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

KEITH CONRAD

1. Introduction

In the integers we can carry out a process of division with remainder, as follows.

Theorem 1.1. *For any integers a and b, with $b \neq 0$ there are unique integers q and r such that*

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

Division with remainder is also possible for polynomials with coefficients in a field (such as $\mathbf{Q}, \mathbf{R}$, or $\mathbf{C}$), and the statement is very similar to Theorem 1.1.

Theorem 1.2. *Let F be a field. For any $f(T)$ and $g(T)$ in $F[T]$, with $g(T) \neq 0$, there are unique $q(T)$ and $r(T)$ in $F[T]$ such that*

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

In both Theorem 1.1 and 1.2, there are two things to be proved: for each a and nonzero b in $\mathbf{Z}$, or $f(T)$ and nonzero $g(T)$ in $F[T]$, a quotient and remainder exist satisfying the conclusions of the theorem *and* there is only one such quotient and remainder. Often when proving such "existence and uniqueness" theorems, it is convenient to split up the proof into an existence part and a uniqueness part.

Not only do Theorems 1.1 and 1.2 look the same, but they can be proved in similar (although not exactly identical) ways. First we will prove Theorem 1.1, and then we will adapt the ideas there for a proof of Theorem 1.2. We will then describe one application of the division theorem in $\mathbf{Z}$ and $F[T]$ that is widely used: base expansions.

2. Proof of Theorem 1.1

Proof. Uniqueness: For a choice of integers a and b with $b \neq 0$, assume there are q_1, r_1 and q_2, r_2 in $\mathbf{Z}$ that both satisfy the conclusion of the theorem for that a and b. 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 that 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: For a choice of integers a and b with $b \neq 0$, we want to prove there are q and r in $\mathbf{Z}$ such that $a = bq + r$ and $0 \leq r < |b|$. We will give *two* proofs. The first one will be very short, while the second may look more fussy and formal. It is the second proof, not the first, whose ideas can 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: $\{0, \pm b, \pm 2b, \dots\}$. Since $b \neq 0$, these multiples are equally spaced all along the real line. The integer a lies in the interval between consecutive multiples of b:

$$bq \leq a < b(q+1) \tag{2.4}$$

for some $q \in \mathbf{Z}$. (Why is $b > 0$ necessary here?) Now subtract bq from all terms in (2.4) to get $0 \leq a - bq < b$. Set $r = a - bq$, so $0 \leq r < b = |b|$. We are done.

For a second proof of existence of q and r when $b > 0$, we treat the cases $a \geq 0$ and $a < 0$ separately. 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. If $0 \leq a < b$ we can use $q = 0$ and $r = a$. Suppose now that $a \geq b$ and for *all* integers a_0 with $0 \leq a_0 < a$ we have the existence of a q_0 and r_0 for a_0 and b in Theorem 1.1. 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 integers q_0 and r_0 such that $a_0 = bq_0 + r_0$ 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 see what's going on. Run through the proof with several choices for a and b, such as $a = 17$ 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: Picking $f(T)$ and $g(T)$ in $F[T]$ with $g(T)$ being nonzero, suppose there are $q_1(T), r_1(T)$ and $q_2(T), r_2(T)$ in $F[T]$ that both satisfy the conclusion of Theorem 1.2:

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

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] on $r_1(T)$ and $r_2(T)$ in (3.1) and (3.2), if $r_1(T) \neq r_2(T)$ then

$$\deg(r_1(T) - r_2(T)) \leq \max(\deg r_1(T), \deg r_2(T)) < \deg g(T). \tag{3.4}$$

Equation (3.3) tells us $r_1(T) - r_2(T)$ is a multiple of $g(T)$, so $\deg(r_1(T) - r_2(T)) \geq \deg g(T)$ if $r_1(T) - r_2(T) \neq 0$. This contradicts (3.4), so we must have $r_1(T) - r_2(T) = 0$, and hence $r_1(T) = r_2(T)$. Then by (3.3), $g(T)(q_1(T) - q_2(T)) = 0$. That implies the difference $q_1(T) - q_2(T)$ is 0 since $g(T) \neq 0$, so $q_1(T) = q_2(T)$.

Existence: Given $f(T)$ and $g(T)$ in $F[T]$ with nonzero $g(T)$, we want to find $q(T)$ and $r(T)$ in $F[T]$ such that

(1) $f(T) = g(T)q(T) + r(T)$,
(2) $r(T) = 0$ or $\deg r(T) < \deg g(T)$.

We will do this by modifying a proof for the analogous situation in $\mathbf{Z}$ (existence of q and r) that we have already discussed in Theorem 1.1.

The first proof of the existence of q and r in Theorem 1.1 does not generalize to polynomials; what would "equally spaced" polynomials mean? However, the second proof of the existence part of Theorem 1.1 will carry over to polynomials, using induction on the *degrees* of polynomials.

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

Fix now a nonconstant $g(T) \in F[T]$. We want to show for all nonzero $f(T)$ in $F[T]$ that there are polynomials $q(T)$ and $r(T)$ in $F[T]$ such that $f(T) = g(T)q(T) + r(T)$ with $r(T) = 0$ or $\deg r(T) < \deg g(T)$. We will prove this by strong induction on $\deg f(T)$. That is, for every integer $n \geq 0$ we will prove the existence of $q(T)$ and $r(T)$ for all $f(T)$ of degree n by strong induction on n.

If $n < \deg g(T)$, then for any $f(T)$ of degree n we can use $q(T) = 0$ and $r(T) = f(T)$.[3] Now assume that $n \geq \deg g(T)$ and the existence of $q(T)$ and $r(T)$ is true for $g(T)$ and all polynomials $f(T)$ of degree less than n. We will use this to show there are $q(T)$ and $r(T)$ for $g(T)$ and every $f(T)$ of degree n. Write the leading terms of $f(T)$ and $g(T)$ as

$$\begin{aligned} f(T) &= a_n T^n + \text{lower order terms}, \\ g(T) &= c_d T^d + \text{lower order terms}. \end{aligned}$$

[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 $b > 0$, where we can use $q = 0$ and $r = a$.

We have $n \geq d$ (why?). Multiplying $g(T)$ by T^{n-d} raises its degree to n, and then scaling by a suitable constant makes its *leading term* match that of $f(T)$:

$$T^{n-d}g(T) = c_dT^n + \text{lower order terms} \Longrightarrow \frac{a_n}{c_d}T^{n-d}g(T) = a_nT^n + \text{lower order terms}.$$

In this step we used the fact that our coefficients are in a field (!) because that is how we know that we can divide by c_d. Since $f(T)$ and $(a_n/c_d)T^{n-d}g(T)$ have the same leading term a_nT^n, the difference $f(T) - (a_n/c_d)T^{n-d}g(T)$ is a polynomial of degree less than n,[4] so by the inductive hypothesis there are $q_0(T)$ and $r_0(T)$ in $F[T]$ such that

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

Bring $(a_n/c_d)T^{n-d}g(T)$ to the other side in (1):

$$f(T) = (a_n/c_d)T^{n-d}g(T) + g(T)q_0(T) + r_0(T) = g(T)(q_0(T) + (a_n/c_d)T^{n-d}) + r_0(T).$$

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 get a polynomial having degree less than $\deg g(T)$ or being the polynomial 0. Putting everything back together, we get $q(T)$ and $r(T)$. This is *exactly* the algorithm taught in 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 $g(T) = T^2 + 5T$. What are $q(T)$ and $r(T)$ making $f(T) = g(T)q(T) + r(T)$ with $r(T) = 0$ or $\deg r(T) < 2$? Since $f(T)$ has the same leading term as $7T^2g(T)$, we compute

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

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 $2 = \deg g(T)$, 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

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

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

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 $2 = \deg g(T)$, so we stop. Feeding the equations into each other gives

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

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

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

Theorem 1.2 is *not* true if we work in $\mathbf{Z}[T]$ instead of $F[T]$. More precisely, the existence part breaks down: if $f(T)$ and $g(T)$ are in $\mathbf{Z}[T]$, the proof of Theorem 1.2 does not generally lead to $q(T)$ and $r(T)$ in $\mathbf{Z}[T]$ such that $f(T) = g(T)q(T) + r(T)$ where $r(T) = 0$ or $\deg r(T) < \deg g(T)$. (The uniqueness part goes through in $\mathbf{Z}[T]$ without a problem.) Look at Example 3.2. The initial data $f(T)$ and $g(T)$ are in $\mathbf{Z}[T]$ while $q(T)$ and $r(T)$ are not in $\mathbf{Z}[T]$. Why is that? Where does the proof break down?

The proof of the existence of $q(T)$ and $r(T)$ has a problem in $\mathbf{Z}[T]$ in exactly one step: when we want to multiply $g(T)$ by a suitable monomial to get the top term to match that of $f(T)$, we want to multiply by a_n/c_d, and that involves *division* by c_d, the leading coefficient of $g(T)$. When a_n and c_d are in $\mathbf{Z}$, the ratio a_n/c_d is usually not an integer. The denominators that get introduced in $q(T)$ and $r(T)$ will come from the leading coefficient c_d 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 $\mathbf{Z}[T]$.

5. Base expansions

Writing positive integers in base 10 is closely related to division by 10. For example,

$$36137 = 36130+7 = 3613\cdot 10+7 = (361\cdot 10+3)\cdot 10+7 = (((3\cdot 10+6)\cdot 10+1)\cdot 10+3)\cdot 10+7.$$

This calculation can be written in terms of successive division by 10:

$$\begin{aligned} 36137 &= 10 \cdot 3613 + 7 \\ 3613 &= 10 \cdot 361 + 3, \\ 361 &= 10 \cdot 36 + 1, \\ 36 &= 10 \cdot 3 + 6, \\ 3 &= 10 \cdot 0 + 3. \end{aligned}$$

The base 10 digits appear as remainders in reverse order. To write 36137 in base 6,

$$\begin{aligned} 36137 &= 6 \cdot 6022 + 5, \\ 6022 &= 6 \cdot 1003 + 4, \\ 1003 &= 6 \cdot 167 + 1, \\ 167 &= 6 \cdot 27 + 5, \\ 27 &= 6 \cdot 4 + 3, \\ 4 &= 6 \cdot 0 + 4. \end{aligned}$$

Check that $36137 = 4 \cdot 6^5 + 3 \cdot 6^4 + 5 \cdot 6^3 + 1 \cdot 6^2 + 4 \cdot 6 + 5 = 435145_6$.

Theorem 5.1. *Fix an integer $b > 1$. Each $n \in \mathbf{Z}^+$ can be written in exactly one way as*

$$n = c_k b^k + c_{k-1} b^{k-1} + \cdots + c_1 b + c_0 \tag{5.1}$$

where $k \geq 0$ and $0 \leq c_i \leq b-1$, with $c_k \neq 0$.

The expression of positive integers in the form (5.1) is called the *base b representation*.

Proof. We break up the theorem into two parts: existence of a base b representation for all positive integers and then its uniqueness for each positive integer.

Existence: We use (strong) induction on n. If $1 \leq n \leq b-1$ then we can use $d = 0$ and $c_0 = n$. Suppose $n \geq b$ and we can write all positive integers less than n in base b. As in the example of 36137 above, we will construct the base representation of n by starting with what will turn out to be its "units" digit c_0. Using division of n by b,

$$n = bq + r \tag{5.2}$$

where $0 \leq r \leq b-1$ and $q \geq 0$ (if $q \leq -1$ then $bq + r \leq -b + r = r - b < 0$, a contradiciton). The integer q is "obviously" smaller than n. Let's check: since $n \geq b$ we don't have $q = 0$, so $q \geq 1$. Then $n = bq + r \geq bq > q$, so $0 < q < n$.

We can now apply the inductive hypothesis to q: it has a base b representation, say

$$q = a_\ell b^\ell + a_{\ell-1} b^{\ell-1} + \cdots + a_1 b + a_0$$

where $\ell \geq 0$ and $0 \leq a_i \leq b-1$ with $a_\ell \neq 0$. Feeding this into (5.2),

$$n = bq + r = b(a_\ell b^\ell + a_{\ell-1} b^{\ell-1} + \cdots + a_1 b + a_0) + r = a_\ell b^{\ell+1} + a_{\ell-1} b^\ell + \cdots + a_1 b^2 + a_0 b + r,$$

which is a base b representation of n: set $k = \ell + 1$, $c_0 = r$, and $c_i = a_{i-1}$ for $i = 1, \ldots, k$.

Uniqueness: We use (strong) induction on n. Using a base b representation of n from the existence part of this theorem, $n \geq c_k b^k \geq b^k$, so if $0 \leq n < b$ then necessarily $k = 0$ and $c_0 = n$.

Suppose $n \geq b$. When $n = c_k b^k + c_{k-1} b^{k-1} + \cdots + c_1 b + c_0$ is a base b representation then $n = bq + c_0$ where $q = c_k b^{k-1} + \cdots + c_1$. Therefore c_0 is the remainder when n is divided by b, and that remainder is unique (for n and b). Thus the base b units digit of n has just one

choice. The integer q is positive and less than n (as in the proof of the existence part), so by induction its base b representation is unique. Since $c_1, \ldots, c_k$ are base b digits of q, we see that these digits for n (including the number of these terms, which is k) are unique for n as well. □

In $F[T]$, elements automatically appear in "base T", but we can get a base representation using other nonconstant polynomials by adapting the proof from $\mathbf{Z}^+$ to the setting of $F[T]$.

Theorem 5.2. *Fix a nonconstant polynomial $b(T)$ in $F[T]$. For each nonzero $f(T)$ in $F[T]$, there is a unique way to write*

$$f(T) = c_k(T)b(T)^k + c_{k-1}(T)b(T)^{k-1} + \cdots + c_1(T)b(T) + c_0(T)$$

where $k \geq 0$ and $c_i(T)$ is 0 or $0 \leq \deg c_i \leq \deg b(T)$, with $c_k(T) \neq 0$.

Proof. First establish existence, then uniqueness. Existence is proved by strong induction on the degree of nonzero polynomials, as in the existence part of the proof of Theorem 5.1 but using division in $F[T]$ instead of in $\mathbf{Z}$. Uniqueness is proved from uniqueness of the remainder in division by $b(T)$ together with strong induction on the degree of the polynomial, as in the uniqueness part of the proof of Theorem 5.1. Details are left to the reader. □

Example 5.3. In $\mathbf{R}[T]$, let's write $T^5 + T + 1$ in base $T^2 + 1$. When the base has degree 2, each "digit" is a polynomial of degree less than 2 (or the digit is 0). Dividing by $T^2 + 1$ repeatedly in the same way that we wrote 36137 in base 6 using repeated division by 6,

$$\begin{aligned} T^5 + T + 1 &= (T^2+1)(T^3 - T) + 2T + 1, \\ T^3 - T &= (T^2+1)(T) - 2T, \\ T &= (T^2+1)(0) + T, \end{aligned}$$

so the digits of $T^5 + T + 1$ in base $T^2 + 1$ are T, $-2T$, and $2T + 1$:

$$\begin{aligned} T^5 + T + 1 &= (T^2+1)(T^3 - T) + (2T + 1) \\ &= (T^2+1)((T^2+1)T - 2T) + (2T + 1) \\ &= \underline{T}(T^2+1)^2 + (\underline{-2T})(T^2+1) + \underline{2T+1}. \end{aligned}$$