*Séminaires & Congrès*
**6**, 2002, p. 43–127

# SEMIGROUP ALGEBRAS AND DISCRETE GEOMETRY

*by*

Winfried Bruns & Joseph Gubeladze

***Abstract.*** — In these notes we study combinatorial and algebraic properties of affine semigroups and their algebras: (1) the existence of unimodular Hilbert triangulations and covers for normal affine semigroups, (2) the Cohen–Macaulay property and number of generators of divisorial ideals over normal semigroup algebras, and (3) graded automorphisms, retractions and homomorphisms of polytopal semigroup algebras.

## Contents



## 1. Introduction

These notes, composed for the Summer School on Toric Geometry at Grenoble, June/July 2000, contain a major part of the joint work of the authors.

In Section 3 we study a problem that clearly belongs to the area of discrete geometry or, more precisely, to the combinatorics of finitely generated rational cones and their Hilbert bases. Our motivation in taking up this problem was the attempt

**2000** ***Mathematics Subject Classification.*** — 13C14, 13C20, 13F20, 14M25, 20M25, 52B20.
***Key words and phrases.*** — Affine semigroup, lattice polytope, Hilbert basis, unimodular covering, integral Carathéodory property, triangulation, divisor class group, Cohen-Macaulay module, number of generators, Hilbert function, automorphism group, elementary transformation, column vector, polytopal algebra, retraction, tame morphism.

to understand the normality of affine semigroups (and their algebras). The counter-example we have found shows that some natural conjectures on the structure of Hilbert bases do not hold, and that there is no hope to explain normality in terms of formally stronger properties. Nevertheless several questions remain open: for example, the positive results end in dimension 3, while the counter-examples live in dimension 6.

Section 4 can be viewed as an intermediate position between discrete geometry and semigroup algebras. Its objects are the sets $T$ of solutions of linear diophantine systems of inequalities relative to the set $S$ of solutions of the corresponding homogeneous systems: $S$ is a normal semigroup and $T$ can be viewed as a module over it. After linearization by coefficients from a field, the vector space $KT$ represents a divisorial ideal over the normal domain $K[S]$ (at least under some assumptions on the system of inequalities). While certain invariants, like number of generators, can be understood combinatorially as well as algebraically, others, like depth, make sense only in the richer algebraic category.

The last part of the notes, Section 5, lives completely in the area of semigroup algebras. More precisely, its objects, namely the homomorphisms of polytopal semigroup algebras, can only be defined after the passage from semigroups to algebras. But there remains the question to what extent the homomorphisms can forget the combinatorial genesis of their domains and targets. As we will see, the automorphism groups of polytopal algebras have a perfect description in terms of combinatorial objects, and to some extent this is still true for retractions of polytopal algebras. We conclude the section with a conjecture about the structure of all homomorphisms of polytopal semigroup algebras.

Polytopal semigroup algebras are derived from lattice polytopes by a natural construction. While normal semigroup algebras in general, or rather their spectra, constitute the affine charts of toric varieties, the polytopal semigroup algebras arise as homogeneous coordinate rings of projective toric varieties. Several of our algebraic results can therefore easily be translated into geometric theorems about embedded projective toric varieties. Most notably this is the case for the description of the automorphism groups.

During the preparation of the final version of these notes the second author was generously supported by a Mercator visiting professorship of the Deutsche Forschungsgemeinschaft.

## 2. Affine and polytopal semigroup algebras

**2.1. Affine semigroup algebras.** — We use the following notation: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ are the additive groups of integral, rational, and real numbers, respectively; $\mathbb{Z}_+$, $\mathbb{Q}_+$ and $\mathbb{R}_+$ denote the corresponding additive subsemigroups of non-negative numbers, and $\mathbb{N} = \{1, 2, \ldots\}$.

*Affine semigroups.* — An *affine semigroup* is a semigroup (always containing a neutral element) which is finitely generated and can be embedded in $\mathbb{Z}^n$ for some $n \in \mathbb{N}$. Groups isomorphic to $\mathbb{Z}^n$ are called *lattices* in the following.

We write $\operatorname{gp}(S)$ for the group of differences of $S$, i. e. $\operatorname{gp}(S)$ is the smallest group (up to isomorphism) which contains $S$. Thus every element $x \in \operatorname{gp}(S)$ can be presented as $s - t$ for some $s, t \in S$.

If $S$ is contained in the lattice $L$ as a subsemigroup, then $x \in L$ is *integral* over $S$ if $cx \in S$ for some $c \in \mathbb{N}$, and the set of all such $x$ is the *integral closure* $\overline{S}_L$ of $S$ in $L$. Obviously $\overline{S}_L$ is again a semigroup. As we shall see in Proposition 2.1.1, it is even an affine semigroup, and can be described in geometric terms.

By a *cone* in a real vector space $V = \mathbb{R}^n$ we mean a subset $C$ such that $C$ is closed under linear combinations with non-negative real coefficients. It is well-known that a cone is finitely generated if and only if it is the intersection of finitely many vector halfspaces. (Sometimes a set of the form $z + C$ will also be called a cone.) If $C$ is generated by vectors with rational or, equivalently, integral components, then $C$ is called *rational*. This is the case if and only if the halfspaces can be described by homogeneous linear inequalities with rational (or integral) coefficients.

This applies especially to the cone $C(S)$ generated by $S$ in the real vector space $L \otimes \mathbb{R}$:

$$(*) \qquad C(S) = \{x \in L \otimes \mathbb{R} : \sigma_i(x) \geqslant 0,\ i = 1, \dots, s\}$$

where the $\sigma_i$ are linear forms on $L \otimes \mathbb{R}$ with integral coefficients.

We consider a single halfspace

$$H_i = \{x \in L \otimes \mathbb{R} : \sigma_i(x) \geqslant 0\}.$$

The semigroup $L \cap H_i$ is isomorphic to $\mathbb{Z}_+ \oplus \mathbb{Z}^{n-1}$ where $n = \operatorname{rank} L$.

Note that the cone $C(S)$ is essentially independent of $L$. The embedding $S \subset L$ induces an embedding $\operatorname{gp}(S) \subset L$ and next an embedding $\operatorname{gp}(S) \otimes \mathbb{R} \subset L \otimes \mathbb{R}$.This embedding induces an isomorphism of the cones $C(S)$ formed with respect to $\operatorname{gp}(S)$ and $L$.

***Proposition 2.1.1***

(a) (Gordan's lemma) *Let $C \subset L \otimes \mathbb{R}$ be a finitely generated rational cone (i. e. generated by finitely many vectors from $L \otimes \mathbb{Q}$). Then $L \cap C$ is an affine semigroup and integrally closed in $L$.*

(b) *Let $S$ be an affine subsemigroup of the lattice $L$. Then*
   - (i) $\overline{S}_L = L \cap C(S)$;
   - (ii) *there exist $z_1, \dots, z_u \in \overline{S}_L$ such that* $\overline{S}_L = \bigcup_{i=1}^{u} z_i + S$;
   - (iii) $\overline{S}_L$ *is an affine semigroup.*

*Proof.* — (a) Note that $C$ is generated by finitely many elements $x_1, \dots, x_m \in L$. Let $x \in L \cap C$. Then $x = a_1 x_1 + \dots + a_m x_m$ with non-negative rational $a_i$. Set $b_i = \lfloor a_i \rfloor$.

Then

$$(*) \qquad x = (b_1x_1 + \cdots + b_mx_m) + (r_1x_1 + \cdots + r_mx_m), \qquad 0 \leqslant r_i < 1.$$

The second summand lies in the intersection of $L$ with a bounded subset of $C$. Thus there are only finitely many choices for it. These elements together with $x_1, \ldots, x_m$ generate $L \cap C$. That $L \cap C$ is integrally closed in $L$ is evident.

(b) Set $C = C(S)$, and choose a system $x_1, \ldots, x_m$ of generators of $S$. Then every $x \in L \cap C$ has a representation $(*)$. Multiplication by a common denominator of $r_1, \ldots, r_m$ shows that $x \in \overline{S}_L$. On the other hand, $L \cap C$ is integrally closed by (a) so that $\overline{S}_L = L \cap C$.

The elements $y_1, \ldots, y_u$ can now be chosen as the vectors $r_1x_1 + \cdots + r_mx_m$ appearing in $(*)$. There number is finite since they are all integral and contained in a bounded subset of $L \otimes \mathbb{R}$. Together with $x_1, \ldots, x_m$ they certainly generate $\overline{S}_L$ as a semigroup. □

See Subsection 4.4 for further results on the finite generation of semigroups.

Proposition 2.1.1 shows that integrally closed affine semigroups can also be defined by finitely generated rational cones $C$: the semigroup $S(C) = L \cap C$ is affine and integrally closed in $L$.

We introduce special terminology in the case in which $L = \operatorname{gp}(S)$. Then the integral closure $\overline{S} = \overline{S}_{\operatorname{gp}(S)}$ is called the *normalization*, and $S$ is *normal* if $S = \overline{S}$. Clearly the semigroups $S(C)$ are normal, and conversely, every normal affine semigroup $S$ has such a representation, since $S = S(C(S))$ (in $\operatorname{gp}(S)$).

Suppose that $L = \operatorname{gp}(S)$ and that representation $(*)$ of $C(S)$ is irredundant. Then the linear forms $\sigma_i$ describe exactly the support hyperplanes of $C(S)$, and are therefore uniquely determined up to a multiple by a non-negative factor. We can choose them to have coprime integral coefficients (with respect to $e_1 \otimes 1, \ldots, e_r \otimes 1$ for some basis $e_1, \ldots, e_r$ of $\operatorname{gp}(S)$), and then the $\sigma_i$ are uniquely determined. We call them the *support forms* of $S$, and write

$$\operatorname{supp}(S) = \{\sigma_1, \ldots, \sigma_s\}.$$

The map

$$\sigma : S \longrightarrow \mathbb{Z}^s, \qquad \sigma(x) = (\sigma_1(x), \ldots, \sigma_s(x)),$$

is obviously a homomorphism that can be extended to $\operatorname{gp}(S)$. Obviously $\operatorname{Ker}(\sigma) \cap \overline{S}$ is the subgroup of $\overline{S}$ formed by its invertible elements: $x, -x \in C(S)$ if and only if $\sigma_i(x) = 0$ for all $i$.

Let $S_i = \{x \in S : \sigma_1(x) + \cdots + \sigma_s(x) = i\}$. Clearly $S = \bigcup_{i=0}^{\infty} S_i$, $S_i + S_j \subset S_{i+j}$ (and $S_0 = \operatorname{Ker}(\sigma) \cap S$). Thus $\sigma$ induces a *grading* on $S$ for which the $S_i$ are the graded components. If we want to emphasize the graded structure on $S$, then we call $\sigma(x)$ the *total degree* of $x$.

We call a semigroup $S$ *positive* if 0 is the only invertible element in $S$. It is easily seen that $\overline{S}$ is positive as well and that positivity is equivalent to the fact that $C(S)$

is a pointed cone with apex 0. Thus $\sigma$ is an injective map, inducing an embedding $\overline{S} \to \mathbb{Z}_+^s$. We call it the *standard embedding* of $\overline{S}$ (or $S$).

One should note that a positive affine semigroup $S$ can even be embedded into $\mathbb{Z}_+^r$, $r = \operatorname{rank}(S)$, such that the image generates $\mathbb{Z}_+^r$ as a group. We can assume that $\operatorname{gp}(S) = \mathbb{Z}^r$, and the dual cone

$$C(S)^* = \{\varphi \in (\mathbb{R}^r)^* : \varphi(x) \geqslant 0 \text{ for all } x \in S\}$$

contains $r$ integral linear forms $\varphi_1, \dots, \varphi_r$ forming a basis of $(\mathbb{Z}^r)^*$ (a much stronger claim will be proved in Subsection 3.3). Then the automorphism $\Phi = (\varphi_1, \dots, \varphi_r)$ of $\mathbb{Z}^r$ yields the desired embedding. (The result is taken from [**Gu2**]; this paper discusses many aspects of affine semigroups and their algebras not covered by our notes).

If $S$ is positive, then the graded components $S_i$ are obviously finite. Moreover, every element of $S$ can be written as the sum of irreducible elements, as follows by induction on the total degree. Since $S$ is finitely generated, the set of irreducible elements is also finite. It constitutes the *Hilbert basis* $\operatorname{Hilb}(S)$ of $S$; clearly $\operatorname{Hilb}(S)$ is the uniquely determined minimal system of generators of $S$. For a cone $C$ the Hilbert basis of $S(C)$ is denoted by $\operatorname{Hilb}(C)$ and called the *Hilbert basis* of $C$.

Especially for normal $S$ the assumption that $S$ is positive is not a severe restriction. In this case $S_0$ (notation as above) is the subgroup of invertible elements of $S$, and the normality of $S$ forces $S_0$ to be a direct summand of $S$. Then the image $S'$ of $S$ under the natural epimorphism $\operatorname{gp}(S) \to \operatorname{gp}(S)/S_0$ is a positive normal semigroup. Thus we have a splitting

$$S = S_0 \oplus S'.$$

*Semigroup algebras.* — Now let $K$ be a field. Then we can form the *semigroup algebra* $K[S]$. Since $S$ is finitely generated as a semigroup, $K[S]$ is finitely generated as a $K$-algebra. When an embedding $S \to \mathbb{Z}^n$ is given, it induces an embedding $K[S] \to K[\mathbb{Z}^n]$, and upon the choice of a basis in $\mathbb{Z}^n$, the algebra $K[\mathbb{Z}^n]$ can be identified with the Laurent polynomial ring $K[X_1^{\pm 1}, \dots, X_n^{\pm 1}]$. Under this identification, $K[S]$ has the monomial basis $X^a$, $a \in S \subset \mathbb{Z}^n$ (where we use the notation $X^a = X_1^{a_1} \cdots X_n^{a_n}$).

If we identify $S$ with the semigroup $K$-basis of $K[S]$, then there is a conflict of notation: addition in the semigroup turns into multiplication in the ring. The only way out would be to avoid this identification and always use the exponential notation as in the previous paragraph. However, this is often cumbersome. We can only ask the reader to always pay attention to the context.

It is now clear that affine semigroup algebras are nothing but subalgebras of $K[X_1^{\pm 1}, \dots, X_n^{\pm 1}]$ generated by finitely many monomials. Nevertheless the abstract point of view has many advantages. When we consider the elements of $S$ as members of $K[S]$, we will usually call them *monomials*. Products $as$ with $a \in K$ and $s \in S$ are called *terms*.

The Krull dimension $\dim K(S)$ of $K[S]$ is given by $\operatorname{rank} S = \operatorname{rank} \operatorname{gp}(S)$, since $\operatorname{rank} S$ is obviously the transcendence degree of the *quotient field* $\mathrm{QF}(K[S]) = \mathrm{QF}(K[\operatorname{gp}(S)])$ over $K$. (For standard notions of commutative algebra we refer the reader to Bruns and Herzog [**BH**], Eisenbud [**Ei**] or Matsumura [**Ma**].)

The semigroup algebra $K[S]$ is a special type of graded object. Therefore we introduce some terminology concerning graded rings and modules. Let $G$ be an abelian group. A *$G$-grading* on a ring $R$ is a decomposition $R = \bigoplus_{g \in G} R_g$ of abelian groups such that $R_g R_h \subset R_{g+h}$ for all $g, h \in G$, and $R$ (together with the grading) is called a *$G$-graded ring.* A $G$-grading on an $R$-module $M$ is a decomposition $M = \bigoplus_{h \in G} M_h$ such that $R_g M_h \subset M_{g+h}$ for all $g, h \in G$. If $H \subset G$ is a semigroup, then we may say that $R$ is $H$-graded if $R_g = 0$ for $g \notin H$. A *positively graded* algebra $R$ over a field $K$ is $\mathbb{Z}$-graded with $R_i = 0$ for $i < 0$ and $R_0 = K$. A *grading* (without further qualification of $G$) is usually a $\mathbb{Z}$-grading. A *multigrading* is a grading by a finitely generated abelian group.

If $S$ is positive, then $\operatorname{Hilb}(S)$ is a minimal set of generators for $K[S]$. Moreover, the total degree on $S$ induces a grading of $K[S]$ that under the standard embedding by $\sigma = (\sigma_1, \dots, \sigma_s)$ is just the grading inherited from the grading by total degree on $K[\mathbb{Z}_+^s] = K[Y_1, \dots, Y_s]$. The embedding $K[S] \subset K[Y_1, \dots, Y_s]$ is also called the *standard embedding* if $S$ is positive. Note that $K[S]$ is a positively graded $K$-algebra for positive $S$ and the total degree.

The reader should note that the usage of the terms "integral over", "integral closure", "normal" and "normalization" is consistent with its use in commutative algebra. So $K[\overline{S}_L]$ is the integral closure of $K[S]$ in the quotient field $\mathrm{QF}(K[L])$ of $K[L]$: it is generated by elements integral over $K[S]$, and it is integrally closed in $\mathrm{QF}(K[L])$. In fact, $K[\overline{S}_L]$ is the intersection of the algebras

$$K[H_i \cap L] \cong K[\mathbb{Z}_+ \oplus \mathbb{Z}^{n-1}] \cong K[Y, Z_1^{\pm 1}, \dots, Z_{n-1}^{\pm 1}].$$

Each of them is integrally closed in its field of fractions $\mathrm{QF}(K[L])$.

If $S$ is normal, then one has a splitting $S = S_0 \oplus S'$ as discussed above. It induces an isomorphism

$$K[S] = K[S_0] \otimes K[S'].$$

Therefore $K[S]$ is a Laurent polynomial extension of $K[S']$. Since Laurent polynomial extensions preserve essentially all ring-theoretic properties, it is in general no restriction to assume that $S$ is positive (if it is normal).

*Monomial prime ideals.* — The prime ideals in $K[S]$ that are generated by monomials can be easily described geometrically. Let $F$ be a face of $C(S)$, i. e. the intersection of $C(S)$ with some of its support hyperplanes. The ideal generated by all the monomials $x \in S$ that do not belong to $F$, has exactly these monomials as a $K$-basis. Thus there is a natural sequence

$$K[F \cap S] \xrightarrow{\iota_F} K[S] \xrightarrow{\pi_F} K[F \cap S]$$

where $\iota_F$ is the embedding induced by $F \cap S \subset S$, and $\pi_F$ is the $K$-linear map sending all elements in $F \cap S$ to themselves and all other elements of $S$ to 0. Obviously $\pi_F$ is a $K$-algebra homomorphism, and $\pi_F \circ \iota_F$ is the identity on $K[F \cap S]$. It follows that

$$\mathfrak{p}_F = \operatorname{Ker} \pi_F$$

is a prime ideal in $K[S]$, and it is not hard to show that the $\mathfrak{p}_F$ are in fact the only monomial prime ideals in $K[S]$ (for example, see [**BH**, 6.1.7]).

Let $R$ be a commutative noetherian ring. For a prime ideal $\mathfrak{p}$ of $R$ one sets $\operatorname{height} \mathfrak{p} = \dim R_{\mathfrak{p}}$, and for an ideal in general

$$\operatorname{height} I = \min\{\operatorname{height} \mathfrak{p} : \mathfrak{p} \supset I\}.$$

If $R$ is a domain finitely generated over a field, then $\operatorname{height} I = \dim R - \dim R/I$ for all ideals $I$. This equation implies

$$\operatorname{height} \mathfrak{p}_F = \operatorname{rank} S - \dim F.$$

It follows from general principles in the theory of graded rings, that every minimal prime overideal of a monomial ideal $I \subset K[S]$ is itself generated by monomials, and thus is one of the ideals $\mathfrak{p}_F$.

If $F$ is a facet (i. e. a face of dimension equal to $\operatorname{rank} S - 1$), then $\mathfrak{p}_F$ is a height 1 prime ideal. If $S$ is normal, then $\mathfrak{p}_F$ is a divisorial prime ideal, and we will also write (especially in Section 5)

$$\operatorname{Div}(F) \quad \text{for} \quad \mathfrak{p}_F.$$

*Inversion of monomials.* — Let us finally discuss the inversion of monomials. Let $S$ be an affine semigroup embedded into the lattice $L = \operatorname{gp}(S)$. Then $K[S][x^{-1}]$ is again a semigroup algebra, namely $K[S[-x]]$ where $S[-x]$ is the subsemigroup of $L$ generated by $S$ and $-x$.

The structure of $S[-x]$ has an easy description if $S$ is normal. Then

$$S[-x] = \{y \in L : \sigma_i(y) \geqslant 0 \text{ if } \sigma_i(x) = 0,\ i = 1, \dots, s\}$$

where again $\operatorname{supp}(S) = \{\sigma_1 \dots, \sigma_s\}$. In fact, the inclusion $\subset$ is evident, and for the converse one observes that $y + mx \in S$ for $m \gg 0$ whenever all the inequalities hold. Namely, $\sigma_i(y + mx) \geqslant 0$ for all $i$ and $m \gg 0$. It follows that $S[-x]$ is again a normal affine semigroup whose support hyperplanes are those support hyperplanes of $S$ that contain $x$.

We call $x \in S$ an *interior* element (or monomial) if $x$ lies in the interior of the cone $C(S)$; in other words: if $\sigma_i(x) > 0$ for all support forms $\sigma_i$ of $S$. Then $S[-x]$ is just $L$, and $K[S][x^{-1}]$ is the Laurent polynomial ring $K[L]$.

The inversion of an extreme element $x \in S$ is further discussed in Subsection 3.5. (Of course, an element of $S$ is called *extreme* if it belongs to an extreme ray of $C(S)$.)

**2.2. Polytopal semigroup algebras.** — Let $M$ be a subset of $\mathbb{R}^n$. We set

$$L_M = M \cap \mathbb{Z}^n,$$
$$E_M = \{(x,1) \colon x \in L_M\} \subset \mathbb{Z}^{n+1};$$

so $L_M$ is the set of lattice points in $M$, and $E_M$ is the image of $L_M$ under the embedding $\mathbb{R}^n \mapsto \mathbb{R}^{n+1}$, $x \mapsto (x,1)$. Very frequently we will consider $\mathbb{R}^n$ as a hyperplane of $\mathbb{R}^{n+1}$ under this embedding; then we may identify $L_M$ and $E_M$. By $S_M$ we denote the subsemigroup of $\mathbb{Z}^{n+1}$ generated by $E_M$.

Now suppose that $P$ is a (finite convex) lattice polytope in $\mathbb{R}^n$, where 'lattice' means that all the vertices of $P$ belong to the integral lattice $\mathbb{Z}^n$. The affine semigroups of the type $S_P$ will be called *polytopal semigroups*. A lattice polytope $P$ is *normal* if $S_P$ is a normal semigroup. In order to simplify notation we set $C(P) = C(S_P)$.

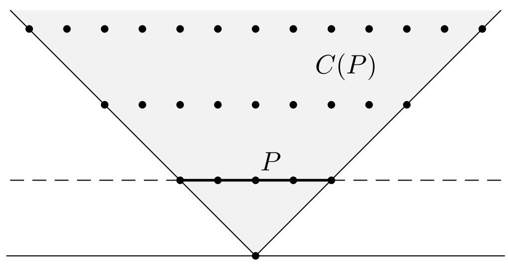


FIGURE 1. Vertical cross-section of a polytopal semigroup

Let $K$ be a field. Then

$$K[P] = K[S_P]$$

is called a *polytopal semigroup algebra* or simply a *polytopal algebra*. Since $\operatorname{rank} S_P = \dim(P) + 1$ and $\dim K[P] = \operatorname{rank} S_P$ as remarked above, we have

$$\dim K[P] = \dim(P) + 1.$$

Note that $S_P$ (or, more generally, $S_M$) is a graded semigroup, i. e. $S_P = \bigcup_{i=0}^{\infty}(S_P)_i$ such that $(S_P)_i + (S_P)_j \subset (S_P)_{i+j}$; its $i$-th graded component $(S_P)_i$ consists of all the elements $(x,i) \in S_P$. Therefore $R = K[P]$ is a graded $K$-algebra in a natural way. Its $i$-th graded component $R_i$ is the $K$-vector space generated by $(S_P)_i$. The elements of $E_P = (S_P)_1$ have degree 1, and therefore $R$ is a homogeneous $K$-algebra in the terminology of Bruns and Herzog [**BH**]. The defining relations of $K[P]$ are the binomials representing the affine dependencies of the lattice points of $P$. Some easy examples:

***Examples 2.2.1***

(a) $P = \operatorname{conv}(1,4) \in \mathbb{R}^1$. (By $\operatorname{conv}(M)$ we denote the convex hull of $M$.) Then $P$ contains the four lattice points $1,2,3,4$, and the relations of the corresponding generators of $K[P]$ are given by

$$X_1X_3 = X_2^2,\ X_1X_4 = X_2X_3,\ X_2X_4 = X_3^2.$$

(b) $P = \operatorname{conv}\big((0,0),(0,1),(1,0),(1,1)\big)$. The lattice points of $P$ are exactly the 4 vertices, and the defining relation of $K[P]$ is $X_1X_4 = X_2X_3$.

(c) $P = \operatorname{conv}\big((1,0),(0,1),(-1,-1)\big)$. There is a fourth lattice point in $P$, namely $(0,0)$, and the defining relation is $X_1X_2X_3 = Y^3$ (in suitable notation).

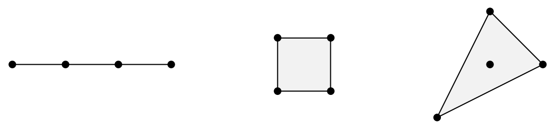

FIGURE 2

Note that the polynomial ring $K[X_1,\dots,X_n]$ is a polytopal algebra, namely $K[\Delta_{n-1}]$ where $\Delta_{n-1}$ denotes the $(n-1)$-dimensional unit simplex.

***Remark 2.2.2***. — If $P$ and $P'$ are two lattice polytopes in $\mathbb{R}^n$ that are integral-affinely equivalent, then $S_P \cong S_{P'}$.

*Integral-affine equivalence* means that $P$ is mapped onto $P'$ by some affine transformation $\psi \in \operatorname{Aff}(\mathbb{R}^n)$ carrying $\mathbb{Z}^n$ onto $\mathbb{Z}^n$. The remark follows from the fact that such an integral-affine transformation of $\mathbb{R}^n$ can be lifted to (a uniquely determined) linear automorphism of $\mathbb{R}^{n+1}$ given by a matrix $\alpha \in \operatorname{GL}_{n+1}(\mathbb{Z})$. (Of course, we understand that $\mathbb{R}^n$ is embedded in $\mathbb{R}^{n+1}$ by the assignment $x \mapsto (x,1)$).

Next we describe the normalization of a semigroup algebra that is 'almost' a polytopal semigroup algebra.

***Proposition 2.2.3***. — *Let $M$ be a finite subset of $\mathbb{Z}^n$. Let $C_M \subset \mathbb{R}^{n+1}$ be the cone generated by $E_M$. Then the normalization of $R = K[S_M]$ is the semigroup algebra $\overline{R} = K[\operatorname{gp}(S_M)\cap C_M]$. Furthermore, with respect to the natural gradings of $R$ and $\overline{R}$, one has $R_1 = \overline{R}_1$ if and only if $M = P \cap \mathbb{Z}^n$ for some lattice polytope $P$.*

*Proof*. — It is an elementary observation that $G\cap C$ is a normal semigroup for every subgroup $G$ of $\mathbb{R}^{n+1}$ and that every element $x \in \operatorname{gp}(S_M)\cap C$ satisfies the condition $cx \in S_M$ for some $c \in \mathbb{N}$.

Consider $\mathbb{R}^n$ as a hyperplane in $\mathbb{R}^{n+1}$ as above. Then the degree 1 elements of $\operatorname{gp}(S_M)\cap C$ are exactly those in the lattice polytope generated by $\operatorname{gp}(S_M)\cap C\cap\mathbb{R}^n$. This implies the second assertion. □

The class of polytopal semigroup algebras can now be characterized in purely ring-theoretic terms.

***Proposition 2.2.4***. — *Let $R$ be a domain. Then $R$ is (isomorphic to) a polytopal semigroup algebra if and only if it has a grading $R = \bigoplus_{i=0}^{\infty} R_i$ such that*

(i) *$K = R_0$ is a field, and $R$ is a $K$-algebra generated by finitely many elements $x_1, \dots, x_m \in R_1$;*

(ii) *the kernel of the natural epimorphism $\varphi\colon K[X_1, \dots, X_m] \to R$, $\varphi(X_i) = x_i$, is generated by binomials $X^{\mathbf{a}} - X^{\mathbf{b}}$ where $X^{\mathbf{a}} = X_1^{a_1} \dots X_m^{a_m}$ for $\mathbf{a} = (a_1, \dots, a_m) \in \mathbb{Z}_+^m$;*

(iii) *$R_1 = \overline{R}_1$ where $\overline{R}$ is the normalization of $R$ (with the grading induced by that of $R$).*

*Proof.* — We have seen above that a polytopal semigroup algebra has properties (i) and (iii). Let $E_M = \{x_1, \dots, x_m\}$. Then the kernel $I_P$ of the natural projection $K[X_1, \dots, X_m] \mapsto K[x_1, \dots, x_m]$, $X_i \mapsto x_i$, is generated by binomials (see Gilmer [**Gi**], §7).

Conversely, a ring with property (ii) is a semigroup algebra over $K$ with semigroup $H$ equal to the quotient of $\mathbb{Z}_+^m$ modulo the congruence relation defined by the pairs $(\mathbf{a}, \mathbf{b})$ associated with the binomial generators of $\operatorname{Ker}\varphi$ ([**Gi**], §7); in particular, $H$ is finitely generated. Since $R$ is a graded domain, $H$ is cancellative and torsion-free, and 0 is its only invertible element. Thus it is a positive affine semigroup and can be embedded in $\mathbb{Z}_+^n$ for a suitable $n$ by its standard embedding. Thus we may consider $x_1, \dots, x_m$ as points of $\mathbb{Z}_+^n$. Set $x_i' = (x_i, 1) \in \mathbb{Z}_+^{n+1}$ and $S$ equal to the semigroup generated by the $x_i'$. We claim that $R$ is isomorphic to $K[S]$. In fact, let $\psi\colon K[X_1, \dots, X_m] \to K[S]$ be the epimorphism given by $\psi(X_i) = x_i'$. We obviously have $\operatorname{Ker}\psi \subset \operatorname{Ker}\varphi$, but the converse inclusion is also true: if $X^{\mathbf{a}} - X^{\mathbf{b}}$ is one of the generators of $\operatorname{Ker}\varphi$, then $X^{\mathbf{a}}$ and $X^{\mathbf{b}}$ have the same total degree, and therefore they are in $\operatorname{Ker}\psi$, too.

Finally it remains to be shown that $x_1', \dots, x_n'$ are exactly the lattice points in the polytope spanned by them. This, however, follows directly from (iii) and 2.2.3 above. □

It is often useful to replace a polytope $P$ by a multiple $cP$ with $c \in \mathbb{N}$. The lattice points in $cP$ can be identified with the lattice points of degree $c$ in the cone $C(S_P)$; in fact, the latter are exactly of the form $(x, c)$ where $x \in L_{cP}$. We quote part of Bruns, Gubeladze and Trung [**BGT1**, 1.3.3]:

***Theorem 2.2.5***. — *Let $P$ be a lattice polytope. Then $cP$ is normal for $c \geqslant \dim P - 1$.*

Polytopal semigroup algebras appear as the coordinate rings of projective toric varieties. We will discuss this connection in Subsection 5.5.

We will indicate in Subsection 3.1 that lattice polytopes of dimension $\leqslant 2$ are always normal. In [**BG2**] the reader can find many concrete examples of normal and non-normal polytopes of dimension 3.

We have started the investigation of polytopal semigroup algebras in our joint paper with Ngo Viet Trung [**BGT1**]. It contains several themes and results mentioned only marginally or not at all in these notes, for example the Koszul property of polytopal semigroup algebras or a detailed investigation of the multiples $cP$.

**2.3. Divisor class groups.** — An extremely useful tool in the exploration of a normal domain $R$ is its divisor class group $\mathrm{Cl}(R)$. For the general theory we refer the reader to Fossum [**Fo**]. In the case of a normal semigroup algebra the computation of the divisor class group is very easy, and the divisor class group carries a great deal of combinatorial information.

Let $R = K[S]$ be a normal affine semigroup algebra. Again we set $\mathrm{supp}(S) = \{\sigma_1, \dots, \sigma_s\}$. Furthermore we let $F_i$ denote the facet of $C(S)$ corresponding to $\sigma_i$ and set $\mathfrak{p}_i = \mathfrak{p}_{F_i}$. As we have seen in Subsection 2.1, the $\mathfrak{p}_i$ are exactly the monomial height 1 prime ideals of $R$.

***Theorem 2.3.1***

(a) *The divisor class group* $\mathrm{Cl}(R)$ *is generated by the classes of the prime ideals* $\mathfrak{p}_1, \dots, \mathfrak{p}_s$.
(b) *Each divisorial ideal of $R$ is isomorphic to an ideal* $\mathfrak{p}_1^{(a_1)} \cap \dots \cap \mathfrak{p}_s^{(a_s)}$, $a_i \in \mathbb{Z}$, $i = 1, \dots, s$.
(c) *The support form $\sigma_i$ extends to the discrete valuation of the quotient field of $R$ associated with the prime ideal $\mathfrak{p}_i$.*
(d) $\mathfrak{p}_1^{(a_1)} \cap \dots \cap \mathfrak{p}_s^{(a_s)}$ *has a $K$-basis by the monomials $x \in L$ such that $\sigma_i(x) \geqslant a_i$ for all $i$.*
(e) $\mathfrak{p}_1^{(a_1)} \cap \dots \cap \mathfrak{p}_s^{(a_s)}$ *and* $\mathfrak{p}_1^{(b_1)} \cap \dots \cap \mathfrak{p}_s^{(b_s)}$ *are isomorphic $R$-modules if and only if there exists $z \in L$ with* $(b_1, \dots, b_s) = \sigma(z) + (a_1, \dots, a_s)$.
(f)
$$\mathrm{Cl}(R) = \mathbb{Z}^s / \sigma(L).$$

*Proof.* — (a) Let $x$ be an interior monomial. As we have seen in Subsection 2.1, the ring $R[x^{-1}]$ is just a Laurent polynomial ring over $K$ and therefore factorial. By Nagata's theorem [**Fo**, 7.1] this implies

$$\mathrm{Cl}(R) = \mathbb{Z}[\mathfrak{p}_1] + \dots + \mathbb{Z}[\mathfrak{p}_s].$$

since $\mathfrak{p}_1, \dots, \mathfrak{p}_s$ are exactly the minimal prime ideals of $x$.

(b) is just a re-statement of (a), since two divisorial ideals belong to the same divisor class if and only if they are isomorphic $R$-modules.

(c) Fix $i$ and let $f \in R$ be an arbitrary element. We write it as a $K$-linear combination of monomials and let $v(f)$ be the minimum over $\sigma_i(x)$ for the monomials $x$ of $f$. It is then easy to check that $v$ is a valuation. It obviously extends $\sigma_i$, and we will write $\sigma_i$ for $v$ in the following.

(d) follows immediately from (c) since $\mathfrak{p}_i^{(a_i)}$ is the ideal of all $f \in R$ such that $\sigma_i(f) \geqslant a_i$.

(e) Two (fractional) monomial ideals $I$ and $J$ are isomorphic if and only if there exists an element $z \in \mathrm{gp}(S)$ with $J = zI$.

(f) This follows immediately from (e). □

The algebra $R$ is the "linearization" (with coefficients in $K$) of the set of solutions to the homogeneous system

$$\sigma_i(x) \geqslant 0$$

of linear diophantine inequalities: its monomial basis is given by the set of solutions. The theorem shows that the divisorial ideals represent the "linearizations" of the associated inhomogeneous systems. We will further pursue this theme in Section 4.

## 3. Covering and normality

**3.1. Introduction.** — In this section we will investigate the question whether the normality of an positive affine semigroup can be characterized in terms of combinatorial conditions on its Hilbert basis. A very natural sufficient condition is (UHC) or *unimodular Hilbert covering*:

(**UHC**) *S is the union (or covered by) the subsemigroups generated by the unimodular subsets of* Hilb($S$).

Here a subset $X$ of a lattice $L$ is called *unimodular* if it is a basis of $L$. In (UHC) $L$ is $\mathrm{gp}(S)$. It is easy to see that (UHC) implies normality:

***Proposition 3.1.1***. — *If $S$ has* (UHC)*, then it is normal. More generally, if $S$ is the union of normal subsemigroups $S_i$ such that* $\mathrm{gp}(S_i) = \mathrm{gp}(S)$*, then $S$ is also normal.*

This follows immediately from the definition of normality (one can also give a relative version in terms of integral closure).

For polytopal semigroups (UHC) has a clear geometric interpretation. Let $P \in \mathbb{R}^n$ be a lattice polytope whose lattice points generate $\mathbb{Z}^n$ affinely (that is, for some (and therefore every) $x_0 \in P \cap \mathbb{Z}^n$ the differences $x - x_0$, $x \in P \cap \mathbb{Z}^n$, generate the lattice $\mathbb{Z}^n$). This is no essential restriction, since we can shrink the lattice if necessary. Then a subset $X$ of Hilb($S$) is unimodular if and only if the corresponding lattice points of $P$ generate $\mathbb{Z}^n$ affinely, or, equivalently, the simplex spanned by them has the smallest possible Euclidean volume $1/n!$, or *normalized volume* 1. Such simplices are likewise called *unimodular*.

Below we will frequently use the fact that any lattice polytope admits a triangulation into empty lattice simplices: a lattice polytope $P \subset \mathbb{R}^n$ is *empty*, if $P \cap \mathbb{Z}^n$ consists exactly of the vertices of $P$.

Lattice *polygons*, i. e. lattice polytopes of dimension $\leqslant 2$ can even be triangulated into unimodular lattice simplices, since a lattice simplex of dimension $\leqslant 2$ that contains no lattice points other than its vertices is necessarily unimodular, as follows from Pick's theorem. Thus polytopes of dimension $\leqslant 2$ are automatically normal. A natural question: *If $P$ is normal, is it covered by unimodular lattice simplices?*

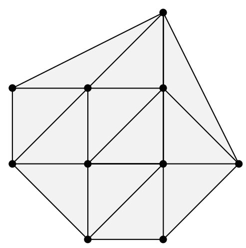

FIGURE 3. Triangulation of a lattice polygon

Since normal affine semigroups are exactly of the form $S(C)$ for finitely generated rational cones $C$, (UHC) has an interpretation in terms of discrete geometry also in the general case, namely: *Is a finitely generated rational cone $C$ covered by the unimodular simplicial subcones that are generated by subsets of* $\operatorname{Hilb}(S(C))$*?*

As a conjecture, (UHC) appears first in Sebö [**Se**, Conjecture B]. We will present a 6-dimensional counterexample to Sebö's conjecture in Subsection 3.6, in which we also describe an algorithm deciding (UHC).

The major positive result supporting (UHC) had been shown by Sebö [**Se**] and, independently, by Aguzzoli and Mundici [**AM**] and Bouvier and Gonzalez-Sprinberg [**BoGo**]: every 3-dimensional rational cone admits a *triangulation* (or *partition*) into unimodular simplicial subcones generated by elements of $\operatorname{Hilb}(C)$. We will discuss this result in Subsection 3.3. (For the cones $C(S_P)$, $P$ a lattice polytope of dimension $\leqslant 2$, this has been indicated above.) This is, of course, a much stronger property than (UHC). However, [**BoGo**] also describes a 4-dimensional cone without such a triangulation.

For algebraic geometry triangulations into simplicial subcones (whose one-dimensional faces are not necessarily spanned by elements of $\operatorname{Hilb}(C)$) are important for the construction of equivariant desingularizations of toric varieties (see [**Oda**]).

Triangulations also provide the connection between discrete geometry and Gröbner bases of the binomial ideal defining a semigroup algebra; see Sturmfels [**Stu**] for this important and interesting theme.

Another positive result in the polytopal case has been proved in [**BGT1**, 1.3.1] and is reproduced in Subsection 3.4: the homothetic multiple $cP$ satisfies (UHC) for $c \gg 0$, regardless of $\dim P$. (It is even known that $dP$ has a triangulation into unimodular simplices for *some* $d$, but the question whether such a triangulation exists for all sufficiently large $d$ seems to be open; see Kempf, Knudsen, Mumford, and

Saint–Donat [**KKMS**].) For elementary reasons one can take $c = 1$ in dimension 1 and 2, and it was communicated by Ziegler that $c = 2$ suffices in dimension 3; see Kantor and Sarkaria [**KS**] where it shown that $4P$ has a unimodular triangulation for all 3-dimensional lattice polytopes. However, in higher dimension no effective lower bound for $c$ seems to be known. (In contrast, $cP$ is normal for $c \geqslant \dim P - 1$; see Theorem 2.2.5 .) Our counterexample to (UHC) is in fact a normal semigroup of type $S_P$ where $P$ is a 5-dimensional lattice polytope. Thus the question about the unimodular covering of normal polytopes has a negative answer.

A natural variant of (UHC), and weaker than (UHC), is the existence of a *free Hilbert cover*:

(**FHC**) *$S$ is the union (or covered by) the subsemigroups generated by the linearly independent subsets of* $\operatorname{Hilb}(S)$.

For (FHC) – in contrast to (UHC) – it is not evident that it implies the normality of the semigroup. Nevertheless it does so, as we will see in Subsection 3.7. A formally weaker – and certainly the most elementary – property is the *integral Carathéodory property*:

(**ICP**) *Every element of $S$ has a representation $x = a_1s_1 + \cdots + a_ms_m$ with $a_i \in \mathbb{Z}_+$, $s_i \in \operatorname{Hilb}(C)$, and $m \leqslant \operatorname{rank} S$.*

Here we have borrowed the well-motivated terminology of Firla and Ziegler [**FZ**]: (ICP) is obviously a discrete variant of Carathéodory's theorem for convex cones. It was first asked in Cook, Fonlupt, and Schrijver [**CFS**] whether all cones have (ICP) and then conjectured in [**Se**, Conjecture A] that the answer is 'yes'.

In joint work with M. Henk, A. Martin and R. Weismantel it has been shown that our counterexample to (UHC) also disproves (ICP) (see [**BGHMW**]). Thus none of the covering properties above is necessary for the normality of affine semigroups.

Later on we will use the *representation length*

$$\rho(x) = \min\{m \mid x = a_1s_1 + \cdots + a_ms_m,\ a_i \in \mathbb{Z}_+,\ s_i \in \operatorname{Hilb}(S)\}$$

for an element $x$ of an affine semigroup $S$. If $\rho(x) \leqslant m$, we also say that $x$ is *$m$-represented*. In order to measure the deviation of $S$ from (ICP), we introduce the notion of *Carathéodory rank* of an affine semigroup $S$,

$$\operatorname{CR}(S) = \max\{\rho(x) \mid x \in S\}.$$

In [**BG3**] we treat some variants of this notion, called asymptotic and virtual Carathéodory rank. See also [**BGT2**], where algorithms for the computation of these Carathéodory ranks (for arbitrary $S$) have been developed.

A short introduction to the theme of this section has been given in Bruns [**Bru**].

**3.2. An upper bound for Carathéodory rank.** — Let $p_1, \ldots, p_n$ be different prime numbers, and set $q_i = \prod_{i \neq j} p_i$. Let $S$ be the subsemigroup of $\mathbb{Z}_+$ generated by $q_1, \ldots, q_n$. Since $\gcd(q_1, \ldots, q_n) = 1$, there exists an $m \in \mathbb{Z}_+$ with $u \in S$ for all $u \geqslant m$. Choose $u \geqslant m$ such that $u$ is not divisible by $p_i$, $i = 1 \ldots, n$. Then all the $q_i$ must be involved in the representation of $u$ by elements of $\operatorname{Hilb}(S)$. This example shows that there is no bound of $\operatorname{CR}(S)$ in terms of $\operatorname{rank} S$ without further conditions on $S$.

For normal $S$ there is a linear bound for $\operatorname{CR}(S)$ as given by Sebö [**Se**]:

***Theorem 3.2.1***. — *Let $S$ be a normal positive affine semigroup of rank $\geqslant 2$. Then* $\operatorname{CR}(S) \leqslant 2(\operatorname{rank}(S) - 1)$.

For the proof we denote by $C'(S)$ the convex hull of $S \smallsetminus \{0\}$ (in $\operatorname{gp}(S) \otimes \mathbb{R}$). Then we define the *bottom* $B(S)$ of $C'(S)$ by

$$B(S) = \bigl\{x \in C'(S) : [0,x] \cap C'(S) = \{x\}\bigr\}$$

($[0,x] = \operatorname{conv}(0,x)$ is the line segment joining 0 and $x$). In other words, the bottom is exactly the set of points of $C'(S)$ that are *visible* from 0 (see Figure 4).

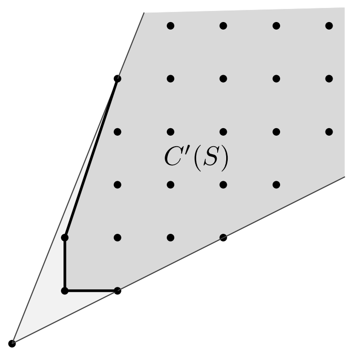


FIGURE 4. The bottom

***Lemma 3.2.2***

(a) *Let $H$ be a support hyperplane of $C'(S)$. Then $H \cap C'(S)$ is compact if and only if $0 \notin H$. The non-compact facets of $C'(S)$ are the intersections $C'(S) \cap G$ where $G$ is a support hyperplane of $C(S)$.*

(b) *Let $F$ be a compact facet of $C'(S)$. Then $F = \operatorname{conv}(\operatorname{Hilb}(S) \cap F)$. In particular, $C'(S)$ has only finitely many (compact) facets.*

(c) *$B(S)$ is the union of the compact facets of $C'(S)$.*

(d) *$B(S) \cap S \subset \operatorname{Hilb}(S)$.*

*Proof.* — (a) Set $F = H \cap C'(S)$. Clearly $ax \in C'(S)$ for every $x \in C'(S)$ and $a \in \mathbb{R}$, $a > 1$. Therefore $F$ cannot be compact if $0 \in H$.

Conversely, suppose that $0 \notin H$. Then we choose a linear form $\gamma$ with $H = \{x : \gamma(x) = a\}$ for some $a \in \mathbb{R}$ and $\gamma(x) \geqslant a$ for $x \in C'(S)$. By hypothesis $a \neq 0$, and it follows that $a > 0$. Every $y \in F$ is a linear combination $y = \sum_{i=1}^{m} b_i z_i$ with $z_1, \dots, z_m \in S$, $a_1, \dots, a_m \geqslant 0$, and $\sum_{i=1}^{m} b_i = 1$. It follows immediately that $z_1, \dots, z_m \in H$, and furthermore that $z_1, \dots, z_m \in \mathrm{Hilb}(S)$. Thus $F = H \cap \mathrm{conv}(\mathrm{Hilb}(S))$ is compact.

Clearly if $H$ is a support hyperplane intersecting $C'(S)$ in a facet and containing 0, then it also a support hyperplane of $C(S)$ intersecting $C(S)$ in a facet.

(b) has just been proved, and (c) and (d) are now obvious. □

Let $H$ be a support hyperplane intersecting $C'(S)$ in a compact facet. Then there exists a unique primitive $\mathbb{Z}$-linear form $\gamma$ on $\mathrm{gp}(S)$ such that $\gamma(x) = a > 0$ for all $x \in H$ (after the extension of $\gamma$ to $\mathrm{gp}(S) \otimes \mathbb{R}$). Since $\mathrm{Hilb}(S) \cap H \neq \varnothing$, one has $a \in \mathbb{Z}$. We call $\gamma$ the *basic grading* of $S$ associated with the facet $H \cap C'(S)$ of $C'(S)$.

*Proof of Theorem 3.2.1.* — As we have seen above, the bottom of $S$ is the union of finitely many lattice polytopes $F$, all of whose lattice points belong to $\mathrm{Hilb}(S)$. We now triangulate each $F$ into empty lattice subsimplices. Choose $x \in S$, and consider the line segment $[0, x]$. It intersects the bottom of $S$ in a point $y$ belonging to some simplex $\sigma$ appearing in the triangulation of a compact facet $F$ of $C'(S)$. Let $z_1, \dots, z_n \in \mathrm{Hilb}(S)$, $n = \mathrm{rank}(S)$, be the vertices of $\sigma$. Then we have

$$x = (a_1 z_1 + \cdots + a_n z_n) + (q_1 z_1 + \cdots + q_n z_n), \qquad a_i \in \mathbb{Z}_+,\ q_i \in \mathbb{Q},\ 0 \leqslant q_i < 1,$$

as in the proof of Gordan's lemma. Set $x' = \sum_{i=1}^{n} q_i z_i$, let $\gamma$ be the basic grading of $S$ associated with $F$, and $a = \gamma(y)$ for $y \in F$. Then $\gamma(x') < na$, and at most $n - 1$ elements of $\mathrm{Hilb}(S)$ can appear in a representation of $x'$. This shows that $\mathrm{CR}(S) \leqslant 2n - 1$.

However, this bound can be improved. Set $x'' = z_1 + \cdots + z_n - x'$. Then $x'' \in S$, and it even belongs to the cone generated by $z_1, \dots, z_n$. If $\gamma(x'') < a$, one has $x'' = 0$. If $\gamma(x'') = a$, then $x''$ is a lattice point of $\sigma$. By the choice of the triangulation this is only possible if $x'' = x_i$ for some $i$, a contradiction. Therefore $\gamma(x'') > a$, and so $\gamma(x') < (n-1)a$. It follows that $\mathrm{CR}(S) \leqslant 2n - 2$. □

The symmetry argument on which the improvement by 1 is based is especially useful in low dimensions, as we will see in the next subsection.

In view of Theorem 3.2.1 it makes sense to set

$$\mathcal{CR}(n) = \max\bigl\{\mathrm{CR}(S) : S \text{ is normal positive and } \mathrm{rank}\, S = n\bigr\}.$$

With this notion we can reformulate Theorem 3.2.1 as $\mathcal{CR}(n) \leqslant 2(\mathrm{rank}(S) - 1)$. On the other hand, the counterexample $S_6$ to (ICP) presented in Subsection 3.6 implies that

$$\mathcal{CR}(n) \geqslant \left\lfloor \frac{7}{6}n \right\rfloor .$$

In fact, $\operatorname{rank} S_6 = 6$ and $\operatorname{CR}(S_6) = 7$. Therefore suitable direct sums $S_6 \oplus \cdots \oplus S_6 \oplus \mathbb{Z}_+^p$ attain the lower bound just stated.

An improvement of both the upper and the lower bound for $\mathcal{CR}(n)$ would be very interesting. It certainly requires a better understanding of Hilbert bases.

**3.3. Dimensions 1,2,3.** — Let $x_1, \dots, x_n$ be linearly independent elements of $\mathbb{Z}^n$ and let $C$ be the cone spanned by them. Then each $y \in S = S(C)$ has a representation

$$y = (a_1x_1 + \cdots + a_nx_n) + (q_1x_1 + \cdots + q_nx_n), \qquad a_i \in \mathbb{Z}_+,\ q_i \in \mathbb{Q},\ 0 \leqslant q_i < 1.$$

Following Sebö we collect the second summands in the set

$$\operatorname{par}(x_1, \dots, x_n) = \mathbb{Z}^n \cap \{q_1x_1 + \cdots + q_nx_n : q_i \in \mathbb{Q},\ 0 \leqslant q_i < 1\}.$$

The notation par is suggested by the fact that its elements are exactly the lattice points in the semi-open parallelepiped spanned by $x_1, \dots, x_n$.

***Lemma 3.3.1.*** — *The set* $\operatorname{par}(x_1, \dots, x_n)$ *contains exactly one representative from each residue class of* $\mathbb{Z}^n$ *modulo* $U = \mathbb{Z}x_1 + \cdots + \mathbb{Z}x_n$. *Therefore*

$$\#\operatorname{par}(x_1, \dots, x_n) = \#(\mathbb{Z}^n/U) = |\det(x_1, \dots, x_n)|.$$

*Proof.* — The first statement is evident and it implies the first equation. The second equation results from the elementary divisor theorem. □

***Remark 3.3.2.*** — Clearly $\operatorname{Hilb}(S) \subset \{x_1, \dots, x_n\} \cup \operatorname{par}(x_1, \dots, x_n)$ (with the notation above). This is used in [**BK**] for an algorithm computing Hilbert bases. A cone generated by elements $y_1, \dots, y_m$ is first triangulated into simplicial subcones spanned by linearly independent elements $x_1, \dots, x_n \in \{y_1, \dots, y_m\}$. For each of the subcones the set $\operatorname{par}(x_1, \dots, x_n)$ is formed, and from their union and $\{y_1, \dots, y_m\}$ the Hilbert basis is selected by checking irreducibility.

A positive affine semigroup of rank 1, for which we can assume that $\operatorname{gp}(S) = \mathbb{Z}$, is either contained in $\mathbb{Z}_-$ or $\mathbb{Z}_+$. If it is normal, then it must contain $-1$ or $1$, so that $S \cong \mathbb{Z}_+$ is free.

In dimension 2 the situation is still very simple:

***Proposition 3.3.3.*** — *Let* $S \subset \mathbb{Z}^2 = \operatorname{gp}(S)$ *be a positive affine semigroup of rank* 2. *Then* $\operatorname{Hilb}(S) = S \cap B(S)$, *and* $C(S)$ *has a (uniquely determined) unimodular Hilbert triangulation.*

*Proof.* — The bottom $B(S)$ is a broken line. It has exactly one triangulation into empty lattice line segments. It is enough to show that the endpoints $x, y$ of each of the line segments are a basis of $\mathbb{Z}^2$. By Lemma 3.3.1 this is equivalent to $\#\operatorname{par}(x, y) = 1$.

Suppose that $z \in \operatorname{par}(x, y)$, $z \neq 0$. Then $z = ax + by$ with $a, b \in \mathbb{Q}$, $0 < a, b < 1$, and $x + y - z \in \operatorname{par}(x, y)$ as well. However, one of the points $z$ or $x + y - z$ must lie in the interior of the simplex $\operatorname{conv}(0, x, y)$ or the interior of the line segment $[x, y]$. This is impossible since $x$ and $y$ span an empty line segment in the bottom of $S$. □

Before we consider dimension 3, let us observe that it is always possible to triangulate a cone generated by finitely many vectors $x \in \mathbb{Z}^n$ into simplicial subcones each of which is spanned by a basis of $\mathbb{Z}^n$. Let $y_1, \dots, y_m \in \mathbb{Z}^n$ generate the cone $C$. Then we first triangulate $C$ into simplicial subcones $\sigma$ each of which is spanned by a linearly independent subset $\{x_1, \dots, x_n\}$ of $\{y_1, \dots, y_m\}$. If $x_1, \dots, x_n$ is not a basis of $\mathbb{Z}^n$, we choose an element $z \in \operatorname{par}(x_1, \dots, x_n)$, $z \neq 0$, and replace $\sigma$ by the union of the subcones spanned by

$$M_i = \{x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n\}, \quad q_i \neq 0, \qquad i = 1, \dots, n,$$

where $z = q_1 x_1 + \dots + q_n x_n$. One has

$$|\det(M_i)| = q_i |\det(x_1, \dots, x_n)| < |\det(x_1, \dots, x_n)|$$

If $q_i = 0$ for some $i$, then $z$ may belong to another simplicial subcone $\sigma' \neq \sigma$. But $\sigma'$ can be subdivided by $z$ as well so that the subdivisions coincide on $\sigma \cap \sigma'$.

This subdivision procedure must stop after finitely many steps, since in each step a simplicial subcone is replaced by the union of strictly "smaller" subcones. We set

$$\operatorname{sdiv}(x_1, \dots, x_n) = \operatorname{par}(x_1, \dots, x_n) \setminus \{0\}.$$

In general one cannot achieve that all vectors $z$ used in the subdivision algorithm belong to $\operatorname{Hilb}(S)$, $S = S(C)$. As mentioned already, there is a counterexample in dimension 4 [**BoGo**].

However, in dimension 3 the elements of $\operatorname{Hilb}(S)$ suffice for the subdivision. We first describe the Hilbert basis in dimension 3.

***Proposition 3.3.4***. — *Let $S$ be a positive normal affine semigroup and $x \in S$, $x \neq 0$. If*

$$\gamma(x) < 2 \min\{\gamma(y) : y \in S\}$$

*for some basic grading $\gamma$ of $S$, then $x \in \operatorname{Hilb}(S)$. If $\operatorname{rank}(S) \leqslant 3$, this condition is also necessary for $x \in \operatorname{Hilb}(S)$.*

*Proof*. — The sufficiency of the condition is trivial. Suppose that $\operatorname{rank} S = 3$ and choose $x \in S$. The line segment $[0, x]$ meets the bottom of $S$ in one of its facets $F$. We triangulate $F$ into empty lattice subsimplices. Then $[0, x]$ meets one of the triangles $\sigma$, say $\sigma = \operatorname{conv}(x_1, x_2, x_3)$, and $x = a_1 x_1 + a_2 x_2 + a_3 x_3 + x'$ with $a_i \in \mathbb{Z}_+$ and $x' \in \operatorname{par}(x_1, x_2, x_3)$.

Let $\gamma$ be the basic grading associated with $F$. Then

$$\gamma(x_i) = \min\{\gamma(y) : y \in S\} = a$$

Clearly $\gamma(x') < 3a$. It is enough to show that $x' = 0$ or $\gamma(x') < 2a$. But this follows from the symmetry argument applied for the proof of Theorem 3.2.1: If $x' \neq 0$ and $\gamma(x') \geqslant 2a$, then $0 < \gamma(x_1+x_2+x_3-x') \leqslant a$. This would imply $x' \in \operatorname{conv}(0, x_1, x_2, x_3)$, and the 4 vertices are the only lattice points in this tetrahedron. □

The previous proof contains a very useful observation: *for a rank 3 positive normal semigroup $S$ one has* $\operatorname{sdiv}(x_1, x_2, x_3) \subset \operatorname{Hilb}(S)$ *if* $\operatorname{conv}(0, x_1, x_2, x_3)$ *is empty and* $x_1, x_2, x_3$ *belong to the same facet of the bottom of $S$.*

Therefore there is no problem in the first subdivision step. However, in order to really achieve a Hilbert triangulation of $C = C(S)$, we must guarantee that the further subdividing vectors also belong to $\operatorname{Hilb}(S)$.

***Lemma 3.3.5***. — *Let $x_1, x_2, x_3 \in \mathbb{Z}^3$ be linearly independent vectors that do not form a basis of $\mathbb{Z}^3$ and suppose that* $\operatorname{conv}(0, x_1, x_2, x_3)$ *is an empty tetrahedron. Then there is $y \in \operatorname{sdiv}(x_1, x_2, x_3)$ such that*

$$\operatorname{sdiv}(x_1, x_2, x_3) = \{y\} \cup \operatorname{sdiv}(y, x_2, x_3) \cup \operatorname{sdiv}(x_1, y, x_3) \cup \operatorname{sdiv}(x_1, x_2, y)$$

*and all the tetrahedra* $\operatorname{conv}(0, y, x_2, x_3)$, $\operatorname{conv}(0, x_1, y, x_3)$, $\operatorname{conv}(0, x_1, x_2, y)$ *are empty and of dimension* 3.

Together with our observation above this lemma completes the proof of Sebö's

***Theorem 3.3.6***. — *Let $S$ be a positive normal semigroup of rank* 3. *Then $C(S)$ has a unimodular Hilbert triangulation.*

*Proof of Lemma 3.3.5.* — Let $C$ be the cone spanned by $x_1, x_2, x_3$ and $S = S(C)$. For $y \in \operatorname{sdiv}(x_1, x_2, x_3)$ let $C_1$ be the cone generated by $y, x_2, x_3$, $S_1 = S(C_1)$, and define $S_2$ and $S_3$ analogously. By symmetry arguments $y$ can not belong to any of the facets of the cone spanned by the tetrahedron $\operatorname{conv}(0, x_1, x_2, x_3)$ at its vertex 0. In other words the cones $C_1$, $C_2$ and $C_3$ are nondegenerate.

We have $S = S_1 \cup S_2 \cup S_3$, and

$$\operatorname{Hilb}(S) \subset \operatorname{Hilb}(S_1) \cup \operatorname{Hilb}(S_2) \cup \operatorname{Hilb}(S_3).$$

By the observation above,

$$\operatorname{Hilb}(S) = \{x_1, x_2, x_3\} \cup \operatorname{sdiv}(x_1, x_2, x_3\}$$

and it is also clear that $\operatorname{Hilb}(S_1) \subset \{y, x_2, x_3\} \cup \operatorname{sdiv}(y, x_2, x_3)$ etc. Thus

$$(*) \qquad \operatorname{Hilb}(S) \subset \{x_1, x_2, x_3, y\} \cup \operatorname{sdiv}(y, x_2, x_3) \cup \operatorname{sdiv}(x_1, y, x_3) \cup \operatorname{sdiv}(x_1, x_2, y).$$

We set $\delta = |\det(x_1, x_2, x_3)|$, $\delta_1 = |\det(y, x_2, x_3)|$, and define $\delta_2$ and $\delta_3$ accordingly. By the next lemma we can choose $y$ such that

$$\delta_1 + \delta_2 + \delta_3 = \delta + 1.$$

Since $\#\operatorname{sdiv}(x_1, x_2, x_3) = \delta - 1$ etc. (by Lemma 3.3.1), $\operatorname{Hilb}(S)$ has $\delta + 2$ elements, whereas the set on the righthand side in $(*)$ can have at most $\delta + 2$ elements. Thus the containment relation implies first that the sets are equal and, second, that the sets on the right hand side are disjoint. Now

$$\operatorname{sdiv}(x_1, x_2, x_3) = \{y\} \cup \operatorname{sdiv}(y, x_2, x_3) \cup \operatorname{sdiv}(x_1, y, x_3) \cup \operatorname{sdiv}(x_1, x_2, y)$$

follows immediately.

The remaining claim is part of the next lemma. □

***Lemma 3.3.7***. — *Let $x_1, x_2, x_3 \in \mathbb{Z}^3$ be linearly independent vectors that do not form a basis of $\mathbb{Z}^3$ and suppose that $\operatorname{conv}(0, x_1, x_2, x_3)$ is an empty tetrahedron. Then there is $y \in \operatorname{sdiv}(x_1, x_2, x_3)$ such that (with the notation of the previous proof)*

$$\delta_1 + \delta_2 + \delta_3 = \delta + 1,$$

*and all the tetrahedra $\operatorname{conv}(0, y, x_2, x_3))$, $\operatorname{conv}(0, x_1, y, x_3)$, $\operatorname{conv}(0, x_1, x_2, y)$ are empty and non-degenerate.*

*Proof*. — For $y \in \mathbb{R}^3$, $y = q_1x_1 + q_2x_2 + q_3x_3$, we set

$$s(y) = q_1 + q_2 + q_3.$$

For $y \in \operatorname{sdiv}(x_1, x_2, x_3)$ one then has $\delta_1 + \delta_2 + \delta_3 = \delta s(y)$ and $1 < s(y) < 2$ (since $\operatorname{conv}(0, x_1, x_2, x_3)$ is empty and by the symmetry argument). In particular, $s(y)$ is not an integer.

By Cramer's rule each $q_i$ can be written as a quotient $a_i/\delta$, $a_i \in \mathbb{Z}_+$. Therefore $\delta s(y)$ can only take one of the $\delta - 1$ values

$$\delta + 1, \dots, 2\delta - 1.$$

Since $\operatorname{sdiv}(x_1, x_2, x_3)$ contains exactly $\delta - 1$ elements, it is enough to show that the $s(y)$, $y \in \operatorname{sdiv}(x_1, x_2, x_3)$, are pairwise different.

Suppose that $s(y) = s(y')$ and set $t = y - y'$. Then $s(t) = 0$. There is a unique representation $t = a_1x_1 + a_2x_2 + a_3x_3 + t'$ with $a_i \in \mathbb{Z}$ and $t' \in \operatorname{par}(x_1, x_2, x_3)$. Since $s(t) \in \mathbb{Z}$, we also have $s(t') \in \mathbb{Z}$, excluding $t' \in \operatorname{sdiv}(x_1, x_2, x_3)$ (as observed above), and so $t' = 0$. This implies that $y$ and $y'$ have the same residue class modulo $\mathbb{Z}x_1 + \mathbb{Z}x_2 + \mathbb{Z}x_3$. This is impossible for $y, y' \in \operatorname{par}(x_1, x_2, x_3)$, unless $y = y'$.

To sum up: we can choose $y$ such that $\delta s(y) = \delta + 1$. That $\operatorname{conv}(0, y, x_2, x_3)$ is empty and non-degenerate, is now easily seen. In fact, $y, x_2, x_3$ are linearly independent, and every point $z$ in $\operatorname{conv}(0, y, x_2, x_3)$ has $s(z) \leqslant s(y)$. But the only lattice points in $\operatorname{conv}(0, x_1, x_2, x_3, y)$ with this property are $0, x_1, x_2, x_3, y$. □

The proof shows that the linear form $\alpha = \delta s$ has the following property: $\alpha(x_i) = \delta$, $i = 1, 2, 3$, and $x \in \mathbb{Z}^3$ belongs to $U = \mathbb{Z}x_1 + \mathbb{Z}x_2 + \mathbb{Z}x_3$ if and only if $\alpha(x) \equiv 0 \ (\delta)$. In particular, $\mathbb{Z}^3/U$ is cyclic, and $\alpha$ separates the residue classes modulo $U$. This is the crucial point.

From a more algebraic perspective it can also be shown as follows. Let $H$ be the vector subspace generated by $x_1 - x_2, x_1 - x_3$. Since the triangle $\operatorname{conv}(x_1, x_2, x_3)$ is empty, this holds as well for $\operatorname{conv}(0, x_1 - x_2, x_1 - x_3)$, and so $x_1 - x_2, x_1 - x_3$ is a basis of $V = \mathbb{Z}^3 \cap H$ (compare the proof of Proposition 3.3.3). Clearly $V$ is a direct summand of $\mathbb{Z}^3$, and $\mathbb{Z}^3/V \cong \mathbb{Z}$. Since $V \subset U$, it follows that $\mathbb{Z}^3/U$ is also cyclic and that there is a unique primitive linear form $\alpha : \mathbb{Z}^3 \to \mathbb{Z}$ such that $V = \operatorname{Ker}\alpha$, and $\alpha(x_1) = \alpha(x_2) = \alpha(x_3) = \delta$. Then $x \in U$ if and only if $\alpha(x) \equiv 0 \ (\delta)$.

**3.4. Unimodular covering of high multiples of polytopes.** — The counterexample discussed in Subsection 3.6 shows that a normal lattice polytope need not be covered by its unimodular lattice subsimplices. However, this always holds for a sufficiently high multiple of $P$ [**BGT1**]:

***Theorem 3.4.1***. — *For every lattice polytope $P$ there exists $c_0 > 0$ such that $cP$ is covered by its unimodular lattice subsimplices (and, hence, is normal by Proposition 3.1.1) for all $c \in \mathbb{N}$, $c > c_0$.*

*Proof.* — We have observed in the previous section that any finitely generated rational cone in $\mathbb{R}^n$ admits a finite subdivision into simplicial cones $C_i$ each of which is generated by a basis of $\mathbb{Z}^n$.

Now let $P$ be a polytope of dimension $n$, and let $v$ be an arbitrary vertex of $P$. Since the properties of $P$ we are dealing with are invariant under integral-affine transformations, we can assume $v = 0 \in \mathbb{Z}^n$. Let $C$ be the cone in $\mathbb{R}^n$ spanned by 0 as its apex and $P$ itself. Let $C = \bigcup_i C_i$ be a subdivision into simplicial cones $C_i$ as above. So the edges of $C_i$ for each $i$ are determined by the radial directions of some basis $\{e_{i1}, \dots, e_{in}\}$ of $\mathbb{Z}^n$. Denote by $\square_i$ the parallelepiped in $\mathbb{R}^n$ spanned by the vectors $e_{i1}, \dots, e_{in} \subset \mathbb{R}^n$. Thus $\operatorname{vol}(\square_i) = 1$ for all $i$. Equivalently, $\square_i \cap \mathbb{Z}^n$ coincides with the vertex set of $\square_i$. Clearly, each of the $C_i$ is covered by parallel translations of $\square_i$ (precisely as $\mathbb{R}^n_+$ is covered by parallel translations of the standard unit $n$-cube).

For each $i$ and each $c \in \mathbb{N}$ let $Q_{ic}$ be the union of the parallel translations of $\square_i$ inside $C_i \cap cP$. Evidently, $Q_{ic}$ is not convex in general. By $c^{-1}Q_{ic}$ we denote the homothetic image of $Q_{ic}$ centered at $v = 0$ with factor $c^{-1}$. The detailed verification of the following claim is left to the reader.

**Claim.** *Let $F_v^{\mathrm{op}}$ denote the union of all the facets of $P$ not containing $v$ (i. e. 0 in our case). Then for any real $\varepsilon > 0$ there exists $c_\varepsilon \in \mathbb{N}$ such that*

$$P \smallsetminus U_\varepsilon(F_v^{\mathrm{op}}) \subset \bigcup_i c^{-1}Q_{ic}$$

*whenever $c > c_\varepsilon$ ($U_\varepsilon(F_v^{\mathrm{op}})$ denotes the $\varepsilon$-neighbourhood of $F_v^{\mathrm{op}}$ in $\mathbb{R}^n$).*

Let us just remark that the crucial point in showing this inclusion is that the covering of each $C_i$ by parallel translations of the $c^{-1}\square_i$ becomes finer in the appropriate sense when $c$ tends to $\infty$. (The finiteness of the collection $\{C_i\}$ is of course essential).

For an arbitrary vertex $w$ of $P$ we define $F_w^{\mathrm{op}}$ analogously.

**Claim.** *There exists $\varepsilon > 0$ such that*

$$\bigcap_w U_\varepsilon(F_w^{\mathrm{op}}) = \varnothing,$$

*where $w$ runs over all vertices of $P$.*

Indeed, first one easily observes that

$$\bigcap_w U_\varepsilon(F_w^{\mathrm{op}}) = \bigcap_F U_\varepsilon(F),$$

where on the right hand side $F$ ranges over the set of facets of $P$, while $U_\varepsilon(F)$ is the $\varepsilon$-neighbourhood of $F$, and then one completes the proof as follows. Consider the function

$$d\colon P \longrightarrow \mathbb{R}_+, \qquad d(x) = \max(\mathrm{dist}(x, F)),$$

where $F$ ranges over the facets of $P$ and $\mathrm{dist}(x, F)$ stands for the (Euclidean) distance from $x$ to $F$. The function $d$ is continuous and strictly positive. So, by the compactness of $P$, it attains its minimal value at some $x_0 \in P$. Now it is enough to choose $\varepsilon < d(x_0)$.

Summing up the two claims, one is directly lead to the conclusion that, for $c \in \mathbb{N}$ sufficiently large, $cP$ is covered by lattice $n$-parallelepipeds which are integral-affinely equivalent to the standard unit cube, i. e. they have volume 1. Now the proof of our theorem is finished by the well-known fact that the standard unit cube has a unimodular triangulation (this is well-known; see [**BGT1**] for a detailed treatment.) □

The algebraic properties of the polytopal semigroup algebras $K[cP]$ have been studied in [**BGT1**].

**3.5. Tight cones.** — In this subsection we introduce the class of tight cones and semigroups and show that they play a crucial rôle for (UHC) and the other covering properties.

***Definition 3.5.1.*** — Let $S$ be a normal affine semigroup, $x \in \mathrm{Hilb}(S)$, and $S'$ the semigroup generated by $\mathrm{Hilb}(S) \setminus \{x\}$. We say that $x$ is *non-destructive* if $S'$ is normal and $\mathrm{gp}(S')$ is a direct summand of $\mathrm{gp}(S)$ (and therefore equal to $\mathrm{gp}(S)$ if $\mathrm{rank}\,\mathrm{gp}(S) = \mathrm{rank}\,\mathrm{gp}(S')$). Otherwise $x$ is *destructive*. We say that $S$ is *tight* if every element of $\mathrm{Hilb}(S)$ is destructive. A cone $C$ is *tight* if $S(C)$ is tight.

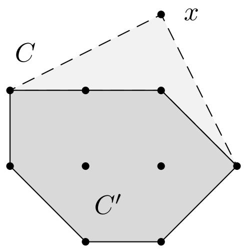


FIGURE 5. Tightening a cone

It is clear that only extreme elements of $\operatorname{Hilb}(S)$ can be non-destructive. Suppose that $x$ is an extreme element of $\operatorname{Hilb}(S)$. Then $S[-x]$ (the subsemigroup of $\operatorname{gp}(S)$ generated by $S$ and $-x$) splits into a product $\mathbb{Z}x \oplus S_x$ where $S_x$ is again a positive normal affine semigroup (for example, see Gubeladze [**Gu0**, Theorem 1.8]). As a consequence one has $C(S)[-x] \cong \mathbb{R} \oplus C(S_x)$.

***Lemma 3.5.2***. — *Let $S$ be a normal positive affine semigroup and $x \in \operatorname{Hilb}(S)$ a non-destructive element. Let $S'$ be the semigroup generated by $\operatorname{Hilb}(S) \smallsetminus \{x\}$ and $S_x$ the quotient $S[-x]/(\mathbb{Z}x)$ introduced above.*

(i) *If $S'$ and $S_x$ both satisfy* (UHC)*, then so does $S$.*
(ii) *If $S'$ and $S_x$ both satisfy* (FHC)*, then so does $S$.*
(iii) *One has* $\operatorname{CR}(S) = \max(\operatorname{CR}(S'), \operatorname{CR}(S_x) + 1)$.

*Proof.* — Suppose $S'$ and $S_x$ both satisfy (UHC). Since $\operatorname{gp}(S')$ is a direct summand of $\operatorname{gp}(S)$ and $\operatorname{Hilb}(S') = \operatorname{Hilb}(S) \smallsetminus \{x\}$ by the hypothesis on $x$, it is clear that all elements of $S'$ are contained in subsemigroups of $S$ generated by subsets $X_i$ of $\operatorname{Hilb}(S)$ such that $X_i$ generates a direct summand of $\operatorname{gp}(S)$. If $\operatorname{rank} S' = \operatorname{rank} S$, then the sets $X_i$ are unimodular with respect to $S$, and if $\operatorname{rank} S' < \operatorname{rank} S$, then $S = S' \oplus \mathbb{Z}_+x$. In proving that $S$ satisfies (UHC), it is therefore enough to consider $S \smallsetminus S'$.

Let $z \in S \smallsetminus S'$. By hypothesis on $S_x$, the residue class of $z$ in $S_x$ has a representation $\overline{z} = a_1\overline{y}_1 + \cdots + a_m\overline{y}_m$ with $a_i \in \mathbb{Z}_+$ and $\overline{y}_i \in \operatorname{Hilb}(S_x)$ for $i = 1, \dots, m$ such that $\overline{y}_1, \dots, \overline{y}_m$ span a direct summand of $\operatorname{gp}(S_x)$. Next observe that $\operatorname{Hilb}(S)$ is mapped onto a system of generators of $S_x$ by the residue class map. Therefore we may assume that the preimages $y_1, \dots, y_m$ belong to $\operatorname{Hilb}(S) \smallsetminus \{x\}$. 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 $\operatorname{Hilb}(S)$ in which the coefficient of $x$ is positive. Thus, if $b < 0$, $z$ has a $\mathbb{Q}_+$-linear representation by the elements of $\operatorname{Hilb}(S) \smallsetminus \{x\}$. This implies $y \in C(S')$, and hence $y \in S'$, a contradiction.

This proves (i), and (ii) and (iii) follow similarly. □

We say that a semigroup $S$ as in the lemma *shrinks* to the semigroup $T$ if there is a chain $S = S_0 \supset S_1 \supset \cdots \supset S_t = T$ of semigroups such that at each step $S_{i+1}$ is generated by $\operatorname{Hilb}(S_i) \smallsetminus \{x\}$ where $x$ is non-destructive. An analogous terminology applies to cones.

***Corollary 3.5.3***. — *A counterexample to* (UHC) *that is minimal with respect to first dimension and then* $\#\operatorname{Hilb}(C)$ *is tight. A similar statement holds for* (FHC).

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 3.5.2. (For (FHC) the argument is the same.) It is therefore clear that one should search for counterexamples only among the tight cones. We will discuss the algorithmic aspects of this strategy in Subsection 3.6.

***Remark 3.5.4***. — It is not hard to see that there are no tight cones of dimension $\leqslant 2$. However, we cannot prove that all 3-dimensional cones $C$ are non-tight; in general, an extreme element of $\operatorname{Hilb}(C)$ can very well be destructive, even if $\dim C = 3$. In dimension 4 there exist 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 $\geqslant 5$ one can easily describe a class of tight cones: let $W$ be a cube whose lattice points are its vertices and its barycenter; then the cone $C_W$ is tight if $\dim W \geqslant 4$.

**3.6. The counterexample.** — Before we give the counterexample, we outline the strategy of the search. It consists of 4 steps:

(G) the choice of the generators of the cone $C$ to be tested;
(T) the shrinking of $C$ to a tight cone;
(C) the computation of several covers of $C$ by simplicial subcones;
(U) the verification that $C$ has a unimodular cover or otherwise.

There is not much to say about step (G). Either the generators of $C$ have been chosen by a random procedure depending on some parameters, especially the dimension, or they have been chosen systematically in order to exhaust a certain class.

Step (T) is carried out as follows. First the Hilbert basis of $C$ is computed and among its elements the set $E$ of extreme ones. Then successively each element $x$ of $E$ is tested for being non-destructive by checking whether (i) $\operatorname{Hilb}(C) \smallsetminus \{x\}$ is a Hilbert basis of the cone $C'$ it generates and (ii) whether the group generated by $\operatorname{Hilb}(C) \smallsetminus \{x\}$ is a direct summand of $\mathbb{Z}^n$. If so, $C$ is replaced by $C'$. Otherwise the next element of $E$ is tested in the same way. The procedure stops with a tight cone (which often is $\{0\}$).

For (T) we use the algorithm mentioned in Remark 3.3.2.

For each of the covers mentioned in step (C) we first compute a triangulation $T$ depending on the order in which $\operatorname{Hilb}(C)$ is given, and for the other covers this order is permuted randomly. None of the simplicial subcones $\sigma \in T$ contains an element of $\operatorname{Hilb}(C)$ different from the extreme generators of $\sigma$. Many of the simplicial subcones $\sigma$ of $T$ will be unimodular and others non-unimodular. We then try to improve the situation as follows: for each non-unimodular $\sigma$ we look at the cones $\sigma_v$ generated by $\sigma$ and $v + \sum(w - v)$ where $v$ is an extreme generator of $\sigma$ and $w$ runs through the set $R$ of the remaining extreme generators of $\sigma$. For each element $y \in \operatorname{Hilb}(C) \cap \sigma_v$ the cone $\sigma$ is covered by the union of the $n-1$ cones $\sigma_1, \dots, \sigma_{n-1}$ generated by $v$, $y$ and $n-2$ elements from $R$. We try to choose $y$ in an 'optimal' way, replace $\sigma$ by $\sigma_1, \dots, \sigma_{n-1}$, and iterate the procedure. Unfortunately the effect of this step depends on the probability that a cone $\sigma_i$ is unimodular. In dimension 6 (or higher) it does usually not improve the situation.

The quality $Q(B)$ of each of the (say, 50) coverings $B$ computed is measured as follows: we sum the absolute values of the determinants of the non-unimodular simplicial subcones of $B$. Among the coverings we choose the 3 best ones $B_1$, $B_2$, $B_3$, and they are the basis for the last step (U) (the number 3 can be varied). First a list of all intersections $\gamma = \sigma_1 \cap \sigma_2 \cap \sigma_3$ is formed where $\sigma_i$ runs through the non-unimodular simplicial subcones of $B_i$. Then each 'critical subcone' obtained in this way is compared to the list $L$ of unimodular simplicial subcones generated by elements of $\operatorname{Hilb}(C)$. First, if $\gamma$ is contained in one of the elements of $L$, then it is discarded. Second, if the interior of $\gamma$ is intersected by some $\sigma \in L$, one of the support hyperplanes of $\sigma$ splits $\gamma$ into two subcones that are then checked recursively. Third, if no unimodular simplicial subcone intersects the interior of $\gamma$, then we have found the desired counterexample. The algorithm stops since the number of unimodular simplicial subcones, and therefore the number of hyperplanes available for the splitting of the critical subcones, is finite.

The output of our implementation of step (U) is a list of subcones $\delta$ such that the relative interior of their union (with respect to $C$) is the complement of the union of the unimodular subcones.

The basis of all computations involved is the dual cone algorithm (see Burger [**Bur**]) that for a given cone $C \subset \mathbb{R}^n$ computes a system of generators of the dual cone

$$C^* = \{\varphi \in (\mathbb{R}^n)^* \mid \varphi(x) \geqslant 0 \text{ for all } x \in C\}.$$

Note that the intersection $C \cap D$ of cones $C$ and $D$ is the dual of the cone generated by the union of $C^*$ and $D^*$.

Although we have an algorithm for general cones, we hoped to find a counterexample within the class of the normal polytopal semigroups $S_P$. We started our search within the class of lattice parallelepipeds $P$ which are automatically normal. The counterexample finally emerged when we applied the shrinking process to cones over 5-dimensional parallelepipeds. Even for generators with 'small' coefficients, the Hilbert bases of these cones can be quite large. We have tried to select examples that are not too 'big'. Nevertheless the task is usually formidable, both in computing time and memory requirements. A typical example: $\#\operatorname{Hilb}(C) = 38$, the minimal value of $Q(B) = 324$, computing time about 24 hours, memory requirement > 100 MB.

Thus we were quite surprised by finding the following 'small' counterexample $C_6$ to (UHC) whose Hilbert basis consists of the following 10 vectors:

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

The cone $C_6$ and the semigroup $S_6 = S(C_6)$ have several remarkable properties:

1. $C_6$ has 27 facets, of which 5 are not simplicial.
2. The automorphism group $\operatorname{Aut}(S_6)$ of $S_6$ has order 20, and it operates transitively on $\operatorname{Hilb}(S_6)$. In particular this implies that $z_1, \dots, z_{10}$ are all extreme generators of $S_6$.
3. The embedding 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 $\mathbb{R}^5$.
4. The vector of lowest degree disproving (UHC) is $t_1 = z_1 + \cdots + z_{10}$. Evidently $t_1$ is invariant for $\operatorname{Aut}(S_6)$, and it can be shown that its multiples are the only such elements.
5. The Hilbert basis is contained in the hyperplane $H$ given by the equation $-5\zeta_1 + \zeta_2 + \cdots + \zeta_6 = 1$. 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).
6. If one removes all the unimodular subcones generated by elements of $\operatorname{Hilb}(C_6)$ from $C_6$, then there remains the interior of a convex cone $N$. While $P_5$ has normalized volume 25, the intersection of $N$ and $P_5$ has only normalized volume 1/1080.
7. The binomial ideal defining the semigroup algebra $K[S_6]$ over an arbitrary field $K$ is generated by 10 binomials of degree 3 and 5 binomials of degree 4 (the latter correspond to the non-simplicial facets).
8. The $h$-vector of $P_5$ is $(1, 4, 10, 10)$ and the $f$-vector is $(1, 10, 40, 80, 75, 27)$.
9. The vector
$$z_1 + 3z_2 + 5z_4 + 2z_5 + z_8 + 5z_9 + 3z_{10}$$
can not be represented by 6 elements of $\operatorname{Hilb}(S)$ (and it is "smallest" with respect to this property.) Moreover, one has $\operatorname{CR}(C_6) = 7$ (as can be seen from a triangulation containing only two non-unimodular simplices).

In particular $C_6$ is even a counterexample to (ICP). This has been shown in cooperation with Henk, Martin and Weismantel; see [**BGHMW**], which also gives more detailed information on properties 2, 3, and 9.

Despite of more than two and a half years of computer time on a fast multi-CPU machine, we have found only one more counterexample to (UHC) essentially different from $C_6$. It is also of dimension 6 and a polytopal semigroup, but its Hilbert basis contains 12 elements. As Henk, Martin, and Weismantel have verified, it violates (ICP), too. Thus the question whether there exist examples satisfying (ICP), but violating (UHC), remains open.

**3.7. (ICP), (FHC), and normality.** — In this subsection we show that (ICP) implies normality and even (FHC).

***Theorem 3.7.1***. — *Let $S \subset \mathbb{Z}^n = \operatorname{gp}(S)$ be a positive affine semigroup of rank $n$. If every element of $S$ can be represented by $n$ elements of* $\operatorname{Hilb}(S)$*, then $S$ is normal and satisfies* (FHC). *Especially* (ICP) *and* (FHC) *are equivalent and they imply the normality of $S$.*

*Proof*. — We need the well-known fact (for instance, see Gubeladze [**Gu2**, Lemma 5.3]) that the conductor ideal $\mathfrak{c}(\overline{S}/S) = \{x \in S \mid x + \overline{S} \subset S\}$ is not empty. This implies that $\overline{S} \setminus S$ *is contained in finitely many hyperplanes parallel to the support hyperplanes of $S$*. Indeed, if $x \in \mathfrak{c}(\overline{S}/S)$, $y \in \overline{S}$, and $\sigma_i(y) \geqslant \sigma_i(x)$ for all support forms $\sigma_i$ of $S$, then $y \in S$, since $y - x \in \overline{S}$.

Next we observe that the set of all $x \in S$ that are not a non-negative linear combination of *linearly independent* elements in $\operatorname{Hilb}(S)$ is "thin". In fact, if $x$ is the linear combination of $n$ elements of $\operatorname{Hilb}(S)$ that are not linearly independent, then it is contained in the proper subspace generated by these elements, and there are only finitely many such subspaces.

Choose $y \in \overline{S}$ and consider all linearly independent subsets $X_i$, $i = 1, \dots, N$, of $\operatorname{Hilb}(S)$ such that $y$ is contained in the cone generated by $X_i$. In view of Carathéodory's theorem we have $N \geqslant 1$. Let $G_i$ be the subgroup of $\mathbb{Z}^n$ generated by $X_i$.

In order to derive a contradiction suppose that $y$ is not contained in one of the subsemigroups of $S$ generated by $X_i$. It is impossible that $y \in \bigcup_{i=1,\dots,N} G_i$. Namely, if $y \in G_i$, it could be written as a $\mathbb{Z}$-linear combination of $X_i$ as well as a linear combination with non-negative coefficients: these must coincide if $X_i$ is linearly independent.

Let $E = y + \bigcap_{i=1,\dots,N} G_i$. Then

$$E \cap \bigcup_{i=1,\dots,N} G_i = \varnothing.$$

Furthermore $E \cap C(S)$ is contained in $\overline{S}$. It is not hard to see that the affine space generated by $E \cap C(S)$ has dimension $n$. Therefore $E \cap C(S)$ is not contained in the union of finitely many proper affine subspaces. This however means that $E \cap C(S)$ contains elements of $S$ that can not be written as linear combination of linearly dependent elements of $\operatorname{Hilb}(S)$. But neither can they be written as a linear combination of linearly independent elements of $\operatorname{Hilb}(S)$ with non-negative coefficients since such elements are always contained in one of the sets $X_i$. □

***Remark 3.7.2***. — The proof of Theorem 3.7.1 suggests an algorithm deciding (FHC) (or (ICP)) for a cone $C$. In addition to the steps (G)–(U) outlined in Subsection 3.6 one applies the following recursive procedure to each of the not unimodularly covered subcones $\delta$ resulting from step (U): (i) Let $X_i$, $i = 1, \dots, N$, be the linearly independent subsets of $\mathrm{Hilb}(C)$ such that $\delta$ is contained in the cone generated by $X_i$. Then we check whether the union of subgroups $G_i$ (notation as in the previous proof) is $\mathbb{Z}^n$. If so, then $\delta$ is 'freely' covered and can be discarded. (ii) Otherwise, if there is a simplicial subcone $\sigma$ generated by elements of $\mathrm{Hilb}(C)$ intersecting $\delta$ in its interior, then we split $\delta$ into two subcones along a suitable support hyperplane of $\sigma$. (iii) If there is no such $\sigma$, then one has found a counterexample to (FHC).

It is crucial for this algorithm that the question whether $\mathbb{Z}^n$ is the union of subgroups $U_1, \dots, U_m$ can be decided algorithmically. For that one forms their intersection $V$ and checks that for each residue class modulo $V$ a representative is contained in one of the $U_i$.

While the algorithm just described only decides whether (ICP) holds for $S$, one can indeed compute $\mathrm{CR}(S)$ for an arbitrary affine semigroup $S$ by suitable "covering algorithms"; see [**BGT2**].

## 4. Divisorial linear algebra

**4.1. Introduction.** — We recall some facts from Subsection 2.1. A normal semigroup $S \subset \mathbb{Z}^n$ can be described as the set of lattice points in a finitely generated rational cone. Equivalently, it is the set

$$(*) \qquad S = \{x \in \mathbb{Z}^n \colon \sigma_i(x) \geqslant 0,\ i = 1, \dots, s\}$$

of lattice points satisfying a system of homogeneous inequalities given by linear forms $\sigma_i$ with integral (or rational) coefficients. For a field $K$ the $K$-algebra $K[S]$ is a normal semigroup algebra. We always assume in Section 4 that $S$ is positive, i.e. 0 is the only invertible element in $S$.

Let $a_1, \dots, a_s$ be integers. Then the set

$$T = \{x \in \mathbb{Z}^n \colon \sigma_i(x) \geqslant a_i,\ i = 1, \dots, s\}$$

satisfies the condition $S + T \subset T$, and therefore the $K$-vector space $KT \subset K[\mathbb{Z}^n]$ is an $R$-module in a natural way.

It is not hard to show that such an $R$-module is a (fractional) ideal of $R$ if the group $\mathrm{gp}(S)$ generated by $S$ equals $\mathbb{Z}^n$. Moreover, if the presentation $(*)$ of $S$ is irredundant, then the $R$-modules $KT$ are even divisorial ideals, as we have seen in Subsection 2.3.1; in fact,

$$KT = \mathfrak{p}_1^{(a_1)} \cap \dots \cap \mathfrak{p}^{(a_s)}$$

where $\mathfrak{p}_1 \dots, \mathfrak{p}_s$ are the divisorial prime ideals of $R$ and $\mathfrak{p}_i$ is generated by all monomials $x$ with $\sigma_i(x) \geqslant 1$. These divisorial ideals represent the full divisor class group $\mathrm{Cl}(R)$. Therefore an irredundant system of homogeneous linear inequalities is the most interesting from the ring-theoretic point of view, and in these notes we restrict ourselves to it. In [**BG7**] the general case has also been treated.

We are mainly interested in two invariants of divisorial ideals $D$, namely their number of generators $\mu(D)$ and their depth as $R$-modules, and in particular in the Cohen–Macaulay property. Our main result, based on combinatorial arguments, is that for each $C \in \mathbb{Z}_+$ there exist, up to isomorphism, only finitely many divisorial ideals $D$ such that $\mu(D) \leqslant C$. It then follows by Serre's numerical Cohen–Macaulay criterion that only finitely many divisor classes represent Cohen–Macaulay modules.

The second main result concerns the growth of Hilbert functions of certain multi-graded algebras and modules. Roughly speaking, it says that the Hilbert function takes values $\leqslant C$ only at finitely many graded components, provided this holds along each arithmetic progression in the grading group. The theorem on Hilbert functions can be applied to the minimal number of generators of divisorial ideals since $R$ can be embedded into a polynomial ring $P$ over $K$ such that $P$ is a $\mathrm{Cl}(R)$-graded $R$-algebra in a natural way. This leads to a second proof of the result on number of generators mentioned above.

Subsection 4.2.1 describes the connection between divisor classes and the standard embedding and contains results on the depth of divisorial ideals. We first show that a divisorial ideal whose class is a torsion element in $\mathrm{Cl}(R)$ is Cohen–Macaulay. (The Cohen–Macaulay property and notions related to it are briefly introduced at the end of this subsection.) This follows easily from Hochster's theorem [**Ho**] that normal semigroup algebras are Cohen–Macaulay. Then we give a combinatorial description of the minimal depth of all divisorial ideals of $R$: it coincides with the minimal number of facets $F_1, \dots, F_u$ of the cone generated by $S$ such that $F_1 \cap \dots \cap F_u = \{0\}$.

Subsection 4.3 contains our main result on number of generators. The crucial point in its proof is that the convex polyhedron $C(D)$ naturally associated with a divisorial ideal $D$ has a compact face of positive dimension if (and only if) the class of $D$ is non-torsion. One can show that $\mu(D) \geqslant M\lambda$ where $M$ is a positive constant only depending on the semigroup $S$ and $\lambda$ is the maximal length of a compact 1-dimensional face of $C(D)$. Moreover, since the compact 1-dimensional faces are in discrete positions and uniquely determine the divisor class, it follows that $\lambda$ has to go to infinity in each infinite family of divisor classes.

The observation on compact faces of positive dimension is also crucial for our second approach to the number of generators via Hilbert functions. Their well-established theory allows us to prove quite precise results about the asymptotic behaviour of $\mu$ and depth along an arithmetic progression in the divisor class group.

Subsection 4.4 finally contains the theorem on the growth of Hilbert functions outlined above. It is proved by an analysis of homomorphisms of affine semigroups and their "modules".

The divisorial ideals can be realized as modules of covariants for an action of a diagonalizable group on the polynomial ring of the standard embedding; see [**BG7**]. The Cohen–Macaulay property of coset modules has been characterized by Stanley

[**St1, St3**] in terms of local cohomology. Brion [**Bri**] has shown that the number of isomorphism classes of Cohen–Macaulay modules of covariants is finite for certain actions of linear algebraic groups; however, the hypotheses of his theorem exclude groups with infinitely many characters. Therefore our result is to some extent complementary to Brion's.

We briefly explain some notions of commutative algebra (see [**BH**] for a detailed account). For a finitely generated $R$-module $M$ and an ideal $I \subset R$ we denote the length of a maximal $M$-sequence in $I$ by $\operatorname{grade}(I, M)$, where $x_1, \dots, x_m$ is an *$M$-sequence* if $(x_1, \dots, x_m)M \neq M$ and $x_i$ is not a zero-divisor on $M/(x_1, \dots, x_{i-1})M$ for $i = 1, \dots, n$. For $\operatorname{grade}(I, R)$ one uses the abbreviation $\operatorname{grade} I$.

Suppose $R$ is a local ring, i. e. has exactly one maximal ideal $\mathfrak{m}$. Then one sets $\operatorname{depth} M = \operatorname{depth}_R M = \operatorname{grade}(\mathfrak{m}, R)$. A *Cohen–Macaulay $R$-module $M$* is characterized by the equation $\operatorname{depth} M = \dim R/\operatorname{Ann}(M)$, where $\operatorname{Ann}(M) = \{x \in R : xM = 0\}$ is the annihilator of $M$. One says that $R$ is a *Cohen–Macaulay ring* if it is a Cohen–Macaulay module over itself. A finitely generated module over a general noetherian ring is Cohen–Macaulay if its localization $M_{\mathfrak{p}}$ is Cohen–Macaulay over $R_{\mathfrak{p}}$ for every prime ideal $\mathfrak{p}$ of $R$. For ideals in Cohen–Macaulay rings one has $\operatorname{grade} I = \operatorname{height} I$. A prime ideal $\mathfrak{p}$ is *associated* to $M$ if $\operatorname{depth} M_{\mathfrak{p}} = 0$. It is an important fact that the set $\operatorname{Ass}(M)$ of associated prime ideals of a finitely generated module over a noetherian ring is finite and that the union of the associated prime ideals is the set of zero divisors of $M$.

This notion depth is also used if $R$ has a distinguished maximal ideal $\mathfrak{m}$, for example if it is a positively $\mathbb{Z}$-graded $K$-algebra with $\mathfrak{m} = \bigoplus_{i>0} R_i$, or a positive affine semigroup algebra $K[S]$ with $\mathfrak{m}$ the maximal ideal generated by all monomials $\neq 1$. In these cases $\operatorname{depth}_R M = \operatorname{depth}_{R_{\mathfrak{m}}} M_{\mathfrak{m}}$. (See [**BH**, 1.5.15] for the proof in the case of $\mathbb{Z}$-graded rings and modules; moreover, note that a positive affine semigroup has a grading as observed in Subsection 2.1.) Similarly a graded $R$-module is a Cohen–Macaulay-module if and only if $M_{\mathfrak{m}}$ is a Cohen–Macaulay $R_{\mathfrak{m}}$-module.

**4.2. The standard embedding, divisor classes, and depth.** — In Subsection 2.1 we have introduced the standard embedding of a positive affine semigroup $S$ into the lattice $\mathbb{Z}^s$ given by the support forms $\sigma_1, \dots, \sigma_s$ and the induced standard embedding of the algebra $K[S]$ into a polynomial ring. This is the optimal tool for the simultaneous study of all the divisor classes. With respect to the standard embedding the divisor classes are realized as *coset modules*, as the proof of the following theorem shows:

***Theorem 4.2.1***. — *Let $K$ be a field, $S$ a positive normal affine semigroup, $R = K[S]$, and $\sigma \colon R \to P = K[Y_1, \dots, Y_s]$ the standard embedding.*

*Then $P$ decomposes as an $R$-module into a direct sum of rank 1 $R$-modules $M_c$, $c \in \operatorname{Cl}(R)$, such that $M_c$ is isomorphic to a divisorial ideal of class $c$.*

*Proof.* — To each divisorial ideal $\mathfrak{p}_1^{(a_1)}\cap\cdots\cap\mathfrak{p}_s^{(a_s)}$, $a_i\in\mathbb{Z}$, we associate $(a_1,\dots,a_s)\in\mathbb{Z}^s$. Under this assignment, the principal divisorial ideal generated by $s\in\operatorname{gp}(S)$ is mapped to $\sigma(s)$. As outlined in Theorem 2.3.1, this yields the isomorphism

$$\operatorname{Cl}(R)\cong\mathbb{Z}^s/\sigma(\mathbb{Z}^r).$$

For each $c\in\mathbb{Z}^s/\sigma(\mathbb{Z}^r)$ we let $M_c$ be the $K$-vector subspace of $P$ generated by all monomials whose exponent vector in $\mathbb{Z}^s/\sigma(\mathbb{Z}^r)$ has residue class $-c$. Then $M_c$ is clearly an $R$-submodule of $P$. Moreover, by construction, $P$ is the direct sum of these $R$-modules.

It remains to show that $M_c$, $c\in\operatorname{Cl}(R)$, is isomorphic to a divisorial ideal of class $c$ (relative to the isomorphism above). We choose a representative $a=(a_1,\dots,a_s)$ of $c$. Then a monomial $s\in\sigma(\mathbb{Z}^r)$ belongs to $\mathfrak{p}_1^{(a_1)}\cap\cdots\cap\mathfrak{p}_s^{(a_s)}$ if and only if $s_i\geqslant a_i$ for all $i$, and this is equivalent to $s_i-a_i\in\mathbb{Z}_+$ for all $i$. Hence the assignment $s\mapsto s-a$, in ring-theoretic terms: multiplication by the monomial $Y^{-a}$, induces an $R$-isomorphism $M_c\cong\mathfrak{p}_1^{(a_1)}\cap\cdots\cap\mathfrak{p}_s^{(a_s)}$. □

***Corollary 4.2.2***. — *Every divisorial ideal $I$ of $R$ whose class in* $\operatorname{Cl}(R)$ *is a torsion element is a Cohen–Macaulay $R$-module.*

*Proof.* — Let $\widehat{S}$ be the integral closure of $S$ in $\mathbb{Z}^s$. Since $K[\widehat{S}]$ is a normal affine semigroup, the ring $K[\widehat{S}]$ is Cohen–Macaulay by Hochster's theorem [**Ho**]. It decomposes into the direct sum of the finitely many and finitely generated $R$-modules $M_c$ where $c$ is a torsion class in $\operatorname{Cl}(R)$. Since $K[\widehat{S}]$ is a Cohen–Macaulay ring, it is a Cohen–Macaulay $K[S]$-module, and so are its direct summands. □

The corollary can be significantly generalized if one applies the idea of its proof to all so-called *pure* embeddings of $S\subset\tilde{S}$ of $S$ into normal affine semigroups $S$. One then obtains the Cohen–Macaulay property for all divisorial ideals $KT$ where $T$ is defined by some $\beta\in\operatorname{gp}(S)\otimes\mathbb{R}$ via

$$T=\{z\in\operatorname{gp}(S):\sigma_i(z)\geqslant\sigma_i(\beta)\text{ for }i=1,\dots,s\}.$$

For evident reasons these divisorial ideals (and their classes) have been termed *conic* in [**BG7**], to which we refer the reader for the details. We will see below that torsion classes are conic, but the converse is wrong as soon as there exist non-torsion classes.

In general a normal semigroup algebra $R$ may very well have Cohen–Macaulay divisorial ideals whose classes are not torsion. This fact and several other aspects of our discussion are illustrated by the following

***Example 4.2.3***. — Consider the Segre product

$$R_{mn}=K[X_iY_j\colon 1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n]\subset P=K[X_1,\dots,X_m,Y_1,\dots,Y_n]$$

of the polynomial rings $K[X_1,\dots,X_m]$, $m\geqslant 2$, and $K[Y_1,\dots,Y_n]$, $n\geqslant 2$, with its standard embedding. It has divisor class group isomorphic to $\mathbb{Z}$, and the two generators of $\mathrm{Cl}(R)$ correspond to the coset modules $M_{-1} = RX_1 + \cdots + RX_m$ and $M_1 = RY_1 + \cdots + RY_n$. Therefore

$$\mu(M_{-i}) = \binom{m+i-1}{m-1}, \qquad \mu(M_i) = \binom{n+i-1}{n-1}.$$

for all $i \geqslant 0$. The Cohen–Macaulay divisorial ideals are represented by

$$M_{-(n-1)}, \ldots, M_0 = R_{mn}, \ldots, M_{m-1}$$

(see Bruns and Guerrieri [**BGu**]), and in particular, their number is finite. However, the finiteness of the number of Cohen–Macaulay classes is not a peculiar property of $R_{mn}$: it holds for all normal semigroup algebras, as we will see in Corollary 4.3.2.

Moreover, one has

$$\inf_{i\geqslant 0} \operatorname{depth} M_{-i} = n, \qquad \inf_{i\geqslant 0} \operatorname{depth} M_i = m,$$

and for $i \gg 0$ the minimal values are attained (see Bruns and Vetter [**BV**, (9.27)]). Set $p(i) = \mu(M_i)$. It follows that the degree of the polynomial $p$ and $\inf_i \operatorname{depth} M_i$ add up to $m+n-1 = \dim R$. This is another instance of a general fact (see Theorem 4.3.5).

That divisorial ideals $I$ of non-torsion class are in general not Cohen-Macaulay, follows already from the asymptotic behaviour of depth $I$ described in the next theorem.

***Theorem 4.2.4***. — *Let $K$ be a field, $S$ a positive normal affine semigroup, $R = K[S]$, and $\sigma\colon R \to P = K[X_1, \ldots, X_s]$ the standard embedding. Furthermore let $\mathfrak{m}$ be the irrelevant maximal ideal of $R$ generated by all non-unit monomials, and $\lambda$ the maximal length of a monomial $R$-sequence. Then*

$$\lambda \leqslant \operatorname{grade} \mathfrak{m}P = \min\{\operatorname{depth} M_c \colon c \in \mathrm{Cl}(R)\}.$$

*Proof.* — For the inequality it is enough to show that a monomial $R$-sequence is also a $P$-sequence. (It is irrelevant whether we consider $P$ as an $R$-module or a $P$-module if elements from $R$ are concerned.) Let $\mu_1, \ldots, \mu_u$ be monomials in $R$ forming an $R$-sequence. Then the subsets $A_i = \operatorname{Ass}_R(R/(\mu_i))$ are certainly pairwise disjoint. On the other hand, $A_i$ consists only of monomial prime ideals of height 1 in $R$, since $R$ is normal. So $A_i = \{\mathfrak{p}_j \colon \sigma_j(\mu_i) > 0\}$, and the sets of indeterminates of $P$ that divide $\sigma(\mu_i)$ in $P$, $i = 1, \ldots, u$, are pairwise disjoint. It follows that $\mu_1, \ldots, \mu_u$ form a $P$-sequence.

In order to prove the equality we first extend the field $K$ to an uncountable one. This is harmless, since all data are preserved by base field extension. Then we can form a maximal $P$-sequence in $\mathfrak{m}P$ by elements from the $K$-vector subspace $\mathfrak{m}$. Such a $P$-sequence of elements in $R$ then has length equal to grade $\mathfrak{m}P$ and is clearly an $M$-sequence for every $R$-direct summand $M$ of $P$, and in particular for each of the modules $M_c$ representing the divisor classes. Thus depth $M_c \geqslant$ grade $\mathfrak{m}P$.

Whereas this argument needs only finite prime avoidance, we have to use countable prime avoidance for the converse inequality. Suppose that $u < \min\{\operatorname{depth} M_c \colon c \in \operatorname{Cl}(R)\}$ and that $x_1, \dots, x_u$ is a $P$-sequence in $\mathfrak{m}$. Then the set

$$A = \bigcup_{c \in \operatorname{Cl}(R)} \operatorname{Ass}\big(M_c/(x_1, \dots, x_u)M_c\big)$$

is a countable set of $K$-vector subspaces of $\mathfrak{m}$. Each prime ideal associated to $M_c/(x_1, \dots, x_u)M_c$ is a proper subspace of $\mathfrak{m}$ because of $u < \operatorname{depth} M_c$. Hence $A$ cannot exhaust $\mathfrak{m}$, as follows from elementary arguments. So we can choose an element $x_{u+1}$ in $\mathfrak{m}$ not contained in a prime ideal associated to any of the $M_c/(x_1, \dots, x_u)M_c$. So $x_{u+1}$ extends $x_1, \dots, x_u$ to an $M_c$-sequence simultaneously for all $c \in \operatorname{Cl}(R)$. ☐

Both the numbers $\lambda$ and $\operatorname{grade} \mathfrak{m}P$ can be characterized combinatorially:

**Proposition 4.2.5.** — *With the notation of the previous theorem, the following hold:*

(a) $\operatorname{grade} \mathfrak{m}P$ *is the minimal number* $u$ *of facets* $F_{i_1}, \dots, F_{i_u}$ *of* $C(S)$ *such that* $F_{i_1} \cap \dots \cap F_{i_u} = \{0\}$.

(b) $\lambda$ *is the maximal number* $\ell$ *of subsets* $\mathcal{F}_1, \dots, \mathcal{F}_\ell$ *of* $\mathcal{F} = \{F_1, \dots, F_s\}$ *with the following properties:*

$$\text{(i)} \quad \mathcal{F}_i \cap \mathcal{F}_j = \varnothing, \qquad \text{(ii)} \bigcap_{F \in \mathcal{F} \smallsetminus \mathcal{F}_i} F \neq \{0\}$$

*for all* $i, j$ *such that* $i \neq j$.

*Proof*

(a) The ideal $\mathfrak{m}P$ of $P$ is generated by monomials. Therefore all its minimal prime ideals are generated by indeterminates of $P$. The ideal generated by $X_{i_1}, \dots, X_{i_u}$ contains $\mathfrak{m}P$ if and only if for each monomial $\mu \in \mathfrak{m}$ there exists a $\sigma_{i_j}$ such that $\sigma_{i_j}(\mu) > 0$. The monomials for which none such inequality holds are precisely those in $F_{i_1} \cap \dots \cap F_{i_u}$.

(b) Let $\mu_1, \dots, \mu_\ell$ be a monomial $R$-sequence. Then the sets

$$\mathcal{F}_i = \{F \colon \mathfrak{p}_F \in \operatorname{Ass}(R/(\mu_i))\}$$

are pairwise disjoint, and moreover $\mu_i \in \bigcap_{F \in \mathcal{F} \smallsetminus \mathcal{F}_i} F$. Thus conditions (i) and (ii) are both satisfied.

For the converse one chooses monomials $\mu_i \in \bigcap_{F \in \mathcal{F} \smallsetminus \mathcal{F}_i} F$. Then

$$\{F \colon \mathfrak{p}_F \in \operatorname{Ass}(R/(\mu_i))\} \subset \mathcal{F}_i,$$

and since the $\mathcal{F}_i$ are pairwise disjoint, the $\mu_i$ form even a $P$-sequence as observed above. ☐

**4.3. The number of generators.** — In this subsection we first prove our main result on the number of generators of divisorial ideals of normal semigroup algebras $R = K[S]$. In its second part we then show that it can also be understood and proved as an assertion about the growth of the Hilbert function of a certain $\mathrm{Cl}(R)$-graded $K$-algebra.

***Theorem 4.3.1***. — *Let $R$ be a positive normal semigroup algebra over the field $K$, and $m \in \mathbb{Z}_+$. Then there exist only finitely many $c \in \mathrm{Cl}(R)$ such that a divisorial ideal $D$ of class $c$ has $\mu(D) \leqslant m$.*

As a consequence of Theorem 4.3.1, the number of Cohen–Macaulay classes is also finite:

***Corollary 4.3.2***. — *There exist only finitely many $c \in \mathrm{Cl}(R)$ for which a divisorial ideal of class $c$ is a Cohen–Macaulay module.*

***Remark 4.3.3***. — (a) One should note that $\mu(D)$ is a purely combinatorial invariant. If $S$ is the underlying semigroup and $T$ is a monomial basis of a monomial representative of $D$, then $\mu(D)$ is the smallest number $g$ such that there exists $x_1, \dots, x_g \in T$ with $T = (S+x_1) \cup \dots \cup (S+x_g)$. Therefore Theorem 4.3.1 can very well be interpreted as a result on the generation of the sets of solutions to inhomogeneous linear diophantine equations and congruences (with fixed associated homogeneous system).

(b) Both the theorem and the corollary hold for all normal affine semigroups $S$, and not only for positive ones. We have observed in Subsection 2.1 that a normal semigroup $S$ splits into a direct summand of its largest subgroup $S_0$ and a positive normal semigroup $S'$. Thus one can write $R = K[S]$ as a Laurent polynomial extension of the $K$-algebra $R' = K[S']$. Each divisor class of $R$ has a representative $D' \otimes_{R'} R$. Furthermore $\mu_{R'}(D') = \mu_R(D' \otimes_{R'} R)$ and the Cohen–Macaulay property is invariant under Laurent polynomial extensions.

We first derive the corollary from the theorem. Let $\mathfrak{m}$ be the irrelevant maximal ideal of $R$. If $M_c$ is a Cohen–Macaulay module, then $(M_c)_\mathfrak{m}$ is a Cohen–Macaulay module (and conversely). Furthermore

$$e((M_c)_\mathfrak{m}) \geqslant \mu((M_c)_\mathfrak{m}) = \mu(M_c).$$

By Serre's numerical Cohen–Macaulay criterion (for example, see [**BH**, 4.7.11]), the rank 1 $R_\mathfrak{m}$-module $(M_c)_\mathfrak{m}$ is Cohen–Macaulay if and only if its multiplicity $e((M_c)_\mathfrak{m})$ coincides with $e(R_\mathfrak{m})$.

*Proof of Theorem 4.3.1.* — Let $D$ be a monomial divisorial ideal of $R$. As pointed out already, there exist integers $a_1, \dots, a_s$ such that the lattice points in the set

$$C(D) = \{x \in \mathbb{R}^r : \sigma_i(x) \geqslant a_i,\ i = 1, \dots, s\}$$

give a $K$-basis of $D$ (again $s = \#\operatorname{supp}(S)$, and the $\sigma_i$ are the support forms). The polyhedron $C(D)$ is uniquely determined by its extreme points since each of its facets is parallel to one of the facets of $C(S)$ and passes through such an extreme point. (Otherwise $C(D)$ would contain a full line, and this is impossible if $S$ is positive.)

Moreover, $D$ is of torsion class if and only if $C(D)$ has a single extreme point. This has been proved in Gubeladze [**Gu2**], but since it is the crucial point (sic!) we include the argument.

Suppose first that $D$ is of torsion class. Then there exists $m \in \mathbb{Z}_+$, $m > 0$, such that $D^{(m)}$ is a principal ideal, $D^{(m)} = xR$ with a monomial $x$. It follows that $C(D^{(m)}) = mC(D)$ has a single extreme point in (the lattice point corresponding to) $x$, and therefore $C(D)$ has a single extreme point. (An extreme point of a convex set $X$ is characterized by the property that $x \notin \operatorname{conv}(X \setminus \{x\})$.)

Conversely, suppose that $C(D)$ has a single extreme point. The extreme point has rational coordinates. After multiplication with a suitable $m \in \mathbb{Z}_+$, $m > 0$, we obtain that $C(D^{(m)}) = mC(D)$ has a single extreme point $x$ which is even a lattice point. All the facets of $C(D^{(m)})$ are parallel to those of $S$ and must pass through the single extreme point. Therefore $C(D^{(m)})$ has the same facets as $C(S) + x$. Hence $C(D^{(m)}) = C(S) + x$. This implies $D^{(m)} = Rx$ (in multiplicative notation), and so $m$ annihilates the divisor class of $D$.

Suppose that $D$ is not of torsion class. We form the line complex $\overline{\mathcal{L}}$ consisting of all 1-dimensional faces of the polyhedron $C(D)$. Then $\overline{\mathcal{L}}$ is connected, and each extreme point is an endpoint of a 1-dimensional face. Since there are more than one extreme points, all extreme points are endpoints of compact 1-dimensional faces, and the line complex $\mathcal{L}(D)$ formed by the *compact* 1-dimensional faces is also connected. Since each facet passes through an extreme point, $D$ is uniquely determined by $\mathcal{L}(D)$ (as a subset of $\mathbb{R}^s$).

Let $\mathcal{C}$ be an infinite family of divisor classes and choose a divisorial ideal $D_c$ of class $c$ for each $c \in \mathcal{C}$. Assume that the minimal number of generators $\mu(D_c)$, $c \in \mathcal{C}$, is bounded above by a constant $C$. By Lemma 4.3.4 below the Euclidean length of all the line segments $\ell \in \mathcal{L}(D_c)$, $c \in \mathcal{C}$, is then bounded by a constant $C'$.

It is now crucial to observe that the endpoints of all the line segments under consideration lie in an overlattice $L = \mathbb{Z}^n[1/d]$ of $\mathbb{Z}^n$. In fact each such point is the unique solution of a certain system of linear equations composed of equations $\sigma_i(x) = a_i$, and therefore can be solved over $\mathbb{Z}[1/d]$ where $d \in \mathbb{Z}$ is a suitable common denominator. (Again we have denoted the support forms of $S$ by $\sigma_i$.)

Let us consider two line segments $\ell$ and $\ell'$ in $\mathbb{R}^n$ as equivalent if there exists $z \in \mathbb{Z}^n$ such that $\ell' = \ell + z$. Since the length of all the line segments under consideration is bounded and their endpoints lie in $\mathbb{Z}^n[1/d]$, there are only finitely many equivalence classes of line segments $\ell \in \mathcal{L}(D_c)$, $c \in \mathcal{C}$.

Similarly we consider two line complexes $\mathcal{L}(D)$ and $\mathcal{L}(D')$ as equivalent if $\mathcal{L}(D') = \mathcal{L}(D) + z$, $z \in \mathbb{Z}^n$. However, this equation holds if and only if $C(D') = C(D) + z$, or, in other words, the divisor classes of $D$ and $D'$ coincide.

Since there are only finitely many equivalence classes of line segments and the number of lines that can appear in a complex $\mathcal{L}(D)$ is globally bounded (for example, by $2^s$ – recall that $s$ is the number of facets of $C(S)$), one can only construct finitely many connected line complexes that appear as $\mathcal{L}(D)$, up to equivalence of line complexes. This contradicts the infinity of the family $\mathcal{C}$. □

***Lemma 4.3.4***. — *Let $S$ be a positive normal semigroup, $K$ a field, $D$ a monomial divisorial ideal whose class is not torsion. Then there exists a constant $M > 0$, which only depends on $S$, such that $\mu(D) \geqslant M\lambda$ where $\lambda$ is the maximal Euclidean length of a compact 1-dimensional face of $C(D)$.*

*Proof.* — We assume that $\mathbb{Z}^n = \operatorname{gp}(S)$ so that the cone $C(S)$ and the polyhedron $C(D)$ are subsets of $\mathbb{R}^n$. Let $\ell$ be a 1-dimensional compact face of $C(D)$. Suppose $D$ is given by the inequalities

$$\sigma_i(x) \geqslant a_i, \qquad i = 1, \dots, s \qquad \big(s = \#\operatorname{supp}(C(S))\big).$$

There exists $\varepsilon > 0$ such that $U_\varepsilon(x) \cap C(D)$ contains a lattice point for each $x \in C(D)$. (In fact, $C(S)$ contains a unit cube, and $x + C(S) \subset C(D)$ for $x \in C(D)$.) Let $x \in \ell$. We can assume that

$$\sigma_i(x) \begin{cases} = a_i, & i = 1, \dots, m, \\ > a_i, & i > m. \end{cases}$$

Let $\tau = \sigma_1 + \dots + \sigma_m$. There exists $C > 0$ such that $\tau(y) < C$ for all $y \in \mathbb{R}^n$ with $|y| < \varepsilon$.

Furthermore we have $\tau(z) > 0$ for all $z \in C(S)$, $z \neq 0$. Otherwise the facets $F_1, \dots, F_m$ would meet in a line contained in $C(S)$, and this is impossible if $\ell$ is compact. In particular there are only finitely many lattice points $z$ in $S$ such that $\tau(z) < C$, and so there exists $\delta > 0$ such that $\tau(z) < C$ for $z \in S$ is only possible with $|z| < \delta$.

Now suppose that $D$ is generated by $x_1, \dots, x_q$. For $x \in \ell$ we choose a lattice point $p \in U_\varepsilon(x) \cap C(D)$. By assumption there exists $z \in C(S)$ such that $p = x_i + z$. Then

$$\tau(z) = \tau(p) - \tau(x_i) \leqslant \tau(p) - \tau(x) = \tau(p - x) < C.$$

Thus $|z| < \delta$, and therefore $|x - x_i| < \delta + \varepsilon$.

It follows that the Euclidean length of $\ell$ is bounded by $2q(\delta + \varepsilon)$. Of course $\delta$ depends on $\tau$, but there exist only finitely many choices for $\tau$ if one varies $\ell$. □

As pointed out, the polyhedron $C(D)$ contains a 1-dimensional compact face if $D$ is not of torsion class, but in general one cannot expect anything stronger. On the other hand, there exist examples for which $C(D)$ for every non-torsion $D$ has a compact face of arbitrarily high dimension; see Example 4.2.3.

If $C(D)$ has a $d$-dimensional face $F$, then the argument in the proof of Lemma 4.3.4 immediately yields that $\mu(D^{(j)}) \geqslant Mj^d$ for a constant $M > 0$: one has only to replace the length of the line segment by the $d$-dimensional volume of $F$. We now give another proof of a slightly more general statement. As we will see, it leads to a quite different proof of Theorem 4.3.1.

Let $S$ be a positive normal affine semigroup. Recall that the polynomial ring $P$ of the standard embedding $\sigma\colon R \to P$ decomposes into the direct sum of modules $M_c$, $c \in \mathrm{Cl}(R)$. In the following $C(M_c)$ stands for any of the congruent polyhedra $C(D)$ where $D$ is a divisorial ideal of class $c$.

***Theorem 4.3.5***. — *Let $c, d \in \mathrm{Cl}(R)$ and suppose that $c$ is not a torsion element.*

(a) *Then $\lim_{j\to\infty} \mu(M_{jc+d}) = \infty$.*

(b) *More precisely, let $m$ be the maximal dimension of the compact faces of $C(M_c)$. Then there exists $e \in \mathbb{N}$ such that*

$$\lim_{j\to\infty} \mu(M_{(ej+k)c+d})\frac{m!}{j^m}$$

*is a positive natural number for each $k = 0, \dots, e-1$.*

(c) *One has $\inf_j \operatorname{depth} M_{jc} = \dim R - m$ and $\inf_j \operatorname{depth} M_{cj+d} \leqslant \dim R - m$.*

*Proof*. — Let

$$\mathcal{D} = \bigoplus_{j=0}^{\infty} M_{jc} \qquad \text{and} \qquad \mathcal{M} = \bigoplus_{j=0}^{\infty} M_{jc+d}.$$

Then $\mathcal{D}$ is a finitely generated $K$-algebra. This follows for general reasons from Theorem 4.4.1 below: $\mathcal{D}$ is the direct sum of graded components of the $\mathrm{Cl}(R)$-graded $R$-algebra $P$, taken over a finitely generated subsemigroup of $\mathrm{Cl}(R)$. Theorem 4.4.1 also shows that $\mathcal{M}$ is a finitely generated $\mathcal{D}$-module. However, these assertions will be proved directly in the following. In particular we will see that $\mathcal{D}$ is a normal semigroup algebra over $K$.

By definition $\mathcal{D}$ is a $\mathbb{Z}_+$-graded $R$-algebra with $\mathcal{D}_0 = M_0 = R$, and $\mathcal{M}$ is a graded $\mathcal{D}$-module if we assign degree $j$ to the elements of $M_{jc+d}$. There exists $e > 0$ such that $\mathcal{D}$ is a finitely generated module over its $R$-subalgebra generated by elements of degree $e$; for example, we can take $e$ to be the least common multiple of the degrees of the generators of $\mathcal{D}$ as an $R$-algebra. Let $\mathcal{E}$ be the $e$th Veronese subalgebra of $\mathcal{D}$. We decompose $\mathcal{M}$ into the direct sum of its $\mathcal{E}$-submodules

$$\mathcal{M}_k = \bigoplus_{j=0}^{\infty} M_{(ej+k)c+d}, \qquad k = 0, \dots, e-1.$$

In view of what has to be proved, we can replace $\mathcal{D}$ by $\mathcal{E}$ and $\mathcal{M}$ by $\mathcal{M}_k$. Then we have reached a situation in which $\mathcal{D}$ is a finitely generated module over the subalgebra generated by its degree 1 elements.

Note that $\mathcal{M}$ is isomorphic to an ideal of $\mathcal{D}$: multiplication by a monomial $X^a$ such that $a$ has residue class $-(d+k)$ in $\mathrm{Cl}(R) \cong \mathbb{Z}^s/\sigma(\mathrm{gp}(S))$ maps $\mathcal{M}$ into $\mathcal{D}$. Since $\mathcal{M}$ is not zero (and $\mathcal{D}$ is an integral domain), we see that $\operatorname{Supp} \mathcal{M} = \operatorname{Spec} \mathcal{D}$.

Let $\mathfrak{m}$ be the irrelevant maximal ideal of $R$; it is generated by all elements $x \in S$, $x \neq 1$ (in multiplicative notation). Then clearly $\overline{\mathcal{M}} = \mathcal{M}/\mathfrak{m}\mathcal{M}$ is a finitely generated $\overline{\mathcal{D}} = \mathcal{D}/\mathfrak{m}\mathcal{D}$-module. Note that $\overline{\mathcal{D}}$ is a $K$-algebra with $\overline{\mathcal{D}}_0 = K$ in a natural way. Furthermore it is a finitely generated module over its subalgebra $\overline{\mathcal{D}}'$ generated by its degree 1 elements. In particular $\overline{\mathcal{M}}$ is a finitely generated $\overline{\mathcal{D}}'$-module. By construction (and Nakayama's lemma) we have

$$\mu(M_{(ej+k)c+d}) = \dim_K M_{(ej+k)c+d}/\mathfrak{m}M_{(ej+k)c+d} = H(\overline{\mathcal{M}}, j)$$

where $H$ denotes the Hilbert function of $\overline{\mathcal{M}}$ as a $\mathbb{Z}_+$-graded $\overline{\mathcal{D}}$- or $\overline{\mathcal{D}}'$-module. For $j \gg 0$ the Hilbert function is given by the Hilbert polynomial. It is a polynomial of degree $\delta - 1$ where $\delta$ is the Krull dimension of $\overline{\mathcal{M}}$. Note that $\operatorname{Supp} \overline{\mathcal{M}} = \operatorname{Spec} \overline{\mathcal{D}}$, since $\operatorname{Supp} \mathcal{M} = \operatorname{Spec} \mathcal{D}$; in particular one has $\dim \overline{\mathcal{M}} = \dim \overline{\mathcal{D}}$. Moreover the leading coefficient of the Hilbert polynomial is $e(\overline{\mathcal{M}})/(\delta-1)!$ and so all the claims for $\mathcal{M}$ follow if $m + 1 = \delta > 1$.

At this point we have to clarify the structure of $\mathcal{D}$ as a normal semigroup algebra over $K$. For convenience we choose a divisorial ideal $I \subset R$ of class $c$ generated by monomials. Then there exists an $R$-module isomorphism $M_c \to I$ mapping monomials to monomials, and such an isomorphism induces a $K$-algebra isomorphism from $\mathcal{D}$ to

$$\mathcal{R} = \bigoplus_{j=0}^{\infty} I^{(j)} T^j \subset R[T] = K[S \oplus \mathbb{Z}_+].$$

There exist $a_1, \dots, a_s \geqslant 0$ such that $I = \mathfrak{p}_1^{(a_1)} \cap \dots \cap \mathfrak{p}_s^{(a_s)}$. The monomial corresponding to $(u, z) \in \mathrm{gp}(S) \oplus \mathbb{Z}$ belongs to $\mathcal{R}$ if and only if

$$z \geqslant 0, \qquad \sigma_i(u) - z a_i \geqslant 0, \qquad i = 1, \dots, s.$$

It follows immediately that $\mathcal{R}$ is a normal semigroup algebra over $K$. Let $\mathcal{S}$ be its semigroup of monomials. One has $\mathrm{gp}(\mathcal{S}) = \mathrm{gp}(S) \oplus \mathbb{Z}$, and the elements with last component $j$ give the monomials of $I^{(j)}$.

It is not hard to show that the faces of $C(\mathcal{S})$ that are not contained in $C(S)$ are the closed envelopes of the $\mathbb{R}_+$-envelopes of the faces of $C(I)' = \{(x, 1) : x \in C(I)\}$.

Moreover, exactly those faces $F$ that do not contain an element from $\mathfrak{m}$ intersect $C(I)'$ in a compact face. In fact, if $F$ contains a monomial $x \in \mathfrak{m}$, then it contains $y + kx$, $k \in \mathbb{Z}_+$, for each $y \in F$, and therefore an unbounded set. If $F$ does not contain an element of $\mathfrak{m}$, then the linear subspace spanned by the elements of $S$ intersects $F$ in a single point, and thus each translate intersects $F$ in a compact set.

Since the dimension of $\mathcal{R}/\mathfrak{m}\mathcal{R}$ is just the maximal dimension of a face $F$ of $C(\mathcal{S})$ not containing an element of $\mathfrak{m}$, we see that $\dim \mathcal{R}/\mathfrak{m}\mathcal{R} = m + 1$. In fact, the largest dimension of a compact face of $C(I)'$ is $m$, and such a face extends to an $m+1$-dimensional face of $C(\mathcal{S})$.

Thus $\delta = m+1$ and $\delta > 1$, since $C(I)$ has at least a 1-dimensional compact face: by hypothesis $I$ is not of torsion class.

For part (c) we note that height $\mathfrak{m}R = \operatorname{grade} \mathfrak{m}\mathcal{R} = \dim \mathcal{R} - \dim \mathcal{R}/\mathfrak{m}\mathcal{R} = \dim R - m$ since $\mathcal{R}$ is Cohen–Macaulay by Hochster's theorem (and all the invariants involved are stable under localization with respect to the maximal ideal of $\mathcal{R}$ generated by monomials). Moreover $\operatorname{grade} \mathfrak{m}\mathcal{R} = \inf_j \operatorname{depth} M_{cj}$, as follows by arguments analogous to those in the proof of Theorem 4.2.4.

By similar arguments the inequality for $\inf_j \operatorname{depth} M_{cj+d}$ results from height $\mathfrak{m}\mathcal{R} = \dim R - m$. □

***Remark 4.3.6***. — The limits in Theorem 4.3.5(b) coincide if and only if the $\overline{\mathcal{D}}'$-modules $\mathcal{M}_k/\mathfrak{m}\mathcal{M}_k$ all have the same multiplicity. However, in general this is not the case. As an example one can take the semigroup algebra

$$R = K[U^2, UV, V^2, XW, YW, XZ, YZ] \subset P = K[U, V, X, Y, Z, W]$$

in its standard embedding. It has divisor class group $\operatorname{Cl}(R) = \mathbb{Z}/(2) \oplus \mathbb{Z}$. The non-zero torsion class is represented by the coset module $M_{(1,0)} = RU + RV$, and $M_{(0,1)} = RX + RY$ represents a generator of the direct summand $\mathbb{Z}$. Let $c \in \operatorname{Cl}(R)$ be the class of $M_{(1,1)}$. As an $R$-module, $M_{jc}$, $j$ odd, is generated by the monomials $U\mu, V\mu$ where $\mu$ is a degree $j$ monomial in $X, Y$, whereas for even $j$ the monomials $\mu$ form a generating system. The limits for $k = 0$ and $k = 1$ therefore differ by a factor of 2 ($d = 0$, $e = 2$).

*Second proof of Theorem 4.3.1.* — Let $P$ be the polynomial ring of the standard embedding of $R$. Then $P$ is a $\operatorname{Cl}(R)$-graded $R$-algebra whose graded component $P_c$, $c \in \operatorname{Cl}(R)$ is the module $M_c$. Passing to residue classes modulo $\mathfrak{m}$ converts the assertion of the theorem into a statement about the Hilbert function (with respect to $K$) of the $\operatorname{Cl}(R)$-graded $K$-algebra $P/\mathfrak{m}P$; note that $(P/\mathfrak{m}P)_0 = R/\mathfrak{m} = K$. By Theorem 4.3.5 the Hilbert function goes to infinity along each arithmetic progression in $\operatorname{Cl}(R)$. Therefore we are in a position to apply Theorem 4.4.3 below. It says that there are only finitely many $c \in \operatorname{Cl}(R)$ where $\mu(M_c) = H(P/\mathfrak{m}P, c)$ does not exceed a given bound $m$. □

This deduction of Theorem 4.3.1 uses the combinatorial hypotheses on $R$ only at a single point in the proof of Theorem 4.3.5, namely where we show that $\dim \mathcal{D}/\mathfrak{m}\mathcal{D} \geqslant 2$. Thus the whole argument can be transferred into a more general setting, provided an analogous condition on dimension holds.

**4.4. On the growth of Hilbert functions.** — We introduce some terminology: if $S$ is a subsemigroup of an abelian group $G$, then $T \subset G$ is an $S$-*module* if $S+T \subset T$ (the case $T = \varnothing$ is not excluded). If $S$ is finitely generated and $T$ is a finitely generated

$S$-module, then every $S$-module $T' \subset T$ is also finitely generated. For example, this follows by "linearization" with coefficients in a field $K$: $M = KT \subset K[G]$ is a finitely generated module over the noetherian ring $R = K[S]$, and so all its submodules are finitely generated over $R$. For $KT'$ this implies the finite generation of $T'$ over $S$.

First we note a result on the finite generation of certain subalgebras of graded algebras and submodules of graded modules. We do not know of a reference covering it in the generality of Theorem 4.4.1.

***Theorem 4.4.1***. — *Let $G$ be a finitely generated abelian group, $S$ a finitely generated subsemigroup of $G$, and $T \subset G$ a finitely generated $S$-module. Furthermore let $R$ be a noetherian $G$-graded ring and $M$ a $G$-graded finitely generated $R$-module. Then the following hold:*

(a) *$R_0$ is noetherian ring, and each graded component $M_g$, $g \in G$, of $M$ is a finitely generated $R_0$–module.*
(b) *$A = \bigoplus_{s\in S} R_s$ is a finitely generated $R_0$-algebra.*
(c) *$N = \bigoplus_{t\in T} M_t$ is a finitely generated $A$-module.*

*Proof*. — (a) One easily checks that $M'R \cap M_g = M'$ for each $R_0$-submodule $M'$ of $M_g$. Therefore ascending chains of such submodules $M'$ of $M_g$ are stationery.

(b) First we do the case in which $G$ is torsionfree, $G = \mathbb{Z}^m$, and $S$ is an integrally closed subsemigroup of $\mathbb{Z}^m$.

Let $\varphi\colon \mathbb{Z}^m \to \mathbb{Z}$ be a non-zero linear form. It induces a $\mathbb{Z}$-grading on $R$ with $\deg_{\mathbb{Z}}(a) = \varphi(\deg_{\mathbb{Z}^M}(a))$ for each non-zero $\mathbb{Z}^m$-homogeneous element of $R$. Let $R'$ denote $R$ with this $\mathbb{Z}$-grading. Set $R'_- = \bigoplus_{k\leqslant 0} R'_k$ and define $R'_+$ analogously. By [**BH**, 1.5.5] the $R'_0$-algebras $R'_+$ and $R'_-$ are finitely generated $R'_0$-algebras, and $R'_0$ is a noetherian ring (by (a)). On the other hand, $R'_0$ is a $(\operatorname{Ker}\varphi)$-graded ring in a natural way, and by induction we can conclude that $R'_0$ is a finitely generated $R_0$-algebra.

If $S = \mathbb{Z}^m$, then it follows immediately that $R$, the sum of $R_-$ and $R_+$ as an $R_0$-algebra, is again a finitely generated $R_0$-algebra.

Otherwise $S = \mathbb{Z}^m \cap C(S)$, and $C(S)$ has at least one support hyperplane:

$$S = \{s \in \mathbb{Z}^m\colon \alpha_i(s) \geqslant 0,\ i = 1, \dots, v\}$$

with $v \geqslant 1$. We use induction on $v$, and the induction hypothesis applies to $R' = \bigoplus_{s\in S'} R_s$,

$$S' = \{s \in \mathbb{Z}^m\colon \alpha_i(s) \geqslant 0,\ i = 1, \dots, v-1\}.$$

Applying the argument above with $\varphi = \alpha_v$, one concludes that $A = R'_+$ is a finitely generated $R_0$-algebra.

In the general case for $G$ and $S$ we set $G' = G/H$ where $H$ is the torsion subgroup of $G$, and denote the natural surjection by $\pi\colon G \to G'$. Let $R'$ be $R$ with the $G'$-grading induced by $\pi$ (its homogeneous components are the direct sums of the components $R_g$ where $g$ is in a fixed fiber of $\pi$). Let $S'$ be the integral closure of $\pi(S)$ in $G'$. Then $A' = \bigoplus_{s'\in S'} R'_{s'}$ is a finitely generated algebra over the noetherian ring $R'_0$, as we have already shown. But $R'_0$ is a finitely generated module over $R_0$ by (a), and so $A'$ is a finitely generated $R_0$-algebra. In particular, $R$ itself is finitely generated over $R_0$.

It is not hard to check that $A'$ is integral over $A$; in fact, each element $s \in \pi^{-1}(S')$ has a power $s^n \in S$ for suitable $n \in \mathbb{N}$. Furthermore it is a finitely generated $A$-algebra, and so a finitely generated $A$-module. But then a lemma of Artin and Tate (see Eisenbud [**Ei**, p. 143]) implies that $A$ is noetherian. As shown above, noetherian $G$-graded rings are finitely generated $R_0$-algebras.

(c) By hypothesis, $T$ is the union of finitely many translates $S + t$. Therefore we can assume that $T = S + t$. Passing to the shifted module $M(-t)$ (given by $M(-t)_g = M_{g-t}$), we can even assume that $S = T$. Now the proof follows the same pattern as that of (b). In order to deal with an integrally closed subsemigroup of a free abelian group $G = \mathbb{Z}^m$, one notes that $M_+$ is a finitely generated module over $R_+$ where $M_+$ is the positive part of $M$ with respect to a $\mathbb{Z}$-grading (induced by a linear form $\varphi\colon \mathbb{Z}^m \to \mathbb{Z}$). This is shown as follows: the extended module $RM_+$ is finitely generated over $R$ , and every of its generating systems $E \subset M_+$ together with finitely many components $M_i$, $i \geqslant 0$, generate $M_+$ over $R_+$; furthermore the $M_i$ are finitely generated over $R_0$ by (a).

For the general situation we consider $N'$ defined analogously as $A'$. It is a finitely generated $A'$-module by the previous argument. Since $A'$ is a finitely generated $A$-module, $N'$ is finitely generated over $A$, and so is its submodule $N$. □

We note a purely combinatorial consequence.

***Corollary 4.4.2***. — *Let $S$ and $S'$ be affine subsemigroups of $\mathbb{Z}^m$, $T \subset \mathbb{Z}^m$ a finitely generated $S$-module, and $T' \subset \mathbb{Z}^m$ a finitely generated $S'$-module. Then $S \cap S'$ is an affine semigroup, and $T \cap T'$ is a finitely generated $S \cap S'$-module.*

*Proof*. — We choose a field $K$ of coefficients and set $R = K[S']$, $M = KT'$. Then the hypotheses of the theorem are satisfied, and it therefore implies the finite generation of

$$A = \bigoplus_{s \in S} R_s = K[S \cap S'], \qquad \text{and} \qquad N = \bigoplus_{t \in T} M_t = K(T \cap T').$$

as a $K = R_0$-algebra and an $A$-module respectively. However, finite generation of the "linearized" objects is equivalent to that of the combinatorial ones. □

The next theorem is our main result on the growth of Hilbert functions. Note that we do not assume that $R_0 = K$; the graded components of $R$ and $M$ may even have infinite $K$-dimension.

***Theorem 4.4.3***. — *Let $K$ be a field, $G$ a finitely generated abelian group, $R$ a noetherian $G$-graded $K$-algebra for which $R_0$ is a finitely generated $K$-algebra, and $M$ a finitely generated $G$-graded $R$-module. Consider a finitely generated subsemigroup $S$ of $G$ containing the elements* $\deg r$, *$r \in R \setminus \{0\}$ homogeneous, and a finitely generated*

*$S$-submodule $T$ of $G$ containing the elements* $\deg x$, $x \in M \smallsetminus \{0\}$ *homogeneous. Furthermore let $H$ be the $G$-graded Hilbert function,* $H(M,t) = \dim_K M_t$ *for all* $t \in G$.

*Suppose* $\lim_{k\to\infty} H(M, kc+d) = \infty$ *for all choices of* $c \in S$, *$c$ not a torsion element of $G$, and* $d \in T$. *Then*

$$\#\{t \in T : H(M,t) \leqslant C\} < \infty$$

*for all* $C \in \mathbb{Z}_+$.

Note that $R$ is a finitely generated $R_0$-algebra by Theorem 4.4.1, and therefore a finitely generated $K$-algebra. Let $S'$ be the subsemigroup of $G$ generated by the elements $\deg r$, $r \in R \smallsetminus \{0\}$ homogeneous, and $T'$ be the $S'$-submodule of $G$ generated by the elements $\deg x$, $x \in M \smallsetminus \{0\}$ homogeneous. Then all the hypotheses are satisfied with $S'$ in place of $S$ and $T'$ in place of $T$. However, for technical reasons the hypothesis of the theorem has to be kept more general. (We are grateful to Robert Koch for pointing out some inaccuracies in previous versions of the theorem and its proof.)

*Proof of Theorem 4.4.3.* — We split $G$ as a direct sum of a torsionfree subgroup $L$ and its torsion subgroup $G_{\mathrm{tor}}$. Let $R' = \bigoplus_{\ell\in L} R_\ell$, and split $M$ into the direct sum

$$M = \bigoplus_{h\in G_{\mathrm{tor}}} M'_h, \qquad M'_h = \bigoplus_{\ell\in L} M_{(\ell,h)}.$$

By Theorem 4.4.1, $R'$ is a finitely generated $R_0$-algebra and $M'_h$ is a finitely generated $L$-graded $R'$-module for all $h \in G_{\mathrm{tor}}$, and since the hypothesis on the Hilbert function is inherited by $M'_h$, it is enough to do the case $G = L = \mathbb{Z}^m$.

We use induction on $m$. In the case $m = 1$ it is not difficult to see (and well-known) that $T$ is the union of finitely many arithmetic progressions that appear in the hypothesis of the theorem.

As a first step we want to improve the hypothesis on the Hilbert function from a "1-dimensional " condition to a "1-codimensional" condition by an application of the induction hypothesis.

Let $U$ be a proper subgroup of $L$ and $u \in L$. Then $U$ is finitely generated as a subsemigroup. We set

$$R' = \bigoplus_{s\in U} R_s \quad \text{and} \quad M' = \bigoplus_{t\in U+u} M_t.$$

Theorem 4.4.1 implies that $R'$ is a finitely generated $K$-algebra, and $M'$ is a finitely generated $R'$-module.

After fixing an origin in $U + u$ we can identify it with $U$. Therefore we can apply the induction hypothesis to $R'$ and $M'$. It follows that

$$(*) \qquad \#\{t \in T \cap (U+u) : H(t,M) \leqslant C\} < \infty.$$

By Theorem 4.4.1, $R$ is a finitely generated $R_0$-algebra and thus a finitely generated $K$-algebra. We represent $R$ as the residue class ring of an $L$-graded polynomial ring $P$ over $K$ in a natural way (in particular the monomials in $P$ are homogeneous in the $L$-grading). The hypothesis that $R_0$ is a finitely generated $K$-algebra is inherited by $P$ since $P_0$ is a (not necessarily positive) normal affine semigroup ring. Thus we may assume that $R$ itself is generated by finitely many algebraically independent elements as a $K$-algebra.

Obviously $M$ has a filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$$

where each successive quotient $M_{i+1}/M_i$ is a cyclic $L$-graded $R$-module, that is

$$M_{i+1}/M_i \cong (R/I_{i+1})(-s_i)$$

with an $L$-graded ideal $I_{i+1}$ in $R$ and a shift $s_i \in L$. As far as the Hilbert function is concerned, we can replace $M$ by the direct sum of these cyclic modules. After the introduction of a term order we can replace $R/(I_{i+1})(-s_i)$ by $R/(\operatorname{in}(I_{i+1}))(-s_i)$ where $\operatorname{in}(I_{i+1})$ is the initial ideal (see [**Ei**, 15.26]). It is well known that $R/\operatorname{in}(I_{i+1})$ has a filtration whose successive quotients are of the form $R/\mathfrak{p}$ with a prime ideal $\mathfrak{p}$ generated by monomials, and therefore by indeterminates of $R$. (For example, see the proof of [**BH**, 4.1.3], and use that associated prime ideals of multigraded modules are generated by indeterminates if the multigrading is that induced by the semigroup of all monomials in $R$.)

Altogether this reduces the problem to the case in which the $K$-vector space $M$ is isomorphic to the direct sum of vector spaces $P_i(-s_i)$ where $P_i$ is a polynomial ring generated by indeterminates with degrees in $L$, and $s_i \in L$. Furthermore we can use that the Hilbert function of $M$ satisfies condition $(*)$. The Hilbert function now counts the total number of monomials in each degree. Replacing the monomials by their exponent vectors, we can deduce the theorem from the next one. □

***Theorem 4.4.4***. — *Let $G$ be a finitely generated abelian group, $S$ a finitely generated subsemigroup of $G$, and $T$ a finitely generated $S$-submodule of $G$. Consider maps*

$$\psi_i \colon A_i \longrightarrow T, \qquad \psi_i(x) = \varphi_i(x) + t_i \quad \textit{for all} \quad x \in A_i$$

*where $A_i$ is an affine semigroup, $\varphi_i : A_i \to S$ is a homomorphism of semigroups, and $t_i \in T$, $i = 1, \dots, v$. Furthermore let*

$$\Psi \colon A_1 \cup \cdots \cup A_v \to T, \qquad \Psi|_{A_i} = \psi_i,$$

*be the map defined on the disjoint union of the $A_i$ by all the $\psi_i$. For $t \in G$ set*

$$H(t) = \#\{x \in A_1 \cup \cdots \cup A_v : \Psi(x) = t\}.$$

*Suppose that $\lim_{k\to\infty} H(kc + d) = \infty$ for all $c \in S$, $c$ not a torsion element of $G$, and all $d \in T$. Then*

$$\#\{t \in T : H(t) \leqslant C\} < \infty$$

*for all $C \in \mathbb{Z}_+$.*

*Proof*. — In step (a) we prove the theorem under the assumption that $G = L = \mathbb{Z}^m$ for some $m$ and that

$$\#\{t \in T \cap (U + u) : H(t) \leqslant C\} < \infty.$$

for all proper subgroups $U$ of $L$. This is enough to complete the proof of Theorem 4.4.3. In step (b) we can then use Theorem 4.4.3.

(a) The first observation is that we can omit all the maps $\psi_i$ that are injective. This reduces the function $H$ in each degree by at most $v$, and has therefore no influence on the hypothesis or the desired conclusion.

The difficult case is $C = 0$, and we postpone it. So suppose that we have already shown that the number of "gaps" (elements in $T$ with no preimage at all) is finite. Then we can restrict ourselves to $\operatorname{Im}\Psi$ if we want to show that there are only finitely many elements with at most $C > 0$ preimages.

It is enough to show that the elements in $\operatorname{Im}\psi_i$ with at most $C$ preimages are contained in the union of finitely many sets of the form $U + u$ where $U$ is a proper direct summand of $L$. Then we can use the hypothesis on the sets $\{x \in T \cap (U + u) : H(x) \leqslant C\}$. We can certainly assume that $v = 1$ and $t_1 = 0$, and have only to consider a *non-injective, surjective* homomorphism $\varphi : A \to S$.

For an ideal (i. e. $S$-submodule) $I \neq \varnothing$ of $S$ we have that $S \setminus I$ is contained in finitely many sets $U + u$. In fact, $S$ contains an ideal $J \neq \varnothing$ of the normalization $\overline{S}$ of $S$, namely the conductor ideal $F = \{s \in S \colon \overline{S} + s \subset S\}$. (Compare the proof of Theorem 3.7.1.) Therefore $S$ contains a set $\overline{S} + s$ with $s \in S$, and so $\overline{S} + s + t \subset I$ for $t \in I$. It follows that $\big(S \setminus I\big) \subset \big(\overline{S} \setminus (\overline{S} + s + t)\big)$. The latter set is contained in finitely many sets of type $U + u$. To sum up, it is enough to find an ideal $I$ in $S$ such that each element of $I$ has at least $C + 1$ preimages.

Now we go to $A$ and choose $a \in A$ such that $\overline{A} + a \subset A$ where $\overline{A}$ is again the normalization. The homomorphism $\varphi$ has a unique extension to a group homomorphism $\operatorname{gp}(A) \to L$, also denoted by $\varphi$. By assumption $\operatorname{Ker}\varphi \neq 0$. A sufficiently large ball $B$ in $\operatorname{gp}(A) \otimes \mathbb{R}$ with center 0 therefore contains $C + 1$ elements of $\operatorname{Ker}\varphi$, and there exists $b \in \overline{A}$ for which $B + b$ is contained in the cone $\mathbb{R}_+ A$. Thus $(B \cap \operatorname{Ker}\varphi) + b \subset \overline{A}$. It follows that each element in $I = \varphi(\overline{A} + a + b)$ has at least $C + 1$ preimages. Since $\overline{A} + a + b$ is an ideal in $A$ and $\varphi$ is surjective, $I$ is an ideal in $S$.

(b) By linearization we now derive Theorem 4.4.4 from Theorem 4.4.3. Let $K$ be a field. Then we set $R_i = K[A_i]$, and the homomorphism $\varphi_i$ allows us to consider $R_i$ as a $G$-graded $K$-algebra. Next we choose a polynomial ring $P_i$ whose indeterminates are mapped to a finite monomial system of generators of $A_i$, and so $P_i$ is also $G$-graded. Set

$$R = P_1 \otimes_K \cdots \otimes_K P_v \qquad \text{and} \qquad M = R_1(-s_1) \oplus \cdots \oplus R_v(-s_v)$$

Evidently $R$ is a finitely generated $G$-graded $K$-algebra; in particular it is noetherian. Moreover $R_i$ is residue class ring of $R$ in a natural way, and therefore $R_i(-s_i)$ can be considered a $G$-graded $R$-module. Therefore $M$ is a $G$-graded $R$-module whose Hilbert function is the function $H$ of the theorem. □

It remains to do the case $C = 0$. For simplicity we only formulate it under the special assumptions of step (a) in the proof of Theorem 4.4.4. We leave the general as well as the commutative algebra version to the reader. The semigroups $A_i$ of Theorem 4.4.4 can now be replaced by their images.

**Proposition 4.4.5**. — *Let $L = \mathbb{Z}^m$, $S$ an affine subsemigroup of $L$, $T$ a finitely generated $S$-submodule of $L$. Consider subsemigroups $A_1, \dots, A_v$ of $L$ and elements $t_1, \dots, t_v \in T$ such that the set*

$$\mathcal{G} = T \setminus ((A_1 + t_1) \cup \cdots \cup (A_v + t_v))$$

*of "gaps" satisfies the following condition: for each proper subgroup $U$ of $L$ and each $u \in L$ the intersection $(U + u) \cap \mathcal{G}$ is finite. Then $\mathcal{G}$ is finite.*

*Proof*. — Note that $T$ is contained in finitely many residue classes modulo $\mathrm{gp}(S)$. Therefore we can replace each $A_i$ by $A_i \cap \mathrm{gp}(S)$: the intersection of $A_i + t_i$ with a residue class modulo $\mathrm{gp}(S)$ is a finitely generated $A_i \cap \mathrm{gp}(S)$-module by Corollary 4.4.2.

We order the $A_i$ in such a way that $A_1, \dots, A_w$ have the same rank as $S$, and $A_{w+1}, \dots, A_v$ have lower rank. Let $W$ be the intersection of $\mathrm{gp}(A_i)$, $i = 1, \dots, w$. Since $\mathrm{gp}(S)/W$ is a finite group, we can replace all the semigroups involved by their intersections with $W$, split the modules into their intersection with the residue classes modulo $W$, and consider every residue class separately. We have now reached a situation where $A_i \subset \mathrm{gp}(S)$ for all $i$, and $\mathrm{gp}(A_i) = \mathrm{gp}(S)$, unless $\operatorname{rank} A_i < \operatorname{rank} S$.

Next one can replace the $A_1, \dots, A_w$ by their normalizations. In this way we fill the gaps in only finitely many $U + u$ (compare the argument in the proof of Theorem 4.4.4), and therefore we fill only finitely many gaps.

At this point we can assume that $A_1, \dots, A_w$ are integrally closed in $L$. Furthermore we must have $C(S) \subset C(A_1) \cup \cdots \cup C(A_w)$ – otherwise an open subcone of $C(S)$ would remain uncovered, and this would remain so in $T$: the lower rank semigroups cannot fill it, and neither can it be filled by finitely many translates $U + u$ where $U$ is a proper subsemigroup of $S$. In fact, $(A_i + t_i) \setminus A_i$ is contained in the union of finitely many such translates, and the same holds for $(C(S) \cap L) \setminus S$. Since $A_1, \dots, A_w$ are integrally closed, we have $S \subset A_1 \cup \cdots \cup A_w$.

Now we choose a system of generators $u_1, \dots, u_q$ of $T$ over $S$. We have

$$T \subset \bigcup_{i,j} A_i + u_j.$$

But $A_i + u_j$ and $A_i + t_i$ only differ in finitely many translates of proper direct subgroups of $L$ parallel to the support hyperplanes of $A_i$. So in the last step we have filled only finitely many gaps. Since no gaps remain, their number must have been finite from the beginning. □

## 5. From vector spaces to polytopal algebras

**5.1. Introduction.** — The category $\mathrm{Vect}(K)$ of finite-dimensional vector spaces over a field $K$ has a natural extension that we call the *polytopal $K$-linear category* $\mathrm{Pol}(K)$. The objects of $\mathrm{Pol}(K)$ are the polytopal algebras. That is, an object $A \in |\mathrm{Pol}(K)|$ is (up to graded isomorphism) a standard graded $K$-algebra $K[P]$ associated with a lattice polytope $P \subset \mathbb{R}^n$ (see §2.2). The morphisms of $\mathrm{Pol}(K)$ are the homogeneous $K$-algebra homomorphisms.

The category $\mathrm{Pol}(K)$ contains $\mathrm{Vect}(K)$ as a full subcategory. In fact, we can identify a vector space with the degree 1 component of its symmetric algebra, which, upon the choice of a basis, can be considered as a polynomial ring $K[X_1, \dots, X_n]$. This polynomial ring is isomorphic to the polytopal algebra $K[\Delta_{n-1}]$ defined by the $(n-1)$-simplex $\Delta_{n-1}$. Vector space homomorphisms extend to homomorphisms of symmetric algebras, and thus to homomorphisms of the corresponding polytopal algebras. In order to have the zero space we have to admit $\varnothing$ as a lattice polytope whose algebra is just $K$.

Our investigation of polytopal algebras is motivated by two closely related goals: (1) to find the connections between the combinatorial structure of $P$ and the algebraic structure of $K[P]$, and (2) to extend theorems valid in $\mathrm{Vect}(K)$ to $\mathrm{Pol}(K)$.

It follows from Gubeladze [**Gu3**] that an algebra isomorphism of $K[P]$ and $K[Q]$ implies the isomorphism of $P$ and $Q$ as lattice polytopes. This result identifies the objects of the category Pol of lattice polytopes with the objects of $\mathrm{Pol}(K)$, but there remains the question to which extent the morphisms in $\mathrm{Pol}(K)$ are determined by those in Pol, namely the $\mathbb{Z}$-affine maps between lattice polytopes. Similarly one must ask whether certain classes of morphisms in $\mathrm{Pol}(K)$ can be described in the same way as the corresponding classes in $\mathrm{Vect}(K)$.

As we will see in Subsections 5.2 – 5.5, there is a total analogy with the linear situation for the *automorphism groups* in $\mathrm{Pol}(K)$ (called *polytopal linear groups* in [**BG1**]): they are generated by elementary automorphisms (generalizing elementary matrices), toric automorphisms (generalizing diagonal invertible matrices) and automorphisms of the underlying polytope; moreover, there are normal forms for such representations of arbitrary automorphisms.

We will apply the main theorem on automorphism groups of polytopal algebras in order to describe the automorphism groups of projective toric varieties; see Subsection 5.5.

In [**BG6**] this analogy has been extended to automorphisms of so-called polyhedral algebras. These algebras, associated with polyhedral complexes, are composed from polytopal algebras, in the same way as Stanley-Reisner rings of simplicial complexes are composed from polynomial rings. However, the combinatorics becomes much more difficult, and the results are not as complete as hoped for.

In Subsections 5.6 – 5.9 we study *retractions* of polytopal algebras, i. e. idempotent homogeneous endomorphisms of polytopal algebras. Our results support the following conjectures:

***Conjecture A.*** — *Retracts of polytopal algebras are again polytopal.*

***Conjecture B.*** — *A codimension 1 retraction factors through either a facet retraction or an affine lattice retraction of the underlying lattice polytope.*

These conjectures generalize the standard facts that every finitely generated vector space has a basis and that an idempotent matrix is conjugate to a sub-unit matrix (ones and zeros on the main diagonal and zeros anywhere else). Conjecture B must be restricted to codimension 1 since there exist counterexamples for higher codimension.

In Subsection 5.8 we discuss the class of *segmentonomial* ideals, that is, ideals generated by polynomials $f$ whose Newton polytope has dimension $\leqslant 1$.

Subsection 5.10 contains a conjecture on the structure of all morphisms in $\mathrm{Pol}(K)$ according to which all morphisms can be obtained by 5 basic operations, namely free extensions, Minkowski sums, homothetic blow-ups, restrictions to subpolytopes (or polytope changes) and compositions. The theorems on automorphism groups and retractions can be viewed as strong versions of this conjecture for special classes of homomorphisms.

A result of [**BG5**] that belongs to the program of this section, but is not discussed in these notes, is the triviality of the Picard group of $\mathrm{Pol}(K)$ for algebraically closed $K$. (The Picard group is the group of covariant "algebraic" autoequivalences.)

Retracts of free modules are projective modules. Therefore the study of algebra retracts can be considered as a non-linear variant of studying the group $K_0$ of a ring. The group $K_1$ compares automorphisms of free modules to the elementary ones, as does our theorem on the automorphisms in $\mathrm{Pol}(K)$. Therefore the latter is a non-linear analogue of $K_1$, and it is natural to push the analogy between $\mathrm{Vect}(K)$ and $\mathrm{Pol}(K)$ further into higher $K$-theory. This will be done in [**BG8**] and [**BG9**].

**5.2. Column structures on lattice polytopes.** — Let $P$ be a lattice polytope as above.

***Definition 5.2.1.*** — An element $v \in \mathbb{Z}^n$, $v \neq 0$, is a *column vector* (for $P$) if there is a facet $F \subset P$ such that $x + v \in P$ for every lattice point $x \in P \setminus F$.

For such $P$ and $v$ the pair $(P, v)$ is called a *column structure*. The corresponding facet $F$ is called its *base facet* and denoted by $P_v$.

One sees easily that for a column structure $(P, v)$ the set of lattice points in $P$ is contained in the union of rays – *columns* – parallel to the vector $-v$ and with end-points in $F$. This is illustrated by Figure 6.

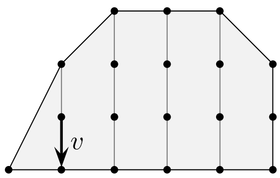


FIGURE 6. A column structure

Moreover, as an affine space over $\mathbb{Z}$, the group $\mathbb{Z}^n$ is the direct sum of the two subgroups generated by $v$ and by the lattice points in $P_v$ respectively. In particular, $v$ is a unimodular element of $\mathbb{Z}^n$. In the following we will identify a column vector $v \in \mathbb{Z}^n$ with the element $(v, 0) \in \mathbb{Z}^{n+1}$. The proof of the next lemma is straightforward. For simplicity of notation we set $C(P) = C(S_P)$.

***Lemma 5.2.2***. — *For a column structure $(P, v)$ and any element $x \in S_P$, such that $x \notin C(P_v)$, one has $x + v \in S_P$ (here $C(P_v)$ denotes the facet of $C(P)$ corresponding to $P_v$).*

One can easily control column structures in such formations as homothetic images and direct products of lattice polytopes. More precisely, let $P_i$ be a lattice $n_i$-polytope, $i = 1, 2$, and let $c$ be a natural number. Then $cP_1$ is the homothetic image of $P_1$ centered at the origin with factor $c$ and $P_1 \times P_2$ is the direct product of the two polytopes, realized as a lattice polytope in $\mathbb{Z}^{n_1+n_2}$ in a natural way. Then one has the following observations:

($*$) For any natural number $c$ the two polytopes $P_1$ and $cP_1$ have the same column vectors.

($**$) The system of column vectors of $P_1 \times P_2$ is the disjoint union of those of $P_1$ and $P_2$ (embedded into $\mathbb{Z}^{n_1+n_2}$).

Actually, ($*$) is a special case of a more general observation on the polytopes defining the same normal fan. The normal fan $\mathcal{N}(P)$ of a (lattice) polytope $P \subset \mathbb{R}^n$ is the family of cones in the dual space $(\mathbb{R}^n)^*$ given by

$$\mathcal{N}(P) = \big(\{\varphi \in (\mathbb{R}^n)^* \mid \mathrm{Max}_P(\varphi) = f\},\ f \text{ a face of } P\big);$$

here $\mathrm{Max}_P(\varphi)$ is the set of those points in $P$ at which $\varphi$ attains its maximal value on $P$ (for example, see Gelfand, Kapranov, and Zelevinsky [**GKZ**]).

For each facet $F$ of $P$ there exists a unique unimodular $\mathbb{Z}$-linear form $\varphi_F : \mathbb{Z}^n \to \mathbb{Z}$ and a unique integer $a_F$ such that $F = \{x \in P \mid \varphi_F(x) = a_F\}$ and

$$P = \{x \in \mathbb{R}^n \mid \varphi_F(x) \geqslant a_F \text{ for all facets } F\},$$

where we denote the natural extension of $\varphi_F$ to an $\mathbb{R}$-linear form on $\mathbb{R}^n$ by $\varphi_F$, too.

That $v$ is a column vector for $P$ with base facet $F$ can now be described as follows: one has $\varphi_F(v) = -1$ and $\varphi_G(v) \geqslant 0$ for all other facets $G$ of $P$. The linear forms

$-\varphi_F$ generate the semigroups of lattice points in the rays (i. e. one-dimensional cones) belonging to $\mathcal{N}(P)$ so that the system of column vectors of $P$ is completely determined by $\mathcal{N}(P)$:

(∗∗∗) Lattice $n$-polytopes $P_1$ and $P_2$ such that $\mathcal{N}(P_1) = \mathcal{N}(P_2)$ have the same systems of column vectors.

We have actually proved a slightly stronger result: if $P$ and $Q$ are lattice polytopes such that $\mathcal{N}(Q) \subset \mathcal{N}(P)$, then $\operatorname{Col}(P) \subset \operatorname{Col}(Q)$.

We further illustrate the notion of column vector by Figure 7: the polytope $P_1$ has 4 column vectors, whereas the polytope $P_2$ has no column vector.


FIGURE 7. Two polytopes and their column structures

Let $(P, v)$ be a column structure. Then for each element $x \in S_P$ we set $\operatorname{ht}_v(x) = m$ where $m$ is the largest non-negative integer for which $x + mv \in S_P$. Thus $\operatorname{ht}_v(x)$ is the 'height' of $x$ above the facet of the cone $C(S_P)$ corresponding to $P_v$ in direction $-v$.

More generally, for any facet $F \subset P$ we define the linear form $\operatorname{ht}_F \colon \mathbb{R}^{n+1} \to \mathbb{R}$ by $\operatorname{ht}_F(y) = \varphi_F(y_1, \dots, y_n) - a_F y_{n+1}$ where $\varphi_F$ and $a_F$ are chosen as above. For $x \in S_P$ and, more generally, for $x \in C(P) \cap \mathbb{Z}^{n+1}$ the height $\operatorname{ht}_F(x)$ of $X$ is a non-negative integer. The kernel of $\operatorname{ht}_F$ is just the hyperplane supporting $C(P)$ in the facet corresponding to $F$, and $C(P)$ is the cone defined by the support functions $\operatorname{ht}_F$.

Clearly, for a column structure $(P, v)$ and a lattice point $x \in P$ we have $\operatorname{ht}_v(x) = \operatorname{ht}_{P_v}(x)$, as follows immediately from Lemma 5.2.2.

In previous sections $\operatorname{ht}_F$ was denoted by $\sigma_i$ for the facet $F = F_i$. Here we have chosen the notion ht because of its geometric significance.

**5.3. The structure of the automorphism group.** — Let $(P, v)$ be a column structure and $\lambda \in K$. We identify the vector $v$, representing the difference of two lattice points in $P$, with the degree 0 element $(v, 0) \in \mathbb{Z}^{n+1}$, and also with the corresponding monomial in $K[\mathbb{Z}^{n+1}]$. Then we define an *injective* mapping from $S_P$ to $\operatorname{QF}(K[P])$, the quotient field of $K[P]$ by the assignment

$$x \mapsto (1 + \lambda v)^{\operatorname{ht}_v(x)} x.$$

Since $\operatorname{ht}_v$ extends to a group homomorphism $\mathbb{Z}^{n+1} \to \mathbb{Z}$ our mapping is a homomorphism from $S_P$ to the multiplicative group of $\operatorname{QF}(K[P])$. Now it is immediate from the definition of $\operatorname{ht}_v$ and Lemma 5.2.2 that the (isomorphic) image of $S_P$ lies actually in $K[P]$. Hence this mapping gives rise to a graded $K$-algebra endomorphism $e_v^\lambda$ of $K[P]$ preserving the degree of an element. By Hilbert function arguments $e_v^\lambda$ is an automorphism.

Here is an alternative description of $e_v^\lambda$. By a suitable integral change of coordinates we may assume that $v = (0, -1, 0, \dots, 0)$ and that $P_v$ lies in the subspace $\mathbb{R}^{n-1}$ (thus $P$ is in the upper halfspace). Now consider the standard unimodular $n$-simplex $\Delta_n$ with vertices at the origin and standard coordinate vectors. It is clear that there is a sufficiently large natural number $c$, such that $P$ is contained in a parallel translate of $c\Delta_n$ by a vector from $\mathbb{Z}^{n-1}$. Let $\Delta$ denote such a parallel translate. Then we have a graded $K$-algebra embedding $K[P] \subset K[\Delta]$. Moreover, $K[\Delta]$ can be identified with the $c$-th Veronese subring of the polynomial ring $K[x_0, \dots, x_n]$ in such a way that $v = x_0/x_1$. Now the automorphism of $K[x_0, \dots, x_n]$ mapping $x_1$ to $x_1 + \lambda x_0$ and leaving all the other variables invariant induces an automorphism $\alpha$ of the subalgebra $K[\Delta]$, and $\alpha$ in turn can be restricted to an automorphism of $K[P]$, which is nothing else but $e_v^\lambda$.

From now on we denote the graded automorphism group of $K[P]$ by

$$\Gamma_K(P).$$

It is clear from this description of $e_v^\lambda$ that it becomes an elementary matrix ($e_{01}^\lambda$ in our notation) in the special case when $P = \Delta_n$, after the identification $\Gamma_K(P) = \mathrm{GL}_{n+1}(K)$.

Therefore the automorphisms of type $e_v^\lambda$ will be called *elementary*.

***Lemma 5.3.1***. — *Let $v_1, \dots, v_s$ be pairwise different column vectors for $P$ with the same base facet $F = P_{v_i}$, $i = 1, \dots, s$.*

(a) *Then the mapping*

$$\varphi\colon \mathbb{A}_K^s \longrightarrow \Gamma_K(P), \qquad (\lambda_1, \dots, \lambda_s) \longmapsto e_{v_1}^{\lambda_1} \circ \cdots \circ e_{v_s}^{\lambda_s},$$

*is an embedding of algebraic groups. In particular, $e_{v_i}^{\lambda_i}$ and $e_{v_j}^{\lambda_j}$ commute for any $i, j \in \{1, \dots, s\}$ and the inverse of $e_{v_i}^{\lambda_i}$ is $e_{v_i}^{-\lambda_i}$.*

(b) *For $x \in L_P$ with $\mathrm{ht}_{v_1}(x) = 1$ one has*

$$e_{v_1}^{\lambda_1} \circ \cdots \circ e_{v_s}^{\lambda_s}(x) = (1 + \lambda_1 v_1 + \cdots + \lambda_s v_s)x.$$

($\mathbb{A}_K^s$ denotes the additive group of the $s$-dimensional affine space.)

*Proof.* — We define a new $K$-algebra automorphism $\vartheta$ of $K[P]$ by first setting

$$\vartheta(x) = (1 + \lambda_1 v_1 + \cdots + \lambda_s v_s)^{\mathrm{ht}_F(x)} x,$$

for $x \in S_P$ and then extending $\vartheta$ linearly. Arguments very similar to those above show that $\vartheta$ is a graded $K$-algebra automorphism of $K[P]$. The lemma is proved once we have verified that $\varphi = \vartheta$.

Choose a lattice point $x \in P$ such that $\mathrm{ht}_F(x) = 1$. (The existence of such a point follows from the definition of a column vector: there is of course a lattice point

$x \in P$ such that $\mathrm{ht}_F(x) > 0$.) We know that $\mathrm{gp}(S_P) = \mathbb{Z}^{n+1}$ is generated by $x$ and the lattice points in $F$. The lattice points in $F$ are left unchanged by both $\vartheta$ and $\varphi$, and elementary computations show that $\varphi(x) = (1 + \lambda_1 v_1 + \cdots + \lambda_s v_s)x$; hence $\varphi(x) = \vartheta(x)$. □

The image of the embedding $\varphi$ given by Lemma 5.3.1 is denoted by $\mathbb{A}(F)$. Of course, $\mathbb{A}(F)$ may consist only of the identity map of $K[P]$, namely if there is no column vector with base facet $F$. In the case in which $P$ is the unit simplex and $K[P]$ is the polynomial ring, $\mathbb{A}(F)$ is the subgroup of all matrices in $\mathrm{GL}_n(K)$ that differ from the identity matrix only in the non-diagonal entries of a fixed column.

For the statement of the main result we have to introduce some subgroups of $\Gamma_K(P)$. First, the $(n+1)$-torus $\mathbb{T}_{n+1} = (K^*)^{n+1}$ acts naturally on $K[P]$ by restriction of its action on $K[\mathbb{Z}^{n+1}]$ that is given by

$$(\xi_1, \ldots, \xi_{n+1})(e_i) = \xi_i e_i, \; i \in [1, n+1],$$

here $e_i$ is the $i$-th standard basis vector of $\mathbb{Z}^{n+1}$. This gives rise to an algebraic embedding $\mathbb{T}_{n+1} \subset \Gamma_K(P)$, and we will identify $\mathbb{T}_{n+1}$ with its image. It consists precisely of those automorphisms of $K[P]$ which multiply each monomial by a scalar from $K^*$.

Second, the automorphism group $\Sigma(P)$ of the semigroup $S_P$ is in a natural way a finite subgroup of $\Gamma_K(P)$. It is the *integral symmetry group* of $P$, i. e. the group of integral affine transformations mapping $P$ onto itself. (In general this group is larger than the group of symmetries with respect to the Euclidean metric.)

Third we have to consider a subgroup of $\Sigma(P)$ defined as follows. Assume $v$ and $-v$ are both column vectors. Then for every point $x \in P \cap \mathbb{Z}^n$ there is a unique $y \in P \cap \mathbb{Z}^n$ such that $\mathrm{ht}_v(x) = \mathrm{ht}_{-v}(y)$ and $x - y$ is parallel to $v$. The mapping $x \mapsto y$ gives rise to a semigroup automorphism of $S_P$: it 'inverts columns' that are parallel to $v$. It is easy to see that these automorphisms generate a normal subgroup of $\Sigma(P)$, which we denote by $\Sigma(P)_{\mathrm{inv}}$.

Finally, $\mathrm{Col}(P)$ is the set of column structures on $P$. Now the main result is

***Theorem 5.3.2***. — *Let $P$ be a convex lattice $n$-polytope and $K$ a field.*

(a) *Every element $\gamma \in \Gamma_K(P)$ has a (not uniquely determined) presentation*

$$\gamma = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_r \circ \tau \circ \sigma,$$

*where $\sigma \in \Sigma(P)$, $\tau \in \mathbb{T}_{n+1}$, and $\alpha_i \in \mathbb{A}(F_i)$ such that the facets $F_i$ are pairwise different and $\#(F_i \cap \mathbb{Z}^n) \leqslant \#(F_{i+1} \cap \mathbb{Z}^n)$, $i \in [1, r-1]$.*

(b) *For an infinite field $K$ the connected component of unity $\Gamma_K(P)^0 \subset \Gamma_K(P)$ is generated by the subgroups $\mathbb{A}(F_i)$ and $\mathbb{T}_{n+1}$. It consists precisely of those graded automorphisms of $K[P]$ which induce the identity map on the divisor class group of the normalization of $K[P]$.*

(c) $\dim \Gamma_K(P) = \# \mathrm{Col}(P) + n + 1$.

(d) *One has* $\Gamma_K(P)^0 \cap \Sigma(P) = \Sigma(P)_{\mathrm{inv}}$ *and* $\Gamma_K(P)/\Gamma_K(P)^0 \cong \Sigma(P)/\Sigma(P)_{\mathrm{inv}}$. *Furthermore, if $K$ is infinite, then $\mathbb{T}_{n+1}$ is a maximal torus of $\Gamma_K(P)$.*

***Remark 5.3.3***

(a) Our theorem is a 'polytopal generalization' of the fact that any invertible matrix with entries from a field is a product of elementary matrices, permutation matrices and diagonal matrices. The normal form in its part (a) generalizes the fact that the elementary transformations $e_{ij}^{\lambda}$, $j$ fixed, can be carried out consecutively.

(b) Observe that we do not claim the existence of a normal form as in (a) for the elements from $\Gamma_K(P)^0$ if we exclude elements of $\Sigma(P)_{\mathrm{inv}}$ from the generating set.

(c) Let $S \subset \mathbb{Z}^{n+1}$ be a normal affine semigroup such that 0 is the only invertible element in $S$. *A priori* $S$ does not have a grading, but there always exists a grading of $S$ such that the number of elements of a given degree is finite, as observed in Subsection 2.1.

One can treat graded automorphisms of such semigroups as follows. It is well known that the cone $C(S)$ spanned by $S$ in $\mathbb{R}^{n+1}$ is a finite rational strictly convex cone. An element $v \in \mathbb{Z}^{n+1}$ of degree 0 is called a column vector for $S$ if there is a facet $F$ of $C(S)$ such that $x + v \in S$ for every $x \in S \smallsetminus F$.

The only disadvantage here is that the condition for column vectors involves an infinite number of lattice points, while for polytopal algebras one only has to look at lattice points in a finite polytope (due to Lemma 5.2.2).

Then one introduces analogously the notion of an elementary automorphism $e_v^{\lambda}$ ($\lambda \in K$). The proof of Theorem 5.3.2 we present below is applicable to this more general situation without any essential change, yielding a similar result for the group of graded $K$-automorphisms of $K[S]$.

(d) In an attempt to generalize the theorem in a different direction, one could consider an arbitrary finite subset $M$ of $\mathbb{Z}^n$ (with $\mathrm{gp}(M) = \mathbb{Z}^n$) and the semigroup $S_M$ generated by the elements $(x, 1) \in \mathbb{Z}^{n+1}$, $x \in M$. However, examples show that there is no suitable notion of column vector in this generality: one can only construct the polytope $P$ spanned by $M$, find the automorphism group of $K[P]$ and try to determine $\Gamma_K(M)$ as the subgroup of those elements of $\Gamma_K(P)$ that can be restricted to $K[S_M]$. (This approach is possible because $K[P]$ is contained in the normalization of $K[S_M]$.)

(e) As a (possibly non-reduced) affine variety $\Gamma_K(P)$ is already defined over the prime field $K_0$ of $K$ since this is true for the affine variety $\operatorname{Spec} K[P]$. Let $S$ be its coordinate ring over $K_0$. Then the dimension of $\Gamma_K(P)$ is just the Krull dimension of $S$ or $S \otimes K$, and part (c) of the theorem must be understood accordingly.

As an application to rings and varieties outside the class of semigroup algebras and toric varieties we determine the groups of graded automorphisms of the determinantal rings, a result which goes back to Frobenius [**Fr**, p. 99] and has been re-proved many times since then. See, for instance, [**Wa**] for a group-scheme theoretical approach for general commutative rings of coefficients, covering also the classes of generic symmetric and alternating matrices. (The generic symmetric case can be done by the same method as the generic one below.)

In plain terms, Corollary 5.3.4 answers the following question: let $K$ be an infinite field, $\varphi\colon K^{mn} \to K^{mn}$ a $K$-automorphism of the vector space $K^{mn}$ of $m \times n$ matrices over $K$, and $r$ an integer, $1 \leqslant r < \min(m,n)$; when is $\operatorname{rank}\varphi(A) \leqslant r$ for all $A \in K^{mn}$ with $\operatorname{rank} A \leqslant r$? This holds obviously for transformations $\varphi(A) = SAT^{-1}$ with $S \in \mathrm{GL}_m(K)$ and $T \in \mathrm{GL}_n(K)$, and for the transposition if $m = n$. Indeed, these are the only such transformations:

***Corollary 5.3.4.*** — *Let $K$ be a field, $X$ an $m \times n$ matrix of indeterminates, and set $R = K[X]/I_{r+1}(X)$ the residue class ring of the polynomial ring $K[X]$ in the entries of $X$ modulo the ideal generated by the $(r+1)$-minors of $X$, $1 \leqslant r < \min(m,n)$. Set $G = \operatorname{gr.aut}_K(R)$.*

(a) *The connected component $G^0$ of unity in $G$ is the image of $\mathrm{GL}_m(K) \times \mathrm{GL}_n(K)$ in $\mathrm{GL}_{mn}(K)$ under the map above, and is isomorphic to $\mathrm{GL}_m(K) \times \mathrm{GL}_n(K)/K^*$ where $K^*$ is embedded diagonally.*

(b) *If $m \neq n$, the group $G$ is connected, and if $m = n$, then $G^0$ has index 2 in $G$ and $G = G^0 \cup \tau G^0$ where $\tau$ is the transposition.*

*Proof.* — The singular locus of $\operatorname{Spec} R$ is given by $V(\mathfrak{p})$ where $\mathfrak{p} = I_r(X)/I_{r+1}(X)$; $\mathfrak{p}$ is a prime ideal in $R$ (see Bruns and Vetter [**BV**, (2.6), (6.3)]). It follows that every automorphism of $R$ must map $\mathfrak{p}$ onto itself. Thus a linear substitution on $K[X]$ for which $I_{r+1}(X)$ is stable also leaves $I_r(X)$ invariant and therefore induces an automorphism of $K[X]/I_r(X)$. This argument reduces the corollary to the case $r = 1$.

For $r = 1$ one has the isomorphism

$$R \to K[Y_iZ_j : i = 1,\dots,m,\ j = 1,\dots,n] \subset K[Y_1,\dots,Y_m,Z_1,\dots,Z_n]$$

induced by the assignment $X_{ij} \mapsto Y_iZ_j$. Thus $R$ is just the Segre product of $K[Y_1,\dots,Y_m]$ and $K[Z_1,\dots,Z_n]$, or, equivalently, $R \cong K[P]$ where $P$ is the direct product of the unit simplices $\Delta_{m-1}$ and $\Delta_{n-1}$. Part (a) follows now from an analysis of the column structures of $P$ (see observation $(**)$ above) and the torus actions.

For (b) one observes that $\mathrm{Cl}(R) \cong \mathbb{Z}$; ideals representing the divisor classes 1 and $-1$ are given by $(Y_1Z_1,\dots,Y_1Z_n)$ and $(Y_1Z_1,\dots,Y_mZ_1)$ [**BV**, 8.4]. If $m \neq n$, these ideals have different numbers of generators; therefore every automorphism of $R$ acts trivially on the divisor class group. In the case $m = n$, the transposition induces the map $s \mapsto -s$ on $\mathrm{Cl}(R)$. Now the rest follows again from the theorem above. (Instead of the divisorial arguments one could also discuss the symmetry group of $\Delta_{m-1} \times \Delta_{n-1}$.) □

**5.4. The Gaussian algorithm for polytopes.** — The Gaussian algorithm tells us how to transform a matrix to a diagonal matrix. Theorem 5.3.2 claims that such a diagonalization is possible for automorphisms of polytopal algebras, and we will carry it out by a procedure generalizing the Gaussian algorithm.

Set $\overline{S}_P = \mathbb{Z}^{n+1} \cap C(P)$. Then $\overline{S}_P$ is the normalization of the semigroup $S_P$ and $K[\overline{S}_P]$ is the normalization of the domain $K[P]$. Let $\overline{\Gamma}_K(P)$ denote the group of graded $K$-algebra automorphisms of $K[\overline{S}_P]$. Since any automorphism of $K[P]$ extends to a unique automorphism of $K[\overline{S}_P]$ we have a natural embedding $\Gamma_K(P) \subset \overline{\Gamma}_K(P)$. On the other hand, $K[\overline{S}_P]$ and $K[P]$ have the same homogeneous components of degree 1. Hence $\Gamma_K(P) = \overline{\Gamma}_K(P)$. Nevertheless we will use the notation $\overline{\Gamma}_K(P)$, emphasizing the fact that we are dealing with automorphisms of $K[\overline{S}_P]$; $\Sigma(P)$ and $\Sigma(P)_{\mathrm{inv}}$ will refer to their images in $\overline{\Gamma}_K(P)$. Furthermore, the extension of an elementary automorphism $e_v^\lambda$ is also denoted by $e_v^\lambda$; it satisfies the rule $e_v^\lambda(x) = (1+\lambda v)^{\mathrm{ht}_v(x)}x$ for all $x \in \overline{S}_P$. (The equation $\Gamma_K(P) = \overline{\Gamma}_K(P)$ shows that the situation considered in Remark 5.3.3(c) really generalizes Theorem 5.3.2; furthermore it explains the difference between polytopal algebras and arbitrary graded semigroup algebras generated by their degree 1 elements.)

In the following it is sometimes necessary to distinguish elements $x \in \overline{S}_P$ from products $\zeta z$ with $\zeta \in K$ and $z \in \overline{S}_P$. As introduced in Subsection 2.1, we call $x$ a monomial and $\zeta z$ a term.

Suppose $\gamma \in \overline{\Gamma}_K(P)$ maps every monomial $x$ to a term $\lambda_x y_x$, $y_x \in \overline{S}_P$. Then the assignment $x \mapsto y_x$ is also a semigroup automorphism of $\overline{S}_P$. Denote it by $\sigma$. It obviously belongs to $\Sigma(P)$. The mapping $\sigma^{-1} \circ \gamma$ is of the type $x \mapsto \xi_x x$, and clearly $\tau = \sigma^{-1} \circ \gamma \in \mathbb{T}_{n+1}$. Therefore, $\gamma = \sigma \circ \tau$.

Let $\mathrm{int}(C(P))$ denote the interior of the cone $C(P)$ and let

$$\omega = (\mathrm{int}(C(P)) \cap \mathbb{Z}^{n+1})K[\overline{S}_P]$$

be the corresponding monomial ideal. (It is known that $\omega$ is the canonical module of $K[\overline{S}_P]$: see Danilov [**Da**], Stanley [**St1**], or [**BH**, Chapter 6].)

***Lemma 5.4.1***

(a) *Suppose $\gamma \in \overline{\Gamma}_K(P)$ leaves the ideal $\omega$ invariant. Then $\gamma = \sigma \circ \tau$ with $\sigma \in \Sigma(P)$ and $\tau \in \mathbb{T}_{n+1}$.*

(b) *One has $\sigma \circ \tau \circ \sigma^{-1} \in \mathbb{T}_{n+1}$ for all $\sigma \in \Sigma(P)$, $\tau \in \mathbb{T}_{n+1}$.*

*Proof*

(a) By the argument above it is enough that $\gamma$ maps monomials to terms.

First consider a non-zero monomial $x \in \overline{S}_P \cap \omega$. We have $\gamma(x) \in \omega$. Since $x$ is an 'interior' monomial, $K[\overline{S}_P]_x = K[\mathbb{Z}^{n+1}]$ (see the end of Subsection 2.1). On the other hand $K[\mathbb{Z}^{n+1}] \subset K[\overline{S}_P]_{\gamma(x)}$. Indeed, since $\mathrm{gp}(\overline{S}_P) = \mathbb{Z}^{n+1}$, it just suffices to observe that for any monomial $z \in \overline{S}_P$ there is a sufficiently large natural number $c$ satisfying the following condition:

> The parallel translate of the Newton polytope $N(\gamma(x)^c)$ by the vector $-z \in \mathbb{R}^{n+1}$ is inside the cone $C(P)$ (here we use additive notation).

(Observe that $N(\gamma(x)^c)$ is the homothetic image of $N(\gamma(x))$, centered at the origin with factor $c$. Instead of Newton polytopes one could also use the minimal prime ideals of $z$, which we will introduce below: they all contain $\gamma(x)$.) Hence all monomials become invertible in $K[\overline{S}_P]_{\gamma(x)}$.

The crucial point is to compare the groups of units $U(K[\mathbb{Z}^{n+1}]) = K^* \oplus \mathbb{Z}^{n+1}$ and $U(K[\overline{S}_P]_{\gamma(x)})$. The mapping $\gamma$ induces an isomorphism

$$\mathbb{Z}^{n+1} \cong U(K[\overline{S}_P]_{\gamma(x)})/K^*.$$

On the other hand we have seen that $\mathbb{Z}^{n+1}$ is embedded into $U(K[\overline{S}_P]_{\gamma(x)})/K^*$ so that the elements from $\overline{S}_P$ map to their classes in the quotient group.

Assume that $\gamma(x)$ is not a term. Then none of the powers of $\gamma(x)$ is a term. In other words, none of the multiples of the class of $\gamma(x)$ is in the image of $\mathbb{Z}^{n+1}$. This shows that $\operatorname{rank}(U(K[\overline{S}_P]_{\gamma(x)})/K^*) > n+1$, a contradiction.

Now let $y \in \overline{S}_P$ be an arbitrary monomial, and $z$ a monomial in $\omega$. Then $yz \in \omega$, and since $\gamma(yz)$ is a term as shown above, $\gamma(y)$ must be also a term.

(b) follows immediately from the fact that $\sigma \circ \tau \circ \sigma^{-1}$ maps each monomial to a multiple of itself. $\square$

In the light of Lemma 5.4.1(a) we see that for Theorem 5.3.2(a) it suffices to show the following claim: for every $\gamma \in \overline{\Gamma}_K(P)$ there exist $\alpha_1 \in \mathbb{A}(F_1), \dots, \alpha_r \in \mathbb{A}(F_r)$ such that

$$\alpha_r \circ \alpha_{r-1} \circ \cdots \circ \alpha_1 \circ \gamma(\omega) = \omega$$

and the $F_i$ satisfy the side conditions of 5.3.2(a).

For a facet $F \subset P$ we have constructed the group homomorphism $\operatorname{ht}_F\colon \mathbb{Z}^{n+1} \to \mathbb{Z}$. Recall that

$$\operatorname{Div}(F) = \{x \in \overline{S}_P \mid \operatorname{ht}_F(x) > 0\} \cdot K[\overline{S}_P].$$

is the divisorial prime ideal of $K[\overline{S}_P]$ associated with the facet $F$. It is clear that $\omega = \bigcap_1^r \operatorname{Div}(F_i)$ where $F_1, \dots, F_r$ are the facets of $P$. This shows the importance of the ideals $\operatorname{Div}(F_i)$ for our goals. In the following Theorem 2.3.1 is an important tool.

Before we prove the claim above (reformulated as Lemma 5.4.4) we collect some auxiliary arguments.

***Lemma 5.4.2***. — *Let $v_1, \dots, v_s$ be column vectors with the common base facet $F = P_{v_i}$, and $\lambda_1, \dots, \lambda_s \in K$. Then*

$$e_{v_1}^{\lambda_1} \circ \cdots \circ e_{v_s}^{\lambda_s}(\operatorname{Div}(F)) = (1 + \lambda_1 v_1 + \cdots + \lambda_s v_s)\operatorname{Div}(F)$$

*and*

$$e_{v_1}^{\lambda_1} \circ \cdots \circ e_{v_s}^{\lambda_s}(\operatorname{Div}(G)) = \operatorname{Div}(G), \qquad G \neq F.$$

*Proof.* — Using the automorphism $\vartheta$ from the proof of Lemma 5.3.1 we see immediately that

$$e_{v_1}^{\lambda_1}\circ\cdots\circ e_{v_s}^{\lambda_s}(\mathrm{Div}(F))\subset(1+\lambda_1v_1+\cdots+\lambda_sv_s)\mathrm{Div}(F).$$

The left hand side is a height 1 prime ideal (being an automorphic image of such) and the right hand side is a proper divisorial ideal inside $K[\overline{S}_P]$. Then, of course, the inclusion is an equality.

For the second assertion it is enough to treat the case $s=1$, $v=v_1$, $\lambda=\lambda_1$. One has

$$e_v^{\lambda}(x)=(1+\lambda v)^{\mathrm{ht}_F(x)}x,$$

and all the terms in the expansion of the right hand side belong to $\mathrm{Div}(G)$ since $\mathrm{ht}_G(v)\geqslant 0$. As above, the inclusion $e_v^{\lambda}(\mathrm{Div}(G))\subset\mathrm{Div}(G)$ implies equality. □

**Lemma 5.4.3.** — *Let $F\subset P$ be a facet, $\lambda_1,\dots,\lambda_s\in K\setminus\{0\}$ and $v_1,\dots,v_s\in\mathbb{Z}^{n+1}$ be pairwise different non-zero elements of degree 0. Suppose $(1+\lambda_1v_1+\cdots+\lambda_sv_s)\mathrm{Div}(F)\subset K[\overline{S}_P]$. Then $v_1,\dots,v_s$ are column vectors for $P$ with the common base facet $F$.*

*Proof.* — If $x\in P\setminus F$ is a lattice point, then $x\in\mathrm{Div}(F)$. Thus $xv_j$ is a degree 1 element of $S_P$; in additive notation this means $x+v_j\in P$. □

The crucial step in the proof of our main result is the next lemma.

**Lemma 5.4.4.** — *Let $\gamma\in\overline{\Gamma}_K(P)$, and enumerate the facets $F_1,\dots,F_r$ of $P$ in such a way that $\#(F_i\cap\mathbb{Z}^n)\leqslant\#(F_{i+1}\cap\mathbb{Z}^n)$ for $i\in[1,r-1]$. Then there exists a permutation $\pi$ of $[1,r]$ such that $\#(F_i\cap\mathbb{Z}^n)=\#(F_{\pi(i)}\cap\mathbb{Z}^n)$ for all $i$ and*

$$\alpha_r\circ\cdots\circ\alpha_1\circ\gamma(\mathrm{Div}(F_i))=\mathrm{Div}(F_{\pi(i)})$$

*with suitable $\alpha_i\in\mathbb{A}(F_{\pi(i)})$.*

In fact, this lemma finishes the proof of Theorem 5.3.2(a): the resulting automorphism $\delta=\alpha_r\circ\cdots\circ\alpha_1\circ\gamma$ permutes the minimal prime ideals of $\omega$ and therefore preserves their intersection $\omega$. By virtue of Lemma 5.4.1(a) we then have $\delta=\sigma\circ\tau$ with $\sigma\in\Sigma(P)$ and $\tau\in\mathbb{T}_{n+1}$. Finally one just replaces each $\alpha_i$ by its inverse and each $F_i$ by $F_{\pi(i)}$.

*Proof of Lemma 5.4.4.* — As mentioned above, the divisorial ideal $\gamma(\mathrm{Div}(F))\subset K[\overline{S}_P]$ is equivalent to some monomial divisorial ideal $\Delta$, i. e. there is an element $\kappa\in\mathrm{QF}(K[\overline{S}_P])$ such that

$$\gamma(\mathrm{Div}(F))=\kappa\Delta.$$

The inclusion $\kappa\in(\gamma(\mathrm{Div}(F)):\Delta)$ shows that $\kappa$ is a $K$-linear combination of some Laurent monomials corresponding to lattice points in $\mathbb{Z}^{n+1}$. We factor out one of the terms of $\kappa$, say $m$, and rewrite the above equality as follows:

$$\gamma(\mathrm{Div}(F))=(m^{-1}\kappa)(m\Delta).$$

Then $m^{-1}\kappa$ is of the form $1+m_1+\cdots+m_s$ for some Laurent terms $m_1,\dots,m_s\notin K$, while $m\Delta$ is necessarily a divisorial monomial ideal of $K[\overline{S}_P]$ (since 1 belongs to the supporting monomial set of $m^{-1}\kappa$). Now $\gamma$ is a *graded* automorphism. Hence

$$(1+m_1+\cdots+m_s)(m\Delta)\subset K[\overline{S}_P]$$

is a graded ideal. This implies that the terms $m_1,\dots,m_s$ are of degree 0. Thus there is always a presentation

$$\gamma(\operatorname{Div}(F))=(1+m_1+\cdots+m_s)\Delta,$$

where $m_1,\dots,m_s$ are Laurent terms of degree 0 and $\Delta\subset K[\overline{S}_P]$ is a monomial ideal (we do not exclude the case $s=0$). A representation of this type will be called *admissible*.

In the following we will have to work with the number of degree 1 monomials in a given monomial ideal $I$. Therefore we let $I_P$ denote the set of such monomials; in other words, $I_P$ is the set of lattice points in $P$ which are elements of $I$. Thus, we have

$$\big(\textstyle\bigcap_1^r \operatorname{Div}(F_i)^{(a_i)}\big)_P=\{x\in P\cap\mathbb{Z}^n \mid \operatorname{ht}_{F_i}(x)\geqslant a_i,\ i\in[1,r]\}$$

for all $a_i\geqslant 0$. (Recall that $\operatorname{ht}_{F_i}$ coincides on lattice points with the valuation of $\operatorname{QF}(K[P])$ defined by $\operatorname{Div}(F_i)$.) Furthermore we set

$$c_i=\#(\operatorname{Div}(F_i)_P).$$

Then $c_i=\#(P\cap\mathbb{Z}^n)-\#(F_i\cap\mathbb{Z}^n)$, and according to our enumeration of the facets we have $c_1\geqslant\cdots\geqslant c_r$.

For $\gamma\in\overline{\Gamma}_K(P)$ consider an admissible representation

$$\gamma(\operatorname{Div}(F_1))=(1+m_1+\cdots+m_s)\Delta.$$

Since $\gamma$ is graded, $\#(\Delta_P)=c_1$: this is the dimension of the degree 1 homogeneous components of the ideals. As mentioned above, there are integers $a_i\geqslant 0$ such that

$$\Delta=\bigcap_1^r \operatorname{Div}(F_i)^{(a_i)}.$$

It follows easily that if $\sum_1^r a_i\geqslant 2$ and $a_{i_0}\neq 0$ for $i_0\in[1,r]$, then $\#(\Delta_P)<c_{i_0}$. This observation along with the maximality of $c_1$ shows that exactly one of the $a_i$ is 1 and all the others are 0. In other words, $\Delta=\operatorname{Div}(G_1)$ for some $G_1\in\{F_1,\dots,F_r\}$ containing precisely $\#(F_1\cap\mathbb{Z}^n)$ lattice points. By Lemmas 5.4.2 and 5.4.3 there exists $\alpha_1\in\mathbb{A}(G_1)$ such that

$$\alpha_1\circ\gamma(\operatorname{Div}(F_1))=\operatorname{Div}(G_1).$$

Now we proceed inductively. Let $1\leqslant t<r$. Assume there are facets $G_1,\dots,G_t$ of $P$ with $\#(G_i\cap\mathbb{Z}^n)=\#(F_i\cap\mathbb{Z}^n)$ and $\alpha_1\in\mathbb{A}(G_1),\dots,\alpha_t\in\mathbb{A}(G_t)$ such that

$$\alpha_t\circ\cdots\circ\alpha_1\circ\gamma(\operatorname{Div}(F_i))=\operatorname{Div}(G_i),\qquad i\in[1,t].$$

(Observe that the $G_i$ are automatically different.) In view of Lemma 5.4.2 we must show there is a facet $G_{t+1} \subset P$, different from $G_1, \dots, G_t$ and containing exactly $\#(F_{t+1} \cap \mathbb{Z}^n)$ lattice points, and an element $\alpha_{t+1} \in \mathbb{A}(G_{t+1})$ such that

$$\alpha_{t+1} \circ \alpha_t \circ \cdots \circ \alpha_1 \circ \gamma(\operatorname{Div}(F_{t+1})) = \operatorname{Div}(G_{t+1}).$$

For simplicity of notation we put $\gamma' = \alpha_t \circ \cdots \circ \alpha_1 \circ \gamma$. Again, consider an admissible representation

$$\gamma'(\operatorname{Div}(F_{t+1})) = (1 + m_1 + \cdots + m_s)\Delta.$$

Rewriting this equality in the form

$$\gamma'(\operatorname{Div}(F_{t+1})) = (m_j^{-1}(1 + m_1 + \cdots + m_s))(m_j\Delta),$$

where $j \in \{0, \dots, s\}$ and $m_0 = 1$, we get another admissible representation. Assume that by varying $j$ we can obtain a monomial divisorial ideal $m_j\Delta$, such that in the primary decomposition

$$m_j\Delta = \bigcap_1^r \operatorname{Div}(F_i)^{(a_i)}$$

there appears a positive power of $\operatorname{Div}(G)$ for some facet $G$ different from $G_1, \dots, G_t$. Then $\#((m_j\Delta)_P) \leqslant c_{t+1}$ (due to our enumeration) and the inequality is strict whenever $\sum_1^r a_i \geqslant 2$. On the other hand $\#(\operatorname{Div}(F_{t+1})_P) = c_{t+1}$. Thus we would have $m_j\Delta = \operatorname{Div}(G)$ and we could proceed as for the ideal $\operatorname{Div}(F_1)$).

Assume to the contrary that in the primary decompositions of all the monomial ideals $m_j\Delta$ there only appear the prime ideals $\operatorname{Div}(G_1), \dots, \operatorname{Div}(G_t)$. We have

$$(1 + m_1 + \cdots + m_s)\Delta \subset \Delta + m_1\Delta + \cdots + m_s\Delta$$

and

$$[(1 + m_1 + \cdots + m_s)\Delta] = [\Delta] = [m_1\Delta] = \cdots = [m_s\Delta]$$

in $\operatorname{Cl}(K[\overline{S}_P])$. Applying $(\gamma')^{-1}$ we arrive at the conclusion that $\operatorname{Div}(F_{t+1})$ is contained in a sum of *monomial* divisorial ideals $\Phi_0, \dots, \Phi_s$, such that the primary decomposition of each of them only involves $\operatorname{Div}(F_1), \dots, \operatorname{Div}(F_t)$. (This follows from the fact that $(\gamma')^{-1}$ maps $\operatorname{Div}(G_i)$ to the monomial ideal $\operatorname{Div}(F_i)$ for $i = 1, \dots, t$; thus intersections of symbolic powers of $\operatorname{Div}(G_1), \dots, \operatorname{Div}(G_t)$ are mapped to intersections of symbolic powers of $\operatorname{Div}(F_1), \dots, \operatorname{Div}(F_t)$, which are automatically monomial.) Furthermore, $\operatorname{Div}(F_{t+1})$ has the same divisor class as each of the $\Phi_i$.

Now choose a monomial $M \in \mathbb{Z}^{n+1} \cap \operatorname{Div}(F_{t+1})$ such that $\operatorname{ht}_{F_1}(M) + \cdots + \operatorname{ht}_{F_t}(M)$ is minimal. Since the monomial ideal $\operatorname{Div}(F_{t+1})$ is contained in the sum of the monomial ideals $\Phi_0, \dots, \Phi_s$, each monomial in it must belong to one of the ideals $\Phi_i$; so we may assume that $M \in \Phi_j$. There is a monomial $d$ with $\operatorname{Div}(F_{t+1}) = d\Phi_j$, owing to the fact that $\operatorname{Div}(F_{t+1})$ and $\Phi_j$ belong to the same divisor class. It is clear that $\operatorname{ht}_{F_i}(d) \leqslant 0$ for $i \in [1, t]$ and $\operatorname{ht}_{F_i}(d) < 0$ for at least one $i \in [1, t]$. In fact,

$$\operatorname{ht}_{F_i}(d) = -a_i \quad \text{for } i = 1 \dots, t,$$

where $\Phi_j = \bigcap_1^t \operatorname{Div}(F_i)^{(a_i)}$. If we had $a_i = 0$ for $i = 1, \dots, t$, then $\Phi_j = K[\overline{S}_P]$, which is evidently impossible. By the choice of $d$ the monomial $N = dM$ belongs to $\operatorname{Div}(F_{t+1})$. But

$$\operatorname{ht}_{F_1}(N) + \cdots + \operatorname{ht}_{F_t}(N) < \operatorname{ht}_{F_1}(M) + \cdots + \operatorname{ht}_{F_t}(M),$$

a contradiction. □

*Proof of Theorem 5.3.2*(b)–(d). — (b) Since $\mathbb{T}_{n+1}$ and the $\mathbb{A}(F_i)$ are connected groups they generate a connected subgroup $U$ of $\overline{\Gamma}_K(P)$ (see Borel [**Bor**, Prop. 2.2]). This subgroup acts trivially on $\operatorname{Cl}(K[\overline{S}_P])$ by Lemma 5.4.2 and the fact that the classes of the $\operatorname{Div}(F_i)$ generate the divisor class group. Furthermore $U$ has finite index in $\overline{\Gamma}_K(P)$ bounded by $\#\Sigma(P)$. Therefore $U = \overline{\Gamma}_K(P)^0$.

Assume $\gamma \in \overline{\Gamma}_K(P)$ acts trivially on $\operatorname{Cl}(K[\overline{S}_P])$. We want to show that $\gamma \in U$. Let $E$ denote the connected subgroup of $\overline{\Gamma}_K(P)$, generated by the elementary automorphisms. Since any automorphism that maps monomials to terms and preserves the divisorial ideals $\operatorname{Div}(F_i)$ is automatically a toric automorphism, by Lemma 5.4.1(a) we only have to show that there is an element $\varepsilon \in E$, such that

$$(1) \qquad \varepsilon \circ \gamma(\operatorname{Div}(F_i)) = \operatorname{Div}(F_i),\ i \in [1, r].$$

By Lemma 5.4.4 we know that there is $\varepsilon_1 \in E$ such that

$$(2) \qquad \varepsilon_1 \circ \gamma(\operatorname{Div}(F_j)) = \operatorname{Div}(F_{i_j}),\ j \in [1, r],$$

where $\{i_1, \dots, i_r\} = \{1, \dots, r\}$. Since $\varepsilon_1$ and $\gamma$ both act trivially on $\operatorname{Cl}(K[\overline{S}_P])$, we get

$$\operatorname{Div}(F_{i_j}) = m_{i_j} \operatorname{Div}(F_j),\ j \in [1, r],$$

for some monomials $m_{i_j}$ of degree 0.

By Lemma 5.4.3 we conclude that if $m_{i_j} \neq 1$ (in additive notation, $m_{i_j} \neq 0$), then both $m_{i_j}$ and $-m_{i_j}$ are column vectors with the base facets $F_j$ and $F_{i_j}$ respectively. Observe that the automorphism

$$\varepsilon_{i_j} = e^1_{m_{i_j}} \circ e^{-1}_{-m_{i_j}} \circ e^1_{m_{i_j}} \in E$$

interchanges the ideals $\operatorname{Div}(F_j)$ and $\operatorname{Div}(F_{i_j})$, provided $m_{i_j} \neq 1$. Now we can complete the proof by successively 'correcting' the equations (2).

(c) We have to compute the dimension of $\overline{\Gamma}_K(P)$. Without loss of generality we may assume that $K$ is algebraically closed, passing to the algebraic closure of $K$ if necessary (see Remark 5.3.3(e)). For every permutation $\rho\colon \{1, \dots, r\} \to \{1, \dots, r\}$ we have the algebraic map

$$\mathbb{A}(F_{\rho(1)}) \times \cdots \times \mathbb{A}(F_{\rho(r)}) \times \mathbb{T}_{n+1} \times \Sigma(P) \to \overline{\Gamma}_K(P),$$

induced by composition. The left hand side has dimension $\#\operatorname{Col}(P) + n + 1$. By Theorem 5.3.2(a) we are given a finite system of constructible sets, covering $\overline{\Gamma}_K(P)$. Hence $\dim \overline{\Gamma}_K(P) \leqslant \#\operatorname{Col}(P) + n + 1$.

To derive the opposite inequality we can additionally assume that $P$ contains an interior lattice point. Indeed, the observation $(*)$ in Subsection 5.2 and Theorem 5.3.2(a) show that the natural group homomorphism $\overline{\Gamma}_K(P) \to \overline{\Gamma}_K(cP)$, induced by restriction to the $c$-th Veronese subring, is surjective for every $c \in \mathbb{N}$ (the surjection for the 'toric part' follows from the fact that $K$ is closed under taking roots). So we can work with $cP$, which contains an interior point provided $c$ is large.

Let $x \in P$ be an interior lattice point and let $v_1, \dots, v_s$ be different column vectors. Then the supporting monomial set of $e_{v_i}^{\lambda_i}(x)$, $\lambda \in K^*$, is not contained in the union of those of $e_{v_j}^{\lambda_j}(x)$, $j \neq i$ (just look at the projections of $x$ through $v_i$ into the corresponding base facets). This shows that we have $\#\operatorname{Col}(P)$ linearly independent tangent vectors of $\overline{\Gamma}_K(P)$ at $1 \in \overline{\Gamma}_K(P)$. Since the tangent vectors corresponding to the elements of $\mathbb{T}_{n+1}$ clearly belong to a complementary subspace and $\overline{\Gamma}_K(P)$ is a smooth variety, we are done.

(d) Assume $v$ and $-v$ both are column vectors. Then the element

$$\varepsilon = e_v^1 \circ e_{-v}^{-1} \circ e_v^1 \in \overline{\Gamma}_K(P)^0 \ (= \Gamma_K(P)^0)$$

maps monomials to terms; more precisely, $\varepsilon$ 'inverts up to scalars' the columns parallel to $v$ so that any $x \in \overline{S}_P$ is sent either to the appropriate $y \in \overline{S}_P$ or to $-y \in K[\overline{S}_P]$. Then it is clear that there is an element $\tau \in \mathbb{T}_{n+1}$ such that $\tau \circ \varepsilon$ is a generator of $\Sigma(P)_{\text{inv}}$. Hence $\Sigma(P)_{\text{inv}} \subset \overline{\Gamma}_K(P)^0$.

Conversely, if $\sigma \in \Sigma(P) \cap \overline{\Gamma}_K(P)^0$ then $\sigma$ induces the identity map on $\operatorname{Cl}(K[\overline{S}_P])$. Hence $\sigma(\operatorname{Div}(F_i)) = m_{i_j}\operatorname{Div}(F_i)$ for some monomials $m_{i_j}$, and the very same arguments we have used in the proof of (b) show that $\sigma \in \Sigma(P)_{\text{inv}}$. Thus $\overline{\Gamma}_K(P)/\overline{\Gamma}_K(P)^0 = \Sigma(P)/\Sigma(P)_{\text{inv}}$.

Finally, assume $K$ is infinite and $\mathbb{T}' \subset \overline{\Gamma}_K(P)$ is a torus, strictly containing $\mathbb{T}_{n+1}$. Choose $x \in \overline{S}_P$ and $\gamma \in \mathbb{T}'$. Then $\tau^{-1} \circ \gamma \circ \tau(x) = \gamma(x)$ for all $\tau \in \mathbb{T}_{n+1}$. Since $K$ is infinite, one easily verifies that this is only possible if $\gamma(x)$ is a term. In particular, $\gamma$ maps monomials to terms. Then, as observed above Lemma 5.4.1, $\gamma = \sigma \circ \tau$ with $\mathbb{T}_{n+1}$, and therefore $\sigma \in \Sigma_0 = \mathbb{T}' \cap \Sigma$. Lemma5.4.1(b) now implies that $\mathbb{T}'$ is the semidirect product $\mathbb{T}_{n+1} \ltimes \Sigma_0$. By the infinity of $K$ we have $\Sigma_0 = 1$. □

**5.5. Projective toric varieties and their groups.** — Having determined the automorphism group of a polytopal semigroup algebra, we show in this subsection that our main result gives the description of the automorphism group of a projective toric variety (over an arbitrary algebraically closed field) via the existence of 'fully symmetric' polytopes.

The description of the automorphism group of a smooth complete toric $\mathbb{C}$-variety given by a fan $\mathcal{F}$ in terms of the roots of $\mathcal{F}$ is due to Demazure in his fundamental work [**De**]. The analogous description of the automorphism group of quasi-smooth complete toric varieties (over $\mathbb{C}$) has been obtained by Cox [**Cox**]. Buehler [**Bue**] generalized Cox' results to arbitrary complete toric varieties. We must restrict ourselves to projective toric varieties, but our method works in arbitrary characteristic.

We start with a brief review of some facts about projective toric varieties. Our terminology follows the standard references (Danilov [**Da**], Fulton [**Fu**], Oda [**Oda**]). To avoid technical complications we suppose from now on that the field $K$ is *algebraically closed*.

Let $P \subset \mathbb{R}^n$ be a polytope as above. Then $\operatorname{Proj}(K[\overline{S}_P])$ is a projective toric variety (though $K[\overline{S}_P]$ needs not be generated by its degree 1 elements). In fact, it is the toric variety defined by the normal fan $\mathcal{N}(P)$, but it may be useful to describe it additionally in terms of an affine covering.

For every vertex $z \in P$ we consider the finite rational polyhedral $n$-cone spanned by $P$ at its corner $z$. The parallel translate of this cone by $-z$ will be denoted by $C(z)$. Thus we obtain a system of the cones $C(z)$, where $z$ runs through the vertices of $P$. It is not difficult to check that $\mathcal{N}(P)$ is the fan in $(\mathbb{R}^n)^*$ whose maximal cones are the dual cones

$$C(z)^* = \{\varphi \in (\mathbb{R}^n)^* \mid \varphi(x) \geqslant 0 \text{ for all } x \in C(z)\}.$$

The affine open subschemes $\operatorname{Spec}(K[\mathbb{Z}^n \cap C(z)])$ cover $\operatorname{Proj}(K[\overline{S}_P])$. The projectivity of $\operatorname{Proj}(K[\overline{S}_P])$ follows from the observation that for all natural numbers $c \gg 0$ the polytope $cP$ is normal (see Subsection 3.4) and, hence, $\operatorname{Proj}(K[\overline{S}_P]) = \operatorname{Proj}(K[cP])$.

A lattice polytope $P$ is called *very ample* if for every vertex $z \in P$ the semigroup $C(z) \cap \mathbb{Z}^n$ is generated by $\{x - z \mid\ x \in P \cap \mathbb{Z}^n\}$.

It is clear from the discussion above that $\operatorname{Proj}(K[\overline{S}_P]) = \operatorname{Proj}(K[P])$ if and only if $P$ is very ample. In particular, normal polytopes are very ample, but not conversely:

***Example 5.5.1***. — Let $\Pi$ be the simplicial complex associated with the minimal triangulation of the real projective plane. It has 6 vertices which we label by the numbers $i \in [1, 6]$. Then the 10 facets of $\Pi$ have the following vertex sets (written as ascending sequences):

$$\begin{matrix} (1,2,3), & (1,2,4), & (1,3,5), & (1,4,6), & (1,5,6) \\ (2,3,6), & (2,4,5), & (2,5,6), & (3,4,5), & (3,4,6). \end{matrix}$$

Let $P$ be the polytope spanned by the indicator vectors of the ten facets (the indicator vector of $(1,2,3)$ is $(1,1,1,0,0,0)$ etc.). All the vertices lie in an affine hyperplane $H \subset \mathbb{R}^6$, and $P$ has indeed dimension 5. Using $H$ as the 'grading' hyperplane, one realizes $R = K[P]$ as the $K$-subalgebra of $K[X_1, \dots, X_6]$ generated by the 10 monomials $\mu_1 = X_1X_2X_3$, $\mu_2 = X_1X_2X_4$, ...

Let $\overline{R}$ be the normalization of $R$. It can be checked by effective methods that $\overline{R}$ is generated as a $K$-algebra by the 10 generators of $R$ and the monomial $\nu = X_1X_2X_3X_4X_5X_6$; in particular $R$ is not normal. Then one can easily compute by hand that the products $\mu_i\nu$ and $\nu^2$ all lie in $R$. It follows that $\overline{R}/R$ is a one-dimensional vector space; therefore $\operatorname{Proj}(R) = \operatorname{Proj}(\overline{R})$ is normal, and $P$ is very ample.

For a very ample polytope $P$ we have a projective embedding

$$\operatorname{Proj}(K[P]) \subset \mathbb{P}^N_K, \qquad N = \#(P \cap \mathbb{Z}^n) - 1.$$

The corresponding very ample line bundle on $\operatorname{Proj}(K[P])$ will be denoted by $L_P$. It is known that any projective toric variety and any very ample equivariant line bundle on it can be realized as $\operatorname{Proj}(K[P])$ and $L_P$ for some very ample polytope $P$. Moreover, any line bundle is isomorphic to an equivariant line bundle, and if $L_Q$ is a very ample equivariant line bundle on $\operatorname{Proj}(K[P])$ (for a very ample polytope $Q$) then $\mathcal{N}(P) = \mathcal{N}(Q)$ (see [**Oda**, Ch. 2] or [**Da**]). Therefore $P$ and $Q$ have the same column vectors (see observation $(***)$ in Subsection 5.2). Furthermore, $L_{Q_1}$ and $L_{Q_2}$ are isomorphic line bundles if and only if $Q_1$ and $Q_2$ differ only by a parallel translation (but they have different equivariant structures if $Q_1 \neq Q_2$).

Let $X$ be a projective toric variety and $L_P, L_Q \in \operatorname{Pic}(X)$ be two very ample equivariant line bundles. Then one has the elegant formula $L_P \otimes L_Q = L_{P+Q}$, where $P + Q$ is the Minkowski sum of $P, Q \subset \mathbb{R}^n$ (see Teissier [**Te**]). (Of course, very ampleness is preserved by the tensor product, and therefore by Minkowski sums.)

In the dual space $(\mathbb{R}^n)^*$ the column vectors $v$ correspond to the integral affine hyperplanes $H$ intersecting exactly one of the rays in $\mathcal{N}(P)$ (this is the condition $\varphi_G(v) \geqslant 0$ for $G \neq F$) and such that there is no lattice point strictly between $H$ and the parallel of $H$ through 0 (this is the condition $\varphi_F(v) = -1$). This shows that the column vectors correspond to Demazure's roots [**De**].

In Figure 8 the arrows represent the rays of the normal fans $\mathcal{N}(P_1)$ and $\mathcal{N}(P_2)$ and the lines indicate the hyperplanes corresponding to the column vectors ($P_1$ and $P_2$ are chosen as in Figure 7).

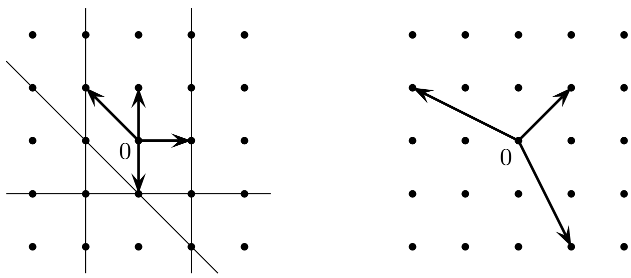


FIGURE 8. The normal fans of the polytopes $P_1$ and $P_2$

***Lemma 5.5.2***. — *If two lattice $n$-polytopes $P_1$ and $P_2$ have the same normal fans, then the quotient groups $\Gamma_K(P_1)^0/K^*$ and $\Gamma_K(P_2)^0/K^*$ are naturally isomorphic.*

*Proof.* — As in Subsection 5.4 we will work with $\overline{\Gamma}_K(P_i)$. Put

$$X = \operatorname{Proj}(K[\overline{S}_{P_1}]) = \operatorname{Proj}(K[\overline{S}_{P_2}])$$

and consider the canonical anti-homomorphisms $\overline{\Gamma}_K(P_i)^0 \to \operatorname{Aut}_K(X)$, $i = 1, 2$. Let $A(P_1)$ and $A(P_2)$ denote the images. We choose a column vector $v$ (for both polytopes) and $\lambda \in K$. We claim that the elementary automorphisms $e_v^\lambda(P_i) \in \overline{\Gamma}_K(P_i)^0$, $i = 1, 2$, have the same images in $\operatorname{Aut}_K(X)$. Denote the images by $e_1$ and $e_2$. For $i = 1, 2$ we can find a vertex $z_i$ of the base facet $(P_i)_v$ such that $C(z_1) = C(z_2)$. Now it is easy to see that $e_1$ and $e_2$ restrict to the same automorphism of the affine subvariety $\operatorname{Spec}(K[C(z_1) \cap \mathbb{Z}^n]) \subset X$, which is open in $X$. Therefore $e_1 = e_2$, as claimed.

It is also clear that for any $\tau \in \mathbb{T}_{n+1}$ the corresponding elements $\tau_i \in \overline{\Gamma}_K(P_i)$, $i = 1, 2$, have the same images in $\operatorname{Aut}_K(X)$. By Theorem 5.3.2(b) we arrive at the equality $A(P_1) = A(P_2)$. It only remains to notice that $K^* = \operatorname{Ker}(\overline{\Gamma}_K(P_i)^0 \to A(P_i))$, $i = 1, 2$. ☐

***Example 5.5.3***. — Lemma 5.5.2 cannot be improved. For example, let $P_1$ be the unit 1-simplex $\Delta_1$ and $P_2 = 2P_1$. Then $\mathbb{C}[S_{P_1}] = \mathbb{C}[X_1, X_2]$, and $\mathbb{C}[S_{P_2}] = \mathbb{C}[X_1^2, X_1X_2, X_2^2]$ is its second Veronese subring. Both polytopes have the same symmetries and column vectors, and moreover the torus action on $\mathbb{C}[S_{P_2}]$ is induced by that on $\mathbb{C}[S_{P_1}]$. Therefore the natural map $\Gamma_{\mathbb{C}}(P_1) \to \Gamma_{\mathbb{C}}(P_2)$ is surjective; in fact, $\Gamma_{\mathbb{C}}(P_1) = \mathrm{GL}_2(\mathbb{C})$ and $\Gamma_{\mathbb{C}}(P_2) = \mathrm{GL}_2(\mathbb{C})/\{\pm 1\}$. If there were an isomorphism between these groups, then $\mathrm{SL}_2(\mathbb{C})$ and $\mathrm{SL}_2(\mathbb{C})/\{\pm 1\}$ would also be isomorphic. This can be easily excluded by inspecting the list of finite subgroups of $\mathrm{SL}_2(\mathbb{C})$.

For a lattice polytope $P$ we denote the group opposite to $\Gamma_K(P)^0/K^*$ by $A_K(P)$, the projective toric variety $\operatorname{Proj}(K[\overline{S}_P])$ by $X(P)$; the symmetry group of a fan $\mathcal{F}$ is denoted by $\Sigma(\mathcal{F})$. ($\Sigma(\mathcal{F})$ is the subgroup of $\mathrm{GL}_n(\mathbb{Z})$ that leaves $\mathcal{F}$ invariant.) Furthermore we consider $A_K(P)$ as a subgroup of $\operatorname{Aut}_K(X(P))$ in a natural way.

***Theorem 5.5.4***. — *For a lattice $n$-polytope $P$ the group $\operatorname{Aut}_K(X(P))$ is generated by $A_K(P)$ and $\Sigma(\mathcal{N}(P))$. The connected component of unity of $\operatorname{Aut}_K(X(P))$ is $A_K(P)$. Furthermore, $\dim(A_K(P)) = \#\operatorname{Col}(P) + n$, and the embedded torus $\mathbb{T}_n = \mathbb{T}_{n+1}/K^*$ is a maximal torus of $\operatorname{Aut}_K(X(P))$.*

*Proof.* — Assume for the moment that $P$ is very ample and $[L_P] \in \operatorname{Pic}(X(P))$ is preserved by every element of $\operatorname{Aut}_K(X(P))$. Then we are able to apply the classical arguments for projective spaces as follows.

We have $K[\overline{S}_P] = \bigoplus_{i \geqslant 0} H^0(X, L_P^i)$. Since $[L_P]$ is invariant under $\operatorname{Aut}_K(X)$, arguments similar to those in Hartshorne [**Ha**, Example 7.1.1, p. 151] show that giving an automorphism of $X$ is equivalent to giving an element of $\Gamma_K(P)$. In other words, the natural anti-homomorphism $\Gamma_K(P) \to \operatorname{Aut}_K(X(P))$ is surjective. Now Theorem 5.3.2 gives the desired result once we notice that $\Sigma(P)$ is mapped to $\Sigma(\mathcal{F})$.

Therefore, and in view of Lemma 5.5.2, the proof is completed once we show that there is a very ample polytope $Q$ having the same normal fan as $P$ and such that $[L_Q]$ is invariant under $\operatorname{Aut}_K(X)$.

The existence of such a 'fully' symmetric polytope is established as follows. First we replace $P$ by the normal polytope $cP$ for some $c \gg 0$ so that we may assume that $P$ is normal. The $K$-vector space of global sections of a line bundle, which is an image of $L_P$ with respect to some element of $\operatorname{Aut}_K(X(P))$, has the same dimension as the space of global sections of $L_P$, which is given by $\#(P \cap \mathbb{Z}^n)$. Easy inductive arguments ensure that the number of polytopes $Q$ such that $\#(Q \cap \mathbb{Z}^n) = \#(P \cap \mathbb{Z}^n)$ and, in addition, $\mathcal{N}(Q) = \mathcal{N}(P)$ is *finite*. It follows that the set $\{[L_{Q_1}], \dots, [L_{Q_t}]\}$ of isomorphism classes of very ample equivariant line bundles to which $[L_P]$ is mapped by an automorphism of $X(P)$ is finite. Since every line bundle is isomorphic to an equivariant one, any element $\alpha \in \operatorname{Aut}_K(X(P))$ must permute the classes $[L_{Q_i}] \in \operatorname{Pic}(X(P))$. In particular, the element

$$[L_{Q_1} \otimes \cdots \otimes L_{Q_t}] \in \operatorname{Pic}(X)$$

is invariant under $\operatorname{Aut}_K(X(P))$. But $L_{Q_1} \otimes \cdots \otimes L_{Q_t} = L_{Q_1 + \cdots + Q_t}$ and, hence, $Q_1 + \cdots + Q_t$ is the desired polytope. $\square$

***Example 5.5.5***. — In general the natural anti-homomorphism $\Gamma_K(P) \to \operatorname{Aut}_K(X(P))$ is not surjective. For example consider the polytopes $P$ and $Q$ in Figure 9. Then

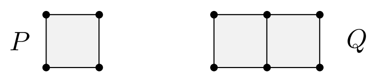


FIGURE 9

$\operatorname{Proj}(K[P]) = \operatorname{Proj}(K[Q]) = \mathbb{P}^1 \times \mathbb{P}^1$. However, the isomorphism corresponding to the exchange of the two factors $\mathbb{P}^1$ cannot be realized in $K[Q]$.

Above we have derived the automorphism group of a projective toric variety from that of the homogeneous coordinate ring of a suitable embedding. This approach has been generalized to arrangements of toric varieties in [**BG6**].

**5.6. Retracts of dimension two.** — A *retract* of a $K$-algebra $A$ is an algebra $B$ such that there exist $K$-homomorphisms $f : B \to A$ and $g : A \to B$ with $g \circ f = 1_B$. This is equivalent to saying that there is an endomorphism $h : A \to A$ such that $h^2 = h$ and $\operatorname{Im}(h) \cong B$. We will call such $g$ and $h$ *retractions* and will frequently make passages between the two equivalent definitions. Moreover, all the retractions considered below are supposed to be graded. For a retraction $h$ as above we put

$$\operatorname{codim}(h) = \dim(A) - \dim(B).$$

The arguments used in the sequel need $K$ to be algebraically closed.

That polytopality of algebras is in general not an invariant property under scalar extension/restriction is exhibited by the following

***Example 5.6.1.*** — Consider the standard graded $\mathbb{R}$-algebra

$$A = \mathbb{R}[X, Y, Z]/(X^2 + Y^2 + Z^2).$$

Then $A$ is a factorial non-polytopal algebra over $\mathbb{R}$ while $\mathbb{C} \otimes A$ is isomorphic to the polytopal algebra $\mathbb{C}[2\Delta_1]$ defined by a lattice segment $2\Delta_1$ of length 2.

The factoriality of $A$ is proved in [Fo,§11]. But the only factorial polytopal algebras (over any field) are polynomial algebras – an easy observation. Hence $A$ is not polytopal because it is singular at the irrelevant maximal ideal. But we have the isomorphism

$$\alpha : \mathbb{C}[U^2, UV, V^2] = \mathbb{C}[2\Delta_1] \to \mathbb{C} \otimes A$$

defined by $U^2 \mapsto X + iY$, $V^2 \mapsto X - iY$, $UV \mapsto iZ$.

Conjecture A holds in Krull dimension $\leqslant 2$:

***Theorem 5.6.2.*** — *A retract $B$ of a polytopal algebra $A$ is polytopal if* $\dim B \leqslant 2$.

The crucial step in the proof is

***Proposition 5.6.3.*** — *Let $K$ be an algebraically closed field and $A$ a standard graded $K$-algebra of dimension 2. If $A$ is a normal domain and the class group* $\mathrm{Cl}(A)$ *is finitely generated, then $A$ is isomorphic to $K[c\Delta_1]$ as a graded $K$-algebra for some $c \in \mathbb{N}$ (as usual, $\Delta_1$ is the unit segment).*

*Proof.* — We have the projectively normal embedding of $\mathrm{Proj}(A)$ given by $A$. Therefore, the projective curve $\mathrm{Proj}(A)$ is normal and thus smooth. Consider the exact sequence

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathrm{Cl}(\mathrm{Proj}(A)) \longrightarrow \mathrm{Cl}(A) \longrightarrow 0.$$

of Weil divisors arising from viewing $\mathrm{Spec}(A)$ as a cone over $\mathrm{Proj}(A)$ ([**Ha**, Ex. II.6.3(b)]). Since $\mathrm{Cl}(A)$ is finitely generated, so is $\mathrm{Cl}(\mathrm{Proj}(A))$. In particular the Jacobian

$$\mathcal{J}(\mathrm{Proj}(A)) \cong \mathrm{Cl}^0(\mathrm{Proj}(A))$$

is trivial ($\mathrm{Cl}^0$ denotes degree zero divisor classes). Therefore the genus of $\mathrm{Proj}(A)$ is 0, or equivalently $\mathrm{Proj}(A) \cong \mathbb{P}^1_K$. Using the normality of $A$ once again we get

$$A \cong \bigoplus_{i=0}^{\infty} H^0(\mathbb{P}^1_K, \mathcal{L}^{\otimes i})$$

for some very ample line bundle $\mathcal{L}$ on $\mathbb{P}^1_K$. But due to the equality $\mathrm{Pic}(\mathbb{P}^1_K) = \mathbb{Z}$ such a line bundle is a positive multiple of $\mathcal{O}(1)$, and hence $A$ is the Veronese subalgebra of the polynomial algebra $K[\Delta_1]$ of some level $c \in \mathbb{N}$. $\square$

*Proof of Theorem 5.6.2.* — In case $\dim(B) = 1$ it is easy to see that $B \cong K[X]$.

Consider the case $\dim(B) = 2$. We write $A = K[P]$ and denote the retraction $A \to B \subset A$ by $g$. Consider the set $(S_P \cap \operatorname{Ker}(g)) \subset K[P]$ of monomials. There is a unique face $F \subset P$ such that $(S_P \cap \operatorname{Ker}(g)) = (S_P \setminus S_F)$ and $K[P]/(S_P \cap \operatorname{Ker}(g))$ is naturally isomorphic to $K[F]$. Then $g$ is a composite of the two retractions

$$K[P] \xrightarrow{\pi} K[F] \xrightarrow{\rho} B$$

where $\rho$ is the homomorphism induced by $g$. Observe that $\rho$ is in fact a retraction as it is split by $\pi|_B$.

Therefore we can from the beginning assume that $(S_P \cap \operatorname{Ker}(g)) = 0$. In this situation $g$ extends (uniquely) to the normalizations

$$\overline{K[P]} = K[\overline{S}_P] \xrightarrow{\overline{g}} \overline{B}. \tag{1}$$

This extension is given by

$$\overline{g}(z) = \frac{g(x)}{g(y)}, \qquad z \in \overline{S}_P,\ x, y \in S_P \text{ and } z = \frac{x}{y}.$$

It is known that the semigroup $S_{nP}$ is normal for all natural numbers $n \geqslant \dim(P) - 1$ (see Subsection 3.4). Therefore, by restricting the retraction (1) to the $n$th Veronese subalgebra for such a number $n$, we get the retraction

$$K[nP] \xrightarrow{\overline{g}_n} \overline{B}_{(n)}. \tag{2}$$

Let us show that $\operatorname{Cl}(\overline{B}_{(n)})$ is finitely generated for $n \geqslant \dim(P) - 1$. We choose a lattice point $x$ of $S_{nP}$ that is in the interior of the cone $C(S_P)$. By localization (2) gives rise to the retraction

$$(x\overline{g}_n(x))^{-1} K[nP] \to (\overline{g}_n(x))^{-1} \overline{B}_{(n)}. \tag{3}$$

Since $(x\overline{g}_n(x))^{-1} K[nP]$ is a localization of the Laurent polynomial ring $x^{-1}K[nP] = K[\operatorname{gp}(S_{nP})]$, it is a factorial ring. Then its retract $(\overline{g}_n(x))^{-1}\overline{B}_{(n)}$ is factorial as well (for example, see Costa [**Cos**]). By Nagata's theorem [**Fo**, 7.1] $\operatorname{Cl}(\overline{B}_{(n)})$ is generated by the classes of the height 1 prime ideals of $\overline{B}_{(n)}$ containing $\overline{g}_n(x)$ – a finite set.

It is also clear from (2) that $\overline{B}_{(n)}$ is generated in degree 1. Consequently, by Proposition 5.6.3 for each $n \geqslant \dim(P) - 1$ there is a natural number $c_n$ and an isomorphism

$$\varphi_n : \overline{B}_{(n)} \to K[c_n \Delta_1].$$

We now fix such a number $n$. Restricting $\varphi_n$ and $\varphi_{n+1}$ to the iterated Veronese subalgebra $\overline{B}_{(n(n+1))} = (\overline{B}_{(n)})_{(n+1)} = (\overline{B}_{(n+1)})_{(n)}$ we obtain two isomorphisms of $\overline{B}_{(n(n+1))}$ with $K[c_{n(n+1)}\Delta_1]$. It follows that there exists $c \in \mathbb{N}$ with $c_n = cn$ and $c_{n+1} = c(n+1)$, and furthermore the restrictions of $\varphi_n$ and $\varphi_{n+1}$ differ by an automorphism of $K[c_{n(n+1)}\Delta_1]$. However, each automorphism of $K[c_{n(n+1)}\Delta_1]$ can be lifted to an automorphism of $K[\Delta_1]$, and then restricted to all Veronese subrings of

$K[\Delta_1]$. (This follows from Theorem 5.3.2.) Therefore we can assume that the restrictions of $\varphi_n$ and $\varphi_{n+1}$ coincide. Then they define an isomorphism of the subalgebra $V$ of $\overline{B}$ generated by the elements in degree $n$ and $n+1$ to the corresponding subalgebra of $K[\Delta_1]$; see Lemma 5.6.4 below. Taking normalizations yields an isomorphism $\overline{B} \cong K[c\Delta_1]$. But then $B = K[c\Delta_1]$ as well, because $\overline{B}$ and $B$ coincide in degree 1 (being retracts of algebras with this property). $\square$

***Lemma 5.6.4***. — *Let $A$ and $B$ be $\mathbb{Z}$-graded rings. Suppose that $B$ is reduced. If the homogeneous homomorphisms $\varphi : A_{(n)} \to B_{(n)}$ and $\psi : A_{(n+1)} \to B_{(n+1)}$ coincide on $A_{(n(n+1))}$, then they have a common extension to a homogeneous homomorphism $\chi : V \to B$, where $V$ is the subalgebra of $A$ generated by $A_{(n)}$ and $A_{(n+1)}$. If, in addition, $A$ is reduced and $\varphi$ and $\psi$ are injective, then $\chi$ is also injective.*

*Proof*. — One checks easily that one only needs to verify the following: if $uv = u'v'$ for homogeneous elements $u, u' \in A_{(n)}$, $v, v' \in A_{(n+1)}$, then $\varphi(u)\psi(v) = \varphi(u')\psi(v')$. As $B$ is reduced, it is enough that $(\varphi(u)\psi(v) - \varphi(u')\psi(v'))^{n(n+1)} = 0$. Since

$$u^p(u')^{n(n+1)-p},\ v^p(v')^{n(n+1)-p} \in A_{(n(n+1))}, \qquad p \in [0, n(n+1)],$$

this follows immediately from the hypothesis that $\varphi$ and $\psi$ coincide on $A_{(n(n+1))}$.

If $A$ is reduced, then every non-zero homogeneous ideal in $A$ intersects $A_{(n(n+1))}$ non-trivially, and this implies the second assertion. $\square$

**5.7. The structure of retractions.** — Now we first consider Conjecture B in detail and then observe that it does not admit a direct extension to codimension $\geqslant 2$.

Let $P \subset \mathbb{R}^n$ be a lattice polytope of dimension $n$ and $F \subset P$ a face. Then there is a uniquely determined retraction

$$\pi_F : K[P] \to K[F], \quad \pi_F(x) = 0 \quad \text{for } x \in L_P \smallsetminus F.$$

Retractions of this type will be called *face retractions* and *facet retractions* if $F$ is a facet or, equivalently, $\operatorname{codim}(\pi_F) = 1$.

Now suppose there are an affine subspace $H \subset \mathbb{R}^n$ and a vector subspace $W \subset \mathbb{R}^n$ with $\dim W + \dim H = n$, such that

$$L_P \subset \bigcup_{x \in L_P \cap H} (x + W).$$

(Observe that $\dim(H \cap P) = \dim H$.) The triple $(P, H, W)$ is called a *lattice fibration of codimension* $c = \dim W$, whose *base polytope* is $P \cap H$; its *fibers* are the maximal lattice subpolytopes of $(x + W) \cap P$, $x \in L_P \cap H$ (the fibers may have smaller dimension than $W$). $P$ itself serves as a *total polytope* of the fibration. If $W = \mathbb{R}w$ is a line, then we call the fibration *segmental* and write $(P, H, w)$ for it. Note that the column structures introduced in Section 5.2 give rise to lattice segmental fibrations in a natural way.

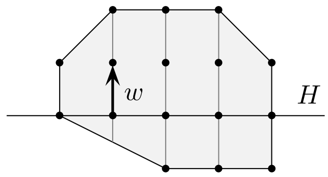


FIGURE 10. A lattice segmental fibration

For a lattice fibration $(P, H, W)$ let $L \subset \mathbb{Z}^n$ denote the subgroup spanned by $L_P$, and let $H_0$ be the translate of $H$ through the origin. Then one has the direct sum decomposition

$$L = (L \cap W) \oplus (L \cap H_0).$$

Equivalently,

$$\mathrm{gp}(S_P) = L \oplus \mathbb{Z} = \big(\mathrm{gp}(S_P) \cap W_1\big) \oplus \mathrm{gp}(S_{P \cap H_1}),$$

where $W_1$ is the image of $W$ under the embedding $\mathbb{R}^n \to \mathbb{R}^{n+1}$, $w \mapsto (w, 0)$, and $H_1$ is the vector subspace of $\mathbb{R}^{n+1}$ generated by all the vectors $(h, 1)$, $h \in H$.

For a fibration $(P, H, W)$ one has the naturally associated retraction:

$$\rho_{(P,H,W)} : K[P] \to K[P \cap H];$$

it maps $L_P$ to $L_{P \cap H}$ so that fibers are contracted to their intersection points with the base polytope $P \cap H$.

Clearly, if $f : K[P] \to K[P]$ is a retraction, then for any graded automorphism $\alpha$ of $K[P]$ the composite map $f^\alpha = \alpha \circ f \circ \alpha^{-1}$ is again a retraction and $\mathrm{Im}(f^\alpha) = \alpha(\mathrm{Im}(f))$ and $\mathrm{Ker}(f^\alpha) = \alpha(\mathrm{Ker}(f))$. Now the exact formulation of Conjecture B is as follows.

***Conjecture B***. — *For a codimension 1 retraction $f : K[P] \to K[P]$ there is $\alpha \in \Gamma_K(P)$ such that $f^\alpha = \iota \circ g$ for a retraction $g$ of type either $\pi_F$ or $\rho_{(P,H,w)}$ and $\iota : \mathrm{Im}(g) \to K[P]$ a graded $K$-algebra embedding.*

In other words this conjecture claims that any codimension 1 retraction can be 'modified' by an automorphism so that the corrected retraction factors through a retraction preserving the monomial structure.

A necessary condition for Conjecture B is that any codimension 1 retraction $f$ can be modified by a graded automorphism $\alpha$ so that $f^\alpha$ has either a homogeneous binomial of degree 1 or a monomial of degree 1 in its kernel. A weaker condition is that $\mathrm{Ker}(f^\alpha)$ contains a homogeneous binomial of degree $\geqslant 1$ (evidently this holds if there is a monomial in $\mathrm{Ker}(f^\alpha)$).

We remark that even an example of just an endomorphism in $\mathrm{Pol}(K)$, such that $\mathrm{Ker}(f^\alpha)$ contains no (homogeneous) binomial for any $\alpha$, is not readily found. However, such exists, even in the class of codimension 2 retractions.

The examples below are constructed from joins of polytopes. A polytope $P \subset \mathbb{R}^{n_P}$ is called a *join* of two polytopes $Q \subset \mathbb{R}^{n_Q}$ and $R \subset \mathbb{R}^{n_R}$ if there are affine embeddings $\varphi_Q : \mathbb{R}^{n_Q} \to \mathbb{R}^{n_P}$ and $\varphi_R : \mathbb{R}^{n_R} \to \mathbb{R}^{n_P}$ such that:

(1) $\operatorname{Im}(\varphi_Q) \cap \operatorname{Im}(\varphi_R) = \varnothing$,

(2) the affine hull of $\operatorname{Im}(\varphi_Q) \cup \operatorname{Im}(\varphi_R)$ is an $(n_Q+n_R+1)$-dimensional affine subspace of $\mathbb{R}^{n_P}$,

(3) $P$ is the convex hull of $\varphi_Q(Q) \cup \varphi_R(R)$.

It is easy to see that joins are uniquely determined up to isomorphism.

The following lemma enables us to describe $\Gamma_K(R)$ for $R = \operatorname{join}(P, Q)$ under a mild assumption on $P$ and $Q$. We identify them with the corresponding faces of $R$.

***Lemma 5.7.1.*** — *Let $P$ and $Q$ be lattice polytopes, both having interior lattice points. Then* $\operatorname{Col}(\operatorname{join}(P, Q)) = \operatorname{Col}(P) \cup \operatorname{Col}(Q)$.

*Proof.* — That each of the column vectors of the polytopes serves as a column vector for $\operatorname{join}(P, Q)$ is clear.

Now let $v \in \operatorname{Col}(\operatorname{join}(P, Q))$. If $v$ is parallel to either $P$ or $Q$ then either $v \in \operatorname{Col}(P)$ or $v \in \operatorname{Col}(Q)$ since $L_{\operatorname{join}(P,Q)} = L_P \cup L_Q$. So without loss of generality we can assume that $v$ is parallel neither to $P$ nor to $Q$. Since $P$ and $Q$ span $\operatorname{join}(P, Q)$ they cannot be contained simultaneously in the base facet of $v$. But then either $p + v \in \operatorname{join}(P, Q)$ or $q + v \in \operatorname{join}(P, Q)$ for suitable vertices $p \in P$ and $q \in Q$. We get a contradiction because one of the points $r + v$ or $s + v$ is outside $\operatorname{join}(P, Q)$ for interior lattice points $r \in P$, $s \in Q$. □

***Example 5.7.2.*** — Let $Q$ be the lattice triangle spanned by $(0, -1)$, $(-1, 0)$, and $(1, 1)$. Then $Q$ contains only one more lattice point, namely $(0, 0)$. Identifying $U$ with $(0, 0)$, $V$ with $(0, -1)$, and $W$ with $(-1, 0)$ we see that the polynomial ring $K[U, V, W]$ can be embedded into $K[Q]$ such that the indeterminates correspond to lattice points. Moreover, $K[\operatorname{gp}(S_Q)]$ is then just the Laurent polynomial ring $K[\mathbb{Z}^3] = K[U^{\pm1}, V^{\pm1}, W^{\pm1}]$.

Let $h' : K[X, Y] \to K[U, V, W]$ be defined by $h'(X) = U + V$, $h'(Y) = U + W$. Then $h'$ induces a retraction $h$ of $K[U, V, W, X, Y]$, namely the retraction mapping $X$ and $Y$ to $h'(X)$ and $h'(Y)$ respectively and leaving $U, V, W$ invariant. This retraction extends in a natural way to retraction of $K[U^{\pm1}, V^{\pm1}, W^{\pm1}, X, Y]$, and can then be restricted to

$$K[Q] \otimes K[\Delta_1] \subset K[U^{\pm1}, V^{\pm1}, W^{\pm1}, X, Y]$$

where we identify $K[X, Y]$ with the polytopal algebra $K[\Delta_1]$ of the unit segment. It can further be restricted to $K[\operatorname{join}(2Q, 2\Delta_1)]$ which is embedded into $K[Q] \otimes K[\Delta_1]$ as the tensor product of the second Veronese subalgebras of the normal algebras $K[Q]$ and $K[\Delta_1]$.

We claim that the just constructed retraction $h$ of $K[P]$, $P = \operatorname{join}(2Q, 2\Delta_1)$, $\dim P = 4$, has no conjugate $h^\alpha$ by an automorphism $\alpha \in \Gamma_K(P)$ such that the kernel of $h^\alpha$ contains a binomial.

The polytope $Q$ has no column structures, a property inherited by $2Q$ (see observation $(*)$ in Subsection 5.2). Moreover, both $2Q$ and $2\Delta_1$ have interior points. Therefore the only column structures on $P$ are those it gets from $2\Delta_1$ (see Lemma 5.7.1). Then every element $\alpha \in \Gamma_K(P)$ is of the form $\tau\circ\beta$, where $\tau$ is a toric automorphism and $\beta = 1\otimes\beta'$ for some $\beta' \in \Gamma_K(2\Delta_1)$. Since $\tau$ does not affect the monomial structure, we can assume $\tau = 1$. Furthermore the graded automorphisms of $K[2\Delta_1]$ are all restrictions of automorphisms of $K[\Delta_1] = K[X, Y]$ so that we have to take into account all automorphisms of $K[P]$ induced by a substitution

$$X \longmapsto a_{11}X + a_{12}Y, \quad Y \longmapsto a_{21}X + a_{22}Y, \quad U \longmapsto U, \quad V \longmapsto V, \quad W \longmapsto W$$

with $\det(a_{ij}) \neq 0$. Then $h^\alpha$ is induced by the substitution

$$a_{11}X + a_{12}Y \longmapsto U + V, \qquad a_{21}X + a_{22}Y \longmapsto U + W,$$

leaving $U, V, W$ invariant. Also $h^\alpha$ extends to a retraction of $K[U^{\pm 1}, V^{\pm 1}, W^{\pm 1}, X, Y]$ and then restricts to $K[U, V, W, X, Y]$. This shows that the kernel of the extension cannot contain a monomial; otherwise it would contain a monomial in $X$ and $Y$, but $h^\alpha$ is injective on $K[X, Y]$. If the kernel contains a binomial $b$, we can assume that $b \in K[U, V, W, X, Y]$. In other words, there is a binomial in the ideal $\mathfrak{p}$ of $K[U, V, W, X, Y]$ generated by

$$a_{11}X + a_{12}Y - (U + V), \qquad a_{21}X + a_{22}Y - (U + W).$$

Since the prime ideal $\mathfrak{p}$ contains no monomials, we can assume that the two terms of $b$ are coprime. But then $b$ reduces to a monomial modulo one of the variables, and since $\mathfrak{p}$ reduces to an ideal generated by linear forms, it reduces to a prime ideal. The reduction of $\mathfrak{p}$ modulo any of the variables cannot contain another variable.

**5.8. Segmentonomial ideals.** — We will use the following theorem of Eisenbud and Sturmfels [**ES**, 2.6] characterizing binomial prime ideals in affine semigroup algebras over algebraically closed fields. (In [**ES**] the theorem is given only for polynomial rings, but the generalization is immediate.)

***Theorem 5.8.1***. — *Let $K$ be an algebraically closed field. A binomial ideal $I$ in an affine semigroup algebra $K[S]$ is prime if and only if the residue class ring $K[S]/I$ contains a (multiplicative) affine semigroup $S'$ such that $K[S]/I = K[S']$ and, moreover, the natural epimorphism $K[S] \to K[S']$ maps the monomials in $K[S]$ to those in $K[S']$.*

For an affine semigroup $S$ an element $f \in K[S]$ will be called *segmentonomial* if the Newton polytope $\mathrm{N}(f) \subset \mathbb{R}\otimes\mathrm{gp}(S)$ has dimension $\leqslant 1$. (Clearly, monomials as well as binomials are segmentonomials.) An ideal $I \subset K[S]$ is called segmentonomial if it is generated by a system of segmentonomials.

It is proved in [**ES**] that every minimal prime ideal over a binomial ideal of $K[X_1,\dots,X_n]$ (a polynomial ring) is again binomial. In this section we derive the same result for segmentonomial ideals in arbitrary affine semigroup algebras.

***Theorem 5.8.2***. — *Let $S$ be an affine semigroup and $I \subset K[S]$ be a segmentonomial ideal.*

(a) *A minimal prime overideal $I \subset \mathfrak{p} \subset K[S]$ is binomial, and $K[S]/I$ is again an affine semigroup algebra.*
(b) *Suppose that $\operatorname{ht}(I) = 1$, $f \in I$, $\dim(\mathrm{N}(f)) = 1$, and $\mathfrak{p}$ is as above. Then for every system of pairwise distinct lattice points $x_1,\dots,x_m \in L_P$, such that none of the pairs $(x_i, x_j)$, $i \neq j$, spans a line in $\mathbb{R}\otimes\operatorname{gp}(S)$ parallel to $\mathrm{N}(f)$, the residue classes $\overline{x}_1,\dots,\overline{x}_m$ constitute a $K$-linearly independent subset of $K[S]/\mathfrak{p}$.*

*Proof*. — We prove claim (a) by induction on $r = \operatorname{rank}(S)$. Claim (b) will follow automatically from the description of $\mathfrak{p}$ derived below.

For $r = 0$ there is nothing to show. Assume the theorem is proved for semigroups of rank $< r$ and choose a segmentonomial $f \in I$. Then $\mathfrak{p}$ contains a minimal prime $\mathfrak{p}_0$ over the principal ideal $(f)$. Assume that $\mathfrak{p}_0$ is a binomial ideal. By Theorem 5.8.1, $K[S]/\mathfrak{p}_0 \cong K[S_1]$ for some affine semigroup $S_1$ and such that monomials in $K[S]$ go to monomials in $K[S_1]$. But then segmentonomials in $K[S]$ are likewise mapped to segmentonomials in $K[S_1]$. This holds true because affinely independent monomials lift to affinely independent monomials. By induction hypothesis the image of $\mathfrak{p}$ in $K[S_1]$ is binomial. Since binomials can be lifted to binomials in $K[S]$, we conclude that $\mathfrak{p}$ is binomial.

The general situation thus reduces to the case in which $I = (f)$ for some segmentonomial $f \in K[S]$ and $\operatorname{ht}(I) = \operatorname{ht}(\mathfrak{p}) = 1$.

If $\mathfrak{p}$ contains a monomial, then $\mathfrak{p}$ is a height 1 monomial prime ideal, and we are done.

Otherwise $S \cap \mathfrak{p} = \varnothing$. Consider the localization $\mathfrak{p}\, K[\operatorname{gp}(S)]$. It is a height 1 prime ideal in the Laurent polynomial ring $K[\operatorname{gp}(S)]$. Therefore, $\mathfrak{p}\, K[\operatorname{gp}(S)] = f_0 K[\operatorname{gp}(S)]$ for some prime element $f_0 \in K[\operatorname{gp}(S)]$.

Also $f_0$ is segmentonomial. In fact, we have $f = f_0 f_1$ for some $f_1 \in K[\operatorname{gp}(S)]$ implying the equality $\mathrm{N}(f) = \mathrm{N}(f_0) + \mathrm{N}(f_1)$ for the corresponding Newton polytopes. Since $\dim N(f_0) = 0$ is excluded, $\dim(\mathrm{N}(f_0)) = 1$. Multiplying $f_0$ by a suitable term from $\operatorname{gp}(S)$ we can achieve that the origin $0 \in \mathbb{R}\otimes\operatorname{gp}(S)$ is one of the end-points of $\mathrm{N}(f_0)$.

Let $\ell \subset \mathbb{R}\otimes\operatorname{gp}(S)$ denote a rational line containing $\mathrm{N}(f_0)$. In a suitable basis of the free abelian group $\operatorname{gp}(S)$ the line $\ell$ becomes a coordinate direction. Therefore, we can assume that

$$K[\operatorname{gp}(S)] = K[X_1, X_1^{-1},\dots,X_n,X_n^{-1}]$$

and that $f_0$ is a monic polynomial in $X_1$. Since $K$ is algebraically closed, if follows that $f_0 = X_1 - a$ for some $a \in K$. Since $\mathfrak{p}$ does not contain a monomial, one has

$$\mathfrak{p} = \mathfrak{p}\, K[\mathrm{gp}(S)] \cap K[S] = (X_1 - a)K[X_1, X_1^{-1}, \dots, X_n, X_n^{-1}] \cap K[S].$$

Thus $\mathfrak{p}$ is the kernel of the composite homomorphism

$$K[S] \hookrightarrow K[X_1, X_1^{-1}, \dots, X_n, X_n^{-1}] \xrightarrow{X_1 \mapsto a} K[X_2, X_2^{-1}, \dots, X_n, X_n^{-1}]$$

This is a homomorphism mapping the elements of $S$ to Laurent monomials in $X_2, X_2^{-1}$, $\dots, X_n, X_n^{-1}$, and therefore $\mathfrak{p}$ is generated by binomials. □

**5.9. Based retractions.** — Throughout this section we suppose that $h : K[P] \to K[P]$ is a retraction and that $A = \mathrm{Im}(h)$. We also assume $P \subset \mathbb{R}^n$, $\dim(P) = n$, $\mathrm{gp}(S_P) = \mathbb{Z}^{n+1}$ (and that $K$ is algebraically closed.)

***Lemma 5.9.1***. — *The following conditions are equivalent:*

(a) *there is a subset $X \subset L_P$ such that the restriction $h : K[S_X] \to A$ is an isomorphism, where $K[S_X] \subset K[P]$ is the subalgebra generated by the semigroup $S_X = \langle X\rangle \subset S_P$,*

(b) *there is a $(\dim(A) - 1)$-dimensional cross section $Q$ of $P$ by a linear subspace $H$ such that $Q$ is a lattice polytope (i. e. the vertices of $Q$ are lattice points) and*

$$h|_{K[Q]} : K[Q] \to A$$

*is an isomorphism. In particular, $A$ is a polytopal algebra.*

*Proof.* — We only need to derive (b) from (a). Let $H$ be the affine hull of $X$ in $\mathbb{R}^n$. We have to show that $Q = H \cap P$ is a lattice polytope with $L_Q = X$. Consider the subsemigroup

$$S'_Q = \{x \in S \mid x \neq 0 \text{ and } \mathbb{R}_+ x \cap P \subset H\} \cup \{0\}.$$

Then $h(K[S'_Q]) = A$ as well. On the other hand

$$\dim K[S'_Q] = \dim H + 1 = \dim K[S_X] = \dim A.$$

Thus the restriction $h : K[S'_Q] \to A$ is also an isomorphism. It follows that $X = L_Q$, and every element in $S'_Q$ is a product of elements of $X$. Furthermore $Q = \mathrm{conv}(X)$ since any rational point of the complement $Q \setminus \mathrm{conv}(X)$ gives rise to elements in $S'_Q \setminus S_X$. □

A subpolytope $Q \subset P$ as in Lemma 5.9.1(b) (if it exists) will be called a *base* of $h$ and $h$ is a *based retraction.* Notice that a base is not necessarily uniquely determined.

***Theorem 5.9.2***. — *Suppose a retraction $h : K[P] \to K[P]$ has a base $Q$ that intersects the interior of $P$. Then $h^\tau = \iota \circ \rho_{(P,H,W)}$ for some toric automorphism $\tau \in \mathbb{T}_K(P)$, a lattice fibration $(P, H, W)$ and a $K$-algebra embedding $\iota : K[H \cap P] \to K[P]$.*

*Proof.* — It is not hard to check that there is no restriction in assuming that $K[Q] = \mathrm{Im}(h)$.

Note that $\mathrm{Ker}(h) \cap S_P = \varnothing$. In fact, if a monomial is mapped to 0 by $h$, then $\mathrm{Ker}(h)$ contains a monomial prime ideal $\mathfrak{p}$ of height 1. Since $\mathfrak{p}$ in turn contains all monomials in the interior of $S_P$, it must also contain monomials from $S_Q$, which is impossible. Thus $h$ can be extended to the normalization $K[\overline{S}_P]$; on $K[\overline{S}_Q] \subset K[\overline{S}_P]$ the extension is the identity.

Set $L = \mathbb{Z}^{n+1}$, and let $U$ be the intersection of the $\mathbb{Q}$-vector subspace of $\mathbb{Q}^{n+1}$ generated by $S_Q$ with $L$. Choose a basis $v_1, ..., v_m$ of a complement of $U$ in $L$. Since $S_Q$ contains elements of degree 1 (given by the last coordinate), we can assume that $\deg v_i = 0$ for $i \in [1, m]$. In sufficiently high degree we can find a lattice point $x$ in $\overline{S}_Q$ such that $xv_i$, $xv_i^{-1} \in \overline{S}_P$. We have the relation $(xv_i)(xv_i^{-1}) = x^2$.

It follows that $h(xv_i) = a_i x_i$, equivalently $h(x(a_i^{-1} v_i)) = x_i$, for some $x_i \in \overline{S}_Q$ and $a_i \in K^*$. After a toric 'correction' leaving $K[\overline{S}_Q]$ fixed we can assume $a_i = 1$ for all $i$.

After the inversion of the elements of $S_P$, we can further extend the homomorphism $h$ to a map defined on the Laurent polynomial ring $K[L]$. Then we have

$$h(v_i x x_i^{-1}) = 1.$$

The vectors $v_i + x - x_i$ are also a basis of a complement of $U$, and thus part of a basis of $L$. Therefore the elements

$$v_i x x_i^{-1} - 1, \qquad i = [1, m],$$

generate a prime ideal of height $m$ in $K[L]$.

It is now clear that $h$ (after the toric correction) is just the retraction $\rho_{(P,H,W)}$ where $H$ is the affine hull of $Q$ in $\mathbb{R}^n$ and $W$ is the sublattice of $\mathbb{Z}^n$ generated by the vectors $v_i + x - x_i$ upon the identification of $\mathbb{Z}^n$ with the degree 0 sublattice of $\mathbb{Z}^{n+1}$. □

Example 5.7.2 shows that even a based retraction $h$ of $K[P]$ need not satisfy Conjecture B if the base does not intersect the interior of $P$ and $h$ has codimension $\geqslant 2$. However, in codimension 1 Conjecture B holds for all based retractions, as follows from Theorem 5.9.2 and

**Theorem 5.9.3**. — *Suppose the codimension* 1 *retraction* $h : K[P] \to K[P]$ *has a base* $F$ *not intersecting the interior of* $P$. *Then* $F$ *is a facet of* $P$ *and* $h^\varepsilon = \iota \circ \pi_F$ *for some* $\varepsilon \in \mathbb{A}(F)$ *and a* $K$*-algebra embedding* $\iota : K[F] \to K[P]$.

See Theorem 5.3.2 for the definition of $\mathbb{A}(F)$. In the proof we will use a general fact on pyramids. Recall that a *pyramid* $\Pi \subset \mathbb{R}^n$ is a polytope which is spanned by a point $v$ and a polytope $B$ such that the affine hull of $B$ does not contain $v$. In this situation $v$ is called an *apex* and $B$ is called a *base* of $\Pi$.

***Lemma 5.9.4***. — *Let* $\Pi \subset \mathbb{R}^n$ *be a pyramid and* $\Pi = \Pi_1 + \Pi_2$ *be a Minkowski sum representation by polytopes* $\Pi_1, \Pi_2 \subset \mathbb{R}^n$. *Then both* $\Pi_1$ *and* $\Pi_2$ *are homothetic images of* $\Pi$ *(with respect to appropriate centers and non-negative factors).*

*Proof*. — The case $\dim(\Pi) = 2$ is an easy exercise.

Now we use induction on $\dim(\Pi)$. Assume $\dim(\Pi) = n$ and assume the claim has been shown for pyramids of dimension $\dim(\Pi) - 1$. Consider any $(n-1)$-dimensional subspace $\Lambda \subset \mathbb{R}^n$ perpendicular to the base $B \subset \Pi$. For a polytope $R \subset \mathbb{R}^n$ let $R_\Lambda$ denote the image of $R$ in $\Lambda$ under the orthogonal projection $\mathbb{R}^n \to \Lambda$. Then $\Pi_\Lambda$ is an $(n-1)$-dimensional pyramid and we have the Minkowski sum representation

$$\Pi_\Lambda = (\Pi_1)_\Lambda + (\Pi_2)_\Lambda.$$

By induction hypothesis there are homothetic transformations of $\Lambda$ transforming $\Pi_\Lambda$ into $(\Pi_1)_\Lambda$ and $(\Pi_2)_\Lambda$ respectively. Considering all the possible subspaces $\Lambda \subset \mathbb{R}^n$ we conclude that

(i) both $\Pi_1$ and $\Pi_2$ are $n$-pyramids (provided none of them is just a point – in this situation the lemma is obvious) such that the cones they span at corresponding vertices are parallel shifts of the cone spanned by $\Pi$ at its apex $v$,

(ii) the corresponding bases of $\Pi_1$ and $\Pi_2$ are parallel to $B$.

That is exactly what we wanted to show. □

*Proof of Theorem 5.9.3*. — As in the proof of 7.2 we can assume $K[F] = \operatorname{Im}(h)$, and, furthermore, $S_P \cap \operatorname{Ker}(h) = \varnothing$, for otherwise $h$ itself passes through a facet retraction. Thus $h$ can be extended to the Laurent polynomial ring $K[L]$, $L = \mathbb{Z}^{n+1}$, and in particular to a retraction of $K[\overline{S}_P]$ with image $K[\overline{S}_Q]$. The latter restricts to retractions $K[iP] \to K[iQ]$ for all $i$. The kernel of the extension $h'$ is a height 1 prime ideal and thus principal; $\operatorname{Ker}(h') = \varphi K[L]$ and $\operatorname{Ker}(h) = (\varphi K[L]) \cap K[P]$ for some element $\varphi \in K[L]$.

Since $F$ is a base of $h$, $\operatorname{Ker}(h)$ contains the elements $x - \ell$, $x \in L_P \setminus F$, $\ell = h(x)$, and $\ell$ is a linear form on the points of $L_F$. Then $\mathrm{N}(\varphi)$ is a Minkowski summand of the pyramid $\mathrm{N}(x - \ell)$ with vertex at $x$. One can shift $\mathrm{N}(\varphi)$ by an integer vector into $\mathrm{N}(x - h(x)) \subset P$ such that the image $R$ satisfies

$$(**) \qquad R \subset P \text{ and } R \cap F \neq \varnothing.$$

Evidently $R$ is the Newton polytope of $y\varphi$ for some $y \in \mathbb{Z}^{n+1}$. Replacing $\varphi$ by $y\varphi$, we can assume that $\mathrm{N}(\varphi)$ satisfies $(**)$.

By Lemma 5.9.4 $\mathrm{N}(\varphi)$ is homothetic to $\mathrm{N}(x - \ell)$. Clearly, $F \cap \mathrm{N}(\varphi)$ is a base of $\mathrm{N}(\varphi)$. The corresponding apex of $\mathrm{N}(\varphi)$ is some $z \in L_P \setminus F$.

Consider the valuation

$$v_F : \mathbb{Z}^{n+1} \longrightarrow \mathbb{Z}$$

determined by the conditions:

$$\operatorname{Im}(v_F) = \mathbb{Z},\ v_F(L_F) = 0,\ v_F(L_P) \geqslant 0$$

We claim: $v_F(z)=1$ and $y+b-z\in P$ for any $b\in L_{F\cap \mathrm{N}(\varphi)}$ and $y\in L_P\smallsetminus F$.

In fact, for $i\in\mathbb{N}$ big enough there is an element $z'\in L_{iP}$ such that $v_F(z')=1$. Since $iF$ is a base of the induced retraction $\overline{h}_i : K[iP]\to K[iP]$ there exists a linear form $\ell'$ on $L_{iF}$ such that $z'-\ell'\in\mathrm{Ker}(h)$. Thus $\mathrm{N}(\varphi)$ is a Minkowski summand of $\mathrm{N}(z'-\ell')$. Because of the condition ($**$) we conclude $v_F(z)\leqslant v_F(z')$. Hence $v_F(z)=1$.

Now choose $y\in L_P\smallsetminus F$. Since $F$ is a base of $h$, we can write $y-\ell''\in\mathrm{Ker}(h)$ for some linear form $\ell''$ on the points of $L_F$. Therefore, the pyramid $\mathrm{N}(\varphi)$ is a Minkowski summand of the pyramid $\mathrm{N}(y-\ell'')$ which has its apex at $y$. By Lemma 5.9.4 the cones spanned by these pyramids at their vertices are the same modulo a parallel shift. This observation in conjunction with the already established equality $v_F(z)=1$ makes the claim clear.

We have shown that the vectors $b-z\in\mathbb{Z}^n$, $b\in L_{F\cap\mathrm{N}(\varphi)}$, are column vectors for $P$. Now, by Lemma 5.3.1(b) there exists $\varepsilon\in\mathbb{A}(F))$ such that $\varepsilon(\varphi)=cz$ for some $c\in K^*$. Therefore, $\mathrm{Ker}(h^\varepsilon)$ is the monomial prime ideal $(L_P\smallsetminus F)K[P]\subset K[P]$, and this finishes the proof of Theorem 5.9.3. □

The next theorem shows that one has a clear picture of all retractions if $P$ is of dimension 2.

***Theorem 5.9.5***. — *Let $P$ be a lattice polygon, i. e. a lattice polytope of dimension 2. Then every codimension 1 retraction $h : K[P]\to K[P]$ is based and, therefore, either $h^\tau=\iota\circ\rho_{(P,H,w)}$ for some lattice segmental fibration $(P,H,w)$, $\tau\in\mathbb{T}_K(P)$ and a $K$-embedding $\iota : K[H\cap P]\to K[P]$, or $h^\varepsilon=\iota\circ\pi_F$ for a facet $F\subset P$, $\varepsilon\in\mathcal{E}_K(P)$ and a $K$-embedding $\iota : K[F]\to K[P]$.*

By Theorems 5.9.2 and 5.9.3 it is enough to find a base for $h$. The first step in its construction is given by

***Proposition 5.9.6***. — *A multiple $c\Delta_1$, $c\in\mathbb{N}$, of the unit segment $\Delta_1$ can be embedded as a lattice polytope into a lattice polytope $P$ if and only if there is a $K$-algebra embedding of $K[c\Delta_1]$ into $K[P]$.*

*Proof*. — Clearly, without loss of generality we can assume $c\geqslant 2$.

Let $\varepsilon : K[c\Delta_1]\to K[P]$ be an embedding. We write

$$L_{c\Delta_1}=\{x_0,x_1,\dots,x_c\}.$$

Thus we have the equations $\varepsilon(x_{i-1})\varepsilon(x_{i+1})=\varepsilon(x_i)^2$ for $i\in[1,c-1]$. Put

$$\varphi=\frac{\varepsilon(x_1)}{\varepsilon(x_0)}=\frac{\varepsilon(x_2)}{\varepsilon(x_1)}=\cdots.$$

In the quotient field of $K[P]=K[\mathbb{Z}^3]$ we can write $\varphi=\varphi_1/\varphi_2$ with coprime $\varphi_1,\varphi_2\in K[P]$. The equality $\varphi_2^c\varepsilon(x_c)=\varphi_1^c\varepsilon(x_0)$ (and the factoriality of $K[\mathbb{Z}^3]$) imply that $\varphi_1^c$ divides $\varepsilon(x_c)$ and $\varphi_2^c$ divides $\varepsilon(x_0)$.

*Case* (a). Both $\varphi_1$ and $\varphi_2$ are monomials in $K[\mathbb{Z}^3]$. In this situation the Newton polygon $\mathrm{N}(\varepsilon(x_c))$ is the parallel shift of $\mathrm{N}(\varepsilon(x_0))$ by the $c$-th multiple of the vector representing the support term of the monomial $\varphi$. But then the existence of the desired embedding $c\Delta_1 \to \operatorname{conv}(\mathrm{N}(\varepsilon(x_0)), \mathrm{N}(\varepsilon(x_c)) \subset P$ is obvious.

*Case* (b). At least one of $\varphi_1$ and $\varphi_2$, say $\varphi_1$, is not a monomial. Then $c\Delta_1$ can be embedded in any of the edges of the polygon $\mathrm{N}(\varphi_1^c) = c\,\mathrm{N}(\varphi_1)$. Since $\varepsilon(x_c) = \psi\varphi_1^c$ for some $\psi \in K[\mathbb{Z}^3]$, we get

$$\mathrm{N}(\varepsilon(x_c)) = \mathrm{N}(\psi) + \mathrm{N}(\varphi_1^c)$$

and the existence of an embedding $c\Delta_1 \to \mathrm{N}(\varepsilon(x_c)) \subset P$ is evident. □

***Remark 5.9.7***. — We expect that Proposition 5.9.6 holds without any restrictions: for lattice polytopes $P$ and $Q$ a $K$-algebra embedding $K[Q] \to K[P]$ should only exist if $Q$ can be embedded into $P$ (as a lattice subpolytope).

One cannot exclude a priori that the retraction $h$ acts injectively on the embedded $K[c\Delta_1]$, and it takes some steps to overcome this difficulty. For the details we refer the reader to [**BG4**].

**5.10. Tame homomorphisms.** — Assume we are given two lattice polytopes $P, Q \subset \mathbb{R}^d$ and a homomorphism $f : K[P] \to K[Q]$ in $\mathrm{Pol}(K)$. Under certain conditions there are several standard ways to derive new homomorphisms from it.

First assume we are given a subpolytope $P' \subset P$ and a polytope $Q' \subset \mathbb{R}^n$, $d \leqslant n$, such that $f(K[P']) \subset K[Q']$. Then $f$ gives rise to a homomorphism $f' : K[P'] \to K[Q']$ in a natural way. (Notice that we may have $Q \subset Q'$.) Also if $P \cong \tilde{P}$ and $Q \cong \tilde{Q}$ are lattice polytope isomorphisms, then $f$ induces a homomorphism $\tilde{f} : K[\tilde{P}] \to K[\tilde{Q}]$. We call these types of formation of new homomorphisms *polytope changes*.

Now consider the situation when $\mathrm{Ker}(f) \cap S_P = \varnothing$. Then $f$ extends uniquely to a homomorphism $\overline{f} : K[\overline{S}_P] \to K[\overline{S}_Q]$ of the normalizations. Here $\overline{S}_P = \{x \in \mathrm{gp}(S_P) \mid x^m \in S_P \text{ for some } m \in \mathbb{N}\}$ and similarly for $\overline{S}_Q$. This extension is given by

$$\overline{f}(x) = \frac{f(y)}{f(z)}, \qquad x \in \overline{S}_P,\ x = \frac{y}{z},\ y \in S_P,\ z \in S_Q.$$

For every natural number $c$ the subalgebra of $K[\overline{S}_P]$ generated by the homogeneous component of degree $c$ is naturally isomorphic to the polytopal algebra $K[cP]$, and similarly for $K[\overline{S}_Q]$. Therefore, the restriction of $\overline{f}$ gives rise to a homomorphism $f^{(c)} : K[cP] \to K[cQ]$. We call the homomorphisms $f^{(c)}$ *homothetic blow-ups* of $f$. (Note that $K[cP]$ is often a proper overring of the $c$th Veronese subalgebra of $K[P]$.)

One more process of deriving new homomorphisms is as follows. Assume that homomorphisms $f, g : K[P] \to K[Q]$ are given such that

$$\mathrm{N}(f(x)) + \mathrm{N}(g(x)) \subset Q \qquad \text{for all } x \in L_P;$$

here N(−) denotes the Newton polytope, and + is the Minkowski sum in $\mathbb{R}^d$. Then we have $z^{-1}f(x)g(x) \in K[Q]$ where $z = (0, \dots, 0, 1) \in S_Q$. Clearly, the assignment

$$x \longmapsto z^{-1}f(x)g(x), \qquad x \in L_P,$$

extends to a Pol($K$)-homomorphism $K[P] \to K[Q]$, which we denote by $f \star g$. We call this process *Minkowski sum* of homomorphisms. (By convention, $\mathrm{N}(0) = \varnothing$, and $P + \varnothing = \varnothing$.)

All the three mentioned recipes have a common feature: the new homomorphisms are defined on polytopal algebras of dimension at most the dimension of the sources of the old homomorphisms. As a result we are not able to really create a non-trivial class of homomorphisms using only these three procedures. This possibility is provided by the fourth (and last in our list) process.

Suppose $P$ is a pyramid with vertex $v$ and basis $P_0$ such that $L_P = \{v\} \cup L_{P_0}$, that is $P = \mathrm{join}(v, P_0)$. Then $K[P]$ is a polynomial extension $K[P_0][v]$. In particular, if $f_0 : K[P_0] \to K[Q]$ is an arbitrary homomorphism and $q \in K[Q]$ is any element, then $f_0$ extends to a homomorphism $f : K[P] \to K[Q]$ with $f(v) = q$. We call $f$ a *free extension* of $f_0$.

***Conjecture 5.10.1***. — *Every homomorphism in* Pol($K$) *is obtained by a sequence of taking free extensions, Minkowski sums, homothetic blow-ups, polytope changes and compositions, starting from the identity mapping $K \to K$. Moreover, there are normal forms of such sequences for idempotent endomorphisms.*

Observe that for general homomorphisms we do not mean that the constructions mentioned in the conjecture are to be applied in certain order so that we get normal forms: we may have to repeat a procedure of the same type at different steps. However, the description of the automorphism group of a polytopal semigroup algebra in Theorem 5.3.2 and Theorem 5.10.3 below show that for special classes of homomorphisms such normal forms are possible.

We could call the homomorphisms obtained in the way described by Conjecture 5.10.1 just *tame*. Then we have the *tame* subcategory $\mathrm{Pol}(K)_{\mathrm{tame}}$ (with the same objects), and the conjecture asserts that actually $\mathrm{Pol}(K)_{\mathrm{tame}} = \mathrm{Pol}(K)$.

***Remark 5.10.2***

(a) The correctness of Conjecture 5.10.1 may depend on whether or not $K$ is algebraically closed. For instance, some of the arguments we have used in the analysis of retractions go through only for algebraically closed fields.

(b) Theorems 5.3.2, 5.9.2, 5.9.3, and 5.9.5 can be viewed as substantial refinements of the conjecture above for the corresponding classes of homomorphisms. Observe that the tameness of elementary automorphisms follows from their alternative description in Subsection 5.3. We also need the tameness of the following classes of homomorphisms: automorphisms that map monomials to monomials, retractions of the type $\rho_{(P,H,w)}$ and $\pi_F$ and the splitting embeddings $\iota$ as in Theorem 5.9.5. This follows from Theorem 5.10.3 and Corollary 5.10.4 below.

The next result shows that certain basic classes of morphisms in $\mathrm{Pol}(K)$ are tame.

***Theorem 5.10.3***. — *Let $K$ be a field (not necessarily algebraically closed), and $c \in \mathbb{N}$. Then*

(a) *every homomorphism from $K[c\Delta_n]$, $n \in \mathbb{N}$, is tame,*
(d) *if $\iota : K[c\Delta_n] \to K[P]$ splits $\rho_{(P,H,W)}$ for some lattice fibration $(P, H, W)$ or $\pi_E$ for some face $E \subset P$ then there is a normal form for representing $\iota$ in terms of certain basic tame homomorphisms.*

***Corollary 5.10.4***. — *For every field $K$ the homomorphisms in $\mathrm{Pol}(K)$ that respect monomial structures are tame.*

*Proof*. — Assume $f : K[P] \to K[Q]$ is a homomorphism respecting the monomial structures and such that $\mathrm{Ker}(f) \cap S_P = \varnothing$. By a polytope change we can assume $P \subset c\Delta_n$ for a sufficiently big natural number $c$, where $n = \dim P$ and $\Delta_n$ is taken in the lattice $\mathbb{Z}L_P$. In this situation there is a bigger lattice polytope $Q' \supset Q$ and a unique homomorphism $g : K[c\Delta_n] \to K[Q']$ for which $g|_{L_P} = f|_{L_P}$. By Theorem 5.10.3 $f$ is tame.

Consider the situation when the ideal $I = (\mathrm{Ker}(f) \cap S_P)K[P]$ is a nonzero prime monomial ideal and there is a face $P_0 \subset P$ such that $\mathrm{Ker}(f) \cap L_{P_0} = \varnothing$ and $f$ factors through the face retraction $\pi : K[P] \to K[P_0]$, that is $\pi(x) = x$ for $x \in L_{P_0}$ and $\pi(x) = 0$ for $x \in L_P \smallsetminus L_{P_0}$. In view of the previous case we are done once the tameness of face retractions has been established.

A face retraction is a composite of facet retractions. Therefore we can assume that $P_0$ is a facet of $P$. Let $(\mathbb{R}P)_+ \subset \mathbb{R}P$ denote the halfspace that is bounded by the affine hull of $P_0$ and contains $P$. There exists a unimodular (with respect to $\mathbb{Z}L_P$) lattice simplex $\Delta \subset (\mathbb{R}P)_+$ such that $\dim \Delta = \dim P$, the affine hull of $P_0$ intersects $\Delta$ in one of its facets and $P \subset c\Delta$ for some $c \in \mathbb{N}$. But then $\pi$ is a restriction of the corresponding facet retraction of $K[c\Delta]$, the latter being a homothetic blow-up of the corresponding facet retraction of the polynomial ring $K[\Delta_n]$ – obviously a tame homomorphism. □

*Proof of Theorem 5.10.3*. — We will use the notation $\{x_0, \dots, x_n\} = L_{\Delta_n}$. Every lattice point $x \in c\Delta_n$ has a unique representation $x = a_0x_0 + \cdots + a_nx_n$ where the $a_i$ are nonnegative integer numbers satisfying the condition $a_0 + \cdots + a_n = c$. The numbers $a_i$ are the *barycentric coordinates* of $x$ in the $x_i$.

Let $f : K[c\Delta_n] \to K[P]$ be an arbitrary homomorphism.

First consider the case when one of the points from $L_{c\Delta_n}$ is mapped to $0 \in K$. In this situation $f$ is a composite of facet retractions and a homomorphism from $K[c\Delta_m]$ with $m < n$. As observed in the proof of Corollary 5.10.4 facet retractions are tame.

Therefore we can assume that none of the $x_i$ is mapped to 0. By a polytope change we can also assume $L_P \subset \{X_1^{a_1} \cdots X_r^{a_r} Y^b Z \mid a_i, b \geqslant 0\}$, $r = \dim P - 1$.

Consider the polynomials $\varphi_x = Z^{-1} f(x) \in K[X_1, \dots, X_r, Y]$, $x \in L_{c\Delta_n}$. Then the $\varphi_x$ are subject to the same binomial relations as the $x$. One the other hand the multiplicative semigroup $K[X_1, \dots, X_r, Y] \smallsetminus \{0\}/K^*$ is a free commutative semigroup and, as such, is an inductive limit of free commutative semigroups of finite rank. Therefore, by Lemma 5.10.5 below there exist polynomials $\psi, \eta_i \in K[X_1, \dots, X_r, Y]$, $i \in [0, n]$, and scalars $t_x \in K^*$, $x \in L_{c\Delta_n}$, such that $\varphi_x = t_x \psi \eta_0^{a_0} \cdots \eta_n^{a_n}$ where the $a_i$ are the barycentric coordinates of $x$. Clearly, $t_x$ are subject to the same binomial relations as the $x \in L_{c\Delta_n}$. Therefore, after the normalizations $\eta_i \mapsto t_{x_i} \eta_i$ ($i \in [0, n]$) we get $\varphi_x = \psi \eta_0^{a_0} \cdots \eta_n^{a_n}$. But the latter equality can be read as follows: $f$ is obtained by a polytope change applied to $\Psi \star \Theta^{(c)}$, where

(i) $\Psi : K[c\Delta_n] \to K[Q]$, $\Psi(x) = \psi Z$, $x \in L_{c\Delta_n}$,

(ii) $\Theta : K[\Delta_n] \to K[Q]$, $\Theta(x_i) = \eta_i Z$, $i \in [0, n]$,

and $Q$ is a sufficiently large lattice polytope so that it contains all the relevant lattice polytopes. Now $\Psi$ is tame because it can be represented as the composite map

$$K[c\Delta_n] \xrightarrow{L_{c\Delta_n} \to t} K[t] \xrightarrow{t \mapsto \psi Z} K[Q]$$

(the first map is the $c$th homothetic blow-up of $K[\Delta_n] \to K[t]$, $x_i \mapsto t$ for all $i \in [0, c]$) and $\Theta$ is just a free extension of the identity embedding $K \to K[Q]$.

(b) First consider the case of lattice segmental fibrations.

Consider the rectangular prism $\Pi = (c\Delta_n) \times (m\Delta_1)$. By a polytope change (assuming $m$ is sufficiently large) we can assume that $P \subset \Pi$ so that $H$ is parallel to $c\Delta_n$: The

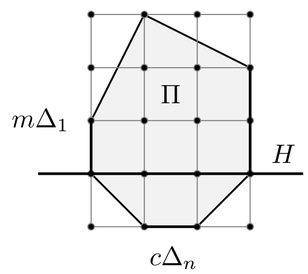


FIGURE 11

lattice point $(x, b) \in \Pi$ will be identified with the monomial $X_1^{a_1} \cdots X_n^{a_n} Y^b Z$ whenever we view it as a monomial in $K[\Pi]$, where the $a_i$ are the corresponding barycentric coordinates of $x$ (see the proof of (a) above). (In other words, the monomial $X_1^{a_1} \cdots X_n^{a_n}$ is identified with the point $x = (c - a_1 - \cdots - a_n)x_0 + a_1 x_1 + \cdots + a_n x_n \in c\Delta_n$.)

Assume $A : K[c\Delta_n] \to K[m'\Delta_1]$ is a homomorphism of the type $A(x_i) = aZ$, $i \in [0, c]$ for some $a \in K[Y]$ satisfying the condition $a(1) = 1$. Consider a homomorphism

$B : K[\Delta_n] \to K[\Pi']$, $\Pi' = \Delta_n \times (m'\Delta_1)$ that splits the projection $\rho' : K[\Pi'] \to K[\Delta_n]$, $\rho'(ZY^b) = Z$ and $\rho'(X_iY^bZ) = X_iZ$ for $i \in [1,n]$, $b \in [0,m']$. The description of such homomorphisms is clear – they are exactly the homomorphisms $B$ for which

$$B(x_0) = B_0 \in Z + (Y-1)(ZK[Y] + X_1ZK[Y] + \cdots + X_nZK[Y]),$$
$$B(x_i) = B_i \in X_iZ + (Y-1)(ZK[Y] + X_1ZK[Y] + \cdots + X_nZK[Y]),\ i \in [1,n],$$
$$\deg_Y B_i \leqslant m' \text{ for all } i \in [0,n].$$

Clearly, all such $B$ are tame.

In case $m \geqslant \max\{m' + c\deg_Y B_i\}_{i=0}^n$ we have the homomorphism $A \star B^{(c)}$ : $K[c\Delta_n] \to K[\Pi]$ which obviously splits the projection $\rho : K[\Pi] \to K[c\Delta_n]$ defined by

$$\rho(X_1^{a_1} \cdots X_n^{a_n} Y^b Z) = X_1^{a_1} \cdots X_n^{a_n} Z.$$

Assume $\iota$ splits $\rho_{(P,H,w)}$. The standard facts on Newton polytopes imply the following: the polynomials $\psi$ and $\eta_i$, mentioned in the proof of (a), that correspond to $\iota$, satisfy the conditions: $\psi \in K[Y]$ and $\eta_i \in K[Y]+X_1K[Y]+\cdots+X_nK[Y]$. Is is also clear that upon evaluation at $Y = 1$ we get $\psi(1), \eta_i(X_1,\dots,X_n,1) \in K^*$, $i \in [0,n]$. Therefore, after the normalizations $\psi \mapsto \psi^{-1}(1)\psi$, $\eta_i \mapsto \eta_i(X_1,\dots X_n,1)^{-1}\eta_i$ we conclude that $\iota$ is obtained by a polytope change applied to $A \star B^{(c)}$ as above (with respect to $a = \psi$, $B_0 = \eta_i Z$, $i \in [0,n]$).

For a lattice fibration $(P,H,W)$ of higher codimension similar arguments show that $\iota$ is obtained by a polytope change applied to $A \star B^{(c)}$, where

> $B$ is a splitting of a projection of the type $\rho_{(P',H',W')}$ such that the base polytope $P' \cap H'$ is a unit simplex and
>
> $A$ is a homomorphism defined by a single polynomial whose Newton polytope is parallel to $W'$.

We skip the details for splittings of face retractions and only remark that similar arguments based on Newton polytopes imply the following: all such splittings are obtained by polytope changes applied to $A \star B^{(c)}$ where $B$ is a splitting of a face retraction onto a polynomial ring and $A$ is again defined by a single polynomial. □

***Lemma 5.10.5***. — *Assume we are given an integral affine mapping $\alpha : c\Delta_n \to \mathbb{R}_+^d$ for some natural numbers $c$, $n$ and $d$. Then there exists an element $v \in \mathbb{Z}_+^d$ and a integral affine mapping $\beta : \Delta_n \to \mathbb{R}_+^d$ such that $\alpha = v + c\beta$.*

*Proof*. — Assume $\alpha(cx_i) = (a_{i1},\dots,a_{id})$, $i \in [0,n]$ (the $x_i$ as above). Consider the vector

$$v = \big(\min\{a_{i1}\}_{i=0}^n, \dots, \min\{a_{id}\}_{i=0}^n\big).$$

It suffices to show that all the vectors $\alpha(cx_i) - v$ are $c$th multiples of integral vectors. But for every index $l \in [1,d]$ the $l$th component of either $\alpha(cx_i) - v$ or $\alpha(cx_j) - v$ for some $j \neq i$ is zero. In the first case there is nothing to prove and in the second case the desired divisibility follows from the fact that $\alpha(cx_i) - \alpha(cx_j) = (\alpha(cx_i) - v) - (\alpha(cx_j) - v)$ is a $c$th multiple of an integral vector (because $\alpha$ is integral affine). □

# Index

## References


[AM] S. Aguzzoli and D. Mundici, *An algorithmic desingularization of* 3*-dimensional toric varieties*, Tôhoku Math. J. **46** (1994), 557–572.

[Bor] A. Borel, *Linear algebraic groups*, Springer, 1991.

[BoGo] C. Bouvier and G. Gonzalez-Sprinberg, *Système generateurs minimal, diviseurs essentiels et G-désingularizations de variétés toriques*, Tôhoku Math. J. **46** (1994), 125–149.

[Bri] M. Brion, *Sur les modules de covariants*, Ann. scient. Éc. Norm. Sup., 4e série, **26** (1993), 1–21.

[Bru] W.Bruns, *Diskrete Kegel und der Satz von Carathéodory.* In: R. Künzel (ed.), Wege der Wissenschaft, Universitätsverlag Rasch, Osnabrück 1999, pp. 361–371. (English version available at `http://www.mathematik.uni-osnabrueck.de/preprints/kommalg.html`)

[BG1] W. Bruns and J. Gubeladze, *Polytopal linear groups*, J. Algebra **218** (1999), 715–737.

[BG2] W. Bruns and J. Gubeladze, *Rectangular simplicial semigroups.* In: D. Eisenbud (ed.), Commutative algebra, algebraic geometry, and computational methods, Springer Singapore 1999, pp. 201–214.

[BG3] W. Bruns and J. Gubeladze, *Normality and covering properties of affine semigroups*, J. Reine Angew. Math. **510**, 151–178, 1999.

[BG4] W. Bruns and J. Gubeladze, *Polytopal linear retractions*, Trans. Amer. Math. Soc., **354** (2002), 179–203.

[BG5] W. Bruns and J. Gubeladze, *Polytopal linear algebra*, Beitr. Algebra Geom., to appear.

[BG6] W. Bruns and J. Gubeladze, *Polyhedral algebras, toric arrangements, and their groups*, Advanced Studies in Pure Mathematics, **33** (2002), 1–51.

[BG7] W. Bruns and J. Gubeladze, *Divisorial linear algebra of normal semigroup rings*, Algebr. Represent. Theory, to appear.

[BG8] W. Bruns and J. Gubeladze, *Polyhedral* $K_2$, preprint.

[BG9] W. Bruns and J. Gubeladze, *Higher polyhedral* $K$*-groups*, preprint.

[BGHMW] W. Bruns, J. Gubeladze, M. Henk, A. Martin, and R. Weismantel, *A counterexample to an integer analogue of Carathéodory's theorem*, J. Reine Angew. Math. **510**, 179–185, 1999.

[BGT1] W. Bruns, J. Gubeladze, and N. V. Trung, *Normal polytopes, triangulations, and Koszul algebras*, J. Reine Angew. Math. **485** (1997), 123–160.

[BGT2] W. Bruns, J. Gubeladze, and N. V. Trung, *Problems and algorithms for affine semigroups*, Semigroup Forum, **64** (2002), 180–212.

[BGu] W. Bruns and A. Guerrieri, *The Dedekind–Mertens lemma and determinantal rings*, Proc. Amer. Math. Soc. **127**, (1999), 657–663.

[BH] W. Bruns and J. Herzog. *Cohen–Macaulay rings* (Rev. Ed.). Cambridge University Press, 1998.

[BK] W. Bruns and R. Koch, *Normaliz*, a program to compute normalizations of semigroups. Available by anonymous `ftp` from `ftp.mathematik.Uni-Osnabrueck.DE/pub/osm/kommalg/software/`.

[BV] W. Bruns and U. Vetter, *Determinantal rings*, Springer Lecture Notes **1327** (1988).

[Bue] D. Buehler, *Homogener Koordinatenring und Automorphismengruppe vollständiger torischer Varietäten*, Diplomarbeit, Basel, 1996.

[Bur] E. Burger, *Über homogene lineare Ungleichungssysteme*, Z. Angew. Math. Mech. **36** (1956), 135–139.

[Ch] L. Chouinard, *Krull semigroups and divisor class groups*, Canad. J. Math. **33** (1981), 1459–1468.

[CFS] W. Cook, J. Fonlupt, and A. Schrijver, *An integer analogue of Carathéodory's theorem*, J. Comb. Theory, Ser. B **40** (1986), 63-70.

[Cos] D. Costa, *Retracts of polynomial rings*, J. Algebra **44** (1977), 492–502.

[Cox] D. A. Cox, *The homogeneous coordinate ring of a toric variety*, J. Algebr. Geom. **4** (1995), 17–50.

[Da] V. Danilov, *The geometry of toric varieties*, Russian Math. Surveys **33** (1978), 97–154.

[De] M. Demazure, *Sous-groupes algébriques de rang maximum du groupe de Cremona*, Ann. Sci. École Norm. Sup. (4) **3** (1970), 507-588.

[Ei] D. Eisenbud, *Commutative algebra with a view towards algebraic geometry*, Springer 1995.

[ES] D. Eisenbud and B. Sturmfels, *Binomial ideals*, Duke Math. J. **84** (1996), 1–45.

[FZ] R. T. Firla and G. M. Ziegler, *Hilbert bases, unimodular triangulations, and binary covers of rational polyhedral cones*, Discrete Comp. Geom. **21** (1999), 205–216.

[Fo] R. M. Fossum, *The divisor class group of a Krull domain*, Springer 1973.

[Fr] F. G. Frobenius, *Gesammelte Abhandlungen*, Hrsg. von J.-P. Serre, Bd. 3, Springer, Berlin, 1968.

[Fu] W. Fulton, *Introduction to toric varieties*, Princeton University Press, 1993.

[GKZ] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, 1994.

[Gi] R. Gilmer, *Commutative Semigroup Rings*, U. Chicago Press, 1984.

[Gu0] J. Gubeladze, *Anderson's conjecture and the maximal monoid class, over which projective modules are free*, Math. USSR Sbornik **63** (1989), 165–180.

[Gu1] J. Gubeladze, *The elementary action on unimodular rows over a monoid ring*, J. Algebra **148** (1992), 135–161.

[Gu2] J. Gubeladze, *Geometric and algebraic representations of commutative cancellative monoids*, Proc. A. Razmadze Math. Inst., Georg. Acad. Sci. **113** (1995), 31-81.

[Gu3] J. Gubeladze, *The isomorphism problem for commutative monoid rings*, J. Pure Appl. Algebra **129** (1998), 35–65.

[Ha] R. Hartshorne, *Algebraic geometry*, Springer, 1977.

[Ho] M. Hochster, *Rings of invariants of tori, Cohen–Macaulay rings generated by monomials, and polytopes*, Ann. Math. **96** (1972), 318–337.

[Hu] J. E. Humphreys, *Linear algebraic groups*, Springer 1975.

[KS] J.-M. Kantor and K. S. Sarkaria, *On primitive subdivisions of an elementary tetrahedron*, preprint.

[KKMS] G. Kempf, F. Knudsen, D. Mumford, and B. Saint-Donat. *Toroidal Embeddings I.* Lecture Notes in Math. **339**, Springer, 1973.

[MW] A. Martin and R. Weismantel, *Conjunctive cuts for Integer Programs*, preprint, Konrad-Zuse-Zentrum Berlin, SC 98-18, 1998.

[Ma] H. Matsumura. *Commutative ring theory.* Cambridge University Press, 1989.

[Oda] T. Oda, *Convex Bodies and algebraic geometry (An introduction to the theory of toric varieties)*, Springer, 1988.

[Se] A. Sebö, *Hilbert bases, Carathéodory's theorem, and combinatorial optimization*, in 'Integer Programming and Combinatorial Optimization' (R. Kannan, W. Pulleyblank, eds.), University of Waterloo Press, Waterloo 1990, 431–456.

[St1] R. P. Stanley, *Hilbert functions of graded algebras*, Adv. in Math. **28** (1978), 57–83.

[St2] R. P. Stanley, *Linear diophantine equations and local cohomology*, Invent. math. **68** (1982), 175–193.

[St3] R. P. Stanley, *Combinatorics and commutative algebra* (2nd ed.), Birkhäuser 1996.

[Stu] B. Sturmfels, *Gröbner bases and convex polytopes*, American Mathematical Society 1996.

[Te] B. Teissier, *Variétés toriques et polytopes*, Séminaire Bourbaki, 33e année (1980/81) n$^0$ 565, 71-84.

[Wa] W. Waterhouse, *Automorphisms of* $\det(X_{ij})$ *: The group scheme approach*, Adv. Math. **65** (1987), 171–203.

W. Bruns, FB Mathematik/Informatik, Universität Osnabrück, 49069 Osnabrück, Germany
*E-mail :* `Winfried.Bruns@mathematik.uni-osnabrueck.de`
*Url :* `http://www.mathematik.uni-osnabrueck.de/staff/phpages/brunsw.rdf.html`

J. Gubeladze, A. Razmadze Mathematical Institute, Alexidze St. 1, 380093 Tbilisi, Georgia
*E-mail :* `gubel@rmi.acnet.ge` • *Url :* `http://www.rmi.acnet.ge/~gubel`