NILPOTENTS, UNITS, AND ZERO DIVISORS FOR POLYNOMIALS

KEITH CONRAD

1. Introduction

Let A be a nonzero commutative ring. We want to describe the following kinds of polynomials in $A[x]$, and more generally in $A[x_1, \ldots, x_d]$ for $d \geq 1$:

- nilpotents,
- units,
- zero divisors.

These are easy to describe when A is an integral domain, since $A[x_1, \ldots, x_d]$ is an integral domain too. The only nilpotent element is 0. The units are the units in A since $\deg(fg) = \deg f + \deg g$ when $f, g \neq 0$, so $fg = 1 \Rightarrow \deg f, \deg g = 0 \Rightarrow f \in A^\times$, and the converse is obvious. The only zero divisor is 0. To extend these results to polynomial rings whose coefficients are in an arbitrary nonzero commutative ring,[1] we will focus first on the case $d = 1$. Most of the essential ideas for general d already appear when looking at $d = 1$, although proving the theorem about zero divisors in $A[x_1, \ldots, x_d]$ when $d > 1$ will involve some new concepts.

2. Properties of polynomials in one indeterminate

Theorem 2.1. *A polynomial in $A[x]$ is nilpotent if and only if all of its coefficients are nilpotent in A.*

Proof. The nilpotent elements in a commutative ring form an ideal, by the binomial theorem, and nilpotent elements of A are nilpotent in $A[x]$. Therefore if a polynomial in $A[x]$ has coefficients that are nilpotent in A, then the polynomial is nilpotent in $A[x]$.

Conversely, suppose $f(x)$ in $A[x]$ is nilpotent. To show all the coefficients of $f(x)$ are nilpotent in A, we may focus on nonzero $f(x)$ and induct on $\deg f$. The case $\deg f = 0$ is easy. Suppose $n \geq 1$ and the result is true for all polynomials of degree less than n. For a nilpotent polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ of degree n, say

$$f(x)^k = 0$$

for some $k \geq 1$. Looking at the coefficient of x^{nk} on both sides tells us $a_n^k = 0$. Therefore a_n is nilpotent in A. Since the nilpotent elements in $A[x]$ form an ideal,

$$f(x) - a_n x^n = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$$

is nilpotent in $A[x]$ and is either 0 or has degree less than n. Either way, by induction all a_i for $i \leq n - 1$ are nilpotent, so all the coefficients of $f(x)$ are nilpotent. □

Theorem 2.2. *A polynomial in $A[x]$ is a unit if and only if its constant term is a unit in A and its higher-degree coefficients are all nilpotent.*

[1] The running hypothesis that $A \neq \{0\}$ will not be regularly repeated.

Proof. In a commutative ring, if u is a unit and b is nilpotent, the sum $u+b$ is a unit: $u+b = u(1+b/u)$, and if $b^k = 0$ then an elementary calculation shows $1+b/u$ is a unit with inverse being the finite geometric series

$$1 - b/u + (-b/u)^2 + (-b/u)^3 + \cdots + (-b/u)^{k-1}.$$

Intuitively, we are expanding $1/(1-r)$ for $r = -b/u$ into a geometric series $\sum_{i\geq 0} r^i$ and truncating it before the term where $i = k$ since $(-b/u)^i = 0$ for $i \geq k$. Thus $u+b = u(1+b/u)$ is a product of two units and therefore is a unit.

By that general argument, if $f(x)$ in $A[x]$ has constant term in $A^\times$ and its higher-degree coefficients are all nilpotent, $f(x)$ is a unit plus a nilpotent in $A[x]$ by the previous theorem, and therefore $f(x) \in A[x]^\times$.

For the converse result, suppose $f(x) = a_0 + a_1x + \cdots + a_nx^n$ is a unit in $A[x]$, say $f(x)g(x) = 1$ for some $g(x) \in A[x]$. Looking at the constant terms on both sides shows $a_0 \in A^\times$. Why are the higher-degree coefficients of $f(x)$ nilpotent? This uses a consequence of Zorn's lemma: in a nonzero commutative ring, the intersection of all the prime ideals is the set of nilpotent elements in the ring.[2] With this in mind, let $\mathfrak{p}$ be a prime ideal in A. The equation $f(x)g(x) = 1$ in $A[x]$ implies after reducing coefficients modulo $\mathfrak{p}$ that $\overline{f}(x)\overline{g}(x) = \overline{1}$ in $(A/\mathfrak{p})[x]$. Since $A/\mathfrak{p}$ is an integral domain, a unit in $(A/\mathfrak{p})[x]$ has degree 0 (see the introduction), so $\deg(\overline{f}(x)) = 0$. Thus all higher-degree coefficients of $f(x)$ belong to $\mathfrak{p}$. Letting $\mathfrak{p}$ run over all the prime ideals of A, each higher-degree coefficient of $f(x)$ is in every prime ideal of A and therefore the higher-degree coefficients of $f(x)$ are nilpotent. □

Example 2.3. In $(\mathbf{Z}/6\mathbf{Z})[x]$, the units are 1 and 5 (units in $\mathbf{Z}/6\mathbf{Z}$): the only nilpotent element of $\mathbf{Z}/6\mathbf{Z}$ is 0, so the higher-degree coefficients of a unit in $(\mathbf{Z}/6\mathbf{Z})[x]$ must be 0.

Example 2.4. In $(\mathbf{Z}/45\mathbf{Z})[x]$, $8 + 15x$ is a unit (it equals $8(1+30x)$, which has inverse $17(1-30x) = 17+30x$), while $8+3x$ is not a unit: 3 is not nilpotent in $\mathbf{Z}/45\mathbf{Z}$.

The following theorem on zero divisors in polynomial rings is due to McCoy [1]. The result is striking: for a polynomial to be a zero divisor means it is annihilated after multiplication by a nonzero polynomial, but in fact it can be annihilated by a nonzero *constant*.

Theorem 2.5 (McCoy)**.** *A polynomial $f(x) \in A[x]$ is a zero divisor if and only if there is some nonzero $a \in A$ such that $af(x) = 0$.*

Proof. Obviously if $af(x) = 0$ for some nonzero $a \in A$, then $f(x)$ is a zero divisor.

To prove the more interesting converse direction, we follow Scott [3]. For a zero divisor $f(x) \in A[x]$, $g(x)f(x) = 0$ for a nonzero $g(x)$. Let $g(x)$ have minimal degree such that $g(x)f(x) = 0$. Assume $\deg(g(x)) > 0$, so at least $f(x) \neq 0$. We will get a contradiction, so $\deg(g(x)) = 0$: $af(x) = 0$ for some nonzero $a \in A$.

Let

$$f(x) = a_0 + a_1x + \cdots + a_nx^n,$$
$$g(x) = b_0 + b_1x + \cdots + b_mx^m$$

where $a_n \neq 0$ and $b_m \neq 0$ with $m \geq 1$. We have $b_mf(x) \neq 0$ by the minimality of $\deg(g(x))$, so $b_ma_i \neq 0$ for some i. Thus $g(x)a_i \neq 0$ for that i. Let j be maximal with $g(x)a_j \neq 0$, so

$$0 = g(x)f(x) = (b_0 + b_1x + \cdots + b_mx^m)(a_0 + a_1x + \cdots + a_jx^j),$$

[2]For a proof, see Theorem 3.3 in https://kconrad.math.uconn.edu/blurbs/zorn1.pdf.

where we drop the terms $a_i x^i$ from $f(x)$ where $i > j$ since $g(x)a_i = 0$ for $i > j$. Looking at the coefficient of x^{m+j}, $b_m a_j = 0$, so $\deg(g(x)a_j) < \deg(g(x))$. (We have $g(x)a_j \neq 0$ by the maximality of j, so $\deg(g(x)a_j)$ makes sense.)

Since $(g(x)a_j)f(x) = a_j(g(x)f(x)) = a_j \cdot 0 = 0$[3] and $\deg(g(x)a_j) < \deg(g(x))$, we have contradicted minimality of $\deg(g(x))$ among nonzero polynomials that annihilate $f(x)$ by multiplication. The assumption that $\deg(g(x)) > 0$ is incorrect, so $\deg(g(x)) = 0$. □

Example 2.6. In $(\mathbf{Z}/6\mathbf{Z})[x]$, $2 + 3x$ is a *not* a zero divisor even though its coefficients 2 and 3 are zero divisors since $a(2 + 3x) \neq 0$ for nonzero $a \in \mathbf{Z}/6\mathbf{Z}$ by a direct calculation.

Corollary 2.7. *For a polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ in $\mathbf{Z}[x]$ and $m \geq 2$, $f(x)$ mod m is a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x]$ if and only if $\gcd(a_0, \ldots, a_n, m) > 1$.*

Proof. If $f(x)$ is a zero divisor, then by Theorem 2.5, $af(x) \equiv 0 \bmod m$ where $a \not\equiv 0 \bmod m$. Then $aa_i \equiv 0 \bmod m$ for all i. Let $d = (a, m)$, $a = da'$, and $m = dm'$. We have $d \neq m$ since $a \not\equiv 0 \bmod m$, so $m' > 1$. Dividing through the congruence $aa_i \equiv 0 \bmod m$ (including the modulus) by d, $a'a_i \equiv 0 \bmod m'$, so $m' \mid a'a_i$. Since $(a', m') = 1$, $m' \mid a_i$ for all i. Therefore m' is a common divisor of each coefficient a_i and of course $m' \mid m$, so $\gcd(a_0, \ldots, a_n, m) \geq m' > 1$.

Set $c = \gcd(a_0, \ldots, a_n, m)$ and suppose $c > 1$. If $c = m$ then all the coefficients of $f(x)$ are divisible by m, so $f(x)$ mod m is 0 in $(\mathbf{Z}/m\mathbf{Z})[x]$ and therefore $f(x)$ mod m is a zero divisor. Now take $1 < c < m$. All coefficients of $f(x)$ are divisible by c, so all coefficients of $(m/c)f(x)$ are divisible by m and $1 < m/c < m$. Thus $(m/c)f(x) \equiv 0 \bmod m$ and $m/c \not\equiv 0 \bmod m$, so $f(x)$ mod m is a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x]$. □

Example 2.8. A zero divisor in $(\mathbf{Z}/6\mathbf{Z})[x]$ must have all coefficients divisible by 2 or all coefficients divisible by 3.

3. Properties of polynomials in d indeterminates

Theorem 3.1. *A polynomial in $A[x_1, \ldots, x_d]$ is nilpotent if and only if all of its coefficients are nilpotent in A.*

Proof. As in the case $d = 1$, a polynomial in $A[x_1, \ldots, x_d]$ whose coefficients are all nilpotent in A is nilpotent in $A[x_1, \ldots, x_d]$ since nilpotent elements form an ideal.

To prove the converse, that a nilpotent polynomial has nilpotent coefficients in A, we induct on d. The case $d = 1$ is Theorem 2.1. Suppose $d \geq 2$ and the converse is proved for polynomials in $d - 1$ indeterminates with coefficients in an arbitrary nonzero commutative ring. Let $f \in A[x_1, \ldots, x_d]$ be nilpotent. Writing $A[x_1, \ldots, x_d]$ as $A[x_1, \ldots, x_{d-1}][x_d]$, we have $f = \sum_{i=0}^n c_i x_d^i$ where $c_i \in A[x_1, \ldots, x_{d-1}]$ for $i = 0, \ldots, n$. By the base case, each c_i is nilpotent in $A[x_1, \ldots, x_{d-1}]$, so by induction the A-coefficients of each c_i are nilpotent. The A-coefficients of all c_i, as i runs from 0 to n, are the A-coefficients of f, so all the A-coefficients of f are nilpotent. □

Theorem 3.2. *A polynomial in $A[x_1, \ldots, x_d]$ is a unit if and only if its constant term is a unit in A and its higher-degree coefficients are all nilpotent.*

Proof. The proof of the"if" direction is identical to the case $d = 1$ in Theorem 2.2.

The "only if" direction is also proved in the same way as the "only if" direction of Theorem 2.2, since $(A/\mathfrak{p})[x_1, \ldots, x_d]$ is an integral domain for each prime ideal $\mathfrak{p}$ of A. □

[3]Here we use commutativity of multiplication in A. The theorem isn't true for some noncommutative A.

To prove an analogue of Theorem 2.5 (about zero divisors) for multivariable polynomials, let's first try the case of two indeterminantes. Suppose $f(x,y)$ is a zero divisor in $A[x,y]$. Viewing $A[x,y]$ as $A[y][x]$, write $f(x,y) = \sum_{i=0}^{n} c_i(y)x^i$ where $c_i(y) \in A[y]$. By Theorem 2.5 for the ring $(A[y])[x]$, there is some nonzero $c(y) \in A[y]$ such that $c(y)f(x,y) = 0$, so $c(y)c_i(y) = 0$ for all $c_i(y)$ by looking at the coefficient of each power of x in $c(y)f(x,y)$. Therefore each $c_i(y)$ is a zero divisor in $A[y]$, so by Theorem 2.5 for the ring $A[y]$, $a_ic_i(y) = 0$ for some $a_i \in A$ where $i = 0, \ldots, n$. Unfortunately, it is not clear that $a_0, \ldots, a_n$ can be chosen as the *same* element of A, and without that we don't get $af(x,y) = 0$ for a nonzero $a \in A$. Nevertheless, that desired result is in fact true, not just in $A[x,y]$ but in $A[x_1, \ldots, x_d]$: a zero divisor in $A[x_1, \ldots, x_d]$ is annihilated by a nonzero element of A. We will prove this in two ways. The first proof, due to McCoy [2], involves proving a *stronger* theorem about ideals rather than polynomials. The second proof will use a total ordering on multivariable monomials that will allow us to generalize the proof of Theorem 2.5 to the multivariable setting rather easily, in contrast to the incomplete attempt at that in $A[x,y]$ above.

Theorem 3.3 (McCoy)**.** *Let I be an ideal in $A[x_1, \ldots, x_d]$ such that $gI = (0)$ for some nonzero $g \in A[x_1, \ldots, x_d]$. Then $aI = (0)$ for some nonzero $a \in A$.*

Proof. We will prove this using induction on d.

Base case $d = 1$. Our argument will be very similar to the proof of Theorem 2.5, with a few differences because ideals in $A[x]$ need not be principal.

Let I be an ideal in $A[x]$ where $g(x)I = (0)$ for a nonzero $g(x) \in A[x]$. Let $g(x)$ have minimal degree subject to the condition $g(x)I = (0)$. Assume $\deg(g(x)) > 0$, so $aI \neq (0)$ for all nonzero $a \in A$. We will get a contradiction, so $\deg(g(x)) = 0$: $aI = (0)$ for some nonzero $a \in A$.

Write $g(x) = b_0 + b_1x + \cdots + b_mx^m$ with $m \geq 1$ and $b_m \neq 0$. Since $b_mI \neq (0)$, $b_mf(x) \neq 0$ for some $f(x) \in I$, so $f(x) \neq 0$. Write $f(x) = a_0 + a_1x + \cdots + a_nx^n$. Since $b_mf(x) \neq 0$, $b_ma_i \neq 0$ for some i, so $g(x)a_i \neq 0$ for some i. Let j be maximal with $g(x)a_j \neq 0$. Since $f(x) \in I$ and $g(x)I = (0)$, the product $g(x)f(x)$ is 0:

$$0 = g(x)f(x) = (b_0 + b_1x + \cdots + b_mx^m)(a_0 + a_1x + \cdots + a_jx^j),$$

where we drop the terms a_ix^i from $f(x)$ where $i > j$ since $g(x)a_i = 0$ for $i > j$. Looking at the coefficient of x^{m+j}, $b_ma_j = 0$, so $\deg(g(x)a_j) < \deg(g(x))$. (We have $g(x)a_j \neq 0$ by the maximality of j, so $\deg(g(x)a_j)$ makes sense.)

For *all* $F(x)$ in I (not just the $f(x)$ above!), $g(x)F(x) = 0$ by the definition of $g(x)$, so $(g(x)a_j)F(x) = a_j(g(x)F(x)) = a_j \cdot 0 = 0$. Therefore $(g(x)a_j)I = (0)$ and $\deg(g(x)a_j) < \deg(g(x))$, which contradicts minimality of $\deg(g(x))$ among nonzero polynomials that annihilate I by multiplication. The assumption that the minimal value of $\deg(g(x))$ is positive is incorrect, so that minimal value is 0: there is a nonzero $a \in A$ such that $aI = (0)$.

Inductive step. For $d \geq 2$, assume the theorem is proved for polynomials in $d-1$ indeterminates over an arbitrary nonzero commutative ring. Let I be an ideal in $A[x_1, \ldots, x_d]$ such that $gI = (0)$ for some nonzero $g \in A[x_1, \ldots, x_d]$ and give g the minimal possible degree. We want to show that minimal degree is 0, so we'll assume instead that the minimal degree is positive and get a contradiction. View $A[x_1, \ldots, x_d]$ as $(A[x_1, \ldots, x_{d-1}])[x_d]$, the ring of polynomials in x_d with coefficients in $A[x_1, \ldots, x_{d-1}]$. That makes I an ideal in $R[x_d]$ where $R = A[x_1, \ldots, x_{d-1}]$. By the base case of single-variable polynomials (with an arbitrary coefficient ring), from $gI = (0)$ we get $c(x_1, \ldots, x_{d-1})I = (0)$ for some nonzero polynomial $c(x_1, \ldots, x_{d-1})$ in R.

For each $f \in I$, write

$$f = \sum_{i=0}^{n} c_i x_d^i \tag{3.1}$$

where $c_i \in R$, and abbreviate $c(x_1, \ldots, x_{d-1})$ to c, so $c, c_0, \ldots, c_n$ are all in R: they don't involve x_d. From $cI = (0)$ we get $cf = 0$, so $cc_i = 0$ by looking at the coefficients of different powers of x_d in cf. This means c multiplied by *each coefficient in R of each polynomial in I is* 0. Let J be the ideal in R generated by the coefficients in R of the polynomials in I (written as polynomials in x_d). Then $cJ = (0)$, where $c \in R$ and J is an ideal in R. Since $R = A[x_1, \ldots, x_{d-1}]$, by the inductive hypothesis $aJ = 0$ for some nonzero $a \in A$. Returning to arbitrary f in I written as in (3.1), the condition $aJ = (0)$ implies $ac_i = 0$ since all c_i's are in J, so $af = \sum_{i=0}^{n} ac_i x_d^i = 0$. This holds for all f in I, so $aI = (0)$. □

Corollary 3.4. *If f is a zero divisor in $A[x_1, \ldots, x_d]$ then $af = 0$ for some nonzero $a \in A$.*

Proof. We are assuming $gf = 0$ for some nonzero $g \in A[x_1, \ldots, x_d]$ and can apply Theorem 3.3 to the principal ideal $I = (f)$: from $gI = (gf) = (0)$, we get $aI = (0)$ for some nonzero $a \in A$, so $af = 0$. □

Our second proof of Corollary 3.4 will not use induction on d, and instead will use an ordering on the monomials in $A[x_1, \ldots, x_d]$ that gives them a total ordering: two different monomials are always comparable (one being greater than the other). This will give us an unambiguous notion of leading term and degree (valued in $\mathbf{N}^d$ rather than $\mathbf{N}$) that will allow the proof of Theorem 2.5 to be carried over rather easily to the multivariable setting.

Definition 3.5. For distinct d-tuples $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$ in $\mathbf{N}^d$, set $\mathbf{i} > \mathbf{j}$ if, for the first r such that $i_r \neq j_r$, we have $i_r > j_r$. Write $\mathbf{i} \geq \mathbf{j}$ if $\mathbf{i} > \mathbf{j}$ or $\mathbf{i} = \mathbf{j}$.

Example 3.6. In $\mathbf{N}^4$, $(5, 1, 1, 3) > (3, 0, 2, 4)$ and $(3, 0, 3, 1) > (3, 0, 2, 4)$.

Example 3.7. In $\mathbf{N}^d$, $\mathbf{0} < \mathbf{i}$ for all $\mathbf{i} \neq \mathbf{0}$.

For all d-tuples $\mathbf{i}$ and $\mathbf{j}$ in $\mathbf{N}^d$, either $\mathbf{i} = \mathbf{j}$, $\mathbf{i} > \mathbf{j}$, or $\mathbf{j} > \mathbf{i}$, so $\mathbf{N}^d$ is totally ordered by $>$. (For example, $\mathbf{i} > \mathbf{0}$ for all $\mathbf{i} \neq \mathbf{0}$.) This way of ordering of d-tuples is called *lexicographic* (*i.e.*, dictionary) *ordering* since it resembles the way words are ordered in the dictionary alphabetically if we think of one word as "greater" than another if it comes later in the dictionary. Alphabetical order compares words by the first letter, if that letter is the same the words are compared by the second letter, and so on. While words in a dictionary have varying length, we are using lexicographic ordering only to compare sequences in $\mathbf{N}$ with the same number of terms.

Lemma 3.8. *Lexicographic ordering on $\mathbf{N}^d$ has the following properties.*

(1) (*Total ordering*) *For all $\mathbf{i}$ and $\mathbf{j}$, exactly one of $\mathbf{i} = \mathbf{j}$ or $\mathbf{i} < \mathbf{j}$ or $\mathbf{j} < \mathbf{i}$ holds.*
(2) (*Transitivity*) *If $\mathbf{i} < \mathbf{j}$ and $\mathbf{j} < \mathbf{k}$ then $\mathbf{i} < \mathbf{k}$. The same is true with $\leq$ in place of $<$.*
(3) (*Compatibility with addition*) *If $\mathbf{i} \leq \mathbf{i}'$ and $\mathbf{j} \leq \mathbf{j}'$ then $\mathbf{i} + \mathbf{j} \leq \mathbf{i}' + \mathbf{j}'$, and if either inequality in the hypothesis is strict then the inequality in the conclusion is strict.*

Proof. (1) If $\mathbf{i} \neq \mathbf{j}$, then there is an r where $i_r \neq j_r$ in $\mathbf{N}$. Let r be the least index where this happens. If $i_r < j_r$ then $\mathbf{i} < \mathbf{j}$, and if $j_r < i_r$ then $\mathbf{j} < \mathbf{i}$.

(2) Let r be the least index where i_r, j_r, and k_r are not all equal. We must have $i_r \neq j_r$ or $j_r \neq k_r$ (if both were equalities then $i_r = j_r = k_r$, which isn't true). Since earlier coordinates in $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are all equal, either $i_r < j_r$ or $j_r < k_r$ because $\mathbf{i} < \mathbf{j}$ and $\mathbf{j} < \mathbf{k}$. Therefore

$i_r \leq j_r \leq k_r$ with at least one inequality being strict, so $i_r < k_r$ and earlier coordinates in $\mathbf{i}$ and $\mathbf{k}$ are equal. Thus $\mathbf{i} < \mathbf{k}$.

This result for $\leq$ is the same argument as with $<$ except we have the extra cases where $\mathbf{i}$ and $\mathbf{j}$ may coincide or $\mathbf{j}$ and $\mathbf{k}$ may coincide, which makes things easier.

(3) Rather than take cases based on where $\mathbf{i}$ and $\mathbf{i}'$ may first differ or where $\mathbf{j}$ and $\mathbf{j}'$ may first differ, observe that $\mathbf{i} \leq \mathbf{i}' \Rightarrow \mathbf{i} + \mathbf{k} \leq \mathbf{i}' + \mathbf{k}$ for all $\mathbf{k}$: this is obvious when $\mathbf{i} = \mathbf{i}'$, and when $\mathbf{i} < \mathbf{i}'$ the only way $i_r + k_r$ differs from $i'_r + k_r$ is if $i_r \neq i'_r$, and the first time this happens we have $i_r < i'_r$, so $i_r + k_r < i'_r + k_r$.

Now we use that twice together with the transitivity in (2). If $\mathbf{i} \leq \mathbf{i}'$ and $\mathbf{j} \leq \mathbf{j}'$, then

$$\mathbf{i} + \mathbf{j} \leq \mathbf{i}' + \mathbf{j} = \mathbf{j} + \mathbf{i}' \leq \mathbf{j}' + \mathbf{i}' = \mathbf{i}' + \mathbf{j}' \Longrightarrow \mathbf{i} + \mathbf{j} \leq \mathbf{i}' + \mathbf{j}'.$$

The case of $<$ in place of $\leq$ is analogous. □

In $\mathbf{N}$ with its usual ordering, there are finitely many elements below a given element, but this is not true in $\mathbf{N}^d$ for $d \geq 2$ with lexicographic ordering: there can be infinitely many d-tuples below some d-tuple. For instance, $(0, b) < (1, 0)$ for all $b \in \mathbf{N}$. But $\mathbf{N}^d$ does share with $\mathbf{N}$ the following important ordering property.

Lemma 3.9. *When $\mathbf{N}^d$ has lexicographic ordering, each nonempty subset of $\mathbf{N}^d$ has a least element.*

The least element in the subset has to be unique since lexicographic ordering is a total ordering on $\mathbf{N}^d$.

Proof. Let S be a nonempty subset of $\mathbf{N}^d$. The idea behind getting a least element of S is as follows. The first coordinates of d-tuples in S are a nonempty subset of $\mathbf{N}$ and thus have a least element ℓ_1. If $d = 1$ then ℓ_1 is the least element of S. For $d \geq 2$, among the elements of S with first coordinate ℓ_1, their second coordinates are a nonempty subset of $\mathbf{N}$ and thus have a least element ℓ_2. If $d = 2$ then (ℓ_1, ℓ_2) is the least element of S. For $d \geq 3$, among the elements of S with first coordinate ℓ_1 and second coordinate ℓ_2, their third coordinates are a nonempty subset of $\mathbf{N}$ and thus have a least element ℓ_3. Continue this way up through the dth coordinate. The d-tuple $(\ell_1, \ell_2, \ldots, \ell_d)$ in S is the least element of S by the definition of lexicographic ordering on $\mathbf{N}^d$.

It is left to the reader to rewrite this argument as a proper proof by induction on d. □

A polynomial $f \in A[x_1, \ldots, x_d]$ is a sum of the form

$$f = \sum_{i_1, \ldots, i_d \geq 0} a_{i_1, \ldots, i_d} x_1^{i_1} \cdots x_d^{i_d}$$

where $a_{i_1,\ldots,i_d} \in A$ and only finitely many coefficients can be nonzero. Abbreviate this sum to multi-index form as $\sum_{\mathbf{i}} a_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$, where $\mathbf{x}^{\mathbf{i}} := x_1^{i_1} \cdots x_d^{i_d}$ for $\mathbf{i} = (i_1, \ldots, i_d)$. Note $\mathbf{x}^{\mathbf{i}} \mathbf{x}^{\mathbf{j}} = \mathbf{x}^{\mathbf{i}+\mathbf{j}}$. (In the notation $\sum_{\mathbf{i}}$, only finitely many terms are nonzero.) When $f \neq 0$, lexicographic ordering lets us compare the different nonzero monomials appearing in f, which leads to the following concepts that generalize degree and leading terms on $A[x]$.

Definition 3.10. Write f in $A[x_1, \ldots, x_d]$ as $\sum_{\mathbf{i}} a_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$. If $f \neq 0$, the *multidegree* of f is the lexicographically largest index in $\mathbf{N}^d$ of a nonzero monomial in f:

$$\operatorname{mdeg} f = \max\{\mathbf{i} : a_{\mathbf{i}} \neq \mathbf{0}\} \in \mathbf{N}^d.$$

The multidegree of the zero polynomial is not defined. If $f \neq 0$ and $\operatorname{mdeg} f = \mathbf{n}$, we call $a_{\mathbf{n}} \mathbf{x}^{\mathbf{n}}$ the *leading term* of f and $a_{\mathbf{n}}$ the *leading coefficient* of f.

We can separate the leading term of a nonzero f of multidegree $\mathbf{n}$ from its other nonzero terms to get

$$f = a_{\mathbf{n}}\mathbf{x}^{\mathbf{n}} + \sum_{\mathbf{i}<\mathbf{n}} a_{\mathbf{i}}\mathbf{x}^{\mathbf{i}}.$$

Because multidegrees are totally ordered, a nonzero polynomial in $A[x_1, \ldots, x_d]$ has a unique leading term and a unique leading coefficient in A.

Example 3.11. In $\mathbf{Z}[x_1, x_2]$, let $f = 7x_1x_2^5 + 3x_2^8 + 9$. Since $\max((1,5),(0,8),(0,0)) = (1,5)$ in $\mathbf{N}^2$, $\mathrm{mdeg}(f) = (1,5)$ and $\mathrm{lead}(7x_1x_2^5 + 3x_2^8 + 9) = 7$.

Example 3.12. In $A[x_1, \ldots, x_d]$, $\mathrm{mdeg}(x_1) = (1, 0, \ldots, 0)$ and $\mathrm{mdeg}(x_d) = (0, 0, \ldots, 1)$.

Example 3.13. The polynomials with multidegree $\mathbf{0}$ are the nonzero constants.

Our definition of multidegree is specific to calling x_1 the "first" indeterminate, x_2 the "second" indeterminate, and so on up to x_d being the "last" one. There is nothing intrinsic about declaring x_1 the "first" indeterminate: there are $d!$ different lexicographic orderings based on which coordinate in $\mathbf{N}^d$ we want to consider first, second, and so on. Having made a choice, the corresponding multidegree function permits us to prove theorems about multivariable polynomials by ordering them according to their multidegree in the same way theorems about single-variable polynomials are proved by ordering them by degree.

Remark 3.14. A simpler way to order nonzero multivariable polynomials f is by their "total degree", which is the largest sum of exponents among the nonzero monomials appearing in f, *e.g.*, $x_1x_2^5 + x_2^8$ has total degree 8. This degree function on nonzero polynomials in $A[x_1, \ldots, x_d]$ has values in $\mathbf{N}$ rather than $\mathbf{N}^d$. It is useful for some purposes, but not for what we want to do because it does not provide a unique leading term for nonzero polynomials $A[x_1, \ldots, x_d]$ when $d > 1$ since different monomials in d indeterminates can have the same total degree. For instance, $x_1x_2^5$, $x_1^3x_2^3$, and x_2^6 all have total degree 6. By contrast, in the lexicographic ordering $x_1^3x_2^3 > x_1x_2^5 > x_2^6$.

Now we'll prove Corollary 3.4 a second way. Compare it to the proof of Theorem 2.5.

Theorem 3.15. *If f is a zero divisor in $A[x_1, \ldots, x_d]$ then $af = 0$ for some nonzero $a \in A$.*

Proof. For a zero divisor $f \in A[x_1, \ldots, x_d]$, $gf = 0$ for some nonzero g. We can pick such g with minimal multidegree in $\mathbf{N}^d$: the set $\{g \in A[x_1, \ldots, x_n] : gf = 0 \text{ and } g \neq 0\}$ is nonempty and the multidegrees of polynomials in this set are a nonempty subset of $\mathbf{N}^d$, so there is a least multidegree for polynomials in that set by Lemma 3.9.

When $\mathrm{mdeg}(g)$ is least, assume $\mathrm{mdeg}(g) \neq \mathbf{0}$. Then $\mathrm{mdeg}(g) > \mathbf{0}$ and $f \neq 0$. We will show $\widetilde{g}f = 0$ where $\widetilde{g} \neq 0$ and $\mathrm{mdeg}(\widetilde{g}) < \mathrm{mdeg}(g)$. That contradicts $\mathrm{mdeg}(g)$ being least, so $\mathrm{mdeg}(g) = \mathbf{0}$, which means $af = 0$ for some nonzero $a \in A$.

Set $\mathbf{n} = \mathrm{mdeg}(f)$ and $\mathbf{m} = \mathrm{mdeg}(g)$ in $\mathbf{N}^d$, so

$$f = \sum_{\mathbf{i}} a_{\mathbf{i}}\mathbf{x}^{\mathbf{i}} = a_{\mathbf{n}}\mathbf{x}^{\mathbf{n}} + \text{lower multidegree terms},$$

$$g = b_{\mathbf{m}}\mathbf{x}^{\mathbf{m}} + \text{lower multidegree terms},$$

where $a_{\mathbf{n}} \neq 0$ and $b_{\mathbf{m}} \neq 0$ in A. Since $af \neq 0$ for all nonzero a in A, $b_{\mathbf{m}}f \neq 0$. Therefore $b_{\mathbf{m}}a_{\mathbf{i}} \neq 0$ for some $\mathbf{i}$, so $ga_{\mathbf{i}} \neq 0$ for that $\mathbf{i}$. Let $\mathbf{j}$ be maximal in $\mathbf{N}^d$ with $ga_{\mathbf{j}} \neq 0$, so $ga_{\mathbf{i}} = 0$ for $\mathbf{i} > \mathbf{j}$. It is crucial that we are using a total ordering on monomials, both to have unique leading terms for f and g and to have a unique maximal $\mathbf{j}$ such that $a_{\mathbf{j}}g \neq 0$. (If we

ordered multivariable polynomials by total degree instead of multidegree, there would not be a unique leading term for nonzero polynomials when $d \geq 2$.)

Since $gf = 0$,

$$0 = gf = (b_{\mathbf{m}}\mathbf{x}^{\mathbf{m}} + \text{lower multidegree terms})\left(a_{\mathbf{j}}\mathbf{x}^{\mathbf{j}} + \text{lower multidegree terms}\right),$$

where we dropped the terms $a_{\mathbf{i}}\mathbf{x}^{\mathbf{i}}$ from f in the product when $\mathbf{i} > \mathbf{j}$ since $ga_{\mathbf{i}} = 0$ for $\mathbf{i} > \mathbf{j}$. By Lemma 3.8, the only $\mathbf{x}^{\mathbf{m}+\mathbf{j}}$-term on the right is $(b_{\mathbf{m}}\mathbf{x}^{\mathbf{m}})(a_{\mathbf{j}}\mathbf{x}^{\mathbf{j}}) = b_{\mathbf{m}}a_{\mathbf{j}}\mathbf{x}^{\mathbf{m}+\mathbf{j}}$, so $b_{\mathbf{m}}a_{\mathbf{j}} = 0$. Since $b_{\mathbf{m}}\mathbf{x}^{\mathbf{m}}$ is the leading term of g and $b_{\mathbf{m}}a_{\mathbf{j}} = 0$, $\mathrm{mdeg}(ga_{\mathbf{j}}) < \mathbf{m} = \mathrm{mdeg}(g)$, where $ga_{\mathbf{j}} \neq 0$ by maximality of $\mathbf{j}$, so $\mathrm{mdeg}(ga_{\mathbf{j}})$ makes sense.

Since $(ga_{\mathbf{j}})f = a_{\mathbf{j}}gf = a_{\mathbf{j}} \cdot 0 = 0$ and $\mathrm{mdeg}(ga_{\mathbf{j}}) < \mathrm{mdeg}(g)$, we have contradicted the minimality of $\mathrm{mdeg}(g)$ among nonzero polynomials g in $A[x_1, \ldots, x_d]$ such that $gf = 0$ if we assume $\mathrm{mdeg}(g) > \mathbf{0}$. Therefore $\mathrm{mdeg}(g) = \mathbf{0}$, so if f is a zero divisor in $A[x_1, \ldots, x_d]$ then $af = 0$ for some nonzero a in A. □

Corollary 3.16. *For $f = \sum_{\mathbf{i}} a_{\mathbf{i}}\mathbf{x}^{\mathbf{i}}$ in $\mathbf{Z}[x_1, \ldots, x_d]$ and $m \geq 2$, $f \bmod m$ is a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x_1, \ldots, x_d]$ if and only if $\gcd(\{a_{\mathbf{i}}\}, m) > 1$.*

Proof. The case $d = 1$ is proved in Corollary 2.7 and that proof works with no changes for all d. □

Corollary 3.17. *For $f = \sum_{\mathbf{i}} a_{\mathbf{i}}\mathbf{x}^{\mathbf{i}}$ in $\mathbf{Z}[x_1, \ldots, x_d]$, $\gcd(\{a_{\mathbf{i}}\}) = 1$ if and only if $f \bmod m$ is not a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x_1, \ldots, x_d]$ for all $m \geq 2$.*

Proof. If $\gcd(\{a_{\mathbf{i}}\}) = 1$, then for $m \geq 2$ we have $\gcd(\{a_{\mathbf{i}}\}, m) = 1$, so $f \bmod m$ is not a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x_1, \ldots, x_d]$ by Corollary 3.16.

If $\gcd(\{a_{\mathbf{i}}\}) > 1$, set $b = \gcd(\{a_{\mathbf{i}}\})$. For all m such that $\gcd(b, m) > 1$ (such as m being a multiple of b), $\gcd(\{a_{\mathbf{i}}\}, m) = \gcd(b, m) > 1$ and therefore $f \bmod m$ is a zero divisor in $(\mathbf{Z}/m\mathbf{Z})[x_1, \ldots, x_d]$ by Corollary 3.16. □

References

[1] N. H. McCoy, "Remarks on divisors of zero," Amer. Math. Monthly **49** (1942), 286–295. Online at https://www.jstor.org/stable/pdf/2303094.

[2] N. H. McCoy, "Annihilators in polynomial rings," Amer. Math. Monthly **64** (1957), 28–29. Online at https://www.jstor.org/stable/2309082.

[3] W. R. Scott, "Divisors of zero in polynomial rings," Amer. Math. Monthly **61** (1954), 336. Online at https://www.jstor.org/stable/pdf/2307474.