2

Affine monoids and their Hilbert bases

Affine monoids are the basic structure on which algebras with coefficients in rings will be built later on. Their finiteness properties allow a rich structure theory, both from the combinatorial and the ring theoretic point of view to be pursued in later chapters. An affine monoid defines a cone in a natural way, and therefore the notions of polyhedral convex geometry will be omnipresent in this chapter.

Within the class of affine monoids, the normal ones represent the discrete counterparts to continuous cones since they can be recovered from their cones and the lattice determining the discrete structure. Part of the theory of general affine monoids will be developed by relating them to their normalizations. By Gordan's lemma the normalizations are also affine.

The last two sections deal with the combinatorics of the Hilbert bases of positive normal affine monoids. On the one hand, we will try to bound the degrees of Hilbert basis elements, and on the other we will investigate whether normality can be explained by conditions on the unimodularity of the Hilbert basis.

2.A Affine monoids

In common usage, a monoid is a set M together with an operation $M \times M \to M$ that is associative and has a neutral element. We are mainly interested in a special class of commutative monoids:

Definition 2.1. A monoid is *affine* if it is finitely generated and isomorphic to a submonoid of a free abelian group $\mathbb{Z}^d$ for some $d \geq 0$.

Very often, especially in the commutative algebra literature, affine monoids are called *affine semigroups*—a line of tradition followed in our joint papers. The usage of "monoids" in this book is more compatible with another tradition, that in the K-theoretic literature.

In view of the definition above it is appropriate to use additive notation for the operation in M. The condition on finite generation then just means that there exist $x_1, \dots, x_n \in M$ for which

W. Bruns, J. Gubeladze, *Polytopes, Rings, and K-Thoery*, Springer Monographs in Mathematics,
DOI 10.1007/b105283_2,

$$M = \mathbb{Z}_+x_1 + \cdots + \mathbb{Z}_+x_n = \{a_1x_1 + \cdots + a_nx_n : a_i \in \mathbb{Z}_+\}.$$

We are of course always free to consider an affine monoid as a submonoid of $\mathbb{Z}^d$ for suitable d.

Within the class of commutative monoids, the affine monoids are characterized by being (i) finitely generated, (ii) *cancellative*, and (iii) *torsionfree*.

Cancellativity means that an equation $x + y = x + z$ for $x, y, z \in M$ implies $y = z$. Equivalently, M can be embedded into a group.

More generally, for every commutative monoid M there exist a group $\mathrm{gp}(M)$ and a monoid homomorphism $\iota : M \to \mathrm{gp}(M)$ solving the following universal problem: every monoid homomorphism $\varphi : M \to H$ from M to a group H factors in a unique way as $\varphi = \psi \circ \iota$ with a group homomorphism $\psi : \mathrm{gp}(M) \to H$. Consequently the *group of differences* $\mathrm{gp}(M)$ is unique up to isomorphism. It is constructed as follows. As a set, $\mathrm{gp}(M)$ consists of the equivalence classes $x - y$ of pairs $(x, y) \in M^2$ with $x - y = u - v$ if and only if $x + v + z = u + y + z$ for some $z \in M$. Addition is defined by $(x - y) + (u - v) = (x + u) - (y + v)$. Then $\mathrm{gp}(M)$ is evidently a group, and the map $\iota : M \mapsto \mathrm{gp}(M)$, $\iota(x) = x - 0$, is a monoid homomorphism that satisfies the universality condition. Clearly, when M is cancellative, then ι is injective.

To be torsionfree for a monoid M means that $ax = ay$ for $a \in \mathbb{N}$ and $x, y \in M$ implies $x = y$. If M is cancellative, this condition is equivalent to the torsion freeness of $\mathrm{gp}(M)$, but in general it is not enough to require torsion freeness for $\mathrm{gp}(M)$ (Exercise 2.1).

We agree on the following

Convention 2.2. In the remainder of this book the term *monoid* always means a commutative, cancellative and torsionfree monoid, unless explicitly stated otherwise.

We see that if a monoid M is finitely generated, then it can be embedded into a group which is finitely generated and torsionfree, i. e., isomorphic to a free abelian group $\mathbb{Z}^r$. If M is given as a submonoid of $\mathbb{Z}^d$, then $\mathrm{gp}(M)$ can be identified with the subgroup $\mathbb{Z}M$ of $\mathbb{Z}^d$ generated by M. The monoids that are isomorphic to $\mathbb{Z}_+^r$, $r \in \mathbb{Z}_+$, are called *free monoids*.

Definition 2.3. The *rank* of a monoid M is the vector space dimension of $\mathbb{Q} \otimes \mathrm{gp}(M)$ over $\mathbb{Q}$. In other words, it is the rank of the abelian group $\mathrm{gp}(M)$.

Clearly, if M is affine and $\mathrm{gp}(M) \cong \mathbb{Z}^r$, then $\operatorname{rank} M = r$. However, note that the definition of rank is not restricted to finitely generated monoids.

Remark 2.4. Every submonoid of $\mathbb{Z}$ is finitely generated (Exercise 2.2). It is isomorphic to a submonoid of $\mathbb{Z}_+$, unless it is a subgroup of $\mathbb{Z}$. Submonoids of $\mathbb{Z}_+$ are often called *numerical semigroups*. A vast amount of research has been devoted to them. See Barucci, Dobbs, and Fontana [16].

In contrast, already $\mathbb{Z}^2$ contains submonoids without a finite system of generators, for example $\{(0,0)\} \cup \{(x, y) : x \geq 1\}$.

Subcones C of $\mathbb{R}^d$ are examples of “continuous” monoids. Unless $C = 0$, such a monoid is not finitely generated. Neither is $C \cap \mathbb{Q}^d$ finitely generated if it contains a nonzero vector.

As in ring theory, it is useful to introduce the notion of module.

Definition 2.5. A set N with an (additively written) operation $M \times N \to N$ is called an *M-module* if $(a + b) + x = a + (b + x)$ and $0 + x = x$ for all $a, b \in M$ and $x \in N$.

A subset I of M is an *ideal* if it is a submodule, i. e., if $M + I \subset I$.

A typical example of a module over $M \subset \mathbb{Z}^d$ is $\mathbb{Z}^d$ itself. Then every subset U of $\mathbb{Z}^d$ with $M + U \subset U$ is a submodule. The empty set is considered an M-module (and an ideal).

Remark 2.6. (a) Let $M \subset \mathbb{Z}^d$ be an affine monoid. Then we define the *interior* of M by $\operatorname{int}(M) = M \cap \operatorname{int}(\mathbb{R}_+M)$. Since $x + y \in \operatorname{int}(\mathbb{R}_+M)$ for $x \in \operatorname{int}(\mathbb{R}_+M)$, $y \in \mathbb{R}_+M$, it follows that $\operatorname{int}(M)$ is an ideal.

One has $0 \in \operatorname{int}(M)$ if and only if M is a group. In that case $\operatorname{int}(M) = M$. In all the other cases $\operatorname{int}(M)$ is not a monoid itself. However, $M_* = \operatorname{int}(M) \cup \{0\}$ is a submonoid of M that will play an important role later on. Note that $M_* = M$ if and only if $\operatorname{rank} M \le 1$ or $M = \operatorname{int}(M)$. Otherwise M_* is not even finitely generated, as the reader may show (Exercise 2.4).

(b) In (a) we have assumed that $M \subset \mathbb{Z}^d$. Nevertheless $\operatorname{int}(M)$ is defined intrinsically in terms of M and does not depend on an embedding of M into $\mathbb{Z}^d$ (or $\mathbb{R}^d$) for some d. The reason is that an isomorphism of monoids $M \subset \mathbb{R}^d$ and $N \subset \mathbb{R}^e$ induces isomorphisms (i) $\operatorname{gp}(M) \cong \operatorname{gp}(N)$, (ii) $\mathbb{R}M \cong \mathbb{R}N$, and (iii) $\mathbb{R}_+M \cong \mathbb{R}_+N$. For a coordinate-free definition one views M as a submonoid of the vector space $\mathbb{R} \otimes_{\mathbb{Z}} \operatorname{gp}(M)$.

Alternatively, one can give intrinsic definitions using only the monoid structure; see Exercise 2.3.

(c) Let M be an affine monoid and F a face of the cone $\mathbb{R}_+M$. Then $F \cap M$, the intersection of M and the monoid F, is a submonoid of M. The submonoids of M that arise in this way are called *extreme submonoids*.

For an arbitrary monoid M and every commutative ring of coefficients R we can form the *monoid algebra* (or *monoid ring*) $R[M]$—the main vehicle in our study of the interactions between discrete geometry, commutative ring theory, and algebraic K-theory. As an R-module $R[M]$ is free with a basis consisting of the symbols X^a, $a \in M$, and the multiplication on $R[M]$ is defined by the R-bilinear extension of $X^aX^b = X^{a+b}$.[1] The elements X^a are called the *monomials* of $R[M]$.

For an M-module N we can analogously define an $R[M]$-module RN that as an R-module is free on the basis X^n, $n \in N$, and on which X^a operates by $X^aX^n = X^{a+n}$. We leave it to the reader to give the proofs of the statements implicitly contained in the descriptions of $R[M]$ and RN.

Proposition 2.7. *Let M be a monoid, N an M-module, and R a ring. Then*

(a) M is finitely generated if and only if $R[M]$ is a finitely generated R-algebra;

[1] Later on, when the use of monoid rings becomes essential, we will switch to a simpler notation identifying the elements $m \in M$ with X^m and writing the monoid operation multiplicatively.

(b) *N is a finitely generated M-module if and only if RN is a finitely generated $R[M]$-module.*

Proof. (a) The implication $\Longrightarrow$ is trivial. For the converse let $f_1, \dots, f_n$ be a system of generators of $R[M]$. There exists a finite subset E of M such that every f_i is an R-linear combination of the elements X^e with $e \in E$. Let $M' = \mathbb{Z}_+ E$. It follows immediately that any R-linear combination of the products $f_1^{a_1} \cdots f_n^{a_n}$ with $a_1, \dots, a_n \in \mathbb{Z}_+^n$ is an R-linear combination of monomials X^b, $b \in M'$. Since $f_1, \dots, f_n$ generate $R[M]$, it follows that $X^a \in R[M']$ for every $a \in M$. This implies $M = M'$.

(b) is proved along the same lines. □

In using Proposition 2.7 for proving assertions about M or N we are free to choose R; for example, we can take R to be a field $\Bbbk$.

Proposition 2.8. *Let M be a finitely generated monoid and N a finitely generated M-module. Then every M-submodule of N is finitely generated.*

Proof. We choose a field $\Bbbk$. Then $\Bbbk[M]$ is a finitely generated $\Bbbk$-algebra and therefore a noetherian ring by Hilbert's basis theorem (see Eisenbud [112, Sect 1.4]). So every submodule of the finitely generated $\Bbbk[M]$-module $\Bbbk N$ is finitely generated. It follows that $\Bbbk U$ is finitely generated for every M-submodule U of X, whence U is finitely generated over M by the previous proposition. □

Gordan's lemma. Very often we will have to round integers:

$$\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \le x\}, \qquad x \in \mathbb{R},$$

denotes the *floor* of x, and

$$\lceil x \rceil = \min\{z \in \mathbb{Z} : z \ge x\}, \qquad x \in \mathbb{R},$$

denotes the *ceiling* of x. For $y \in \mathbb{R}^n$ we set

$$\lfloor y \rfloor = (\lfloor y_1 \rfloor, \dots, \lfloor y_n \rfloor) \quad \text{and} \quad \lceil y \rceil = (\lceil y_1 \rceil, \dots, \lceil y_n \rceil).$$

The first opportunity to use $\lfloor x \rfloor$ is the proof of *Gordan's lemma*:

Lemma 2.9. *Let C be a rational cone in $\mathbb{R}^d$, and $L \subset \mathbb{Q}^d$ a lattice. Then $C \cap L$ is an affine monoid.*

Proof. Set $C' = C \cap \mathbb{R}L$. Then C' is a rational cone as well. Moreover every element x of $\mathbb{Q}^d \cap \mathbb{R}L$ is a rational linear combination of elements of L and so there exists a positive integer a with $ax \in L$. We choose a finite system of generators $x_1, \dots, x_n$ of C'. As just seen, we can assume that $x_1, \dots, x_n \in L$. Let M' be the affine monoid generated by $x_1, \dots, x_n$.

Every element x of $C' \cap L$ has a representation $x = a_1 x_1 + \dots + a_n x_n$ with $a_i \in \mathbb{R}_+$. (We can choose $a_1, \dots, a_n \in \mathbb{Q}_+$, but this is irrelevant here.) Therefore

$$x = \big(\lfloor a_1 \rfloor x_1 + \cdots + \lfloor a_n \rfloor x_n\big) + \big(q_1 x_1 + \cdots + q_n x_n\big),$$
$$0 \le q_i = a_i - \lfloor a_i \rfloor < 1, \; i = 1, \ldots, n.$$

The first summand on the right hand side is in M', the second is an element of $C' \cap L$ that belongs to a bounded subset B of $\mathbb{R}^n$. It follows that $C' \cap L$ is generated as an M'-module by the finite set $B \cap C' \cap L$. Being a finitely generated module over an affine monoid, the monoid $C \cap L$ is itself finitely generated. □

Another proof of Gordan's lemma will be given in the framework of graded rings; see Lemma 4.12.

The essential point in the hypothesis of Lemma 2.9 is the existence of a lattice L' containing both L and the generators of C: after a change of coordinates we can assume that $L' \subset \mathbb{Z}^d$; then C is rational, and we have arrived at the hypothesis of Lemma 2.9.

The reader may show that the intersection of $\mathbb{Z}^2$ with the cone generated by $(1, 0)$ and $(1, a)$ is a finitely generated monoid if and only if a is rational (Exercise 2.6).

Corollary 2.10. *Let M be a submonoid of $\mathbb{R}^d$, L a lattice in $\mathbb{R}^d$ containing M, and $C = \mathbb{R}_+ M$. Then the following are equivalent:*

(a) M is an affine monoid;
(b) $\widehat{M}_L = C \cap L$ is an affine monoid;
(c) C is a cone.

Moreover, if M is affine, then $\widehat{M}_L$ is a finitely generated M-module.

Proof. It is obvious that both (a) and (b) imply (c), The implication (c) $\Longrightarrow$ (b) is essentially Gordan's lemma. To complete the proof of the equivalence by the implication (c) $\Longrightarrow$ (a) one notes that the affine monoid M' in the proof of the lemma can be chosen as a submonoid of M if the cone under consideration is generated by M. Thus M is a submodule of the finitely generated M'-module $\widehat{M}_L$, and therefore itself finitely generated by 2.8.

Finally, since $\widehat{M}_L$ is finitely generated over M', it is even more so over M. □

The monoids of type $\widehat{M}_L$ will be further investigated in the next subsection.

Corollary 2.11. *Let M and N be affine submonoids of $\mathbb{R}^d$, and let C be a cone generated by elements of* $\operatorname{gp}(M)$*. Then*

(a) $M \cap N$ is an affine monoid;
(b) $M \cap C$ is an affine monoid;
(c) the extreme submonoids of M are affine.

Proof. (a) The group generated by $M \cup N$ is a finitely generated torsionfree abelian group and therefore isomorphic to $\mathbb{Z}^n$ for some n (possibly, $n > d$). Replacing $\mathbb{R}^d$ by $\mathbb{R}^n$, we may, right from the start, assume that M and N are submonoids of $\mathbb{Z}^d$. The conical set $\mathbb{R}_+ M \cap \mathbb{R}_+ N$ is the intersection of the rational cones $\mathbb{R}_+ M$ and $\mathbb{R}_+ N$. Such an intersection is itself a rational cone by Proposition 1.69. So the claim follows from the previous corollary once we have shown that $\mathbb{R}_+ M \cap \mathbb{R}_+ N = \mathbb{R}_+(M \cap N)$.

The inclusion $\mathbb{R}_+(M \cap N) \subset \mathbb{R}_+M \cap \mathbb{R}_+N$ is trivial. For the converse it is enough to consider rational elements $x \in \mathbb{R}_+M \cap \mathbb{R}_+N$. Then there exist $\alpha, \beta \in \mathbb{Z}$, $\alpha, \beta > 0$, such that $\alpha x \in M$ and $\beta x \in N$. Thus $\alpha\beta x \in M \cap N$, and $x \in \mathbb{R}_+(M \cap N)$.

(b) By Corollary 2.9 $\operatorname{gp}(M) \cap C$ is an affine monoid. Therefore $M \cap C = M \cap \operatorname{gp}(M) \cap C$ is affine by (a).

(c) is an immediate consequence of (b) since every face F of $\mathbb{R}_+M$ is generated by $M \cap F$. □

These results provide us with a wealth of examples of affine monoids. Gordan's lemma has an inhomogeneous generalization. It can be considered as the discrete analogue of Motzkin's theorem 1.27 on the finite generation of polyhedra.

Theorem 2.12. *Let $P \subset \mathbb{R}^d$ be a rational polyhedron, C the recession cone of P, and $L \subset \mathbb{Q}^d$ a lattice. Then $P \cap L$ is a finitely generated module over the affine monoid $C \cap L$.*

Proof. We form the cone $C(P) \subset \mathbb{R}^{d+1}$ over P and extend the lattice L to $L' = L \oplus \mathbb{Z}$. By Gordan's lemma $C(P) \cap L'$ is an affine monoid. Its generators at height 0 generate $C \cap L$, and its generators at height 1 generate $P \cap L$ as a $C \cap L$-module. □

The theorem contains (and is equivalent to) a statement about the set N of solutions of a homogeneous system of linear diophantine inequalities and congruences given as

$$\begin{aligned} &a_{i1}x_1 + \cdots + a_{id}x_d \geq u_i, && i = 1, \dots, m, \quad a_{ij}, u_i \in \mathbb{Z}, \\ &b_{i1}x_1 + \cdots + b_{id}x_d \equiv v_i \quad (w_i), && i = 1, \dots, n, \quad b_{ij}, v_i, w_i \in \mathbb{Z}. \end{aligned} \tag{2.1}$$

(We can split equations into a pair of inequalities.) The inequalities define a rational polyhedron P. The homogeneous congruences $b_{i1}x_1 + \cdots + b_{id}x_d \equiv 0 \ (w_i)$ define a sublattice L of $\mathbb{Z}^d$, and we claim that N is a finite module over L. In fact, the set of solutions of the inhomogeneous system of congruences is an affine lattice of the form $x' + L$ (unless it is empty). Let $P' = P - x'$; then $N = x' + (P' \cap L)$, and to $P' \cap L$ we can apply the theorem: there exists a finite set E generating $P' \cap L$ as an $P \cap L$-module. To find the system of generators of N, we replace each $x \in E$ by $x' + x$. To sum up:

Corollary 2.13. *Let $N \subset \mathbb{Z}^d$ be the set of solutions of the system* (2.1)*, and let $M \subset \mathbb{Z}^d$ be the (affine) monoid of solutions of the corresponding homogeneous system. Then N is a finitely generated M-module.*

Irreducible elements, standard map and total degree. The nonzero elements of an integral domain form a commutative cancellative monoid with respect to multiplication. In analogy to number-theoretic nomenclature, let us call an element x of a monoid M a *unit* if x has an inverse in M. Clearly, the units of M form a group, denoted by $\mathrm{U}(M)$. One calls x *irreducible* if in every decomposition $x = y + z$ one of the summands y, z must be a unit.

It is not difficult to analyze affine monoids M in these terms. To this end, and for many other purposes, we introduce the standard map on an affine monoid as follows.

The group $\operatorname{gp}(M)$ is isomorphic to $\mathbb{Z}^r$, $r = \operatorname{rank} M$. We identify $\operatorname{gp}(M)$ and $\mathbb{Z}^r$. Let $C = \mathbb{R}_+M \subset \mathbb{R}^r$ the cone generated by M. This cone has a representation

$$C = H^+_{\sigma_1} \cap \cdots \cap H^+_{\sigma_s}$$

as an irredundant intersection of halfspaces defined by linear forms on $\mathbb{R}^r$. Each of the hyperplanes H_{σ_i} is generated as a vector space by integral vectors. Therefore we can assume that σ_i is the $\mathbb{Z}^r$-height above H_{σ_i} (with nonnegative values on C; see Remark 1.72). After this standardization we call the σ_i the *support forms* of M and

$$\sigma : M \to \mathbb{Z}^s_+, \qquad \sigma(x) = (\sigma_1(x), \dots, \sigma_s(x))$$

the *standard map* on M. The hyperplanes H_{σ_i} are called the *support hyperplanes* of M.

The standard map has a natural extension to $\mathbb{R}^r$ with values in $\mathbb{R}^s$, also denoted by σ. It restricts to a $\mathbb{Z}$-linear map $\operatorname{gp}(M) \to \mathbb{Z}^s$. Furthermore $\sigma(C) \subset \mathbb{R}^s_+$. Theorem 1.16 implies that $\sigma_1, \dots, \sigma_s$ is a minimal set of generators of the dual cone C^*.

Note that the standard map depends only on the order of $\sigma_1, \dots, \sigma_s$. The σ_i (as $\mathbb{Z}$-linear forms on $\operatorname{gp}(M)$) are defined intrinsically by M; see Remark 2.6(b).

We call $\tau = \sigma_1 + \cdots + \sigma_s$ the *total degree* on M. The term "total degree" is justified: for any ring R of coefficients, σ induces an homomorphism $R[M] \to R[Y_1, \dots, Y_s]$ of R-algebras, namely the R-linear extension of the map $X^m \mapsto Y^{\sigma(m)}$ and $\tau(m)$ is the total degree of the monomial $Y^{\sigma(m)}$.

Proposition 2.14. *Let M be an affine monoid with standard map σ. Then:*

(a) the units of M are precisely the elements x with $\sigma(x) = 0$, or, equivalently, $\tau(x) = 0$;
(b) every element $x \in M$ has a presentation $x = u + y_1 + \cdots + y_m$ in which u is a unit and $y_1, \dots, y_m$ are irreducible;
(c) up to differences by units, there exist only finitely many irreducible elements in M.

Proof. (a) Clearly if $x, -x \in M$, then $\sigma(x) = 0$, since 0 is the only unit in $\mathbb{Z}^s_+$. Conversely, let $\sigma(x) = 0$. Then $\sigma(-x) = 0$ as well, and so $-x \in C$, where $C = \mathbb{R}_+M$ as above. Thus there exists a positive integer m with $m(-x) \in M$. Then $x' = (m-1)x + m(-x) \in M$, too, and $x + x' = 0$.

(b) Suppose that x is neither a unit nor irreducible. Then there exists a decomposition $x = y + z$ in which neither y nor z is a unit, and since $\tau(x) = \tau(y) + \tau(z)$, assertion (a) implies $\tau(y), \tau(z) < \tau(x)$. So we are done by induction.

(c) By hypothesis, M is finitely generated, say by $x_1, \dots, x_p$. We apply (b) to each of the x_i and obtain a collection H of finitely many irreducible elements. Evidently, every other irreducible element of M has the form $h + u$ with $h \in H$ and a unit u. □

The total degree τ is an example of a *grading* γ on M; we use this notion as a synonym for "monoid homomorphism from M to $\mathbb{Z}$." The choice of this name is motivated by the fact that a grading on M induces a grading on the algebra $R[M]$ in which all elements of R have degree 0 and every monomial X^s has degree $\gamma(s)$ (see Remark 4.5).

Positive monoids. Suppose that 0 is the only unit in M. It follows immediately from Proposition 2.14 that M has only finitely many irreducible elements in this case. Not only do they constitute a system of generators—they must be contained in every system of generators. This observation justifies the definition of the Hilbert basis:

Definition 2.15. A monoid is called *positive* if 0 is its only unit. The unique minimal system of generators of a positive affine monoid M given by its irreducible elements is called the *Hilbert basis* of M and denoted by $\operatorname{Hilb}(M)$.

For positive affine monoids the standard map is injective, and therefore we call it *standard embedding* in this case:

Proposition 2.16. *Let M be an affine monoid with* $\operatorname{gp}(M) = \mathbb{Z}^r$ *and $C = \mathbb{R}_+ M \subset \mathbb{R}^r$. Then the following are equivalent:*

(a) M is positive;
(b) the standard map σ is injective on M;
(c) $\sigma : \mathbb{R}^r \to \mathbb{R}^s$ is injective;
(d) C is pointed.

Proof. (a) $\Longrightarrow$ (d) Let U be the vector subspace of $\mathbb{R}^r$ consisting of all the elements $x \in C$ with $-x \in C$. It is evidently the intersection of the support hyperplanes of C and therefore a rational vector subspace. It is enough to show that $x = 0$ for $x \in U \cap \mathbb{Q}^r$. For such x there exists $a \in \mathbb{N}$, with $ax, a(-x) \in M$. Hence $x = 0$.

(d) $\Longrightarrow$ (c) If C is pointed, the dual cone C^* has dimension d (see Proposition 1.19). Thus the support forms generate $(\mathbb{R}^r)^*$, and a suitable collection $\sigma_{i_1}, \dots, \sigma_{i_r}$ of r support forms is linearly independent. If $\sigma_{i_j}(x) = 0$ for $j = 1, \dots, r$, then $x = 0$.

The remaining implications (c) $\Longrightarrow$ (b) and (b) $\Longrightarrow$ (a) are trivial. □

The total degree τ on a positive affine monoid is a grading under which only 0 has degree 0, as follows immediately from the injectivity of σ; it is an example of a *positive grading* γ on M, by definition a homomorphism $\gamma : M \to \mathbb{Z}_+$ such that $\gamma(x) = 0$ implies $x = 0$. A positive grading on M induces a positive grading on the algebra $R[M]$ in which all elements of R have degree 0 and every monomial X^s, $s \neq 0$, has positive degree.

We can now characterize positive affine monoids as submonoids of $\mathbb{Z}_+^n$, $n \geq 0$:

Proposition 2.17. *Let M be an affine monoid of rank r and with s support forms. Then the following are equivalent:*

(a) M is positive;
(b) M is isomorphic to a submonoid of $\mathbb{Z}_+^d$ for some d;
(c) M is isomorphic to a submonoid M' of $\mathbb{Z}_+^s$ such that the intersections $H_i \cap \mathbb{R}M'$ of the coordinate hyperplanes $H_1, \dots, H_s$ are exactly the support hyperplanes of M';
(d) M is isomorphic to a submonoid M' of $\mathbb{Z}_+^r$ such that the intersections $H_i \cap \mathbb{R}M'$ of the coordinate hyperplanes $H_1, \dots, H_r$ are among the support hyperplanes of M';
(e) M is isomorphic to a submonoid M' of $\mathbb{Z}_+^r$ with $\operatorname{gp}(M') = \mathbb{Z}^r$;
(f) M has a positive grading.

Proof. Each of (b)–(f) implies (a) for obvious reasons, and each of (c)–(e) implies (b). That (a) $\Longrightarrow$ (f) has already been observed.

For the implication (a) $\Longrightarrow$ (c) we use the injectivity of the standard embedding $\sigma : M \to \mathbb{Z}^s$. Set $M' = \sigma(M)$. Then $M' = \mathrm{gp}(M') \cap \mathbb{R}^s_+$, and since $M' \cong M$ has s support hyperplanes, all the coordinate hyperplanes must be support hyperplanes of M'. For (a) $\Longrightarrow$ (d) we simply choose r linearly independent ones among the support forms, as in the proof of Proposition 2.16(d) $\Longrightarrow$ (c).

For the implication (a) $\Longrightarrow$ (e) we have to find r linear forms $\rho_1, \dots, \rho_r$ with integral coefficients in $(\mathbb{R}_+ M)^*$ that form a basis of $(\mathbb{Z}^r)^*$. The existence of such a basis will be shown in Corollary 2.74. (However, note that the additional conditions in (d) and (e) cannot always be satisfied simultaneously, since $\rho_1, \dots, \rho_r$ cannot always be chosen as support forms of M. Exercise 2.8 asks for an example.) □

Polytopal monoids. We now introduce a special class of affine monoids, whose investigation in combinatorial and algebraic terms will be important for us.

Definition 2.18. Let L be an affine lattice in $\mathbb{R}^d$ and P an L-polytope. The *polytopal affine monoid* $M(P)$ associated with P is the monoid

$$\mathbb{Z}_+\{(x,1) : x \in \mathrm{lat}(P)\}.$$

in $\mathbb{R}^{d+1}$. The set $\{(x,1) : x \in \mathrm{lat}(P)\}$ generating $M(P)$ is denoted by $E(P)$.

Since the set $\mathrm{lat}(P)$ is finite, $M(P)$ is an affine monoid. It is evidently positive and its Hilbert basis is the set $\{(x,1) : x \in \mathrm{lat}(P)\}$.

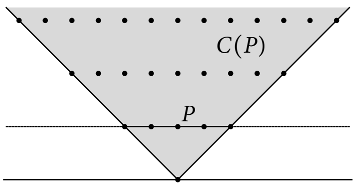

Fig. 2.1. Vertical cross-section of a polytopal monoid

The lattice L is not uniquely determined by P, and therefore the notations $E(P)$ and $M(P)$ are somewhat ambiguous. It will however always be clear which lattice L is to be considered.

Remark 2.19. More generally, we can associate a submonoid of $\mathbb{R}^d \oplus \mathbb{Z}_+ \subset \mathbb{R}^{d+1}$ with every subset X of $\mathbb{R}^d$, replacing the generating set $E(P)$ by $\{(x,1) : x \in X\}$. The structure of this monoid depends only on the affine structure of X. In fact, suppose that α is an affine isomorphism of $\mathbb{R}^d$ onto itself. Then α induces a *linear* isomorphism α' of the vector space $\mathbb{R}^{d+1}$ given as follows:

$$\alpha'(x,h) = (\alpha(x) - \alpha(0), h), \qquad x \in \mathbb{R}^d,\ h \in \mathbb{R}.$$

The restriction of α' to the monoid M over X maps M isomorphically onto the monoid over $\alpha(X)$.

In particular, if $L' = x + L$ is an affine lattice associated with the lattice L, and P' is an L'-polytope, then $P = P' - x$ is an L-polytope such that $M(P') \cong M(P)$. Whenever it should be convenient for the analysis of $M(P)$, we can therefore assume that P has its vertices in a lattice (and not just in an affine lattice).

Polytopal monoids are special instances of *homogeneous* affine monoids; such monoids M are positive and admit a positive grading in which every irreducible element has degree 1. The proof of the following proposition is left to the reader.

Proposition 2.20. *Let M be an affine monoid. Then the following are equivalent:*
(a) M is homogeneous;
(b) there exists a hyperplane H of $\mathbb{R}^d$, not passing through 0, such that M is generated by elements of H;
(c) M is positive, and the number of summands in every representation of an element $x \in M$ as a sum of irreducibles is constant.

Monoids with property (c) are called *half-factorial*; see Geroldinger and Halter–Koch [141] for this notion.

2.B Normal affine monoids

Definition 2.21. Let M be a submonoid of a commutative monoid N. The *saturation* or *integral closure* of M in N is the submonoid

$$\widehat{M}_N = \{x \in N : mx \in M \text{ for some } m \in \mathbb{N}\}$$

of N. One calls M *saturated* or *integrally closed* in N if $M = \widehat{M}_N$.

The *normalization* $\bar{M}$ of a cancellative monoid M is the integral closure of M in $\operatorname{gp}(M)$, and if $\bar{M} = M$, then M is called *normal*.

The terms “integral closure” and “normalization” are borrowed from commutative algebra, to which we will connect them in Section 4.E.

For submonoids of $\mathbb{Q}^d$ the integral closure has a clear geometric interpretation. It shows that the use of the notation $\widehat{M}_L$ in Corollary 2.10 was justified.

Proposition 2.22. *Let $M \subset N$ be submonoids of $\mathbb{Q}^d$ and $C = \mathbb{R}_+M$. Then $\widehat{M}_N = C \cap N$. If M and N are affine monoids, then so is $\widehat{M}_N$.*

Proof. The inclusion $\widehat{M}_N \subset C \cap N$ is trivial.

For the converse inclusion we choose an element $x \in C \cap N$. Then x has a representation as a $\mathbb{Q}_+$-linear combination of elements of M. Let m be the least common multiple of the denominators of the coefficients. Then $mx \in M$ and, hence, $x \in \widehat{M}_N$.

For the last statement we set $L = \operatorname{gp}(N)$. Then $\widehat{M}_L$ is affine by Corollary 2.10, and the intersection $\widehat{M}_N = N \cap \widehat{M}_L$ is affine by Corollary 2.11. □

Example 2.23. Consider the extension of affine monoids $2\mathbb{Z}_+ \subset \mathbb{Z}_+$. Then $2\mathbb{Z}_+$ is normal and, hence, smaller than its integral closure in $\mathbb{Z}_+$, which is $\mathbb{Z}_+$ itself.

The most important consequence of Proposition 2.22 is the characterization of affine normal monoids. It combines Proposition 2.22 with Gordan's lemma.

Corollary 2.24. *Let $M \subset \mathbb{Z}^r$ be a monoid such that* $\operatorname{gp}(M) = \mathbb{Z}^r$. *Then $\mathbb{Z}^r \cap \mathbb{R}_+M$ is the normalization of M.*

Moreover, the following are equivalent:

(a) M is normal and affine;
(b) $\mathbb{R}_+M$ is finitely generated and $M = \mathbb{Z}^r \cap \mathbb{R}_+M$;
(c) there exist finitely many rational halfspaces $H_i^+ \subset \mathbb{R}^r$ such that $M = \bigcap_i H_i^+ \cap \mathbb{Z}^r$.

We will sometimes refer to the monoids $H^+ \cap \mathbb{Z}^r$ (with H^+ a rational halfspace) as *discrete halfspaces*. Their structure is given by Corollary 2.27 below.

Before we record another useful corollary, we remind the reader that $N_* = \{0\} \cup \operatorname{int}(N)$ (see Remark 2.6).

Corollary 2.25. *Let M be a (not necessarily affine) integrally closed submonoid of an affine monoid N. If* $\operatorname{rank} M = \operatorname{rank} N$, *then* $\operatorname{gp}(M) = \operatorname{gp}(N)$. *In particular,* $\operatorname{gp}(N_*) = \operatorname{gp}(N)$.

Proof. We can assume that $\operatorname{gp}(N) = \mathbb{Z}^r$. Note that $\mathbb{R}M = \mathbb{R}N = \mathbb{R}^r$ since $\operatorname{rank} M = r$. There is nothing to show if $r = 0$. Suppose now that $r > 0$, and choose elements $x_1, \dots, x_r \in M$ generating $\mathbb{R}^r$ as a vector space. Let $M' = \mathbb{Z}_+x_1 + \cdots + \mathbb{Z}_+x_r$, and set $M'' = \mathbb{R}_+M' \cap N$. Then M'' is the integral closure of M' in N and an affine monoid itself by Proposition 2.22. Since $M' \subset M$, we obtain that $M'' \subset M$ by hypothesis on M. We can now replace M by M'' and assume that M is itself affine.

Choose $x \in \operatorname{int}(M)$. Then all support forms of M have positive value on x, and they are indeed linear forms on $\mathbb{R}N = \mathbb{R}M$. For $y \in N$ it therefore follows that $y + kx \in \mathbb{R}_+M$ for $k \gg 0$. Proposition 2.22 implies that $y + kx$ is integral over M. So $y + kx \in M$ by hypothesis, and $y \in \operatorname{gp}(M)$. □

In the previous subsection we have introduced the standard map $\sigma : M \to \mathbb{Z}^s$ on an affine monoid M. Proposition 2.14(a) shows that $\sigma(x) = 0$ if and only if x is a unit in M. For normal monoids M we can say even more:

Proposition 2.26. *Let M be a normal affine monoid,* $\mathrm{U}(M)$ *its subgroup of units, and* $\sigma : \operatorname{gp}(M) \to \mathbb{Z}^s$ *the standard map on M. Then M is isomorphic to* $\mathrm{U}(M) \oplus \sigma(M)$.

Proof. Let $L = \operatorname{gp}(M)$. We claim that $\mathrm{U}(M)$ is the kernel of σ. Clearly $\mathrm{U}(M) \subset \operatorname{Ker}\sigma$. Conversely let $x \in \operatorname{Ker}\sigma$. Then $x \in C \cap L$, where C is the cone generated by M. The normality of M then shows $x \in M$, and so $x \in \mathrm{U}(M)$.

Since $\mathrm{U}(M)$ is a direct summand of L, there exists a projection $\pi : L \to \mathrm{U}(M)$, i. e. a surjective $\mathbb{Z}$-linear map π with $\pi^2 = \pi$. Let $x \in M$. Then $(\pi(x), \sigma(x)) \in \mathrm{U}(M) \oplus \sigma(M)$. Conversely, given $(x_0, y') \in \mathrm{U}(M) \oplus \sigma(M)$ we choose $y \in M$ with $y' = \sigma(y)$. Then $x_0 + y - \pi(y) \in M$, $\pi(x_0 + y - \pi(y)) = x_0$ and $\sigma(x_0 + y - \pi(y)) = y'$. □

Corollary 2.27. *Let H^+ be a rational linear halfspace in $\mathbb{Z}^r$. Then $H^+ \cap \mathbb{Z}^r \cong \mathbb{Z}^{r-1} \oplus \mathbb{Z}_+$.*

Proof. Evidently $M = H^+ \cap \mathbb{Z}^r$ is normal, and $\mathrm{U}(M) = H \cap \mathbb{Z}^r \cong \mathbb{Z}^{r-1}$ since H is rational. Since $\sigma(M) \subset \mathbb{Z}_+$ is normal, we have $\sigma(M) \cong \mathbb{Z}_+$. □

The direct sum of a monoid M with a group G can be considered a trivial extension of M in almost every context treated in this book. Therefore one can usually restrict the discussion of normal affine monoids to the positive ones.

In the proof of Proposition 2.26 the normality of M is only used to the extent that $\mathrm{U}(M)$ is a direct summand of $\mathrm{gp}(M)$. However, if this condition is violated, we cannot expect an isomorphism $M \cong \mathrm{U}(M) \oplus \sigma(M)$. For example, choose $M = \{(x, y) \in \mathbb{Z}^2 : y > 0, \text{ or } y = 0 \text{ and } x \equiv 0 \ \ (2)\}$.

Polytopal monoids (see Definition 2.18) can be characterized in terms of their normalizations:

Proposition 2.28. *Let M be an affine monoid. Then the following are equivalent:*

(a) M is polytopal;
(b) M is homogeneous and coincides with $\bar{M}$ in degree 1.

Proof. The implication (a) $\Longrightarrow$ (b) is (almost) the definition of polytopal monoid. In fact, let P be a lattice polytope. The height 1 lattice points of $\mathbb{R}_+ M(P)$ are exactly the generators of $M(P)$, and so are contained in $M(P)$.

For the converse let $\mathrm{gp}(M) = \mathbb{Z}^d$. By hypothesis M has a grading γ. We can first extend it to $\mathbb{Z}$-linear form on $\mathbb{Z}^d$ and then to a linear form on $\mathbb{R}^d$. Since M is generated by elements of degree 1, $\mathbb{R}_+ M$ is generated by integral vectors of degree 1. Their convex hull is the lattice polytope $P = \{x \in \mathbb{R}_+ M : \gamma(x) = 1\}$. Since all the lattice points of P correspond to elements of M by hypothesis, $M \cong M(P)$. □

Normal monoids as pure submonoids. We have seen that every positive affine monoid can be realized as a submonoid of $\mathbb{Z}_+^n$ for suitable n. This statement can be considerably strengthened if M is normal.

Theorem 2.29. *The following are equivalent for a positive affine monoid M:*

(a) M is normal;
(b) there exists $m \in \mathbb{Z}_+$ and a subgroup U of $\mathbb{Z}^m$ of rank r such that M is isomorphic to $\mathbb{Z}_+^m \cap U$;
(c) there exist $p, q \in \mathbb{Z}_+$ and a $\mathbb{Z}$-linear map $\lambda : \mathbb{Z}^p \to \mathbb{Z}^q$ such that M is isomorphic to $\mathbb{Z}_+^p \cap \operatorname{Ker} \lambda$.

If M is normal of rank r and with s support forms, then one can choose $m = r$, $p = s$, and $q = r + s$. Moreover, an isomorphism as in (b) is given by the standard embedding.

Proof. It is obvious that the monoids in (b) and (c) are normal—those in (c) are even integrally closed in $\mathbb{Z}^p$.

Now suppose that M is normal. We identify $\mathrm{gp}(M)$ with $\mathbb{Z}^r$, apply the standard embedding $\sigma : M \to \mathbb{Z}_+^s$, and set $M' = \sigma(M)$, $U = \mathrm{gp}(\sigma(M))$. The monoid isomorphism $\sigma : M \to M'$ extends to an isomorphism $\sigma : \mathrm{gp}(M) \to U$. Proposition 2.22 shows that $x \in M$ if and only if $\sigma(x) \in \mathbb{Z}_+^s$.

For the more difficult implication (b) $\Longrightarrow$ (c) we have to use the *elementary divisor theorem* (for example, see Lang [238, p. 153]): let $L \cong \mathbb{Z}^n$ be a free abelian group, and L' a subgroup of rank m; then there exist a basis $e_1, \dots, e_n$ of L and positive integers $d_1, \dots, d_m$ such that $d_1 e_1, \dots, d_m e_m$ is a basis of L' (and $d_1 \mid d_2 \mid \cdots \mid d_m$).

We can assume that $M = \mathbb{Z}_+^s \cap U$, $U = \operatorname{gp}(M)$, $r = \operatorname{rank} M$, and apply the elementary divisor theorem with $L = \mathbb{Z}^s$ and $L' = U$. Let ρ_i, $i = 1, \dots, s$, be the linear form on $\mathbb{Z}^s$ that assigns each vector its ith coordinate with respect to the basis $e_1, \dots, e_s$. Then an element $x = (x_1, \dots, x_s) \in \mathbb{Z}^s$ belongs to M if and only if (i) $x_1, \dots, x_s \geq 0$, (ii) $\rho_i(x) = 0$ for $i = r+1, \dots, s$, (iii) $\rho_i(x) \equiv 0 \ \ (d_i)$ for $i = 1, \dots, r$.

In order to achieve our goal we have to convert the congruences into linear equations. We can change the coefficients of $\rho_1, \dots, \rho_r$ (with respect to the given coordinates of $\mathbb{Z}^s$) by adding multiples of d_i, and so we may assume that $\rho_1, \dots, \rho_r$ have nonnegative values on $\mathbb{Z}_+^s$. For elements $x \in \mathbb{Z}_+^s$ the congruence $\rho_i(x) \equiv 0 \ \ (d_i)$ is then equivalent to the solvability of the equation $\rho_i(x) = y_i d_i$ with $y_i \in \mathbb{Z}_+$, and the solution $y_i = y_i(x)$ is uniquely determined by x.

Define the map $\pi : U \to \mathbb{Z}^{s+r}$ by $x \mapsto \big(x, y_1(x), \dots, y_r(x)\big)$, and set $V = \pi(U)$. Then π is injective, and $\pi(x) \in \mathbb{Z}_+^{s+r}$ if and only if $x \in M$. It remains to show that V is the kernel of a suitable map $\mathbb{Z}^{s+r} \to \mathbb{Z}^s$. This is equivalent (by the elementary divisor theorem) to the torsion freeness of $\mathbb{Z}^{s+r}/V$.

Let $z \in \mathbb{Z}^{s+r}$, $z = (z', z'')$, $z' \in \mathbb{Z}^s$, $z'' \in \mathbb{Z}^r$, and suppose that $mz \in V$ for some positive integer m. Then $x = mz' \in U$. Moreover,

$$\rho_j(z') = m^{-1}\rho_j(x) = m^{-1} y_j(x) d_j = m^{-1} m z''_j d_j = z''_j d_j.$$

It follows that $z' \in U$ and $(z', z'') \in V$. This shows the torsion freeness of $\mathbb{Z}^{s+r}/V$. $\square$

Suppose M is a submonoid of N. Then M is called *pure* in N if $M = N \cap \operatorname{gp}(M)$. Equivalently we can require that $N \setminus M$ is an M-submodule of N. The notion of purity develops its power in the context of monoid algebras, where the purity of M in N means that $R[M]$ is a direct summand of $R[N]$ as an $R[M]$-module. Theorem 2.29 shows that a normal affine monoid can be realized not only as a pure submonoid of $\mathbb{Z}_+^n$ for suitable n (this is (b)), but also as a pure, integrally closed submonoid of a free monoid (this is (c)).

Remark 2.30. Let M be an positive affine normal monoid. The standard embedding $\sigma : M \to \mathbb{Z}^s$ realizes M as a pure submonoid as we have seen in Theorem 2.29(b). The quotient $\mathbb{Z}^s/\sigma(\operatorname{gp}(M))$ is an intrinsic invariant of M, namely the divisor class group $\operatorname{Cl}(M)$ of M (or any monoid algebra $\Bbbk[M]$ in which $\Bbbk$ is a factorial ring). See Corollary 4.56. So $\sigma(\operatorname{gp}(M))$ is a direct summand of $\mathbb{Z}^s$ if and only if $\operatorname{Cl}(M)$ is torsionfree.

Let M be a pure submonoid of the monoid N. Then N decomposes naturally into a disjoint union of M-submodules given by the intersections of N with the cosets of $\operatorname{gp}(M)$ in $\operatorname{gp}(N)$:

$$N = \bigcup \big(y + \operatorname{gp}(M)\big) \cap N,$$

where y runs through a system of representatives of the cosets $z + \operatorname{gp}(M)$, $z \in N$. We call the modules $\big(y + \operatorname{gp}(M)\big) \cap N$ the *coset modules* of M in N. Under suitable finiteness conditions they are finitely generated:

Proposition 2.31. *Let M be a pure submonoid of an affine monoid N. Then M is affine, too, and the coset modules of M in N are finitely generated over M. Moreover, each coset module is isomorphic to an M-submodule of* $\operatorname{gp}(M)$.

Proof. Since M is the intersection of affine monoids, namely N and $\operatorname{gp}(M)$, it is itself affine (Corollary 2.11).

Let us first assume that N is normal. Then M is likewise normal. After the identification of $\operatorname{gp}(N)$ with $\mathbb{Z}^n$, we can understand N as the set of solutions of a homogeneous linear diophantine system of inequalities, and M is cut out from N by a system of such equations and congruences. The coset module is then given as the set of solutions of an inhomogeneous variant of the homogeneous system, and so we can apply Corollary 2.13.

Now let N be arbitrary, and observe that $\bar{M} \subset \bar{N}$ is again pure. Since $\bar{M}$ is a finitely generated M-module, the coset modules of $\bar{M}$ in $\bar{N}$ are finite also over M. But each coset module of M in N is an M-submodule of a coset module of $\bar{M}$ in $\bar{N}$, and the desired finiteness follows.

The last statement is evident: if $w \in \big(y+\operatorname{gp}(M)\big)\cap N$, then $-w+\big(y+\operatorname{gp}(M)\big)\cap N \subset \operatorname{gp}(M)$. □

The coset modules of the standard embedding play a special role: they represent the divisorial ideals of M; see Theorem 4.62.

Adjoining inverse elements. In the case of a normal affine monoid M, certain extensions of M can be controlled by the support hyperplanes of M. For a monoid M contained in a group G and a subset $N \subset M$ we let $M[-N]$ denote the smallest submonoid of G containing M and all the elements $-x$ with $x \in N$. (If $N = \{x\}$, then we write $M[-x]$ instead of $M[-N]$.) One calls $M[-N]$ the *localization* with respect to N.

Proposition 2.32. *Let M be a normal affine monoid with* $\operatorname{gp}(M) = \mathbb{Z}^d$, *$x \in M$, and $H_1, \dots, H_s$ its support hyperplanes. Furthermore let F be the face of $\mathbb{R}_+M$ with $x \in$* $\operatorname{int}(F)$. *Then*

$$M[-x] = M[-(F\cap M)] = \bigcap\{H_i^+ : x \in H_i\} \cap \mathbb{Z}^d.$$

Furthermore $M[-x]$ splits into a direct sum $L \oplus M'$ where $L \cong \mathbb{Z}^e$, $e = \dim F$.

If M is positive, then M' is positive.

Proof. Let N be the normal affine monoid on the right hand side. Clearly $M[-x] \subset M[-(F\cap M)] \subset N$, since $M \subset N$ and $-(F\cap M) \subset N$. Conversely suppose that $y \in N$. For all the support forms σ_j with $x \notin H_j$ we have $\sigma_j(x) > 0$. Therefore $\sigma_j(y + kx) \geq 0$ for all such j if $k \gg 0$. It follows that $y + kx \in M$ for $k \gg 0$, and so $y \in M[-x]$. Altogether we have $M[-x] = N$.

Let $U = \mathbb{R}F \cap \mathbb{Z}^d$. Then U is a direct summand of $\mathbb{Z}^d$, and, moreover, $U \subset M[-x]$. In fact, $U \subset H_i$ for each support hyperplane H_i of M with $x \in H_i$. By the same argument as in the proof of Proposition 2.26, U splits off M.

For the last assertion one has to show that U is the group of units of $M[-x]$ if M is positive. Let $y \in M[-x]$ be an invertible element, say $y = y' - px$ with $y' \in M$ and $p \in \mathbb{Z}$. Then $-y = y'' - qx$, $y'' \in M$, $q \in \mathbb{Z}$. The equation $y' + y'' - (p+q)x = 0$ shows that $y' = 0$ if $p + q \le 0$, since y' is a unit in M in this case. If $p + q > 0$, then both y' and y'' must belong to the face F, and we are again done. □

The conductor. Let M be an affine monoid with normalization $\bar{M}$. Then the set of "gaps" $\bar{M} \setminus M$ is, in a sense, small.

Proposition 2.33. *Let M be an affine monoid. Then the ideal*

$$\mathfrak{c}(\bar{M}/M) = \{x \in M : x + \bar{M} \subset M\}$$

is nonempty. Moreover, the set $\bar{M} \setminus M$ is contained in finitely many hyperplanes parallel to the facets of M.

Proof. We have seen in Corollary 2.10 that $\bar{M}$ is a finitely generated M-module. Since $\bar{M} \subset \operatorname{gp}(M)$, there exist elements $z_1, \dots, z_n \in \operatorname{gp}(M)$ with $\bar{M} = (z_1+M) \cup \dots \cup (z_n+M)$. Write $z_i = y_i - x_i$ with $x_i, y_i \in M$, $i = 1, \dots, n$, and set $x = x_1 + \dots + x_n$. Then $x \in \mathfrak{c}(\bar{M}/M)$, and $\mathfrak{c}(\bar{M}/M) \neq \emptyset$.

Pick $x \in \mathfrak{c}(\bar{M}/M)$. If $z \in \bar{M} \setminus M$, then there exists at least one support form σ_i of M with $\sigma_i(z) < \sigma_i(x)$. So z belongs to one of the hyperplanes $\{y : \sigma_i(y) = k\}$, $k \in \mathbb{Z}_+$, $k < \sigma_i(x)$. □

The common name for $\mathfrak{c}(\bar{M}/M)$ is *conductor*. The reader may check that $\mathfrak{c}(\bar{M}/M)$ is the largest ideal of M that is also an ideal of $\bar{M}$ (Exercise 2.9).

Our first corollary shows that monoid elements in the interior of $\mathbb{R}_+M$ are nilpotent modulo any nonempty ideal of M. More precisely:

Corollary 2.34. *Let M be an affine monoid, L a lattice containing* $\operatorname{gp}(M)$*, $x \in M$, and $y \in \widehat{M}_L \cap \operatorname{int}(\mathbb{R}_+M)$. Then $ky \in x + \operatorname{int}(M)$ for some $k \in \mathbb{N}$.*

If, in addition, $y \in \operatorname{gp}(M)$, then $ky \in x + \operatorname{int}(M)$ for $k \gg 0$.

Proof. Suppose first that $y \in \operatorname{gp}(M)$. Then $y \in \bar{M} = \operatorname{gp}(M) \cap \mathbb{R}_+M$. Choose $z \in \mathfrak{c}(\bar{M}/M)$. Then $x + z + \bar{M} \subset x + M$, and after replacing x by $x + z$ we are left with the case in which M is normal. Since $y \in \operatorname{int}(\mathbb{R}_+M)$ we have $\sigma_i(y) > 0$ for all support forms σ_i of M. Thus $\sigma_i(ky) > \sigma_i(x)$ for all i and $k \gg 0$. It follows that $ky - x \in \operatorname{int}(\mathbb{R}_+M) \cap \operatorname{gp}(M) = M$ and therefore $ky \in x + \operatorname{int}(M)$.

In the general case there exists $q \in \mathbb{N}$ with $qy \in M$ for some $q \in \mathbb{N}$. Now we apply to qy what has just been shown. □

One can investigate the set $\bar{M} \setminus M$ in more detail. The next proposition can be derived rather easily from standard facts of commutative algebra. Imitating terminology of this field, we set

$$\operatorname{Rad}(I) = \{x \in M : ax \in I \text{ for some } a \in \mathbb{N}\}$$

for an ideal I in a monoid M, and call it the *radical* of I.

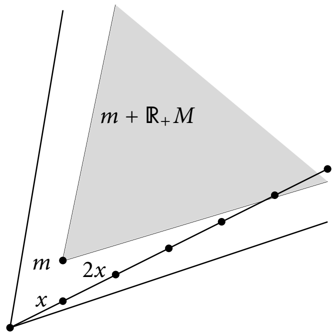

Fig. 2.2. Nilpotence of the interior modulo a nonempty ideal

Proposition 2.35. *Let M be an affine monoid. Then:*

(a) $\operatorname{Rad}\big(\mathfrak{c}(\bar{M}/M)\big)$ *is the set of all $x \in M$ such that $M[-x]$ is normal.*

(b) $\bar{M} \setminus M$ *is the union of a finite family of sets $x + (F \cap M)$ where $x \in M$ and F is a face such that $F \cap \mathfrak{c}(\bar{M}/M) = \varnothing$. Moreover, if F is maximal among these faces, then at least one set of type $x + (F \cap M)$ must appear.*

We will obtain the proposition as a consequence of a more general result on modules over affine monoids; see Proposition 4.37.

Prime and radical ideals. Let I be an ideal in a monoid M. One calls I a *radical ideal* if $I = \operatorname{Rad}(I)$ and a *prime ideal* if $I \neq M$, and $m + n \in I$ for $m, n \in M$ is only possible if $m \in I$ or $n \in I$. An affine monoid has only finitely many radical ideals and they are completely determined by the geometry of $\mathbb{R}_+M$:

Proposition 2.36. *Let M be an affine monoid and $I \subset M$ an ideal.*

(a) I is a radical ideal if and only if I is the intersection of the sets $M \setminus F$ where F is a face of $\mathbb{R}_+M$ with $F \cap I = \varnothing$.

(b) I is a prime ideal if and only if there exists a face F with $I = M \setminus F$.

Proof. Let F be a face of $\mathbb{R}_+M$. Then $M \setminus F$ is obviously a prime ideal, and since the intersection of prime ideals is radical, the intersection of sets of type $M \setminus F$ is a radical ideal.

Conversely, suppose that I is a radical ideal, and let $x \in M \setminus I$. There is a unique face F with $x \in \operatorname{int}(F)$, and it is enough to show that $F \cap I = \varnothing$. On the contrary, assume there exists $y \in F \cap I$. Let $N = F \cap M$. Then $kx \in y + N \subset y + M \subset I$ for some $k > 0$ by Corollary 2.34. Since I is a radical ideal, this implies $x \in I$, a contradiction.

For (b) one must show that the set of faces F with $F \cap I = \varnothing$ has a unique maximal element if I is a prime ideal. Assume that $F_1, \dots, F_n$ are the maximal faces in the complement of I. If $n \geq 2$, then a sum $x_1 + \cdots + x_n$ with $x_i \in \operatorname{int}(F_i) \cap M$ for all i neither belongs to I if I is prime nor belongs to the union $F_1 \cup \cdots \cup F_n$. This is a contradiction. □

By Proposition 2.36(b) the extreme submonoids of M (compare Remark 2.6) are exactly the complements of the prime ideals.

Special instances of prime ideals I are those generated by a single element x; in this case $I = M + x$. Then x is called a *prime element*. We say that prime elements x, y are *nonassociated* if $M + x \neq M + y$, or, in other words, $x - y \notin M$ and $y - x \notin M$.

Proposition 2.37. *Let M be an affine monoid, and N a submonoid generated by prime elements $x_1, \dots, x_n$ that are pairwise nonassociated. Then there exists a single face F of $\mathbb{R}_+ M$ such that $x_i \notin F$ for all i, and $M = \mathbb{Z}_+ x_1 \oplus \cdots \oplus \mathbb{Z}_+ x_n \oplus (M \cap F)$.*

Proof. We use induction, starting with the case $n = 1$. So let x be a prime element in M. Since x is not a unit, there exists a facet F not containing x. The ideal $M \setminus F$ is a prime ideal that does not contain any other prime ideal of M, and since $x \notin F$, we must have $M + x = M \setminus F$.

Let y be an element of M. Then there exists $k \in \mathbb{Z}_+$ such that $y - kx \in M$, but $y - (k+1)x \notin M$ (simply because $\sigma_F(x) > 0$). So $y - kx \in F$, for otherwise $y - kx \in M + x$, and $y - (k+1)x \in M$. We conclude that $M = \mathbb{Z}_+ x + (M \cap F)$. Since $\mathbb{Z}x \cap F = \{0\}$, we must even have $M = \mathbb{Z}_+ x \oplus (M \cap F)$.

Now suppose that $n > 1$. Using x_1 one splits M as $\mathbb{Z}_+ x_1 \oplus (M \cap F_1)$. Since $x_2, \dots, x_n \notin M + x$, they all lie in F, and we can conclude by induction since $x_2, \dots, x_n$ are prime elements also in the monoid $M \cap F_1$ and pairwise nonassociated, as the reader may check. □

Unions of normal monoids. A monoid M that is a union of normal submonoids M_i with $\operatorname{gp}(M_i) = \operatorname{gp}(M)$ is evidently itself normal. If M is affine and only finitely many submonoids are brought into play, then a much stronger statement is possible:

Theorem 2.38. *Let M be an affine monoid and $M_1, \dots, M_n$ submonoids of M such that $M = M_1 \cup \cdots \cup M_n$ and M_i is normal whenever* rank M_i = rank M. *Then M is normal, and it is the union of those M_i for which* rank M_i = rank M.

Proof. We can assume that $\operatorname{gp}(M) = \mathbb{Z}^d$. It is our first goal to reduce the theorem to its special case in which $\mathbb{R}_+ M_i = \mathbb{R}_+ M$ for all i with $\operatorname{rank} M_i = d$. To this end we consider all hyperplanes H in $\mathbb{R}^d$ that support one of the cones $C_i = \mathbb{R}_+ M_i$ with $\operatorname{rank} M_i = d$. This set of hyperplanes induces a dissection Γ of the cone $C = \mathbb{R}_+ M$ into rational subcones (see Proposition 1.69). Each cone $D \in \Gamma$ is generated by $D \cap M$, and $M = \bigcup\{D \cap M : D \in \Gamma,\ \dim D = d\}$.

Furthermore, if $D \in \Gamma$ with $\dim D = d$, then $\operatorname{gp}(D \cap M) = \mathbb{Z}^d$ by Corollary 2.25. So the normality of M follows from the normality of the intersections $D \cap M$. Now we can consider all the monoids $D \cap M$, $D \in \Gamma$, $\dim D = d$, and their submonoids $D \cap M_i$ separately. By construction of Γ we may therefore assume that $\mathbb{R}_+ M_i = \mathbb{R}_+ M$ for all submonoids M_i with $\operatorname{rank} M_i = d$.

Renumbering the M_i if necessary we can further assume that $\operatorname{rank} M_i = d$ for $i = 1, \dots, p$, and $\operatorname{rank} M_i < d$ for $i > p$. Set $G_i = \operatorname{gp}(M_i)$. If our claim does not hold, then we can find $x \in \bar{M} \setminus \bigcup_{i=1}^p M_i = \bar{M} \setminus \bigcup_{i=1}^p G_i$. Set $E = (x + \bigcap_{i=1}^p G_i) \cap C$. Then, on the one hand, $E \subset \bar{M} \setminus \bigcup_{i=1}^p G_i$. On the other hand, E is not contained in the union of finitely many hyperplanes. However, $\bar{M} \setminus M$ (by Proposition 2.33) and each submonoid M_i with rank $M_i < d$ is contained in such a union, and the conclusion is that E contains elements of $M \setminus \bigcup_{i=1}^{n} M_i$. This is a contradiction. □

Seminormal monoids. A property close to normality and of importance in K-theory (see Section 4.G and in particular Chapter 8) is seminormality:

Definition 2.39. A monoid is *seminormal* if every element $x \in \operatorname{gp}(M)$ with $2x, 3x \in M$ (and therefore $mx \in M$ for $m \in \mathbb{Z}_+$, $m \geq 2$) is itself in M. The *seminormalization* $\operatorname{sn}(M)$ of M is the intersection of all seminormal submonoids of $\operatorname{gp}(M)$ containing M.

It follows immediately from Corollary 2.10 that the seminormalization $\operatorname{sn}(M)$ of an affine monoid M is also affine; in fact, it is contained in the normalization $\bar{M}$, and therefore $\mathbb{R}_+ \operatorname{sn}(M) = \mathbb{R}_+ M$ is a cone.

A normal monoid is obviously seminormal, but the converse does not hold. There even exist seminormal, nonnormal polytopal monoids. We will give an example in Remark 2.56(b). For affine monoids M, the relationship between normality and seminormality is made precise by the next proposition. Recall that $M_* = \operatorname{int}(M) \cup \{0\}$ (Remark 2.6).

Proposition 2.40. *An affine monoid M is seminormal if and only if $(M \cap F)_*$ is a normal monoid for every face F of $\mathbb{R}_+ M$. Thus $M_* = \bar{M}_*$ if M is seminormal.*

Proof. Suppose that M is seminormal. Then $M \cap F$ is obviously seminormal for each face F of $\mathbb{R}_+ M$. Therefore it is enough to show that M_* is normal. By Corollary 2.25, $\operatorname{gp}(M_*) = \operatorname{gp}(M)$. Let $x \in \operatorname{gp}(M_*)$, $x \neq 0$, with $ax \in M_*$ for some $a \in \mathbb{N}$. Then $x \in \operatorname{int}(\mathbb{R}_+ M)$ and, by Corollary 2.34, $kx \in M$ for all $k \gg 0$. Let m be the largest integer for which $mx \notin M$. If $m > 0$, we have $2mx, 3mx \in M$, and so $mx \in M$ by seminormality, an obvious contradiction.

For the converse implication let $x \in \operatorname{gp}(M)$ be such that $2x, 3x \in M$. Let F be the unique face of $\mathbb{R}_+ M$ with $2x \in \operatorname{int}(F)$. Then certainly $3x \in \operatorname{int}(F)$, too, and so $x \in \operatorname{gp}\big(\operatorname{int}(F) \cap M\big)$. By hypothesis $x \in \operatorname{int}(F) \cap M$.

If M is seminormal, then M_* is normal, as just seen, and the normality of M_* implies $M_* = \bar{M}_*$. □

While M_* is almost never finitely generated, it is the filtered union of affine submonoids, and if M is seminormal, these can be chosen to be normal. To be more precise: there exists a family M_i of affine submonoids, indexed by the elements of a set I such that (i) $M = \bigcup_{i \in I} M_i$, and (ii) for all $i, j \in I$ there exists $k \in I$ such that $M_i, M_j \subset M_k$. In such a situation we will simply speak of a *filtered union*. For simplicity we restrict ourselves to positive affine monoids.

Proposition 2.41. *Let M be a positive affine monoid. Then M_* is the filtered union of affine submonoids. If M_* is normal, then these submonoids can be chosen to be normal.*

Proof. We embed M into $\mathbb{Z}^r = \operatorname{gp}(M)$, choose a rational cross-section P of $\mathbb{R}_+M$, and a rational point $z \in \operatorname{int}(P)$. In the affine space $\operatorname{aff}(P)$ we consider the homothety ϑ_λ with center z and factor $\lambda \in (0,1) \cap \mathbb{Q}$. Set $M_\lambda = M \cap \mathbb{R}_+\vartheta_\lambda(P)$. Then M_* is the filtered union of the affine monoids M_λ, and M_λ is normal if M_* is. We leave the detailed proof of this claim to the reader. □

We want to give another characterization of seminormal monoids that will turn out useful in proving the seminormality of their associated algebras and is similar in spirit to the characterization of normal affine monoids in Corollary 2.24.

Proposition 2.42. *Let $M \subset \mathbb{Z}^r$ be a monoid with* $\operatorname{gp}(M) = \mathbb{Z}^r$. *Then the following are equivalent:*

(a) M is seminormal affine;

(b) there exist finitely many rational halfspaces H_i^+ and subgroups $U_i \subset H_i \cap \mathbb{Z}^r$ such that $\operatorname{rank} U_i = r - 1$ *and*

$$M = \bigcap_i \big(U_i \cup (H_i^> \cap \mathbb{Z}^r)\big).$$

We leave the detailed proof to the reader. Also see Exercises 2.11 and 2.12 for more results on seminormal monoids. However, we want to indicate the construction of the halfspaces and the subgroups U_i: for each *face* F we choose a rational hyperplane H_i intersecting $\mathbb{R}_+M$ exactly in F_i, and set $U_i = \operatorname{gp}(M \cap F) \oplus V_i$ where V_i is a subgroup of $H_i \cap \mathbb{Z}^r$ such that $\operatorname{rank} V_i = r - 1 - \operatorname{rank} U_i$ and $U_i \cap V_i = 0$. Note that the proposition includes the statement that the monoids $U_i \cup (H_i^> \cap \mathbb{Z}^r)$ are affine.

2.C Generating normal affine monoids

Let $C \subset \mathbb{R}^d$ be a pointed rational cone, and L a sublattice of $\mathbb{Q}^d$. Then $C \cap L$ is an integrally closed submonoid of L. In this section we study the Hilbert basis of $C \cap L$. It is no restriction to assume that $\dim C = \operatorname{rank} L$. Otherwise we can replace C by the cone $C \cap \mathbb{R}L$. However, we do not insist on $d = \dim C$.

Simplicial cones and multiplicities. Let $v_1, \dots, v_r$ be linearly independent vectors in the lattice L. By

$$\operatorname{par}(v_1, \dots, v_r) = \{q_1v_1 + \dots + q_rv_r : 0 \le q_i < 1,\ i = 1, \dots, n\}$$

we denote the *semi-open parallelotope* spanned by $v_1, \dots, v_r$ (see Figure 2.3). Let U denote the sublattice generated by $v_1, \dots, v_r$. It is clear that every residue class $x + U$ for an element x in the saturation $\widehat{U}_L = \mathbb{Q}U \cap L = \mathbb{R}U \cap L$ of U in L has a unique representative in $\operatorname{par}(v_1, \dots, v_r)$, namely

$$x' = (a_1 - \lfloor a_1 \rfloor)v_1 + \dots + (a_r - \lfloor a_r \rfloor)v_r,$$

where $x = \sum_{i=1}^r a_iv_i$, $a_i \in \mathbb{Q}$. Furthermore, if $x \in C \cap L$ where C is the cone generated by $v_1, \dots, v_r$, then $a_1, \dots, a_r \ge 0$, and $x \in x' + M$ where $M = \mathbb{Z}_+v_1 + \dots + \mathbb{Z}_+v_r$. We have thus proved:

Proposition 2.43. *With the notation just introduced, the following hold:*
(a) $E = L \cap \operatorname{par}(v_1, \dots, v_r)$ *is a system of generators of the* M*-module* $C \cap L$*;*
(b) $(x + M) \cap (y + M) = \varnothing$ *for* $x, y \in E$, $x \neq y$*;*
(c) $\#E = \big[(\mathbb{Q}U \cap L) : U\big]$*;*
(d) $\operatorname{Hilb}(C \cap L) \subset \{v_1, \dots, v_r\} \cup E$.

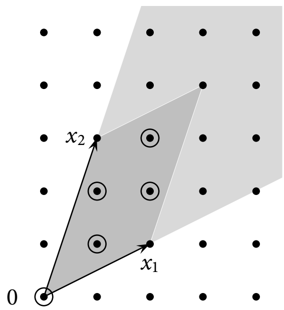

Fig. 2.3. The semi-open parallelotope and the set E

In part (c) we have used $\big[(\mathbb{Q}U \cap L) : U\big]$ for the index of the subgroup U in the group $\mathbb{Q}U \cap L$. The proposition can be interpreted as saying that $C \cap L$ is a free module over M with basis E. The rank of this free module, namely $\big[(\mathbb{Q}U \cap L) : U\big]$, measures how far U is from $\mathbb{Q}U \cap L$. Before continuing the main theme, namely the generation of normal affine monoids, we want to transfer this measure to lattice simplices, which, unlike the simplex $\operatorname{conv}(0, v_1, \dots, v_r)$ in Proposition 2.43, may not have 0 among their vertices.

Let $L = x + L_0$ be an affine lattice with associated lattice L_0 and U an affine sublattice, say $U = y + U_0$. Then we have $y \in L$ and $U_0 \subset L_0$. Then we set $L/U = L_0/U_0$. Since L_0 and U_0 do not depend on the choice of x and y, this definition is justified. However, it is reasonable beyond the formal justification: L decomposes into a disjoint union of translates of $V + y$ where V runs through the cosets of U_0 in L_0. Consequently we define the index of U in L by

$$[L : U] = [L_0 : U_0].$$

Let L be an affine lattice and Δ an L-simplex (see Definition 1.73). Then the roles of the lattices compared in Proposition 2.43(c) are played by the lattices introduced after Definition 1.73, namely

$$L_\Delta = L \cap \operatorname{aff}(\Delta)$$

and its affine sublattice

$$\mathscr{L}(\Delta) = v_0 + \sum_{v \in \operatorname{vert}(\Delta)} \mathbb{Z}(v - v_0),$$

where v_0 is an arbitrary vertex of Δ.

Definition 2.44. The *multiplicity* of Δ (with respect to L) is the index $\mu_L(\Delta) = [L_\Delta : \mathscr{L}(\Delta)]$.

The multiplicity has a very useful geometric interpretation that will allow us to define it for an arbitrary lattice polytope. In order to simplify the discussion we choose the origin in a vertex v_0 of Δ so that we are back in the situation of Proposition 2.43 and L_Δ and $\mathscr{L}(\Delta)$ are lattices of the same rank r. By the elementary divisor theorem we can find a basis $e_1, \dots, e_r$ of L_Δ and integers $d_1, \dots, d_r \geq 1$ such that $d_1 e_1, \dots, d_r e_r$ is a basis of $\mathscr{L}(\Delta)$. Then 2.43(c) shows that $\mu_L(\Delta) = d_1 \cdots d_r$. On $V = \mathbb{R}L_\Delta$ we introduce the r-dimensional volume function that gives volume 1 to the parallelotope spanned by $e_1, \dots, e_r$. This volume function depends only on L_Δ. We call it the *L-volume* vol_L on V. In fact, every change of basis in L_Δ is by an integral matrix of determinant ± 1. It follows that $\mu_L(\Delta)$ is exactly the L-volume of the parallelotope P spanned by $v_1, \dots, v_r$. By Exercise 3.1 the parallelotope P decomposes into $r!$ simplices that have the same volume as Δ:

Corollary 2.45. *For an L-simplex of dimension r we have* $\mu_L(\Delta) = r! \operatorname{vol}_L(\Delta)$.

Often the multiplicity of Δ is called its *normalized volume*. We can also pass to the monoid over the L-simplex $\Delta \subset \mathbb{R}^d$ and use it to measure the multiplicity of Δ. If $v_0, \dots, v_r$ are the vertices of Δ, then $M(\Delta)$ is generated by the vectors $v_i' = (v_i, 1) \in \mathbb{R}^{d+1}$, $i = 0, \dots, r$. Together with 0 they span again a simplex Δ'. Its multiplicity must be taken with respect to the lattice over L, namely the sublattice L' of $\mathbb{R}^{d+1}$ generated by the vectors $(x, 1)$, $x \in L$. We will write $L' = \mathbb{Z}(L, 1)$. (If $0 \in L$, then $L' = L \oplus \mathbb{Z}$.)

Proposition 2.46. *With the notation introduced, we have* $\mu_L(\Delta) = \mu_{L'}(\Delta')$.

Proof. We can assume that $v_0 = 0$ since all data are invariant with respect to the choice of origin in the affine space containing Δ. Next we are allowed to assume that $\operatorname{aff}(\Delta) = \mathbb{R}^r$, and finally that $L = \mathbb{Z}^r$. Then the L-volume is just the euclidean volume on $\mathbb{R}^r$, and $L' = \mathbb{Z}^{r+1}$. Clearly

$$\begin{aligned}\mu(\Delta) = r! \operatorname{vol}(\Delta) &= |\det(v_1, \dots, v_r)| \\ &= |\det(v_0', \dots, v_r')| = (r+1)! \operatorname{vol}(\Delta') = \mu(\Delta'). \end{aligned} \tag{2.2}$$

Alternatively, one establishes a bijection between $E = \operatorname{par}(v_1, \dots, v_r) \cap \mathbb{Z}^r$ and $E' = \operatorname{par}(v_0', \dots, v_r') \cap \mathbb{Z}^{r+1}$ using $v_0' = (0, 1)$ to lift the elements of E to their proper heights in E'. □

Example 2.47. Let $\Delta \subset \mathbb{R}^n$ be the convex hull of the vectors $x_0 = 0, x_1 = e_1, \dots, x_{n-1} = e_{n-1}$ (here e_i is the ith unit vector) and $x_n = (1, 1, \dots, 1, n)$. Then we have $\mu(\Delta) = n$ (with respect to $L = \mathbb{Z}^n$), as formula (2.2) shows. We consider the simplicial cone generated by $x_i' = (x_i, 1)$. Then $\operatorname{par}(x_1', x_2', \dots, x_n')$ contains exactly n lattice vectors, namely $0 \in \mathbb{R}^{n+1}$ and

$$\begin{aligned} &(1, 1, \dots, 1, a, n-a), && 1 \leq a \leq n/2, \\ &(1, 1, \dots, 1, a, n-a+1), && n/2 < a < n. \end{aligned}$$

The reader should compute this example, including all details. In Example 2.58 it will be used again.

In the situation of Proposition 2.43 the monoid $C \cap L$ depends only on C (and L), and if one wants to determine $\operatorname{Hilb}(C \cap L)$ it is certainly advisable to take $v_1, \dots, v_r$ in such a way that $\#E$ becomes as small as possible. There is a unique such choice: for each extreme ray R of (an arbitrary rational cone) C the monoid $R \cap L$ is normal and of rank 1. Therefore it is generated by a single element e; we call e an *extreme L-generator* of C, or an *extreme integral generator* if $L = \mathbb{Z}^d$. It is obvious that the extreme L-generators are irreducible elements of $C \cap L$ and therefore are elements of the Hilbert basis.

Definition 2.48. Let C be a simplicial rational cone. The simplex $\Delta_L(C)$ with vertices in 0 and the extreme L-generators of C is called the *basic L-simplex* of C. The *L-multiplicity* of C is $\mu_L(C) = \mu_L(\Delta_L(C))$.

Simplices and simplicial cones with the smallest possible multiplicity have a special name:

Definition 2.49. An L-simplex Δ is *unimodular* if $\mu_L(\Delta) = 1$. A simplicial rational cone C is *unimodular* (with respect to L) if $\Delta_L(C)$ is unimodular.

Proposition 2.50. *Let $\Delta = \operatorname{conv}(x_0, \dots, x_r)$ be an L-simplex where $L \subset \mathbb{R}^d$ is an affine lattice, and let $L' = \mathbb{Z}(L, 1)$. Then the following are equivalent:*

(a) Δ is unimodular;
(b) the sublattice generated by $x_1 - x_0, \dots, x_r - x_0$ is a direct summand of $L - x_0$;
(c) the submonoid of L' generated by $x'_i = (x_i, 1)$, $i = 0, \dots, r$, is integrally closed in L';
(d) the cone $C(\Delta)$ is unimodular with respect to L'.

Cones over polytopes. We generalize Proposition 2.43 in two steps, first to cones $C(P)$ where P is an L-polytope, and later on to arbitrary rational cones.

There is a simple strategy of how to make Proposition 2.43 useful for $C(P)$: we triangulate P, consider the simplicial cone over each simplex, and unite the systems of generators obtained in this way. Let us first introduce some special types of triangulations:

Definition 2.51. An L-polytope P is *empty* (with respect to L) if it contains no elements from L other than its vertices. A triangulation Σ of an L-polytope P is *full* (with respect to L) if $\operatorname{vert}(\Sigma) = \operatorname{lat}(P)$, or, equivalently, all its simplices are empty.

A triangulation Σ of an L-polytope is *unimodular* (with respect to L) if all simplices in Σ are unimodular.

It is immediately observed that every unimodular triangulation is full: a unimodular L-simplex contains no point of L different from its vertices, and every point $x \in \operatorname{lat}(P)$ must be contained in one of the simplices of Σ.

The natural degree on $\mathbb{Z}(L, 1) \subset \mathbb{R}^{d+1}$ associated with a lattice $L \subset \mathbb{R}^d$ is given by the last coordinate. If P is an L-polytope, then the natural generators of the polytopal monoid $M(P)$ have degree 1. (Polytopal monoids have been introduced in Definition 2.18.)

Theorem 2.52. *Let P be an L-polytope and $C = C(P)$. Then the integral closure $\widehat{M}(P) = C \cap \mathbb{Z}(L,1)$ of $M(P)$ in $\mathbb{Z}(L,1)$ is generated as an $M(P)$-module by elements of degree $\leq \dim P - 1$.*

Proof. We choose a full triangulation Σ of P, which exists by Theorem 1.51. For each simplex $\Delta \in \Sigma$ we apply Proposition 2.43 to the cone $C(\Delta)$. The union of the corresponding sets E_Δ evidently generates $\widehat{M}(P) = \bigcup_\Delta \widehat{M}(\Delta)$ over $M(P)$.

It remains to show that each element $y \in E_\Delta$ has degree at most $\dim P - 1$. Let $x_1, \dots, x_r$, $r = \dim P + 1$, be the vertices of Δ. Then $y = q_1 x_1' + \dots + q_r x_r'$, $0 \leq q_i < 1$, $x_i' = (x_i, 1)$, $i = 1, \dots, r$. Clearly

$$\deg y = q_1 + \dots + q_r < r = \dim P + 1,$$

and if we had chosen a coarse triangulation of Δ we might not have been able to reach the bound $\dim P - 1$. However, the basic simplices of the cones $C(\Delta)$ are empty. If $\deg y = \dim P$, then $y' = (x_1' + \dots + x_r') - y$ has degree 1 and belongs to the basic simplex of $C(\Delta)$ (more precisely, to its face opposite of 0). This is a contradiction, since $y' \neq x_i'$ for all i. □

Remark 2.53. In the proof of Theorem 2.52 we have constructed a system of generators of $\widehat{M}(P)$ as the union of the sets E_Δ. Note that

$$\sum_\Delta \#E_\Delta = \sum_\Delta \operatorname{vol}_L(\Delta)(\dim P)! = \operatorname{vol}_L(P)(\dim P)!$$

is independent of the triangulation. We call $\mu_L(P) = \operatorname{vol}_L(P)(\dim P)!$ the *multiplicity* of P (with respect to L). Theorem 6.54 will show that the multiplicity of P can indeed be interpreted as the multiplicity of a graded algebra, and this will finally justify the terminology.

Exercise 2.19 contains a variant of Theorem 2.52 taking into account the degrees of the elements in $\operatorname{int}(\widehat{M}(P))$. For $\dim P \leq 2$ the proposition has a strong consequence:

Corollary 2.54. *Let P be an L-polytope of dimension ≤ 2. Then $M(P)$ is integrally closed in $\mathbb{Z}(L,1)$.*

An empty L-simplex of dimension ≤ 2 is unimodular, and every full triangulation of an L-polytope of dimension ≤ 2 is unimodular.

Remark 2.55. In dimension $d \geq 3$ an empty lattice simplex need not be unimodular. For example, let $d = 3$ and consider the simplex

$$\Delta_{pq} = \operatorname{conv}\big((0,0,0), (0,1,0), (0,0,1), (p,q,1)\big)$$

with coprime integers $0 < q < p$. Then Δ_{pq} is empty (with respect to $\mathbb{Z}^3$), but has multiplicity p. It has been proved by White [377] that every empty 3-simplex is isomorphic to Δ_{pq} for some p and q, and that $\Delta_{pq} \cong \Delta_{uv}$ if and only if $p = u$ (an obviously necessary condition) and $v = q$ or $v = p - q$.

A similar classification of empty simplices in higher dimension is unknown. The classification in dimension 3 follows readily if one has shown that there exists a $\mathbb{Z}$-linear form α and an integer m such that Δ is sitting between the hyperplanes given by $\alpha(x) = m$ and $\alpha(x) = m+1$: Δ has *lattice width* 1. This is no longer true in dimension > 3. See Haase and Ziegler [174] and Sebő [313] for a discussion of empty simplices and further references.

Example 2.56. (a) Let Δ_{pq} be the simplex introduced in the preceding remark, $p > 1$. Then the basic simplex of $C(\Delta_{pq})$ is not unimodular, and the monoid $M(\Delta_{pq})$ is generated by the vectors $(x_i, 1)$ where x_i, $i = 1, \dots, 4$, runs through the vertices of Δ_{pq}. By Proposition 2.50 $M(\Delta_{pq})$ is not integrally closed in $\mathbb{Z}^4$. Thus $M(P)$ need not be integrally closed if $\dim P \geq 3$. However, note that Δ_{pq}, being empty, is unimodular with respect to $\mathscr{L}(\Delta_{pq})$. Therefore $M(\Delta_{pq})$ is normal.

(b) Let $v_1, \dots, v_4$ be the vertices of $\Delta = \Delta_{21}$ and $w = (v_1 + \dots + v_4)/2$. Furthermore we set $u_{ij} = (v_i + v_j)$, $1 \leq i \leq j \leq 4$. Then the u_{ij} and w are the lattice points of 2Δ, an integrally closed polytope according to Corollary 2.57 below. Now we form a 4-dimensional polytope P as the convex hull of $F_0 = (\Delta, 0)$ and $F_1 = (2\Delta, 1)$. So Δ and 2Δ are two parallel facets of P.

Set $M = M(P)$. A run of normaliz [74] shows that $\operatorname{Hilb}(\widehat{M})$ consists of the 15 elements corresponding to the lattice points of P, namely $(v_i, 0, 1)$, $(u_{ij}, 1, 1)$, $(w, 1, 1)$, and one more additional element $(w, 0, 2) \notin (P, 1)$. (This element must appear since $M(\Delta)$ is not integrally closed.) We claim that M is seminormal (see Definition 2.39) but not normal. In fact, suppose that $2x, 3x \in M$. Then $x \in \widehat{M}$, and so x can be represented by the 16 elements of $\operatorname{Hilb}(\widehat{M})$. If $(w, 0, 2)$ appears with the coefficient 0, then $x \in M$. If $2x, 3x$ in $\widehat{M}(F_0)$, then $x \in \operatorname{gp}(M(F_0))$, and so $x \in M(F_0) \subset M(P)$ since $M(F_0)$ is normal (albeit not integrally closed). The only possibility remaining is that $(w, 0, 2)$ and at least one of the generators $(u_{ij}, 1, 1)$ or $(w, 1, 1)$ of $M(F_1)$ appear in the presentation of x with positive coefficients. But since $(w, 0, 2) + (u_{ij}, 1, 1) \in M(P)$ as well as $(w, 0, 2) + (w, 1, 1) \in M(P)$ (as the reader may check), we finally conclude that $x \in M(P)$. Therefore the polytopal monoid M is seminormal. It is not normal since $\operatorname{gp}(M) = \mathbb{Z}^4$, and so $\overline{M} = \widehat{M} \neq M$.

(c) There exist 3-dimensional non-(semi)normal polytopes P. Let

$$P = \{x \in \mathbb{R}^3_+ : x_1/2 + x_2/3 + x_3/5 \leq 1\}.$$

Then P is a simplex with vertices $(0,0,0)$, $(2,0,0)$, $(0,3,0)$, and $(0,0,5)$. Obviously $\mathbb{Z}^3$ is the smallest lattice containing the vertices of P. The monoid $M = M(P)$ is not normal: $(1,2,4,2)$ belongs to $C(P)$, but not to M. Already in M_* normality is violated, so M is not even seminormal (see Proposition 2.40).

We refer the reader to [59] for a detailed discussion of the normality of simplices like P.

Nevertheless, Corollary 2.54 can be generalized if we consider multiples of lattice polytopes.

Corollary 2.57. *Let P be an L-polytope and $v \in \operatorname{vert}(P)$. Then $M(cP)$ is integrally closed in $\mathbb{Z}((c-1)v + L, 1)$ and, hence, normal for $c \in \mathbb{N}$, $c \geq \dim P - 1$.*

Proof. After a parallel translation of P we can assume that $v = 0$. Then the surrounding lattice is simply $L \oplus \mathbb{Z}$.

The integral closure of $M(cP)$ in $L \oplus \mathbb{Z}$ is isomorphic to the intersection of the integral closure of $M(P)$ with $L \oplus c\mathbb{Z}$. Under the isomorphism $L \oplus \mathbb{Z} \to L \oplus c\mathbb{Z}$, $(x, h) \mapsto (x, ch)$ the set of generators $E(cP)$ of $M(cP)$ is identified with the set of degree c elements in $\widehat{M}(P)$. Therefore we must show that every element x in $\widehat{M}(P)$ of degree nc, $n \in \mathbb{Z}_+$, is the sum of elements of degree c. There is nothing to show for $n \leq 1$. Suppose that $n \geq 2$. Then, by Theorem 2.52, $x = x' + x''$ where $x' \in M(P)$ and $\deg x'' \leq c$. Since x' decomposes into a sum of degree 1 elements, we can modify the representation of x to $x = x_1 + x_2$ where x_2 has degree c. The summand x_1 can be subdivided into a sum of degree c elements. □

Example 2.58. Suppose that $P \in \mathbb{R}^n$ is a lattice polytope with $z = 0 \in \operatorname{vert}(P)$. Then $M(P)[-(z,1)] = M' \oplus \mathbb{Z}$ where M' is the submonoid of $\mathbb{Z}^n$ generated by the lattice points in P (considered as vectors). If $M(P)$ is integrally closed in $\mathbb{Z}^{n+1}$, then the monoid M' is integrally closed in $\mathbb{Z}^n$. In other words, the monoid $C \cap \mathbb{Z}^n$, $C = \mathbb{R}_+ P$, is generated by the vectors $x \in C \cap P$.

We apply this argument to $P = c\Delta$ where Δ is the simplex constructed in Example 2.47, $n \geq 3$, $c \leq n-2$. The vector $w = (1, \dots, 1)$ belongs to C, but it is not contained in M'. First, it is not contained in P, since one has $\sum a_i = n - 1$ in a representation $w = a_1 x_1 + \dots + a_n x_n$ (take $a = 1$ in 2.47), and, moreover, every lattice vector in C with last coordinate 1 has all its coordinates positive: w is irreducible in $C \cap \mathbb{Z}^n$.

This example shows that Theorem 2.52 or Corollary 2.57 cannot be improved. It is not even possible to improve the bound $\dim P - 1$ if one replaces "integrally closed in $L \oplus \mathbb{Z}$" by "normal." In fact, $\operatorname{gp}(M(c\Delta)) = \mathbb{Z}^{n+1}$ for $c \geq 2$, as the reader may show (take $a = n - 1$ in 2.47).

We transfer the attributes "integrally closed" and "normal" from monoids to polytopes:

Definition 2.59. Let P be an L-polytope. We say that P is *L-integrally closed* if $M(P)$ is integrally closed in $\mathbb{Z}(L, 1)$ and that it is *normal* if $M(P)$ is normal.

The integral closedness of 2-dimensional lattice polytope P has been derived from the fact that the simplices in a full triangulation of P are unimodular. More generally, by the same argument one has the following:

Proposition 2.60. *Let P be an L-polytope that is covered by its unimodular subsimplices. Then P is integrally closed.*

In the next section and in Chapter 3 we will discuss to what extent the proposition can be reversed: does the integral closedness of P imply the existence of a triangulation into unimodular simplices or at least that P is covered by its unimodular subsimplices? What can be said about cP in this respect?

Pointed rational cones. It is not difficult to generalize the arguments above to arbitrary pointed rational cones. First we have to find a replacement for the polytope P, or rather $(P,1)$, spanning the cone. Now the notion of bottom introduced in Section 1.C is very useful. We set

$$C'_L = \operatorname{conv}(x \in L \cap C,\ x \neq 0\}.$$

Then $C'_L = \operatorname{conv}(\operatorname{Hilb}(C \cap L)) + C$, and so C'_L is a polyhedron by Theorem 1.27. Its bottom $B_L(C)$ is the union of its bounded faces (since C does not contain a line). Moreover, the line segment connecting 0 with a point in C'_L intersects $B_L(C)$ (all the nonbounded faces are faces of C). Let F be a facet of $B_L(C)$ and H_0 the hyperplane through 0 and parallel to F. Then the L-height above H_0 is a linear form; we denote it by β_F and call it a *basic grading*; the *basic F-degree* of C is the constant value $b_F = \beta_F(x)$, $x \in \operatorname{aff}(F)$.

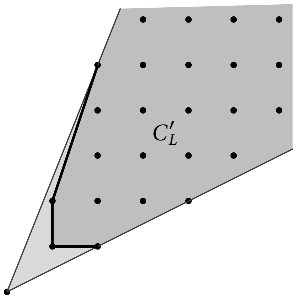

Fig. 2.4. The bottom

Example 2.61. Let $C \subset \mathbb{R}^3$ be generated by $x_1 = (0,1,0)$, $x_2 = (0,0,1)$, and $x_3 = (2,1,1)$. Together with $(0,0,0)$ these vectors span the empty simplex Δ_{21} of multiplicity 2 (see Remark 2.79). Thus the simplicial cone C has a single basic grading, corresponding to the single facet F of the bottom, the triangle spanned by x_1, x_2, x_3. As the reader may check, one has $b_F = 2$.

One needs one more element for $\operatorname{Hilb}(C)$, namely $y = (1,1,1)$: the cones generated by y and two of x_1, x_2, x_3 are unimodular so that C is the union of three unimodular cones.

Proposition 2.62. *Let $C \subset \mathbb{R}^d$ be a pointed rational cone of dimension d, L a lattice in $\mathbb{Q}^d$. Furthermore, let $B = B_L(C) \cap L$ and $M = \mathbb{Z}_+ B$. Then:*

(a) if $x \in C \cap L$, $x \neq 0$, and there exists a facet F of $B_L(C)$ with $\beta_F(x) < 2b_F$, then $x \in \operatorname{Hilb}(C)$; in particular $B \subset \operatorname{Hilb}(C)$;

(b) for each element y in the minimal system of generators G of $C \cap L$ as an M-module there exists a facet F of $B_L(C)$ such that $\beta_F(y) < (d-1)b_F$.

Proof. For (a) it is enough to note that an element $x \in C \cap L$, $x \neq 0$, with $\beta_F(x) < 2b_F$ must be irreducible since b_F is the minimal value of β_F on C'_L.

For (b) we choose a triangulation Σ of the conical complex, formed by the cones over the faces of the bottom $B_L(C)$, such that the rays of Σ constitute the set $\{\mathbb{R}_+ x : x \in B\}$ (Theorem 1.54). Let $D \in \Sigma$ be a simplicial cone. Then the basic L-simplex $\Delta_L(D)$ is empty. There is a facet $F \subset B_L(C)$, containing the bottom $B_L(D)$. As in the proof of Theorem 2.52 (where $b_F = 1$) we obtain that $D \cap L$ is generated by elements y with $\beta_F(y) < (d-1)b_F$ as a module over the monoid spanned by $D \cap B$. It only remains to unite all these vectors y to a system of generators of $C \cap L$ over B. □

In dimension 2, the lattice points in the bottom form the Hilbert basis because all the cones over the segments of the unique triangulation Σ of $B_L(C)$ with $\operatorname{vert}(\Sigma) = B_L(C) \cap L$ are unimodular by Corollary 2.54:

Corollary 2.63. *If* $\dim C = 2$, *then* $\operatorname{Hilb}(C \cap L) = B_L(C) \cap L$.

Remark 2.64. There is a remarkable connection between Hilbert bases in dimension 2 and continued fraction expansions, found by van der Corput [360]. Also see Oda [275, 1.6].

In dimension 3, $B_L(C) \cap L$ need not be the Hilbert basis of $C \cap L$, as Example 2.61 shows. Nevertheless one can easily describe the Hilbert basis: $\operatorname{Hilb}(C \cap L)$ consists exactly of $B = B_L(C) \cap L$ and the nonzero elements in the minimal system of generators of $C \cap L$ as a module over $\mathbb{Z}_+ B$, as follows immediately from Proposition 2.62. For later application we record

Corollary 2.65. *Let C be a cone of dimension* 3, *and $D \subset C$ a rational subcone.*

(a) If $B_L(D)$ is contained in $B_L(C)$, then $\operatorname{Hilb}(D \cap L) = D \cap \operatorname{Hilb}(C \cap L)$.
(b) If D is generated by elements $v_1, v_2, v_3 \in B_L(C) \cap L$ that span an empty simplex in a facet F of $B_L(C)$, then

$$\operatorname{Hilb}(D \cap L) = \{v_1, v_2, v_3\} \cup \big(\operatorname{par}(v_1, v_2, v_3) \cap L \setminus \{0\}\big).$$

Proof. (a) An element of $D \cap L$ that is irreducible in $C \cap L$, is certainly irreducible in $D \cap L$. This shows the containment $\supset$, which holds whenever $D \subset C$.

If $\dim D \le 2$, then the converse containment follows from the previous corollary. So let $\dim D = 3$ and consider $x \in \operatorname{Hilb}(D \cap L)$. Then there exists a basic grading β_F of D with $\beta_F(x) < 2b_F$. But β_F is a basic grading of C as well, and so $x \in \operatorname{Hilb}(C)$.

(b) This follows from the argument in the proof of the proposition: $\beta_F(x) < 2b_F$ for each $x \in \operatorname{par}(v_1, v_2, v_3) \cap L$. □

We can give up some precision in measuring the size of the elements of $\operatorname{Hilb}(C)$, using only the extreme L-generators of C. Since the bottom $B_L(C)$ is contained in the convex hull of 0 and the extreme generators, we obtain this corollary:

Corollary 2.66. *Let X be the set of extreme L-generators of the cone C. Then* $\operatorname{Hilb}(C \cap L) \subset (d-1)\operatorname{conv}(0, X)$ *if* $d = \dim C \ge 2$.

Carathéodory rank. So far we have tried to bound the systems of generators or the Hilbert basis in terms of degree. Another type of measure is given by the minimal number of elements in the Hilbert basis of a monoid M that, given $x \in M$, are needed to represent x:

Definition 2.67. Let M be a positive affine monoid. The *representation length* $\rho(x)$ of $x \in M$ is the smallest number k of elements $x_1, \dots, x_k \in \operatorname{Hilb}(M)$ such that x is a $\mathbb{Z}_+$-linear combination of $x_1, \dots, x_k$.

The *Carathéodory rank* of M, $\operatorname{CR}(M)$, is the maximum of the representation lengths $\rho(x)$, $x \in M$.

Since M is finitely generated, $\operatorname{CR}(M)$ is finite, namely $\operatorname{CR}(M) \le \#\operatorname{Hilb}(M)$. Without further hypothesis a better bound for $\operatorname{CR}(M)$ is impossible.

Example 2.68. Let $p_1, \dots, p_n$ be pairwise different prime numbers, and set $q_i = \prod_{j \ne i} p_j$. Then $q_1, \dots, q_n$ are coprime, and every $m \in \mathbb{Z}$, $m \gg 0$ is in the additive monoid generated by $q_1, \dots, q_n$. However, unless m is divisible by one of the prime numbers p_i, all the generators $q_1, \dots, q_n$ appear in every presentation of m.

If M is normal, then one can bound Carathéodory rank by a linear function of rank:

Theorem 2.69 (Sebő). *Let M be a positive normal affine monoid of rank r. Then* $\operatorname{CR}(M) = r$ *if* $r \le 3$, *and* $r \le \operatorname{CR}(M) \le 2r - 2$ *if* $r \ge 4$.

Proof. We identify $\operatorname{gp}(M)$ with $\mathbb{Z}^r$ and set $C = \mathbb{R}_+ M$.

It is easy to see that $\operatorname{CR}(M) \ge r$. In fact, the elements of M that are $\mathbb{Z}$-linear combinations of at most $r-1$ elements of $\operatorname{Hilb}(M)$ are contained in the union of finitely many hyperplanes in $\mathbb{R}^r$, and such a union cannot contain M.

If $r = 1$, then $M \cong \mathbb{Z}_+$, and there is nothing to show. If $r = 2$, then the bottom B of C' (constructed with respect to $L = \mathbb{Z}^2$) decomposes into a sequence of line segments $[x, y]$ such that $(x, y) \cap \mathbb{Z}^2 = \varnothing$. As we have seen in the proof of Proposition 2.62, every lattice point in the cone generated by $[x, y]$ belongs to the monoid generated by x and y. The case $r = 3$ will be discussed in Theorem 2.78.

So let $r \ge 4$ and $x \in M$. As in the proof of 2.62 (and the case $r = 2$ above) we choose a $B_L(C)$-bottom compatible triangulation of C whose rays pass through all lattice points in $B_L(C)$. The line segment $[0, x]$ intersects a facet F of the bottom, and we have seen that we can write $x = x' + x''$ where (i) x' is a linear combination of those (at most r) elements of $B \cap \mathbb{Z}^r$ that generate the smallest cone in Σ containing x, and (ii) $\beta_F(x') \le r b_F - 2$. Therefore x'' can be written as a $\mathbb{Z}_+$-linear combination of at most $r - 2$ elements in $\operatorname{Hilb}(M)$. □

We will continue the discussion of Carathéodory rank in the next section. At this point it is not easy to find monoids M as in the theorem with $\operatorname{CR}(M) > \operatorname{rank} M$.

Remark 2.70. In [58] two variants of CR have been introduced, *asymptotic* and *virtual* Carathéodory rank. The asymptotic Carathéodory rank $\operatorname{CR}^{\mathrm{a}}(M)$ is the smallest number m such that the proportion

$$\frac{\#\{x \in M, \|x\| \le t : \rho(x) > m\}}{\#\{x \in M, \|x\| \le t\}}$$

goes to 0 with $t \to \infty$. Here $\| \ \|$ refers to an arbitrarily fixed norm on the ambient real vector space $M \subset \mathbb{R} \otimes_{\mathbb{Z}} \operatorname{gp}(M)$. The virtual Carathéodory rank $\operatorname{CR}^{\mathrm{v}}(M)$ is the smallest number m such that $\rho(x) \le m$ with only finitely many exceptions.

One can show that $\operatorname{CR}^{\mathrm{a}}(M) \le 2 \operatorname{rank} M - 3$ if M is a normal monoid of rank ≥ 3. For a proof and further results we refer the reader to [58].

2.D Normality and unimodular covering

In the previous sections we used triangulations to construct systems of generators for normal monoids. In this section we will discuss the existence of triangulations into—and, more generally, of covers by unimodular simplices and unimodular cones. While the notions to be introduced below make sense with respect to arbitrary lattice structures, *we will work only with the lattice $\mathbb{Z}^d$ in this section*. This does not restrict the generality in an essential way, but simplifies the formulations somewhat. Consequently, we set $\operatorname{Hilb}(C) = \operatorname{Hilb}(C \cap \mathbb{Z}^d)$ if C is a pointed rational cone.

Definition 2.71. Let $\mathscr{F}$ be a rational fan in $\mathbb{R}^d$ consisting of pointed cones. It is called *unimodular* if all its cones are unimodular.

Theorem 2.72. *Let $\mathscr{F}$ be a rational fan in $\mathbb{R}^d$. Then $\mathscr{F}$ has a regular unimodular triangulation Σ.*

Proof. In the first step $\mathscr{F}$ is triangulated regularly as described in Corollary 1.66 after choosing a rational generating system in each $C \in \mathscr{F}$. By the transitivity of regular subdivisions (the conical analogue of Proposition 1.61) we may then assume that $\mathscr{F}$ consists of rational cones.

It remains to refine the triangulation if there exists a nonunimodular cone $D \in \mathscr{F}$. We use induction on (i) the maximal multiplicity of a d-dimensional cone D in $\mathscr{F}$, and on (ii) the number of such D with maximal multiplicity.

Let $m = \max\{\mu(D) : D \in \mathscr{F}\}$. If $m > 1$, we choose D with $\mu(D) = m$. Let $y_1, \dots, y_d$ be the extreme integral generators of D. Since $M = C \cap \mathbb{Z}^d$ the vectors $y_1, \dots, y_d$ belong to M. By Proposition 2.43 there exists $x \in \operatorname{par}(y_1, \dots, y_d) \cap \mathbb{Z}^d$. Now we apply stellar subdivision by x to $\mathscr{F}$. Figure 2.5 shows a typical situation after two generations of subdivisions in the cross-section of a 3-cone.

In this process every d-dimensional cone $D \in \mathscr{F}$ with $x \in D$ is replaced by the union of subcones $D'' = \mathbb{R}_+ x + D'$, D' a facet of D with $x \notin D'$. Then there exists j such that the y_i with $i \ne j$ generate D'. One has

$$\mu(D'') \le |\det(y_1, \dots, y_{j-1}, x, y_{j+1}, \dots, y_d)| < |\det(y_1, \dots, y_d)| = \mu(D)$$

since the coefficient of y_j in the representation of x is in the interval $[0, 1)$. Since at least one $D \in \mathscr{F}$ with $\mu(D) = m$ is properly subdivided, we are done by induction.

The regularity of the refined triangulation results from the conical analogue of Lemma 1.65 (see also Exercise 1.21). □

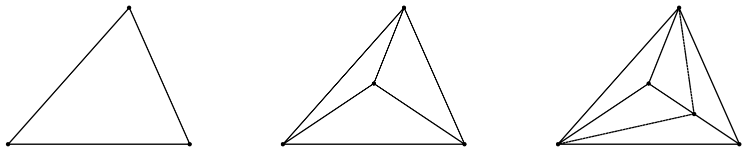

Fig. 2.5. Successive stellar triangulations

In the case in which $\mathscr{F}$ is a subdivision of a (not necessarily pointed) rational cone we can strengthen Theorem 2.72.

Corollary 2.73. *Let C be a rational cone in $\mathbb{R}^d$, and Γ a rational subdivision of C. Then there exists a regular unimodular triangulation of C that refines Γ.*

Proof. By Exercise 1.21 Γ has a rational regular refinement $\mathscr{F}$ to which we can apply Theorem 2.72. The regularity of the resulting triangulation follows from the conical analogue of Proposition 1.61. □

A much weaker consequence that has already been used in the proof of Proposition 2.17 is

Corollary 2.74. *Let C be a rational cone. Then* $\operatorname{int}(C)$ *contains a unimodular subcone.*

For the proof one chooses a simplicial subcone of $\operatorname{int}(C)$ and applies Theorem 2.72 to the fan of its faces. Another byproduct of the theorem is a criterion for the integral closedness of affine monoids:

Theorem 2.75. *Let $M \subset \mathbb{Z}^d$ be a positive affine monoid. Then the following are equivalent:*

(a) M is integrally closed in $\mathbb{Z}^d$;
(b) there exists a (regular) triangulation Σ of $\mathbb{R}_+ M$ such that each $D \in \Sigma$ is a unimodular cone with $\operatorname{Hilb}(D) \subset M$.

Proof. The implication (a) $\implies$ (b) follows immediately from Theorem 2.72 applied to (the fan of faces of) the cone $\mathbb{R}_+M$: the extreme integral generators of the unimodular cones belong to M.

The converse is a (trivial) special case of Theorem 2.38. □

Despite the simplicity of their proofs, 2.72, 2.73, and 2.75 have very important applications, as we will see later on.

Hilbert triangulations and covers. In the previous section we have seen much stronger results for special types of cones. For example, if P is a 2-dimensional lattice polytope, then $C(P)$ has a unimodular triangulation with rays through the Hilbert basis $E(P)$ of $M(P)$. The following definition covers this additional condition.

Definition 2.76. We say that Σ is a *Hilbert triangulation* of a pointed rational cone C if each ray of Σ is of type $\mathbb{R}_+x$ with $x \in \operatorname{Hilb}(C)$. It is a *full* Hilbert triangulation if all rays $\mathbb{R}_+x$ with $x \in \operatorname{Hilb}(C)$ are rays of Σ.

A very strong unimodularity condition for $\operatorname{Hilb}(C)$ is: *Every full Hilbert triangulation of C is unimodular.* All cones of dimension 2 and polytopal cones $C(P)$ with $\dim P = 2$ have this property. Another class is given by the cones over direct products of two unimodular simplices of arbitrary dimension (see Sturmfels [336, Sect. 6]).

Remark 2.77. Relaxing the Hilbert condition, one may ask whether every positive rational cone C contains a finite subset X of $\mathbb{Z}^d$ such that each triangulation using all the rays $\mathbb{R}_+x$, $x \in X$, is unimodular. We leave it to the reader that X has this property if and only if it is *supernormal* in the sense of Hoşten, Maclagan, and Sturmfels [245]: $D \cap X$ contains a Hilbert basis of D for every cone D generated by a subset X' of X. In general, a cone C does not contain a supernormal subset; see [245] for a counterexample.

A substantially weaker and much more important property is

(UHT) *C has a unimodular Hilbert triangulation.*

Theorem 2.78 (Sebő). *Let $C \subset \mathbb{R}^3$ be a positive rational cone. Then C has a unimodular Hilbert triangulation.*

The crucial step in the proof is the following lemma.

Lemma 2.79. *Let v_1, v_2, v_3 be vectors in $\mathbb{Z}^3$, for which* $\operatorname{conv}(0, v_1, v_2, v_3)$ *is an empty 3-simplex. Furthermore let β be the primitive linear form on $\mathbb{Z}^3$ that has constant value $b > 0$ on v_1, v_2, v_3. Then:*

(a) β is the (unique) basic grading of C;
(b) $\operatorname{Ker}\beta = \mathbb{Z}(v_2 - v_1) + \mathbb{Z}(v_3 - v_1)$*;*
(c) $b = \mu(C)$;
(d) on $\operatorname{par}(v_1, v_2, v_3) \cap \mathbb{Z}^3$ *the linear form β takes exactly the values* $0, b+1, \dots, 2b-1$.

Proof. (a) is clear from the choice of β. For (b) we observe that $\mathbb{Z}(v_2 - v_1) + \mathbb{Z}(v_3 - v_1) \subset \operatorname{Ker}\beta$. On the other hand 0, $v_2 - v_1$, and $v_3 - v_1$ are the vertices of an empty lattice triangle. Such a triangle is unimodular by Corollary 2.54 or, in other words, $v_2 - v_1$ and $v_3 - v_1$ generate a direct summand of $\mathbb{Z}^3$.

Now (c) is clear: since $\operatorname{Ker}\beta \subset U = \mathbb{Z}v_1 + \mathbb{Z}v_2 + \mathbb{Z}v_3$, we have $\mathbb{Z}/\beta(U) \cong \mathbb{Z}^3/U$, and $\mu(C) = \#(\mathbb{Z}^3/U)$. For part (d) one notes that the elements in $\operatorname{par}(v_1, v_2, v_3) \cap \mathbb{Z}^3$ represent the residue class modulo U. On the one hand, they have separate values under β and, simultaneously, only the values given are possible since Δ is empty. □

Proof of Theorem 2.78. We start with a triangulation of the bottom B of C into empty simplices Δ. By Corollary 2.65(a) we have $\operatorname{Hilb}(\mathbb{R}_+\Delta) = \operatorname{Hilb}(C) \cap \mathbb{R}_+\Delta$, and this allows us to replace C by $\mathbb{R}_+\Delta$. In other words, we can assume that C is generated by $v_1, v_2, v_3 \in \mathbb{Z}^3$ for which $\operatorname{conv}(0, v_1, v_2, v_3)$ is an empty simplex.

Assume C is not unimodular. With the notation of the lemma, we choose $w \in \operatorname{par}(v_1, v_2, v_3)$ such that $\beta(w) = b + 1$. Write $w = q_1v_1 + q_2v_2 + q_3v_3$ with rational numbers q_i, $0 \le q_i < 1$. Then $\beta(w) = (q_1 + q_2 + q_3)b$. Furthermore let C_i, $i = 1, 2, 3$ be the cone spanned by w and the v_j with $j \neq i$. Since $\operatorname{conv}(0, v_1, v_2, v_3)$ is empty and $\beta(w)$ is a generator of $\mathbb{Z}^3/U$, all three cones are full-dimensional. We have $\mu(C_i) = q_i b$, and $\sum_{i=1}^{3} \mu(C_i) = b + 1$.

Set $\operatorname{sdiv}(C) = \operatorname{par}(v_1, v_2, v_3) \cap (\mathbb{Z}^3 \smallsetminus \{0\})$. With analogous notation for C_i we claim

$$\operatorname{sdiv}(C) = \{w\} \cup \operatorname{sdiv}(C_1) \cup \operatorname{sdiv}(C_2) \cup \operatorname{sdiv}(C_3).$$

If this equation holds, we are done; the basic simplices of the cones C_i are empty (there is no integer between b and $b+1$), and an application of the same argument to C_i, $i = 1, 2, 3$, shows that all further subdividing vectors can be chosen in $\operatorname{sdiv}(C) \subset \operatorname{Hilb}(C)$ (compare 2.65(b)).

Note that the sets on the right hand side are disjoint. Since $\#\operatorname{sdiv}(C) = b - 1$ and $\#\operatorname{sdiv}(C_i) = q_i b - 1$, the choice of w guarantees that both sides have the same cardinality. Thus it is enough to show that $\operatorname{sdiv}(C)$ is contained in the right hand side R. This holds since $\operatorname{sdiv}(C) \subset \operatorname{Hilb}(C)$ by Corollary 2.65(b), and R, together with v_1, v_2, v_3, generates the monoid $C \cap \mathbb{Z}^3$. □

Remark 2.80. Theorem 2.78 cannot be extended to higher dimension. A polytopal counterexample in dimension 4 can be constructed as follows. We choose an empty 3-simplex Δ_{pq} (see Remark 2.55) with $1 < q < p-1$ (and p, q coprime) and set $P = 2\Delta$. By Corollary 2.57 the polytopal monoid $M(P)$ is integrally closed in $\mathbb{Z}^4$ so that $E(P)$ is a Hilbert basis of $C(P)$. If $C(P)$ admits a unimodular Hilbert triangulation, then P has a triangulation into unimodular simplices. However, such does not exist according to Kantor and Sarkaria [217]. (The first, nonpolytopal counterexample was given by Bouvier and Gonzalez–Sprinberg [40].)

While the example just discussed shows the failure of (UHT) in dimension ≥ 4, it does not exclude that a weaker unimodularity condition is always satisfied, namely:

(UHC) *Let C be a positive rational cone. Then C is the union of the unimodular simplicial cones generated by elements of* $\operatorname{Hilb}(C)$. *That is, a unimodular Hilbert cone.*

As we will see below, (UHC) also fails: there exists an integrally closed polytope P of dimension 5 such that $C(P)$, a cone of dimension 6, violates (UHC); equivalently, P is not covered by its unimodular subsimplices.

While the unimodular subcones generated by elements of $\operatorname{Hilb}(C)$ may fail to cover C, there always exist such unimodular subcones.

Proposition 2.81. *Let C be a pointed rational cone and $0 \subset F_1 \subset \cdots \subset F_d = C$ a complete flag of faces of C. Then there exist $x_1, \dots, x_d \in \operatorname{Hilb}(C)$ with the following properties:*

(a) $\mathbb{R}_+x_1 + \cdots + \mathbb{R}_+x_d$ *is unimodular;*

(b) $x_1, \dots, x_i \in F_i$ *for all* $i = 1, \dots, n$.

Proof. By induction we can assume that $x_1, \dots, x_{d-1} \in F_{d-1}$ have been found such that they satisfy the conditions (a) and (b) for F_{d-1}. Let σ be the support form of C associated with the facet F_{d-1}. Then there exists $y \in C \cap \mathbb{Z}^d$ with $\sigma(y) = 1$. Therefore $\sigma(x) = 1$ for at least one element $x \in \operatorname{Hilb}(C)$. Since $\operatorname{Hilb}(C) \cap F_i = \operatorname{Hilb}(F_i)$, the choice $x_d = x$ satisfies our needs. □

Isomorphism classes of lattice polytopes with given multiplicity. One interesting consequence of Proposition 2.81 (together with Proposition 2.57) is the following finiteness result for polytopes of bounded multiplicity.

Corollary 2.82. *Let d and μ be natural numbers. Then there are only finitely many isomorphism classes of lattice polytopes $P \subset \mathbb{R}^d$ such that $\mu(P) \le \mu$.*

Proof. In view of the definition of morphism of lattice polytopes (see Definition 1.73), we can assume that $\operatorname{aff}(P) = \mathbb{R}^d$, replacing $\mathbb{R}^d$ by $\operatorname{aff}(P)$ and $\mathbb{Z}^d$ by $\operatorname{aff}(P) \cap \mathbb{Z}^d$ if necessary.

It is sufficient to consider only normal d-polytopes P with $\operatorname{gp}(M(P)) = \mathbb{Z}^{d+1}$. In fact, the $(d-1)$th multiple of a lattice d-polytope in $\mathbb{R}^d$ always satisfies these conditions (Corollary 2.57) and, therefore, the multiplication by $d-1$ injects the class of lattice polytopes P in $\mathbb{R}^d$ of multiplicity $\le \mu$ into that of normal lattice polytopes Q in $\mathbb{R}^d$ of multiplicity $\le \mu(d-1)^d$ with $\operatorname{gp}(M(Q)) = \mathbb{Z}^{d+1}$. We leave it to the reader to show that P and P' are isomorphic if and only if $(d-1)P$ and $(d-1)P'$ are isomorphic.

According to Proposition 2.81 every polytope P from our class contains a unimodular d-simplex. This simplex is isomorphic to $\Delta = \operatorname{conv}(0, e_1, \dots, e_d)$, and therefore we can assume that $\Delta \subset P$. But then the condition $\mu(P) \le \mu$ implies that P is contained in the cube

$$\{(x_1, \dots, x_d) : |x_i| \le \mu,\ i = 1, \dots, d\} \subset \mathbb{R}^d. \qquad \square$$

See Bárány and Vershik [15] for an asymptotically sharp bound on the number of isomorphism classes, settling a conjecture of Arnold.

The integral Carathéodory property. If a pointed rational cone $C \subset \mathbb{R}^d$ has (UHC), then it satisfies the *integral Carathéodory property*, a name explained by the obvious analogy to Carathéodory's theorem 1.55:

(ICP) $\operatorname{CR}(C \cap \mathbb{Z}^d) = \dim C$.

Below we will give a counterexample to (ICP) (and therefore to (UHC) as well) and a counterexample to (UHC) that satisfies (ICP). The validity of (UHC) or (ICP) seems to be open in dimensions 4 and 5.

We can formulate all the conditions above for positive affine monoids M and their Hilbert bases: M has (UHT) if $\mathbb{R}_+M$ has a triangulation Σ whose simplices are generated by unimodular subsets of $\operatorname{Hilb}(M)$ (with respect to $\operatorname{gp}(M)$). In a similar manner, (UHC) is transferred to monoids, and (ICP) simply means that $\operatorname{CR}(M) = \operatorname{rank} M$. While the flexibility of the monoid language will be useful in the next subsection, it does not lead to a proper generalization. In fact, the weakest of these properties implies the normality of M:

Proposition 2.83. *Let M be an affine monoid and X a finite subset of M such that every element of M can be represented as a $\mathbb{Z}_+$-linear combination of $r = \operatorname{rank} M$ elements of X. Then M is normal, and every element of M is a $\mathbb{Z}_+$-linear combination of linearly independent elements of X.*

In particular, M is normal if $\operatorname{CR}(M) = r$.

Proof. We consider the submonoids N of M generated by at most r elements of X. If N has rank r, then its Hilbert basis is linear independent, and so N is normal.

By hypothesis, M is the union of all these submonoids N. By Theorem 2.38 it is normal and the union of those N that have rank r. □

Tight cones. We introduce the class of tight cones and monoids, and show that they play a crucial role for (UHC) and (ICP).

Definition 2.84. Let M be a positive normal affine monoid, $x \in \operatorname{Hilb}(M)$, and M' the monoid generated by $\operatorname{Hilb}(M) \setminus \{x\}$. We say that x is *nondestructive* if M' is normal and $\operatorname{gp}(M')$ is a direct summand of $\operatorname{gp}(M)$ (and therefore equal to $\operatorname{gp}(M)$ if $\operatorname{rank}\operatorname{gp}(M) = \operatorname{rank}\operatorname{gp}(M')$). Otherwise x is *destructive*. We say that M is *tight* if every element of $\operatorname{Hilb}(M)$ is destructive. A pointed rational cone $C \subset \mathbb{R}^d$ is *tight* if $C \cap \mathbb{Z}^d$ is tight.

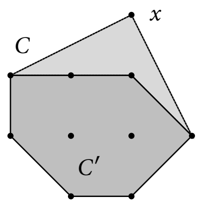

Fig. 2.6. Cross-section of shrinking of a cone

It is clear that only extreme elements of $\operatorname{Hilb}(M)$ can be nondestructive. Suppose that x is an extreme element of $\operatorname{Hilb}(M)$. Then the localization $M[-x]$ splits into a product $\mathbb{Z}x \oplus M_x$ where M_x is again a positive normal affine monoid (see Proposition 2.32).

Lemma 2.85. *Let M be a normal positive affine monoid and $x \in \operatorname{Hilb}(M)$ a nondestructive element. Let M' be the monoid generated by $\operatorname{Hilb}(M) \setminus \{x\}$.*

(a) If M' and M_x both satisfy (UHC), then so does M.
(b) One has $\operatorname{CR}(M) \le \max(\operatorname{CR}(M'), \operatorname{CR}(M_x) + 1)$.

Proof. Suppose M' and M_x both satisfy (UHC). Since $\operatorname{gp}(M')$ is a direct summand of $\operatorname{gp}(M)$ and $\operatorname{Hilb}(M') = \operatorname{Hilb}(M) \setminus \{x\}$ by the hypothesis on x, it is clear that all elements of M' are contained in submonoids of M generated by unimodular subsets X_i of $\operatorname{Hilb}(M)$ (with respect to $\operatorname{gp}(M)$). We have to show that this holds for elements of $M \setminus M'$, too.

Let $z \in M \setminus M'$. By hypothesis on M_x, the residue class of z in M_x has a representation $\bar{z} = a_1\bar{y}_1 + \cdots + a_m\bar{y}_m$ with $a_i \in \mathbb{Z}_+$ and $\bar{y}_i \in \operatorname{Hilb}(M_x)$ for $i = 1, \dots, m$ such that $\bar{y}_1, \dots, \bar{y}_m$ span a direct summand of $\operatorname{gp}(M_x)$. Next observe that $\operatorname{Hilb}(M)$ is mapped onto a system of generators of M_x by the residue class map. Therefore we may assume that the preimages $y_1, \dots, y_m$ belong to $\operatorname{Hilb}(M) \setminus \{x\}$. Furthermore, $z = a_1y_1 + \cdots + a_my_m + bx$ with $b \in \mathbb{Z}$.

It only remains to show that $b \in \mathbb{Z}_+$. There is a representation of z as a $\mathbb{Z}_+$-linear combination of the elements of $\operatorname{Hilb}(M)$ in which the coefficient of x is positive. Thus, if $b < 0$, z has a $\mathbb{Q}_+$-linear representation by the elements of $\operatorname{Hilb}(M) \setminus \{x\}$. This implies $z \in \mathbb{R}_+M'$, and hence $z \in M'$, a contradiction.

This proves (a), and (b) follows similarly. □

For a partial converse of the inequality in (b) see Exercise 2.31. We say that a monoid M as in the lemma *shrinks* to the monoid T if there is a chain $M = M_0 \supset M_1 \supset \cdots \supset M_t = T$ of monoids such that at each step M_{i+1} is generated by $\operatorname{Hilb}(M_i) \setminus \{x\}$ where x is nondestructive. An analogous terminology applies to cones.

Corollary 2.86. *A counterexample to* (UHC) *that is minimal, first with respect to dimension and then with respect to* $\#\operatorname{Hilb}(C)$, *is tight. A similar statement holds for* (ICP).

In fact, suppose that the cone C is a minimal counterexample to (UHC) with respect to dimension, and that C shrinks to D. Then D is also a counterexample to (UHC) according to Lemma 2.85. For (ICP) the argument is the same. It is therefore clear that one should search for counterexamples only among the tight cones. In fact, experiments have shown that such counterexamples are extremely rare, and without the restriction to tight cones such may never have been found.

Remark 2.87. It is not hard to see that there are no tight cones of dimension ≤ 2. However, tight cones exist in all dimensions ≥ 3. The first such example in dimension 3 was found by P. Dueck. The smallest example found by the authors has a Hilbert basis of 19 elements. Its cross-section is a regular hexagon (with respect to the action of the group of 3×3 invertible integer matrices). In dimension 4 there exist plenty of tight cones but none of the examples we have found is of the form C_P with a 3-dimensional lattice polytope P. In dimension ≥ 5 one can easily describe a class of tight cones; see Exercise 2.27.

The shrinking of cones and tests for (UHC) and (ICP) have been implemented as computer programs. See [53], [58], and [64] for a description of the search for counterexamples and the algorithms used for testing (UHC) and (ICP).

Counterexamples to (UHC) and (ICP). Let $C_6 \subset \mathbb{R}^6$, be the cone generated by the vectors $z_1, \dots, z_{10}$:

$$\begin{aligned} z_1 &= (0,1,0,0,0,0), & z_6 &= (1,0,2,1,1,2),\\ z_2 &= (0,0,1,0,0,0), & z_7 &= (1,2,0,2,1,1),\\ z_3 &= (0,0,0,1,0,0), & z_8 &= (1,1,2,0,2,1),\\ z_4 &= (0,0,0,0,1,0), & z_9 &= (1,1,1,2,0,2),\\ z_5 &= (0,0,0,0,0,1), & z_{10} &= (1,2,1,1,2,0). \end{aligned}$$

The cone C_6 and the monoid $M_6 = C_6 \cap \mathbb{Z}^6$ have several remarkable properties (see [68] for more detailed information):

(a) $\mathrm{Hilb}(C_6) = \{z_1, \dots, z_{10}\}$.
(b) C_6 has 27 facets, of which 5 are not simplicial.
(c) The automorphism group $\Sigma(M_6)$ of M_6 is the Frobenius group of order 20, the semidirect product of $\mathbb{Z}_5$ with its automorphism group $\mathbb{Z}_5^* \cong \mathbb{Z}^4$. It operates transitively on $\mathrm{Hilb}(M_6)$. In particular, this implies that $z_1, \dots, z_{10}$ are all extreme generators of C_6.
(d) The embedding $C_6 \subset \mathbb{R}^6$ above has been chosen in order to make visible the subgroup U of those automorphisms that map each of the sets $\{z_1, \dots, z_5\}$ and $\{z_6, \dots, z_{10}\}$ to itself; U is isomorphic to the dihedral group of order 10. However, C_6 can even be realized as the cone over a 0-1-polytope in $\mathbb{R}^5$.
(e) The vector of lowest degree disproving (UHC) is $t_1 = z_1 + \dots + z_{10}$. Evidently t_1 is invariant for $\Sigma(M_6)$, and it can be shown that its multiples are the only such elements.
(f) The Hilbert basis is contained in the hyperplane H given by the equation $-5\zeta_1 + \zeta_2 + \dots + \zeta_6 = 1$. Thus $z_1, \dots, z_{10}$ are the vertices of a normal 5-dimensional lattice polytope P_5 that is not covered by its unimodular lattice subsimplices (and contains no other lattice points).
(g) If one removes all the unimodular subcones generated by elements of $\mathrm{Hilb}(C_6)$ from C_6, then there remains the interior of a convex cone N. While P_5 has volume $25/120$, the intersection of N and P_5 has only volume $1/(1080 \times 120)$.
(h) The vector
$$z_1 + 3z_2 + 5z_4 + 2z_5 + z_8 + 5z_9 + 3z_{10}$$
cannot be represented by 6 elements of $\mathrm{Hilb}(M)$ (and it is smallest with respect to this property.) Moreover, one has $\mathrm{CR}(C_6) = 7$ (as can be seen from a triangulation containing only two nonunimodular simplices).

In particular, C_6 is even a counterexample to (ICP). If one adds the vectors

$$z_{11} = (0,-1,2,-1,-1,2) \quad \text{and} \quad z_{12} = (1,0,3,0,0,3),$$

one obtains a tight cone C_6' that satisfies (ICP), but fails (UHC). Remarkably, the Frobenius group of order 20 is the automorphism group of the monoid $M_6' = C_6' \cap \mathbb{Z}^6$, too, but its action on C_6' does not extend the action on M_6! Only a subgroup of order 4 restricts to M_6. Therefore M_6 has 5 conjugate embeddings into C_6', each of which contains $\{z_1, z_5\}$.

Exercises

2.1. Give an example of a monoid with torsion for which $\mathrm{gp}(M)$ is torsionfree.

2.2. Show that every submonoid of $\mathbb{Z}$ is finitely generated.

2.3. The notions of interior and extreme submonoid of a monoid M can be defined without reference to convex geometry. In this exercise we collect the corresponding assertions according to Swan [351, Sect. 5]. Let M be an affine monoid. Prove:

(a) A submonoid E of M is extreme if and only if $x + y \in E$ implies $x, y \in E$ for all $x, y \in M$.

(b) $\mathrm{int}(M) = \{x \in M : \text{for every } y \in M \text{ there exist } n \in \mathbb{N} \text{ and } z \in M \text{ such that } y + z = nx\}$.

For arbitrary monoids one can now use the properties in (a) and (b) to define extreme submonoids and the interior. So let M be an arbitrary monoid. Show:

(c) Every maximal submonoid of $M \setminus \mathrm{int}(M)$ is extreme.

(d) If E is an extreme submonoid and N an arbitrary submonoid. If $E \cap \mathrm{int}(N) \neq \emptyset$, then $N \subset E$.

2.4. Prove the following claims for an affine monoid M:

(a) $M = M_*$ if and only if $\operatorname{rank} M \leq 1$.

(b) If $M \neq M_*$, then M_* is not affine.

2.5. Let M be an affine monoid M. Show that the following are equivalent:

(i) M is a group;

(ii) the normalization $\bar{M}$ is a group;

(iii) $\mathbb{R}_+ M$ is a vector space.

2.6. Show that the intersection of $\mathbb{Z}^2$ with the cone generated by $(1, 0)$ and $(1, a)$ is a finitely generated monoid if and only if a is rational.

2.7. Let M be a normal affine monoid and $x \in M$. Then $M[-x] \cong \mathbb{Z}^e \oplus M'$ as discussed in Proposition 2.32. Let $\sigma_1, \dots, \sigma_e$ be those support forms of M that vanish on x. Show that the linear map $y \mapsto (\sigma_1(y), \dots, \sigma_t(y))$, $y \in \mathrm{gp}(M)$, restricts to the standard map on M'.

2.8. Show by means of an example that the additional conditions in Proposition 2.17(d) and (e) cannot always be satisfied simultaneously. (Rank 2 is enough.)

2.9. Show that $\mathfrak{c}(\bar{M}/M)$ is the largest ideal of M that is also an ideal of $\bar{M}$.

2.10. Let M be an affine monoid and $x \in \mathrm{int}(M)$. Show:

(a) $mx \in \mathfrak{c}(\bar{R}/R)$ for $m \gg 0$;

(b) $M[-x] = \mathrm{gp}(M)$.

Hint: (a) reduces (b) to the case in which $M = \bar{M}$.

2.11. Let M be a monoid. Show that $\mathrm{sn}(M)$ consists of all $x \in \mathrm{gp}(M)$ such that $mx, nx \in M$ for some coprime $m, n \in \mathbb{Z}_+$.

2.12. Let M be an affine monoid and let N be an overmonoid of M contained in $\mathrm{sn}(M)$. Show that there exists a chain

$$M = M_0 \subset M_1 \subset \cdots \subset M_n = N$$

of monoids such that $M_i = M_{i-1} \cup (M_{i-1} + x)$ for some x with $2x, 3x \in M_{i-1}$, $i = 1, \dots, n$.

Hint: first construct a strictly ascending chain of extensions as required, and note that M-submodules of $\bar{M}$ satisfy the ascending chain condition.

(In ring-theoretic terminology the extension $M_{i-1} \subset M_i$ is called *elementary subintegral* and the extension $M \subset N$ is *subintegral*; see p. 157.)

2.13. Let M be a positive normal affine monoid. We call M *Gorenstein* if there exists $x \in M$ such that $\operatorname{int}(M) = x + M$. (The nomenclature will be justified in Remark 6.34.) Show that M is Gorenstein if and only if there exists $x \in M$ such that $\sigma_F(x) = 1$ for each facet F of $\mathbb{R}_+M$ and the associated support form σ_F, in which case $\operatorname{int}(M) = x + M$.

2.14. Let M be as in Exercise 2.13 and suppose that $\operatorname{gp}(M) = \mathbb{Z}^d$. The *dual* monoid M^* is the set of all linear forms $\gamma \in (\mathbb{Z}^d)^*$ such that $\gamma(y) \geq 0$ for all $y \in M$. (Clearly, $M^* = (\mathbb{R}_+M)^* \cap (\mathbb{Z}^d)^*$.) Show the following are equivalent:

(i) M is polytopal;
(ii) M^* is Gorenstein.

Moreover, prove $M = M^{**}$.

Hint: consider the linear form on $\mathbb{Z}^d$ with respect to which M is polytopal.

2.15. Let $P \subset \mathbb{R}^d$ a lattice polytope containing 0 as an interior lattice polytope. One says P is *reflexive* if the dual polytope is again a lattice polytope (see Exercise 1.14). Show that P is reflexive if and only if 0 has height 1 over each facet of P.

Moreover, prove that the following are equivalent for an arbitrary lattice polytope:

(i) the normalization $\bar{M}(P)$ of $M(P)$ is Gorenstein;
(ii) there exists $x \in \mathbb{Z}^d$ and $k \in \mathbb{N}$ such that $kP - x$ is reflexive.

Batyrev [24] has used reflexive polytopes in the construction of mirror symmetric Calabi–Yau hypersurfaces. See Kreuzer and Skarke [232], [233] for classifications in dimensions 3 and 4. For further results we refer to Haase and Melnikov [172] and Haase and Nill [173].

2.16. (a) Let P be a lattice polygon with exactly one interior lattice point $x = 0$. Show P is reflexive.

(b)Try to find all reflexive lattice polygons. (There are 16 such polygons; see Kreuzer and Skarke [231].)

2.17. Let P be a lattice polygon with no interior lattice point. Show that $\operatorname{int}(M(P))$ is generated by elements of degree 2.

2.18. (a) Let M be a normal positive affine monoid. Suppose $x_1, \dots, x_m \in M$ generate $\operatorname{int}(M)$ as an ideal. Show that for each support form σ_F there exists i with $\sigma_F(x_i) = 1$.

(b) Consider the lattice polytope $P \in \mathbb{R}^3$ with vertices $(1,0,2)$, $(0,1,2)$, $(1,1,0)$, $(-1,-1,-1)$, $(1,1,2)$. Using normaliz or by giving a unimodular triangulation show that P is integrally closed, has exactly one lattice point, but is not reflexive, and $M(P)$ is not Gorenstein.

2.19. Let P be a lattice polytope, $M = \widehat{M}(P)$ be the integral closure of $M(P)$ and $m = \min\{\deg x : x \in \operatorname{int}(M)\}$. Show that M is generated by elements of degree $\leq \dim P + 1 - m$, and formulate a consequence analogous to Corollary 2.57.

Hint: reduce the problem to the simplicial case and apply a similar argument as in the proof of Theorem 2.52.

2.20. With the notation of Exercise 2.19 show that $\operatorname{int}(M)$ is generated by elements of degree $\leq \dim P + 1$ as an $M(P)$-module.

2.21. (a) Show that every positive rational cone is the intersection of finitely many unimodular cones.

(b) Show that every lattice polytope is the intersection of finitely many multiples of unimodular simplices.

2.22. Show that every lattice parallelotope is integrally closed.

2.23. Let $P \subset \mathbb{R}^d$ be a lattice polytope. For a vertex x of P set $C(x) = \mathbb{R}_+(P - x)$. We say that P is *very ample* if for every vertex x of P the vectors $y - x$, $y \in \operatorname{lat}(P)$, generate the monoid $C(x) \cap \mathbb{Z}^d$. (The terminology will be explained in Section 10.B.)

Set $M = M(P) \subset \mathbb{Z}^{d+1}$ and denote by $\widehat{M}$ the integral closure of M in $\mathbb{Z}^{d+1}$. Moreover, for each vertex x of P, let $\tilde{x} = (x, 1) \in M(P)$. Prove that the following are equivalent:

(i) P is very ample;
(ii) $k\tilde{x} + \widehat{M} \subset M$ for all $x \in \operatorname{vert}(P)$ and all $k \gg 0$;
(iii) $\operatorname{Rad} \mathfrak{c}(\widehat{M}/M) = M \setminus \{0\}$;
(iv) $M[-\tilde{x}]$ is integrally closed in $\mathbb{Z}^{d+1}$ for all $x \in \operatorname{vert}(P)$;
(v) $\widehat{M} \setminus M$ is a finite set.

It follows that an integrally closed lattice polytope is very ample.

2.24. Let I_i, $i = 1, 2, 3, 4$ be intervals in $\mathbb{Z}$, each containing at least two integers. Define the lattice polytope $P \subset \mathbb{R}^3$ as the convex hull of the set

$$\big((0,0) \times I_1\big) \cup \big((0,1) \times I_2\big) \cup \big((1,1) \times I_3\big) \cup \big((1,0) \times I_4\big).$$

(a) Prove that P is very ample.

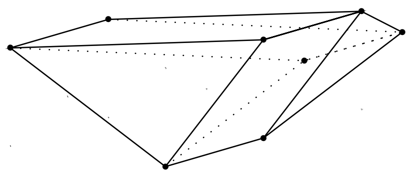

Fig. 2.7. A very ample, not integrally closed polytope

(b) Prove that P is not integrally closed for $I_1 = \{0,1\}$, $I_2 = \{2,3\}$, $I_3 = \{1,2\}$, $I_4 = \{3,4\}$ (see Figure 2.7).

2.25. Let $H \subset \mathbb{R}^d$ be a purely irrational hyperplane through 0 (i. e., containing no lattice point different from 0). Show that $H^>$ is the union of an ascending chain of unimodular subcones.

Hint: in dimension 2 choose a basis $x_1, x_2 \in H^>$ of $\mathbb{Z}^2$. Then exactly one of the cones spanned by $x_1, x_2 - x_1$, and $x_1 - x_2, x_2$ is contained in $H^>$. Choose it as the next member of the chain and iterate. In higher dimension consider pairs of basis elements.

2.26. Let P be a lattice polytope and suppose that $\mathscr{C}_0$ is a collection of unimodular subsimplices of P that covers P (in particular, P is integrally closed). Let $\mathscr{C}$ be the set of all faces of the simplices of $\mathscr{C}_0$, and let us say that $\delta \in \mathscr{C}$ is an *interior face* if $\delta \not\subset \partial P$.

Show that $\min\{\dim \delta : \delta \in \mathscr{C} \text{ is an interior face}\} + 1$ is the minimal degree of an element in $\operatorname{int}(M(P))$ [69, 1.2.5].

2.27. Let W be a cube of dimension d whose lattice points are its vertices and its barycenter. (For example, one can choose the cube with vertices in $\{\pm 1\}^d$ and consider the lattice generated by them and the origin.) Prove:

(a) The polytope W is normal in all dimensions.

(a) The cone $C(W)$ is tight if $d \geq 4$.

Hint: consider the midpoint x of the line segment joining the barycenter and a vertex. It belongs to the convex hull of the other vertices if $d \geq 4$. What follows for $2x$?

2.28. Let M_1 and M_2 be positive affine monoids and set $M = M_1 \oplus M_2$. Prove:
(a) $\mathrm{CR}(M) = \mathrm{CR}(M_1) + \mathrm{CR}(M_2)$.
(b) M satisfies (UHC) if and only if M_1 and M_2 do so.

2.29. Let M_1 and M_2 be homogeneous monoids, and let $M = M_1 \# M_2 = \{(x, y) \in M_1 \oplus M_2 : \deg x = \deg y\}$ be their *Segre product*. Prove:
(a) M is homogeneous;
(b) $\mathrm{CR}(M) \leq \mathrm{CR}(M_1) + \mathrm{CR}(M_2) - 1$;
(c) M satisfies (UHC) if and only if M_1 and M_2 do so.
Hint for (c): use what has been said about products of unimodular simplices. Indirectly it can also be used for (b).

2.30. Let P and Q be lattice polytopes. Prove $M(P \times Q) \cong M(P) \# M(Q)$.
There is also a polytope $\tilde{P}$ such that $M(\tilde{P}) = M(P) \oplus M(Q)$. Identify $\tilde{P}$.

2.31. Let M be a positive affine monoid, x an extreme integral generator, and suppose that $M[-x] \cong \mathbb{Z} \oplus M'$ in such a way that the associated homomorphism maps $\mathrm{Hilb}(M) \setminus \{x\}$ bijectively onto $\mathrm{Hilb}(M')$. Prove:
(a) $\mathrm{CR}(M) \geq \mathrm{CR}(M') + 1$;
(b) if M satisfies (UHC), then M' does so;
(c) a trivial realization of the situation considered in this exercise is $M = \mathbb{Z}_+ \oplus M'$. Show it is the only one if M and M' are homogeneous.

Notes

In 1872 Gordan [145] proved that the monoid of nonnegative solutions of a linear diophantine homogeneous system is finitely generated by its irreducible elements. In the context of invariant theory where it was found, the theorem indeed served as a lemma, namely in Gordan's second, simplified proof of the finite generation of the ring of invariants of binary forms. See M. Noether's obituary of his friend Gordan [274, p. 14]. Hilbert [190, p. 117] reproduced Gordan's theorem and proof and applied it to the same problem (in a much more transparent way).

Like many other theorems, Gordan's theorem has been proved several times. In 1903 Elliott [115] published an algorithm for the computation of the nonnegative solutions of a homogeneous linear diophantine equation. It was generalized to systems of such equations by MacMahon [247, Sect. VIII]. The Elliott–MacMahon algorithm actually yields the multigraded Hilbert series of the monoid of solutions (see Chapter 6). It was analyzed by Stanley in modern terms [323]: the algorithm amounts to computing a unimodular triangulation of the cone C of solutions starting from the unimodular positive orthant containing C.

In the generality of Lemma 2.9, the finite generation of the monoid of lattice points in a rational cone was proved by van der Corput in [360] (and two preceding articles), together with the uniqueness of the minimal system of generators in the positive case.

Normal affine monoids were thoroughly investigated by Hochster [195]. In his terminology pure submonoids are called “full” and integrally closed ones appear as “expanded” submonoids. Purity is a key concept for the investigation of rings of invariants. Results like Theorem 2.29 are modeled after [195]. Theorem 2.38 is a specialization of a theorem proved by the authors [58, Th. 6.1],

In the K-theoretic investigation of monoid rings (Chapter 8), the notion of seminormality, defined by Traverso, suggests itself, and the characterization of seminormal monoids consequently is to be found in [157]. Our treatment is influenced by [160] and Swan’s version [351] of [157]. See also Reid and Roberts [299] for a study of seminormality and related properties in monoids.

Theorem 2.52 is essentially due to Ewald and Wessels [119]. It was also proved by Liu, Trotter, and Ziegler [242] and Bruns, Gubeladze, and Trung [69]. Example 2.47 is a slight modification of an example in [119]. An ideal-theoretic variant of Theorem 2.52 was given by Reid, Roberts, and Vitulli [300]. See Bruns, Vasconcelos, and Villarreal [79] for related results.

The notion of Carathéodory rank was suggested by a result of Cook, Fonlupt, and Schrijver [91], later on improved by Sebő [314] to the bound given in Theorem 2.69.

The existence of a unimodular Hilbert triangulation in dimension 3 (Theorem 2.78) was proved independently by Sebő [314], Aguzzoli and Mundici [1], and Bouvier and Gonzalez–Sprinberg [40]. We have reproduced Sebő’s proof. In [40] it is furthermore shown that each two such triangulations are connected by a series of “flips.”

References for counterexamples to (UHC) and (ICP) and for the strategy of their search have been included in Section 2.D. A weaker version of (UHC) was discussed by Firla and Ziegler [124].

http://www.springer.com/978-0-387-76355-2

Polytopes, Rings, and K-Theory
Bruns, W.; Gubeladze, J.
2009, XIV, 461 p. 52 illus., Hardcover
ISBN: 978-0-387-76355-2