

Polytopal Linear Groups

Winfried Bruns

FB Mathematik/ Informatik, Universität Osnabrück, 49069 Osnabrück, Germany

E-mail: Winfried.Bruns@mathematik.uni-osnabrueck.de

and

Joseph Gubeladze

A. Razmadze Mathematical Institute, Alexidze St. 1, 380093 Tbilisi, Georgia

E-mail: gubel@rmi.acnet.ge

Communicated by Craig Huneke

Received June 24, 1998

1. INTRODUCTION

The main objects of this paper are the graded automorphisms of polytopal semigroup rings, i.e., semigroup rings $k[S_P]$ where k is a field and S_P is the semigroup associated with a lattice polytope P (Bruns, Gubeladze, and Trung [BGT]). The generators of $k[S_P]$ correspond bijectively to the lattice points in P , and their relations are the binomials representing the affine dependencies of the lattice points.

The simplest examples of such rings are the polynomial rings $k[X_1, \dots, X_n]$ with the standard grading. They are associated to the unit simplices Δ_{n-1} . The graded automorphism group $\text{GL}_n(k)$ of $k[X_1, \dots, X_n]$ is generated by diagonal and elementary automorphisms. Our main result is a generalization of this classical fact to the graded automorphism group $\Gamma_k(P)$ of an arbitrary polytopal semigroup ring: It says that each automorphism $\gamma \in \Gamma_k(P)$ has a (non-unique) normal form as a composition of toric and elementary automorphisms and symmetries of the underlying polytope. In view of this analogy we call the groups $\Gamma_k(P)$ *polytopal linear groups*. As in the case of $\text{GL}_n(k)$, the elementary and toric automorphisms generate $\Gamma_k(P)$ if (and only if) this group is connected. The elementary automorphisms are defined in terms of so-called column structures on P .



The proof of the main result consists of two major steps. First we prove that an arbitrary automorphism that leaves the “interior” of $k[S_P]$ invariant preserves the monomial structure, and is therefore a composition of a toric automorphism and a symmetry. (If $k[S_P]$ is normal, the interior is just the canonical module.) Second we show how to “correct” a graded automorphism γ by elementary automorphisms such that the composition preserves the interior of $k[S_P]$. This correction is based on the action of γ on the divisorial ideals of the normalization of $k[S_P]$.

Our approach applies to arbitrary fields k , and it can even be generalized to graded automorphisms of an arbitrary normal affine semigroup ring (see Remark 3.3). A further application is a generalization from a single polytope to (lattice) polyhedral complexes [BG]. (Rings defined by lattice polyhedral complexes generalize polytopal semigroup rings in the same way as Stanley–Reisner rings generalize polynomial rings.

As an application outside the class of semigroup rings we determine the graded automorphism groups of the determinantal rings.

The geometric objects associated to polytopal semigroup rings are projective toric varieties. 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 recently been obtained by Cox [Co]. As we learnt after this work had been completed, Bühler [Bu] generalized Cox’s results to arbitrary complete toric varieties. In the last section we derive a description of the automorphism group of an arbitrary projective toric variety from our theorem on polytopal linear groups. This method has the advantage of working in arbitrary characteristic, and furthermore it generalizes to arrangements of toric varieties (see [BG]).

Notation and Terminology. Let n be a natural number and let P be a finite, convex, n -dimensional lattice polytope in \mathbb{R}^n ; i.e., P has vertices in $\mathbb{Z}^n \subset \mathbb{R}^n$. We will assume that the lattice points of P affinely generate the whole group \mathbb{Z}^n (this can always be achieved by a suitable change of the lattice of reference).

Let k be a field. The *polytopal algebra* (or *polytopal semigroup ring*) associated with P and k is the semigroup algebra $k[S_P]$, where S_P is the additive sub-semigroup of \mathbb{Z}^{n+1} generated by $\{(x, 1) \mid x \in P \cap \mathbb{Z}^n\}$ [BGT]. The cone spanned by S_P will be denoted by $C(P)$.

The k -algebra $k[S_P]$ and, more generally, $k[\mathbb{Z}^{n+1}] = k[\text{gp}(S_P)]$ is naturally graded in such a way that the last component of (the exponent vector of) a monomial is its degree. The semigroup S_P consists of homogeneous elements, and lattice points of P correspond to degree 1 elements of S_P . In order to avoid cumbersome notation we will in the following identify the

lattice points of P with the degree 1 elements in S_P , and, more generally, elements of \mathbb{Z}^{n+1} with Laurent monomials in $k[\mathbb{Z}^{n+1}]$.

It follows immediately from the definitions that $\Gamma_k(P)$ is an affine k -group: It can be identified with the closed subgroup of $GL_N(k)$, $N = \#(P \cap \mathbb{Z}^n)$, given by the equations that reflect the relations between the degree 1 monomials in S_P .

Recall that a sub-semigroup $S \subset \mathbb{Z}^m$, $m \in \mathbb{N}$, is called normal if $x \in \text{gp}(S)$ (the group of differences of S) and $cx \in S$ for some natural c imply $x \in S$. This is equivalent to the normality of $k[S]$ (for example, see Bruns and Herzog [BH, Chap. 6]). A lattice polytope P is called normal if S_P is a normal semigroup. Observe that P is normal if and only if $S_P = \mathbb{Z}^{n+1} \cap C(P)$.

A convention: when a semigroup is considered as the semigroup of monomials in the corresponding ring, we use multiplicative notation for the semigroup operation; otherwise additive notation is used.

2. COLUMN STRUCTURES ON LATTICE POLYTOPES

Let P be a lattice polytope as above.

DEFINITION 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 Fig. 1.

Moreover, the group \mathbb{Z}^n is a 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.

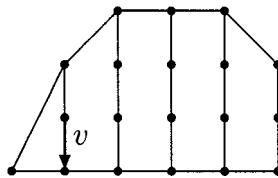


FIG. 1. A column structure.

LEMMA 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

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

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

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

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

where we denote the natural extension of φ_F to an \mathbb{R} -linear form on \mathbb{R}^n by φ_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 $\varphi_G(v) \geq 0$ 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 $\text{Col}(P) \subset \text{Col}(Q)$.

We further illustrate the notion of column vector by Fig. 2: the polytope P_1 has four column vectors, whereas the polytope P_2 has no column vector.

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

More generally, for any facet $F \subset P$ we define the linear form $\text{ht}_F: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ by $\text{ht}_F(y) = \varphi_F(y_1, \dots, y_n) - a_F y'_{n+1}$, where φ_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 $\text{ht}_F(x)$ of x is a non-negative integer. The kernel of 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 ht_F .

Clearly, for a column structure (P, v) and a lattice point $x \in P$ we have $\text{ht}_v(x) = \text{ht}_{P_v}(x)$, as follows immediately from Lemma 2.2.

3. ELEMENTARY AUTOMORPHISMS AND THE MAIN RESULT

Let (P, v) be a column structure and let $\lambda \in k$. 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 $\text{QF}(k[S_P])$, the quotient field of $k[S_P]$, by the assignment

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

Since ht_v extends to a group homomorphism $\mathbb{Z}^{n+1} \rightarrow \mathbb{Z}$ our mapping is a homomorphism from S_P to the multiplicative group of $\text{QF}(k[S_P])$. Now it is immediate from the definition of ht_v and Lemma 2.2 that the (isomorphic) image of S_P lies actually in $k[S_P]$. Hence this mapping gives rise to a graded k -algebra endomorphism e_v^λ of $k[S_P]$ preserving the degree of an element. By Hilbert function arguments e_v^λ is an automorphism.

Here is an alternative description of e_v^λ . By a suitable integral change of coordinates we may assume that $v = (0, -1, 0, \dots, 0)$ and that P_v lies in



FIG. 2. Two polytopes and their column structures.

the subspace \mathbb{R}^{n-1} (thus P is in the upper halfspace). Now consider the standard unimodular n -simplex Δ_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 Δ denote such a parallel translate. Then we have a graded k -algebra embedding $k[S_P] \subset k[S_\Delta]$. Moreover, $k[S_\Delta]$ 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 α of the subalgebra $k[S_\Delta]$, and α in turn can be restricted to an automorphism of $k[S_P]$, which is nothing else but e_v^λ .

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

Therefore the automorphisms of type e_v^λ will be called *elementary*.

LEMMA 3.1. *Let v_1, \dots, v_s be pairwise different column vectors for P with the same base facet $F = P_{v_i}$, $i = 1, \dots, s$. Then the mapping*

$$\varphi: \mathbb{A}_k^s \rightarrow \Gamma_k(P), \quad (\lambda_1, \dots, \lambda_s) \mapsto e_{v_1}^{\lambda_1} \circ \dots \circ e_{v_s}^{\lambda_s},$$

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 $e_{v_i}^{-\lambda_i}$.

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

Proof. We define a new k -algebra automorphism ϑ of $k[S_P]$ by first setting

$$\vartheta(x) = (1 + \lambda_1 v_1 + \dots + \lambda_s v_s)^{\text{ht}_F(x)} x$$

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

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

The image of the embedding φ given by Lemma 3.1 is denoted by $\mathbb{A}(F)$. Of course, $\mathbb{A}(F)$ may consist only of the identity map of $k[S_P]$, namely if there is no column vector with base facet F . In the case in which P is the

unit simplex and $k[S_P]$ is the polynomial ring, $\mathbb{A}(F)$ is the subgroup of all matrices in $GL_n(k)$ 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 $k[S_P]$ by restriction of its action on $k[\mathbb{Z}^{n+1}]$ given by

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

where 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 $k[S_P]$ which multiply each monomial by a scalar from k^* .

Second, the automorphism group $\Sigma(P)$ of the semigroup S_P is a finite subgroup of $\Gamma_k(P)$ in a natural way. It is the group of integral affine transformations mapping P onto itself.

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 $\text{ht}_v(x) = \text{ht}_{-v}(y)$ 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 $\Sigma(P)_{\text{inv}}$.

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

THEOREM 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*

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

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 $\#(F_i \cap \mathbb{Z}^n) \leq \#(F_{i+1} \cap \mathbb{Z}^n)$, $i \in [1, r - 1]$.

(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 $k[S_P]$ which induce the identity map on the divisor class group of the normalization of $k[S_P]$.*

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

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

Remark 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}^λ , 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 $\Sigma(P)_{\text{inv}}$ 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 (for example, see [BH, Chap. 6]).

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 $x \in S \setminus F$.

The only disadvantage here is that the condition for column vectors involves an infinite number of lattice points, while for polytopal rings one only has to look at lattice points in a finite polytope (due to Lemma 2.2).

Then one introduces analogously the notion of an elementary automorphism e_v^λ ($\lambda \in k$). The proof of Theorem 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 $\text{gp}(M) = \mathbb{Z}^n$) 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 $k[S_P]$, 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 $k[S_P]$ 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 $\text{Spec } k[S_P]$. 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 rings and toric varieties we determine the groups of graded automorphisms of the determinantal rings. In plain terms, Corollary 3.4 answers the

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

COROLLARY 3.4. *Let k be a field, X an $m \times n$ matrix of indeterminates, and $R = K[X]/I_{r+1}(X)$ 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 , $1 \leq r < \min(m, n)$.*

(a) *The connected component G^0 of unity in $G = \text{gr.aut}_k(R)$ is the image of $\text{GL}_m(k) \times \text{GL}_n(k)$ in $\text{GL}_{mn}(k)$ under the map indicated above, and is isomorphic to $\text{GL}_m(k) \times \text{GL}_n(k)/k^*$ 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 τ is the transposition.*

Proof. The singular locus of $\text{Spec } R$ 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 $R \rightarrow k[Y_i Z_j : i = 1, \dots, m, j = 1, \dots, n] \subset k[Y_1, \dots, Y_m, Z_1, \dots, Z_n]$ induced by the assignment $X_{ij} \mapsto Y_i Z_j$. Thus R is just the Segre product of $k[Y_1, \dots, Y_m]$ and $k[Z_1, \dots, Z_n]$, or, equivalently, $R \cong k[S_P]$ where P is the direct product of the unit simplices Δ_{m-1} and Δ_{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 $\text{Cl}(R) \cong \mathbb{Z}$; ideals representing the divisor classes 1 and -1 are given by $(Y_1 Z_1, \dots, Y_1 Z_n)$ and $(Y_1 Z_1, \dots, Y_m Z_1)$ [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 $\text{Cl}(R)$. Now the rest follows again from the theorem above. (Instead of the divisorial arguments one could also discuss the symmetry group of $\Delta_{m-1} \times \Delta_{n-1}$.) ■

4. PROOF OF THE MAIN RESULT

Set $\bar{S}_P = \mathbb{Z}^{n+1} \cap C(P)$. Then \bar{S}_P is the normalization of the semigroup S_P and $k[\bar{S}_P]$ is the normalization of the domain $k[S_P]$. Let $\bar{\Gamma}_k(P)$ denote

the group of graded k -algebra automorphisms of $k[\bar{S}_P]$. Since any automorphism of $k[S_P]$ extends to a unique automorphism of $k[\bar{S}_P]$ we have a natural embedding $\Gamma_k(P) \subset \bar{\Gamma}_k(P)$. On the other hand, $k[\bar{S}_P]$ and $k[S_P]$ have the same homogeneous components of degree 1. Hence $\Gamma_k(P) = \bar{\Gamma}_k(P)$. Nevertheless we will use the notation $\bar{\Gamma}_k(P)$, emphasizing the fact that we are dealing with automorphisms of $k[\bar{S}_P]$; $\Sigma(P)$ and $\Sigma(P)_{\text{inv}}$ will refer to their images in $\bar{\Gamma}_k(P)$. Furthermore, the extension of an elementary automorphism e_v^λ is also denoted by e_v^λ ; it satisfies the rule $e_v^\lambda(x) = (1 + \lambda v)^{\text{ht}_v(x)} x$ for all $x \in \bar{S}_P$. (The equation $\Gamma_k(P) = \bar{\Gamma}_k(P)$ shows that the situation considered in Remark 3.3(c) really generalizes Theorem 3.2; furthermore it explains the difference between polytopal algebras and arbitrary graded semigroup rings generated by their degree 1 elements.)

In the following it is sometimes necessary to distinguish elements $x \in \bar{S}_P$ from products ζz with $\zeta \in k$ and $z \in \bar{S}_P$. We will call x a *monomial* and ζz a *term*.

Suppose $\gamma \in \bar{\Gamma}_k(P)$ maps every monomial x to a term $\lambda_x y_x$, $y_x \in \bar{S}_P$. Then the assignment $x \mapsto y_x$ is also a semigroup automorphism of \bar{S}_P . Denote it by σ . 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 $\text{int}(C(P))$ denote the interior of the cone $C(P)$ and let $\omega = (\text{int}(C(P)) \cap \mathbb{Z}^{n+1})k[\bar{S}_P]$ be the corresponding monomial ideal. (It is known that ω is a canonical module of $k[\bar{S}_P]$; see Danilov [Da], Stanley [St], or [BH, Chap. 6].)

LEMMA 4.1. (a) Suppose $\gamma \in \bar{\Gamma}_k(P)$ preserves the ideal ω . 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 γ maps monomials to terms.

First consider a non-zero monomial $x \in \bar{S}_P \cap \omega$. We have $\gamma(x) \in \omega$. Since x is an ‘‘interior’’ monomial, $k[\bar{S}_P]_x = k[\mathbb{Z}^{n+1}]$. On the other hand $k[\mathbb{Z}^{n+1}] \subset k[\bar{S}_P]_{\gamma(x)}$. Indeed, since $\text{gp}(\bar{S}_P) = \mathbb{Z}^{n+1}$, it just suffices to observe that for any monomial $z \in \bar{S}_P$ 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 $k[\bar{S}_P]_{\gamma(x)}$.

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

$$\mathbb{Z}^{n+1} \approx U\left(k[\bar{S}_P]_{\gamma(x)}\right)/k^*.$$

On the other hand we have seen that \mathbb{Z}^{n+1} is embedded into $U(k[\bar{S}_P]_{\gamma(x)})/k^*$ so that the elements from \bar{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 $\text{rank}(U(k[\bar{S}_P]_{\gamma(x)})/k^*) > n + 1$, a contradiction.

Now let $y \in \bar{S}_P$ be an arbitrary monomial and z a monomial in ω . 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. ■

In the light of Lemma 4.1(a) we see that for Theorem 3.2(a) it suffices to show the following claim: For every $\gamma \in \bar{\Gamma}_k(P)$ there exist $\alpha_1 \in \mathbb{A}(F_1), \dots, \alpha_r \in \mathbb{A}(F_r)$ such that

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

and the F_i satisfy the side conditions of Theorem 3.2(a).

For a facet $F \subset P$ we have constructed the group homomorphism $\text{ht}_F: \mathbb{Z}^{n+1} \rightarrow \mathbb{Z}$. We now define the divisorial ideal of F by

$$\text{Div}(F) = \{x \in \bar{S}_P \mid \text{ht}_F(x) > 0\}k[\bar{S}_P].$$

It is clear that $\omega = \bigcap'_i \text{Div}(F_i)$, where F_1, \dots, F_r are the facets of P . This shows the importance of the ideals $\text{Div}(F_i)$ for our goals. We will use the following information about them.

(a) The ideals $\text{Div}(F_i)$ are precisely the height 1 prime ideals generated by monomials. Moreover, the minimal prime overideals of any height 1 monomial ideal are among the ideals $\text{Div}(F_i)$. Note that $\text{ht}_{F_i}(x)$ is precisely the discrete valuation of $\mathbb{Q}F(k[\bar{S}_P])$ that corresponds to $\text{Div}(F_i)$ evaluated at x .

(b) Any divisorial fractional ideal of $k[\bar{S}_P]$ is equivalent (i.e., defines the same element in the divisor class group $\text{Cl}(k[\bar{S}_P])$) to a divisorial monomial ideal contained in $k[\bar{S}_P]$.

(c) Every divisorial monomial ideal $I \subset k[\overline{S}_P]$ has a *unique* presentation of the type

$$I = \bigcap_1^r \text{Div}(F_i)^{(a_i)}, \quad a_i \geq 0$$

($\text{Div}(F)^{(j)}$ is the j th symbolic power of $\text{Div}(F)$).

For (a) see [BH, Chap. 6]; (b) is due to Chouinard [Ch]. (It is in fact an immediate consequence of Nagata's theorem on divisor class groups: if one inverts an "interior" monomial of \overline{S}_P , then the ring of quotients is the Laurent monomial ring $k[\mathbb{Z}^{n+1}]$; according to (a), $\text{Cl}(k[\overline{S}_P])$ is generated by the $\text{Div}(F_i)$.) Statement (c) is a consequence of (a) since in a normal domain every divisorial ideal is an intersection of symbolic powers of the divisorial prime ideals in which it is contained. (For all the facts about divisorial ideals that we will need in addition to (a) above we refer the reader to Fossum [Fo].)

Before we prove the claim above (reformulated as Lemma 4.4) we collect some auxiliary arguments.

LEMMA 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*

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

and

$$e_{v_1}^{\lambda_1} \circ \dots \circ e_{v_s}^{\lambda_s}(\text{Div}(G)) = \text{Div}(G), \quad G \neq F.$$

Proof. Using the automorphism ϑ from the proof of Lemma 3.1, we see immediately that

$$e_{v_1}^{\lambda_1} \circ \dots \circ e_{v_s}^{\lambda_s}(\text{Div}(F)) \subset (1 + \lambda_1 v_1 + \dots + \lambda_s v_s) \text{Div}(F).$$

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

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

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

LEMMA 4.3. *Let $F \subset P$ be a facet and let $\lambda_1, \dots, \lambda_s \in k \setminus \{0\}$ and $v_1, \dots, v_s \in \mathbb{Z}^{n+1}$ be pairwise different non-zero elements of degree 0. Suppose $(1 + \lambda_1 v_1 + \dots + \lambda_s v_s) \text{Div}(F) \subset k[\bar{S}_P]$. 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 $x \in \text{Div}(F)$. 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 4.4. *Let $\gamma \in \bar{\Gamma}_k(P)$, and enumerate the facets F_1, \dots, F_r of P in such a way that $\#(F_i \cap \mathbb{Z}^n) \leq \#(F_{i+1} \cap \mathbb{Z}^n)$ for $i \in [1, r - 1]$. Then there exists a permutation π of $[1, r]$ such that $\#(F_i \cap \mathbb{Z}^n) = \#(F_{\pi(i)} \cap \mathbb{Z}^n)$ for all i and*

$$\alpha_r \circ \dots \circ \alpha_1 \circ \gamma(\text{Div}(F_i)) = \text{Div}(F_{\pi(i)}),$$

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

In fact, this lemma finishes the proof of Theorem 3.2(a): the resulting automorphism $\delta = \alpha_r \circ \dots \circ \alpha_1 \circ \gamma$ permutes the minimal prime ideals of ω and therefore preserves their intersection ω . By virtue of Lemma 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 α_i by its inverse and each F_i by $F_{\pi(i)}$.

Proof of Lemma 4.4. As mentioned above, the divisorial ideal $\gamma(\text{Div}(F)) \subset k[\bar{S}_P]$ is equivalent to some monomial divisorial ideal Δ ; i.e., there is an element $\kappa \in \text{QF}(k[\bar{S}_P])$ such that

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

The inclusion $\kappa \in (\gamma(\text{Div}(F)) : \Delta)$ shows that κ 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 κ , say m , and rewrite the above equality as

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

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

$$(1 + m_1 + \dots + m_s)(m\Delta) \subset k[\bar{S}_P]$$

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

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

where m_1, \dots, m_s are Laurent terms of degree 0 and $\Delta \subset k[\bar{S}_P]$ 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

$$\left(\bigcap_1^r \text{Div}(F_i)^{(a_i)} \right)_P = \{x \in P \cap \mathbb{Z}^n \mid \text{ht}_{F_i}(x) \geq a_i, i \in [1, r]\}$$

for all $a_i \geq 0$. (Recall that ht_{F_i} coincides on lattice points with the valuation of $\text{QF}(k[S_P])$ defined by $\text{Div}(F_i)$.) Furthermore we set

$$c_i = \#(I_{F_i}).$$

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

For $\gamma \in \bar{\Gamma}_k(P)$ consider an admissible representation

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

Since γ 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 $a_i \geq 0$ such that

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

It follows easily that if $\sum_1^r a_i \geq 2$ 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, $\Delta = \text{Div}(G_1)$ for some $G_1 \in \{F_1, \dots, F_r\}$ containing precisely $\#(F_1 \cap \mathbb{Z}^n)$ lattice points. By Lemmas 4.2 and 4.3 there exists $\alpha_1 \in \mathbb{A}(G_1)$ such that

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

Now we proceed inductively. Let $1 \leq t < r$. 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

$$\alpha_t \circ \dots \circ \alpha_1 \circ \gamma(\text{Div}(F_i)) = \text{Div}(G_i), \quad i \in [1, t].$$

(Observe that the G_i are automatically different.) In view of Lemma 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

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

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

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

Rewriting this equality in the form

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

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

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

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

Assume to the contrary that in the primary decompositions of all the monomial ideals $m_j\Delta$ there only appear the prime ideals $\text{Div}(G_1), \dots, \text{Div}(G_t)$. 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 $\text{Cl}(k[\overline{S}_p])$. Applying $(\gamma')^{-1}$ we arrive at the conclusion that $\text{Div}(F_{t+1})$ is contained in a sum of monomial divisorial ideals Φ_0, \dots, Φ_s , such that the primary decomposition of each of them only involves $\text{Div}(F_1), \dots, \text{Div}(F_t)$. (This follows from the fact that $(\gamma')^{-1}$ maps $\text{Div}(G_i)$ to the monomial ideal $\text{Div}(F_i)$ for $i = 1, \dots, t$; thus intersections of symbolic powers of $\text{Div}(G_1), \dots, \text{Div}(G_t)$ are mapped to intersections of symbolic powers of $\text{Div}(F_1), \dots, \text{Div}(F_t)$, which are automatically monomial.) Furthermore, $\text{Div}(F_{t+1})$ has the same divisor class as each of the Φ_i .

Now choose a monomial $M \in \mathbb{Z}^{n+1} \cap \text{Div}(F_{t+1})$ such that $\text{ht}_{F_1}(M) + \dots + \text{ht}_{F_t}(M)$ is minimal. Since the monomial ideal $\text{Div}(F_{t+1})$ is contained in the sum of the monomial ideals Φ_0, \dots, Φ_s , each monomial in it must belong to one of the ideals Φ_j ; so $M \in \Phi_j$ for some j . There is a monomial d with $\text{Div}(F_{t+1}) = d\Phi_j$, owing to the fact that $\text{Div}(F_{t+1})$ and Φ_j belong to the same divisor class. It is clear that $\text{ht}_{F_i}(d) \leq 0$ for $i \in [1, t]$ and $\text{ht}_{F_i}(d) < 0$ for at least one $i \in [1, t]$. In fact,

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

where $\Phi_j = \bigcap_1^t \text{Div}(F_i)^{(a_i)}$. If we had $a_i = 0$ for $i = 1, \dots, t$, then $\Phi_j = k[\bar{S}_P]$, which is impossible since Φ_j has no non-zero elements in degree 0. By the choice of d the monomial $N = dM$ belongs to $\text{Div}(F_{t+1})$. But

$$\text{ht}_{F_1}(N) + \dots + \text{ht}_{F_t}(N) < \text{ht}_{F_1}(M) + \dots + \text{ht}_{F_t}(M),$$

a contradiction. ■

Proof of Theorem 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 $\bar{\Gamma}_k(P)$ (see Borel [Bo, Proposition 2.2]). This subgroup acts trivially on $\text{Cl}(k[\bar{S}_P])$ by Lemma 4.2 and the fact that the classes of the $\text{Div}(F_i)$ generate the divisor class group. Furthermore U has finite index in $\bar{\Gamma}_k(P)$ bounded by $\#\Sigma(P)$. Therefore $U = \bar{\Gamma}_k(P)^0$.

Assume $\gamma \in \bar{\Gamma}_k(P)$ acts trivially on $\text{Cl}(k[\bar{S}_P])$. We want to show that $\gamma \in U$. Let E denote the connected subgroup of $\bar{\Gamma}_k(P)$, generated by the elementary automorphisms. Since any automorphism that maps monomials to terms and preserves the divisorial ideals $\text{Div}(F_i)$ is automatically a toric automorphism, by Lemma 4.1(a) we only have to show that there is an element $\varepsilon \in E$, such that

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

By Lemma 4.4 we know that there is $\varepsilon_1 \in E$ such that

$$\varepsilon_1 \circ \gamma(\text{Div}(F_j)) = \text{Div}(F_j), \quad j \in [1, r], \quad (2)$$

where $\{i_1, \dots, i_r\} = \{1, \dots, r\}$. Since ε_1 and γ both act trivially on $\text{Cl}(k[\bar{S}_P])$, we get

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

for some monomials m_{i_j} of degree 0.

By Lemma 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

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

interchanges the ideals $\text{Div}(F_j)$ and $\text{Div}(F_{i_j})$, 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 $\bar{\Gamma}_k(P)$. Without loss of generality we may assume that k is algebraically closed, passing to the algebraic closure of k if necessary (see Remark 3.3(e)). For every permutation $\rho: \{1, \dots, r\} \rightarrow \{1, \dots, r\}$ we have the algebraic map

$$\mathbb{A}(F_{\rho(1)}) \times \dots \times \mathbb{A}(F_{\rho(r)}) \times \mathbb{T}_{n+1} \times \Sigma(P) \rightarrow \bar{\Gamma}_k(P),$$

induced by composition. The left-hand side has dimension $\#\text{Col}(P) + n + 1$. By Theorem 3.2(a) we are given a finite system of constructible sets, covering $\bar{\Gamma}_k(P)$. Hence $\dim \bar{\Gamma}_k(P) \leq \#\text{Col}(P) + n + 1$.

To derive the opposite inequality, we can additionally assume that P contains an interior lattice point. Indeed, the observation (*) in Section 2 and Theorem 3.2(a) show that the natural group homomorphism $\bar{\Gamma}_k(P) \rightarrow \bar{\Gamma}_k(cP)$, 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 $\#\text{Col}(P)$ linearly independent tangent vectors of $\bar{\Gamma}_k(P)$ at $1 \in \bar{\Gamma}_k(P)$. Since the tangent vectors corresponding to the elements of \mathbb{T}_{n+1} clearly belong to a complementary subspace and $\bar{\Gamma}_k(P)$ is a smooth variety, we are done.

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

$$\varepsilon = e_v^1 \circ e_{-v}^{-1} \circ e_v^1 \in \bar{\Gamma}_k(P)^0 \quad (= \Gamma_k(P)^0)$$

maps monomials to terms; more precisely, ε ‘‘inverts up to scalars’’ the columns parallel to v so that any $x \in \bar{S}_p$ is sent either to the appropriate $y \in \bar{S}_p$ or to $-y \in k[\bar{S}_p]$. 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 $\Sigma(P)_{\text{inv}} \subset \bar{\Gamma}_k(P)^0$.

Conversely, if $\sigma \in \Sigma(P) \cap \bar{\Gamma}_k(P)^0$ then σ induces the identity map on $\text{Cl}(k[S_p])$. Therefore $\sigma(\text{Div}(F_i)) = m_{i_j} \text{Div}(F_i)$ 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 $\bar{\Gamma}_k(P)/\bar{\Gamma}_k(P)^0 = \Sigma(P)/\Sigma(P)_{\text{inv}}$.

Finally, assume k is infinite and $\mathbb{T}' \subset \bar{\Gamma}_k(P)$ is a torus, strictly containing \mathbb{T}_{n+1} . Choose $x \in \bar{S}_P$ 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, γ maps monomials to terms. Then, as observed above Lemma 4.1, $\gamma = \sigma \circ \tau$ with $\tau \in \mathbb{T}_{n+1}$, and therefore $\sigma \in \Sigma_0 = \mathbb{T}' \cap \Sigma$. Lemma 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. PROJECTIVE TORIC VARIETIES AND THEIR GROUPS

Having determined the automorphism group of a polytopal semigroup ring, we show in this section 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. (See the Introduction for references to previous work on the automorphism groups of toric varieties.)

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

Let $P \subset \mathbb{R}^n$ be a polytope as above. Then $\text{Proj}(k[\bar{S}_P])$ is a projective toric variety (though $k[\bar{S}_P]$ need 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 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

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

The affine open subschemes $\text{Spec}(k[\mathbb{Z}^n \cap C(z)])$ cover $\text{Proj}(k[\bar{S}_P])$. The projectivity of $\text{Proj}(k[\bar{S}_P])$ follows from the observation that for all natural numbers $c \gg 0$ the polytope cP is normal (see [Oda], or [BGT] for a sharp bound on c) and, hence, $\text{Proj}(k[\bar{S}_P]) = \text{Proj}(k[S_{cP}])$.

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 the $\text{Proj}(k[\bar{S}_P]) = \text{Proj}(k[S_P])$ if and only if P is very ample. In particular, normal polytopes are very ample, but not conversely:

EXAMPLE 5.1. Let Π be the simplicial complex associated with the minimal triangulation of the real projective plane. It has six vertices which we label by the numbers $i \in [1, 6]$. Then the 10 facets of Π have the following vertex sets (written as ascending sequences):

$$(1, 2, 3), (1, 2, 4), (1, 3, 5), (1, 4, 6), (1, 5, 6), \\ (2, 3, 6), (2, 4, 5), (2, 5, 6), (3, 4, 5), (3, 4, 6).$$

Let P be the polytope spanned by the indicator vectors of the 10 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 $R = k[S_P]$ as the k -subalgebra of $k[X_1, \dots, X_6]$ generated by the 10 monomials $\mu_1 = X_1X_2X_3$, $\mu_2 = X_1X_2X_4$, \dots .

Let \bar{R} be the normalization of R . It can be checked by effective methods that \bar{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 ν^2 all lie in R . It follows that \bar{R}/R is a one-dimensional vector space; therefore $\text{Proj}(R) = \text{Proj}(\bar{R})$ is normal, and P is very ample.

For a very ample polytope P we have a projective embedding

$$\text{Proj}(k[S_P]) \subset \mathbb{P}_k^N, \quad N = \#(P \cap \mathbb{Z}^n) - 1.$$

The corresponding very ample line bundle on $\text{Proj}(k[S_P])$ 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 $\text{Proj}(k[S_P])$ 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 $\text{Proj}(k[S_P])$ (for a very ample polytope Q) then $\mathcal{N}(P) = \mathcal{N}(Q)$ (see [Oda, Chap. 2] or [Da]). Therefore P and Q have the same column vectors (see observations (***) in Section 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 $L_P, L_Q \in \text{Pic}(X)$ 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 interesting exactly one of the rays in $\mathcal{N}(P)$ (this is the condition $\varphi_G(v) \geq 0$ 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 De-mazure's roots [De].

In Fig. 3 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 Fig. 2).

LEMMA 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 Section 4 we will work with $\bar{\Gamma}_k(P_i)$. Put $X = \text{Proj}(k[\bar{S}_{P_1}] = \text{Proj}(k[\bar{S}_{P_2}])$ and consider the canonical anti-homomorphisms $\bar{\Gamma}_k(P_i)^0 \rightarrow \text{Aut}_k(X)$, $i = 1, 2$. 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 $e_v^\lambda(P_i) \in \bar{\Gamma}_k(P_i)^0$, $i = 1, 2$, have the same images in $\text{Aut}_k(X)$. 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 $\text{Spec}(k[C(z_1) \cap \mathbb{Z}^n]) \subset X$, 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 $\tau_i \in \bar{\Gamma}_k(P_i)$, $i = 1, 2$, have the same images in $\text{Aut}_k(X)$. By Theorem 3.2(b) we arrive at the equality $A(P_1) = A(P_2)$. It only remains to notice that $k^* = \text{Ker}(\bar{\Gamma}_k(P_i)^0 \rightarrow A(P_i))$, $i = 1, 2$. ■

EXAMPLE 5.3. Lemma 5.2 cannot be improved. For example, let P_1 be the unit 1-simplex Δ_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

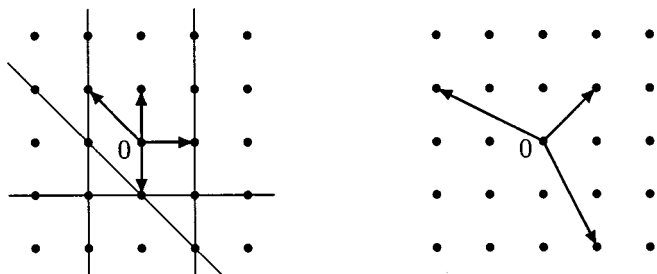


FIG. 3. The normal fans of the polytopes P_1 and P_2 .

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 $\Gamma_{\mathbb{C}}(P_1) \rightarrow \Gamma_{\mathbb{C}}(P_2)$ is surjective; in fact, $\Gamma_{\mathbb{C}}(P_1) = \text{GL}_2(\mathbb{C})$ and $\Gamma_{\mathbb{C}}(P_2) = \text{GL}_2(\mathbb{C})/\{\pm 1\}$. If there were an isomorphism between these groups, then $\text{SL}_2(\mathbb{C})$ and $\text{SL}_2(\mathbb{C})/\{\pm 1\}$ would also be isomorphic. This can be easily excluded by inspecting the list of finite subgroups of $\text{SL}_2(\mathbb{C})$.

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

THEOREM 5.4. *For a lattice n -polytope P the group $\text{Aut}_k(X(P))$ is generated by $A_k(P)$ and $\Sigma(\mathcal{N}(P))$. The connected component of unity of $\text{Aut}_k(X(P))$ is $A_k(P)$, $\dim(A_k(P)) = \#\text{Col}(P) + n$, and the embedded torus $\mathbb{T}_n = \mathbb{T}_{n+1}/k^*$ is a maximal torus of $\text{Aut}_k(X(P))$.*

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

We have $k[\bar{S}_P] = \bigoplus_{i \geq 0} H^0(X, L_P^i)$. Since $[L_P]$ is invariant under $\text{Aut}_k(X)$, 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 $\Gamma_k(P) \rightarrow \text{Aut}_k(X(P))$ is surjective. Now Theorem 3.2 gives the desired result once we notice that $\Sigma(P)$ is mapped to $\Sigma(\mathcal{F})$.

Therefore, and in view of Lemma 5.2, 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 $\text{Aut}_k(X)$.

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 $\text{Aut}_k(X(P))$, 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 $\{[L_{Q_1}], \dots, [L_{Q_r}]\}$ 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 $\alpha \in \text{Aut}_k(X(P))$ must permute the classes $[L_{Q_i}] \in \text{Pic}(X(P))$. In



FIGURE 4

particular, the element

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

is invariant under $\text{Aut}_k(X(P))$. 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. ■

EXAMPLE 5.5. In general the natural anti-homomorphism $\Gamma_k(P) \rightarrow \text{Aut}_k(X(P))$ is not surjective. For example consider the polytopes P and Q in Fig. 4. Then $\text{Proj}(k[S_P]) = \text{Proj}(k[S_Q]) = \mathbb{P}^1 \times \mathbb{P}^1$. However, the isomorphism corresponding to the exchange of the two factors \mathbb{P}^1 cannot be realized in $k[S_Q]$.

ACKNOWLEDGMENTS

This paper was written while the second author was an Alexander von Humboldt Research Fellow at Universität Osnabrück. The generous grant of the Alexander von Humboldt Foundation is gratefully acknowledged. The second author was also supported by CRDF Grant GM1-115. We thank Matei Toma for stimulating discussions and Klaus Warneke for his technical help in the preparation of this paper.

REFERENCES

- [Bo] A. Borel, "Linear Algebraic Groups," Springer-Verlag, Berlin/New York, 1991.
- [BG] W. Bruns and J. Gubeladze, Polyhedral algebras, toric arrangements, and their groups, preprint.
- [BGT] W. Bruns, J. Gubeladze, and N. V. Trung, Normal polytopes, triangulations, and Koszul algebras, *J. Reine Angew. Math.* **485** (1997), 123–160.
- [BH] W. Bruns and J. Herzog, "Cohen–Macaulay Rings," Cambridge Univ. Press, Cambridge, UK, 1993.
- [BV] W. Bruns and U. Vetter, "Determinantal Rings," Lecture Notes in Mathematics, Vol. 1327, Springer-Verlag, Berlin/New York, 1988.
- [Bu] D. Bühler, "Homogener Koordinatenring und Automorphismengruppe vollständiger torischer Varietäten," Diplomarbeit, Basel, 1996.
- [Ch] L. Chouinard, Krull semigroups and divisor class groups, *Canad. J. Math.* **33** (1981), 1459–1468.
- [Co] D. A. Cox, The homogeneous coordinate ring of a toric variety, *J. Algebraic 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.
- [Fo] R. M. Fossum, “The Divisor Class Group of a Krull Domain,” Springer-Verlag, Berlin/New York, 1993.
- [Fu] W. Fulton, “Introduction to Toric Varieties,” Princeton Univ. Press, Princeton, NJ, 1993.
- [GKZ] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, “Discriminants, Resultants, and Multidimensional Determinants,” Birkhäuser, Basel, 1994.
- [Ha] R. Hartshorne, “Algebraic Geometry,” Springer-Verlag, Berlin/New York, 1977.
- [Oda] T. Oda, “Convex Bodies and Algebraic Geometry” (An Introduction to the Theory of Toric Varieties), Springer-Verlag, Berlin/New York, 1988.
- [St] R. Stanley, Hilbert functions of graded algebras, *Adv. in Math.* **28** (1978), 57–83.
- [Te] B. Teissier, Variétés toriques et polytopes, *Sém. Bourbaki*, **33** (1980/1981), 71–84.