the Courant functions over $\mathbb{R}$, and that the induced mapping

$$\varphi : \mathbb{R}[\Delta]/\langle x_1 + \cdots + x_q - 1 \rangle \to C^0(\Delta)$$

(see part b) is an isomorphism of $\mathbb{R}$-algebras. See [Bil2].

§4 The Gröbner Fan of an Ideal

Gröbner bases for the same ideal but with respect to different monomial orders have different properties and can look very different. For example, the ideal

$$I = \langle z^2 - x + y - 1, x^2 - yz + x, y^3 - xz + 2\rangle \subset \mathbb{Q}[x, y, z]$$

has the following three Gröbner bases.

1. Consider the *grevlex* order with $x > y > z$. Since the leading terms of the generators of I are pairwise relatively prime,
$$\{z^2 - x + y - 1, x^2 - yz + x, y^3 - xz + 2\}$$
is a monic (reduced) Gröbner basis for I with respect to this monomial order. Note that the basis has three elements.
2. Consider the weight order $>_{\mathbf{w},grevlex}$ on $\mathbb{Q}[x, y, z]$ with $\mathbf{w} = (2, 1, 5)$. This order compares monomials first according to the weight vector $\mathbf{w}$ and breaks ties with the *grevlex* order. The monic Gröbner basis for I with respect to this monomial order has the form:
$$\begin{aligned}\{&xy^3 + y^2 - xy - y + 2x + y^3 + 2, yz - x^2 - x,\\ &y^6 + 4y^3 + yx^2 + 4 - y^4 - 2y, x^2y^2 + 2z + xy - x^2 - x + xy^2,\\ &x^3 - y^4 - 2y + x^2, xz - y^3 - 2, z^2 + y - x - 1\}.\end{aligned}$$
This has seven instead of three elements.
3. Consider the *lex* order with $x > y > z$. The monic Gröbner basis for this ideal is:
$$\begin{aligned}\{&z^{12} - 3z^{10} - 2z^8 + 4z^7 + 6z^6 + 14z^5 - 15z^4 - 17z^3 + z^2 + 9z + 6,\\ &y + \tfrac{1}{38977}(1055z^{11} + 515z^{10} + 42z^9 - 3674z^8 - 12955z^7 + 5285z^6\\ &- 1250z^5 + 36881z^4 + 7905z^3 + 42265z^2 - 63841z - 37186),\\ &x + \tfrac{1}{38977}(1055z^{11} + 515z^{10} + 42z^9 - 3674z^8 - 12955z^7 + 5285z^6\\ &- 1250z^5 + 36881z^4 + 7905z^3 + 3288z^2 - 63841z + 1791)\}\end{aligned}$$
This basis of three elements has the triangular form described by the Shape Lemma (Exercise 16 of Chapter 2, §4).

Many of the applications discussed in this book make crucial use of the different properties of different Gröbner bases. At this point, it is natural to ask the following questions about the collection of *all* Gröbner bases of a fixed ideal I.

- Is the collection of possible Gröbner bases of I finite or infinite?
- When do two different monomial orders yield the same monic (reduced) Gröbner basis for I?

- Is there some geometric structure underlying the collection of Gröbner bases of I that can help to elucidate properties of I?

Answers to these questions are given by the construction of the *Gröbner fan* of an ideal I. A *fan* consists of finitely many closed convex polyhedral cones with vertex at the origin (as defined in §2) with the following properties.

a. Any face of a cone in the fan is also in the fan. (A *face* of a cone σ is $\sigma \cap \{\ell = 0\}$, where $\ell = 0$ is a nontrivial linear equation such that $\ell \geq 0$ on σ. This is analogous to the definition of a face of a polytope.)
b. The intersection of two cones in the fan is a face of each.

These conditions are similar to the definition of the polyhedral complex given in Definition (3.5). The Gröbner fan encodes information about the different Gröbner bases of I and was first introduced in the paper [MR] of Mora and Robbiano. Our presentation is based on theirs.

The first step in this construction is to show that for each fixed ideal I, as $>$ ranges over all possible monomial orders, the collection of monomial ideals $\langle \mathrm{LT}_>(I)\rangle$ is finite. We use the notation

$$\mathrm{Mon}(I) = \{\langle \mathrm{LT}_>(I)\rangle : > \text{ a monomial order}\}.$$

(4.1) Theorem. *For an ideal $I \subset k[x_1, \ldots, x_n]$, the set* $\mathrm{Mon}(I)$ *is finite.*

Proof. Aiming for a contradiction, suppose that $\mathrm{Mon}(I)$ is an infinite set. For each monomial ideal N in $\mathrm{Mon}(I)$, let $>_N$ be any one particular monomial order such that $N = \langle \mathrm{LT}_{>_N}(I)\rangle$. Let Σ be the collection of monomial orders $\{>_N : N \in \mathrm{Mon}(I)\}$. Our assumption implies that Σ is infinite.

By the Hilbert Basis Theorem we have $I = \langle f_1, \ldots, f_s\rangle$ for polynomials $f_i \in k[x_1, \ldots, x_n]$. Since each f_i contains only a finite number of terms, by a pigeonhole principle argument, there exists an infinite subset $\Sigma_1 \subset \Sigma$ such that the leading terms $\mathrm{LT}_>(f_i)$ agree for all $>$ in Σ_1 and all i, $1 \leq i \leq s$. We write N_1 for the monomial ideal $\langle \mathrm{LT}_>(f_1), \ldots, \mathrm{LT}_>(f_s)\rangle$ (taking any monomial order $>$ in Σ_1).

If $F = \{f_1, \ldots, f_s\}$ were a Gröbner basis for I with respect to some $>_1$ in Σ_1, then we claim that F would be a Gröbner basis for I with respect to *every* $>$ in Σ_1. To see this, let $>$ be any element of Σ_1 other than $>_1$, and let $f \in I$ be arbitrary. Dividing f by F using $>$, we obtain

$$f = a_1 f_1 + \cdots + a_s f_s + r, \tag{4.2}$$

where no term in r is divisible by any of the $\mathrm{LT}_>(f_i)$. However, both $>$ and $>_1$ are in Σ_1, so $\mathrm{LT}_>(f_i) = \mathrm{LT}_{>_1}(f_i)$ for all i. Since $r = f - a_1 f_1 - \cdots - a_s f_s \in I$, and F is assumed to be a Gröbner basis for I with respect to $>_1$, this implies that $r = 0$. Since (4.2) was obtained using the division algorithm, $\mathrm{LT}_>(f) = \mathrm{LT}_>(a_i f_i)$ for some i, so $\mathrm{LT}_>(f)$ is divisible by $\mathrm{LT}_>(f_i)$. This shows that F is also a Gröbner basis for I with respect to $>$.

However, this cannot be the case since the original set of monomial orders $\Sigma \supset \Sigma_1$ was chosen so that the monomial ideals $\langle \text{LT}_>(I)\rangle$ for $>$ in Σ were all distinct. Hence, given any $>_1$ in Σ_1, there must be some $f_{s+1} \in I$ such that $\text{LT}_{>_1}(f_{s+1}) \notin \langle \text{LT}_{>_1}(f_1), \ldots, \text{LT}_{>_1}(f_s)\rangle = N_1$. Replacing f_{s+1} by its remainder on division by $f_1, \ldots, f_s$, we may assume in fact that no term in f_{s+1} is divisible by any of the monomial generators for N_1.

Now we apply the pigeonhole principle again to find an infinite subset $\Sigma_2 \subset \Sigma_1$ such that the leading terms of $f_1, \ldots, f_{s+1}$ are the same for all $>$ in Σ_2. Let $N_2 = \langle \text{LT}_>(f_1), \ldots, \text{LT}_>(f_{s+1})\rangle$ for all $>$ in Σ_2, and note that $N_1 \subset N_2$. The argument given in the preceding paragraph shows that $\{f_1, \ldots, f_{s+1}\}$ cannot be a Gröbner basis with respect to any of the monomial orders in Σ_2, so fixing $>_2 \in \Sigma_2$, we find an $f_{s+2} \in I$ such that no term in f_{s+2} is divisible by any of the monomial generators for $N_2 = \langle \text{LT}_{>_2}(f_1), \ldots, \text{LT}_{>_2}(f_{s+1})\rangle$.

Continuing in the same way, we produce a descending chain of infinite subsets $\Sigma \supset \Sigma_1 \supset \Sigma_2 \supset \Sigma_3 \supset \cdots$, and an infinite strictly ascending chain of monomial ideals $N_1 \subset N_2 \subset N_3 \subset \cdots$. This contradicts the ascending chain condition in $k[x_1, \ldots, x_n]$, so the proof is complete. □

We can now answer the first question posed at the start of this section. To obtain a precise result, we introduce some new terminology. It is possible for two monic Gröbner bases of I with respect to different monomial orders to be equal as sets, while the leading terms of some of the basis polynomials are different depending on which order we consider. Examples where I is principal are easy to construct; also see (4.9) below. A *marked Gröbner basis for I* is a set G of polynomials in I, together with an identified leading term in each $g \in G$ such that G is a monic Gröbner basis with respect to some monomial order selecting those leading terms. (More formally, we could define a marked Gröbner basis as a set GM of ordered pairs (g, m) where $\{g : (g, m) \in GM\}$ is a monic Gröbner basis with respect to some order $>$, and $m = \text{LT}_>(g)$ for each (g, m) in GM.) The idea here is that we do not want to build a specific monomial order into the definition of G. It follows from Theorem (4.1) that each ideal in $k[x_1, \ldots, x_n]$ has only finitely many marked Gröbner bases.

(4.3) Corollary. *The set of marked Gröbner bases of I is in one-to-one correspondence with the set* Mon(I).

PROOF. The key point is that if the leading terms of two marked Gröbner bases generate the same monomial ideal, then the Gröbner bases must be equal. The details of the proof are left to the reader as Exercise 4. □

Corollary (4.3) also has the following interesting consequence.

Exercise 1. Show that for any ideal $I \subset k[x_1, \ldots, x_n]$, there exists a finite $U \subset I$ such that U is a Gröbner basis simultaneously for all monomial orders on $k[x_1, \ldots, x_n]$.

A set U as in Exercise 1 is called a *universal Gröbner basis* for I. These were first studied by Weispfenning in [Wei], and that article gives an algorithm for constructing universal Gröbner bases. This topic is also discussed in detail in [Stu2].

To answer our other questions we will represent monomial orders using the matrix orders $>_M$ described in Chapter 1, §2. Recall that if M has rows $\mathbf{w}_i$, then $x^\alpha >_M x^\beta$ if there is an ℓ such that $\alpha \cdot \mathbf{w}_i = \beta \cdot \mathbf{w}_i$ for $i = 1, \ldots, \ell - 1$, but $\alpha \cdot \mathbf{w}_\ell > \beta \cdot \mathbf{w}_\ell$.

When $>_M$ is a matrix order, the first row of M plays a special role and will be denoted $\mathbf{w}$ in what follows. We may assume that $\mathbf{w} \neq 0$.

Exercise 2.

a. Let $>_M$ be a matrix order with first row $\mathbf{w}$. Show that

$$\mathbf{w} \in (\mathbb{R}^n)^+ = \{(a_1, \ldots, a_n) : a_i \geq 0, \text{ all } i\}.$$

We call $(\mathbb{R}^n)^+$ the *positive orthant* in $\mathbb{R}^n$. Hint: $x_i >_M 1$ for all i since $>_M$ is a monomial order.

b. Prove that every nonzero $\mathbf{w} \in (\mathbb{R}^n)^+$ is the first row of some matrix M such that $>_M$ is a monomial order.

c. Let M and M' be matrices such that the matrix orders $>_M$ and $>_{M'}$ are equal. Prove that their first rows satisfy $\mathbf{w} = \lambda \mathbf{w}'$ for some $\lambda > 0$.

Exercise 2 implies that each monomial order determines a well-defined ray in the positive orthant $(\mathbb{R}^n)^+$, though different monomial orders may give the same ray. (For example, all graded orders give the ray consisting of positive multiples of $(1, \ldots, 1)$.) Hence it should not be surprising that our questions lead naturally to cones in the positive orthant.

Now we focus on a single ideal I. Let $G = \{g_1, \ldots, g_t\}$ be one of the finitely many marked Gröbner bases of I, with $\mathrm{LT}(g_i) = x^{\alpha(i)}$, and $N = \langle x^{\alpha(1)}, \ldots, x^{\alpha(t)} \rangle$ the corresponding element of $\mathrm{Mon}(I)$. Our next goal is to understand the set of monomial orders for which G is the corresponding marked Gröbner basis of I. This will answer the second question posed at the start of this section. We write

$$g_i = x^{\alpha(i)} + \sum_\beta c_{i,\beta} x^\beta,$$

where $x^{\alpha(i)} > x^\beta$ whenever $c_{i,\beta} \neq 0$. By the above discussion, each such order $>$ comes from a matrix M, so in particular, to find the leading terms we compare monomials first according to the first row $\mathbf{w}$ of the matrix.

If $\alpha(i) \cdot \mathbf{w} > \beta \cdot \mathbf{w}$ for all β with $c_{i,\beta} \neq 0$, the single weight vector $\mathbf{w}$ selects the correct leading term in g_i as the term of highest weight. As we know, however, we may have a tie in the first comparison, in which case we would have to make further comparisons using the other rows of M. This suggests that we should consider the following set of vectors:

$$\begin{aligned} C_G &= \{\mathbf{w} \in (\mathbb{R}^n)^+ : \alpha(i) \cdot \mathbf{w} \ \geq \ \beta \cdot \mathbf{w} \ \text{ whenever } c_{i,\beta} \neq 0\} \\ &= \{\mathbf{w} \in (\mathbb{R}^n)^+ : (\alpha(i) - \beta) \cdot \mathbf{w} \geq 0 \text{ whenever } c_{i,\beta} \neq 0\}. \end{aligned} \tag{4.4}$$

It is easy to see that C_G is an intersection of closed half-spaces in $\mathbb{R}^n$, hence is a closed convex polyhedral cone contained in the positive orthant. There are many close connections between this discussion and other topics we have considered. For example, we can view the process of finding elements of C_G as finding points in the feasible region of a linear programming problem as in §1 of this chapter. Moreover, given a polynomial, the process of finding its term(s) of maximum weight with respect to a given vector $\mathbf{w}$ is equivalent to an integer programming maximization problem on a feasible region given by the Newton polytope $NP(f)$.

The cone C_G has the property that if $>_M$ is a matrix order such that G is the marked Gröbner basis of I with respect to $>_M$, then the first row $\mathbf{w}$ of M lies in C_G. However, you will see below that the converse can fail, so that the relation between C_G and monomial orders for which G is a marked Gröbner basis is more subtle than meets the eye.

In the following example we determine the cone corresponding to a given marked Gröbner basis for an ideal.

(4.5) Example. Consider the ideal

$$I = \langle x^2 - y, xz - y^2 + yz\rangle \subset \mathbb{Q}[x, y, z]. \tag{4.6}$$

The marked Gröbner basis with respect to the *grevlex* order with $x > y > z$ is

$$G^{(1)} = \{\underline{x^2} - y, \underline{y^2} - xz - yz\},$$

where the leading terms are underlined. Let $\mathbf{w} = (a, b, c)$ be a vector in the positive orthant of $\mathbb{R}^3$. Then $\mathbf{w}$ is in $C_{G^{(1)}}$ if and only if the following inequalities are satisfied:

$$\begin{aligned} (2, 0, 0) \cdot (a, b, c) &\geq (0, 1, 0) \cdot (a, b, c) \quad \text{ or } \quad 2a \geq b \\ (0, 2, 0) \cdot (a, b, c) &\geq (1, 0, 1) \cdot (a, b, c) \quad \text{ or } \quad 2b \geq a + c \\ (0, 2, 0) \cdot (a, b, c) &\geq (0, 1, 1) \cdot (a, b, c) \quad \text{ or } \quad 2b \geq b + c. \end{aligned}$$

To visualize $C_{G^{(1)}}$, slice the positive orthant by the plane $a + b + c = 1$ (every nonzero weight vector in the positive orthant can be scaled to make this true). The above inequalities are pictured in Figure 8.11, where the a-axis, b-axis, and c-axis are indicated by dashed lines and you are looking toward the origin from a point on the ray through $(1, 1, 1)$.

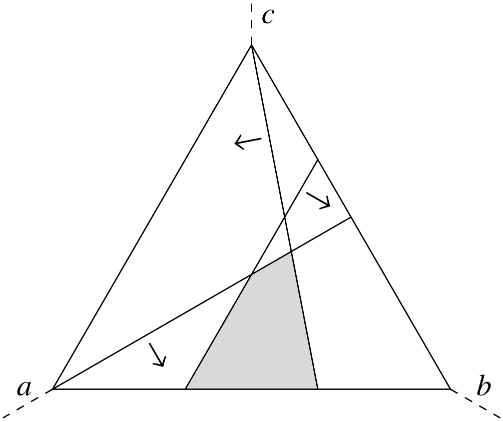

FIGURE 8.11. A slice of the cone $C_{G^{(1)}}$

In this figure, the inequality $2a \geq b$ gives the region in the slice to the left (as indicated by the arrow), the line segment connecting $(0,0,1)$ at the top of the triangle to $(\frac{1}{3}, \frac{2}{3}, 0)$ on the base. The other two inequalities are represented similarly, and their intersection in the first orthant gives the shaded quadrilateral in the slice. Then $C_{G^{(1)}}$ consists of all rays emanating from the origin that go through points of the quadrilateral.

Any weight **w** corresponding to a point in the interior of $C_{G^{(1)}}$ (where the inequalities above are strict) will select the leading terms of elements of $G^{(1)}$ exactly; a weight vector on one of the boundary planes in the interior of the positive orthant will yield a "tie" between terms in one or more Gröbner basis elements. For instance, $(a,b,c) = (1,1,1)$ satisfies $2b = a + c$ and $2b = b + c$, so it is on the boundary of the cone. This weight vector is not sufficient to determine the leading terms of the polynomials.

Now consider a different monomial order, say the *grevlex* order with $z > y > x$. For this order, the monic Gröbner basis for I is

$$G^{(2)} = \{\underline{x^2} - y, \underline{yz} + xz - y^2\},$$

where again the leading terms are underlined. Proceeding as above, the slice of $C_{G^{(2)}}$ in the plane $a + b + c = 1$ is a triangle defined by the inequalities

$$2a \geq b, \ \ b \geq a, \ \ c \geq b.$$

You should draw this triangle carefully and verify that $C_{G^{(1)}} \cap C_{G^{(2)}}$ is a common face of both cones (see also Figure 8.12 below).

Exercise 3. Consider the *grlex* order with $x > y > z$. This order comes from a matrix with $(1, 1, 1)$ as the first row. Let I be the ideal from (4.6).

a. Find the marked Gröbner basis G of I with respect to this order.
b. Identify the corresponding cone C_G and its intersections with the two cones $C_{G^{(1)}}$ and $C_{G^{(2)}}$. Hint: The Gröbner basis polynomials contain more terms than in the example above, but some work can be saved by the observation that if $x^{\beta'}$ divides x^{β} and $\mathbf{w} \in (\mathbb{R}^n)^+$, then $\alpha \cdot \mathbf{w} \geq \beta \cdot \mathbf{w}$ implies $\alpha \cdot \mathbf{w} \geq \beta' \cdot \mathbf{w}$.

Example (4.5) used the *grevlex* order with $z > y > x$, whose matrix has the same first row $(1, 1, 1)$ as the *grlex* order of Exercise 3. Yet they have very different marked Gröbner bases. As we will see in Theorem (4.7) below, this is allowed to happen because the weight vector $(1, 1, 1)$ is on the *boundary* of the cones in question.

Here are some properties of C_G in the general situation.

(4.7) Theorem. *Let I be an ideal in $k[x_1, \ldots, x_n]$, and let G be a marked Gröbner basis of I.*

a. *The interior* $\mathrm{Int}(C_G)$ *of the cone C_G is a nonempty open subset of* $\mathbb{R}^n$.
b. *Let $>_M$ be any matrix order such that the first row of M lies in* $\mathrm{Int}(C_G)$. *Then G is the marked Gröbner basis of I with respect to $>_M$.*
c. *Let G' be a marked Gröbner basis of I different from G. Then the intersection $C_G \cap C_{G'}$ is contained in a boundary hyperplane of C_G, and similarly for $C_{G'}$.*
d. *The union of all the cones C_G, as G ranges over all marked Gröbner bases of I, is the positive orthant* $(\mathbb{R}^n)^+$.

Proof. To prove part a, fix a matrix order $>_M$ such that G is a marked Gröbner basis of I with respect to $>_M$ and let $\mathbf{w}_1, \ldots, \mathbf{w}_m$ be the rows of M. We will show that $\mathrm{Int}(C_G)$ is nonempty by proving that

$$\mathbf{w} = \mathbf{w}_1 + \epsilon\mathbf{w}_2 + \cdots + \epsilon^{m-1}\mathbf{w}_m \in \mathrm{Int}(C_G) \tag{4.8}$$

provided $\epsilon > 0$ is sufficiently small. In Exercise 5, you will show that given exponent vectors α and β, we have

$$x^\alpha >_M x^\beta \Rightarrow \alpha \cdot \mathbf{w} > \beta \cdot \mathbf{w} \text{ provided } \epsilon > 0 \text{ is sufficiently small,}$$

where "sufficiently small" depends on α, β, and M. It follows that we can arrange this for any finite set of pairs of exponent vectors. In particular, since $x^{\alpha(i)} = \mathrm{LT}_{>_M}(x^{\alpha(i)} + \sum_{i,\beta} c_{i,\beta}x^\beta)$, we can pick ϵ so that

$$\alpha(i) \cdot \mathbf{w} > \beta \cdot \mathbf{w} \text{ whenever } c_{i,\beta} \neq 0$$

in the notation of (4.4). Furthermore, using $x_i >_M 1$ for all i, we can also pick ϵ so that $\mathbf{e}_i \cdot \mathbf{w} > 0$ (where $\mathbf{e}_i$ is the ith standard basis vector). It follows that w is in the interior of the positive orthant. From here, $\mathbf{w} \in \mathrm{Int}(C_G)$ follows immediately.

For part b, let $>_M$ be a matrix order such that the first row of M lies in $\mathrm{Int}(C_G)$. This easily implies that for every $g \in G$, $\mathrm{LT}_{>_M}(g)$ is the marked term of g. From here, it is straightforward to show that G is the marked Gröbner basis of I with respect to $>_M$. See Exercise 6 for the details.

We now prove part c. In Exercise 7, you will show that if $C_G \cap C_{G'}$ contains interior points of either cone, then by part a it contains interior points of both cones. If $\mathbf{w}$ is such a point, we take any monomial order $>_M$ defined by a matrix with first row $\mathbf{w}$. Then by part b, G and G' are both the marked Gröbner bases of I with respect to $>_M$. This contradicts our assumption that $G \neq G'$.

Part d follows immediately from part b of Exercise 2. □

With more work, one can strengthen part c of Theorem (4.7) to show that $C_G \cap C_{G'}$ is a face of each (see [MR] or [Stu2] for a proof). It follows that as G ranges over all marked Gröbner bases of I, the collection formed by the cones C_G and their faces is a fan, as defined earlier in the section. This is the *Gröbner fan* of the ideal I.

For example, using the start made in Example (4.5) and Exercise 3, we can determine the Gröbner fan of the ideal I from (4.6). In small examples like this one, a reasonable strategy for producing the Gröbner fan is to find the monic (reduced) Gröbner bases for I with respect to "standard" orders (e.g., *grevlex* and *lex* orders with different permutations of the set of variables) first and determine the corresponding cones. Then if the union of the known cones is not all of the positive orthant, select some $\mathbf{w}$ in the complement, compute the monic Gröbner basis for $>_{\mathbf{w},grevlex}$, find the corresponding cone, and repeat this process until the known cones fill the positive orthant.

For the ideal of (4.6), there are seven cones in all, corresponding to the marked Gröbner bases:

$$\begin{aligned}
G^{(1)} &= \{\underline{x^2} - y, \underline{y^2} - xz - yz\}\\
G^{(2)} &= \{\underline{x^2} - y, \underline{yz} + xz - y^2\}\\
G^{(3)} &= \{\underline{x^4} - x^2z - xz, \underline{y} - x^2\}\\
G^{(4)} &= \{\underline{x^2} - y, \underline{xz} - y^2 + yz, \underline{y^2z} + xy^2 - y^3 - yz\}\\
G^{(5)} &= \{\underline{y^4} - 2y^3z + y^2z^2 - yz^2, \underline{xz} - y^2 + yz,\\
&\qquad \underline{xy^2} - y^3 + y^2z - yz, \underline{x^2} - y\}\\
G^{(6)} &= \{\underline{y^2z^2} - 2y^3z + y^4 - yz^2, \underline{xz} - y^2 + yz,\\
&\qquad \underline{xy^2} - y^3 + y^2z - yz, \underline{x^2} - y\}\\
G^{(7)} &= \{\underline{y} - x^2, \underline{x^2z} - x^4 + xz\}.
\end{aligned} \tag{4.9}$$

(Note that $G^{(5)}$ is the Gröbner basis from Exercise 3.)

Figure 8.12 below shows a picture of the slice of the Gröbner fan in the plane $a+b+c=1$, following the discussion from Example (4.5). The cones are labeled as in (4.9).

For instance, if the Gröbner bases $G^{(1)}, \ldots, G^{(6)}$ in this example are known, the "missing" region of the positive orthant contains (for instance) the vector $\mathbf{w} = (1/10, 2/5, 1/2)$ (see Figure 4.2). Using this weight vector, we find $G^{(7)}$, and the corresponding cone completes the Gröbner fan.

When the number of variables is larger and/or the ideal generators have more terms, this method becomes much less tractable. Mora and Robbiano propose a "parallel Buchberger algorithm" in [MR] which produces the Gröbner fan by considering all potential identifications of leading terms in the computation and reduction of S-polynomials. But their method is certainly not practical on larger examples either. Gröbner fans can be extremely complicated! Fortunately, Gröbner fans are used primarily as conceptual tools—it is rarely necessary to compute large examples.

If we relax our requirement that $\mathbf{w}$ lie in the first orthant and only ask that $\mathbf{w}$ pick out the correct leading terms of a marked Gröbner basis of I, then we can allow weight vectors with negative entries. This leads to a larger "Gröbner fan" denoted $GF(I)$ in [Stu2]. Then the Gröbner fan of Theorem (4.7) (sometimes called the *restricted Gröbner fan*) is obtained by intersecting the cones of $GF(I)$ with the positive orthant. See [MR] and [Stu2] for more about what happens outside the positive orthant.

We close this section with a comment about a closely related topic. In the article [BaM] which appeared at the same time as [MR], Bayer and

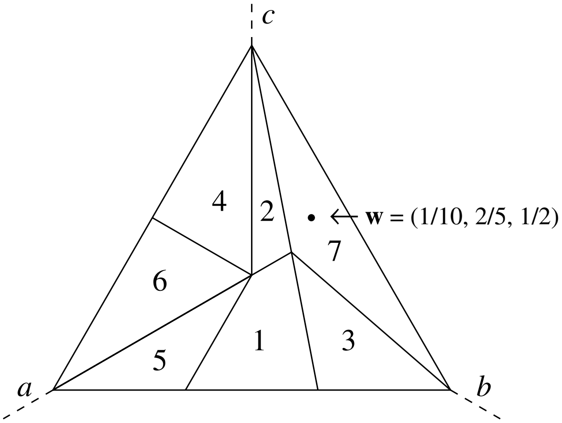

FIGURE 8.12. A slice of the Gröbner fan

Morrison introduced the *state polytope* of a homogeneous ideal. In a sense, this is the dual of the Gröbner fan $GF(I)$ (more precisely, the vertices of the state polytope are in one-to-one correspondence with the elements of $\mathrm{Mon}(I)$, and $GF(I)$ is the normal fan of the state polytope). The state polytope may also be seen as a generalization of the Newton polytope of a single homogeneous polynomial. See [BaM] and [Stu2] for more details.

In the next section, we will see how the Gröbner fan can be used to develop a general Gröbner basis conversion algorithm that, unlike the FGLM algorithm from Chapter 2, does not depend on zero-dimensionality of I.

Additional Exercises for §4

Exercise 4. Using the proof of Proposition (4.1), prove Corollary (4.3).

Exercise 5. Assume that $x^\alpha >_M x^\beta$, where M is an $m \times n$ matrix giving the matrix order $>_M$. Also define $\mathbf{w}$ as in (4.8). Prove that $\alpha \cdot \mathbf{w} > \beta \cdot \mathbf{w}$ provided that $\epsilon > 0$ is sufficiently small.

Exercise 6. Fix a marked Gröbner basis G of an ideal I and let $>$ be a monomial order such that for each $g \in G$, $\textsc{lt}_>(g)$ is the marked term of the polynomial g. Prove that G is the marked Gröbner basis of I with respect to $>$. Hint: Divide $f \in I$ by G using the monomial order $>$.

Exercise 7. Show that if the intersection of two closed, n-dimensional convex polyhedral cones C, C' in $\mathbb{R}^n$ contains interior points of C, then the intersection also contains interior points of C'.

Exercise 8. Verify the computation of the Gröbner fan of the ideal from (4.6) by finding monomial orders corresponding to each of the seven Gröbner bases given in (4.9) and determining the cones $C_{G^{(k)}}$.

Exercise 9. Determine the Gröbner fan of the ideal of the affine twisted cubic curve: $I = \langle y - x^2, z - x^3 \rangle$. Explain why all of the cones have a common one-dimensional edge in this example.

Exercise 10. This exercise will determine which terms in a polynomial $f = \sum_{i=1}^k c_i x^{\alpha(i)}$ can be $\textsc{lt}(f)$ with respect to some monomial order.

a. Show that $x^{\alpha(1)}$ is $\textsc{lt}(f)$ for some monomial order if and only if there is some vector $\mathbf{w}$ in the positive orthant such $(\alpha(1) - \alpha(j)) \cdot \mathbf{w} > 0$ for all $j = 2, \ldots, k$.
b. Show that such a $\mathbf{w}$ exists if and only if the origin is *not* in the convex hull of the set of all $(\alpha(1) - \alpha(j))$ for $j = 2, \ldots, k$, together with the standard basis vectors $\mathbf{e}_i$, $i = 1, \ldots, n$ in $\mathbb{R}^n$.

c. Use the result of part b to determine which terms in $f = x^2yz + 2xyw^2 + x^2w - xw + yzw + y^3$ can be $\mathrm{LT}(f)$ for some monomial order. Determine an order that selects each of the possible leading terms.

Exercise 11. Determine the Gröbner fan of the following ideals:
a. $I = \langle x^3yz^2 - 2xy^3 - yz^3 + y^2z^2 + xyz\rangle$.
b. $I = \langle x - t^4, y - t^2 - t\rangle$.

§5 The Gröbner Walk

One interesting application of the Gröbner fan is a general Gröbner basis conversion algorithm known as the *Gröbner Walk*. As we saw in the discussion of the FGLM algorithm in Chapter 2, to find a Gröbner basis with respect to an "expensive" monomial order such as a *lex* order or another elimination order, it is often simpler to find some other Gröbner basis first, then convert it to a basis with respect to the desired order. The algorithm described in Chapter 2 does this using linear algebra in the quotient algebra $k[x_1, \ldots, x_n]/I$, so it applies only to zero-dimensional ideals.

In this section, we will present the Gröbner Walk introduced by Collart, Kalkbrener, and Mall in [ColKM]. This method converts a Gröbner basis for any ideal $I \subset k[x_1, \ldots, x_n]$ with respect to any one monomial order into a Gröbner basis with respect to any other monomial order. We will also give examples showing how the walk applies to elimination problems encountered in implicitization.

The basic idea of the Gröbner Walk is pleasingly simple. Namely, we assume that we have a marked Gröbner basis G for I, say the marked Gröbner basis with respect to some monomial order $>_s$. We call $>_s$ the *starting order* for the walk, and we will assume that we have some matrix M_s with first row $\mathbf{w}_s$ representing $>_s$. By the results of the previous section, G corresponds to a cone C_G in the Gröbner fan of I.

The goal is to compute a Gröbner basis for I with respect to some other given *target order* $>_t$. This monomial order can be represented by some matrix M_t with first row $\mathbf{w}_t$. Consider a "nice" (e.g., piecewise linear) path from $\mathbf{w}_s$ to $\mathbf{w}_t$ lying completely in the positive orthant in $\mathbb{R}^n$. For instance, since the positive orthant is convex, we could use the straight line segment between the two points, $(1 - u)\mathbf{w}_s + u\mathbf{w}_t$ for $u \in [0, 1]$, though this is not always the best choice. The Gröbner Walk consists of two basic steps:

- Crossing from one cone to the next;
- Computing the Gröbner basis of I corresponding to the new cone.

These steps are done repeatedly until the end of the path is reached, at which point we have the Gröbner basis with respect to the target order. We will discuss each step separately.

Crossing Cones

Assume we have the marked Gröbner basis G_{old} corresponding to the cone C_{old}, and a matrix M_{old} with first row $\mathbf{w}_{old}$ representing $>_{old}$. As we continue along the path from $\mathbf{w}_{old}$, let $\mathbf{w}_{new}$ be the *last point* on the path that lies in the cone C_{old}.

The new weight vector $\mathbf{w}_{new}$ may be computed as follows. Let $G_{old} = \{x^{\alpha(i)} + \sum_{i,\beta} c_{i,\beta} x^{\beta} : 1 \leq i \leq t\}$, where $x^{\alpha(i)}$ is the leading term with respect to $>_{M_{old}}$. To simplify notation, let $v_1, \ldots, v_m$ denote the vectors $\alpha(i) - \beta$ where $1 \leq i \leq t$ and $c_{i,\beta} \neq 0$. By (4.4), C_{old} consists of those points in the positive orthant $(\mathbb{R}^n)^+$ for which

$$\mathbf{w} \cdot v_j \geq 0, \quad 1 \leq j \leq m.$$

For simplicity say that the remaining portion of the path to be traversed consists of the straight line segment from $\mathbf{w}_{old}$ to $\mathbf{w}_t$. Parametrizing this line as $(1-u)\mathbf{w}_{old} + u\mathbf{w}_t$ for $u \in [0,1]$, we see that the point for the parameter value u lies in C_{old} if and only if

$$(5.1) \qquad (1-u)(\mathbf{w}_{old} \cdot v_j) + u(\mathbf{w}_t \cdot v_j) \geq 0, \quad 1 \leq j \leq m.$$

Then $\mathbf{w}_{new} = (1-u_{last})\mathbf{w}_{old} + u_{last}\mathbf{w}_t$, where u_{last} is computed by the following algorithm.

```
         Input: $\mathbf{w}_{old}, \mathbf{w}_t, v_1, \ldots, v_m$
         Output: $u_{last}$
         $u_{last} = 1$
(5.2)    FOR $j = 1, \ldots, m$ DO
             IF $\mathbf{w}_t \cdot v_j < 0$ THEN $u_j := \dfrac{\mathbf{w}_{old} \cdot v_j}{\mathbf{w}_{old} \cdot v_j - \mathbf{w}_t \cdot v_j}$
             IF $u_j < u_{last}$ THEN $u_{last} := u_j$
```

The idea behind (5.2) is that if $\mathbf{w}_t \cdot v_j \geq 0$, then (5.1) holds for all $u \in [0,1]$ since $\mathbf{w}_{old} \cdot v_j \geq 0$. On the other hand, if $\mathbf{w}_t \cdot v_j < 0$, then the formula for u_j gives the largest value of u such that (5.1) holds for this particular j. Note that $0 \leq u_j < 1$ in this case.

Exercise 1. Prove carefully that $\mathbf{w}_{new} = (1-u_{last})\mathbf{w}_{old} + u_{last}\mathbf{w}_t$ is the last point on the path from $\mathbf{w}_{old}$ to $\mathbf{w}_t$ that lies in C_{old}.

Once we have $\mathbf{w}_{new}$, we need to choose the next cone in the Gröbner fan. Let $>_{new}$ be the weight order where we first compare $\mathbf{w}_{new}$-weights and break ties using the *target* order. Since $>_t$ is represented by M_t, it follows that $>_{new}$ is represented by $\binom{\mathbf{w}_{new}}{M_t}$. This gives the new cone C_{new}.

Furthermore, if we are in the situation where M_t is the bottom of the matrix representing $>_{old}$ (which is what happens in the Gröbner Walk), the following lemma shows that whenever $\mathbf{w}_{old} \neq \mathbf{w}_t$, the above process is guaranteed to move us closer to $\mathbf{w}_t$.

(5.3) Lemma. *Let u_{last} be as in (5.2) and assume that $>_{old}$ is represented by $\binom{\mathbf{w}_{old}}{M_t}$. Then $u_{last} > 0$.*

Proof. By (5.2), $u_{last} = 0$ implies that $\mathbf{w}_{old} \cdot v_j = 0$ and $\mathbf{w}_t \cdot v_j < 0$ for some j. But recall that $v_j = \alpha(i) - \beta$ for some $g = x^{\alpha(i)} + \sum_{i,\beta} c_{i,\beta} x^\beta \in G$, where $x^{\alpha(i)}$ is the leading term for $>_{old}$ and $c_{i,\beta} \neq 0$. It follows that

$$\mathbf{w}_{old} \cdot \alpha(i) = \mathbf{w}_{old} \cdot \beta \quad \text{and} \quad \mathbf{w}_t \cdot \alpha(i) < \mathbf{w}_t \cdot \beta. \tag{5.4}$$

Since $>_{old}$ is represented by $\binom{\mathbf{w}_{old}}{M_t}$, the equality in (5.4) tells us that $x^{\alpha(i)}$ and x^β have the same $\mathbf{w}_{old}$-weight, so that we break the tie using M_t. But $\mathbf{w}_t$ is the first row of M_t, so that the inequality in (5.4) implies that $x^{\alpha(i)}$ is not the leading term for $>_{old}$. This contradiction proves the lemma. □

Converting Gröbner Bases

Once we have crossed from C_{old} into C_{new}, we need to convert the marked Gröbner basis G_{old} into a Gröbner basis for I with respect to the monomial order $>_{new}$ represented by $\binom{\mathbf{w}_{new}}{M_t}$. This is done as follows.

The key feature of $\mathbf{w}_{new}$ is that it lies on the boundary of C_{old}, so that some of the inequalities defining C_{old} become equalities. This means that the leading term of some $g \in G_{old}$ has the same $\mathbf{w}_{new}$-weight as some other term in g. In general, given a weight vector $\mathbf{w}$ is the positive orthant $(\mathbb{R}^n)^+$ and a polynomial $f \in k[x_1, \ldots, x_n]$, the *initial form* of f for $\mathbf{w}$, denoted $\mathrm{in}_{\mathbf{w}}(f)$, is the sum of all terms in f of maximum $\mathbf{w}$-weight. Also, given a set S of polynomials, we let $\mathrm{in}_{\mathbf{w}}(S) = \{\mathrm{in}_{\mathbf{w}}(f) : f \in S\}$.

Using this notation, we can form the ideal

$$\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old}) \rangle$$

of $\mathbf{w}_{new}$-initial forms of elements of G_{old}. Note that $\mathbf{w}_{new} \in C_{old}$ guarantees that the marked term of $g \in G_{old}$ appears in $\mathrm{in}_{\mathbf{w}_{new}}(g)$. The important thing to realize here is that in nice cases, $\mathrm{in}_{\mathbf{w}_{new}}(G_{old})$ consists mostly of monomials, together with a small number of polynomials (in the best case, only one binomial together with a collection of monomials).

It follows that finding a monic Gröbner basis

$$H = \{h_1, \ldots, h_s\}$$

of $\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old}) \rangle$ with respect to $>_{new}$ may usually be done very quickly. The surprise is that once we have H, it is relatively easy to convert G_{old} into the desired Gröbner basis.

(5.5) Proposition. *Let G_{old} be the marked Gröbner basis for an ideal I with respect to $>_{old}$. Also let $>_{new}$ be represented by $\binom{\mathbf{w}_{new}}{M_t}$, where $\mathbf{w}_{new}$*

is any weight vector in C_{old}, and let H be the monic Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old})\rangle$ with respect to $>_{new}$ as above. Express each $h_j \in H$ as

$$h_j = \sum_{g\in G_{old}} p_{j,g}\, \mathrm{in}_{\mathbf{w}_{new}}(g). \tag{5.6}$$

Then replacing the initial forms by the g themselves, the polynomials

$$\overline{h}_j = \sum_{g\in G_{old}} p_{j,g}\, g, \quad 1 \le j \le s, \tag{5.7}$$

form a Gröbner basis of I with respect to $>_{new}$.

Before giving the proof, we need some preliminary observations about weight vectors and monomial orders. A polynomial f is $\mathbf{w}$-*homogeneous* if $f = \mathrm{in}_{\mathbf{w}}(f)$. In other words, all terms of f have the same $\mathbf{w}$-weight. Furthermore, every polynomial can be written uniquely as a sum of $\mathbf{w}$-homogeneous polynomials that are its $\mathbf{w}$-*homogeneous components* (see Exercise 5).

We say that a weight vector $\mathbf{w}$ is *compatible* with a monomial order $>$ if $\mathrm{LT}_>(f)$ appears in $\mathrm{in}_{\mathbf{w}}(f)$ for all nonzero polynomials f. Then we have the following result.

(5.8) Lemma. *Fix $\mathbf{w} \in (\mathbb{R}^n)^+ \setminus \{0\}$ and let G be the marked Gröbner basis of an ideal I for a monomial order $>$.*

a. *If $\mathbf{w}$ is compatible with $>$, then $\mathrm{LT}_>(I) = \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(I)) = \mathrm{LT}_>(\langle \mathrm{in}_{\mathbf{w}}(I)\rangle)$.*

b. *If $\mathbf{w} \in C_G$, then $\mathrm{in}_{\mathbf{w}}(G)$ is a Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}}(I)\rangle$ for $>$. In particular,*

$$\langle \mathrm{in}_{\mathbf{w}}(I)\rangle = \langle \mathrm{in}_{\mathbf{w}}(G)\rangle.$$

Proof. For part a, the first equality $\mathrm{LT}_>(I) = \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(I))$ is obvious since the leading term of any $f \in k[x_1, \ldots, x_n]$ appears in $\mathrm{in}_{\mathbf{w}}(f)$. For the second equality, it suffices to show $\mathrm{LT}_>(f) \in \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(I))$ whenever $f \in \langle \mathrm{in}_{\mathbf{w}}(I)\rangle$. Given such an f, write it as

$$f = \sum_{i=1}^{t} p_i\, \mathrm{in}_{\mathbf{w}}(f_i), \quad p_i \in k[x_1, \ldots, x_n],\ f_i \in I.$$

Each side is a sum of $\mathbf{w}$-homogeneous components. Since $\mathrm{in}_{\mathbf{w}}(f_i)$ is already $\mathbf{w}$-homogeneous, this implies that

$$\mathrm{in}_{\mathbf{w}}(f) = \sum_{i=1}^{t} q_i\, \mathrm{in}_{\mathbf{w}}(f_i),$$

where we can assume that q_i is $\mathbf{w}$-homogeneous and f and $q_i f_i$ have the same $\mathbf{w}$-weight for all i. It follows that $\mathrm{in}_{\mathbf{w}}(f) = \mathrm{in}_{\mathbf{w}}(\sum_{i=1}^{t} q_i\, f_i) \in \mathrm{in}_{\mathbf{w}}(I)$. Then compatibility implies $\mathrm{LT}_>(f) = \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(f)) \in \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(I))$.

Turning to part b, first assume that $\mathbf{w}$ is compatible with $>$. Then

$$\langle \mathrm{LT}_>(I)\rangle = \langle \mathrm{LT}_>(G)\rangle = \langle \mathrm{LT}_>(\mathrm{in}_{\mathbf{w}}(G))\rangle,$$

where the first equality follows since G is a Gröbner basis for $>$ and the second follows since $\mathbf{w}$ is compatible with $>$. Combining this with part a, we see that $\langle \textsc{lt}_>(\langle \mathrm{in}_{\mathbf{w}}(I)\rangle)\rangle = \langle \textsc{lt}_>(\mathrm{in}_{\mathbf{w}}(G))\rangle$. Hence $\mathrm{in}_{\mathbf{w}}(G)$ is a Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}}(I)\rangle$ for $>$, and the final assertion of the lemma follows.

It remains to consider what happens when $\mathbf{w} \in C_G$, which does not necessarily imply that $\mathbf{w}$ is compatible with $>$ (see Exercise 6 for an example). Consider the weight order $>'$ which first compares $\mathbf{w}$-weights and breaks ties using $>$. Note that $\mathbf{w}$ is compatible with $>'$.

The key observation is that since $\mathbf{w} \in C_G$, the leading term of each $g \in G$ with respect to $>'$ is the marked term. By Exercise 6 of §4, it follows that G is the marked Gröbner basis of I for $>'$. Since $\mathbf{w}$ is compatible with $>'$, the earlier part of the argument implies that $\mathrm{in}_{\mathbf{w}}(G)$ is a Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}}(I)\rangle$ for $>'$. However, for each $g \in G$, $\mathrm{in}_{\mathbf{w}}(g)$ has the same leading term with respect to $>$ and $>'$. Using Exercise 6 of §4 again, we conclude that $\mathrm{in}_{\mathbf{w}}(G)$ is a Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}}(I)\rangle$ for $>$. □

We can now prove the proposition.

Proof of Proposition (5.5). We will give the proof in three steps. Since $>_{new}$ is represented by $\binom{\mathbf{w}_{new}}{M_t}$, $\mathbf{w}_{new}$ is compatible with $>_{new}$. By part a of Lemma (5.8), we obtain

$$\textsc{lt}_{>_{new}}(I) = \textsc{lt}_{>_{new}}(\langle \mathrm{in}_{\mathbf{w}_{new}}(I)\rangle).$$

The second step is to observe that since $\mathbf{w}_{new} \in C_{old}$, the final assertion of part b of Lemma (5.8) implies

$$\langle \mathrm{in}_{\mathbf{w}_{new}}(I)\rangle = \langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old})\rangle.$$

For the third step, we show that

$$\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old})\rangle = \langle \textsc{lt}_{>_{new}}(H)\rangle = \langle \textsc{lt}_{>_{new}}(\overline{H})\rangle,$$

where $H = \{h_1, \ldots, h_t\}$ is the given Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old})\rangle$ and $\overline{H} = \{\overline{h}_1, \ldots, \overline{h}_t\}$ is the set of polynomials described in the statement of the proposition. The first equality is obvious, and for the second, it suffices to show that for each j, $\textsc{lt}_{>_{new}}(h_j) = \textsc{lt}_{>_{new}}(\overline{h}_j)$. Since the $\mathrm{in}_{\mathbf{w}_{new}}(g)$ are $\mathbf{w}_{new}$-homogeneous, Exercise 7 below shows that the same is true of the h_j and the $q_{j,g}$. Hence for each g, all terms in $q_{j,g}(g - \mathrm{in}_{\mathbf{w}_{new}}(g))$ have smaller $\mathbf{w}_{new}$-weight than those in the initial form. Lifting as in (5.7) to get $\overline{h}_j$ adds only terms with smaller $\mathbf{w}_{new}$-weight. Since $>_{new}$ is compatible with $\mathbf{w}_{new}$, the added terms are also smaller in the new order, so the $>_{new}$-leading term of $\overline{h}_j$ is the same as the leading term of h_j.

Combining the three steps, we obtain

$$\langle \textsc{lt}_{>_{new}}(I)\rangle = \langle \textsc{lt}_{>_{new}}(\overline{H})\rangle.$$

Since $\overline{h}_j \in I$ for all j, we conclude that $\overline{H}$ is a Gröbner basis for I with respect to $>_{new}$, as claimed. □

The Gröbner basis $\overline{H}$ from Proposition (5.5) is minimal, but not necessarily reduced. Hence a complete interreduction is usually necessary to obtain the marked Gröbner basis G_{new} corresponding to the next cone. In practice, this is a relatively quick process.

In order to use Proposition (5.5), we need to find the polynomials $p_{j,g}$ in (5.6) expressing the Gröbner basis elements h_j in terms of the ideal generators of $\mathrm{in}_{\mathbf{w}_{new}}(G_{old})$. This can be done in two ways:

- First, the $p_{j,g}$ can be computed along with H by an extended Buchberger algorithm (see for instance [BW], Chapter 5, Section 6);
- Second, since $\mathrm{in}_{\mathbf{w}_{new}}(G_{old})$ is a Gröbner basis of $\langle \mathrm{in}_{\mathbf{w}_{new}}(G_{old})\rangle$ with respect to $>_{old}$ by part b of Lemma (5.8), the $p_{j,g}$ can be obtained by dividing h_j by $\mathrm{in}_{\mathbf{w}_{new}}(G_{old})$ using $>_{old}$.

In practice, the second is often more convenient to implement. The process of replacing the $\mathbf{w}_{new}$-initial forms of the g by the g themselves to go from (5.6) to (5.7) is called *lifting* the initial forms to the new Gröbner basis.

The Algorithm

The following algorithm is a basic Gröbner Walk, following the straight line segment from $\mathbf{w}_s$ to $\mathbf{w}_t$.

(5.9) Theorem. *Let*

1. **NextCone** *be a procedure that computes u_{last} from (5.2). Recall that $\mathbf{w}_{new} = (1 - u_{last})\mathbf{w}_{old} + u_{last}\mathbf{w}_t$ is the last weight vector along the path that lies in the cone C_{old} of the previous Gröbner basis G_{old};*
2. **Lift** *be a procedure that lifts a Gröbner basis for the $\mathbf{w}_{new}$-initial forms of the previous Gröbner basis G_{old} with respect to $>_{new}$ to the Gröbner basis G_{new} following Proposition (5.5); and*
3. **Interreduce** *be a procedure that takes a given set of polynomials and interreduces them with respect to a given monomial order.*

Then the following algorithm correctly computes a Gröbner basis for I with respect to $>_t$ and terminates in finitely many steps on all inputs:

```
Input: $M_s$ and $M_t$ representing start and target orders with first
        rows $\mathbf{w}_s$ and $\mathbf{w}_t$, $G_s$ = Gröbner basis with respect to $>_{M_s}$
Output: last value of $G_{new}$ = Gröbner basis with respect to $>_{M_t}$
$M_{old} := M_s$
$G_{old} := G_s$
$\mathbf{w}_{new} := \mathbf{w}_s$
```

```
$M_{new} := \begin{pmatrix} \mathbf{w}_{new} \\ M_t \end{pmatrix}$
$done := false$
WHILE $done = false$ DO
    $In := \mathrm{in}_{\mathbf{w}_{new}}(G_{old})$
    $InG := \mathrm{gbasis}(In, >_{M_{new}})$
    $G_{new} := \mathrm{Lift}(InG, G_{old}, In, M_{new}, M_{old})$
    $G_{new} := \mathrm{Interreduce}(G_{new}, M_{new})$
    $u := \mathrm{NextCone}(G_{new}, \mathbf{w}_{new}, \mathbf{w}_t)$
    IF $\mathbf{w}_{new} = \mathbf{w}_t$ THEN
        $done := true$
    ELSE
        $M_{old} := M_{new}$
        $G_{old} := G_{new}$
        $\mathbf{w}_{new} := (1 - u)\mathbf{w}_{new} + u\mathbf{w}_t$
        $M_{new} := \begin{pmatrix} \mathbf{w}_{new} \\ M_t \end{pmatrix}$
RETURN($G_{new}$)
```

Proof. We traverse the line segment from $\mathbf{w}_s$ to $\mathbf{w}_t$. To prove termination, observe that by Corollary (4.3), the Gröbner fan of $I = \langle G_s \rangle$ has only finitely many cones, each of which has only finitely many bounding hyperplanes as in (4.4). Discarding those hyperplanes that contain the line segment from $\mathbf{w}_s$ to $\mathbf{w}_t$, the remaining hyperplanes determine a finite set of distinguished points on our line segment.

Now consider $u_{last} = \mathrm{NextCone}(G_{new}, \mathbf{w}_{new}, \mathbf{w}_t)$ as in the algorithm. This uses (5.2) with $\mathbf{w}_{old}$ replaced by the current value of $\mathbf{w}_{new}$. Furthermore, notice that the monomial order always comes from a matrix of the form $\binom{\mathbf{w}_s}{M_t}$. It follows that the hypothesis of Lemma (5.3) is always satisfied. If $u_{last} = 1$, then the next value of $\mathbf{w}_{new}$ is $\mathbf{w}_t$, so that the algorithm terminates after one more pass through the main loop. On the other hand, if $u_{last} = u_j < 1$, then the next value of $\mathbf{w}_{new}$ lies on the hyperplane $\mathbf{w} \cdot v_j = 0$, which is one of our finitely many hyperplanes. However, (5.2) implies that $\mathbf{w}_t \cdot v_j < 0$ and $\mathbf{w}_{new} \cdot v_j \geq 0$, so that the hyperplane meets the line segment in a single point. Hence the next value of $\mathbf{w}_{new}$ is one of our distinguished points. Furthermore, Lemma (5.3) implies that $u_{last} > 0$, so that if the current $\mathbf{w}_{new}$ differs from $\mathbf{w}_t$, then we must move to a distinguished point farther along the line segment. Hence we must eventually reach $\mathbf{w}_t$, at which point the algorithm terminates.

To prove correctness, observe that in each pass through the main loop, the hypotheses of Proposition (5.5) are satisfied. Furthermore, once the value of $\mathbf{w}_{new}$ reaches $\mathbf{w}_t$, the next pass through the loop computes a Gröbner basis of I for the monomial order represented by $\binom{\mathbf{w}_t}{M_t}$. Using Exercise 6 of §4, it follows that the final value of G_{new} is the marked Gröbner basis for $>_t$. □

The complexity of the Gröbner Walk depends most strongly on the number of cones that are visited along the path through the Gröbner fan, and the number of different cones that contain the point $\mathbf{w}_{new}$ at each step. We will say more about this in the examples below.

Examples

We begin with a simple example of the Gröbner Walk in action. Consider the ideal $I = \langle x^2 - y, xz - y^2 + yz\rangle \subset \mathbb{Q}[x, y, z]$ from (4.6). We computed the full Gröbner fan for I in §4 (see Figure 8.12). Say we know

$$G_s = G^{(1)} = \{\underline{x^2} - y, \underline{y^2} - xz - yz\}$$

from (4.9). This is the Gröbner basis of I with respect to $>_{(5,4,1),grevlex}$ (among many others!). Suppose we want to determine the Gröbner basis with respect to $>_{(6,1,3),lex}$ (which is $G^{(6)}$). We could proceed as follows. Let

$$M_s = \begin{pmatrix} 5 & 4 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

so $\mathbf{w}_s = (5, 4, 1)$. Following Exercise 6 from Chapter 1, §2, we have used a square matrix defining the same order instead of the 4×3 matrix with first row $(5, 4, 1)$ and the next three rows from a 3×3 matrix defining the *grevlex* order (as in part b of Exercise 6 of Chapter 1, §2). Similarly,

$$M_t = \begin{pmatrix} 6 & 1 & 3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and $\mathbf{w}_t = (6, 1, 3)$. We will choose square matrices defining the appropriate monomial orders in all of the following computations by deleting appropriate linearly dependent rows.

We begin by considering the order defined by

$$M_{new} = \begin{pmatrix} 5 & 4 & 1 \\ 6 & 1 & 3 \\ 1 & 0 & 0 \end{pmatrix}$$

(using the weight vector $\mathbf{w}_{new} = (5,4,1)$ first, then refining by the target order). The $\mathbf{w}_{new}$-initial forms of the Gröbner basis polynomials with respect to this order are the same as those for G_s, so the basis does not change in the first pass through the main loop.

We then call the NextCone procedure (5.2) with $\mathbf{w}_{new}$ in place of $\mathbf{w}_{old}$. The cone of $>_{M_{new}}$ is defined by the three inequalities obtained by comparing x^2 vs. y and y^2 vs. xz and yz. By (5.2), u_{last} is the largest u such that $(1-u)(5,4,1)+u(6,1,3)$ lies in this cone and is computed as follows:

$$\begin{aligned}
&x^2 \text{ vs. } y: \\
&\quad v_1 = (2,-1,0), \mathbf{w}_t \cdot v_1 = 6 \geq 0 \Rightarrow u_1 = 1 \\
&y^2 \text{ vs. } xz: \\
&\quad v_2 = (-1,2,-1), \mathbf{w}_t \cdot v_2 = -7 < 0 \Rightarrow u_2 = \tfrac{\mathbf{w}_{new}\cdot v_2}{\mathbf{w}_{new}\cdot v_2 - (-7)} = \tfrac{2}{9} \\
&y^2 \text{ vs. } yz: \\
&\quad v_2 = (0,-,-1), \mathbf{w}_t \cdot v_3 = -2 < 0 \Rightarrow u_3 = \tfrac{\mathbf{w}_{new}\cdot v_3}{\mathbf{w}_{new}\cdot v_3 - (-2)} = \tfrac{3}{5}.
\end{aligned}$$

The smallest u value here is $u_{last} = \frac{2}{9}$. Hence the new weight vector is $\mathbf{w}_{new} = (1 - \frac{2}{9})(5,4,1) + \frac{2}{9}(6,1,3) = (47/9, 10/3, 13/9)$, and M_{old} and

$$M_{new} = \begin{pmatrix} 47/9 & 10/3 & 13/9 \\ 6 & 1 & 3 \\ 1 & 0 & 0 \end{pmatrix}$$

are updated for the next pass through the main loop.

In the second pass, $In = \{y^2 - xz, x^2\}$. We compute the Gröbner basis for $\langle In\rangle$ with respect to $>_{new}$ (with respect to this order, the leading term of the first element is xz), and find

$$H = \{-y^2 + xz, x^2, xy^2, y^4\}.$$

In terms of the generators for $\langle In\rangle$, we have

$$\begin{aligned}
-y^2 + xz &= -1 \cdot (y^2 - xz) + 0 \cdot (x^2) \\
x^2 &= 0 \cdot (y^2 - xz) + 1 \cdot (x^2) \\
xy^2 &= x \cdot (y^2 - xz) + z \cdot (x^2) \\
y^4 &= (y^2 + xz) \cdot (y^2 - xz) + z^2 \cdot (x^2).
\end{aligned}$$

So by Proposition (5.5), to get the next Gröbner basis, we lift to

$$\begin{aligned}
-1 \cdot (y^2 - xz - yz) + 0 \cdot (x^2 - y) &= xz + yz - y^2 \\
0 \cdot (y^2 - xz - yz) + 1 \cdot (x^2 - y) &= x^2 - y \\
x \cdot (y^2 - xz - yz) + z \cdot (x^2 - y) &= xy^2 - xyz - yz \\
(y^2 + xz) \cdot (y^2 - xz - yz) + z^2 \cdot (x^2 - y) &= y^4 - y^3 z - xyz^2 - yz^2.
\end{aligned}$$

Interreducing with respect to $>_{new}$, we obtain the marked Gröbner basis G_{new} given by

$$\{\underline{xz} + yz - y^2, \underline{x^2} - y, \underline{xy^2} - y^3 + y^2z - yz, \underline{y^4} - 2y^3z + y^2z^2 - yz^2\}.$$

(This is $G^{(5)}$ in (4.9).) For the call to NextCone in this pass, we use the parametrization $(1-u)(47/9, 10/3, 13/9) + u(6, 1, 3)$. Using (5.2) as above, we obtain $u_{last} = 17/35$, for which $\mathbf{w}_{new} = (28/5, 11/5, 11/5)$.

In the third pass through the main loop, the Gröbner basis does not change as a set. However, the leading term of the initial form of the last polynomial $y^4 - 2y^3z + y^2z^2 - yz^2$ with respect to $>_{M_{new}}$ is now y^2z^2 since

$$M_{new} = \begin{pmatrix} 28/5 & 11/5 & 11/5 \\ 6 & 1 & 3 \\ 1 & 0 & 0 \end{pmatrix}.$$

Using Proposition (5.5) as usual to compute the new Gröbner basis G_{new}, we obtain

$$(5.10)\quad \{\underline{xz} + yz - y^2, \underline{x^2} - y, \underline{xy^2} - y^3 + y^2z - yz, \underline{y^2z^2} - 2y^3z + y^4 - yz^2\},$$

which is $G^{(6)}$ in (4.9). The call to NextCone returns $u_{last} = 1$, since there are no pairs of terms that attain equal weight for any point on the line segment parametrized by $(1-u)(28/5, 11/5, 11/5) + u(6, 1, 3)$. Thus $\mathbf{w}_{new} = \mathbf{w}_t$. After one more pass through the main loop, during which G_{new} doesn't change, the algorithm terminates. Hence the final output is (5.10), which is the marked Gröbner basis of I with respect to the target order.

We note that it is possible to modify the algorithm of Theorem (5.9) so that the final pass in the above example doesn't occur. See Exercise 8.

Exercise 2. Verify the computation of u_{last} in the steps of the above example after the first.

Exercise 3. Apply the Gröbner Walk to convert the basis $G^{(3)}$ for the above ideal to the basis $G^{(4)}$ (see (4.9) and Figure (4.2)). Take $>_s = >_{(2,7,1),grevlex}$ and $>_t = >_{(3,1,6),grevlex}$.

Many advantages of the walk are lost if there are many terms in the $\mathbf{w}_{new}$-initial forms. This tends to happen if a portion of the path lies in a face of some cone, or if the path passes through points where many cones intersect. Hence in [AGK], Amrhein, Gloor, and Küchlin make systematic use of perturbations of weight vectors to keep the path in as general a position as possible with respect to the faces of the cones. For example, one possible variant of the basic algorithm above would be to use (4.8) to obtain a perturbed weight vector in the interior of the corresponding cone each time a new marked Gröbner basis is obtained, and resume the walk to the target monomial order from there. Another variant designed for elimination problems is to take a "sudden-death" approach. If we want a Gröbner basis with respect to a monomial order eliminating the variables $x_1, \ldots, x_n$, leaving $y_1, \ldots, y_m$, and we expect a single generator for the elimination ideal, then we could terminate the walk as soon as some polynomial in $k[y_1, \ldots, y_m]$ appears in the current G_{new}. This is only guaranteed to be a multiple of the generator of the elimination ideal, but even a polynomial satisfying that condition can be useful in some circumstances. We refer the interested reader to [AGK] for a discussion of other implementation issues.

In [Tran], degree bounds on elements of Gröbner bases are used to produce weight vectors in the interior of each cone of the Gröbner fan, which gives a deterministic way to find good path perturbations. A theoretical study of the complexity of the Gröbner Walk and other basis conversion algorithms has been made by Kalkbrener in [Kal].

Our next example is an application of the Gröbner Walk algorithm to an implicitization problem inspired by examples studied in robotics and computer-aided design. Let C_1 and C_2 be two curves in $\mathbb{R}^3$. The *bisector surface* of C_1 and C_2 is the locus of points P equidistant from C_1 and C_2 (that is, P is on the bisector if the closest point(s) to P on C_1 and C_2 are the same distance from P). See, for instance, [EK]. Bisectors are used, for example, in motion planning to find paths avoiding obstacles in an environment. We will consider only the case where C_1 and C_2 are smooth complete intersection algebraic curves $C_1 = \mathbf{V}(f_1, g_1)$ and $C_2 = \mathbf{V}(f_2, g_2)$. (This includes most of the cases of interest in solid modeling, such as lines, circles, and other conics, etc.) $P = (x, y, z)$ is on the bisector of C_1 and C_2 if there exist $Q_1 = (x_1, y_1, z_1) \in C_1$ and $Q_2 = (x_2, y_2, z_2) \in C_2$ such that the distance from P to C_i is a minimum at Q_i, $i = 1, 2$ and the distance from P to Q_1 equals the distance from P to Q_2. Rather than insisting on an absolute minimum of the distance function from P to C_i at Q_i, it is simpler to insist that the distance function simply have a critical point there. It is easy to see that this condition is equivalent to saying that the line segment from P to Q_i is *orthogonal* to the tangent line to C_i at Q_i.

Exercise 4. Show that the distance from C_i to P has a critical point at Q_i if and only if the line segment from P to Q_i is *orthogonal* to the tangent line to C_i at Q_i, and show that this is equivalent to saying that

$$(\nabla f_i(Q_i) \times \nabla g_i(Q_i)) \cdot (P - Q_i) = 0,$$

where $\nabla f_i(Q_i)$ denotes the gradient vector of f_i at Q_i, and $\times$ is the cross product in $\mathbb{R}^3$.

By Exercise 4, we can find the bisector as follows. Let (x_i, y_i, z_i) be a general point Q_i on C_i, and $P = (x, y, z)$. Consider the system of equations

$$\begin{aligned}
0 &= f_1(x_1, y_1, z_1)\\
0 &= g_1(x_1, y_1, z_1)\\
0 &= f_2(x_2, y_2, z_2)
\end{aligned}$$

$$\begin{aligned}
0 &= g_2(x_2, y_2, z_2)\\
0 &= (\nabla f_1(x_1, y_1, z_1) \times \nabla g_1(x_1, y_1, z_1)) \cdot (x - x_1, y - y_1, z - z_1)\\
0 &= (\nabla f_2(x_2, y_2, z_2) \times \nabla g_2(x_2, y_2, z_2)) \cdot (x - x_2, y - y_2, z - z_2)\\
0 &= (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2\\
&\quad - (x - x_2)^2 - (y - y_2)^2 - (z - z_2)^2.
\end{aligned} \tag{5.11}$$

Let $J \subset \mathbb{R}[x_1, y_1, z_1, x_2, y_2, z_2, x, y, z]$ be the ideal generated by these seven polynomials. We claim the bisector will be contained in $\mathbf{V}(I)$, where I is the elimination ideal $I = J \cap \mathbb{R}[x, y, z]$. A proof proceeds as follows. $P = (x, y, z)$ is on the bisector of C_1 and C_2 if and only if there exist $Q_i = (x_i, y_i, z_i)$ such that $Q_i \in C_i$, Q_i is a minimum of the distance function to P, restricted to C_i, and $PQ_1 = PQ_2$. Thus P is in the bisector if and only if the equations in (5.11) are satisfied for some $(x_i, y_i, z_i) \in C_i$. Therefore, P is the projection of some point in $\mathbf{V}(J)$, hence in $\mathbf{V}(I)$. Note that (5.11) contains seven equations in nine unknowns, so we expect that $\mathbf{V}(J)$ and its projection $\mathbf{V}(I)$ have dimension 2 in general.

For instance, if C_1 is the twisted cubic $\mathbf{V}(y - x^2, z - x^3)$ and C_2 is the line $\mathbf{V}(x, y - 1)$, then our ideal J is

$$\begin{aligned}
J = \langle y_1 - x_1^2, z_1 - x_1^3, x_2, y_2 - 1,&\\
x - x_1 + 2x_1(y - y_1) + 3x_1^2(z - z_1), z - z_2,&\\
(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2&\\
- (x - x_2)^2 - (y - y_2)^2 - (z - z_2)^2\rangle.&
\end{aligned} \tag{5.12}$$

We apply the Gröbner Walk with $>_s$ the *grevlex* order with $x_1 > y_1 > z_1 > x_2 > y_2 > z_2 > x > y > z$, and $>_t$ the $>_{\mathbf{w},grevlex}$ order, where $\mathbf{w} = (1, 1, 1, 1, 1, 1, 0, 0, 0)$, which has the desired elimination property to compute $J \cap \mathbb{R}[x, y, z]$.

Using our own (somewhat naive) implementation of the Gröbner Walk based on the `Groebner` package in Maple, we computed the $>_{\mathbf{w},grevlex}$ basis for J as in (5.12). As we expect, the elimination ideal is generated by a single polynomial: $J \cap \mathbb{R}[x, y, z] =$

$$\begin{aligned}
&\langle 5832z^6y^3 - 729z^8 - 34992x^2y - 14496yxz - 14328x^2z^2\\
&+ 24500x^4y^2 - 23300x^4y + 3125x^6 - 5464z^2 - 36356z^4y\\
&+ 1640xz^3 + 4408z^4 + 63456y^3xz^3 + 28752y^3x^2z^2\\
&- 201984y^3 - 16524z^6y^2 - 175072y^2z^2 + 42240y^4xz - 92672y^3zx\\
&+ 99956z^4y^2 + 50016yz^2 + 90368y^2 + 4712x^2 + 3200y^3x^3z\\
&+ 6912y^4xz^3 + 13824y^5zx + 19440z^5xy^2 + 15660z^3x^3y + 972z^4x^2y^2\\
&+ 6750z^2x^4y - 61696y^2z^3x + 4644yxz^5 - 37260yz^4x^2\\
&- 85992y^2x^2z^2 + 5552x^4 - 7134xz^5 + 64464yz^2x^2
\end{aligned}$$

$$
\begin{aligned}
&- 5384zyx^3 + 2960zy^2x^3 - 151z^6 + 1936\\
&+ 29696y^6 + 7074z^6y + 18381z^4x^2 - 2175z^2x^4 + 4374xz^7\\
&+ 1120zx - 7844x^3z^3 - 139264y^5 - 2048y^7 - 1024y^6z^2\\
&- 512y^5x^2 - 119104y^3x^2 - 210432y^4z^2 + 48896y^5z^2\\
&- 104224y^3z^4 + 28944y^4z^4 + 54912y^4x^2 - 20768y + 5832z^5x^3\\
&- 8748z^6x^2 + 97024y^2x^2 + 58560y^2zx + 240128y^4 + 286912y^3z^2\\
&+ 10840xyz^3 + 1552x^3z - 3750zx^5\rangle.
\end{aligned}
$$

The computation of the full Gröbner basis (including the initial computation of the *grevlex* Gröbner basis of J) took 43 seconds on a 866 MHz Pentium III using the Gröbner Walk algorithm described in Theorem (5.9). Apparently the cones corresponding to the two monomial orders $>_s, >_t$ are very close together in the Gröbner fan for J, a happy accident. The $\mathbf{w}_{new}$-initial forms in the second step of the walk contained a large number of distinct terms, though. With the "sudden-death" strategy discussed above, the time was reduced to 23 seconds and produced the same polynomial (not a multiple). By way of contrast, a direct computation of the $>_{\mathbf{w},grevlex}$ Gröbner basis using the `gbasis` command of the `Groebner` package was terminated after using 20 minutes of CPU time and over 200 Mb of memory. In our experience, in addition to gains in speed, the Gröbner Walk tends also to use much less memory for storing intermediate polynomials than Buchberger's algorithm with an elimination order. This means that even if the walk takes a long time to complete, it will often execute successfully on complicated examples that are not feasible using the Gröbner basis packages of standard computer algebra systems. Similarly encouraging results have been reported from several experimental implementations of the Gröbner Walk.

As of this writing, the Gröbner Walk has not been included in the Gröbner basis packages distributed with general-purpose computer algebra systems such as Maple or *Mathematica*. An implementation is available in Magma, however. The CASA Maple package developed at RISC-Linz (see `http://www.risc.uni-linz.ac.at/software/casa/`) also contains a Gröbner Walk procedure.

ADDITIONAL EXERCISES FOR §5

Exercise 5. Fix a nonzero weight vector $\mathbf{w} \in (\mathbb{R}^n)^+$. Show that every $f \in k[x_1, \dots, x_n]$ can be written uniquely as a sum of $\mathbf{w}$-homogeneous polynomials.

Exercise 6. Fix a monomial order $>$ and a nonzero weight vector $\mathbf{w} \in (\mathbb{R}^n)^+$. Also, given an ideal $I \subset k[x_1, \dots, x_n]$, let $C_>$ be the cone in the Gröbner fan of I corresponding to $\langle \textsc{lt}_>(I)\rangle \in \mathrm{Mon}(I)$.

a. Prove that $\mathbf{w}$ is compatible with $>$ if and only if $\mathbf{w}\cdot\alpha > \mathbf{w}\cdot\beta$ always implies $x^\alpha > x^\beta$.
b. Prove that if $\mathbf{w}$ is compatible with $>$, then $\mathbf{w} \in C_>$.
c. Use the example of $>_{lex}$ for $x > y$, $I = \langle x+y\rangle \subset k[x,y]$ and $\mathbf{w} = (1,1)$ to show that the naive converse to part b is false. (See part d for the real converse.)
d. Prove that $\mathbf{w} \in C_>$ if and only if there is a monomial order $>'$ such that $C_{>'} = C_>$ and $\mathbf{w}$ is compatible with $>'$. Hint: See the proof of part b of Lemma (5.8).

Exercise 7. Suppose that J is an ideal generated by $\mathbf{w}$-homogeneous polynomials. Show that every reduced Gröbner basis of I consists of $\mathbf{w}$-homogeneous polynomials. Hint: This generalizes the corresponding fact for homogeneous ideals. See [CLO], Theorem 2 of Chapter 8, §3.

Exercise 8. It is possible to get a slightly more efficient version of the algorithm described in Theorem (5.9). The idea is to modify (5.2) so that u_{last} is allowed to be greater than 1 if the ray from $\mathbf{w}_{old}$ to $\mathbf{w}_t$ leaves the cone at a point beyond $\mathbf{w}_t$.
a. Modify (5.2) so that it behaves as described above and prove that your modification behaves as claimed.
b. Modify the algorithm described in Theorem (5.9) in two ways: first, $\mathbf{w}_{new}$ is defined using $\min\{1, u_{last}\}$ and second, the IF statement tests whether $u_{last} > 1$ or $\mathbf{w}_{new} = \mathbf{w}_t$. Prove that this modified algorithm correctly converts G_s to G_t.
c. Show that the modified algorithm, when applied to the ideal $I = \langle x^2 - y, y^2 - xz - yz\rangle$ discussed in the text, requires one less pass through the main loop than without the modificiation.

Exercise 9. In a typical polynomial implicitization problem, we are given $f_i \in k[t_1,\ldots,t_m]$, $i = 1,\ldots,n$ (the coordinate functions of a parametrization) and we want to eliminate $t_1,\ldots,t_m$ from the equations $x_i = f_1(t_1,\ldots,t_m)$, $i = 1,\ldots,n$. To do this, consider the ideal

$$J = \langle x_1 - f_1(t_1,\ldots,t_m),\ldots,x_n - f_n(t_1,\ldots,t_m)\rangle$$

and compute $I = J \cap k[x_1,\ldots,x_n]$ to find the implicit equations of the image of the parametrization. Explain how the Gröbner Walk could be applied to the generators of J directly to find I without any preliminary Gröbner basis computation. Hint: They are already a Gröbner basis with respect to a suitable monomial order.

Exercise 10. Apply the Gröbner Walk method suggested in Exercise 9 to compute the implicit equation of the parametric curve

$$\begin{cases} x = t^4 \\ y = t^2 + t. \end{cases}$$

(If you do not have access to an implementation of the walk, you will need to perform the steps “manually” as in the example given in the text.) Also see part b of Exercise 11 in the previous section.