COUNTING ROOTS OF POLYNOMIALS

KEITH CONRAD

In $\mathbf{R}[T]$, a linear polynomial $aT+b$ has exactly one root in $\mathbf{R}$: $at+b=0$ if and only if $t=-b/a$. By the quadratic formula, a quadratic polynomial in $\mathbf{R}[T]$ has at most 2 roots in $\mathbf{R}$. Even though there is not an analogue of the quadratic formula for roots of all polynomials (especially in degree 5 and up), the bound we described on the number of roots in degrees 1 and 2 in $\mathbf{R}[T]$ is valid in all degrees when the coefficients are in an arbitrary field and we will prove this by induction on the degree.

Theorem 1. *Let $f(T)$ be a nonzero polynomial of degree d with coefficients in a field F. Then $f(T)$ has at most d roots in F.*

We can't replace "at most d roots" with "exactly d roots" since there are nonconstant polynomials with no roots: T^2+1 in $\mathbf{R}[T]$ has no roots in $\mathbf{R}$ and T^3-2 in $\mathbf{Q}[T]$ has no roots in $\mathbf{Q}$.

Proof. We induct on the degree of polynomials. Each step in the induction is about all polynomials of a common degree: the theorem in degree 0, then in degree 1, then in degree 2, then in degree 3, and so on.

The base case is degree 0. A polynomial of degree 0 in $F[T]$ is a nonzero constant polynomial, so it has no roots at all.

Now assume the theorem is true for all polynomials in $F[T]$ of degree d for some $d \geq 0$. We will prove the theorem is true for all polynomials in $F[T]$ of degree $d+1$.

A polynomial of degree $d+1$ in $F[T]$ has the form

$$f(T) = c_{d+1}T^{d+1} + c_dT^d + \cdots + c_1T + c_0, \tag{1}$$

where $c_0, \ldots, c_{d+1} \in F$ and $c_{d+1} \neq 0$. To bound the number of roots of $f(T)$ in F, we consider two cases.

Case 1. If $f(T)$ has no root in F, then we're done since $0 \leq d+1$.

Case 2. If $f(T)$ has a root in F, say r, then

$$0 = c_{d+1}r^{d+1} + c_dr^d + \cdots + c_1r + c_0. \tag{2}$$

From this condition we can show $T-r$ is a factor of $f(T)$: $f(T) = (T-r)Q(T)$ for some $Q(T)$ in $F[T]$. Here are two different ways of doing that.

Method 1. Divide $f(T)$ by $T-r$ using the division algorithm in $F[T]$. The remainder is 0 or is nonzero with degree less than $\deg(T-r) = 1$, so either way the remainder is constant:

$$f(T) = (T-r)Q(T) + c$$

for some $c \in F$. To find c, set $T=r$: $0 = 0 \cdot Q(0) + c = c$, so $f(T) = (T-r)Q(T)$.

Method 2. Subtract (2) from (1). The constant terms c_0 cancel and we get

$$f(T) = c_{d+1}(T^{d+1} - r^{d+1}) + c_d(T^d - r^d) + \cdots + c_1(T-r). \tag{3}$$

Each difference $T^j - r^j$ for $j = 1, 2, \ldots, d+1$ has $T - r$ as a factor:

$$T^j - r^j = (T - r)(T^{j-1} + rT^{j-2} + \cdots + r^iT^{j-1-i} + \cdots + r^{j-2}T + r^{j-1}).$$

Write the more complicated second factor, a polynomial of degree $j-1$, as $Q_{j,r}(T)$. So

$$T^j - r^j = (T - r)Q_{j,r}(T), \tag{4}$$

and substituting (4) into (3) gives

$$f(T) = \sum_{j=1}^{d+1} c_j(T - r)Q_{j,r}(T) = (T - r)\sum_{j=1}^{d+1} c_jQ_{j,r}(T) = (T - r)Q(T),$$

where $Q(T) = \sum_{j=1}^{d+1} c_jQ_{j,r}(T)$.

By either method, from $f(T) = (T - r)Q(T)$ we take degrees on both sides to see $d + 1 = 1 + \deg Q$, so $\deg Q = d$.

A root of $f(T)$ in F is either r or is a root of $Q(T)$. Indeed, for $s \in F$ we have

$$f(s) = (s - r)Q(s),$$

so if $f(s) = 0$ then $(s - r)Q(s) = 0$, which means $s - r = 0$ or $Q(s) = 0$: $s = r$ or s is a root of $Q(s)$. By the inductive hypothesis, $Q(T)$ has at most d roots in F, so $f(T)$ has at most $d + 1$ roots: s and the roots of $Q(T)$ in F.

Since $f(T)$ was an arbitrary polynomial of degree $d+1$ in $F[T]$, we have shown that the d-th case of the theorem being true implies the $(d+1)$-th case is true. By induction on the degree, the theorem is true for all nonconstant polynomials. □

Corollary 2. *If F is a field and $f(T) \in F[T]$ is nonconstant, then for each $c \in F$ the equation $f(t) = c$ has at most $\deg f$ solutions in F.*

Proof. A solution t to $f(t) = c$ is a root of the polynomial $f(T) - c$, and $\deg(f(T) - c) = \deg(f(T))$ since $f(T)$ is not constant. By Theorem 1 the number of roots of $f(T) - c$ in F is at most $\deg(f(T) - c) = \deg(f(T))$. □

Example 3. For a nonconstant polynomial $f(T) \in \mathbf{Z}[T]$ and $c \in \mathbf{Z}$, the equation $f(n) = c$ has finitely many integer solutions n since it has finitely many rational solutions n.

This corollary is not true in general for polynomials whose coefficients are not in a field: the polynomial T^2 has degree 2 and if it is viewed as a polynomial with coefficients in $\mathbf{Z}/(8)$ the equation $t^2 = 1$ has 4 solutions in $\mathbf{Z}/(8)$: 1, 3, 5, and 7. Note $\mathbf{Z}/(8)$ is not a field.

The most important qualitative consequence of Theorem 1 is that a polynomial in $F[T]$ has *finitely many roots* in F.

Corollary 4. *If F is an infinite field and two polynomials $f(T)$ and $g(T)$ in $F[T]$ satisfy $f(t) = g(t)$ for infinitely many t in F then $f(t) = g(t)$ for all $t \in F$.*

As an example, if two polynomials in $\mathbf{R}[T]$ are equal at all numbers in the interval $(0, 1)$ then they are equal at all real numbers.

Proof. Look at the difference polynomial $f(T) - g(T)$. By hypothesis, this polynomial has infinitely many roots in F, so by Theorem 1 it can't be a nonzero polynomial . Thus $f(T) - g(T)$ is the zero polynomial, so $f(T) = g(T)$. Thus $f(t) = g(t)$ for all $t \in F$. □

When p is prime, $\mathbf{F}_p = \mathbf{Z}/(p)$ is a field of size p. This is a finite field.

Corollary 5. *For a prime p, a polynomial $f(T)$ in $\mathbf{F}_p[T]$ of degree less than p is not identically zero on $\mathbf{F}_p$: there's some $t \in \mathbf{F}_p$ such that $f(t) \not\equiv 0 \bmod p$.*

Proof. By Theorem 1, $f(T)$ has at most $\deg f$ roots in $\mathbf{F}_p$. Since $\deg f < p$, the set of roots of $f(T)$ in $\mathbf{F}_p$ is not all of $\mathbf{F}_p$, so there's some $t \in \mathbf{F}_p$ such that $f(t) \not\equiv 0 \bmod p$. $\square$

To appreciate this corollary, we have $t^3(t^2-1) = 0$ for all t in $\mathbf{Z}/(8)$: $t = 0, 2, 4, 6$ satisfy $t^3 \equiv 0 \bmod 8$ and $t = 1, 3, 5, 7$ satisfy $t^2 - 1 \equiv 0 \bmod 8$. Therefore the polynomial $T^3(T^2-1)$ of degree 5 is identically 0 on the 8 elements of $\mathbf{Z}/(8)$. Note $\mathbf{Z}/(8)$ is not a field.

Theorem 6. *Let $f(T)$ be a nonconstant polynomial in $\mathbf{Z}[T]$. For each $k \geq 1$ there is an integer n such that $f(n)$ has at least k different prime factors.*

The meaning of this theorem is that it's impossible for a polynomial with integral coefficients to have its values all be of the form $\pm 2^a 3^b$ or some other product of a *fixed* set of primes.

Proof. The argument below is from Jorge Miranda. It is a proof by induction on k.

First, since the equations $f(n) = 1$, and $f(n) = -1$ each have only finitely many solutions in $\mathbf{Z}$ (see Example 3), some value $f(n)$ is divisible by a prime. This settles the case $k = 1$.

Now suppose $k \geq 2$ and there are primes $p_1, \ldots, p_{k-1}$ and a positive integer m such that $f(m)$ is divisible by $p_1, \ldots, p_{k-1}$. We will find a new prime p_k and a value $f(n)$ divisible by $p_1, \ldots, p_{k-1}, p_k$.

If $f(0) = 0$ then as a polynomial $f(T)$ has no constant term:

$$f(T) = c_d T^d + c_{d-1} T^{d-1} + \cdots + c_1 T$$

with $c_j \in \mathbf{Z}$. Therefore $f(n)$ is divisible by n for all n, so letting p_k be a prime other than $p_1, \ldots, p_{k-1}$, the number $f(p_1 \cdots p_k)$ is divisible by $p_1, \ldots, p_k$.

Now suppose $f(0) \neq 0$. Write $f(T) = c_d T^d + \cdots + c_1 T + c_0 = T g(T) + c_0$, where $c_0 = f(0)$ and $g(T)$ is a nonzero polynomial. Factor $f(0)$ into primes as $\pm p_1^{e_1} \cdots p_{k-1}^{e_{k-1}}$. For each positive integer n,

$$f(n) = n g(n) + f(0) = n g(n) \pm p_1^{e_1} \cdots p_{k-1}^{e_{k-1}}.$$

If n is divisible by $p_1^{e_1+1} \cdots p_{k-1}^{e_{k-1}+1}$ then the power of each p_i in $f(n)$ is e_i (why?). Therefore $f(n) = f(0)N$ where N is not divisible by any of $p_1, \ldots, p_{k-1}$. The equation $f(n) = \pm f(0)$ has only finitely many solutions in $\mathbf{Z}$, while there are infinitely many multiples of $p_1^{e_1+1} \cdots p_{k-1}^{e_{k-1}+1}$, so there is an n that's a multiple of $p_1^{e_1+1} \cdots p_{k-1}^{e_{k-1}+1}$ such that $f(n) \neq \pm f(0)$. Therefore $f(n) = f(0)N$ where $|N| \geq 2$. A prime factor of N is not any of $p_1, \ldots, p_{k-1}$, so $f(n)$ has k prime factors. $\square$