BERNSTEIN'S THEOREM

M. B. YOUNG

Abstract. We investigate some topics in the theory of polytopes in $\mathbb{R}^n$, which leads naturally to the introduction of Newton polytopes. These ideas are then used to prove Bernstein's Theorem, which gives an upper bound to the number of common solutions in $(\mathbb{C}^*)^n$ to a system of n Laurent polynomials in n variables. Moreover, in the case that the coefficeints of the polynomials are generic, the bound in fact gives the exact number of common solutions. A short discussion is included about the extensions of these results to common solutions in $\mathbb{C}^n$.

1. Introduction

In both pure and applied settings, it is often desirable to know the number of common solutions to a system of equations. In general, the answer cannot be found and in most cases, bounds on the number of solutions cannot even be found. When we restrict the class and number of equations to, say, a single polynomial in one variable, the Fundamental Theorem of Algebra tells us that we have the same number of roots as the degree of the polynomial, where we consider roots in $\mathbb{C}$ and count multiplicities. Such a restriction of systems is far too strong for a number of interesting situations. The question presents itself as to how one may estimate the number of solutions to more general systems. Bezout's Theorem ([1], [2]) states that the number of common solutions in $(\mathbb{C}^*)^n$ to the system $f_1 = \cdots = f_n = 0$ where $f \in \mathbb{C}[x_1, \ldots, x_n]$ is bounded above by $\prod_{i=1}^n deg(f_i)$. This bound, however, is often much larger than the actual number of common solutions. Bernstein's Theorem gives a much more effective bound, that in most cases, is not only a bound but instead gives precisely the number of common solutions.[1]

We begin by introducing the concept of a polytope, and proceed to derive some of its properties. A natural link to between polynomials and polytopes is encountered and forms the basis of the proof of Bernstein's Theorem. Following the proof of Bernstein's Theorem is a short discussion of some of the more technical details of the proof. Also inclused is a short discussion on the possibility of extending the results of the theorem to solutions in $\mathbb{C}^n$. An appendix contains a result required in the proof of Bernstein's Theorem that lies slightly outside the flow of the paper.

2. Polytopes in $\mathbb{R}^n$

Definition. *Let $A \subset \mathbb{R}^n$. A is **convex** if it contains the line connecting any two points in A. The **convex hull** of A is the smallest convex set containing A, and is denoted by* $\operatorname{conv}(A)$.

It can be shown that

$$\operatorname{conv}(A) = \left\{ \lambda_1 a_1 + \cdots + \lambda_m a_m | a_i \in A, \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

This provides a useful characterization of convex hulls and will be used below to show that different sets may generate the same convex hull.

Definition. *A polytope P is the convex hull of a finite set $A \in \mathbb{R}^n$. Furthermore, if $A \subset \mathbb{Z}^n$, then P is said to be a **lattice polytope**.*

[1] This bound is sometimes call the **BKK bound**, as it was first suggested by Anatolii Georgievich Kushnirenko and Askold Khovarskii.

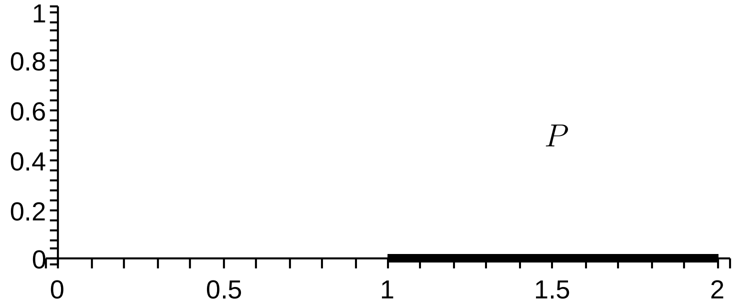

FIGURE 1. The Newton polytope P of $A = \{1, 2\} \subset \mathbb{R}$ is illustrated above.

If $P \subset \mathbb{R}^n$ is a polytope, then we define the **n−volume** of P to be its usual euclidean volume

$$\mathrm{Vol}_n(P) := \int \cdots \int_P dx_1 \cdots dx_n.$$

Example. *Let* $A = \{1, 2\} \subset \mathbb{R}$*. It is easy to see that* $P = \mathrm{conv}(A)$ *is the line segment with endpoints* 1 *and* 2*. If* $B = \{1, b, 2\} \subset \mathbb{R}$*, with* $b \in [1, 2]$*, then* $\mathrm{conv}(B) = P$*, because*

$$b = (2 - b) \cdot 1 + (-1 + b) \cdot 2$$

with $(2 - b) + (-1 + b) = 1$*. See Figure 1. Hence, different sets can produce the same convex hull. The* 1*-dimensional volume of* P *is*

$$\mathrm{Vol}_1(P) = \int_1^2 dx = \int_1^2 dx = 1.$$

Note that the two dimensional volume of P*, with* A *viewed as a subset of* $\mathbb{R}^2$*, vanishes:*

$$\mathrm{Vol}_2(P) = \int_P \int_0^0 dx_2 dx_1 = 0.$$

Example. *Let* $A = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 1, 1)\} \subset \mathbb{R}^3$*. Then,* $P = \mathrm{conv}(A)$ *is a triangular pyramid with its base in the* $z = 0$ *plane. See Figure 2. We may compute the volume of* P *by directly evaluating the required integral or by using the usual formula for the volume of a pyramid. Either way, we find*

$$\mathrm{Vol}_3(P) = \iiint_P dx_1 dx_2 dx_3 = \frac{1}{6}.$$

The natural geometry of polytopes leads to the definition of faces of a polytope. Intuitively, faces should be of dimension less than the polytope itself and should be a subset of the polytope. To

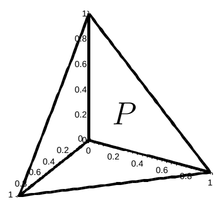

FIGURE 2. The Newton polytope P of $A = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 1, 1)\} \subset \mathbb{R}^3$ is illustrated above.

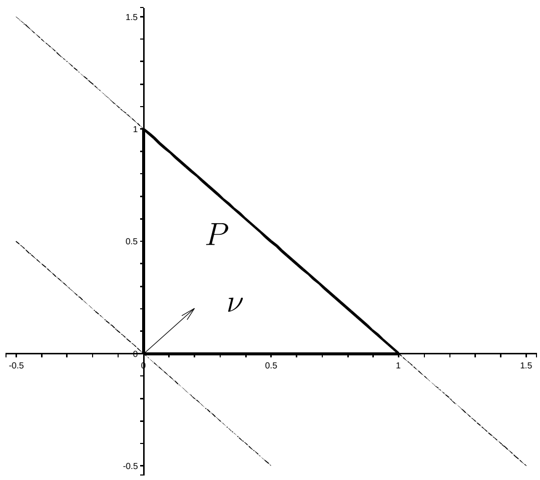

FIGURE 3. The Newton polytope P of $A = (\{0,0\}, \{1,0\}, \{0,1\}) \subset \mathbb{R}^2$ is illustrated above by thick lines. The thin lines, including the axes, are supporting hyperplane of P. For the hyperplane at the vertex $(0,0)$, we include the inward normal $\nu = (1,1)$. Note however that there are infintely many supporting hyperplanes at $(0,0)$.

precisely define a face of a polytope $P \subset \mathbb{R}^n$, consider a non-null vector $\nu \in \mathbb{R}^n$, and define the **supporting hyperplane** of P to be the set of $m \in \mathbb{R}^n$ satisfying

$$m \cdot \nu = -a_P(\nu)$$

where

$$a_P(\nu) := -\min_{m \in P} m \cdot \nu.$$

We call ν the **inward pointing normal**. Define the **face of P determined by** ν to be the set

$$P_\nu := P \cap \{m \in \mathbb{R}^n | m \cdot \nu = -a_P(\nu)\}.$$

Note that P_ν is necessarily non-empty. To see this, observe that P is n-dimensional, while the equation $m \cdot \nu = -a_P(\nu)$ defines an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. Since P is compact and the mapping $m \mapsto m \cdot \nu$ is continuous, $\min_{m \in P}(m \cdot \nu)$, or equivalently $a_P(\nu)$, exists. Thus, P_ν being the intersection of an n-dimensional compact set with an $(n-1)$-dimensional subspace of $\mathbb{R}^n$ must be non-empty.

For a polytope $P \subset \mathbb{R}^n$ of dimension n, we call the $n-1$ dimensional faces of P **facets**. In fact, P can be characterized entirely by its facets. Let $\{\mathcal{F}_i\}_{i=1}^N$ be the facets of P with corresponding inward pointing normal vectors $\{\nu_i\}_{i=1}^N$. These vectors are unique up to a scalar factor. Then, we can write

$$P = \{m \in \mathbb{R}^n | m \cdot \nu_i \geq -a_P(\nu_i), 1 \leq i \leq N\}. \tag{1}$$

Example. *Consider the set $A = (\{0,0\}, \{1,0\}, \{0,1\}) \subset \mathbb{R}^2$ and the associated polytope $P =$ conv(A). See Figure 3. We can see that P has three facets with inward pointing normals* $(0,1), (1,0)$ *and* $(-1,-1)$*. P also has three verticies, one at each point of A. Note there are infinitely many hyperplanes that support each vertex.*

We would like to relate the concept of polytopes to polynomials, or more generally to Laurent polynomials. There will be a strong connection between polytopes and the solutions of a system of Laurent polynomials.

Definition. *Let $\alpha \in \mathbb{Z}^n$ and k be a field. A **Laurent monomial** in the variables $x_1, \ldots, x_n$ is*
$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$
*A **Laurent polynomial** is a finite sum of Laurent monomials;*
$$f = \sum_{\alpha \in A \subset \mathbb{Z}^n} c_\alpha x^\alpha, \quad c_\alpha \in k.$$
The ring of Laurent polynomials with coeffecients in k is denoted by $k[x_1^\pm, \ldots, x_n^\pm]$.

Given a finite set $A = \{\alpha_1, \ldots, \alpha_m\} \subset \mathbb{Z}^n$, it will be useful below to introduce the notation $L(A)$ for the set of Laurent polynomials corresponding to A,
$$L(A) := \{c_1 x^{\alpha_1} + \cdots + c_m x^{\alpha_m} | c_i \in k\}.$$
The above discussions, we see that a Laurent polynomials has a naturally associated polytope.

Definition. *Let $f = \sum_{\alpha \in \mathbb{Z}^n} c_\alpha x^\alpha \in k[x_1^\pm, \ldots, x_n^\pm]$. The **Newton polytope** of f is the polytope*
$$NP(f) := \operatorname{conv}(\{\alpha \in \mathbb{Z}^n | c_\alpha \neq 0\}).$$

We see that there is a unique Newton polytope corresponding to each set $L(A)$. In particular, the Newton polytope does not distinguish between Laurent polynomials with the same exponents but different (non-zero) coefficients.

Remark 2.1. *Note that $NP(f)$ is necessarily a lattice polytope.*

For a Newton polytope $P \subset \mathbb{R}^n$, it can be shown that the inward pointing normal $\nu_\mathcal{F}$ of a facet $\mathcal{F}$ of P may be rescaled such that $0 \neq \nu_\mathcal{F} \in \mathbb{Z}^n$ with the entries of $\nu_\mathcal{F}$ pairwise relatively prime. In this case, $\nu_\mathcal{F}$ is said to be **primitive**; we will assume that all such inward pointing normal vectors have been rescaled to be primitive.

We have seen that to each Laurent polynomial there corresponds a Newton polytope. Below we introduce a tool to study the Newton polytope of a product of Laurent polynomials.

Definition. *Let $P, Q \subset \mathbb{R}^n$ be polytopes, and let $\lambda \in \mathbb{R}$, $\lambda \geq 0$. Then,*
1. *$P + Q := \{p + q | p \in P, q \in Q\}$,*
2. *$\lambda P := \{\lambda p | p \in P\}$.*

*$P + Q$ is called the **Minkowski sum** of P and Q.*

Remark 2.2. *The Minkowski sum of two (Newton) polytopes is again a (Newton) polytope. By induction, this extends to finite Minkowski sums of polytopes as well.*

Example. *Let $f = a_{30}x^3 + a_{22}x^2y^2 + a_{12}xy^2 + a_{00}$. Then $NP(f)$ is the trapezoid with vertices at $(0,0), (1,2), (2,2)$ and $(3,0)$ and has a $Vol_2(NP(f)) = 5$. If $g = b_{11}xy + b_{02}y^2 + b_{04}y^4$, then $NP(g)$ is the triangle with vertices at $(1,1), (0,2)$ and $(0,4)$ and has $Vol_2(NP(g)) = 1$. See Figure 4. Continuing, we compute the Minkowski sum $NP(f) + NP(g)$ to be pentagon with vertices $(1,1), (5,1), (2,6), (1,6)$ and $(0,4)$. A computation shows that $Vol_2(NP(f) + NP(g)) = 14$.*

When the polytopes in question are Newton polytopes, corresponding to $f_1, \ldots, f_m \in k[x_1^\pm, \ldots, x_n^\pm]$, it is simple to check that
$$NP(f_1) + \cdots + NP(f_m) = NP(f_1 \cdots f_m).$$
That is, the geometric operation of Minkowski summation corresponds to the algebraic operation of polynomial multiplication.

With the introduction of Minkowski sums, we may prove a result central to the proof of Bernstein's Theorem. Note that this result is not restricted to Newton polytopes.

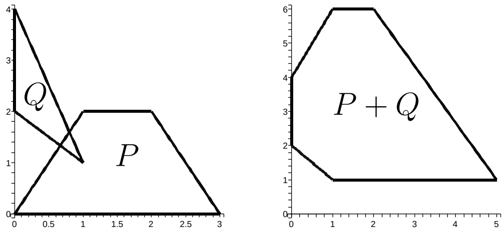

FIGURE 4. The plot to the left shows $P = NP(a_{30}x^3 + a_{22}x^2y^2 + a_{12}xy^2 + a_{00})$ and $Q = NP(b_{11}xy + b_{02}y^2 + b_{04}y^4)$. To the right, $P + Q$ is shown.

Proposition 2.1. *Let $P_1, \ldots, P_r \subset \mathbb{R}^n$ be polytopes and define $P = P_1 + \cdots + P_r$. Then every face P' of P can be expressed as a Minkowski sum of faces of $P_1, \ldots, P_r$:*

$$P' = P_1' + \cdots + P_r'.$$

Proof. There exists a normal vector $0 \neq \nu \in \mathbb{R}^n$ such that

$$P' = P_\nu = P \cap \{m \in \mathbb{R}^n | m \cdot \nu = -a_P(\nu)\}.$$

We claim

$$P_\nu = (P_1 + \cdots + P_r)_\nu = (P_1)_\nu + \cdots + (P_r)_\nu.$$

For $r = 2$, we have

$$(P_1 + P_2)_\nu = \{m \in \mathbb{R}^n | m \cdot \nu = a_{P_1+P_2}(\nu)\}.$$

Now,

$$a_{P_1+P_2}(\nu) = -\min_{m \in P_1+P_2} m \cdot \nu = -\min_{m \in P_1} m \cdot \nu - \min_{m' \in P_2} m' \cdot \nu = a_{P_1}(\nu) + a_{P_2}(\nu).$$

From this and the definition of the Minkowski sum we see that

$$(P_1 + P_2)_\nu = (P_1)_\nu + (P_2)_\nu.$$

The case for general r follows by induction. $\square$

Example. *Consider in $\mathbb{R}^2$ the square $P = conv(\{(0, 0, (1, 0), (1, 1), (0, 1)\})$ and the triangle $Q = conv(\{(0, 0), (0, 1), (0, 1)\})$. The facet $\mathcal{F}_1 = conv(\{(0, 0), (1, 0)\})$ of $P + Q$ is also a facet of P, and hence Proposition 2.1 holds trivially. Consider now the facet $\mathcal{F}_2 = conv(\{(0, 2), (1, 1)\})$ of $P + Q$. We may write $\mathcal{F}_2 = conv(\{(1, 0), (1, 1)\}) + conv(\{(0, 1), (1, 0)\})$, where the first term on the right is a facet of P and the second is a facet of Q, again showing Proposition 2.1 holds.*

With the above proposition in mind, it is natural to ask whether the volume of a polytope $P \subset \mathbb{R}^n$ can be expressed in terms of the volumes of its facets, $\mathcal{F}$. The answer is in fact yes, but we must first introduce some new notation before this can be done.

Using Eq.(1), we may write P in terms of its facets $\mathcal{F}$ as

$$P = \bigcap_{\mathcal{F}} \{m \in \mathbb{R}^n | m \cdot \nu_{\mathcal{F}} \geq -a_P(\nu_{\mathcal{F}})\}, \tag{2}$$

where the intersection is over all facets $\mathcal{F}$ of P. By $\nu_{\mathcal{F}}^\perp$ we denote the $(n-1)$-dimensional subspace defined by the affine hyperplane $m \cdot \nu_{\mathcal{F}} = 0$. It can be shown that there exists unique vectors $w_1, \ldots, w_{n-1} \in \nu_{\mathcal{F}}^\perp \cap \mathbb{Z}^n$ such that any element in $\nu_{\mathcal{F}}^\perp \cap \mathbb{Z}^n$ can be written as a linear combination of $\{w_i\}_{i=1}^{n-1}$ with integer coefficients. The set

$$\mathcal{P} = \{\lambda_1 w_1 + \cdots + \lambda_{n-1} w_{n-1} | 0 \leq \lambda_i \leq 1, 1 \leq i \leq n-1\}$$

is call the **fundamental lattice parallelotope** of $\nu_{\mathcal{F}}^{\perp} \cap \mathbb{Z}^n$. Using this, we define the **normalized volume** of a facet $\mathcal{F}$ of P to be

$$\mathrm{Vol}'_{n-1}(\mathcal{F}) := \frac{\mathrm{Vol}_{n-1}(\mathcal{F})}{\mathrm{Vol}_{n-1}(\mathcal{P})}.$$

The normalized volume is the usual volume rescaled so that the normalized volume of the fundamental lattice parallelotope is unity. We will use below the result that $\mathrm{Vol}_{n-1}(\mathcal{P}) = \|\nu_{\mathcal{F}}\|$.

With the above definitions, we state a proposition relating the volume of P in terms of the normalized volumes of its facets. For a proof and further discussion see [3] and references therein.

Proposition 2.2. *Let $P \subset \mathbb{R}^n$ be a lattice polytope. If P is written as in Eq.(2), then*

$$\mathrm{Vol}_n(P) = \frac{1}{n} \sum_{\mathcal{F}} a_P(\nu_{\mathcal{F}}) \mathrm{Vol}'_{n-1}(\mathcal{F}),$$

the sum running over all facets $\mathcal{F}$ of P.

Example. *We claim that if P is the n-cube, $P = conv(\{(1,0,\dots,0),\dots,(0,\dots,0,1)\})$, then $Vol_n(P) = 1$. We will confirm this commonly known result with the above proposition. In the case $n = 1$, the 1-cube is defined by $m \cdot \nu_1 \geq 0$ and $m \cdot \nu_2 \geq -1$ where $1 = \nu_1 = -\nu_2$. Hence $a_P(\nu_{\mathcal{F}_1}) = 0$ and $a_P(\nu_{\mathcal{F}_2}) = 1$. We compute, using the proposition,*

$$\mathrm{Vol}_n(P) = \frac{1}{1}(0 \cdot 1 + 1 \cdot 1) = 1.$$

Assume the claim is true for $n-1$. A n-cube is defined similarly to the 1-cube:

$$m \cdot \nu_1 \geq 0, m \cdot \nu_2 \geq -1, \dots, m \cdot \nu_{2n-1} \geq 0, m \cdot \nu_{2n} \geq -1$$

with

$$e_1 = \nu_1 = -\nu_2, \dots, e_n = \nu_{2n-1} = -\nu_{2n}$$

where e_i is the vector with 1 in the i^{th} slot and zero in all others. By the inductive hypothesis, and since the facets of an n-cube are $n-1$-cubes, we see $\mathrm{Vol}'_{n-1}(\mathcal{F}) = 1$. Since n of the $a_P(\nu_{\mathcal{F}})$ are unity while n vanish, we see that

$$Vol_n(P) = \frac{1}{n}(0 \cdot 1 + 1 \cdot 1 + \cdots + 0 \cdot 1 + 1 \cdot 1) = \frac{1}{n} n = 1$$

as desired.

Before proceeding to Bernstein's Theorem, we must introduce one more idea related to polytopes and their volumes.

Definition. *The **n-dimensional mixed volume** of the polytopes $P_1, \dots, P_n \subset \mathbb{R}^n$, written $MV_n(P_1, \dots, P_n)$, is the coefficient of the monomial $\lambda_1 \cdots \lambda_n$ in $\mathrm{Vol}_n(\lambda_1 P_1 + \cdots + \lambda_n P_n)$.*

Intrinsic to this definition is the fact that $\mathrm{Vol}_n(\lambda_1 P_1 + \cdots + \lambda_n P_n)$ is indeed a polynomial; in fact, it is a homogeneous polynomal of degree n in the λ_i. See [3].

Mixed volumes are central to the statement and proof of Bernstein's Theorem. Thus, it is useful to have some methods for computing the mixed volume of a set of polytopes. We state two methods here without proof. The first will play a fundamental rôle in the proof of Bernstein's Theorem, while the second provides a computational method for low dimensional cases.[2] For more properties of the mixed volume and references to their proofs, see [3].

[2]By no means are Propositions 2.3 and 2.4 the most efficient methods to compute mixed volumes. A more refined method uses a **mixed subdivision** of Minkowski sums.

Proposition 2.3. *Let $P_1, \ldots, P_n \subset \mathbb{R}^n$ be any collection of polytopes. Then,*

$$MV_n(P_1, \ldots, P_n) = \sum_{\nu} a_{P_1}(\nu) MV'_{n-1}\left((P_2)_\nu, \ldots, (P_2)_\nu\right),$$

where the sum is over all primitive vectors ν such that $(P_i)_\nu$ has dimension at least 1 for $i \geq 2$.

Proposition 2.4. *Let $P_1, \ldots, P_n \subset \mathbb{R}^n$ be any collection of polytopes. Then,*

$$MV_n(P_1, \ldots, P_n) = \sum_{k=1}^{n} (-1)^{n-k} \sum_{\substack{|I| \;=k \\ I \subset \;\{1,\ldots,n\}}} \mathrm{Vol}_n\left(\sum_{i \in I} P_i\right) \tag{3}$$

Example. *Again, let $f = a_{30}x^3 + a_{22}x^2y^2 + a_{12}xy^2 + a_{00}$ and $g = b_{11}xy + b_{02}y^2 + b_{04}y^4$. Using Proposition 2.4, we have*

$$MV_2(NP(f), NP(g)) = -\mathrm{Vol}_2(NP(f)) - \mathrm{Vol}_2(NP(g)) + \mathrm{Vol}_2(NP(fg)).$$

Using the work done in the examples above, we find $MV_2(NP(f), NP(g)) = -5 - 1 + 14 = 8$.

3. Bernstein's Theorem

As stated above, Bernstein's Theorem provides the exact number of common solutions to a system of equations provided the polynomials a sufficiently *generic*. We provide a precise defintion here.

Definition. *Let $A_1, \ldots, A_m \subset \mathbb{Z}^n$ be finite sets. A property is said to hold* ***generically*** *for Laurent polynomials $(f_1, \ldots, f_m) \in L(A_1) \times \cdots \times L(A_m)$ if there exists a non-zero polynomial in the coefficients of the f_i such that the property holds for all $f_1, \ldots, f_m$ for which the polynomial is non-vanishing.*

We are now in a position to prove Bernstein's Theorem. The proof given here follows that given in [3]. We will not prove Bernstein's Theorem in its full generality as we omit parts of the proof dealing with the genericity of the system in question. In practice it may often be assumed that a system of polynomials are generic. Only for sufficiently rare cases is the system not generic. However, following the proof we will see that systems that are not generic do exist.

Theorem 3.1 (Bernstein, 1975)**.** *Let $f_1, \ldots, f_n \in \mathbb{C}[x_1^{\pm}, \ldots, x_n^{\pm}]$ with finitely many common zeros in $(\mathbb{C}^*)^n$ and let $P_i = NP(f_i) \subset \mathbb{R}^n$. Then the number of common zeros of the f_i in $(\mathbb{C}^*)^n$ is bounded above by the mixed volume $MV_n(P_1, \ldots, P_n)$. Moreover, for generic choices of the coefficients in the f_i, the number of solutions is exactly $MV_n(P_1, \ldots, P_n)$.*

Proof. We proceed by induction on the size of the system, n. For $n = 1$, let $f = \sum_{\alpha \in \mathbb{Z}} c_\alpha x^\alpha$ be a Laurent polynomial in x. Let

$$\hat{\alpha} = \min_{\alpha} \{0\} \cup \{-\alpha | c_\alpha \neq 0\}.$$

Then, $\hat{f} := x^{\hat{\alpha}} f$ is a polynomial in x. Note that f and $\hat{f}$ have the same number of solutions in $\mathbb{C}^*$. Furthermore, by the choice of $\hat{\alpha}$,

$$\hat{f} = c_d x^d + \cdots + c_0$$

for some $d \in \mathbb{N}$. By the Fundamental Theorem of Algebra, $\hat{f}$ has d roots in $\mathbb{C}^*$, counting multiplicities. Observe that by the definition of $\hat{f}$, $NP(\hat{f})$ is a translation of $NP(f)$ by $\hat{\alpha}$ in the x direction, and hence $\mathrm{Vol}_1(NP(f)) = \mathrm{Vol}_1(NP(\hat{f}))$. From the defintion of volume,

$$\mathrm{Vol}_1(NP(\hat{f})) = \int_{NP(\hat{f})} dx = \int_0^d dx = d.$$

This gives

$$MV_1(NP(f)) = \mathrm{Vol}_1(NP(f)) = \mathrm{Vol}_1(NP(\hat{f})) = d,$$

where we used Proposition 2.4. This proves the base case.

Now, assume the theorem is true for $n = k$. Let $P = P_1 + \cdots + P_{k+1}$ and let the inward pointing normal to a facet P_ν of P be ν. By Proposition 2.1 we can decompose P_ν as

$$P_\nu = (P_1)_\nu + \cdots + (P_{k+1})_\nu$$

where $(P_i)_\nu$ is the face of P_i determined by ν. For all monomials x^α of f_i, we have

$$a \cdot \nu \geq -a_{P_i}(\nu),$$

with equality holding if an only if $\alpha \in (P_i)_\nu$. Decompose f_i as

$$f_i(x_1, \ldots, x_{k+1}) = \sum_{\{\alpha | \alpha \cdot \nu = -a_{P_i}(\nu)\}} c_{i,\alpha} x^\alpha + \sum_{\{\alpha | \alpha \cdot \nu > -a_{P_i}(\nu)\}} c_{i,\alpha} x^\alpha. \tag{4}$$

Observe that we may assume $a_{P_1}(\nu) \geq 0$. To see this, multiply f_1 by $x^{-\alpha}$ for some $\alpha \in P_1$ so that f_1 has a non-zero constant term. Thus $0 \in P_1$ which implies $a_{P_1} \geq 0$ as desired. As discussed above, such a multiplication does not affect the common solutions of $f_1 = \cdots = f_{k+1} = 0$ in $(\mathbb{C}^*)^{k+1}$. Neither does this affect the mixed volume of the Newton polytopes as such a multiplication corresponds only to a translation in P_1.

Let $B = (b_{ij})$ be an integer matrix such that $b_{1j} = \nu_j$, ν_j being the j^{th} coordinate of ν. We may also take B to have an inverse that is also an integer matrix; for a proof that such a matrix B exists see the appendix. Consider the change of coordinates

$$x_j \mapsto \prod_{i=1}^{k+1} y_i^{-b_{ij}}.$$

Appying this transformation to f_i, the monomials in the first sum in Eq. (4) become

$$y^{-B\alpha} = y_1^{a_{P_i}(\nu)} y_2^{\beta_2} \cdots y_{k+1}^{\beta_{k+1}}$$

where $\beta_2, \ldots, \beta_{k+1}$ are integers dependent on the entries of B. Monomials in the second sum in Eq.(4) become

$$y^{-B\alpha} = y_1^{\beta_1} \cdots y_{k+1}^{\beta_{k+1}}, \quad \beta_1 < a_{P_i}(\nu).$$

with $\beta_1 < a_{P_i}(\nu)$, which follows by the restion of the summation in the first term of Eq.(4). Hence, Eq.(4) in the new coordinates is

$$f_i(y_1, \ldots, y_{k+1}) = g_{i\nu}(y_2, \ldots, y_{k+1}) y_1^{a_{P_i}(\nu)} + \sum_{j < a_{P_i}(\nu)} g_{ij\nu}(y_2, \ldots, y_{k+1}) y_1^j.$$

The original system of equations is thus transformed into the new system

$$0 = g_{i\nu}(y_2, \ldots, y_{k+1}) y_1^{a_{P_i}(\nu)} + \sum_{j < a_{P_i}(\nu)} g_{ij\nu}(y_2, \ldots, y_{k+1}) y_1^j, \quad 1 \leq i \leq k+1. \tag{5}$$

By the form of the coordinate change, it is clear that the Newton polytope of $g_{i\nu}$ is the image of face $(P_i)_\nu$ under B, which we denote by $B\,(P_i)_\nu$

We are now in a position to deform the equations in such a way that we may apply the inductive hypothesis. To deform the equations, substitute

$$c_1 \mapsto \frac{c_1}{t^{a_{P_1}(\nu)}} \text{ and } y_1 \mapsto \frac{y_1}{t}$$

for some new variable t, where c_1 is the constant term of f_1.

Proceed to multiply the i^{th} equation of Eq.(5) by $t^{a_{P_i}(\nu)}$. Consider first the case in which $a_{P_1}(\nu) > 0$. Then, Eq.(5), now deformed, is of the form

$$(6)\qquad \begin{aligned} 0 &= g_{1\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_1}(\nu)} + c_1 + O(t) \\ 0 &= g_{i\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_i}(\nu)} + O(t), \quad 2 \le i \le k+1. \end{aligned}$$

The notation $O(t)$ above denotes a sum of terms divisible by t. In the particular case that $t = 1$, the deformation is trivial in that Eq.(6) reduces to Eq.(5).

If we let $t \to 0$ along a general path in $\mathbb{C}$, Eq.(6) becomes

$$(7)\qquad \begin{aligned} 0 &= g_{1\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_1}(\nu)} + c_1 \\ 0 &= g_{i\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_i}(\nu)}, \quad 2 \le i \le k+1. \end{aligned}$$

Restricting to solutions in $(\mathbb{C}^*)^{k+1}$, it suffices to consider solutions in $(\mathbb{C}^*)^{k+1}$ to the system

$$(8)\qquad \begin{aligned} 0 &= g_{1\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_1}(\nu)} + c_1 \\ 0 &= g_{i\nu}(y_2,\ldots,y_{k+1}), \quad 2 \le i \le k+1. \end{aligned}$$

Now, for a sufficiently generic set of original equations $f_1 = \cdots = f_{k+1} = 0$, the equations $g_{2\nu} = \cdots = g_{(k+1)\nu} = 0$ are generic with respect to the transformed polytopes $B\left(P_2\right)_\nu, \ldots, B\left(P_{k+1}\right)_\nu$. Applying the inductive hypothesis to the k equations corresponding to $i = 2, \ldots, k+1$ in Eq.(8), we see that there are

$$MV_k\left(B\left(P_2\right)_\nu, \ldots, B\left(P_{k+1}\right)_\nu\right)$$

solutions $(y_2, \ldots, y_{k+1}) \in (\mathbb{C}^*)^k$. But,

$$MV_k\left(B\left(P_2\right)_\nu, \ldots, B\left(P_{k+1}\right)_\nu\right) = MV_k'\left(\left(P_2\right)_\nu, \ldots, \left(P_{k+1}\right)_\nu\right)$$

where the equality follows from proposition whose proof we omit here.

For each solution $(y_2, \ldots, y_{k+1})$ of the last k equations of Eq.(8), there corresponds $a_{P_1(\nu)}$ values of $y_1 \in \mathbb{C}^*$, such that the first equation in Eq.(8) holds, provided that $g_{1\nu}(y_2, \ldots, y_{k+1}) \neq 0$ and $c_1 \neq 0$, which can be shown to hold generically, something we omit here. Thus, there are

$$(9)\qquad a_{P_1(\nu)} MV_k'\left(\left(P_2\right)_\nu, \ldots, \left(P_{k+1}\right)_\nu\right)$$

solutions to the system in question.

Given solutions $(y_1, \ldots, y_{k+1})$ of Eq.(8), we claim that we can find parameterized solutions $(y_1(t), \ldots, y_{k+1}(t))$ of Eq.(6) such that $(y_1(0), \ldots, y_{k+1}(0)) = (y_1, \ldots, y_{k+1})$. Then, $(y_1(1), \ldots, y_{k+1}(1))$ is a solution of Eq.(5). Using B^{-1}, $(y_1(1), \ldots, y_{k+1}(1))$ corresponds to a solution $(x_1, \ldots, x_{k+1}) \in (\mathbb{C}^*)^{k+1}$ of the original system; the fact that B^{-1} is an integer matrix ensures this correspondence is injective. Hence, there are Eq.(9) solutions to the original system of equations for each ν. Summing over all inward pointing normal vectors ν corresponding to the facets of P, we see that there are

$$\sum_\nu a_{P_1(\nu)} MV_k'\left(\left(P_2\right)_\nu, \ldots, \left(P_{k+1}\right)_\nu\right) = MV_{k+1}\left(P_1, \ldots, P_{k+1}\right)$$

solutions, with the equality following from Proposition 2.3. This completes the proof for the case in which $a_{P_1}(\nu) > 0$.

Now, consider the case in which $a_{P_1}(\nu) = 0$. Then, instead of Eq.(6), we get

$$\begin{aligned} 0 &= g_{1\nu}(y_2,\ldots,y_{k+1}) + O(t) \\ 0 &= g_{i\nu}(y_2,\ldots,y_{k+1})y_1^{a_{P_i}(\nu)} + O(t), \quad 2 \le i \le k+1. \end{aligned}$$

The constant term c_1 is now included in $g_{1\nu}$. This follows because $a_{P_1}(\nu) = 0$ implies that c_1 lies in the first sum of Eq.(4). Again, taking $t \to 0$ along a general path in $\mathbb{C}$, we get

$$0 = g_{i\nu}(y_2, \ldots, y_{k+1})y_1^{a_{P_i}}, \quad 1 \le i \le k+1.$$

Considering only solutions in $(\mathbb{C}^*)^{k+1}$, it suffices to consider the non-trivial solutions to the system

$$0 = g_{i\nu}(y_2, \ldots, y_{k+1}), \quad 1 \le i \le k+1.$$

It can be shown that $g_{1\nu}(y_2, \ldots, y_{k+1})$ is generic and thus doesn not vanish at the solutions $(y_2, \ldots, y_{k+1})$ of $g_{2\nu} = \cdots = g_{(k+1)\nu} = 0$. Hence, the $t \to 0$ limit of the deformed system has no solutions. This agrees with Eq.(9), and hence the theorem holds for the case in which $a_{P_1}(\nu) = 0$ as well. This completes the proof. $\square$

Example. *Consider again $f = a_{30}x^3 + a_{22}x^2y^2 + a_{12}xy^2 + a_{00}$ and $g = b_{11}xy + b_{02}y^2 + b_{04}y^4$. Then, for any chocies of $a_{00}, a_{12}, \ldots, b_{11}$, the number of common solutions of $f = g = 0$ in $(\mathbb{C}^*)^2$ is bounded above by $MV_2(NP(f), NP(g))$, which equals 8 by an example given above. Moreover, in the case that the $a_{00}, a_{12}, \ldots, b_{11}$ are generic, the number of common solutions in $(\mathbb{C}^*)^2$ is exactly 8.*

Example. *Consider the case in which we have the trivial system $f_1(x) = 0$ and wish to solve $f_1 = 0$. Clearly, $f_1(x) = 0$ for any $x \in \mathbb{C}^*$. It is simple to check that $MV_1(f_1) = 0$. Hence, the BKK bounds is zero, while there are infinitely many solutions to the system in $\mathbb{C}^*$. This illustrates that the requirement in Berstein's Theorem that the system have only finitely many solutions is necessary.*

A slightly less trivial example of the necessity of this requirement is given by the polynomials $f_1(x, y) = x$ and $f_2(x, y) = xy$. Then, for any $y \in \mathbb{C}^*$, $(0, y)$ is a solution of the system $f_1 = f_2 = 0$ in $(\mathbb{C}^*)^2$. Hence, again there are infinitely many solutions while $MV_2(f_1, f_2)$ vanishes.

Example. *Consider the case in which $f_1(x) = (x-1)^m$ for $m \in \mathbb{Z}^*$. Clearly, the system has only one solution in $\mathbb{C}^*$, namely $x = 1$. We compute*

$$MV_1(f_1) = |\int_0^m dx| = m$$

where the absolute value is required as we may have $m < 0$. Hence, for $|m| > 1$, the system has strictly less solutions than the BKK bound. This is an example of a polynomial that is not generic, which explains why the BKK bound is not acheived.

Example. *The above example hints at the following. In the $n = 1$ case, a polynomial of degree two, $f = ax^2 + bx + c$ is generic if its coefficients satisfy the requirement*

$$b^2 - 4ac \neq 0;$$

that is the disciminant of f is non-zero. Then, Bernstein's Theorem guarentees that $f_1 = 0$ has two solutions in $\mathbb{C}^*$ in the case that $a \neq 0$ and one in the case that $a = 0$. It the discriminant of f does vanish, then the BKK bound will in fact be a strict bound.

For a general polynomial $f(x)$ of degree d, we see that f is generic if its discriminant vanishes. To see this, note that by the Fundamental Theorem of Algebra, f has d roots in $\mathbb{C}$ counting multiplicities. If the discriminant of f vanishes, then f has a root of multiplicity two. Hence, f cannot have d distinct solutions in $\mathbb{C}^*$, as would be predicted by Bernstein's Theorem, assuming the constant term of f is non-zero.

4. Discussion and the Restriction to Solutions in $(\mathbb{C}^*)^n$

We have seen how the mixed volume of Newton polytopes relates the number of common solutions in $(\mathbb{C}^*)^n$. The proof given above is not that originally given by Bernstein. In his original proof [4], Bernstein instead uses Puiseux series which are beyond the scope of this paper.

We have side stepped a number of problems in our proof of Bernstein's Theorem. The most obvious is the ommision of the part of the proof that shows that the number of solutions in the original system is *generically* equal to the mixed volume in question. A second, more subtle omission is how one actually parameterizes the solutions $(y_1(t), \ldots, y_n(t))$. We should note that $y_i(t)$ is not a polynomial in t, which makes the task of parameterization more difficult. Such a parameterization is done via *homotopy continuation methods* and can in certain cases be reduced to solving an initial value problem in the theory of ordinary differential equations. We are given the initial condition $(y_1(0), \ldots, y_n(0))$, and thus we may use numerical techniques to find approximate values of $(y_1(1), \ldots, y_n(1))$, which solve the original system in question. A more detailed heuristic discussion is given in [3], which also contains references to more rigorous treatments of this method.

One may wonder whether we may extend the above work to find common solutions in $\mathbb{C}^n$. However, this in general cannot be done. The proof of Bernstein's Theorem given above fails in the $n = 1$ case, as multiplying f by $x^{\hat{\alpha}}$ may change the solutions in $\mathbb{C}^n$. Aside from this, there are a number of other major problems. Hence, we find that the mixed volume is fundamentally related to solutions in $(\mathbb{C}^*)^n$ and *not* to solutions in $\mathbb{C}^n$. There are also significant difficulties in the genericity of the system in question when we allow solutions in $\mathbb{C}^n$. For example, a system of polynomials equations may generically have m solutions in $(\mathbb{C}^*)^n$ while having generically $p \neq m$ solutions in $\mathbb{C}^n$. Recently, much work has been done on this problem; the reader is encouraged to see [3] for appropriate references.

5. Appendix

Below we prove a claim made during the proof of Bernstein's Theorem. Because it lies outside the line of thought of the paper, it is included in this appendix. This claim is an exercise in [3].

Proposition 5.1. *Let $\nu \in \mathbb{Z}^n$ be primitive. Then there exists an integer matrix B with ν as its first row. Moreover, B has an inverse that is also an integer matrix.*

Proof. View ν as a column vector. It suffices to find an integer matrix A with integer inverse such that $A\nu = e_1$, where $e_1 = (1, 0, \ldots, 0)^T$. For if such an A exists, then

$$\nu = A^{-1} e_1$$

and thus the matrix A^{-1} has the desired properties of B, since the above shows that the first row of A^{-1} is ν. Consider now integer row operations: these consist of switching two rows, multiplying a row by ± 1 and adding an integral multiple of a row to another row. It is easy to show that matrices that are elementary with respect to interger row operations are integer matrices having integer inverses. This is done by writing out the matrix corresponding to each operation, finding that its determinant is an integer and then using the distributivity of the determinant. Hence, if we can use integer row operations to transform ν into e_1, then there exists corresponding integer row operation matrices $E_1, \ldots E_k$ such that

$$E_1 \cdots E_k \nu = e_1.$$

Observe that $\det(E_1 \cdots E_k) = \det(E_1) \cdots \det(E_l) \in \mathbb{Z}$ because each $\det(E_i) \in \mathbb{Z}$, and $(E_1 \cdots E_k)^{-1} = E_k^{-1} \cdots E_1^{-1}$ is an intger matrix because each E_i^{-1} is. So, $E_1 \cdots E_k$ has the desired properties of B, and hence we have reduced the problem to transforming ν to e_1 via integer row operations.

Put ν into the form $(b_1, \ldots, b_n)^T$ where $b_1 > 0$ and $b_i \geq b_1$ when $b_i \neq 0$, using only integer row operations; multiply any negative entries of ν by -1 and then switch the smallest non-zero element of ν, which exists because ν is not the null vector, with the first entry in ν. With $(b_1, \ldots, b_n)^T$ as desired, subtract multiples of b_1 from some $b_i \neq 0$, $1 < i \leq n$ until either the entry vanishes or is positive and less than b_1.

Now, again put this vector into the form $(c_1, \ldots, c_n)^T$ where $c_1 > 0$ and $c_i \geq c_1$ when $c_i \neq 0$, and subtract multiples of c_1 form some $c_i \neq 0$ until it vanishes or is positive and less than c_1. Iterating this procedure, we come to a vector $\tilde{\nu}$ that is a non-zero scalar multiple of e_1:

$$\tilde{\nu} = \lambda e_1, \quad \lambda \in \mathbb{Z} \setminus \{0\}.$$

From the above discussion, there exists integer row operation matrices $E_1, \ldots, E_k$ such that

$$E_1 \cdots E_k \nu = \lambda e_1;$$

in other words, there exists an integer matrix A with integer inverse such that $A\nu = \lambda e_1$. Multiplying from the left by A^{-1} gives $\nu = \lambda A^{-1} e_1$. Since A^{-1} is an integer matrix, we see that λ divides all elements of ν. Since ν is primitive, its elements are relativlely prime, and hence $\lambda = 1$. Thus, $A\nu = e_1$, which completes the proof. □

6. Acknowledgements

This paper was written as a requirement of MATH 413 Topics in Commutative Algebra at Queen's University at Kingston, fall term 2006. The author wishes to thank Nathan Grieve, Sarah Mayes and Dr. Greg Smith for their helpful comments with regards to previous drafts.

References

[1] B. Sturmfels, Amer. Math. Monthly. **105, 10**, 907 (1998).
[2] D. Cox, J. Little and D. O'Shea. *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, New York, N.Y. 1997.
[3] D. Cox, J. Little and D. O'Shea. *Using Algebraic Geometry*. Springer, New York, N.Y. 2005.
[4] D. Bernstein, Functional Anal. Appli. **9**, 1 (1975).

E-mail address: 3my@qlink.queensu.ca

Department of Mathematics and Statistics, Queen's University at Kingston, Kingston, Ontario, K7L 3N6, Canada