THE DIVISION THEOREM IN Z AND $F[T]$

KEITH CONRAD

1. Introduction

In both $\mathbf{Z}$ and $F[T]$, we can carry out a process of division with remainder.

Theorem 1.1. *For any integers a and b, with b nonzero, there are unique integers q and r such that*

$$a = bq + r, \quad 0 \leq r < |b|.$$

Theorem 1.2. *For any $f(T)$ and $g(T)$ in $F[T]$, with $g(T)$ nonzero, there are unique $q(T)$ and $r(T)$ in $F[T]$ such that*

$$f(T) = g(T)q(T) + r(T), \quad r(T) = 0 \text{ or } \deg r(T) < \deg g(T).$$

In both theorems, there are two things to be proved: a quotient and remainder exist satisfying the conclusions, and there is only one such pair. Often when proving such "existence and uniqueness" theorems, it is simpler to split up the proof into an existence part and a uniqueness part, which is what we will do.

2. Proof of Theorem 1.1

First we show uniqueness, then existence.

Proof. Uniqueness: Assume there are q_1, r_1 and q_2, r_2 in $\mathbf{Z}$ which both satisfy the conclusion. That is,

$$a = bq_1 + r_1, \quad 0 \leq r_1 < |b| \tag{2.1}$$

and

$$a = bq_2 + r_2, \quad 0 \leq r_2 < |b|. \tag{2.2}$$

Comparing the equations in (2.1) and (2.2), we have $bq_1 + r_1 = bq_2 + r_2$. Subtracting,

$$b(q_1 - q_2) = r_2 - r_1. \tag{2.3}$$

This implies the difference $r_2 - r_1$ is a multiple of b.

Because r_1 and r_2 range from 0 to $|b| - 1$, the difference $r_2 - r_1$ is smaller in absolute value than $|b|$. (Why?) Feeding this into (2.3) implies

$$|b(q_1 - q_2)| = |r_2 - r_1| < |b|.$$

The only integer multiple of b which is smaller in absolute value than $|b|$ is 0, so $b(q_1 - q_2) = 0$. Because $b \neq 0$ (aha...), we must have $q_1 - q_2 = 0$, so $q_1 = q_2$. Then, returning to (2.3), $r_2 - r_1 = b \cdot 0 = 0$ and we get $r_1 = r_2$.

Existence: We give two proofs. The first one is very short, while the second looks more fussy and formal. It is the *second* proof whose ideas will generalize to the polynomial setting of Theorem 1.2.

The most interesting case is $b > 0$, so we treat this first. Consider all the integer multiples of b: $\{bq : q \in \mathbf{Z}\}$. Since $b \neq 0$, these multiples are equally spaced all along the real line. The integer a lies in the interval between two consecutive multiples of b:

$$bq \leq a < b(q+1)$$

for some $q \in \mathbf{Z}$. (Why is $b > 0$ necessary here?) Now subtract bq from all terms to get $0 \leq a - bq < b$. Let $r = a - bq$. Then $0 \leq r < b = |b|$.

For the second proof of existence when $b > 0$, we treat the cases $a \geq 0$ and $a < 0$ separately. That is, we *fix* $b > 0$ and will show for each $a \geq 0$ there are appropriate q and r, and then we will show for each $a < 0$ there are appropriate q and r.

When $a \geq 0$, we argue by (strong) induction on a. The case $a = 0$ is trivial: let $q = 0$ and $r = 0$. In fact, if $0 \leq a < b$ we can use $q = 0$ and $r = a$. Suppose now that $a \geq b$ and for *all* $0 \leq a_0 < a$ we have the existence of a q_0 and r_0 for a_0 and b. To get q and r for a and b, consider the number $a_0 := a - b$. Since $a \geq b > 0$, we have $0 \leq a_0 < a$. Therefore there are q_0 and r_0 such that $a_0 = bq_0 + r'$ and $0 \leq r_0 < b$. Writing this as

$$a - b = bq_0 + r_0, \quad 0 \leq r_0 < b,$$

add b to both sides: $a = b(q_0 + 1) + r_0$. Use $q = q_0 + 1$ and $r = r_0$. This completes the second proof of existence for $b > 0$ and $a \geq 0$.

If $a < 0$ and $b > 0$, then consider $-a$ and b. Both are positive, so by the previous case we can write

$$-a = bQ + R, \quad 0 \leq R < b.$$

Negating, we have $a = b(-Q) - R$ with $-b < -R \leq 0$. If $R = 0$ then $a = b(-Q)$ so we can use $q = -Q$ and $r = 0$. If $R > 0$, so $-b < -R < 0$, we want to add b to $-R$ to make it positive (and still small), so write $a = b(-Q - 1) + (b - R)$ with $0 < b - R \leq b$. We can use $q = -Q - 1$ and $r = b - R$

Finally, if $b < 0$ and a is arbitrary, then consider a and $-b$. From what we already showed, we can write $a = -bQ + R$ where $0 \leq R < b$. Writing this as $a = b(-Q) + R$, we can use $q = -Q$ and $r = R$. □

Reread this proof until you understand the basic strategy and really see what's going on. You might try running through the proof with several choices for a and b, say $a = 17$ and $b = 5$, or $a = -17$ and $b = 3$.

3. Proof of Theorem 1.2

As with the proof of Theorem 1.1, we first show uniqueness and then existence.

Proof. Uniqueness: Assume, for polynomials $f(T)$ and $g(T)$, that there are $q_1(T), r_1(T)$ and $q_2(T), r_2(T)$ in $F[T]$ which both satisfy the conclusion. That is,

$$f(T) = g(T)q_1(T) + r_1(T), \quad r_1(T) = 0 \text{ or } \deg r_1(T) < \deg g(T) \tag{3.1}$$

and

$$f(T) = g(T)q_2(T) + r_2(T), \quad r_2(T) = 0 \text{ or } \deg r_2(T) < \deg g(T). \tag{3.2}$$

Comparing the equations in (3.1) and (3.2), we have $g(T)q_1(T)+r_1(T) = g(T)q_2(T)+r_2(T)$. Subtracting,

$$g(T)(q_1(T) - q_2(T)) = r_2(T) - r_1(T). \tag{3.3}$$

This implies the difference $r_2(T) - r_1(T)$ is a polynomial multiple of $g(T)$.[1]

From the degree bounds[2] in (3.1) and (3.2), if $r_1(T) \neq r_2(T)$ then

$$\deg(r_1(T) - r_2(T)) \leq \max(\deg r_1(T), \deg r_2(T)) < \deg g(T).$$

But (3.3) tells us $r_1(T) - r_2(T)$ is a multiple of $g(T)$, so its degree (if $r_1(T) - r_2(T) \neq 0$) is at least $\deg g(T)$. This is a contradiction, so we must have $r_1(T) = r_2(T)$. Then by (3.3), $g(T)(q_1(T) - q_2(T)) = 0$, so (since $g(T) \neq 0$) $q_1(T) - q_2(T) = 0$ and hence $q_1(T) = q_2(T)$.

Existence: Looking back at the first proof of the existence part of Theorem 1.1, we see that argument does not extend so easily to polynmials, since inequalities with integers don't carry over as nicely to polynomials. (Also, what would "equally spaced" polynomials mean?). However, there are inequalities on degrees of polynomials, since degrees are integers. We will work with inequalities on the degrees of polynomials and mimic the idea in the second proof of the existence part of Theorem 1.1.

The case when $g(T)$ is constant (that is, $\deg g(T) = 0$) is easy: if $g(T) = c$ is a nonzero constant then for any $f(T)$ we can use $q(T) = (1/c)f(T)$ and $r(T) = 0$.

Now fix a nonconstant $g(T)$. We will argue by (strong) induction on $\deg f(T)$. That is, for all $f(T)$ of a given degree we will explain in a uniform way how to find polynomials $q(T)$ and $r(T)$ such that $f(T) = g(T)q(T) + r(T)$ with $r(T) = 0$ or $0 \leq \deg r(T) < \deg g(T)$.

First suppose $\deg f(T) < \deg g(T)$. In this case use $q(T) = 0$ and $r(T) = f(T)$.[3]

Now assume that for some integer $n \geq \deg g(T)$ we have constructed a polynomial $q(T)$ and $r(T)$ for each $f(T)$ of degree less than n. We want to construct $q(T)$ and $r(T)$ for all polynomials $f(T)$ of degree n. For a polynomial $f(T)$ of degree n, write its leading term as a_nT^n. Let $g(T)$ have leading term c_dT^d, so $n \geq d$. Then $f(T)$ has the same leading term as $(a_n/c_d)T^{n-d}g$ (check!), which means the difference $f(T) - (a_n/c_d)T^{n-d}g(T)$ is a polynomial with degree less than n,[4] so by the inductive hypothesis there are polynomials $q_0(T)$ and $r_0(T)$ such that

$$f(T) - (a_n/c_d)T^{n-d}g(T) = g(T)q_0(T) + r_0(T), \quad r_0(T) = 0 \text{ or } \deg r_0(T) < \deg g(T).$$

Bring the $g(T)$-term on the left over to the other side by addition:

$$f(T) = g(T)(q_0(T) + (a_n/c_d)T^{n-d}) + r_0(T), \quad r_0(T) = 0 \text{ or } \deg r_0 < \deg g(T).$$

Therefore $f(T) = g(T)q(T)+r(T)$ where $q(T) = q_0(T)+(a_n/c_d)T^{n-d}$ and $r(T) = r_0(T)$. □

In the last part of the proof (the 'reduction' part), we multiplied $g(T)$ by a monomial to make the top term match the top term of $f(T)$ (in both degree and coefficient), so the difference has lower degree. This process can be repeated to drop the degree further, until we finally get a polynomial that has degree less than $\deg g(T)$ or is the polynomial 0. Putting everything back together, we get $q(T)$ and $r(T)$. This is *exactly* the algorithm taught in high school to divide one polynomial by another, but perhaps said in a slightly different way.

Example 3.1. Let $f(T) = 7T^4 - 1$ and $g(T) = T^2 + 5T$. Since $f(T)$ has the same leading term as $7T^2g$, we compute

$$f(T) - 7T^2g(T) = -35T^3 - 1.$$

[1]The argument so far is just like the proof of uniqueness in $\mathbf{Z}$.

[2]Here we need a slightly different argument than in $\mathbf{Z}$, since polynomials don't have absolute values. We use the degree as a measure of size instead.

[3]This corresponds to the case $0 \leq a < b$ in the proof for $\mathbf{Z}$ when $b > 0$.

[4]This is the analogue, in the proof for $\mathbf{Z}$, of considering $a_0 = a - b$ in place of a when $a \geq b > 0$.

Since $-35T^3 - 1$ has the same leading term as $-35Tg(T)$, we compute

$$(-35T^3 - 1) - (-35Tg(T)) = 175T^2 - 1.$$

Since $175T^2 - 1$ has the same leading term as $175g(T)$, we compute

$$(175T^2 - 1) - 175g(T) = -875T - 1,$$

whose degree is less than $\deg g$, so we stop. Feeding each equation into the previous ones gives

$$\begin{aligned} f(T) &= 7T^2g(T) - 35T^3 - 1 \\ &= 7T^2g(T) - 35Tg(T) + (175T^2 - 1) \\ &= 7T^2g(T) - 35Tg(T) + 175g(T) - 875T - 1 \\ &= g(T)(7T^2 - 35T + 175) - 875T - 1. \end{aligned}$$

Thus $q(T) = 7T^2 - 35T + 175$ and $r(T) = -875T - 1$.

Example 3.2. Let $f(T) = 2T^4 + T^2 + 6$ and $g(T) = 3T^2 + 1$. Since $f(T)$ has the same leading term as $\frac{2}{3}T^2g(T)$, we compute

$$f(T) - \frac{2}{3}T^2g(T) = \frac{1}{3}T^2 + 6.$$

The right side has the same leading term as $\frac{1}{9}g(T)$, so we compute

$$\left(\frac{1}{3}T^2 + 6\right) - \frac{1}{9}g(T) = \frac{53}{9},$$

whose degree is less than $\deg g$, so we stop. Feeding the equations into each other gives

$$\begin{aligned} f(T) &= \frac{2}{3}T^2g(T) + \frac{1}{3}T^2 + 6 \\ &= \frac{2}{3}T^2g(T) + \frac{1}{9}g(T) + \frac{53}{9} \\ &= g(T)\left(\frac{2}{3}T^2 + \frac{1}{9}\right) + \frac{53}{9}, \end{aligned}$$

so $q(T) = \frac{2}{3}T^2 + \frac{1}{9}$ and $r(T) = \frac{53}{9}$.

4. Division theorem in $\mathbf{Z}[T]$

Is Theorem 1.2 correct if we work in $\mathbf{Z}[T]$? In other words, if $f(T)$ and $g(T)$ are in $\mathbf{Z}[T]$, does the proof of Theorem 1.2 go through and give us unique $q(T)$ and $r(T)$ in $\mathbf{Z}[T]$ such that $f(T) = g(T)q(T) + r(T)$ where $r(T) = 0$ or $0 \leq \deg r(T) < \deg g(T)$? The uniqueness part goes through without a problem, but the existence part need not work! Look at Example 3.2. The initial data $f(T)$ and $g(T)$ are in $\mathbf{Z}[T]$ while the q and r that come out are not in $\mathbf{Z}[T]$. Why is that? Where does the proof break down?

The proof has a problem in exactly one place: at the end of the existence part of the proof, we want to multiply $g(T)$ by a suitable monomial to get the top term to match that of $f(T)$, and for this we *divide* by c_d, the leading coefficient of $g(T)$. This usually requires rational numbers, unless $c_d = \pm 1$. It is the division by the leading coefficient of $g(T)$ which tells us the division theorem for $\mathbf{Z}[T]$ is not generally valid. More precisely, the denominators that get introduced in $q(T)$ and $r(T)$ will come from the leading coefficient of

$g(T)$. For instance, in the second example after the proof of Theorem 1.2, $g(T)$ has leading coefficient 3 and $q(T)$ and $r(T)$ have coefficients with denominators 3 and 9.

There is a special (and important!) case where division in $\mathbf{Z}[T]$ *is* valid: if the leading coefficient of $g(T)$ is 1. Division by 1 does not introduce denominators. Therefore, when $g(T)$ has leading coefficient 1, the difficulty in the proof of Theorem 1.2 for $\mathbf{Z}[T]$ does not arise. So there is a *restricted* division theorem in $\mathbf{Z}[T]$, as follows.

Theorem 4.1. *For any $f(T)$ and $g(T)$ in $\mathbf{Z}[T]$, with $g(T)$ having leading coefficient 1, there are unique $q(T)$ and $r(T)$ in $\mathbf{Z}[T]$ such that*

$$f(T) = g(T)q(T) + r(T), \qquad r(T) = 0 \text{ or } \deg r(T) < \deg g(T).$$

It is left to the reader, as an exercise, to check that the proof of Theorem 1.2 carries over to the setting of Theorem 4.1.

We already saw an example of Theorem 4.1 in Example 3.1. There $g(T)$ has leading coefficient 1 and the resulting $q(T)$ and $r(T)$ are in $\mathbf{Z}[T]$.