Theory and Applications of Lattice Point Methods for Binomial Ideals

Ezra Miller

Abstract This survey of methods surrounding lattice point methods for binomial ideals begins with a leisurely treatment of the geometric combinatorics of binomial primary decomposition. It then proceeds to three independent applications whose motivations come from outside of commutative algebra: hypergeometric systems, combinatorial game theory, and chemical dynamics. The exposition is aimed at students and researchers in algebra; it includes many examples, open problems, and elementary introductions to the motivations and background from outside of algebra.

Keywords binomial ideal, primary decomposition, polynomial ring, affine semigroup, commutative monoid, lattice point, convex polyhedron, monomial ideal, combinatorial game, lattice game, rational strategy, misère quotient, Horn hypergeometric system, mass-action kinetics

1 Introduction

Binomial ideals in polynomial rings over algebraically closed fields admit *binomial primary decompositions*: expressions as intersections of primary binomial ideals. The algebra of these decompositions is governed by the geometry of lattice points in polyhedra and related lattice-point combinatorics arising from congruences on commutative monoids. The treatment of this geometric combinatorics is terse at the source [DMM10]. Therefore, a primary goal of this exposition is to provide a more leisurely tour through the relevant phenomena; this is the concern of Sects. 2, 3, and 4.

That the geometry of congruences should govern binomial primary decomposition was a realization made in the context of classical multivariate hypergeometric

Ezra Miller
Mathematics Department, Duke University, Durham, NC 27708, USA
e-mail: ezra@math.duke.edu

G. Fløystad et al. (eds.), *Combinatorial Aspects of Commutative Algebra and Algebraic Geometry*, Abel Symposia 6, DOI 10.1007/978-3-642-19492-4_8,

series, going back to Horn, treated in Sect. 5. Lattice-point combinatorics related to monoids and congruences has recently been shown relevant to the theory of combinatorial games, and is sure to play a key role in algorithms for computing rational strategies and misère quotients, as discussed in Sect. 6. Finally, binomial commutative algebra is central to a long-standing conjecture on the dynamics of chemical reactions under mass-action kinetics. The specifics of this connection are briefly outlined in Sect. 7, along with the potential relevance of combinatorial methods for binomial primary decomposition. Limitations of time and space prevented the inclusion of algebraic statistics in this survey; for an exposition of binomial aspects of Markov bases and conditional independence models, such as graphical models, as well as applications to phylogenetics, see [DSS09] and the references therein.

Sections 2, 3, and 4 are complete in the sense that statements are made in full generality, and precise references are provided for the details of any argument that is only sketched. In contrast, Sects. 5, 6, and 7 are more expository. The results there are sometimes stated in less than full generality—but still mathematically precisely—to ease the exposition. In addition, Sects. 5, 6, and 7 are independent of one another, and to a large extent independent of Sects. 2, 3, and 4, as well; readers interested in the applications should proceed to the relevant sections and refer back as necessary.

Acknowledgements. I am profoundly grateful to the organizers and participants of the International School on Combinatorics at Sevilla, Spain in January 2010, where these notes were presented as five lectures. That course was based on my Abel Symposium talk at Voss, Norway in June 2009; I am similarly indebted to that meeting's organizers. Thanks also go to my coauthors, from whom I learned so much while working on various projects mentioned in this survey. Funding was provided by NSF CAREER grant DMS-0449102 = DMS-1014112 and NSF grant DMS-1001437.

Theory

2 Affine Semigroups and Prime Binomial Ideals

2.1 Affine Semigroups

Let $\mathbb{Z}^d \subset \mathbb{R}^d$ denote the integer points in a real vector space of dimension d. Any integer point configuration

$$A = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{Z}^d \quad \longleftrightarrow \quad A = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix} \in \mathbb{Z}^{d\times n}$$

can be identified with a $d \times n$ integer matrix.

Definition 2.1. A *monoid* is a set with an associative binary operation and an identity element. An *affine semigroup* is a monoid that is isomorphic to

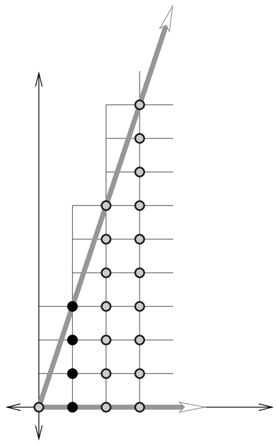
$$\mathbb{N}A = \mathbb{N}\{\mathbf{a}_1,\ldots,\mathbf{a}_n\} = \{c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n \mid c_1,\ldots,c_n \in \mathbb{N}\}$$

for some lattice point configuration $A \subset \mathbb{Z}^d$.

Thus a monoid is a group without inverses. Although a *semigroup* is generally not required to have an identity element, standard terminology from the literature dictates that an affine semigroup is a monoid, and in particular (isomorphic to) a finitely generated submonoid of an integer lattice $\mathbb{Z}^d$ for some d.

Example 2.2. The configuration in the plane $\mathbb{Z}^2$ drawn as solid dots in the following diagram generates the affine semigroup comprising all lattice points in the real cone bounded by the thick horizontal ray and the diagonal ray.

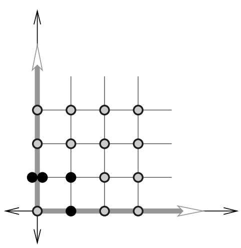
$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \quad \longleftrightarrow$$



This example will henceforth be referred to as "0123".

Example 2.3. A point configuration is allowed to have repeated elements, such as

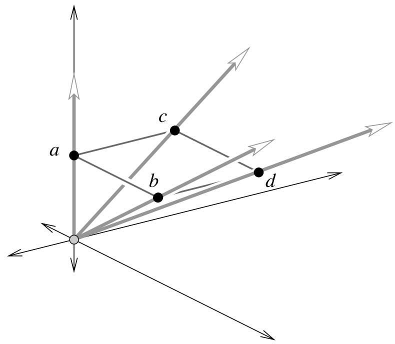
$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \longleftrightarrow$$



in $\mathbb{Z}^2$, which generates the affine semigroup $\mathbb{N}A = \mathbb{N}^2$ of all lattice points in the nonnegative quadrant. This example will henceforth be referred to as "$\substack{1100\\0111}$".

Example 2.4. It will be helpful, later on, to have a three-dimensional example ready. Consider a square at height 1 parallel to the horizontal plane. It can be represented as a matrix and point configuration in $\mathbb{Z}^3$ as follows:

$$A = \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \quad \longleftrightarrow$$



The affine semigroup $\mathbb{N}A$ comprises all of the lattice points in the real cone generated by the vertices of the square.

In all of these examples, the affine semigroups are *normal*: each one equals the set of all lattice points from a rational polyhedral cone. General affine semigroups need not be normal, though they always comprise "most" of the lattice points in a cone.

Example 2.5. In Example 2.2, the outer columns of the matrix, namely $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\3\end{smallmatrix}\right]$, correspond to the extremal rays; they therefore generate the rational polyhedral cone whose lattice points constitute the 0123 affine semigroup. Consequently, the configuration $\left[\begin{smallmatrix}1&1&1\\0&2&3\end{smallmatrix}\right]$ generates the same rational polyhedral cone, but the affine semigroup it generates is different—and not normal—because the point $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ does not lie in it, even though the configuration still generates $\mathbb{Z}^2$ as a group.

The geometry of binomial primary decomposition is based on the sort of geometry that arises from the projection determined by A. To be more precise, A determines a monoid morphism $\mathbb{Z}^d \leftarrow \mathbb{N}^n$. This morphism can be expressed as the restriction of the group homomorphism (the linear map) $\mathbb{Z}^d \leftarrow \mathbb{Z}^n$ induced by A. The diagram is as follows:

$$\begin{array}{ccccccc}
\mathbf{a}_j & \longleftarrow\!\!\shortmid & \mathbf{e}_j & & & & \\
& \boxed{\begin{array}{cc} & n \\ d & A \end{array}} & & & & & \\
\mathbb{Z}^d & \longleftarrow & \mathbb{N}^n & & & & \\
\| & & \cap\!| & & & & \\
\mathbb{Z}^d & \longleftarrow & \mathbb{Z}^n & \longleftarrow & L & \leftarrow & 0 \\
\cap\!| & & \cap\!| & & \cap\!| & & \\
\mathbb{R}^d & \longleftarrow & \mathbb{R}^n & \longleftarrow & \mathbb{R}L & \leftarrow & 0
\end{array}$$

The kernel L of the homomorphism $\mathbb{Z}^d \leftarrow \mathbb{Z}^n$ induced by A is a *saturated* lattice in $\mathbb{Z}^n$, meaning that $\mathbb{Z}^n/L$ is torsion-free, or equivalently that $L = \mathbb{R}L \cap \mathbb{Z}^n$, where $\mathbb{R}L = \mathbb{R} \otimes_{\mathbb{Z}} L$ is the real subspace of $\mathbb{R}^n$ generated by L.

It makes little sense to say that the monoid morphism $\mathbb{Z}^d \leftarrow \mathbb{N}^n$ has a kernel: it is often the case that $L \cap \mathbb{N}^n = \{0\}$, even when the monoid morphism is far from injective. However, when $\mathbb{Z}^d \leftarrow \mathbb{N}^n$ fails to be injective, the fibers admit clean geometric descriptions, inherited from the fact that the fibers of the vector space map $\mathbb{R}^d \leftarrow \mathbb{R}^n$ are the cosets of $\mathbb{R}L$ in $\mathbb{R}^n$.

Definition 2.6. A *polyhedron* in a real vector space is an intersection of finitely many closed real half-spaces.

This survey assumes basic knowledge of polyhedra. Readers for whom Definition 2.6 is not familiar are urged to consult [Zie95, Chaps. 0, 1, and 2].

Lemma 2.7. *The fiber of the monoid morphism* $\mathbb{Z}^d \xleftarrow{A} \mathbb{N}^n$ *over a given lattice point* $\alpha \in \mathbb{Z}^d$ *is the set*

$$F_\alpha = \mathbb{N}^n \cap (\mathbf{u} + L) = \mathbb{N}^n \cap P_\alpha$$

of lattice points in the polyhedron

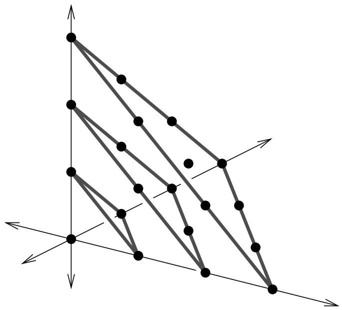
$$P_\alpha = (\mathbf{u} + \mathbb{R}L) \cap \mathbb{R}^n_{\geq 0}$$

for any vector $\mathbf{u} \in \mathbb{Z}^d$ *satisfying* $A\mathbf{u} = \alpha$*, where* $L = \ker A$.

This description is made particularly satisfying by the fact that the polyhedra for various $\alpha \in \mathbb{Z}^d$ are all related to one another.

Example 2.8. When $A = [1\ 1\ 1]$ is the "coordinate-sum" map $\mathbb{N} \leftarrow \mathbb{N}^3$, the polyhedra P_α for $\alpha \in \mathbb{N}$ are equilateral triangles, the lattice points in them corresponding to the monomials of total degree α in three variables:

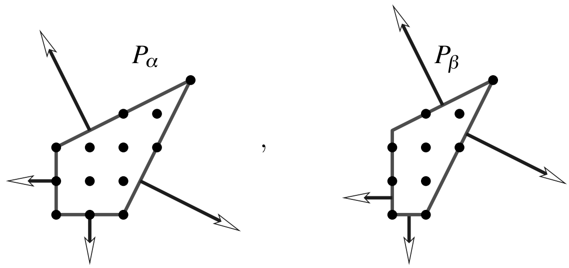
fibers of $A = [1\ 1\ 1]$ $\longleftrightarrow$



The polyhedra in Example 2.8 are all scalar multiples of one another, but this phenomenon is special to codimension 1. In general, when $n - d > 1$, the polyhedra P_α in the family indexed by $\alpha \in \mathbb{N}A$ have facet normals chosen from the same fixed set of possibilities—namely, the images in L^* of the dual basis vectors of $(\mathbb{Z}^n)^*$—so their shapes feel roughly similar, but faces can shrink or disappear.

Example 2.9. In the case of 0123, a basis for the kernel $\ker A$ can be chosen so that the inclusion $\mathbb{Z}^n \leftarrow \ker A$ is given by the matrix

$$B = \begin{bmatrix} 1 & 0 \\ -2 & 1 \\ 1 & -2 \\ 0 & 1 \end{bmatrix} \quad \rightsquigarrow$$



so that, for example, the depicted polytopes P_α and P_β for $\alpha = \left[\begin{smallmatrix} 8 \\ 12 \end{smallmatrix}\right]$ and $\beta = \left[\begin{smallmatrix} 7 \\ 12 \end{smallmatrix}\right]$ both have outer normal vectors that are the negatives of the rows of B. Moving to P_γ for $\gamma = \left[\begin{smallmatrix} 6 \\ 12 \end{smallmatrix}\right]$ would shrink the bottom edge entirely. The corresponding fibers F_α, F_β, and F_γ comprise the lattice points in these polytopes.

2.2 Affine Semigroup Rings

Definition 2.10. The *affine semigroup ring* of $\mathbb{N}A$ over a field $\Bbbk$ is

$$\Bbbk[\mathbb{N}A] = \bigoplus_{\alpha \in \mathbb{N}A} \Bbbk \cdot \mathbf{t}^\alpha,$$

a subring of the Laurent polynomial ring

$$\Bbbk[\mathbb{Z}^d] = \Bbbk[t_1^{\pm 1}, \ldots, t_d^{\pm 1}],$$

in which

$$\mathbf{t}^\alpha = t_1^{\alpha_1} \cdots t_d^{\alpha_d} \quad \text{and} \quad \mathbf{t}^{\alpha+\beta} = \mathbf{t}^\alpha \mathbf{t}^\beta.$$

The definition could be made with $\Bbbk$ an arbitrary commutative ring, but in fact the case we care about most is $\Bbbk = \mathbb{C}$, the field of complex numbers. The reason is that the characteristic zero and algebraically closed hypotheses enter at key points; these notes intend to be precise about which hypotheses are needed where.

Definition 2.11. Denote by π_A the surjection

$$\begin{aligned} \Bbbk[\mathbb{N}A] &\overset{\pi_A}{\twoheadleftarrow} \Bbbk[\mathbf{x}] \\ t^{\mathbf{a}_i} &\leftarrow\!\shortmid x_i \end{aligned}$$

onto the affine semigroup ring $\Bbbk[\mathbb{N}A]$ from the polynomial ring $\Bbbk[\mathbf{x}] := \Bbbk[x_1 \ldots, x_n]$.

The next goal is to calculate the kernel $I_A = \ker(\pi_A)$. To do this it helps to note that both $\Bbbk[\mathbb{N}A]$ and the polynomial ring are graded, in the appropriate sense.

Definition 2.12. Let $A \in \mathbb{Z}^{d\times n}$, so $\mathbb{Z}A \subseteq \mathbb{Z}^d$ is a subgroup. A ring R is A-*graded* if R is a direct sum of *homogeneous components*

$$R = \bigoplus_{\alpha\in\mathbb{Z}A} R_\alpha, \quad \text{such that} \quad R_\alpha R_\beta \subseteq R_{\alpha+\beta}.$$

An ideal in an A-graded ring is A-*graded* if it is generated by homogeneous elements.

Example 2.13. The affine semigroup ring $\Bbbk[\mathbb{N}A]$ is A-graded, with $\Bbbk[\mathbb{N}A]_\alpha = \Bbbk\cdot\mathbf{t}^\alpha$ if $\alpha\in\mathbb{N}A$, and $\Bbbk[\mathbb{N}A]_\alpha = 0$ otherwise. The polynomial ring $\Bbbk[\mathbf{x}]$ is also A-graded, with

$$\Bbbk[\mathbf{x}]_\alpha = \Bbbk\cdot\{F_\alpha\},$$

the vector space spanned by the fiber F_α from Lemma 2.7.

Proposition 2.14. *The kernel of the surjection π_A from Definition 2.11 is*

$$\begin{aligned} I_A &= \langle \mathbf{x}^\mathbf{u} - \mathbf{x}^\mathbf{v} \mid \mathbf{u},\mathbf{v}\in\mathbb{N}^n \text{ and } A\mathbf{u} = A\mathbf{v}\rangle \\ &= \langle \mathbf{x}^\mathbf{u} - \mathbf{x}^\mathbf{v} \mid \mathbf{u},\mathbf{v}\in\mathbb{N}^n \text{ and } \mathbf{u}-\mathbf{v}\in\ker A\rangle. \end{aligned}$$

Proof. The "$\supseteq$" containment follows simply because $\pi_A(\mathbf{x}^\mathbf{u}) = \mathbf{x}^{A\mathbf{u}}$. The reverse containment uses the A-grading: in the ring $R = \Bbbk[\mathbf{x}]/\langle \mathbf{x}^\mathbf{u} - \mathbf{x}^\mathbf{v} \mid A\mathbf{u} = A\mathbf{v}\rangle$, the dimension of the image of $\Bbbk[\mathbf{x}]_\alpha$ as a vector space over $\Bbbk$ is either 0 or 1 since $A\mathbf{u} = A\mathbf{v} = \alpha$ if $\mathbf{u}$ and $\mathbf{v}$ lie in the same fiber F_α. On the other hand, $R \twoheadrightarrow \Bbbk[\mathbb{N}A]_\alpha$ maps surjectively onto the affine semigroup ring by the "$\supseteq$" containment already proved. The surjection must be an isomorphism because, as we noted in Example 2.13, $\Bbbk[\mathbb{N}A]_\alpha$ has dimension 1 whenever F_α is nonempty. □

Corollary 2.15. *The* toric ideal I_A *is prime.*

Proof. $\Bbbk[\mathbb{N}A]$ is an integral domain, being contained in $\Bbbk[\mathbb{Z}^d]$. □

Example 2.16. In the 0123 case, with variables $\{a,b,c,d\}$ instead of $\{x_1,x_2,x_3,x_4\}$,

$$A = \begin{bmatrix} 1&1&1&1\\ 0&1&2&3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1&0\\ -2&1\\ 1&-2\\ 0&1 \end{bmatrix} \implies I_A = \langle ac-b^2, bd-c^2, ad-bc\rangle,$$

where $\ker A$ is the image of B in $\mathbb{Z}^4$. The presence of binomials $ac-b^2$ and $bd-c^2$ in I_A translates the statement "the columns of the matrix B lie in the kernel of A". However, note that I_A is not generated by these two binomials; it is a complicated problem, in general, to determine a minimal generating set for I_A.

Example 2.17. In the $^{1100}_{0111}$ case, using variables $\{a,b,c,d\}$ instead of $\{x_1,x_2,x_3,x_4\}$,

$$A = \begin{bmatrix} 1&1&0&0\\ 0&1&1&1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1&1\\ -1&-1\\ 1&0\\ 0&1 \end{bmatrix} \implies I_A = \langle ac-b, c-d\rangle,$$

where again $\ker A$ is the image of B in $\mathbb{Z}^4$. In this case, I_A is indeed generated by two binomials corresponding to a basis for the kernel of A, but not the given basis appearing as the columns of B. In Sect. 5, we shall be interested in the ideal generated by the two binomials corresponding to the columns of B.

Example 2.18. In the square cone case, using $\{a,b,c,d\}$ instead of $\{x_1,x_2,x_3,x_4\}$,

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \implies I_A = \langle ad - bc \rangle,$$

where $\ker A$ is the image of B. In the codimension 1 case, when $\ker A$ has rank 1, the toric ideal is always principal, just as any codimension 1 prime ideal in $\Bbbk[\mathbf{x}]$ is.

2.3 Prime Binomial Ideals

At last it is time to define binomial ideals precisely.

Definition 2.19. An ideal $I \subseteq \Bbbk[\mathbf{x}]$ is a *binomial ideal* if it is generated by *binomials*

$$\mathbf{x}^{\mathbf{u}} - \lambda \mathbf{x}^{\mathbf{v}} \quad \text{with} \quad \mathbf{u}, \mathbf{v} \in \mathbb{N}^n \text{ and } \lambda \in \Bbbk.$$

Note that $\lambda = 0$ is allowed: monomials are viable generators of binomial ideals. This may seem counterintuitive, but it is forced by allowing arbitrary nonzero constants λ, and in any case, even ideals generated by differences of monomials ("pure-difference binomials") have associated primes containing monomials.

Example 2.20. The ideal $\langle x^3 - y^2, x^3 - 2y^2 \rangle \subseteq \Bbbk[x,y]$ is generated by binomials but equals the monomial ideal $\langle x^3, y^2 \rangle$, no matter the characteristic of $\Bbbk$. Worse, the ideal $I = \langle x^2 - xy, xy - 2y^2 \rangle \subseteq \Bbbk[x,y]$ is generated by "honest" binomials that are not linear combinations of monomials in I, and yet I contains monomials, because I contains both of $x^2y - xy^2$ and $x^2y - 2xy^2$, so the monomials x^2y and xy^2 lie in I.

Example 2.21. The pure-difference binomial ideal $I = \langle x^2 - xy, xy - y^2 \rangle \subseteq \Bbbk[x,y]$ has a monomial associated prime ideal $\langle x,y \rangle$, since $I = \langle x - y \rangle \cap \langle x^2, y \rangle$.

In general, which binomial ideals are prime? We have seen that toric ideals I_A are prime, but for binomial primary decomposition in general it is important to know all of the other binomial primes, as well. The answer was given by Eisenbud and Sturmfels [ES96, Corollary 2.6].

Theorem 2.22. *When $\Bbbk$ is algebraically closed, a binomial ideal $I \subseteq \Bbbk[x_1,\ldots,x_m]$ is prime if and only if it is the kernel of a surjective A-graded homomorphism π : $\Bbbk[\mathbf{x}] \twoheadrightarrow \Bbbk[\mathbb{N}A]$ in which the variables are homogeneous.*

The surjection π in Theorem 2.22 need not equal π_A. For example, $\pi(x_j) = 0$ is allowed, so I could contain monomials. Furthermore, even when $\pi(x_j) \neq 0$, the image of x_j could be $\lambda \mathbf{t}^\alpha$ for some $\lambda \neq 1$. Thus, if $\mathbf{x}^\mathbf{u} \mapsto \lambda \mathbf{t}^\alpha$ and $\mathbf{x}^\mathbf{v} \mapsto \mu \mathbf{t}^\alpha$, then $\mathbf{x}^\mathbf{u} - \frac{\mu}{\lambda}\mathbf{x}^\mathbf{v} \in I$: the binomial generators of I need not be pure differences. However, assuming I is prime, we are not free to assign the coefficients λ and μ at will:

$$\mathbf{x}^\mathbf{u} - \tfrac{\mu}{\lambda}\mathbf{x}^\mathbf{v} \in I \implies \mathbf{x}^{\mathbf{u}+\mathbf{w}} - \tfrac{\mu}{\lambda}\mathbf{x}^{\mathbf{v}+\mathbf{w}} \in I \text{ for } \mathbf{w} \in \mathbb{N}^n$$
$$\text{and } \mathbf{x}^{r\mathbf{u}} - \tfrac{\mu^r}{\lambda^r}\mathbf{x}^{r\mathbf{v}} \in I \text{ for } r \in \mathbb{N}.$$

The first line means that the coefficient on $\mathbf{x}^\mathbf{v}$ in $\mathbf{x}^\mathbf{u} - \mathbf{x}^\mathbf{v}$ depends only on $\mathbf{u}-\mathbf{v}$, and the second means essentially that the assignment $\mathbf{u}-\mathbf{v} \mapsto \mu/\lambda$ constitutes a homomorphism $\ker A \to \Bbbk^*$. The precise statement requires a definition.

Definition 2.23. A *character* on a sublattice $L \subseteq \mathbb{Z}^n$ is a homomorphism $\rho : L \to \Bbbk^*$. If $L \subseteq \mathbb{Z}^J$ for some subset $J \subseteq \{1,\dots,n\}$, then

$$\begin{aligned} I_{\rho,J} &= I_\rho + \mathfrak{m}_J, \\ \text{where} \quad I_\rho &= \langle \mathbf{x}^\mathbf{u} - \rho(\mathbf{u}-\mathbf{v})\mathbf{x}^\mathbf{v} \mid \mathbf{u}-\mathbf{v} \in L\rangle \\ \text{and} \quad \mathfrak{m}_J &= \langle x_i \mid i \notin J\rangle. \end{aligned}$$

Corollary 2.24. *A binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ with $\Bbbk$ algebraically closed is prime if and only if $I = I_{\rho,J}$ for a character $\rho : L \to \Bbbk^*$ defined on a saturated lattice $L \subseteq \mathbb{Z}^J$.*

In other words, every prime binomial ideal in the polynomial ring $\Bbbk[\mathbf{x}]$ over an algebraically closed field $\Bbbk$ is toric after forgetting some of the variables (those outside of J) and rescaling the rest (by the character ρ).

Remark 2.25. Given any sublattice $L \subseteq \mathbb{Z}^n$, a character ρ is defined as a homomorphism $L \to \Bbbk^*$. On the other hand, rescaling the variables x_j for $j \in J$ amounts to a homomorphism $\mathbb{Z}^J \to \Bbbk^*$. When $L \subsetneq \mathbb{Z}^J$, there is usually no unique way to extend ρ to a character $\mathbb{Z}^J \to \Bbbk^*$ (there can be a unique way if $\Bbbk$ has positive characteristic). However, there is always at least one way when L is saturated—so the inclusion $L \hookrightarrow \mathbb{Z}^n$ is split—because the natural map $\mathrm{Hom}(\mathbb{Z}^J, \Bbbk^*) \to \mathrm{Hom}(L, \Bbbk^*)$ is surjective.

Example 2.26. Let $\omega = \frac{1+\sqrt{-3}}{2} \in \mathbb{C}$ be a primitive cube root of 1. If $L \subseteq \mathbb{Z}^4 \subseteq \mathbb{Z}^5$ is spanned by the columns of the matrix B, below, and the character ρ takes the indicated values on these generators of L, then $I_{\rho,J}$ for $J = \{1,2,3,4\}$ is as indicated.

$$\begin{array}{c} B = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \\ -1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \\ \rho : \ \omega^2 \quad \omega \quad \omega \end{array} \implies I_{\rho,\{1,2,3,4\}} = \langle bc - \omega^2 ad, b^2 - \omega ac, c^2 - \omega bd, e\rangle.$$

For instance, when $\mathbf{u} = (0,1,1,0)$ and $\mathbf{v} = (1,0,0,1)$, we get $\rho(\mathbf{u}-\mathbf{v}) = \omega^2$. Compare this example to the 0123 case in Example 2.16.

3 Monomial Ideals and Primary Binomial Ideals

The lattice-point geometry of binomial primary decomposition generalizes the geometry of monomial ideals. For primary binomial ideals, the connection is particularly clear. To highlight it, this section discusses what it looks like for a primary binomial ideal to have a monomial associated prime. The material in this section is developed in the context of an arbitrary affine semigroup ring, because that generality will be crucial in the applications to binomial ideals in polynomial rings $\Bbbk[\mathbf{x}]$. In return for the generality, there are no restrictive hypotheses on the characteristic or algebraic closure of $\Bbbk$ to contend with; except in Example 3.19 and Theorem 3.20, $\Bbbk$ can be arbitrary.

Definition 3.1. An ideal $I \subseteq \Bbbk[Q]$ in the monoid algebra of an affine semigroup Q over an arbitrary field is a *monomial ideal* if it is generated by *monomials* $\mathbf{t}^\alpha$, and I is a *binomial ideal* if it is generated by *binomials* $\mathbf{t}^\alpha - \lambda\mathbf{t}^\beta$ with $\alpha, \beta \in Q$ and $\lambda \in \Bbbk$.

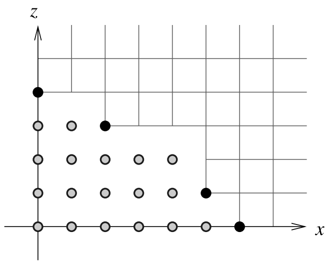
3.1 Monomial Primary Ideals

Given a monomial ideal, it is convenient to have terminology and notation for certain sets of monomials and lattice points.

Definition 3.2. If $I \subseteq \Bbbk[Q]$ is a monomial ideal, then write $\mathrm{std}(I)$ for the set of exponent vectors on its *standard monomials*, meaning those outside of I.

Example 3.3. Here is a monomial ideal in $\Bbbk[Q]$ for $Q = \mathbb{N}^2$:

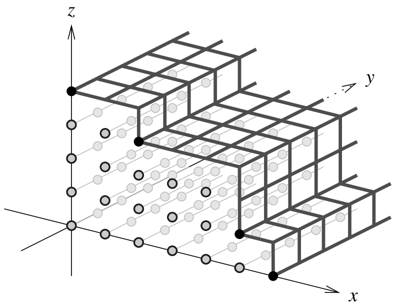
$$I = \langle x^6, x^5z, x^2z^3, z^4 \rangle \subseteq \Bbbk[x,z] \qquad \longleftrightarrow$$



The reason for using z instead of y will become clear in Example 3.4. The bottom of the cross-hatched region is the *staircase* of I; its lower corners are the (lattice points corresponding to) the generators of I. The lattice points below the staircase correspond to the standard monomials of I.

Example 3.4. The same monomial generators can result in a higher-dimensional picture if the ambient monoid is different. Consider the generators from Example 3.3 but in $\Bbbk[Q]$ for $Q = \mathbb{N}^3$:

$$I = \langle x^6, x^5 z, x^2 z^3, z^4 \rangle \subseteq \Bbbk[x,y,z] \qquad \longleftrightarrow$$



The "front face" of the picture—corresponding to the xz-plane—coincides with Example 3.3. The surface cross-hatched in thick lines is the staircase of I; its minimal elements again correspond to the generators of I, drawn as solid dots. Below the staircase sit the lattice points corresponding to the standard monomials of I. There are infinitely many standard monomials, but they occur along finitely many rays parallel to the y-axis, each emanating from a point drawn as a bold hollow dot; these are the points in $\mathrm{std}(I)$ in the xz-plane.

Example 3.5. In the square cone case, Example 2.4, let a,b,c,d be the generators of the affine semigroup ring, as indicated in the figure there. Thus $\Bbbk[Q] = \Bbbk[a,b,c,d]/\langle ad - bc \rangle$. The monomial ideal $\langle c,d \rangle \subseteq \Bbbk[Q]$ is prime; in fact, the composite map $\Bbbk[a,b] \hookrightarrow \Bbbk[Q] \twoheadrightarrow \Bbbk[Q]/\langle c,d \rangle$ is an isomorphism.

The phenomenon in Example 3.5 is general; the statement requires a definition.

Definition 3.6. A *face* $F \subseteq Q$ of an affine semigroup $Q \subseteq \mathbb{Z}^d$ is a subset $F = Q \cap H$ obtained by intersecting Q with a halfspace $H \subseteq \mathbb{R}^d$ such that $Q \subseteq H^+$ is contained in one of the two closed halfspaces H^+, H^- defined by H in $\mathbb{R}^d$.

Lemma 3.7. *A monomial ideal I in an affine semigroup ring is prime $\Leftrightarrow$ $\mathrm{std}(I)$ is a face. The prime $\mathfrak{p}_F$ for $F \subseteq Q$ induces an isomorphism $\Bbbk[F] \hookrightarrow \Bbbk[Q] \twoheadrightarrow \Bbbk[Q]/\mathfrak{p}_F$.*

For a proof of the lemma, and lots of additional background on the connections between faces of cones and the algebra of affine semigroup rings, see [MS05, §7.2].

Before jumping to the question of when an arbitrary binomial ideal is primary, let us first consider the monomial case. In a polynomial ring $\Bbbk[\mathbf{x}]$, there is an elementary algebraic description as well as a satisfying geometric one. Both will be important in later sections, but it is the geometric description that generalizes most easily to arbitrary affine semigroup rings.

Proposition 3.8. *A monomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ is primary if and only if*

$$I = \langle x_{i_1}^{m_1}, \ldots, x_{i_r}^{m_r}, \textit{ some other monomials in } x_{i_1}, \ldots, x_{i_r} \rangle.$$

A monomial ideal $I \subseteq \Bbbk[Q]$ for an arbitrary affine semigroup $Q \subseteq \mathbb{Z}^d$ is $\mathfrak{p}_F$-primary for a face $F \subseteq Q$ if and only if there are elements $\alpha_1, \ldots, \alpha_\ell \in Q$ with

$$\operatorname{std}(I) = \bigcup_{k=1}^{\ell} (\alpha_k + \mathbb{Z}F) \cap Q,$$

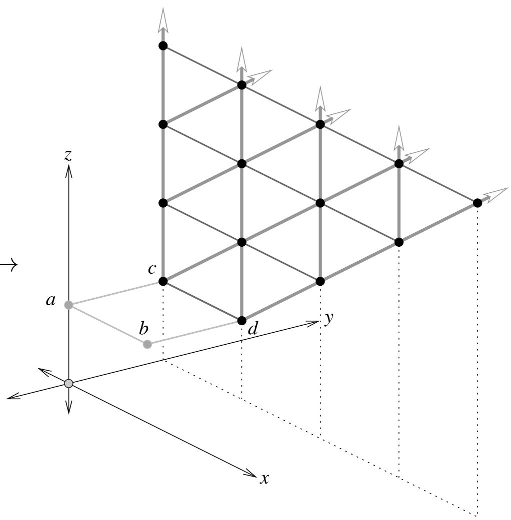
where $\mathbb{Z}F$ is the subgroup of $\mathbb{Z}^d$ generated by F.

Proof. The statement about $\Bbbk[\mathbf{x}]$ is a standard exercise in commutative algebra. The statement about $\Bbbk[Q]$ is the special case of Theorem 3.23, below, in which the binomial ideal I is generated by monomials. $\square$

What does a set of the form $(\alpha + \mathbb{Z}F) \cap Q$ look like? Geometrically, it is roughly the (lattice points in the) intersection of an affine subspace with a cone. In the polynomial ring case, where $Q = \mathbb{N}^n$, a set $(\alpha + \mathbb{Z}F) \cap Q$ is always $\beta + F$ for some lattice point $\beta \in \mathbb{N}^n$: the intersection of a translate of a coordinate subspace with the nonnegative orthant is a translated orthant. In fact, $F = \mathbb{N}^J = \mathbb{N}^n \cap \mathbb{Z}^J$ for some subset $J \subseteq \{1, \ldots, n\}$, and then β is obtained from α by setting all coordinates from J to 0 (if α has negative coordinates outside of J, then $\alpha + \mathbb{Z}F$ fails to meet $\mathbb{N}^n$). For general Q, on the other hand, $(\alpha + \mathbb{Z}F) \cap Q$ need not be a translate of F.

Example 3.9. The prime ideal in Example 3.5 corresponds to the face F consisting of the nonnegative integer combinations of $e_1 + e_3$ and e_3, where e_1, e_2, e_3 are the standard basis of $\mathbb{Z}^3$. In terms of the depiction in Example 2.4, these are the lattice points in Q that lie in the xz-plane. The subgroup $\mathbb{Z}F \subseteq \mathbb{Z}^3$ comprises all lattice points in the xz-plane. Now suppose that $\alpha = e_2$, the first lattice point along the y-axis. Then

$$\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \mathbb{Z} \cdot \{xz\}\text{-plane} \right) \cap Q \quad \longleftrightarrow$$



is a union of two translates of F. For reference, the square over which Q is the cone is drawn lightly, while dotted lines fill out part of the vertical plane $\alpha + \mathbb{Z}F$.

Proposition 3.10. *Every monomial ideal $I \subseteq \Bbbk[Q]$ in an affine semigroup ring has a unique minimal primary decomposition $I = P_1 \cap \cdots \cap P_r$ as an intersection of monomial primary ideals P_i with distinct associated primes.*

Proof sketch. I has a unique irredundant decomposition $I = W_1 \cap \cdots \cap W_s$ as an intersection of *irreducible monomial ideals* W_j. The existence of an irredundant irreducible decomposition can be proved the same way irreducible decompositions are produced for arbitrary submodules of noetherian modules. The uniqueness of such a decomposition, on the other hand, is special to monomial ideals in affine semigroup rings [MS05, Corollary 11.5]; it follows from the uniqueness of *irreducible resolutions* [Mil02, Theorem 2.4]. See [MS05, Chap. 11] for details.

Given the uniqueness properties of minimal monomial irreducible decompositions, the (unique) monomial primary components are obtained by intersecting all irreducible components sharing a given associated prime. □

Remark 3.11. In polynomial rings, uniqueness of monomial irreducible decomposition occurs for approximately the same reason that monomial ideals have unique minimal monomial generating sets: the partial order on irreducible ideals is particularly simple [Mil09, Proposition 1.4]. See [MS05, §5.2] for an elementary derivation of existence and uniqueness of monomial irreducible decomposition by Alexander duality.

3.2 Congruences on Monoids

The uniqueness of irreducible and primary decomposition of monomial ideals rests, in large part, on the fine grading on $\Bbbk[Q]$, in which the nonzero components $\Bbbk[Q]_\alpha$ have dimension 1 as vector spaces over $\Bbbk$. Similar gradings are available for quotients modulo binomial ideals, except that the gradings are by general noetherian commutative monoids, rather than by free abelian groups or by affine semigroups. Our source for commutative monoids is Gilmer's excellent book [Gil84]. For the special case of affine semigroups, by which we mean finitely generated submonoids of free abelian groups, see [MS05, Chap. 7].

For motivation, recall from Lemma 2.7 that the fibers of a monoid morphism from $\mathbb{N}^n$ to $\mathbb{Z}^d$ have nice structure, and that the polynomial ring $\Bbbk[\mathbf{x}]$ becomes graded by $\mathbb{Z}^d$ via such a morphism. The fibers are the equivalence classes in an equivalence relation, as is the case for any map $\pi : Q \to Q'$ of sets; but when π is a morphism of monoids, the equivalence relation satisfies an extra condition.

Definition 3.12. A *congruence* on a commutative monoid Q is an equivalence relation $\sim$ that is *additively closed*, in the sense that

$$u \sim v \;\Rightarrow\; u+w \sim v+w \quad \text{for all } w \in Q.$$

The *quotient* $Q/{\sim}$ is the set of equivalence classes under addition.

Lemma 3.13. *The quotient $\overline{Q} = Q/\!\sim$ of a monoid by a congruence is a monoid. Any congruence $\sim$ on Q induces a $\overline{Q}$-grading on the* monoid algebra $\Bbbk[Q] = \bigoplus_{u\in Q} \Bbbk \cdot \mathbf{t}^u$ *in which the* monomial $\mathbf{t}^u$ *has* degree $\overline{u} \in \overline{Q}$ *whenever* $u \mapsto \overline{u}$.

Proof. This is an easy exercise. It uses that the multiplication on $\Bbbk[Q]$ is given by $\mathbf{t}^u\mathbf{t}^v = \mathbf{t}^{u+v}$ for $u, v \in Q$. ☐

Definition 3.14. In any monoid algebra $\Bbbk[Q]$, a binomial ideal $I \subseteq \Bbbk[Q]$ generated by binomials $\mathbf{t}^u - \lambda\mathbf{t}^v$ with $\lambda \in \Bbbk$ *induces* a congruence $\sim$ (often denoted by $\sim_I$) in which

$$u \sim v \text{ if } \mathbf{t}^u - \lambda\mathbf{t}^v \in I \text{ for some } \lambda \neq 0.$$

Lemma 3.15. *Fix a binomial ideal $I \subseteq \Bbbk[Q]$ in a monoid algebra. Then I and $\Bbbk[Q]/I$ are both graded by $\overline{Q} = Q/\!\sim$. The* Hilbert function $\overline{Q} \mapsto \mathbb{N}$, *which for any $\overline{Q}$-graded vector space M takes $\overline{q} \mapsto$* $\dim_\Bbbk M_{\overline{q}}$, *satisfies*

$$\dim_\Bbbk (\Bbbk[Q]/I)_{\overline{q}} = \begin{cases} 0 & \textit{if } \overline{q} = \{u \in Q \mid \mathbf{t}^u \in I\} \\ 1 & \textit{otherwise.} \end{cases}$$

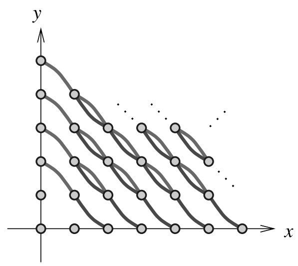
The proof of the lemma is another simple exercise. To rephrase, it says that every pair of monomials in a given congruence class under $\sim_I$ are equivalent up to a nonzero scalar modulo I, and the only monomials sent to 0 lie in I. A slightly less set-theoretic and more combinatorial way to think about congruences uses graphs.

Definition 3.16. Any binomial ideal $I \subseteq \Bbbk[Q]$ defines a graph G_I whose vertices are the elements of the monoid Q and whose (undirected) edges are the pairs $(u, v) \in Q \times Q$ such that $\mathbf{t}^u - \lambda\mathbf{t}^v \in I$ for some nonzero $\lambda \in \Bbbk$. Write $\pi_0 G_I$ for the set of connected components of G_I.

Thus $C \in \pi_0 G_I$ is the same thing as a congruence class under $\sim_I$. The moral of the story is that combinatorics of the graph G_I controls the (binomial) primary decomposition of I.

Example 3.17. Each of the two binomial generators of the ideal

$$I = \langle x^2 - xy, xy - y^2 \rangle \subseteq \Bbbk[x, y] \quad \Longrightarrow \quad G_I =$$



determines a collection of edges of the graph G_I, indicated in the figure, by additivity of the congruence $\sim_I$. In reality, G_I has many more edges than those depicted: since $\sim_I$ is an equivalence relation, every connected component is a complete graph on its vertex set. However, in examples, it is convenient to draw—and more helpful to see—only edges determined by monomial multiples of generating binomials.

The connected components of G_I are the fibers of the monoid morphism of $\mathbb{N}^2$ to

$$\mathbb{N}^2/\sim_I \; =$$

the monoid $\mathbb{N}$ with the element 1 "doubled". The ideal I has primary decomposition

$$I = \langle x^2, xy, y^2 \rangle \cap \langle x - y \rangle.$$

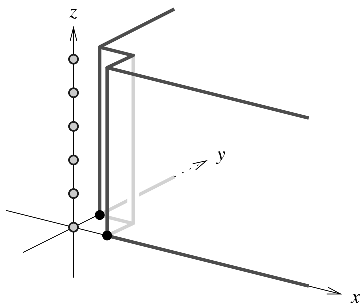
The first primary component reflects the three singleton components of G_I near the origin. The other primary component reflects the diagonal connected components of G_I marching off to infinity.

Although the $\mathbb{N}$-graded Hilbert function of $\Bbbk[x,y]/I$ takes values $1,2,1,1,1,\ldots$, the $\overline{\mathbb{N}}^2$-graded Hilbert function takes only the value 1.

Example 3.18. When $I = I_A$ is the toric ideal for a matrix A, the connected components of G_I are the fibers F_α for $\alpha \in \mathbb{N}A$.

Example 3.19. If $I = I_{\rho,J}$ is a binomial prime in a polynomial ring $\Bbbk[\mathbf{x}]$ over an algebraically closed field $\Bbbk$, and $C \in \pi_0 G_I$ is a connected component, then either $C = \mathbb{N}^n \smallsetminus \mathbb{N}^J$ or else $C = (u+L) \cap \mathbb{N}^J$ for some $u \in \mathbb{N}^J$. When ρ is the trivial (only) character on the lattice $L = \{0\} \subseteq \mathbb{Z}^{\{3\}}$ and $J = \{3\} \subseteq \{1,2,3\}$, for instance, then

$$I = I_{\rho,J} = \langle x, y \rangle \subseteq \Bbbk[x,y,z] \qquad \Longrightarrow \qquad G_I \;=$$



In this case $\mathbb{N}^n \smallsetminus \mathbb{N}^J$ consists of the monomials off of the vertical axis (i.e., those in the region outlined by bold straight lines), whereas every component $(u+L) \cap \mathbb{N}^J$ of G_I is simply a single lattice point on the vertical axis.

3.3 Binomial Primary Ideals with Monomial Associated Primes

The algebraic characterization of monomial primary ideals in the first half of Proposition 3.8 has an approximate analogue for primary binomial ideals, although it requires hypotheses on the base field $\Bbbk$.

Theorem 3.20. *Fix $\Bbbk$ algebraically closed of characteristic* 0*. If $I \subseteq \Bbbk[\mathbf{x}]$ is an $I_{\rho,J}$-primary binomial ideal, then $I = I_\rho + B$ for some binomial ideal $B \supseteq \mathfrak{m}_J^\ell$ with $\ell > 0$.*

Proof. This is the characterization of primary decomposition [ES96, Theorem 7.1] applied to a binomial ideal that is already primary. □

The content of the theorem is that I contains both I_ρ and a power of each variable x_i for $i \notin J$. (In positive characteristic, I contains a Frobenius power of I_ρ, but not necessarily I_ρ itself.) A more precise analogue of the algebraic part of Proposition 3.8 would characterize which binomial ideals B result in primary ideals. However, there is no simple way to describe such binomial ideals B in terms of generators. The best that can be hoped for is an answer to a pair of questions:

(a) What analogue of standard monomials allows us to ascertain when I is primary?
(b) What property of the standard monomials characterizes the primary condition?

Preferably the answers should be geometric, and suitable for binomial ideals in arbitrary affine semigroup rings, as in the second half of Proposition 3.8.

Lemma 3.15 answers the first question. Indeed, if $I \subseteq \Bbbk[Q]$ is a monomial ideal, then $\mathrm{std}(I)$ is exactly the subset of Q such that $\Bbbk[Q]/I$ has Q-graded Hilbert function $\dim_\Bbbk(\Bbbk[Q]/I)_u = 1$ for $u \in \mathrm{std}(I)$ and 0 for $\mathbf{t}^u \in I$. In the monomial case, we could still define the monoid quotient $\overline{Q} = Q/{\sim_I}$, whose classes are all singleton monomials except for the class of monomials in I. Therefore, for general binomial ideals I, the set of non-monomial classes of the congruence $\sim_I$ plays the role of $\mathrm{std}(I)$.

Note that Proposition 3.8 answers the second question for monomial ideals in affine semigroup rings, whose associated primes are automatically monomial. The next step relaxes the condition on I but not on the associated prime: consider a binomial ideal I in an affine semigroup ring $\Bbbk[Q]$, and ask when it is $\mathfrak{p}_F$-primary for a face $F \subseteq Q$. The answer in this case relies, as promised, on the combinatorics of the graph G_I from Definition 3.16 and its set $\pi_0 G_I$ of connected components; however, as Proposition 3.8 hints, the group generated by F enters in an essential way.

Definition 3.21. For a face F of an affine semigroup Q, and any $\Bbbk[Q]$-module M,

$$M[\mathbb{Z}F] = M \otimes_{\Bbbk[Q]} \Bbbk[Q + \mathbb{Z}F]$$

is the localization by inverting all monomials not in $\mathfrak{p}_F$. If $I \subseteq \Bbbk[Q]$ is a binomial ideal, then a connected component $C \in \pi_0 G_I$ is *F-finite* if $C = C' \cap Q$ for some finite connected component C' of the graph $G_{I[\mathbb{Z}F]}$ for the localization $I[\mathbb{Z}F] \subseteq \Bbbk[Q][\mathbb{Z}F]$.

Thus, for example, $\Bbbk[Q][\mathbb{Z}F] = \Bbbk[Q + \mathbb{Z}F]$ is the monoid algebra for the affine semigroup $Q + \mathbb{Z}F$ obtained by inverting the elements of F in Q. In Proposition 3.8, where $I \subseteq \Bbbk[Q]$ is a monomial ideal, all of the connected components of G_I inside of $\mathrm{std}(I)$ are singletons, and the same is true of $G_{I[\mathbb{Z}F]}$. See Sect. 4.1 for additional information and examples concerning geometry and combinatorics of localization.

Example 3.22. Connected components of G_I can be finite but not F-finite for a given face F of Q. For instance, if $I = \langle x - y \rangle \subseteq \Bbbk[x, y] = \Bbbk[\mathbb{N}^2]$, then each connected component of G_I is finite—comprising the set of monomials in $\Bbbk[x, y]$ of some fixed total degree—but not F-finite if F is the horizontal axis of $Q = \mathbb{N}^2$: once x is inverted, $\pi_0 G_{I[x^{-1}]}$ consists of infinite northwest-pointing rays in the upper half-plane.

Theorem 3.23. *Fix a monomial prime ideal $\mathfrak{p}_F = \langle \mathbf{t}^u \mid u \notin F \rangle$ in an affine semigroup ring $\Bbbk[Q]$ for a face $F \subseteq Q$. A binomial ideal $I \subseteq \Bbbk[Q]$ is $\mathfrak{p}_F$-primary if and only if*

(a) Every connected component of G_I other than $\{u \in Q \mid \mathbf{t}^u \in I\}$ is F-finite.
(b) F acts on the set of F-finite components semifreely with finitely many orbits.

Thus the primary condition is fundamentally a finiteness condition—or really a pair of finiteness conditions. A proof is sketched after Example 3.26; but first, the terminology requires precise explanations. Semifreeness, for example, guarantees that the set of F-finite components is a subset of a set acted on freely by $\mathbb{Z}F$; this is part of the characterization of semifree actions in [KM10].

Definition 3.24. An *action* of a monoid F on a set T is a map $F \times T \to T$, written $(f,t) \mapsto f+t$, that satisfies $0+t = t$ for all $t \in T$ and respects addition: $(f+g)+t = f+(g+t)$. The monoid action is *semifree* if $t \mapsto f+t$ is an injection $T \hookrightarrow T$ for each $f \in F$, and $f \mapsto f+t$ is an injection $F \hookrightarrow T$ for each $t \in T$.

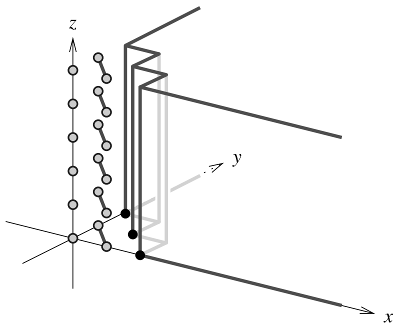
In contrast to group actions, monoid actions do not a priori define equivalence relations, because the relation $t \sim f+t$ can fail to be symmetric. The relation is already reflexive and transitive, however, precisely by the two axioms for monoid actions.

Definition 3.25. An *orbit* of a monoid action of F on T is an equivalence class under the symmetrization of the relation $\{(s,t) \mid f+s = t \text{ for some } f \in F\} \subseteq T \times T$.

Combinatorially, if F acts on T, one can construct a directed graph with vertex set T and an edge from s to t if $t = f+s$ for some $f \in F$. Then an orbit is a connected component of the underlying undirected graph.

Example 3.26. The ideal

$$I = \langle x-y, x^2 \rangle \subseteq \Bbbk[x,y,z] \quad \Longrightarrow \quad G_I \;=$$



is $\mathfrak{p}$-primary for $\mathfrak{p} = \mathfrak{p}_F = \langle x,y \rangle$, where F is the z-axis of $\mathbb{N}^3$. The monoid F acts on the F-finite connected components of G_I with two orbits: one on the z-axis, where each connected component is a singleton; and one adjacent orbit, where every connected component is a pair. The monomial class in this example (outlined by bold straight lines) is the set of monomials in $\langle x^2, xy, y^2 \rangle$.

Example 3.27. Fix notation as in Examples 3.5 and 3.9. The ideal $I = \langle c^2, cd, d^2\rangle \subseteq \Bbbk[Q]$ is $\mathfrak{p}_F$-primary. In contrast to Example 3.26, this time the F-finite connected components are all singletons, but the two orbits are not isomorphic as sets acted on by the face F: one orbit is F itself, while the other is the set depicted in Example 3.9.

Proof sketch for Theorem 3.23. This theorem is the core conclusion of [DMM10, Theorem 2.15 and Proposition 2.13]. The argument is summarized as follows.

For any set T, let $\Bbbk\{T\}$ denote the vector space over $\Bbbk$ with basis T. If $\pi_0 G_I$ satisfies the two conditions, then $\Bbbk[Q]/I$ has finite a filtration, as a $\Bbbk[Q]$-module, whose associated graded pieces are the vector spaces $\Bbbk\{T\}$ for the finitely many F-orbits T of F-finite components of G_I. In fact, semifreeness guarantees that for each orbit T, the vector space $\Bbbk\{T\}$ is naturally a torsion-free module over $\Bbbk[F] = \Bbbk[Q]/\mathfrak{p}_F$. Finiteness of the number of orbits guarantees that the associated graded module of $\Bbbk[Q]/I$ is a finite direct sum of modules $\Bbbk\{T\}$, so it has only one associated prime, namely $\mathfrak{p}_F$. Consequently $\Bbbk[Q]/I$ itself has just one associated prime.

For the other direction, when I is $\mathfrak{p}_F$-primary, one proves that inverting the monomials and binomials outside of $\mathfrak{p}_F$ annihilates the $\overline{Q}$-graded pieces of $\Bbbk[Q]/I$ for which the connected component in $G_{I[\mathbb{Z}F]}$ is infinite [DMM10, Lemmas 2.9 and 2.10]. Since the elements outside of $\mathfrak{p}_F$ act injectively on $\Bbbk[Q]/I$ by definition of $\mathfrak{p}_F$-primary, every class of $\sim_I$ that is not F-finite must therefore already consist of monomials in I. The semifree action of F on the F-finite components derives simply from the fact that $\Bbbk[Q]/I$ is torsion-free as a $\Bbbk[F]$-module, where the $\Bbbk[F]$-action is induced by the inclusion $\Bbbk[F] \subseteq \Bbbk[Q]$. Generalities about $\overline{Q}$-gradings of this sort imply that $\Bbbk[Q]/I$ possesses a filtration whose associated graded pieces are as in the previous paragraph. The minimality of $\mathfrak{p}_F$ over I implies that the length of the filtration is finite. $\square$

4 Binomial Primary Decomposition

General binomial ideals induce more complicated congruences than primary binomial ideals. This section completes the combinatorial analysis of binomial primary decomposition by describing how to pass from an arbitrary binomial ideal to its primary components. There are crucial points where characteristic zero or algebraically closed hypotheses are required of the field $\Bbbk$, but those will be mentioned explicitly; if no mention is made, then $\Bbbk$ is assumed to be arbitrary.

4.1 Monomial Primes Minimal over Binomial Ideals

The first step is to consider again the setting from the previous section, particularly Theorem 3.23, where a monomial prime ideal $\mathfrak{p}_F$ is associated to I in an arbitrary affine semigroup ring $\Bbbk[Q]$, except that now the binomial ideal I is not assumed to be primary. The point is to construct its $\mathfrak{p}_F$-primary component. The nature of Theorem 3.23 splits the construction into two parts:

(a) ensuring that the (non-monomial) connected components are all F-finite, and
(b) forcing the face F to act in the correct manner on the components.

These operations will be carried out in reverse order, with part 2 being accomplished by localization, and then part 1 being accomplished by simply lumping all of the connected components that are not F-finite together.

Definition 4.1. For a face F of an affine semigroup Q and a binomial ideal $I \subseteq \Bbbk[Q]$,

$$(I : \mathbf{t}^F) = I[\mathbb{Z}F] \cap \Bbbk[Q]$$

is the kernel of the composite map $\Bbbk[Q] \to \Bbbk[Q]/I \to (\Bbbk[Q]/I)[\mathbb{Z}F]$.

Remark 4.2. The notation $(I : \mathbf{t}^F)$ is explained by an equivalent construction of this ideal. Indeed, the usual meaning of the colon operation for an element $y \in \Bbbk[Q]$ is that $(I : y) = \{z \in \Bbbk[Q] \mid yz \in I\}$. Here, $(I : \mathbf{t}^F) = (I : \mathbf{t}^f)$ for any lattice point f lying sufficiently far in the relative interior of F. Equivalently, $(I : \mathbf{t}^F) = (I : (\mathbf{t}^f)^\infty) = \bigcup_{r\in\mathbb{N}}(I : \mathbf{t}^{rf})$ for any lattice point $f \in F$ that does not lie on a proper subface of F.

Combinatorially, the passage from I to $(I : \mathbf{t}^F)$ has a concrete effect.

Lemma 4.3. *The connected components of the graph $G_{(I:\mathbf{t}^F)}$ defined by $(I : \mathbf{t}^F)$ are obtained from $\pi_0 G_I$ by joining together all pairs of components C_u and C_v such that $C_{u+f} = C_{v+f}$ for some $f \in F$, where C_u is the component containing $u \in Q$.*

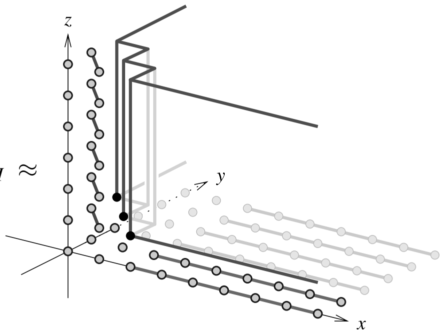
Proof. Two lattice points $u, v \in Q$ lie in the same component of $G_{(I:\mathbf{t}^F)}$ exactly when there is a binomial $\mathbf{t}^u - \lambda\mathbf{t}^v \in \Bbbk[Q]$ such that $\mathbf{t}^f(\mathbf{t}^u - \lambda\mathbf{t}^v) \in I$ for some $f \in F$. $\square$

Roughly speaking: join the components if they become joined after moving them up by an element $f \in F$. Illustrations of lattice point phenomena related to binomial primary decomposition become increasingly difficult to draw in two dimensions as the full nature of the theory develops, but a small example is possible at this stage.

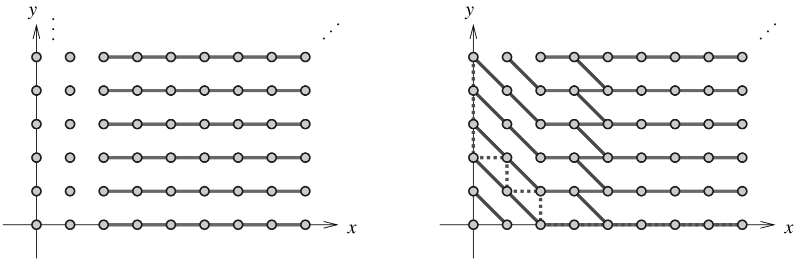
Example 4.4. The ideal $I = \langle xz - yz, x^2 - x^3\rangle \subseteq \Bbbk[x, y, z]$ yields the same graph as Example 3.4 except for two important differences:

- here there are many fewer edges in the xy-plane ($z = 0$); and
- every horizontal ($z = \text{constant} \geq 1$) slice of the big region is a separate connected component, in contrast to Example 3.4, where the entire big region was a single connected component corresponding to the monomials in the ideal.

$$I = \langle xz - yz, x^2 - x^3\rangle \subseteq \Bbbk[x, y, z] \implies G_I \approx$$



Only the second generator, $x^2 - x^3$, is capable of joining pairs of points in the xy-plane, and it does so parallel to the x-axis, starting at $x = 2$. Of course, $x^2 - x^3$ also joins pairs of points in the same manner at positive heights $z \geq 1$, but again only starting at $x = 2$. The first generator, $xz - yz$, has the same effect as $x - y$ did in Example 3.4, except that $xz - yz$ only joins pairs of lattice points at height $z = 1$ or more. In summary, every connected component of G_I in this example is contained in a single horizontal slice, and the horizontal slices of G_I are



$z = 0$ and $z \geq 1$.

The outline of the big region is drawn as a dotted line in the $z \geq 1$ slice illustration, which depicts only enough of the edges to elucidate its three connected components.

Let F be the part of $\mathbb{N}^3$ in the xy-plane, so $\mathfrak{p}_F = \langle z \rangle$. The ideal $(I : \mathbf{t}^F) = (I : z^\infty) = (I : z) = \langle x - y, x^2 - x^3 \rangle$ again has the property that every connected component of $G_{(I:z)}$ is contained in a single horizontal slice, but now all of these slices look like the $z \geq 1$ slices of G_I. Compare this to the statement of Lemma 4.3.

Theorem 4.5. *Fix a monomial prime $\mathfrak{p}_F = \langle \mathbf{t}^u \mid u \notin F \rangle$ in an affine semigroup ring $\Bbbk[Q]$ for a face $F \subseteq Q$. If $\mathfrak{p}_F$ is minimal over a binomial ideal $I \subseteq \Bbbk[Q]$ and $\overline{\pi}_0 G_{I[\mathbb{Z}F]}$ is the set of finite components of the graph $G_{I[\mathbb{Z}F]}$, then I has $\mathfrak{p}_F$-primary component*

$$(I : \mathbf{t}^F) + \langle \mathbf{t}^u \mid C_u \notin \overline{\pi}_0 G_{I[\mathbb{Z}F]} \rangle.$$

The exponents on the monomials in this primary component are precisely the elements of Q that lie in infinite connected components of the graph $G_{I[\mathbb{Z}F]}$.

Proof. This is [DMM10, Theorem 2.15]. The $\mathfrak{p}_F$-primary component of I is equal to the $\mathfrak{p}_F$-primary component of $(I : \mathbf{t}^F)$ because primary decomposition is preserved by localization (see [AM69, Proposition 4.9], for example), so we may as well assume that $I = (I : \mathbf{t}^F)$. It is elementary to check that F acts on the connected components. The action is semifree on the F-finite components, for if $f + u \sim g + u$ for some $f, g \in F$, then $u \sim n(f - g) + u$ for all $n \in \mathbb{N}$, whence $C_u \in \pi_0 G_{I[\mathbb{Z}F]}$ is infinite; and if $f + u \sim f + v$ then $u \sim v$, because $\mathbf{x}^f$ is a unit on $I[\mathbb{Z}F]$. It is also elementary, though nontrivial, to check that the kernel of the usual localization homomorphism $\Bbbk[Q]/I \to \Bbbk[Q]_{\mathfrak{p}_F}/I_{\mathfrak{p}_F}$—inverting all polynomials outside of $\mathfrak{p}_F$, not just monomials—contains every monomial $\mathbf{t}^u$ for which $C_u \notin \overline{\pi}_0 G_{I[\mathbb{Z}F]}$ [DMM10, Lemmas 2.9 and 2.10]. Now note that $(I : \mathbf{t}^F) + \langle \mathbf{t}^u \mid C_u \notin \overline{\pi}_0 G_{I[\mathbb{Z}F]} \rangle$ is already primary by Theorem 3.23. ∎

Remark 4.6. The graph of $I[\mathbb{Z}F]$ has vertex set $Q[\mathbb{Z}F]$, which naturally contains Q. That is why, in Definition 3.21 and Theorem 4.5, it makes sense to say that an element of Q lies in a connected component of $G_{I[\mathbb{Z}F]}$.

Example 4.7. The $\mathfrak{p}_F$-primary component of the ideal I in Example 4.4 is the ideal I in Example 3.4: the differences that remain, after the localization operation in Example 4.4 is complete, are erased by lumping together the infinite connected components into a single monomial component.

Example 4.8. Starting with $(I : \mathbf{t}^F)$ in Theorem 4.5, it is not enough to throw in the monomials whose exponents lie in infinite components of $G_{(I:\mathbf{t}^F)}$; i.e., a connected component of $G_{(I:\mathbf{t}^F)}$ could be finite but nonetheless equal to the intersection with Q of an infinite component of $G_{I[\mathbb{Z}F]}$. This occurs for $I = \langle xz - yz\rangle \subseteq \Bbbk[x,y,z]$, with F being the xy-coordinate plane of $Q = \mathbb{N}^3$, so $\mathfrak{p}_F = \langle z\rangle$. Every connected component of G_I is finite, even though $I = (I : \mathbf{t}^F)$ is not primary. When x and y are inverted to form $I[\mathbb{Z}F]$, the components at height $z \geq 1$ become cosets of the line spanned by $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$, whose intersections with $\mathbb{N}^3$ are bounded. Hence the $\langle z\rangle$-primary component of I is $I + \langle z\rangle = \langle z\rangle$, as is clear from the primary decomposition $I = \langle z\rangle \cap \langle x - y\rangle$.

4.2 Primary Components for Arbitrary Given Associated Primes

For this subsection, fix a binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ in a polynomial ring with a binomial associated prime $I_{\rho,J}$ for some character $\rho : L \to \Bbbk^*$ defined on a saturated sublattice $L \subseteq \mathbb{Z}^J$. Now it is important to assume that the field $\Bbbk$ is algebraically closed of characteristic 0, for these hypotheses are crucial to the truth of Theorem 3.20, and that theorem is the tool that reduces the current general situation to the special case in Sect. 4.1. The logic is as follows.

Every binomial $I_{\rho,J}$-primary ideal contains I_ρ by Theorem 3.20. Since we are trying to construct a binomial $I_{\rho,J}$-primary component of I starting from I itself, the first step should therefore be to enlarge I by throwing in I_ρ. The following is a formal statement.

Proposition 4.9. *Fix a binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ with $\Bbbk$ algebraically closed of characteristic* 0*. If P is any binomial $I_{\rho,J}$-primary component of I, then P is the preimage in $\Bbbk[\mathbf{x}]$ of a binomial $(I_{\rho,J}/I_\rho)$-primary component of $(I + I_\rho)/I_\rho \subseteq \Bbbk[\mathbf{x}]/I_\rho$.*

An alternate phrasing makes the point of considering the quotient $\Bbbk[\mathbf{x}]/I_\rho$ in Proposition 4.9 clearer.

Proposition 4.10. *If P is an $I_{\rho,J}$-primary binomial ideal in $\Bbbk[\mathbf{x}]$, with $\Bbbk$ algebraically closed of characteristic* 0*, then the image of P in the affine semigroup ring $\Bbbk[Q] = \Bbbk[\mathbf{x}]/I_\rho$ is a binomial ideal $\mathfrak{p}_F$-primary to the monomial prime $\mathfrak{p}_F = I_{\rho,J}/I_\rho$ in $\Bbbk[Q]$.*

Proof. This is an immediate consequence of Theorem 3.20 and Theorem 2.22 along with Corollary 2.24: the affine semigroup Q is $(\mathbb{N}^J/L) \times \mathbb{N}^{\bar{J}}$, and the face F is the copy of $\mathbb{N}^J/L = (\mathbb{N}^J/L) \times \{0\}$ in Q. ◻

Thus the algebra of general binomial associated primes for polynomial rings is lifted from the algebra of monomial associated primes in affine semigroup rings. The final step is isolating how the combinatorics, namely Theorem 3.23, lifts. Since the algebra of quotienting $\Bbbk[\mathbf{x}]$ modulo I_ρ corresponds to the quotient of $\mathbb{N}^n$ modulo L, we expect the lifted finiteness conditions to involve cosets of L.

Definition 4.11. A subset of $\mathbb{N}^n$ is *L-bounded* for a sublattice $L \subseteq \mathbb{Z}^n$ if the subset is contained in a finite union of cosets of L.

Corollary 4.12. *Fix a binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ with $\Bbbk$ algebraically closed of characteristic 0. If $I_{\rho,J}$ is minimal over I, then the $I_{\rho,J}$-primary component of I is*

$$P = I' + \langle \mathbf{x}^u \mid C_u \in \pi_0 G_{I'[\mathbb{Z}^J]} \text{ is not } L\text{-bounded}\rangle,$$

where $I'[\mathbb{Z}^J]$ is the localization along $\mathbb{N}^J$, and I' is defined, using $\mathbf{x}_J = \prod_{j\in J} x_j$, to be

$$I' = \big((I+I_\rho) : \mathbf{x}_J^\infty\big) = (I+I_\rho)[\mathbb{Z}^J] \cap \Bbbk[\mathbf{x}].$$

If $I_{\rho,J}$ is associated to I but not minimal over I, then for any monomial ideal K containing a sufficiently high power of $\mathfrak{m}_J = \langle x_i \mid i \notin J\rangle$, an $I_{\rho,J}$-primary component of I is defined as P is, above, but using I'_K in place of I', where

$$I'_K = \big((I+I_\rho+K) : \mathbf{x}_J^\infty\big).$$

Proof sketch. This is [DMM10, Theorem 3.2]. The key is to lift the monomial minimal prime case for affine semigroup rings in Theorem 4.5 to the current binomial associated prime case in polynomial rings using Propositions 4.9 and 4.10. For an embedded prime $I_{\rho,J}$, one notes that any given $I_{\rho,J}$-primary component of I must contain a sufficiently high power of $\mathfrak{m}_J$, so it is logical to begin the search for an $I_{\rho,J}$-primary component by simply throwing such monomials along with I_ρ into I. But then $I_{\rho,J}$ is minimal over the resulting ideal $I+I_\rho+K$, so the minimal prime case applies. □

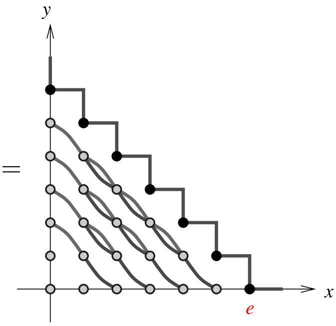
The definition of P in the theorem says that G_P has two types of connected components: the ones that are L-bounded upon localization along $\mathbb{Z}^J$, and the connected component consisting of exponents on monomials in P. The theorem says that G_P shares all but its monomial component with the graph $G_{I'}$, and that the other connected components of $G_{I'}$ fail to remain L-bounded upon localization along $\mathbb{Z}^J$. In the case where $I_{\rho,J}$ is an embedded prime, the graph-theoretic explanation is that $G_{I'}$ has too many connected components that remain L-bounded upon localization; in fact, there are infinitely many $\mathbb{N}^J$-orbits. The hack of adding K throws all but finitely many $\mathbb{N}^J$-orbits into the L-infinite "big monomial" connected component.

Example 4.13. A primary decomposition of the ideal $I = \langle x^2 - xy, xy - y^2\rangle$ was already given in Example 3.17. Analyzing it from the perspective of Corollary 4.12 completes the heuristic insight.

First let $I_{\rho,J} = \langle x - y\rangle$, so $J = \{1,2\}$ and $\rho : L \to \Bbbk^*$ is trivial on the lattice L generated by $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$. Then $I' = I + \langle x-y\rangle$ is already prime. Passing from I to I' has the sole effect of joining the two isolated points (the basis vectors) on the axes together.

Now let $I_{\rho,J} = \langle x,y\rangle$, so $J = \varnothing$ and ρ is the trivial (only) character defined on $L = \mathbb{Z}^J = \{0\}$. Every connected component of G_I in Example 3.17 remains L-bounded upon localization along $\mathbb{Z}^J$, but there are infinitely many such components. Choosing $K = \langle x,y\rangle^e$, so that $I'_K = I' + \langle x,y\rangle^e$, kills off all but finitely many, to get

$$P = \langle x^2 - xy, xy - y^2\rangle + \langle x,y\rangle^e \subseteq \Bbbk[x,y] \implies G_P =$$



In particular, taking $e = 2$ recovers the primary decomposition from Example 3.17.

4.3 Finding Associated Primes Combinatorially

The constructions of binomial primary components in previous sections assume that a monomial or binomial associated prime of a binomial ideal has been given. To conclude the discussion of primary decomposition of binomial ideals, it remains to examine the set of associated primes. The existence of binomial primary decompositions hinges on a fundamental result, due to Eisenbud and Sturmfels [ES96, Theorem 6.1], that was a starting point for all investigations involving primary decomposition of binomial ideals.

Theorem 4.14. *Every associated prime of a binomial ideal in $\Bbbk[\mathbf{x}]$ is a binomial prime if the field $\Bbbk$ is algebraically closed.*

Although the statement is for polynomial rings, a simple reduction implies the existence of binomial primary decomposition in the generality of monoid algebras as defined in Sect. 3.2, given the construction of binomial primary components.

Corollary 4.15. *Fix a finitely generated commutative monoid Q and an algebraically closed field $\Bbbk$ of characteristic 0. Every binomial ideal I in $\Bbbk[Q]$ admits a binomial primary decomposition: $I = P_1 \cap \cdots \cap P_r$ for binomial ideals $P_1,\ldots,P_r$.*

Proof. Choose a presentation $\mathbb{N}^n \twoheadrightarrow Q$. The kernel of the induced presentation $\Bbbk[\mathbf{x}] \twoheadrightarrow \Bbbk[Q]$ is a binomial ideal in $\Bbbk[\mathbf{x}]$. Therefore the preimage of I in $\Bbbk[\mathbf{x}]$ is a binomial ideal I'. The image in $\Bbbk[Q]$ of any binomial primary decomposition of I' is a binomial primary decomposition of I. Therefore it suffices to prove the case where $Q = \mathbb{N}^n$ and $I' = I$. Since every associated prime of I is binomial by Theorem 4.14, the result follows from Corollary 4.12. $\square$

Corollary 4.15 is stated only for characteristic 0 to demonstrate the connection between prior results in this survey. However, the restriction is unnecessary.

Theorem 4.16. *Corollary 4.15 holds for fields of positive characteristic, as well.*

Proof. For polynomial rings this is [ES96, Theorem 7.1], and the case of general monoids Q follows by the argument in the proof of Corollary 4.15. □

What's missing in positive characteristic cases is combinatorics of primary ideals.

Open Problem 4.17. Characterize primary binomial ideals and primary components of binomial ideals combinatorially in positive characteristic.

Note, however, that a solution to this problem would still not say how to discover—from the combinatorics—which primes are associated. The same is true in characteristic 0. Thus Corollary 4.12 is unsatisfactory for two reasons:

- it requires strong hypotheses on the field $\Bbbk$; and
- it assumes we know which primes $I_{\rho,J}$ are associated to I.

Fortunately, there is a combinatorial, lattice-point method to recognize associated primes—or at least, to reduce the recognition to a finite problem. The main point is Theorem 4.26: the combinatorics of the graph G_I can be used to construct a decomposition of I as an intersection of "primary-like" binomial ideals in a manner requiring no hypotheses on the characteristic or algebraic closure of the base field. The statement employs some additional concepts.

Definition 4.18. A subset of $\mathbb{Z}^n$ is *J-bounded* if it intersects only finitely many cosets of $\mathbb{Z}^J$ in $\mathbb{Z}^n$.

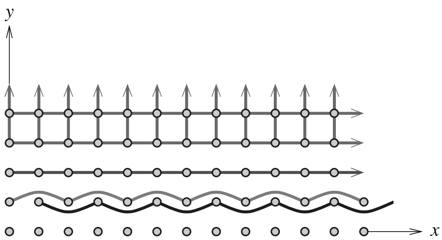
Lemma 4.19. *If $I \subseteq \Bbbk[\mathbf{x}]$ is a binomial ideal, then $\mathbb{Z}^J$ acts on the set of J-bounded components of the graph $G_{I[\mathbb{Z}^J]}$ on $\mathbb{Z}^J \times \mathbb{N}^{\bar{J}}$ induced by localizing I along $\mathbb{N}^J$.*

Proof. In fact, $\mathbb{Z}^J$ acts on all of the connected components, because the Laurent monomials $\mathbf{x}^u$ for $u \in \mathbb{Z}^J$ are units modulo $I[\mathbb{Z}^J] = I[\mathbb{Z}F]$ for the face $F = \mathbb{N}^J$. □

Definition 4.20. A *witness* for a sublattice $L \subseteq \mathbb{Z}^J$ *potentially associated* to I is any element in a J-bounded connected component of $G_{I[\mathbb{Z}^J]}$ whose stabilizer is L.

Example 4.21. The binomial ideal

$$I = \langle y - x^2y, y^2 - xy^2, y^3 \rangle \subseteq \Bbbk[x,y] \quad \longleftrightarrow$$



induces the depicted congruence. Its potentially associated lattices are all contained in $\mathbb{Z} = \mathbb{Z}^{\{1\}}$, parallel to the x-axis. The lattices are generated by 0, by $\left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$, and by $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$.

The subset J is part of the definition of potentially associated sublattice; it is not enough to specify L alone. The notion of *associated lattice*, without the adverb "potentially", would require further discussion of primary decomposition of congruences on monoids; see the definition of associated lattice in [KM10]. That said, the set of potentially associated lattices, which contains the set of associated ones, suffices for the purposes here, although sharper results could be stated with the more precise notion.

Proposition 4.22. *Every binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ has finitely many potentially associated lattices $L \subseteq \mathbb{Z}^J$. If $K = (I : \mathbf{x}^u) \subseteq \Bbbk[\mathbf{x}]$ is the annihilator of $\mathbf{x}^u$ in $\Bbbk[\mathbf{x}]/I$ for a witness $u \in \mathbb{N}^n$, then $K[\mathbb{Z}^J] + \mathfrak{m}_J = I_{\sigma,J}[\mathbb{Z}^J]$ for a uniquely determined* witness character $\sigma : L \to \Bbbk^*$. *Given I, each $L \subseteq \mathbb{Z}^J$ determines finitely many witness characters.*

Proof. This is proved in [KM10] on the way to the existence theorem for combinatorial mesoprimary decomposition. The finiteness of the set of potentially associated lattices traces back to the noetherian property for congruences on finitely generated commutative monoids. The conclusion concerning K is little more than the characterization of binomial ideals in Laurent polynomial rings [ES96, Theorem 2.1]. The finiteness of the number of witness characters occurs because witnesses for $L \subseteq \mathbb{Z}^J$ with distinct witness characters are forced to be incomparable in $\mathbb{N}^n$. □

In Proposition 4.22, the domain L of the character σ appearing in $I_{\sigma,J}$ need not be saturated (see Definition 2.23), and no hypotheses are required on the field $\Bbbk$.

Deducing combinatorial statements about associated primes or primary decompositions of binomial ideals is often most easily accomplished by reducing to the case of ideals with the simplest possible structure in this regard.

Definition 4.23. A binomial ideal with a unique potentially associated lattice is called *mesoprimary*. A *mesoprimary decomposition* of a binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ is an expression of I as an intersection of finitely many mesoprimary binomial ideals.

Example 4.24. Primary binomial ideals in polynomial rings over algebraically closed fields of characteristic 0 are mesoprimary. That is the content of Theorem 3.23 for such fields, given Proposition 4.10. More precisely, combinatorics of mesoprimary ideals is just like that of primary ideals, except that instead of an affine semigroup acting semifreely, an arbitrary finitely generated cancellative monoid acts semifreely; see the characterizations of mesoprimary congruences in [KM10].

Definition 4.23 stipulates constancy of the combinatorics, not the arithmetic—meaning the witness characters—but the arithmetic constancy is automatic.

Lemma 4.25. *If I is a mesoprimary ideal, then the witnesses for the unique potentially associated lattice all share the same witness character.*

Proof. This follows from the same witness incomparability that appeared in the proof of Proposition 4.22. The statement is equivalent to one direction of the characterization of mesoprimary binomial ideals as those with precisely one associated mesoprime; see [KM10], where a complete proof can be found. □

Theorem 4.26. *Every binomial ideal $I \subseteq \Bbbk[\mathbf{x}]$ admits a mesoprimary decomposition in which the unique associated lattice and witness character of each mesoprimary component is potentially associated to I.*

Proof. This is a weakened form of the existence theorem for combinatorial mesoprimary decomposition in [KM10]. □

The power of Theorem 4.26 lies in the crucial conceit that the combinatorics of the graph G_I controls everything, so the lattices associated to the mesoprimary components are severely restricted. Over an algebraically closed field of characteristic 0, for instance, every primary decomposition is a mesoprimary decomposition, but usually the lattices are not associated to I. This is the case for a lattice ideal I_L, as long as the lattice L is not saturated: the ideal I_L is already mesoprimary, but the associated lattice of every associated prime is the *saturation* $L_{\text{sat}} = (L \otimes_{\mathbb{Z}} \mathbb{Q}) \cap \mathbb{Z}^n$, the smallest saturated sublattice of $\mathbb{Z}^n$ containing L. In general, the combinatorial control is what allows Theorem 4.26 to be devoid of hypotheses on the field.

As in any expression of an ideal I as an intersection of larger ideals, information about associated primes of I can just as well be read off of the intersectands. For mesoprimary decompositions this is especially effective because primary decomposition of mesoprimary ideals [KM10] is essentially as simple as that of lattice ideals [ES96, Corollary 2.5]. In particular, when the field is algebraically closed, potentially associated lattices yield associated primes by way of saturation. The point is that only finitely many characters $L_{\text{sat}} \to \Bbbk^*$ restrict to a given fixed character $L \to \Bbbk^*$. In fact, when $\Bbbk$ is algebraically closed, these characters are in bijection with the finite group $\operatorname{Hom}(L_{\text{sat}}/L, \Bbbk^*)$. Thus Theorem 4.26 reduces the search for associated primes of I to the combinatorics of the graph G_I, along with a minimal amount of arithmetic.

Corollary 4.27. *If $\Bbbk$ is algebraically closed, then every associated prime of $I \subseteq \Bbbk[\mathbf{x}]$ is $I_{\rho,J}$ for some character $\rho : L_{\text{sat}} \to \Bbbk^*$ whose restriction to L is one of the finitely many witness characters defined on a potentially associated lattice $L \subseteq \mathbb{Z}^J$ of I.*

Proof. Every associated prime of I is associated to a mesoprime in the decomposition from Theorem 4.26. Now apply either the primary decomposition of mesoprimary ideals [KM10] or the witness theorem for cellular ideals [ES96, Theorem 8.1], using the fact that mesoprimary ideals are cellular [KM10]. □

Remark 4.28. In the special case where I is *cellular*, meaning that every variable is either nilpotent or a nonzero divisor modulo I, Corollary 4.27 coincides with [ES96, Theorem 8.1]. Every binomial ideal in any polynomial ring over any field is an intersection of cellular ideals [ES96, Theorem 6.2], with at most one cellular component for each subset $J \subseteq \{1,\dots,n\}$, so it suffices for many purposes to understand the combinatorics of cellular ideals. (Theorem 4.26 strengthens this approach, since mesoprimary ideals are cellular and their combinatorics is substantially simpler.) The way witnesses and witness characters are defined above, however, it is not quite obvious that the information extracted from witnesses for the original ideal I and those for its cellular components coincides. That this is indeed the case constitutes a key ingredient proved in preparation for the existence theorem for combinatorial mesoprimary decomposition in [KM10].

Exercise 4.29. The ideal $I = \langle xz - yz, x^2 - x^3 \rangle \subseteq \Bbbk[x,y,z]$ from Example 4.4 has primary decomposition $I = \langle x-y, x^2, xy, y^2 \rangle \cap \langle x-1, y-1 \rangle \cap \langle x^2, z \rangle \cap \langle x-1, z \rangle$. The reader is invited to find all associated lattices $L \subseteq \mathbb{Z}^J$ of I and match them to the associated primes of I. Then the reader can verify, using Corollary 4.12, that the given primary decomposition of I really is one. Hint: take $J \in \big\{\{3\}, \{1,2,3\}, \{2\}, \{1,2\}\big\}$.

The upshot of Sects. 2–4 is that the lattice-point combinatorics of congruences on monoids lifts to combinatorics of monomial and binomial primary and mesoprimary decompositions of binomial ideals in monoid algebras. From there, binomial primary decomposition is a small arithmetic step, having to do with group characters for finitely generated abelian groups.

Applications

5 Hypergeometric Series

The idea for lattice-point methods in binomial primary decomposition originated in the study of hypergeometric systems of differential equations, particularly their series solutions. The literature on these systems and series is so vast—owing to its connections with physics, numerical analysis, combinatorics, probability, number theory, complex analysis, and algebraic geometry—that one section in a survey lacks the ability to lend proper perspective. Therefore, the goal of this section is to make a beeline for the connections to binomial primary decomposition, with just enough background along the way to allow the motivations and conclusions to shine through. Much of the exposition is borrowed from [DMM07, DMM10′], sometimes nearly verbatim. The extended abstract [DMM07] presents a broader, more complete historical overview.

5.1 Binomial Horn Systems

Horn systems are certain sets of linear partial differential equations with polynomial coefficients. Their development grew out of the ordinary univariate hypergeometric theory going back to Gauss (see [SK85], for example) and Kummer [Kum1836], through the bivariate versions of Appell, Horn, and Mellin [App1880, Hor1889, Hor31, Mel21]. These formulations had no apparent connection to binomials, but through a relatively simple change of variables, Gelfand, Graev, Kapranov, and Zelevinsky brought binomials naturally into the picture [GGZ87, GKZ89].

The data required to write down a binomial Horn system consist of a basis for a sublattice $L \subseteq \mathbb{Z}^n$ and a homomorphism $\beta : \mathbb{Z}^n/L \to \mathbb{C}$. Focus first on the basis, which is traditionally arranged in an integer $n \times m$ matrix B, where $m = \operatorname{rank}(L)$. If $\mathbf{b} \in \mathbb{Z}^n$ is a column of B, then $\mathbf{b}$ determines a binomial:

$$\mathbf{b} \in \mathbb{Z}^n \ \rightsquigarrow\ \partial^{\mathbf{b}_+} - \partial^{\mathbf{b}_-} \in \mathbb{C}[\partial] = \mathbb{C}[\partial_1, \ldots, \partial_n],$$

where $\mathbf{b} = \mathbf{b}_+ - \mathbf{b}_-$ expresses the vector $\mathbf{b}$ as a difference of nonnegative vectors with disjoint support. Elements of the polynomial ring $\mathbb{C}[\partial]$ are to be viewed as differential operators on functions $\mathbb{C}^n \to \mathbb{C}$. Therefore the matrix B determines a system of m binomial differential operators, one for each column. The interest is a priori in solutions to differential systems, not really the systems themselves, so it is just as well

$$I(B) = \langle \partial^{\mathbf{b}_+} - \partial^{\mathbf{b}_-} \mid \mathbf{b} = \mathbf{b}_+ - \mathbf{b}_- \text{ is a column of } B \rangle \subseteq \mathbb{C}[\partial]$$

generated by these binomials, because any function annihilated by the m binomials is annihilated by all of $I(B)$.

Example 5.1. In the 0123 situation from Examples 2.2 and 2.16, with variables $\partial = \partial_1, \partial_2, \partial_3, \partial_4$ instead of a, b, c, d or x_1, x_2, x_3, x_4 yields $I(B) = \langle \partial_1\partial_3 - \partial_2^2, \partial_2\partial_4 - \partial_3^2 \rangle$.

Example 5.2. In the $\substack{1100\\0111}$ situation from Examples 2.3 and 2.17 with ∂ variables, $I(B) = \langle \partial_1\partial_3 - \partial_2, \partial_1\partial_4 - \partial_2 \rangle$.

The set of homomorphisms $\mathbb{Z}^n/L \to \mathbb{C}$ is a complex vector space $\operatorname{Hom}(\mathbb{Z}^n/L, \mathbb{C})$ of dimension $d := n - m$. Choosing a basis for this vector space is the same as choosing a basis for $(\mathbb{Z}^n/L) \otimes_{\mathbb{Z}} \mathbb{C}$, which is the same as choosing a $d \times n$ matrix A with $AB = 0$. Let us now, once and for all, fix such a matrix A with entries a_{ij} for $i = 1, \ldots, d$ and $j = 1, \ldots, n$. The situation is therefore just as it was in Examples 2.16 and 2.17, and our homomorphism $\mathbb{Z}^n/L \to \mathbb{C}$ becomes identified with a complex vector $\beta \in \mathbb{C}^d$. Together, A and β determine d differential operators $E_1 - \beta_1, \ldots, E_d - \beta_d$, where

$$E_i = a_{i1}x_1\partial_1 + \cdots + a_{in}x_n\partial_n.$$

Note that $a_{ij}x_j\partial_j$ is the operator on functions $f(x_1, \ldots, x_n) : \mathbb{C}^n \to \mathbb{C}$ that takes the partial derivative with respect to x_j and multiplies the resulting function by $a_{ij}x_j$.

Definition 5.3. The *binomial Horn system* $H(B, \beta)$ is the system

$$\begin{aligned} I(B)f &= 0 \\ E_1 f &= \beta_1 f \\ &\vdots \\ E_d f &= \beta_d f \end{aligned}$$

of differential equations on functions $f(\mathbf{x}) : \mathbb{C}^n \to \mathbb{C}$ determined by the *lattice basis ideal* $I(B)$ and the *Euler operators* $E_1 - \beta_1, \ldots, E_d - \beta_d$.

The goal is to find, characterize, or otherwise understand the solutions to $H(B, \beta)$.

Example 5.4. In the 0123 case from Example 2.16, $H(B,\beta)$ has lattice basis part

$$\begin{aligned} I(B)f = 0 \Leftrightarrow (\partial_1\partial_3 - \partial_2^2)f &= 0 \\ \text{and } (\partial_2\partial_4 - \partial_3^2)f &= 0 \end{aligned}$$

and the Euler operators yield the following equations:

$$\begin{aligned} (x_1\partial_1 + x_2\partial_2 + \ \ x_3\partial_3 + \ \ x_4\partial_4)f &= \beta_1 f \\ (\qquad\quad x_2\partial_2 + 2x_3\partial_3 + 3x_4\partial_4)f &= \beta_2 f. \end{aligned}$$

Example 5.5. In the ${}^{1100}_{0111}$ case from Example 5.2, $H(B,\beta)$ has lattice basis part

$$\begin{aligned} I(B)f = 0 \Leftrightarrow (\partial_1\partial_3 - \partial_2)f &= 0 \\ \text{and } (\partial_1\partial_4 - \partial_2)f &= 0 \end{aligned}$$

and the Euler operators yield the following equations:

$$\begin{aligned} (x_1\partial_1 + x_2\partial_2 \qquad\qquad\qquad\quad\ )f &= \beta_1 f \\ (\qquad\quad x_2\partial_2 + x_3\partial_3 + x_4\partial_4)f &= \beta_2 f. \end{aligned}$$

Since Horn systems are linear, their solution spaces are complex vector spaces. More precisely, the term *solution space* in what follows means the vector space of local holomorphic solutions defined in a neighborhood of a (fixed, but arbitrary) point in $\mathbb{C}^n$ that is nonsingular for the Horn system.

Example 5.6. In the 0123 case from Example 5.4, for any parameter vector β, the Puiseux monomial $f = x_1^{\beta_1/3}x_4^{\beta_2/3}$ is a solution of $H(B,\beta)$. Indeed,

$$\partial_1\partial_3(f) = \partial_2(f) = \partial_2\partial_4(f) = \partial_3^2(f) = 0,$$

so $I(B)f = 0$, and $(E_1 - \beta_1)f = (E_2 - \beta_2)f = 0$ because

$$\begin{aligned} x_1\partial_1(f) &= (\beta_1 - \tfrac{1}{3}\beta_2)f \\ x_2\partial_2(f) &= 0 \\ x_3\partial_3(f) &= 0 \\ x_4\partial_4(f) &= \tfrac{1}{3}\beta_2 f. \end{aligned}$$

Erdélyi produced this solution and similar ones in other examples [Erd50], but he furnished no explanation for why it should exist or how he found it. In this particular example, the Horn system has, in addition to the Puiseux monomial f, three linearly independent fully supported solutions, in the following sense.

Definition 5.7. A Puiseux series solution f to a Horn system $H(B,\beta)$ is *fully supported* if there is a normal affine semigroup Q of dimension m and a vector $\gamma \in \mathbb{C}^n$ such that the translate $\gamma + Q$ consists of vectors that are exponents on monomials with nonzero coefficient in f.

The integer $m = \operatorname{rank}(L)$ in the definition is the maximum possible: the Euler operator equations impose homogeneity on Puiseux series solutions, meaning that every solution must be supported on a translate of $L \otimes_{\mathbb{Z}} \mathbb{C}$. In fact, the translate is by any vector $\gamma \in \mathbb{C}^n$ satisfying $A\gamma = \beta$.

Questions 5.8. Consider the family of Horn systems B determines with varying β.

(a) For which parameters β does $H(B,\beta)$ have finite-dimensional solution space?
(b) What is a combinatorial formula for the minimum solution space dimension, over all possible choices of the parameter β?
(c) Which β are generic in the sense that the minimum dimension is attained?
(d) Which monomials occur in solutions expanded as series centered at the origin?

These questions arise from classical work done in the 1950s, such as Erdélyi's, and earlier. Implicit in Question 3 is that the dimension of the solution space rises above the minimum for only a "small" subset of parameters β.

Example 5.9. In the $\substack{1100\\0111}$ case from Example 5.5, if $\beta_1 = 0$, then any (local holomorphic) bivariate function $f(x_3,x_4)$ satisfying $x_3\partial_3 f + x_4\partial_4 f = \beta_2 f$ is a solution of the Horn system $H(B,\beta)$. The space of such functions is infinite-dimensional; in fact, it has uncountable dimension, as it contains all Puiseux monomials $x_3^{w_3}x_4^{w_4}$ with $w_3, w_4 \in \mathbb{C}$ and $w_3 + w_4 = \beta_2$. When $\beta_1 \neq 0$, the solution space has finite dimension.

The $\substack{1100\\0111}$ example has vast numbers of linearly independent solutions expressible as Puiseux series with small support, but only for special values of β. In contrast, in the 0123 case there are many fewer series solutions of small support, but they appear for arbitrary values of β. This dichotomy is central to the interactions of Horn systems with binomial primary decomposition.

5.2 True Degrees and Quasidegrees of Graded Modules

The commutative algebraic version of the dichotomy just mentioned arises from elementary (un)boundedness of Hilbert functions of A-graded modules (recall Definition 2.12, Example 2.13, and Lemma 3.15); see Definition 5.16. For the remainder of this section, fix a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d whose affine semigroup $\mathbb{N}A \subseteq \mathbb{Z}^d$ is pointed (Definition 6.10).

Lemma 5.10. *A binomial ideal $I \subseteq \mathbb{C}[\partial]$ is A-graded if and only if it is generated by binomials $\mathbf{x}^{\mathbf{u}} - \lambda\mathbf{x}^{\mathbf{v}}$ for which $A\mathbf{u} = A\mathbf{v}$.* □

It is of course not necessary—and it almost never happens—that every binomial of the form $\mathbf{x}^{\mathbf{u}} - \lambda\mathbf{x}^{\mathbf{v}}$ with $A\mathbf{u} = A\mathbf{v}$ lies in I.

Example 5.11. $I = I(B)$ is always A-graded when B is a matrix for $\ker(A)$.

The set of degrees where a graded module is nonzero should, for the purposes of the applications to Horn systems, be considered geometrically.

Definition 5.12. For any A-graded module M,

$$\operatorname{tdeg}(M) = \{\alpha \in \mathbb{Z}^d \mid M_\alpha \neq 0\}$$

is the set of *true degrees* of M. The set of *quasidegrees* of M is

$$\operatorname{qdeg}(M) = \overline{\operatorname{tdeg}(M)},$$

the Zariski closure in $\mathbb{C}^d$ of the true degree set of M.

The Zariski closure here warrants some discussion. By definition, the *Zariski closure* of a subset $T \subseteq \mathbb{C}^d$ is the largest set $\overline{T}$ of points in $\mathbb{C}^d$ such that every polynomial vanishing on T also vanishes on $\overline{T}$. All of the sets T that we shall be interested in are sets of lattice points in $\mathbb{Z}^d \subseteq \mathbb{C}^d$. When T consists of lattice points on a line, for example, its Zariski closure $\overline{T}$ is the whole line precisely when T is infinite. When T is contained in a plane, its Zariski closure is the whole plane only if T is not contained in any algebraic curve in the plane. In the cases that interest us, $\overline{T}$ will always be a finite union of translates of linear subspaces of $\mathbb{C}^d$.

Lemma 5.13. *If M is a finitely generated A-graded module over $\mathbb{C}[\partial]$, then* $\operatorname{qdeg}(M)$ *is a finite arrangement of affine subspaces of $\mathbb{C}^d$, each one parallel to $\mathbb{Z}A_J$ for some $J \subseteq \{1,\ldots,n\}$, where A_J is the submatrix of A comprising the columns indexed by J.*

Proof. This is [DMM10′, Lemma 2.5]. Since M is noetherian, it has a finite filtration whose successive quotients are A-graded translates of $\mathbb{C}[\partial]/\mathfrak{p}$ for various A-graded primes $\mathfrak{p}$. Finiteness of the filtration implies that $\operatorname{qdeg}(M)$ is the union of the quasidegrees of these A-graded translates. But $\operatorname{tdeg}(\mathbb{C}[\partial]/\mathfrak{p})$ is the affine semigroup generated by $\{\deg(\partial_i) \mid \partial_i \notin \mathfrak{p}\}$, because $\operatorname{tdeg}(\mathbb{C}[\partial]/\mathfrak{p})$ is an integral domain. $\square$

Example 5.14. If $\mathfrak{p} = I_L + \mathfrak{m}_J$ is a binomial prime ideal, then $\operatorname{qdeg}(\mathbb{C}[\partial]/\mathfrak{p}) = \mathbb{C}A_J$ is the complex vector subspace of $\mathbb{C}^d$ spanned by the columns of A indexed by J.

Example 5.15. In the $\begin{smallmatrix}1100\\0111\end{smallmatrix}$ case from Examples 2.3 and 2.17, consider the A-graded module $M = \mathbb{C}[\partial_1,\partial_2,\partial_3,\partial_4]/I$ for varying ideals I. Then the true degree sets and quasidegree sets can be depicted as follows.

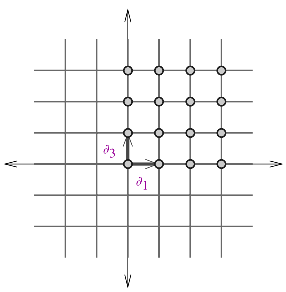
(a) When $I = I_A = \langle \partial_1\partial_3 - \partial_2, \partial_3 - \partial_4 \rangle$, so that $M \cong \mathbb{C}[\partial_1,\partial_3]$ by the homomorphism sending $\partial_2 \mapsto \partial_1\partial_3$ and $\partial_4 \mapsto \partial_3$,

$M = \mathbb{C}[\partial]/\langle \partial_1\partial_3 - \partial_2, \partial_3 - \partial_4 \rangle \cong \mathbb{C}[\partial_1,\partial_3]$

$\deg(\partial_1) = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\deg(\partial_3) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ $\qquad\longleftrightarrow$

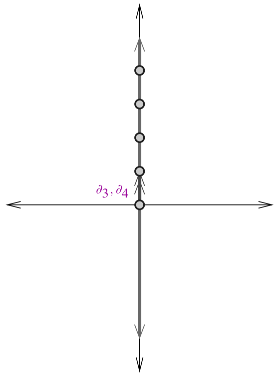
$\operatorname{tdeg}(M) = \mathbb{N}\left\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\} = \mathbb{N}A_{\{1,3\}}$

$\operatorname{qdeg}(M) = \mathbb{C}^2$



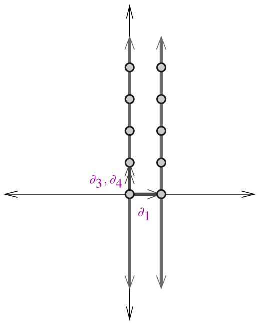
(b) When $I = \langle \partial_1, \partial_2 \rangle$, so that $M \cong \mathbb{C}[\partial_3, \partial_4]$ by the homomorphism sending $\partial_1 \mapsto 0$ and $\partial_2 \mapsto 0$,

$$M = \mathbb{C}[\partial]/\langle \partial_1, \partial_2 \rangle \cong \mathbb{C}[\partial_3, \partial_4]$$
$$\deg(\partial_3) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \text{ and } \deg(\partial_4) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \quad \longleftrightarrow$$
$$\mathrm{tdeg}(M) = \mathbb{N}\left\{\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\} = \mathbb{N}A_{\{3,4\}}$$
$$\mathrm{qdeg}(M) = \text{vertical axis in } \mathbb{C}^2$$



(c) When $I = \langle \partial_1^2, \partial_2 \rangle$, so that $M \cong \mathbb{C}[\partial_3, \partial_4] \oplus \partial_1 \mathbb{C}[\partial_3, \partial_4]$,

$$M = \mathbb{C}[\partial]/\langle \partial_1^2, \partial_2 \rangle \cong \mathbb{C}[\partial_3, \partial_4] \oplus \partial_1 \mathbb{C}[\partial_3, \partial_4]$$
$$\deg(\partial_1) = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \deg(\partial_3) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \text{ and } \deg(\partial_4) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \quad \longleftrightarrow$$
$$\mathrm{tdeg}(M) = \mathbb{N}A_{\{3,4\}} \cup \left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] + \mathbb{N}A_{\{3,4\}}\right)$$
$$\mathrm{qdeg}(M) = \text{vertical axis} \cup \left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] + \text{vertical axis}\right) \text{ in } \mathbb{C}^2$$



Definition 5.16. An A-graded prime ideal $\mathfrak{p} \subseteq \mathbb{C}[\partial]$ is

- *toral* if the Hilbert function $\mathbf{u} \mapsto \dim_{\mathbb{C}} (\mathbb{C}[\partial]/\mathfrak{p})_{\mathbf{u}}$ is bounded for $\mathbf{u} \in \mathbb{Z}^d$, and
- *Andean* if the Hilbert function is unbounded.

The adjective "Andean" indicates that Andean A-graded components sit like a high, thin mountain range on $\mathbb{Z}^d$ of unbounded elevation over cosets of lattices $\mathbb{Z}A_J$.

Example 5.17. Consider the situation from Example 5.15.

(a) The prime ideal I_A in Example 5.15.1 is toral because the Hilbert function of $\mathbb{C}[\partial]/I_A$ only takes the value 1 on the true degrees. By definition, the Hilbert function vanishes outside of the true degree set.
(b) The prime ideal $I = \langle \partial_1, \partial_2 \rangle$ in Example 5.15.2 is Andean because the Hilbert function of $\mathbb{C}[\partial]/I$ is unbounded: $\left[\begin{smallmatrix}0\\k\end{smallmatrix}\right] \mapsto k$.

Example 5.18. In the 0123 situation from Examples 2.2 and 2.16, the ideal $I_A = \langle \partial_1\partial_3 - \partial_2^2, \partial_2\partial_4 - \partial_3^2, \partial_1\partial_4 - \partial_2\partial_3 \rangle$ is toral under the A-grading. Indeed, for any matrix A the toric ideal I_A is toral under the A-grading by Lemma 3.15: the Q-graded Hilbert function of the algebra $\Bbbk[Q]$ takes the constant value 1 and is thus bounded.

Theorem 5.19. *Fix an A-graded ideal $I \subseteq \mathbb{C}[\partial]$. Given that $\mathbb{N}A$ is pointed, every associated prime of I is A-graded, and I admits a decomposition as an intersection of A-graded primary ideals. The intersection I_{Andean} of the primary components of I with Andean associated primes is well-defined.*

Proof. The A-graded conclusion on the associated primes is [MS05, Prop. 8.11]. The Andean part is well-defined because if $\mathfrak{p} \supseteq \mathfrak{q}$ for some Andean $\mathfrak{p}$ then $\mathfrak{q}$ is also Andean. ("The set of Andean primes is closed under going down.") □

Corollary 5.20. *If $I \subseteq \mathbb{C}[\partial]$ is an A-graded ideal, then I admits a decomposition*

$$I = I_{\text{toral}} \cap I_{\text{Andean}}$$

into toral and Andean parts, where I_{toral} is the intersection of the primary components of I with toral associated primes in any fixed primary decomposition of I.

Definition 5.21. The *Andean arrangement* of an ideal I is $\operatorname{qdeg}(\mathbb{C}[\partial]/I_{\text{Andean}})$.

The Andean arrangement is a union of affine subspaces of $\mathbb{C}^d$ by Lemma 5.13. It is well-defined by Theorem 5.19.

Example 5.22. The ${}^{1100}_{0111}$ lattice ideal $I(B) = \langle \partial_1\partial_3 - \partial_2, \partial_1\partial_4 - \partial_2 \rangle$ from Example 5.2 has primary decomposition

$$\begin{aligned} I(B) = {} & I_A \cap \langle \partial_1, \partial_2 \rangle \\ = {} & I_{\text{toral}} \cap I_{\text{Andean}}, \end{aligned}$$

with toral part $I_{\text{toral}} = I_A$ and Andean part $I_{\text{Andean}} = \langle \partial_1, \partial_2 \rangle$ by Example 5.17. Therefore the Andean arrangement of $I(B)$ is the thick vertical line in Example 5.15.2.

5.3 Counting Series Solutions

The distinction between toral and Andean primes provides the framework for the answers to Questions 5.8. Throughout the remainder of this section, fix a matrix $B \in \mathbb{Z}^{n\times m}$ of rank $m = n - d$ such that $AB = 0$. Assume that B is *mixed*, meaning that every nonzero integer vector in the span of the columns of B has two nonzero entries of opposite sign. The mixed condition is a technical hypothesis arising while constructing series solutions to $H(B, \beta)$; its main algebraic consequence is that it forces $\mathbb{N}A$ to be pointed.

The *multiplicity* of a prime ideal $\mathfrak{p}$ in an ideal I is, by definition, the length of the largest submodule of finite length in the localization $\mathbb{C}[\partial]_{\mathfrak{p}}/I_{\mathfrak{p}}$. This number is nonzero precisely when $\mathfrak{p}$ is associated to I. Combinatorially, when $\mathfrak{p}$ and I are binomial ideals, the multiplicity of $\mathfrak{p}$ in I counts connected components of graphs related to G_I, such as those in Corollary 4.12. For the purposes of Horn systems, the most relevant number is derived from multiplicities of prime ideals as follows.

Definition 5.23. The *multiplicity* $\mu(L,J)$ of a saturated sublattice $L \subseteq \mathbb{Z}^J$ is the product $\iota\mu$, where ι is the index $|L/(\mathbb{Z}B \cap \mathbb{Z}^J)|$ of the sublattice $\mathbb{Z}B \cap \mathbb{Z}^J$ in L, and μ is the multiplicity of the binomial prime ideal $I_L + \mathfrak{m}_J$ in the lattice ideal $I(B)$.

The factor $\iota = |L/(\mathbb{Z}B \cap \mathbb{Z}^J)|$ counts the number of partial characters $\rho : L \to \mathbb{C}^*$ for which $I_{\rho,J}$ is associated to $I(B)$. It is the penultimate combinatorial input required to count solutions to Horn systems. The final one is polyhedral.

Definition 5.24. For any subset $J \subseteq \{1,\ldots,n\}$, write $\mathrm{vol}(A_J)$ for the *volume* of the convex hull of A_J and the origin, normalized so a lattice simplex in $\mathbb{Z}A_J$ has volume 1.

Example 5.25. In the 0123 case, $\mathrm{vol}(A) = 3$, since the columns of A span $\mathbb{Z}^2$ and the convex hull of A with the origin is a triangle that is a union of three lattice triangles.

$\mathrm{vol}(A) = 3$ $\qquad$ $\mathrm{vol}(A_{\{1,4\}}) = 1$

When $J = \{1,4\}$, in contrast, $\mathrm{vol}(A_J) = 1$, since the first and last columns of A form a basis for the lattice (of index 3 in $\mathbb{Z}^2$) that they span.

Answers 5.26. The answers to Questions 5.8 for the systems $H(B,\beta)$ are as follows.

(a) The dimension is finite exactly when β lies in the Andean arrangement of $I(B)$.
(b) The generic (minimum) dimension is $\sum \mu(L,J) \cdot \mathrm{vol}(A_J)$, the sum being over all saturated $L \subseteq \mathbb{Z}^J$ such that $I_L + \mathfrak{m}_J$ is a toral binomial prime with $\mathbb{C}A_J = \mathbb{C}^d$.
(c) The minimum rank is attained precisely when β lies outside of an affine subspace arrangement determined by certain local cohomology modules, with the same flavor as (and containing) the Andean arrangement.
(d) If the configuration A lies in an affine hyperplane not containing the origin, and β is general, then the solution space of $H(B,\beta)$ has a basis containing precisely $\sum_J \mu(L,J) \cdot \mathrm{vol}(A_J)$ Puiseux series supported on finitely many cosets of L.

Example 5.27. To illustrate Answer 5.26.1 in the $\substack{1100\\0111}$ case, compare Example 5.9 to Example 5.22: the solution space has finite dimension precisely when the parameter lies off the vertical axis, which is the Andean arrangement in this case.

Example 5.28. In contrast, both associated primes are toral in the 0123 case, where

$$I(B) = I_A \cap \langle \partial_2, \partial_3 \rangle.$$

Indeed, the quotient of $\mathbb{C}[\partial]$ modulo each of these components is an A-graded affine semigroup ring, with the second component yielding $\mathbb{C}[\partial]/\langle \partial_2, \partial_3 \rangle \cong \mathbb{C}[\mathbb{N}A_{\{1,4\}}]$. It follows that the solution space has finite dimension for all parameters β.

On the other hand, Answer 5.26.2 is interesting in this 0123 case: Example 5.25 implies that $H(B,\beta)$ has generic solution space of dimension

$$3+1=\mu\big(\mathbb{Z}B,\{1,2,3,4\}\big)+\mu\big(\{0\},\{1,4\}\big),$$

with the first summand giving rise to solution series of full support, and the second summand giving rise to one solution series with finite support—that is, supported on finitely many cosets of $\{0\}$—by Answer 5.26.4. Compare Example 5.6.

Proof of Answers 5.26. These are some of the main results of [DMM10′], namely:

(a) Theorem 6.3.
(b) Theorem 6.10.
(c) Definition 6.9 and Theorem 6.10.
(d) Theorem 6.10, Theorem 7.14, and Corollary 7.25.

The basic idea is to filter $\mathbb{C}[\partial]/I(B)$ with successive quotients that are A-graded translates of $\mathbb{C}[\partial]/\mathfrak{p}$ for various binomial primes $\mathfrak{p}$. There is a functorial ("Euler–Koszul") way to lift this to a filtration of a corresponding D-module canonically constructed from $H(B,\beta)$. The successive quotients in this lifted filtration are *A-hypergeometric systems* of Gelfand, Graev, Kapranov, and Zelevinsky [GGZ87, GKZ89]. The solution space dimension equals the volume for such hypergeometric systems, and the factor $\mu(L,J)$ simply counts how many times a given such hypergeometric system appears as a successive quotient in the D-module filtration. That, together with series solutions constructed by GGKZ, proves Answers 2 and 4.

When a successive quotient is $\mathbb{C}[\partial]/\mathfrak{p}$ for an Andean prime $\mathfrak{p}$, the Euler operators span a vector space of too small dimension; they consequently fail to cut down the solution space to finite dimension: at least one "extra" Euler operator is needed. Without this extra Euler operator, its (missing) continuous parameter allows an uncountable family of solutions as in Example 5.9; this proves Answer 1. Answer 3 is really a corollary of the main results of [MMW05], and is beyond the scope of this survey. □

To explain Erdélyi's observation (Example 5.6), note that only $|\ker A/\mathbb{Z}B|\cdot\mathrm{vol}(A)$ many of the series solutions from Answer 5.26.4 have full support, where $\ker(A)=(\mathbb{Z}B)_{\mathrm{sat}}$ is the saturation of the image of B. The remaining solutions have smaller support. Most lattice basis ideals have associated primes other than $I_{\mathbb{Z}B}$ [HS00], so most Horn systems $H(B,\beta)$ have spurious solutions, whether they be of the toral kind (finite-dimensional, but small support, perhaps for special parameters β) or Andean kind (uncountable dimensional).

The "hyperplane not containing the origin" condition in Answer 5.26.4 amounts to a homogeneity condition on $I(B)$: the generators should be homogeneous under the standard $\mathbb{N}$-grading of $\mathbb{C}[\partial]$, as in Example 5.28. More deeply, this condition is equivalent to regular holonomicity of the corresponding D-module [SW08]. As soon as the support of a Puiseux series solution is specified, hypergeometric recursions determine the coefficients up to a global scalar.

The recursive rules governing the coefficients of series solutions to Horn hypergeometric systems force the combinatorics of lattice-point graphs upon binomial primary decomposition of lattice basis ideals, via the arguments in the proof of Answers 5.26. Granted the generality of Sects. 2–4, it subsequently follows that the questions as well as the answers work essentially as well for arbitrary A-graded binomial ideals, with little adjustment [DMM10′].

6 Combinatorial Games

Combinatorial games are two-player affairs in which the sides alternate moves, both with complete information and no element of chance. The germinal goal of Combinatorial Game Theory (CGT) is to find strategies for such games. After briefly reviewing the foundations and history of CGT using some key examples (Sect. 6.1), this section gives an overview of how to phrase the theory in terms of lattice points in polyhedra (Sect. 6.2). Exploring data structures for strategies via generating functions (Sect. 6.3) or misère quotients (Sect. 6.4) leads to conjectures and computational open problems involving binomial ideals and related combinatorics.

6.1 Introduction to Combinatorial Game Theory

There are many different ways to represent games and winning strategies by combinatorial structures. To understand their formal definitions, it is best to have in mind some concrete examples.

Example 6.1. The quintessential combinatorial game is NIM. The players—you and I, say—are presented with a finite number of heaps of beans, such as

when there are three heaps, of sizes 3, 7, and 4. Any finite set of heaps is a *position* in the game of NIM. The game is played by alternating turns, where each turn consists of picking one of the heaps and removing at least one bean from it. For instance, if you play first, then you could remove one bean from the 3-heap, or three beans from the 4-heap, or all of the beans from the 7-heap, to get one of the following positions:

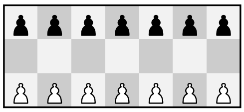
 or  or 

The goal of the game is to play last. As it turns out, if I play first in the 3–7–4 game, then you can always force a win by ensuring that you play last. How? Take the *nim sum* of the heap sizes: express each heap size in binary and add these binary numbers digit by digit, as elements of the field $\mathbb{F}_2$ of cardinality 2:

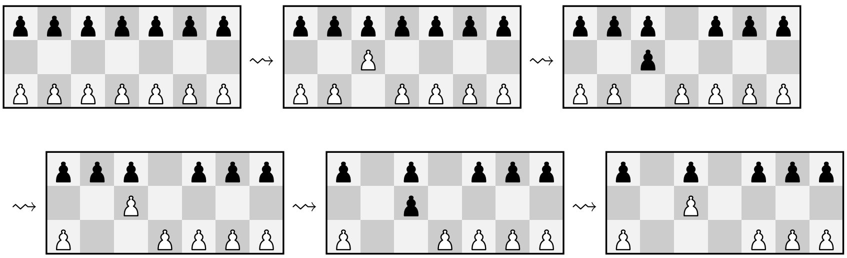
$$\begin{array}{r} 1\ 1 \\ 1\ 1\ 1 \\ 1\ 0\ 0 \\ \hline 0\ 0\ 0 \end{array}$$

Whatever move I make will alter only one of the summands and hence will leave a nonzero nim sum, at which point you can always remove beans from a heap to reset the nim sum to zero; I can't win because removing the last bean leaves a zero nim sum. This general solution to NIM is one of the oldest formal contributions to combinatorial game theory [Bou1902]. More general "heap games", in which players take beans from heaps according to specified rules, constitute a core class of examples for the theory.

Example 6.2. CHESS and its variants give rise to a rich bounty of combinatorial games. One of the most famous, other than CHESS itself, is DAWSON'S CHESS, played on a $3 \times d$ board with an initial position of opposing pawns facing one another with a blank rank (row) between [Daw34]. When $d = 7$, the initial position is:



Moves in DAWSON'S CHESS are as usual for CHESS pawns. The only additional rule is that a capture must be made if one is possible. For example, if white moves first (as usual) and chooses to push the third pawn, then a game might start as follows:



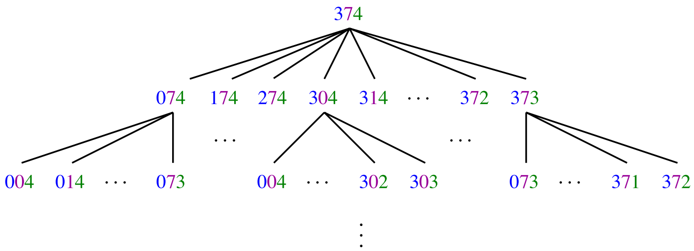
And now it is black's turn; with no captures available, black is allowed to push any pawn in file (column) 1, 5, 6, or 7 down to the middle rank. Another bloodbath ensues, and then it is white's turn to push a pawn freely.

The goal of DAWSON'S CHESS is to force your opponent to play last. Thus, in contrast to NIM, a player has won when their turn arrives and no moves are available.

Definition 6.3. An *impartial combinatorial game* is a rooted directed graph on which two players alternate *moves* along edges. One *wins* by moving to a node (*position*) with no outgoing edges. A game is *finite* if its graph is finite with no directed cycles.

The root of the directed graph corresponds to the *initial position*. Any finite impartial game can equivalently be represented as a rooted tree in which the children of each position correspond to its *options*: the endpoints of its outgoing edges in any directed graph representation. A given position might be repeated in the tree representation. Given a tree representation, an optimally efficient directed graph representation can be constructed by identifying all vertices whose descendant subtrees are isomorphic.

Example 6.4. In the language of combinatorial game theory, NIM is not a finite game but a family of finite games that specifies a consistent "rule set" for how to play starting from any particular initial position among an infinite number of possibilities. In the situation of Example 6.1, using ijk for $0 \le i, j, k \le 9$ to represent three heaps of sizes i, j, and k, the top of the game tree is

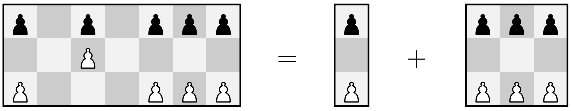


Note that positions such as 372, 004 and 073 are repeated in the tree (even in the small bit of the tree depicted here), at the same level or at multiple levels; their descendant subtrees can be identified to form another directed graph representation of this game of NIM. Every leaf of the tree corresponds to the position 000.

Remark 6.5. It is natural to wonder why the games in Definition 6.3 are called "impartial". The term is meant to indicate that the players have the same options available from each position, as opposed to *partizan* games, where the players can have distinct sets of options. "But only one player is allowed to play from each position," you may argue, "so how can you tell the difference between impartial and partizan games?" The answer is to put two games side by side; this is called the *disjunctive sum* of the two games: the player whose turn it is chooses one game and plays any of their legal moves in that game. The result is that a player can end up making consecutive moves in a single game, if the intervening move took place in the other game. In a partizan game, such as ordinary CHESS, white's options from any given position are different from black's options. When white makes two consecutive moves on the same board, they are two consecutive moves *of white pieces only*. In contrast, in an impartial game such as NIM, a move by either player could have been made by the other player, if the other player had the chance.

Beyond the ability to distinguish between impartial and partizan games, what are disjunctive sums good for? The answer is that these sums arise naturally when positions decompose, in the course of play, into smaller independent subgames.

Example 6.6. The final position in Example 6.2 can be represented as a disjoint union of two Dawson's Chess boards, one with one file and one with three,



except that now it is black's turn to move. Similarly, an initial move at either end of the board obliterates the two columns at that end, leaving the other player to move. In addition, any move on a 3×1 or 3×2 board obliterates the entire board, as does a move on the middle file of a 3×3 board.

This description implies that Dawson's Chess is a heap game, by restricting to the ordinary (non-capturing) pawn moves. Indeed, any connected board in its initial position is a heap whose size is its number of files (that is, its width), and any position between bloodbaths is a disjunctive sum of such boards. The rules allow any player to

- eliminate any heap of size 1, 2, or 3;
- take two or three from any heap of size at least 3; or
- split any heap of size $d \geq 3$ into two heaps of sizes k and $d-3-k$.

Dawson's choice for the ending of his fairy chess game was both unfortunate and fortuitous. It was unfortunate because it made the game hard: over three quarters of a century after Dawson published his little game, its solution remains elusive, both computationally and in a closed form akin to Bouton's Nim solution.

Open Problem 6.7. Determine a winning strategy for Dawson's Chess and find a polynomial-time algorithm to calculate it.

Dawson's choice was fortuitous because the simple change of ending uncovered a remarkable phenomenon: the vast difference between trying to lose and trying to win.

Definition 6.8. Given a finite combinatorial game, the *misère play* version declares the winner to be the player who does not move last.

Thus misère play is what happens when both players try to lose under the *normal play* rules. It fosters amusing titles such as "Advances in losing" [Pla09]. Dawson's Chess motivated substantial portions of the development of CGT over the past few decades.

Misère games are generally much more complex than their normal-play counterparts. Heuristically, the reason is that, in contrast to the unique "zero position" in normal play, the multiple "penultimate positions" that become winning positions in misère play cause ramifications in positions expanding farther from the zero position, and these ramifications interfere with one another in relatively unpredictable ways.

Regardless of the reason, aspects of the fact of misère difficulty were formalized by Conway in the 1970s (see [Con01]). There are, for example, many more non-isomorphic impartial misère games than impartial normal play games of any given *birthday* (the height of the game tree). For comparison, note that a complete structure theory for normal play games was formulated in the late 1930s [Spr36, Gru39]. It is based on the *Sprague–Grundy theorem*, building on Bouton's solution of NIM by reducing all finite impartial games to it: every impartial game under normal play is, in a precise sense, equivalent to a single NIM heap of some size. (The details of this theory would be more appropriate for a focused exposition on the foundations of CGT, such as Siegel's highly recommended lecture notes [Sie06], which proceed quickly to the substantive aspects from an algebraic perspective. Additional background and details can be found in [ANW07, BCG82].) Because the "zero position" is declared off-limits in misère structure theory, the elegant additivity of normal play under disjunctive sum fails for misère play, and what results is algebraically complicated in that case; see Sect. 6.4.

6.2 Lattice Games

Impartial combinatorial games admit a reformulation in terms of lattice points in polyhedra. For the purpose of Open Problem 6.7, the idea is to bring to bear the substantial algorithmic theory of rational polyhedra [BW03]. The transformation begins by a simple change of perspective on NIM, DAWSON'S CHESS, and other heap games.

Example 6.9. For certain families of games, the game tree is an inefficient encoding. For heap games, it is better to arrange the numbers of heaps of each size into a nonnegative integer vector whose i^{th} entry is the number of heaps of size i. Thus the 374 and 373 positions from Example 6.4 become

$$\begin{aligned} 3\ 7\ 4 &\leftrightarrow (0,0,1,1,0,0,1) \in \mathbb{N}^7 \\ 3\ 7\ 3 &\leftrightarrow (0,0,2,0,0,0,1) \in \mathbb{N}^7 \end{aligned}$$

The moves "make a heap of size j into a heap of size $i < j$" and "remove a heap of size j" correspond to other (not necessarily positive) integer vectors, namely

$$\begin{aligned} (\ldots,1,\ldots,-1,\ldots) &= e_i - e_j \text{ for } i < j, \text{ and} \\ (\ldots,-1,\ldots) &= -e_j \text{ for all } j \geq 1, \end{aligned}$$

where $e_1,\ldots,e_d$ is the standard basis of $\mathbb{Z}^d$. All entries in the moves are zero except for the 1 and -1 entries indicated. NIM looks curiously like it could be connected to root systems of type A, but nothing has been made of this connection.

Example 6.9 says that NIM positions with heaps of size at most d are points in $\mathbb{N}^d$, and moves between them are vectors in $\mathbb{Z}^d$. The idea behind lattice games is to polyhedrally formalize the relationship between the positions and moves. To that end, for the rest of this section, fix a pointed rational cone $C \subseteq \mathbb{Z}^d$ of dimension d, and write $Q = C \cap \mathbb{Z}^d$ for the normal affine semigroup of integer points in C. As in earlier sections, basic knowledge of polyhedra is assumed; see [Zie95] for additional background. For simplicity, this survey restricts attention to lattice games that are played on (the lattice points in) cones instead of arbitrary polyhedra, and to special rule sets for which every position has a path to the origin (cf. [GM10, Lemma 3.5]); see [GM10, §2] for full generality.

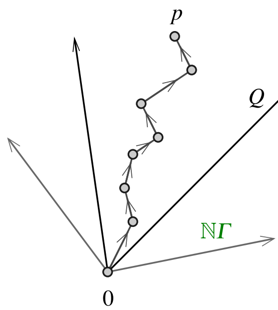
Briefly, a lattice game is played by moving a token on a game board comprising all but finitely many of the lattice points in a polyhedral cone. The allowed moves come from a finite rule set consisting of vectors that generate a pointed cone containing the game board cone. The game ends when no legal moves are available; the winner is the last player to move. The misère condition is encoded by the finitely many disallowed lattice point positions. To define these properly, it is necessary to define rule sets first.

Definition 6.10. A *rule set* is a finite subset $\Gamma \subset \mathbb{Z}^d \smallsetminus \{0\}$ such that

(a) the affine semigroup $\mathbb{N}\Gamma$ is *pointed*, meaning that its unit group is trivial, and
(b) every lattice point $p \in Q$ has a *Γ-path to* 0 in Q, meaning a sequence

$$0 = p_0, \dots, p_r = p \text{ in } Q, \text{ with } p_{i+1} - p_i \in \Gamma,$$

as illustrated in the following figure.



With these conventions, moves correspond to elements of $-\Gamma$ rather than of Γ itself. The sign is a choice that must be made, and neither option is fully convenient. The choice in Definition 6.10 prevents unpleasant signs in the next lemma.

Lemma 6.11. *$\mathbb{N}\Gamma$ contains Q and induces a partial order on $\mathbb{Z}^d$ in which $p \preceq q$ whenever $q - p \in \mathbb{N}\Gamma$.*

Proof. The containment is immediate from Definition 6.10. The partial order occurs because $\mathbb{N}\Gamma$ is a pointed affine semigroup. □

Definition 6.12. A *lattice game* played on a normal affine semigroup $Q = C \cap \mathbb{Z}^d$ has

- a rule set Γ,
- *defeated positions* $D \subseteq Q$ that constitute a finite Γ-order ideal, and
- *game board* $B = Q \smallsetminus D$.

Lattice points in Q are referred to as *positions*; note that these might lie off the game board. A position $p \in Q$ has a *move to* q if $p - q \in \Gamma$; the move is *legal* if $q \in B$. The order ideal condition, which means by definition that $q \in D \Rightarrow p \in D$ if $p \preceq q$, guarantees that a legal move must originate from a position on the game board B.

Example 6.13. A heap game in which the heaps have size at most d is played on $Q = \mathbb{N}^d$. Under normal play, the game board is all of $\mathbb{N}^d$, so $D = \varnothing$. To get misère play, let the set of defeated positions be $D = \{0\}$, so $B = \mathbb{N}^d \smallsetminus \{0\}$. Larger sets of defeated positions allow generalizations of misère play not previously considered.

Example 6.14. DAWSON'S CHESS is a lattice game on $\mathbb{N}^d$ when the heap sizes are bounded by d. The game board in this case is $\mathbb{N}^d \smallsetminus \{0\}$, corresponding to misère play. The rule set is composed of the following vectors, by Example 6.6:

- e_1;
- e_2 and $e_j - e_{j-2}$ for $j \geq 3$;
- e_3 and $e_j - e_{j-3}$ for $j \geq 4$ and $e_j - e_i - e_{j-3-i}$ for $j \geq 5$.

Historically, the abstract theory of combinatorial games was developed more with set theory than combinatorics. Formally, a finite impartial combinatorial game is often defined as a set consisting of its options, each being, recursively, a finite impartial combinatorial game. Using this language, the disjunctive sum of games G and H is the game $G+H$ whose options comprise the union of $\{G'+H \mid G' \text{ is an option of } G\}$ and $\{G+H' \mid H' \text{ is an option of } H\}$. A set of games is *closed* if it is closed under taking options and under disjunctive sum. In particular, the *closure* of a single game G is the free commutative monoid on G and its *followers*, meaning the games obtained recursively as an option, or an option of an option, etc. See [PS07] and its references for more details on closure and on the historical development of CGT.

Theorem 6.15. *Any position in a lattice game determines a finite impartial combinatorial game. Conversely, the closure of an arbitrary finite impartial combinatorial game, in normal or misère play, can be encoded as a lattice game played on* $\mathbb{N}^d$.

Proof. The first sentence is a consequence of Lemma 6.11. The second is [GM10, Theorem 5.1]: if the game graph has d nodes, then lattice points in $Q = \mathbb{N}^d$ correspond to disjunctive sums of node-positions. □

The proof of Theorem 6.15 clarifies an important point about the connection between lattice games and games given by graphs: lattice game encodings are efficient only when the nodes of the graph represent "truly different" positions.

Example 6.16. The encoding of the 374 NIM game in Example 6.4 by using all d of the followers of 374 as coordinate directions in $\mathbb{N}^d$ is woefully inefficient. On the other hand, the 374 position is encoded efficiently in $\mathbb{N}^7$ because it lies in the closure of the single NIM heap of size 7, whose followers are "truly different" from one another.

This explains part of the reason for allowing arbitrary normal affine semigroups as game boards: more classes of combinatorial games beyond heap games can be encoded efficiently. That said, heap games are now—and have been for decades—key sources of motivation and examples. As such, the encoding of heap games is particularly efficient for the following class of games [GM10, §6–§7].

Definition 6.17. A lattice game is *squarefree* if it is played on $\mathbb{N}^d$ and the maximum entry of any vector in the rule set is 1. Equivalently, a squarefree game represents a heap game in which each move destroys at most one heap of each size.

Multiple heaps of different sizes can be destroyed, and a destroyed heap can be replaced with multiple heaps of other sizes, as long as the moves still form a rule set.

Example 6.18. Squarefree games are the natural limiting generalizations of *octal games*, invented by Guy and Smith [GS56′] with DAWSON'S CHESS as a motivating example. For each $k \in \mathbb{N}$, an octal game specifies whether or not any heap of size

- k may be destroyed;
- $j \geq k+1$ may be turned into a heap of $j-k$; and
- $j \geq k+2$ may be turned into two heaps of sizes summing to $j-k$.

These constitute three binary choices, and hence are conveniently represented by an octal digit $0 \leq d_k \leq 7$. DAWSON'S CHESS is ".137" as an octal game, where the digits correspond, in order, to the types of moves in Example 6.14. For example, $7 = 111$ in binary indicates that all options are allowed for $k = 3$, while $3 = 011$ indicates that only the top two are available for $k = 2$. The dot in ".137" is a place holder.

6.3 Rational Strategies

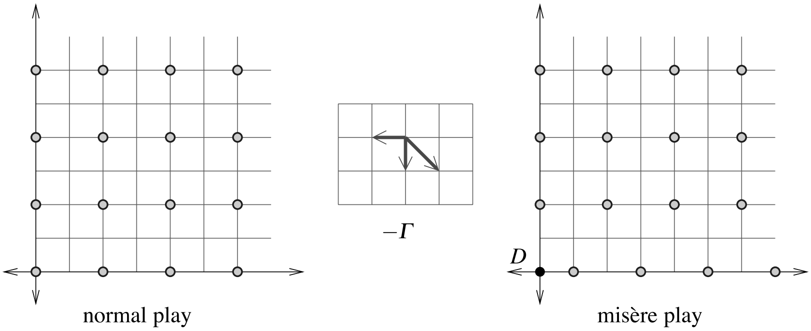
What does a strategy for a combinatorial game look like? Abstractly, the finiteness condition ensures that one of the players can force a win. The argument explaining why is recursive and elementary. But how does one describe such a strategy? Lattice games provide malleable data structures for this purpose.

Definition 6.19. Two subsets $W, L \subseteq B$ are *winning* and *losing* positions for a lattice game with game board B if

- $B = W \mathbin{\dot\cup} L$ is the disjoint union of W and L; and
- $(W + \Gamma) \cap B = L$.

Winning positions are the desired spots to move to. The first condition says that every position in B is either a winning position (the player who moved to that spot can force a win) or a losing position (the player who moves from that spot can force a win by moving to a winning position). The second condition says that losing positions are precisely those with (legal) moves to winning positions.

Example 6.20. Consider the game NIM$_2$ of NIM with heaps of size at most 2. The rule set in this case is $\left\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right]\right\}$. The negatives of these vectors—representing the legal moves from a generic position—are depicted in the following figure, along with the winning positions in NIM$_2$ for both for normal and misère play. An easy way to verify that the depicted sets W are forced is to figure out what happens on the bottom row first, from left to right, and then proceed upward, row by row.



The defeated position, labeled by D in the misère play diagram, causes the bottom row of winning positions to be shifted over one unit to the right.

Remark 6.21. The disarray caused by the defeated position in Example 6.20 becomes substantially worse with more complicated rule sets in higher dimensions. Much of the study of misère combinatorial games amounts to analyzing, quantifying, computing, and controlling the disarray.

Theorem 6.22. *Given a lattice game with rule set $\Gamma \subset \mathbb{Z}^d$ and game board B, there exist unique sets W and L of winning and losing positions for B.*

Proof. This is [GM10, Theorem 4.6]. The main point is that the cones generated by Q and $\mathbb{N}\Gamma$ point in the same direction, so recursion is possible after declaring the Γ-minimal positions in B to be winning. □

Thus everything there is to know about a lattice game is encoded by its set of winning positions: given a pointed normal affine semigroup, Theorem 6.22 implies that specifying a rule set and defeated positions is the same as specifying a valid set of winning positions, at least abstractly. But the rule set encodes winning strategies only implicitly, while the set of winning positions—or better, a generating function $f_W(\mathbf{t}) = \sum_{w \in W} \mathbf{t}^w$ for the winning positions—encodes the strategy explicitly. The following is [GM10, Conjecture 8.5].

Conjecture 6.23. *Every lattice game has a* rational strategy*: a generating function for its winning positions expressed as a ratio of polynomials with integer coefficients.*

Example 6.24. Resume Example 6.20. In normal play $\textsc{Nim}_2$, a rational strategy is

$$f_W(a,b) = \frac{1}{(1-a^2)(1-b^2)},$$

the rational generating function for the affine semigroup $2\mathbb{N}^2$. In misère play, a rational strategy is

$$f_W(a,b) = \frac{a}{1-a^2} + \frac{b^2}{(1-a^2)(1-b^2)},$$

where the first term enumerates the odd lattice points on the horizontal axis, and the second enumerates normal play winning positions that lie off the horizontal axis.

Example 6.25. For a squarefree game, if $W_0 = W \cap \{0,1\}^d \subseteq \mathbb{N}^d$ then

$$W = W_0 + 2\mathbb{N}^d.$$

This is [GM10, Theorem 6.11]. It implies that the game has a rational strategy

$$f_W(\mathbf{t}) = \sum_{w \in W_0} \frac{\mathbf{t}^w}{(1-t_1^2)\cdots(1-t_d^2)}.$$

The reader is encouraged to reconcile this statement with Example 6.1.

A rational strategy has a reasonable claim to the title of "solution to a lattice game" since it can be manipulated algorithmically and has potential to be compact.

Theorem 6.26. *A rational strategy for a lattice game produces algorithms to*

- *determine whether a position is winning or losing, and*
- *compute a legal move to a winning position, given any losing position.*

These algorithms are efficient when the rational strategy is a short *rational function, in the sense of Barvinok and Woods [BW03].*

Proof. This is a straightforward application of the theory developed by Barvinok and Woods; see [GM10′] for details. □

The efficiency in the theorem is in the sense of complexity theory. Short rational generating functions have not too many terms in their numerator and denominator polynomials. They are algorithmically efficient to manipulate and—when they enumerate lattice points in polyhedra—to compute. Since computations of lattice points in rational polyhedra are efficient, it would be better to get a polyhedral decomposition of the set W of winning positions. In fact, examples of lattice games exhibit a finer structural phenomenon than is indicated [GM10, Conjecture 8.9] & [GMW09].

Conjecture 6.27. *Every lattice game has an* affine stratification*: an expression of its winning positions as a finite union of translates of affine semigroups.*

Roughly speaking, winning positions should be finite unions of sets of the form (lattice ∩ cone). This definition of affine stratification differs from [GM10, Definition 8.6] but is equivalent [Mil10, Theorem 2.6]; it would also be equivalent to require the union to be disjoint, or (independently of disjointness) the affine semigroups to be normal.

Example 6.28. Consider again the situation from Examples 6.20 and 6.24. An affine stratification for this game is $W = 2\mathbb{N}^2$; that is, the entire set of winning positions forms an affine semigroup. In misère play, $W = \big((1,0)+\mathbb{N}(2,0)\big) \cup \big((0,2)+2\mathbb{N}^2\big)$

is the disjoint union of $W_1 = 1 + 2\mathbb{N}$ (along the first axis) and W_2, which equals the translate by twice the second basis vector of the affine semigroup $2\mathbb{N}^2$.

Remark 6.29. Conjecture 6.27 bears a resemblance to statements about local cohomology of finitely generated $\mathbb{Z}Q$-graded modules M over an affine semigroup ring $\Bbbk[Q]$ with support in a monomial ideal: the local cohomology $H^i_I(M)$ is supported on a finite union of translates of affine semigroups [HM05]. If Conjecture 6.27 is true, then perhaps it would be possible to develop a homological theory for winning positions in combinatorial games that explains why.

Theorem 6.30. *A rational strategy can be efficiently computed from any given affine stratification.*

Proof. See [GM10′, §5]. As with Theorem 6.26, this is reasonably straightforward, applying the methods from [BW03] with care. □

Algorithms for dealing with affine stratifications and rational strategies are stepping stones toward a higher aim, which would be to prove the existence of affine stratifications (Conjecture 6.27), and hence rational strategies (Conjecture 6.23), in an efficient algorithmic manner and in enough generality for DAWSON'S CHESS (Open Problem 6.7). Part of the problem to overcome for DAWSON'S CHESS is the need to deal with increasing heap size d. That problem is key, since the algorithm for DAWSON'S CHESS is supposed to be polynomial in d as $d \to \infty$, but affine stratifications and rational strategies at present are designed for fixed d.

6.4 Misère Quotients

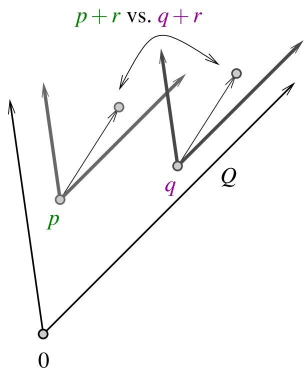
The development of lattice games and rational strategies outlined in the previous subsections were motivated by—and continue to take cues from—exciting recent advances in misère theory by Plambeck and Siegel [Pla05, PS07] pertaining to misère quotients (Definition 6.31). The lattice point methods are also beginning to return the favor, spawning new effective methods for misère quotients. This final subsection on combinatorial games ties together the lattice point perspectives on games and binomial ideals with misère quotients, particularly in Theorem 6.36.

Definition 6.31. Fix a lattice game with winning positions $W \subseteq Q$ in a pointed normal affine semigroup. Two positions $p, q \in Q$ are *indistinguishable*, written $p \sim q$, if

$$(p+Q) \cap W - p = (q+Q) \cap W - q.$$

In other words, $p + r \in W \Leftrightarrow q + r \in W$ for all $r \in Q$. The *misère quotient* of G is the quotient $\overline{Q} = Q/\!\sim$ of the affine semigroup Q modulo indistinguishability.

Geometrically, indistinguishability means that the winning positions in the cone above p are the same as those above q, up to translation by $p - q$.



Lemma 6.32. *Indistinguishability is a congruence in the sense of Definition 3.12, so the misère quotient $\overline{Q}$ is a monoid.* □

Example 6.33. In the situation of Example 6.25, the misère quotient is a quotient of $(\mathbb{Z}/2\mathbb{Z})^d = \mathbb{N}^d/2\mathbb{N}^d$ [GM10, Proposition 6.8]. As with Example 6.25, the reader is encouraged to reconcile this statement with Example 6.1.

Example 6.34. For NIM$_2$ (Examples 6.20, 6.24, and 6.28), the misère quotient is the commutative monoid with presentation $\overline{Q} = \langle a, b \mid a^2 = 1, b^3 = b\rangle$, in multiplicative notation. This monoid has six elements because it is $\langle a \mid a^2 = 1\rangle \times \langle b \mid b^3 = b\rangle$, and the second factor has order 3. The presentation of $\overline{Q}$ can be seen geometrically in the right-hand figure from Example 6.20: translating the grid two units to the right moves W bijectively to the part of W outside of the leftmost two columns (this is $a^2 = 1$), and translating the grid up by two units takes the part of W above the first row bijectively to the part of W above the third row (this is $b^3 = b$).

Misère quotients were introduced by Plambeck [Pla05] as a less stringent way to collapse the set of games than had been proposed earlier by Grundy and Smith [GS56], in view of Conway's proof that very little simplification results when the collapsing is attempted in too large a universe of games [Con01, Theorem 77]. Misère quotients have subsequently been studied and applied to computations by Plambeck and Siegel [PS07, Sie07], the point being that taking quotients often leaves a much smaller—and sometimes finite—set of positions to consider, when it comes to strategies. The Introduction of [PS07] contains an excellent account of the history, including personal accounts from some of the main players.

The first contribution of lattice games to misère theory is the following.

Proposition 6.35. *A short rational strategy for a game played on Q results in an efficient algorithm for determining the indistinguishability of any pair of positions in Q. In particular, an affine stratification results in such an algorithm.*

Proof. This is the main result in [GM10′, §6]. If $f = \sum_{q\in Q} \varphi_q \mathbf{t}^q$ and $g = \sum_{q\in Q} \psi_q \mathbf{t}^q$ are two short rational functions, then their *Hadamard product* $f \star g = \sum_{q\in Q} \varphi_q \psi_q \mathbf{t}^q$ can be efficiently computed as a short rational function [BW03]. □

The final result in this section combines three main themes thus far in the survey: lattice games, misère quotients, and—for the proof—binomial combinatorics.

Theorem 6.36. *Lattice games with finite misère quotients have affine stratifications.*

Proof. This is [Mil10, Corollary 4.5], given that we work with games played on normal affine semigroups. The proof proceeds via a general result [Mil10, Theorem 3.1] of interest here: the fibers of any projection $Q \to \overline{Q}$ from an affine semigroup Q to a monoid $\overline{Q}$ all possess affine stratifications. This is proved using combinatorial mesoprimary decompositions of congruences—the (simpler) monoid analogues of binomial mesoprimary decomposition—whose combinatorics, as in Example 4.24, give rise to affine stratifications of fibers. When the misère quotient of a lattice game is finite, the winning positions automatically comprise a finite union of fibers. □

Open Problem 6.37. Find an algorithm to compute the misère quotient of any lattice game starting from an affine stratification.

Algorithms for computing finite misère quotients are known and useful [PS07, Appendix A]. In addition, Weimerskirch has algorithmic methods that apply in the presence of certain known periodicities [Wei08], although for infinite quotients the methods fail to terminate.

Binomial primary decomposition, or at least the combinatorial aspects present in mesoprimary decomposition of congruences on monoids, is likely to play a role in further open questions on misère quotients, including when finite quotients occur, and more complex "algebraic periodicity" questions, which have yet to be formulated precisely [PS07, Appendix A.5].

7 Mass-Action Kinetics in Chemistry

Toss some chemicals into a vat. Stir. What products are produced? How fast? If the process is repeated, can the result differ? These questions belong to the study of chemical reaction dynamics. One of the earliest theories of such dynamics, the law of mass action, was formulated by Guldberg and Waage in 1864 [GW1864]. It is widely observed to hold in real-life chemical systems (as distinguished from, say, biochemical systems; see Remark 7.9).

Over the years, mass-action kinetics has matured, especially certain mathematical aspects following seminal work by Horn, Jackson, and Feinberg [HJ72, Fei87] from the 1970s and onward. In the past decade, the resulting mathematical formalizations have seen increasing amounts of algebra, particularly of the binomial sort. This section provides a brief overview of mass-action kinetics (Sect. 7.1), covering just enough basics to understand the relevance of binomial algebra. From there, the main goal is to explain the Global Attractor Conjecture (Sect. 7.2), which posits that a system of reversible chemical reactions always reaches the same steady state if one exists, with a view to how binomial primary decomposition could be relevant to its solution.

Length constraints prevent many substantial details, as well as examples demonstrating key phenomena, from being included. For an elementary introduction to chemical reaction network theory, in mathematical language, the reader is referred to the well-written notes by Gunawardena [Gun03]. For details on an abstract formalization of the law of mass-action in terms of binomials, see [AGHMR09].

7.1 Binomials from Chemical Reactions

Before presenting mass-action kinetics in general, it is worthwhile to study a small sample reaction.

Example 7.1. Consider the breakdown of hydrogen peroxide into water and oxygen:

$$2H_2O_2 \underset{\mu}{\overset{\lambda}{\rightleftharpoons}} 2H_2O + O_2$$

The λ and μ here are *rate constants*: λ indicates that two molecules of peroxide decompose into two molecules of water and one molecule of oxygen at some rate, and μ indicates that the reverse reaction also occurs, though at another (in this case, slower) rate: two molecules of water and one molecule of oxygen react to from two molecules of peroxide. To be precise, let $x = [H_2O_2]$, $y = [H_2O]$, and $z = [O_2]$ be the concentrations of peroxide, water, and oxygen in some medium. These concentrations are viewed as functions of time, and as such, they satisfy a system of ordinary differential equations:

$$\begin{aligned}\dot{x} &= 2\mu y^2 z - 2\lambda x^2\\ \dot{y} &= 2\lambda x^2 - 2\mu y^2 z\\ \dot{z} &= \lambda x^2 - \mu y^2 z.\end{aligned}$$

The right-hand side of $\dot{x}$ says that after an infinitesimal unit of time,

- for every two molecules of H_2O_2 that came together (this is the x^2 term), two molecules of H_2O_2 disappear (this is the -2 in the coefficient of x^2) some fraction of the time (this is the meaning of λ in the coefficient of x^2); and
- for every two H_2O molecules and one O_2 that came together (the y^2z term), two molecules of H_2O_2 are formed (the 2 on y^2z) some fraction of the time (μ).

The fact that x and y are squared in all of the right-hand sides is for the same reason that y^2 is multiplied by z: the products represent concentrations of chemical complexes. Thus y^2z can be thought of as an "effective concentration" of $2H_2O + O_2$.

Example 7.2. For comparison, it is instructive to see what happens when a new species is introduced to the chemical equation. Suppose that the reaction $2H_2O_2 \rightleftharpoons 2H_2O + O_2$ in Example 7.1 had a fictional additional term on the right-hand side:

$$2H_2O_2 \underset{\mu}{\overset{\lambda}{\rightleftharpoons}} 2H_2O + O_2 + 3A$$

The right-hand sides of the equations governing evolution of the species H_2O_2, H_2O, and O_2 would remain binomial:

$$\begin{aligned}\dot{x} &= 2\mu y^2 z a^3 - 2\lambda x^2\\ \dot{y} &= 2\lambda x^2 - 2\mu y^2 z a^3\\ \dot{z} &= \lambda x^2 - \mu y^2 z a^3\end{aligned}$$

as would the new equation $\dot{a} = 3\lambda x^2 - 3\mu y^2 z a^3$ governing evolution of the species A.

In general, a chemical reaction involves *species* $s_1, \ldots, s_n$ with corresponding *concentrations* $[s_i] = x_i$, each viewed as a function $x_i = x_i(t)$ of time. In Example 7.1, the species are peroxide, water, and oxygen. A reaction $A \rightleftharpoons B$ occurs between *chemical complexes* $A = a_1 s_1 + \cdots + a_n s_n$ and $B = b_1 s_1 + \cdots + b_n s_n$. Thus the complex A is composed of a_i molecules of s_i for $i = 1, \ldots, n$, and similarly for B. In Example 7.1, the complexes are H_2O_2 and $2H_2O + O_2$.

Definition 7.3. A reaction $A \rightleftharpoons B$ between complexes $A = a_1 s_1 + \cdots + a_n s_n$ and $B = b_1 s_1 + \cdots + b_n s_n$ evolves under *mass action kinetics* [GW1864] if species s_i is lost at a_i times a rate proportional to the concentration $x_1^{a_1} \cdots x_n^{a_n}$ of A, and gained at b_i times a rate proportional to the concentration $x_1^{b_1} \cdots x_n^{b_n}$ of B. The reaction is *reversible* if the reaction rates in both directions are strictly positive.

As in Example 7.1, the concentration of a complex A is the product of species concentrations with exponents corresponding to the multiplicities of the species in A.

Proposition 7.4. *The differential equation governing the evolution of the reversible reaction $A \rightleftharpoons B$ from Definition 7.3 under mass-action kinetics is*

$$\dot{x}_i = (b_i - a_i)(\lambda \mathbf{x}^{\mathbf{a}} - \mu \mathbf{x}^{\mathbf{b}}),$$

with rate constants $\lambda, \mu > 0$. In vector form, with $\lambda = \lambda_{\mathbf{ab}}$ and $\mu = \lambda_{\mathbf{ba}}$, this becomes

$$\dot{\mathbf{x}} = (\mathbf{b} - \mathbf{a})(\lambda_{\mathbf{ab}} \mathbf{x}^{\mathbf{a}} - \lambda_{\mathbf{ba}} \mathbf{x}^{\mathbf{b}})$$

Proof. This is merely a translation of Definition 7.3 into symbols. □

The factor of $\lambda \mathbf{x}^{\mathbf{a}} - \mu \mathbf{x}^{\mathbf{b}}$ in Proposition 7.4 is a scalar quantity; the only vector quantity on the right-hand side is $\mathbf{b} - \mathbf{a}$.

A single reaction under mass-action kinetics reaches a steady state when the binomial on the right-hand side of its evolution equation vanishes. Thus the set of steady states for a single reaction is the zero set of a binomial. General reaction systems involve more than one reaction at a time: in a given vat of chemicals, simultaneous transformations take place involving different pairs of chemical complexes using the given set of species in the vat.

Definition 7.5. For multiple reactions on a set of species, in which each reaction $A \rightleftharpoons B$ involves complexes with species vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$, the *law of mass-action* is obeyed if the species evolve according to the binomial sum:

$$\dot{\mathbf{x}} = \sum_{A \rightleftharpoons B} (\mathbf{b}-\mathbf{a})(\lambda_{\mathbf{ab}}\mathbf{x}^{\mathbf{a}} - \lambda_{\mathbf{ba}}\mathbf{x}^{\mathbf{b}}).$$

Thus, when more than one reaction is involved, the i^{th} entry of the vector field on the right-hand side is not a binomial but a sum of binomials, one for each reaction in which species s_i occurs. Consequently, the set of steady states need not be binomial [DM10].

Definition 7.6. A point $\xi = (\xi_1,\ldots,\xi_n) \in \mathbb{R}^n$ is a *detailed balanced equilibrium* for a reversible reaction system if it is *strictly positive*, meaning $\xi_i > 0$ for all i, and every binomial summand on the right-hand side in Definition 7.5 vanishes at ξ. A system is *detailed balanced* if it is reversible and has a detailed balanced equilibrium.

Definition 7.6 creates the bridge from chemistry to binomial algebra: the chemical interest lies in equilibria, and these are varieties of binomial ideals. Detailed balanced equilibria lie interior to the positive orthant in $\mathbb{R}^n$. For such concentrations of the species, each reaction $A \rightleftharpoons B$ in the system rests at equilibrium with both A and B present at some nonzero concentration. In fact, more is true.

Theorem 7.7. *A detailed balanced equilibrium is a locally attracting steady state.*

Proof. The main point is that a detailed balanced equilibrium possesses an explicit strict Lyapunov function (given by Helmholtz free energy) [HJ72, Fei87]. □

Remark 7.8. Chemical reaction network theory (CRNT) works in more general settings than systems of reactions each of which is reversible; see [Gun03] for an introduction. The theory is most successful when the system of reactions is *weakly reversible*: each individual reaction $A \to B$ need not be reversible, but it must be possible to reach the reactant complex A from the product complex B through a sequence of reactions in the system. See also Remark 7.11.

Remark 7.9. Mass-action kinetics fails for more complicated chemical systems, such as biochemical ones. Indeed, it must fail, for life abhors chemical equilibrium: an organism whose chemical reactions are at steady-state is otherwise known as dead. The failure of mass-action kinetics in biochemical systems occurs for a number of reasons. For one, the reaction medium is not homogeneous—that is, the reactants are not well-mixed). In addition, the molecules are often too big, and the number of them too small, for the natural discreteness to be smoothed; see [Gun03, §2].

7.2 Global Attractor Conjecture

Polynomial dynamical systems—linear ordinary differential equations with polynomial right-hand sides—behave quite poorly and unpredictably, in general. The famous chaotic *Lorenz attractor*, for example, is defined by a vector field whose entries are simple 3-term cubics. However, the binomial nature of mass-action chemistry lends a striking tameness to the dynamics.

The reversibility hypothesis for detailed balanced systems is natural from the perspective of chemistry: every reaction can, in principle, be reversed (although the activation energy required might be prohibitive under standard conditions). Overwhelming experience says that typical chemical reactions—well-mixed, at constant temperature, as in chemical manufacturing—approach balanced steady states, and the same products really do emerge every time. But this is surprisingly unknown theoretically for detailed balanced systems under mass-action kinetics, even though their equilibria are local attractors by Theorem 7.7.

Conjecture 7.10 (Global Attractor Conjecture [HJ72, Hor74])**.** *If a reversible reaction system as in Definition 7.5 has a detailed balanced equilibrium, then every trajectory starting from strictly positive initial concentrations reaches it in the limit.*

The Global Attractor Conjecture is "the fundamental open question in the field" [AGHMR09, §1], since it would close the book on fundamentally justifying mass-action kinetics. It is known that the conjecture holds when the binomial ideal is prime [Gop09]. It is also known, for detailed balanced reaction systems with fixed positive initial species concentrations, that

(a) the detailed balanced equilibrium is unique [Fei87], and
(b) each trajectory tends toward some equilibrium [Son01, Cha03].

Thus it suffices to bound every strictly positive trajectory away from all boundary equilibria, for then the trajectory limits are forced toward the detailed balanced one.

Remark 7.11. Detailed balancing is a stronger hypothesis than required for the known results listed above, including Theorem 7.7. Weak reversibility as in Remark 7.8, or even a less stringent condition, often suffices. Detailed balancing is also weaker than the hypothesis in the strongest (and still widely believed) form of Conjecture 7.10, which stipulates a condition called *complex-balancing* that implies weak reversibility; see, for instance, [AS09, §4.2] for a precise statement.

What do the boundary equilibria look like? The restriction of a detailed balanced reaction system to a coordinate subspace of $\mathbb{R}^n$ amounts to forcing the concentrations of some reactant species to be zero. Such restrictions still constitute detailed balanced reaction systems, and the binomials whose vanishing describes the equilibria come from the binomials in the original system. This discussion can be rephrased as follows.

Proposition 7.12. *Boundary equilibria of detailed balanced reaction systems are zeros of associated primes of the ideal generated by the binomials in Definition 7.5. Conjecture 7.10 holds if and only if every trajectory with positive initial concentrations remains bounded away from the zero set of every associated prime.* □

Proposition 7.12 brings binomial primary decomposition to bear on the chemistry of mass-action kinetics. The details of how this occurs are illustrated by certain special cases of Conjecture 7.10 whose proofs are known. The characterizations of these cases rely on a polyhedral concept hiding in the dynamics.

Definition 7.13. The *stoichiometric compatibility subspace* of the reaction system in Definition 7.5 is the real span S of the vectors $\mathbf{a}-\mathbf{b}$ over all reactions $A \rightleftharpoons B$ in the system. The *stoichiometric compatibility class* (or *invariant polyhedron*) of a species concentration vector $\xi \in \mathbb{R}^n$ is the intersection of the nonnegative orthant with $\xi + S$.

Lemma 7.14. *Trajectories for any reaction system are constrained to lie in the invariant polyhedron of the vector of initial species concentrations.*

Proof. This is immediate from the equation for $\dot{\mathbf{x}}$ in Definition 7.5. □

The known cases of Conjecture 7.10 include all systems whose initial species concentration vectors have invariant polyhedra of dimension 2 or less [AS09, Corollary 4.7]. In the language of Proposition 7.12, the method of proof is to bound all trajectories away from the zero set of every associated prime whose intersection with the relevant invariant polyhedron is

- a vertex [And08, CDSS09] or
- interior to a facet [AS09].

For a comprehensive review of known cases of the Global Attractor Conjecture, see [AS09, §1 and §4].

The combinatorics of binomial primary decomposition might contribute further than merely the statement of Proposition 7.12. For example, mesoprimary decomposition (Definition 4.23; see [KM10]) provides decompositions of binomial ideals over the rational or real numbers, and therefore takes steps toward primary decomposition over the reals. Mesoprimary decomposition also characterizes the associated lattices combinatorially, without a priori knowing the primary decomposition. Both could be important for applications to the Global Attractor Conjecture: perhaps finiteness conditions surrounding associated lattices (Example 4.24) indicates how to produce the desired trajectory bounds, with the reality (i.e., defined over $\mathbb{R}$) of the components forcing progress away from the boundary, as opposed to (say) periodicity of some kind.

In an amazing convergence, graphs associated to *event systems* [AGHMR09, Definition 2.9] provide chemical interpretations of the graphs G_I from Sects. 3 and 4 (particularly Definition 3.16) in this survey. Reversibility of the reaction system means that it is correct for G_I to be undirected. The characterization of *naturality* in [AGHMR09, Theorem 5.1] is a condition on the primary decomposition of the *event ideal*. This convergence is cause for optimism that lattice-point point combinatorics will be instrumental in proving Conjecture 7.10 via Proposition 7.12.

References

[AGHMR09] Leonard Adleman, Manoj Gopalkrishnan, Ming-Deh Huang, Pablo Moisset, and Dustin Reishus, *On the mathematics of the law of mass action*, SIAM Review, to appear, 2009.

[And08] David F. Anderson, *Global asymptotic stability for a class of nonlinear chemical equations*, SIAM J. Appl. Math. **68**, no. 5 (2008), 1464–1476.

[AS09] David F. Anderson and Anne Shiu, *The dynamics of weakly reversible population processes near facets*, preprint. arXiv:math.DS/0903.0901

[App1880] Paul Appell, *Sur les séries hypergéometriques de deux variables et sur des équations différentielles linéaires aux dérivées partielles*, Comptes Rendus **90** (1880), 296–298.

[ANW07] Michael H. Albert, Richard J. Nowakowski, and David Wolfe, *Lessons in play: An introduction to the combinatorial theory of games*, A K Peters, Wellesley, 2007.

[AM69] M.F. Atiyah and I.G. Macdonald, *Introduction to commutative algebra*, Addison–Wesley, Reading, MA–London–Don Mills, ON, 1969.

[BW03] Alexander Barvinok and Kevin Woods, *Short rational generating functions for lattice point problems*, J. Amer. Math. Soc. **16**, no. 4 (2003), 957–979. (electronic)

[BCG82] Elwyn R. Berlekamp, John H. Conway, and Richard K. Guy, *Winning ways for your mathematical plays. Vol. 1: Games in general*, Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers), London–New York, 1982.

[Bou1902] Charles L. Bouton, *Nim, a game with a complete mathematical theory*, Ann. of Math. (2) **3**, no. 1–4 (1901/02), 35–39.

[Cha03] Madalena Chavez, *Observer design for a class of nonlinear systems, with applications to chemical and biological networks*, Ph.D. Thesis, Rutgers University, New Brunswick, NJ, 2003.

[Con01] John H. Conway, *On numbers and games*, Second edition, A K Peters, Ltd., Natick, MA, 2001.

[CDSS09] Gheorghe Craciun, Alicia Dickenstein, Anne Shiu, and Bernd Sturmfels, *Toric dynamical systems*, J. Symbolic Computation **44** (2009), 1551–1565.

[Daw34] Thomas Dawson, *Fairy Chess Supplement*, The Problemist: British Chess Problem Society **2**, no. 9 (1934), p. 94, Problem No. 1603.

[DMM07] Alicia Dickenstein, Laura Felicia Matusevich, and Ezra Miller, *Extended abstract: Binomial D-modules*, Proceedings MEGA (Effective Methods in Algebraic Geometry), Strobl, Austria, 2007, 13 pages. http://www.ricam.oeaw.ac.at/mega2007/electronic/electronic.html

[DMM10] Alicia Dickenstein, Laura Felicia Matusevich, and Ezra Miller, *Combinatorics of binomial primary decomposition*, Mathematische Zeitschrift **264**, no. 4 (2010), 745–763.

[DMM10′] Alicia Dickenstein, Laura Felicia Matusevich, and Ezra Miller, *Binomial D-modules*, Duke Mathematical Journal **151**, no. 3 (2010), 385–429.

[DM10] Alicia Dickenstein and Mercedes Pérez Millán, *How far is complex balancing from detailed balancing?*, preprint. arXiv:math.DS/1001.0947

[DSS09] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Oberwolfach seminars, vol. 39, Springer, (Birkhäuser, Basel), 2009.

[ES96] David Eisenbud and Bernd Sturmfels, *Binomial ideals*, Duke Math. J. **84**, no. 1 (1996), 1–45.

[Erd50] Arthur Erdélyi, *Hypergeometric functions of two variables*, Acta Math. **83** (1950), 131–164.

[Fei87] Martin Feinberg, *Chemical reaction network structure and the stability of complex isothermal reactors, I. The deficiency zero and deficiency one theorems*, Chem. Eng. Sci., **42**, no. 10 (1987), 2229–2268.

[GGZ87] I.M. Gelfand, M.I. Graev, and A.V. Zelevinskiĭ, *Holonomic systems of equations and series of hypergeometric type*, Dokl. Akad. Nauk SSSR **295**, no. 1 (1987), 14–19.

[GKZ89] I.M. Gelfand, A.V. Zelevinskiĭ, and M.M. Kapranov, *Hypergeometric functions and toric varieties*, Funktsional. Anal. i Prilozhen. **23**, no. 2 (1989), 12–26. Correction in ibid, **27**, no. 4 (1993), 91.

[Gil84] Robert Gilmer, *Commutative semigroup rings*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1984.

[Gop09] Manoj Gopalkrishnan, *An algebraic generalization of the atomic hypothesis*, preprint, 2009.

[Gru39] Patrick M. Grundy, *Mathematics and games*, Eureka **2** (1939), 6–8; reprinted **27** (1964), 9–11.

[GS56] Patrick M. Grundy and C.A.B. Smith, *Disjunctive games with the last player losing*, Proc. Cambridge Philos. Soc. **52** (1956), 527–533.

[GW1864] Cato M. Guldberg and Peter Waage (translated by H.I. Abrash), *Studies concerning affinity*, Journal of chemical education **63** (1986), 1044–1047.

[Gun03] Jeremy Gunawardena, *Chemical reaction network theory for in-silico biologists*, preprint, 2003. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.126&rep=rep1&type=pdf

[GM10] Alan Guo and Ezra Miller, *Lattice point methods for combinatorial games*, Adv. in Appl. Math., 19 pages, to appear. arXiv:math.AC/0908.3473

[GM10′] Alan Guo and Ezra Miller, *Algorithms for lattice games*, in preparation, 2010.

[GMW09] Alan Guo, Ezra Miller, and Mike Weimerskirch, *Potential applications of commutative algebra to combinatorial game theory*, in *Kommutative Algebra*, abstracts from the April 19–25, 2009 workshop, organized by W. Bruns, H. Flenner, and C. Huneke, Oberwolfach rep. **22** (2009), 23–26.

[GS56′] R.K. Guy and C.A.B. Smith, The G-values of various games, Proc. Cambridge Philos. Soc. **52** (1956), 514–526.

[HM05] David Helm and Ezra Miller, *Algorithms for graded injective resolutions and local cohomology over semigroup rings*, J. Symbolic Computation **39** (2005), 373–395.

[HJ72] Friedrich Horn and Roy Jackson, *General mass action kinetics*, Arch. Ration. Mech. Anal., **47**, no. 2 (1972), 81–116.

[Hor74] Friedrich Horn, *The dynamics of open reaction systems*, in *Mathematical aspects of chemical and biochemical problems and quantum chemistry* (Proc. SIAM–AMS Sympos. Appl. Math., New York, 1974), pp. 125–137. SIAM–AMS Proceedings, vol. VIII, Amer. Math. Soc., Providence, R.I., 1974.

[Hor1889] J. Horn, *Über die konvergenz der hypergeometrischen Reihen zweier und dreier Veränderlichen*, Math. Ann. **34** (1889), 544–600.

[Hor31] J. Horn, *Hypergeometrische Funktionen zweier Veränderlichen*, Math. Ann. **105**, no. 1 (1931), 381–407.

[HS00] Serkan Hoşten and Jay Shapiro, *Primary decomposition of lattice basis ideals*, J. Symbolic Comput. **29**, no. 4–5 (2000), 625–639, Symbolic computation in algebra, analysis, and geometry (Berkeley, CA, 1998).

[KM10] Thomas Kahle and Ezra Miller, *Decompositions of commutative monoid congruences and binomial ideals*, in preparation, 2010.

[Kum1836] Ernst Eduard Kummer, *Über die hypergeometrische Reihe* $F(\alpha,\beta,x)$, J. Reine Angew. Math. **15** (1836).

[Mel21] Hjalmar Mellin, *Résolution de l'équation algébrique générale à l'aide de la fonction* Γ, C.R. Acad. Sc. **172** (1921), 658–661.

[MMW05] Laura Felicia Matusevich, Ezra Miller, and Uli Walther, *Homological methods for hypergeometric families*, J. Amer. Math. Soc. **18**, no. 4 (2005), 919–941.

[Mil02] Ezra Miller, *Cohen–Macaulay quotients of normal semigroup rings via irreducible resolutions*, Math. Res. Lett. **9**, no. 1 (2002), 117–128.

[Mil09] Ezra Miller, *Alexander duality for monomial ideals and their resolutions*, Rejecta Mathematica **1**, no. 1 (2009), 18–57. arXiv:math.AC/9812095

[Mil10] Ezra Miller, *Affine stratifications from finite misère quotients*, preprint, 2010. arXiv:math.CO/1009.2199

[MS05] Ezra Miller and Bernd Sturmfels, *Combinatorial commutative algebra*, Graduate Texts in Mathematics, vol. 227, Springer-Verlag, New York, 2005.

[Pla05] Thane E. Plambeck, *Taming the wild in impartial combinatorial games*, Integers **5**, no. 1 (2005), G5, 36 pp. (electronic)

[Pla09] Thane E. Plambeck, *Advances in losing*, In *Games of no chance 3*, papers from the Combinatorial Game Theory Workshop held in Banff, AB, June 2005, edited by Michael H. Albert and Richard J. Nowakowski, MSRI Publications, Cambridge University Press, Cambridge, forthcoming. arXiv:math.CO/0603027

[PS07] Thane E. Plambeck and Aaron N. Siegel, *Misère quotients for impartial games*, J. Combin. Theory Ser. A **115**, no. 4 (2008), 593–622. arXiv:math.CO/0609825v5

[SW08] Mathias Schulze and Uli Walther, *Irregularity of hypergeometric systems via slopes along coordinate subspaces*, Duke Math. J. **142**, no. 3 (2008), 465–509.

[Sie06] Aaron N. Siegel, *Misère games and misère quotients*, unpublished lecture notes. arXiv:math.CO/0612616

[Sie07] Aaron N. Siegel, *The structure and classification of misère quotients*, preprint. arXiv:math.CO/0703070

[Son01] Eduardo Sontag, *Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction*, IEEE Trans. Automat. Control, **46** (2001), 1028–1047.

[Spr36] Roland P. Sprague, *Über mathematische Kampfspiele [On mathematical war games]*, Tôhoku Math. Journal **41** (1935–1936), 438–444.

[SK85] H.M. Srivastava and Per W. Karlsson, *Multiple Gaussian hypergeometric series*, Ellis Horwood Series: Mathematics and its Applications, Ellis Horwood Ltd., Chichester, 1985.

[Wei08] Michael Weimerskirch, *An algorithm for computing indistinguishability quotients in misère impartial combinatorial games*, preprint, 31 August 2008.

[Zie95] Günter M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995.

http://www.springer.com/978-3-642-19491-7

Combinatorial Aspects of Commutative Algebra and
Algebraic Geometry
The Abel Symposium 2009
Fløystad, G.; Johnsen, T.; Knutsen, A.L. (Eds.)
2011, XIV, 174 p., Hardcover
ISBN: 978-3-642-19491-7