Irish Math. Soc. Bulletin
Number 91, Summer 2023, 73–76
ISSN 0791-5578

There are infinitely many primes: two ring-theoretic variations on Euclid

ALAN ROCHE

Abstract. Using elementary ring theory, we present two proofs in the mode of Euclid that there are infinitely many primes.

1. Introduction

Euclid's proof of the infinitude of primes is a paragon of incisive mathematical reasoning. It's the first entry—deservedly—in Aigner and Ziegler's compilation, their terrestrial approximation to the celestial BOOK [1, p. 3]. The result (infinitude of primes) has been re-proved over and over. Aigner and Ziegler, for example, discuss six proofs in their first chapter and infinitely many more (in a sense) in an appendix.

We use elementary ring theory to show, yet again, that there are infinitely many primes. The argument's strategy is simple: if $p_1, \ldots, p_n$ is the complete list of primes, then the ring of rational numbers $\mathbb{Q}$ is obtained from the ring of integers $\mathbb{Z}$ by adjoining the single element $1/p_1 \cdots p_n$. The task then is to show that this is an untenable structure for $\mathbb{Q}$ which we do in two overlapping ways. In each case, the proof makes use of the key Euclidean manoeuvre: given the list of primes $p_1, \ldots, p_n$, consider $p_1 \cdots p_n + 1$.

We conclude with some comments on Euclid's classic argument.

2. First Proof

Given nonzero integers $a_1, \ldots, a_n$, we write $\mathbb{Z}[1/a_1, \ldots, 1/a_n]$ for the smallest subring of $\mathbb{Q}$ containing $\mathbb{Z}$ and each $1/a_i$. Equivalently, it's the smallest subring of $\mathbb{Q}$ with identity in which $a_1, \ldots, a_n$ are invertible. As the notation suggests, it consists of all $f(1/a_1, \ldots, 1/a_n)$ for $f(\mathrm{X}_1, \ldots, \mathrm{X}_n) \in \mathbb{Z}[\mathrm{X}_1, \ldots, \mathrm{X}_n]$.

Note that

$$\mathbb{Z}[1/a_1, \ldots, 1/a_n] = \mathbb{Z}[1/a_1 \cdots a_n]. \tag{1}$$

Indeed, $a_1 \cdots a_n$ is invertible (in a subring of $\mathbb{Q}$ with identity) if and only if each a_i is invertible (in that subring), and so the two rings coincide.

Suppose now that there are only finitely many primes, say $p_1, \ldots, p_n$. Since each positive integer m is a product of primes, our supposition implies that $1/m$ is in $\mathbb{Z}[1/p_1, \ldots, 1/p_n]$, and therefore

$$\mathbb{Q} = \mathbb{Z}[1/p_1, \ldots, 1/p_n].$$

Equivalently, by (1), $\mathbb{Q} = \mathbb{Z}[1/p_1 \cdots p_n]$. To simplify the notation, we set $a = p_1 \cdots p_n$, so that $\mathbb{Q} = \mathbb{Z}[1/a]$.

2020 *Mathematics Subject Classification.* 11A41, 13A05.
Key words and phrases. primes, polynomials, maximal ideal.
Received on 4-4–2023; revised 15-5-2023.
DOI:10.33232/BIMS.0091.73.76.

In particular, $1/(a+1) \in \mathbb{Z}[1/a]$. This means there exist integers $c_0, c_1, \ldots, c_m$ such that

$$\frac{1}{a+1} = c_0 + c_1\frac{1}{a} + \cdots + c_m\frac{1}{a^m}.$$

Multiplying through by a^m, we have

$$\frac{a^m}{a+1} = c_0a^m + c_1a^{m-1} + \cdots + c_m \in \mathbb{Z}.$$

That is, $a+1$ divides a^m. Now $1 = [(a+1) - a]^m$. Expanding the right side, we see that

$$1 = A(a+1) + (-1)^m a^m,$$

for some integer A. Since $a+1$ divides a^m, it follows that $a+1$ divides 1 which is absurd. We've proved that there are infinitely many primes. □

3. Second Proof

Assume once more that there are only finitely many primes $p_1, \ldots, p_n$. As above, it follows that $\mathbb{Q} = \mathbb{Z}[1/a]$ for $a = p_1 \cdots p_n$. In other words, the homomorphism of rings

$$f(\mathrm{X}) \mapsto f(1/a) : \mathbb{Z}[\mathrm{X}] \to \mathbb{Q} \tag{2}$$

is surjective. We write I_a for its kernel, so that (2) induces an isomorphism of rings

$$\overline{f(\mathrm{X})} \longmapsto f(1/a) : \mathbb{Z}[\mathrm{X}]/I_a \xrightarrow{\simeq} \mathbb{Q}. \tag{3}$$

In particular, $\mathbb{Z}[\mathrm{X}]/I_a$ is a field, or equivalently I_a is a maximal ideal in $\mathbb{Z}[\mathrm{X}]$.

To finish the argument, we could appeal to a property of maximal ideals in $\mathbb{Z}[\mathrm{X}]$—that each such ideal contains some nonzero constant polynomial. Indeed, as I_a contains no nonzero constants, we see that I_a cannot be maximal, a contradiction.

This approach, however, is unsatisfying: the property that maximal ideals in $\mathbb{Z}[\mathrm{X}]$ contain nonzero constants lies deeper than the existence of infinitely many primes. Instead, we'll use only our bare hands to prove the following: if $\mathbb{Z}[\mathrm{X}]/I_a$ is a field then $a+1$ must divide 1 (as in the first proof). Our path to this absurdity rests on identifying the structure of the ideal I_a.

Lemma. *We have $I_a = (a\mathrm{X} - 1)$, the principal ideal generated by $a\mathrm{X} - 1$.*

The ideal of elements of $\mathbb{Q}[\mathrm{X}]$ that vanish at $1/a$ is generated by $\mathrm{X} - 1/a$ and so also by $a\mathrm{X} - 1$. The proof that I_a is generated by $a\mathrm{X} - 1$ is then a short exercise using Gauss's Lemma—a product of primitive polynomials is primitive. (Recall an element of $\mathbb{Z}[\mathrm{X}]$ is primitive if the greatest common divisor of its coefficients is 1.) We prefer, however, a still more elementary, albeit ad hoc approach. We want to avoid all tools beyond the most basic properties of polynomials, even one as fundamental as Gauss's Lemma.

Proof. Let $f(\mathrm{X}) = c_0 + c_1\mathrm{X} + \cdots + c_m\mathrm{X}^m \in \mathbb{Z}[\mathrm{X}]$ with $c_m \neq 0$, so $f(\mathrm{X})$ has degree m. We have

$$c_0 + c_1\frac{1}{a} + \cdots + c_m\frac{1}{a^m} = \frac{c_0a^m + c_1a^{m-1} + \cdots + c_m}{a^m}.$$

Thus $f(1/a) = 0$ if and only if $\widetilde{f}(a) = 0$ where

$$\begin{aligned} \widetilde{f}(\mathrm{X}) &= \mathrm{X}^m f(1/\mathrm{X}) \\ &= c_0\mathrm{X}^m + c_1\mathrm{X}^{m-1} + \cdots + c_m. \end{aligned} \tag{4}$$

We call $\widetilde{f}(\mathrm{X})$ the *reverse* of $f(\mathrm{X})$ and going from $f(\mathrm{X})$ to $\widetilde{f}(\mathrm{X})$ *reversing.* Visibly, the reverse of the reverse of $f(\mathrm{X})$ is $f(\mathrm{X})$: reversing is an involution on the set of nonzero elements of $\mathbb{Z}[\mathrm{X}]$. Moreover, it follows readily from (4) that reversing is multiplicative: that is, $\widetilde{f_1 f_2}(\mathrm{X}) = \widetilde{f}_1(\mathrm{X})\widetilde{f}_2(\mathrm{X})$ for nonzero $f_i(\mathrm{X}) \in \mathbb{Z}[\mathrm{X}]$ $(i = 1, 2)$.

Remember the division algorithm for polynomials applies to monic elements of $\mathbb{Z}[\mathrm{X}]$. Hence, for $g(\mathrm{X}) \in \mathbb{Z}[\mathrm{X}]$, we have $g(a) = 0$ if and only if $\mathrm{X} - a$ divides $g(\mathrm{X})$ in $\mathbb{Z}[\mathrm{X}]$. In particular,

$$\widetilde{f}(a) = 0 \iff \widetilde{f}(\mathrm{X}) = (\mathrm{X} - a)\, h(\mathrm{X}),$$

for some $h(\mathrm{X})$. Reversing the polynomial equation and noting that the reverse of $\mathrm{X} - a$ is $-(a\mathrm{X} - 1)$, we see that

$$\widetilde{f}(a) = 0 \iff f(\mathrm{X}) = (a\mathrm{X} - 1)\left(-\widetilde{h}(\mathrm{X})\right).$$

Thus $f(1/a) = 0$ if and only if $a\mathrm{X} - 1$ divides $f(\mathrm{X})$. We've proved the lemma. □

Now, since $a + 1 \notin I_a$, the coset $(a + 1) + I_a$ is invertible in the field $\mathbb{Z}[\mathrm{X}]/I_a$. Hence there is an $h(\mathrm{X}) \in \mathbb{Z}[\mathrm{X}]$ such that $(a + 1)\, h(\mathrm{X}) + I_a = 1 + I_a$. Using the lemma, it follows that

$$(a + 1)\, h(\mathrm{X}) = 1 + (a\mathrm{X} - 1)\, k(\mathrm{X}), \tag{5}$$

for some $k(\mathrm{X})$. Substituting $\mathrm{X} = a$, we obtain

$$(a + 1)\, h(a) = 1 + \left(a^2 - 1\right) k(a),$$

and so

$$(a + 1)\, [h(a) - (a - 1)\, k(a)] = 1.$$

Again, we've reached the absurdity that $a + 1$ divides 1. We've proved once more that there are infinitely many primes. □

4. Comments on Euclid's Proof

First, let's recast Euclid's argument in the language of ring theory.

Proof. Let a be a nonunit in $\mathbb{Z}$, that is, $a \neq \pm 1$. Then a has a prime divisor p, or equivalently $a \in (p)$ for some prime p. We assume that there are only finitely many primes, say $p_1, \ldots, p_n$. It follows that each nonunit in $\mathbb{Z}$ is contained in some (p_i), and therefore

$$\mathbb{Z} \setminus \{\pm 1\} \;=\; \bigcup_{i=1}^{n} (p_i). \tag{6}$$

Now $p_1 \cdots p_n + 1$ is not divisible by p_i, for $i = 1, \ldots, n$. That is,

$$p_1 \cdots p_n + 1 \;\notin\; \bigcup_{i=1}^{n} (p_i).$$

Using (6), we have $p_1 \cdots p_n + 1 = \pm 1$. Nonsense! We conclude that there are infinitely many primes. □

Remark 1. We've presented our variants of Euclid's argument in terms of contradiction. In this form, they give the *existence* of infinitely many primes. As many have noted, however, Euclid's reasoning is *constructive* (see, for example, [2, p. 31]): given a finite list of primes $p_1, \ldots, p_n$, Euclid gives a way (an inefficient way) of adjoining a new prime to the list—namely, any prime factor of $p_1 \cdots p_n + 1$.

Having dressed Euclid's proof in ring-theoretic garb, we can use some set theory to obtain a small generalization. First, some notation. For R a ring with identity, we write $R^\times$ for the group of units of R.

Proposition. *Let R be a PID that is not a field and suppose the cardinality of $R^\times$ is strictly smaller than that of R. Then R contains infinitely many irreducible elements (up to multiplication by units).*

The result applies, in particular, if $R^\times$ is finite.

Proof. We assume that R has only finitely many irreducible elements $\varpi_1, \ldots, \varpi_n$ (up to multiplication by units) and will show that $R^\times$ and R have the same cardinality.

By hypothesis, each nonunit in R is divisible by some ϖ_i. Therefore

$$R \setminus R^\times = \bigcup_{i=1}^{n} (\varpi_i).$$

Now, for $r \in R$, the element $1 + r\varpi_1 \cdots \varpi_n$ is not contained in any (ϖ_i), and so belongs to $R^\times$. Hence we have a map

$$r \mapsto 1 + r\varpi_1 \cdots \varpi_n : R \to R^\times$$

which is injective (as R is a domain). By the Schröder-Bernstein Theorem, $R^\times$ and R have the same cardinality. □

Remark 2. The proposition is not sharp—it was too easy to prove to expect it to be sharp! That is, there are PIDs R with infinitely many irreducible elements (up to multiplication by units) for which $R^\times$ has the same cardinality as R. Example: $R = \mathbb{Z}[\sqrt{2}]$. Indeed, as $(\sqrt{2}+1)(\sqrt{2}-1) = 1$, we see that $R^\times$ contains the infinite cyclic group generated by $\sqrt{2}+1$, and so is countably infinite.

Remark 3. Which PIDs R contain infinitely many irreducible elements (up to multiplication by units)? The note [3] gives a characterization in terms of the polynomial ring $R[\mathrm{X}]$: a PID R has the given property if and only if each maximal chain of prime ideals in $R[\mathrm{X}]$ has length two, that is, has the form $\{0\} \subsetneqq \mathfrak{p}_1 \subsetneqq \mathfrak{p}_2$, for prime ideals $\mathfrak{p}_i$ in $R[\mathrm{X}]$ $(i = 1, 2)$.

References

[1] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, 6th ed. Springer, 2018.
[2] J. Stillwell, *The Story of Proof—Logic and the History of Mathematics.* Princeton University Press, 2022.
[3] F. Zanello, *When are there infinitely many irreducible elements in a principal ideal domain?* Amer. Math. Monthly. 111(2) (2004), 150–152.

Alan Roche holds degrees in mathematics from University College Dublin and the University of Chicago. He has taught and thought at the University of Oklahoma since 2001.

Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA.
E-mail address: aroche@ou.edu