Noether normalization and Hilbert's Nullstellensatz

We prove the Noether normalization theorem over a field and, more generally, over an integral domain. We then deduce Hilbert's Nullstellensatz.

The following result implies that, after a change of variables, any nonzero polynomial in $R = K[x_1, \ldots, x_n]$, the polynomial ring in in n variables over a field, becomes a nonzero scalar times a polynomial that is monic in x_n with coefficients in $A = K[x_1, \ldots, x_{n-1}] \subseteq R$, where we think of R as $A[x_n]$. We may also do this with any one of the other variables. This simple trick, or method, provides a wealth of information about algebras finitely generated over a field. It will be the key to our proofs of the Noether normalization theorem and Hilbert's Nullstellensatz.

Consider this example: the polynomial x_1x_2 is not monic in either variable. But there is an automorphism of the polynomial ring in two variables that fixes x_2 and maps x_1 to $x_1 + x_2$. (Its inverse fixes x_2 and maps x_1 to $x_1 - x_2$.) The image of x_1x_2 is $(x_1 + x_2)x_2 = x_2^2 + x_1x_2$. As a polynomial in x_2 over $K[x_1]$, this is monic. Note that we may also think of the effect of applying an automorphism as a change of variables.

More generally, note that if $g_1(x_n), \ldots, g_{n-1}(x_n)$ are arbitrary elements of $K[x_n] \subseteq R$, then there is a K-automorphism ϕ of R such that $x_i \mapsto y_i = x_i + g_i(x_n)$ for $i < n$ and while $x_n = y_n$ is fixed. The inverse automorphism is such that $x_i \mapsto x_i - g_i(x_n)$ while x_n is again fixed. This means that the elements y_i are algebraically independent and generate $K[x_1, \ldots, x_n]$. They are "just as good" as our original indeterminates.

Lemma. *Let D be a domain and let $f \in D[x_1, \ldots, x_n]$. Let $N \geq 1$ be an integer that bounds all the exponents of the variables occurring in the terms of f. Let ϕ be the D-automorphism of $D[x_1, \ldots, x_n]$ such that $x_i \mapsto x_i + x_n^{N^i}$ for $i < n$ and such that x_n maps to itself. Then the image of f under ϕ is a polynomial whose highest degree term involving x_n has the form cx_n^m, where c is a nonzero element of D. In particular, if $D = K$ is a field, then the image of f is a nonzero scalar of the field times a polynomial that is monic in x_n when considered as a polynomial over $K[x_1, \ldots, x_{n-1}]$.*

Proof. Consider any nonzero term of f, which will have the form $c_\alpha x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$, where $\alpha = (a_1, \ldots, a_n)$ and c_α is a nonzero element in D. The image of this term under ϕ is

$$c_\alpha (x_1 + x_n^N)^{a_1} (x_2 + x_n^{N^2})^{a_2} \cdots (x_{n-1} + x_n^{N^{n-1}})^{a_{n-1}} x_n^{a_n},$$

and this contains a unique highest degree term: it is the product of the highest degree terms coming from all the factors, and it is

$$c_\alpha (x_n^N)^{a_1} (x_n^{N^2})^{a_2} \cdots (x_n^{N^{n-1}})^{a_{n-1}} x_n^{a_n} = c x_n^{a_n + a_1 N + a_2 N^2 + \cdots + a_{n-1} N^{n-1}}.$$

The exponents that one gets on x_n in these largest degree terms coming from distinct terms of f are all distinct, because of uniqueness of representation of integers in base N. Thus, no two exponents are the same, and no two of these terms can cancel. Therefore, the degree m of the image of f is the same as the largest of the numbers

$$a_n + a_1 N + a_2 N^2 + \cdots + a_{n-1} N^{n-1}$$

as $\alpha = (a_1, \ldots, a_n)$ runs through n-tuples of exponents occurring in nonzero terms of f, and for the choice α_0 of α that yields m, $c_{\alpha_0} x_n^m$ occurs in $\phi(f)$, is the only term of degree m, and and cannot be canceled. When $D = K$ is a field, it follows that $c_{\alpha_0}^{-1}\phi(f)$ is monic of degree m in x_n when viewed as a polynomial in $A[x_n]$, as required. $\square$

Let R be an A-algebra and $z_1, \ldots, z_d \in R$. We shall say that the elements $z_1, \ldots, z_d$ are *algebraically independent* over A if the unique A-algebra homomorphism from the polynomial ring $A[x_1, \ldots, x_d] \to R$ that sends x_i to z_i for $1 \le i \le n$ is an isomorphism. An equivalent statement is that the mononomials $z_1^{a_1} \cdots z_d^{a_d}$ as $(a_1, \ldots, a_d)$ varies in $\mathbb{N}^d$ are all distinct and span a free A-submodule of R: of course, this free A-submodule is $A[z_1, \ldots, z_d]$. The failure of the z_j to be algebraically independent means precisely that there is some nonzero polynomial $f(x_1, \ldots, x_d) \in A[x_1, \ldots, x_d]$ such that $f(z_1, \ldots, z_d) = 0$. The following is now easy:

In the proof result below, we localize successively at several nonzero elements in a domain D. Note that if we localize D at an element $c \neq 0$, and then localize D_c at an element $b/c^k \neq 0$, we get the same result D_{bc} as if we had localized D at the single element bc. Therefore, by induction, the effect of a finite number of localizations at nonzero elements is the same as the result of localizing the original domain at one nonzero element.

Noether normalization theorem. *Let D be an integral domain and let R be any finitely generated D-algebra extension of D. Then there is a nonzero element $c \in D$ and elements $z_1, \ldots, z_d$ in R_c algebraically independent over D_c such that R_c is module-finite over its subring $D_c[z_1, \ldots, z_d]$, which is isomorphic to a polynomial ring (d may be zero) over D_c. In particular, if $D = K$, a field, then it is not necessary to invert an element: every finitely generated K-algebra is isomorphic with a module-finite extension of a polynomial ring!*

Proof. We use induction on the number n of generators of R over D. If $n = 0$ then $R = D$. We may take $d = 0$. Now suppose that $n \ge 1$ and that we know the result for algebras generated by $n-1$ or fewer elements. Suppose that $R = D[\theta_1, \ldots, \theta_n]$ has n generators. If the θ_i are algebraically independent over K then we are done: we may take $d = n$ and $z_i = \theta_i$, $1 \le i \le n$. Therefore we may assume that we have a nonzero polynomial $f(x_1, \ldots, x_n) \in D[x_1, \ldots, x_n]$ such that $f(\theta_1, \ldots, \theta_n) = 0$. Instead of using the original θ_j as generators of our K-algebra, note that we may use instead the elements

$$\theta_1' = \theta_1 - \theta_n^N,\, \theta_2' = \theta_2 - \theta_n^{N^2},\, \ldots,\, \theta_{n-1}' = \theta_{n-1} - \theta_n^{N^{n-1}},\, \theta_n' = \theta_n$$

where N is chosen for f as in the preceding Lemma. With ϕ as in that Lemma, we have that these new algebra generators satisfy $\phi(f) = f(x_1 + x_n^N, \ldots, x_{n-1} + x_n^{N^{n-1}}, x_n)$ which we shall write as g. We replace D and R by their localizations at D_c and R_c, where c is the coefficient of the highest power of x_n occurring, so that the polynomial may be replaced by a multiple that is monic in x_n. After multiplying by a unit of D_c, we have that g is monic in x_n with coefficients in $D_c[x_1, \ldots, x_{n-1}]$. This means that θ_n' is integral over $D_c[\theta_1', \ldots, \theta_{n-1}'] = R_0$, and so R_c is module-finite over R_0. Since R_0 has $n-1$ generators over R_c, we have by the induction hypothesis that R_0 is module-finite over a polynomial $R_{cc'}[z_1, \ldots, z_d] \subseteq R_0$, and then $R_{cc'}$ is module-finite over $D_{cc'}[z_1, \ldots, z_d]$ as well. $\square$

Note that if $K \subseteq L$ are fields, the statement that L is module-finite over K is equivalent to the statement that L is a finite-dimensional vector space over K, and both are equivalent to the statement that L is a finite algebraic extension of K.

Also notice that the polynomial ring $R = K[x_1, \ldots, x_d]$ for $d \geq 1$ has dimension at least d: $(0) \subset (x_1)R \subset (x_1, x_2)R \subset \cdots \subset (x_1, \ldots, x_d)R$ is a strictly increasing chain of prime ideals of length d. Later we shall show that the dimension of $K[x_1, \ldots, x_d]$ is exactly d. But for the moment, all we need is that $K[x_1, \ldots, x_d]$ has dimension at least one for $d \geq 1$.

Corollary (Zariski's Lemma). *Let R be a finitely generated algebra over a field K, and suppose that R is a field. Then R is a finite algebraic extension of K, i.e., R is module-finite over K.*

Proof. By the Noether normalization theorem, R is module-finite over some polynomial subring $K[z_1, \ldots, z_d]$. If $d \geq 1$, the polynomial ring has dimension at least one, and then R has dimension at least one, a contradiction. Thus, $d = 0$, and R is module-finite over K. Since R is a field, this means precisely that R is a finite algebraic extension of K. $\square$

Corollary. *Let K be an algebraically closed field, let R be a finitely generated K-algebra, and let m be a maximal ideal of R. Then the composite map $K \to R \twoheadrightarrow R/m$ is an isomorphism.*

Proof. R/m is a finitely generated K-algebra, since R is, and it is a field. Thus, $K \to R/m$ gives a finite algebraic extension of K. Since K is algebraically closed, it has no proper algebraic extension, and so $K \to R/m$ must be an isomorphism.

Corollary (Hilbert's Nullstellensatz, weak form). *Let $R = K[x_1, \ldots, x_n]$ be a polynomial ring over and algebraically closed field K. Then every maximal ideal m of R is the kernel of a K-homomorphism $K[x_1, \ldots, x_n] \to K$, and so is determined by the elements $\lambda_1, \ldots, \lambda_n \in K$ to which $x_1, \ldots, x_n$ map. This maximal ideal is the kernel of the evaluation map $f(x_1, \ldots, x_n) \mapsto f(\lambda_1, \ldots, \lambda_n)$. It may also be described as the ideal $(x_1 - \lambda_1, \ldots, x_n - \lambda_n)R$.*

Proof. Since $\gamma : K \cong R/m$, the K-algebra map $R \to R/m$, composed with γ^{-1}, gives a map $R \twoheadrightarrow K$ whose kernel is m. $\square$

Thus, when K is algebraically closed, we have a bijection between the points of K^n and the maximal ideals of $K[x_1, \ldots, x_n]$.

Corollary (Hilbert's Nullstellensatz, alternate weak form). *Let $f_1, \ldots, f_n$ be polynomials in $K[x_1, \ldots, x_n]$, where K is algebraically closed. Then then the f_i generate the unit ideal (i.e., we have $1 = \sum_t g_t f_t$ for suitable polynomials g_t) if and only if the polynomials f_i do not vanish simultaneously, i.e., if and only if the algebraic set $V(f_1, \ldots, f_n) = \emptyset$.*

Proof. If the f_i do not generate the unit ideal, the ideal they generate is contained in some maximal ideal of $K[x_1, \ldots, x_n]$. But the functions in that maximal ideal all vanish at one point of K^n, a contradiction. On the other hand, if the f_i all vanish simultaneously at a point of K^n, they are in the maximal ideal of polynomials that vanish at that point: this direction does not need that K is algebraically closed. $\square$

We have two uses of the notation $V(S)$: one is for any subset S of any ring, and it is the set of all primes containing S. The other use is for polynomial rings $K[x_1, \dots, x_n]$, and then it is the set of points where the given polynomials vanish. For clarity, suppose that we use $\mathcal{V}$ for the second meaning. If we think of these points as corresponding to a subset of the maximal ideals of the ring (it corresponds to all maximal ideals when the field is algebraically closed), we have that $\mathcal{V}(S)$ is the intersection of $V(S)$ with the maximal ideals corresponding to points of K^n, thought of as a subset of K^n. Suppose that for every $y \in K^n$ we let $m_y = \{f \in K[x_1, \dots, x_n] : f(y) = 0\}$. Then m_y is a maximal ideal of $K[x_1, \dots, x_n]$ whether K is algebraically closed or not. When K is algebraically closed, we know that all maximal ideals have this form. This gives an injection $K^n \to \mathrm{Spec}\,(R)$ that sends y to m_y. The closed algebraic sets of K^n are simply the closed sets of $\mathrm{Spec}\,(R)$ intersected with the image of K^n, if we identify that image with K^n. Thus, the algebraic sets are the closed sets of a topology on K^n, which is called the *Zariski topology.* It is the inherited Zariski topology from $\mathrm{Spec}\,(R)$. Note that $\mathcal{V}(I) = \{y \in Y : m_y \in V(I)\}$.

In this course, I will continue from here on to use the alternate notation $\mathcal{V}$ when discussing algebraic sets. However, people often use the same notation for both, depending on the context to make clear which is meant.

Theorem (Hilbert's Nullstellensatz, strong form. *Let K be an algebraically closed field and let $R = K[x_1, \dots, x_n]$ be the polynomial ring in n variables over K. Suppose that $g, f_1, \dots, f_s \in R$. Then $g \in \mathrm{Rad}\,(f_1, \dots, f_s)$ if and only if $\mathcal{V}(g) \supseteq V(f_1, \dots, f_s)$, i.e., if and only if g vanishes at every point where the f_i vanish simultaneously.*

Proof. It is clear that $g^N = \sum_{i=1}^{s} g_i f_i$ implies that g vanishes wherever the all of the f_i vanish: at such a point y, we have that $g(y)^N = 0$ and so $g(y) = 0$.

The more interesting implication is the statement that if g does vanish whenever all the f_i vanish then g has a power that is in the ideal generated by the f_i. The following method of proof is called Rabinowitsch's trick. Introduce an extra variable z and consider the polynomials $f_1, \dots, f_s, 1 - gz \in K[x_1, \dots, x_n, z]$. There is no point of K^{n+1} where these all vanish: at any point where the f_i vanish (this only depends on what the first n coordinates of the point are), we have that g vanishes as well, and therefore $1 - gz$ is $1 - 0 = 1$. This means that $f_1, \dots, f_s, 1 - gz$ generate the unit ideal in $K[x_1, \dots, x_n, z]$, by the weak form of Hilbert's Nullstellensatz that we have already established. This means that there is an equation

$$1 = H_1(z)f_1 + \cdots + H_s(z)f_s + H(z)(1 - gz)$$

where $H_1(z), \dots, H_s(z)$ and $H(z)$ are polynomials in $K[x_1, \dots, x_n, z]$: all of them may involve all of the variables x_j and z, but we have chosen a notation that emphasizes their dependence on z. But note that $f_1, \dots, f_s$ and g do not depend on z. We may assume that $g \neq 0$ or the result is obvious. We now define a $K[x_1, \dots, x_n]$-algebra map ϕ from $K[x_1, \dots, x_n, z]$, which we think of as $K[x_1, \dots, x_n][z]$, to the ring $K[x_1, \dots, x_n][1/g] = K[x_1, \dots, x_n]_g$, which we may think of as a subring of the fraction field of $K[x_1, \dots, x_n]$. This ring is also the localization of $K[x_1, \dots, x_n]$ at the multiplicative system $\{1, g, g^2, \dots\}$ consisting of all powers of g. Note that every element of $K[x_1, \dots, x_n]_g$ can be written in the form u/g^h, where $u \in K[x_1, \ldots, x_n]$ and h is some nonnegative integer. We define the $K[x_1, \ldots, x_n]$-algebra map ϕ simply by specifying that the value of z is to be $1/g$. Applying this homomorphism to the displayed equation, we find that

$$1 = H_1(1/g)f_1 + \cdots + H_s(1/g)f_s + H(1/g)(1-1)$$

or

$$1 = H_1(1/g)f_1 + \cdots + H_s(1/g)f_s.$$

Since each of the $H_i(1/g)$ is in $K[x_1, \ldots, x_n]_g$, we can choose a positive integer N so large that each of the $g_i = g^N H_i(1/g) \in K[x_1, \ldots, x_n]$: there are only finitely many denominators to clear. Multiplying the most recently displayed equation by g^N gives the equation $g^N = g_1 f_1 + \cdots + g_n f_n$ with $g_i \in K[x_1, \ldots, x_n]$, which is exactly what we wanted to prove. $\square$

Corollary. *Let $R \to S$ be a homomorphism of finitely generated K-algebras. Then every maximal ideal of S contracts to a maximal ideal of R.*

Proof. Suppose that the maximal ideal n of S contracts to the prime P in R, so that $K \subseteq R/P \subseteq S/n$. Then S/n is a finite algebraic extension of K, i.e., a finite dimensional K-vector space, and so the domain R/P is a finite-dimensional K-vector space, i.e., it is module-finite over K, and therefore it is a domain of dimension 0, which forces it to be a field. $\square$