Preprints.org

Article

Not peer-reviewed version

A Proof of the Infinitude of Prime Numbers

Md. Shouvik Iqbal *

Posted Date: 25 February 2025

doi: 10.20944/preprints202406.1296.v2

Keywords: Prime Numbers; Infinitude of Primes; Euclid's theorem; Proof of Euclid's theorem; Proof of Infinitude of Primes; Set Theory; Proof by Contradiction

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2025 doi:10.20944/preprints202406.1296.v2

Article

A Proof of the Infinitude of Prime Numbers

Md. Shouvik Iqbal

shouvikiqbal@gmail.com

Abstract: This paper presents a new proof of the infinitude of prime numbers, also known commonly as "Euclid's theorem." Grounded in fundamental Set theory, the proof uses the method of contradiction to demonstrate the absurdity of assuming the set of all prime numbers to be finite.

Keywords: prime numbers; infinitude of primes; euclid's theorem; proof of euclid's theorem; proof of infinitude of primes; set theory; proof by contradiction

---

1. Introduction

Among the natural numbers, there exists a prime category of numbers that are so prime that they may best be described as "prime numbers." The defining property of such numbers is that they cannot be broken down into smaller whole parts other than 1 and themselves. It was Euclid[1], who appears to have provided the earliest *documented* definition of prime numbers in his work as a collection of books called the "Elements," dating back to circa 300 BCE. In Book IX of *Elements*, Proposition 20, Euclid proved that there are an infinite number of such prime numbers[2] [2**?** ,3]. It was quite the proof that stands for such a high consequence that it is still valued in modernity [4]. From that point onward, the nonexistence of the largest prime number became evident, as evidenced by Euclid's proof. Post-Euclid, many mathematicians have supported this proposition through their own respective proofs. Several such proofs can be found in [5–8]. In this paper, we will also give another proof of this proposition. For that, the motivation is visible when the following is assumed,

$$\mathbb{P} = \{\text{the set of } \textit{all} \text{ prime numbers}\}$$

$$\mathbb{C} = \{\text{the set of } \textit{all} \text{ composite numbers}\}$$

This implies, $\mathbb{N}^* = \mathbb{P} \cup \mathbb{C}$, where $\mathbb{N}^* = \{n \in \mathbb{N} \mid n \geq 2\}$. However, if $|\mathbb{P}|$ is assumed finite, then an interesting point may have been ignored easily that $\mathbb{N}^* = \mathbb{P} \cup \mathbb{C} \cup \mathbb{S}$, where $\mathbb{S}$ is the set of all natural numbers that are neither prime nor composite, which is contradictorily impossible as every element in that set would be greater than 1. Thus, by contradiction, it is possible to prove that the assumption of a finite set containing all prime numbers is inherently false.

If the core idea of the motivation were to be explicitly stated without any formalism, it would assume that there exists a finite number of prime numbers—say only $2, 3, 5$ and 7, then consider the set of all natural numbers greater than or equal to 2 as follows,

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, \cdots$$

Given the assumption that $2, 3, 5,$ and 7 are *only* prime numbers, they are thereupon removed from the set as follows,

$$4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, \cdots$$

---

1 an ancient Greek mathematician, who lived in Alexandria, Egypt, circa 300 BCE.

2 it was done by showing that prime numbers are more than any assigned multitude of prime numbers. Note that, the proof was specifically for only three prime numbers. Euclid did not consider any arbitrary finite set of prime numbers, as is commonly done nowadays [1].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2025 doi:10.20944/preprints202406.1296.v2

Now, given that prime numbers are those numbers greater than 1 that are divisible only by 1 and themselves, it follows that every composite number is a multiple of a prime number. Therefore, every multiple of $2, 3, 5$, and 7 are removed from the set as follows,

$$11, 13, 17, 19, 23, 29, 31, 37, \cdots, \underbrace{11 \cdot 11}_{121}, \cdots, \underbrace{11 \cdot 13}_{143}, \cdots, \underbrace{11 \cdot 17}_{187}, \cdots, \underbrace{11 \cdot 19}_{209}, \cdots$$

Now, this is a set of numbers that are neither prime (as the only prime numbers are assumed to be 2, 3, 5, and 7) nor composite (as they are not multiples of any assumed prime number), and all of them are greater than 1. This is a contradiction!

Having outlined the basic idea, we now give the complete proof of the infinitude of prime numbers as follows.

2. The Theorem and Its Proof

Theorem 1. *There are infinitely many prime numbers.*

Proof. Assume—for the sake of contradiction—that the elements in the set of all prime numbers are finite, and thus define the following sets,

$$\mathbb{N}^* = \{N \in \mathbb{N} \mid N \geq 2\}$$

$$\mathbb{P} = \{p_1, p_2, p_3, \cdots, p_k\}$$

$$\mathbb{N}^* \setminus \mathbb{P} = \{N \in \mathbb{N}^* \mid N \notin \mathbb{P}\}$$

$$\mathbb{C} = \{N \in (\mathbb{N}^* \setminus \mathbb{P}) \mid N = np, \text{ where, } p \in \mathbb{P}, n \in \mathbb{N}^*\}$$

Here, $\mathbb{N}^* \setminus \mathbb{P} \subset \mathbb{N}^*$ and $\mathbb{C} \subseteq \mathbb{N}^* \setminus \mathbb{P}$. To show that $\mathbb{C} \subsetneq \mathbb{N}^* \setminus \mathbb{P}$, notice that $\mathbb{C}$ is the set of all natural numbers of the form np, where $n \in \mathbb{N}^*$ and $p \in \{p_1, p_2, p_3, \cdots, p_k\}$; whereas $\mathbb{N}^* \setminus \mathbb{P}$ includes all natural numbers of the form nk, where $n, k \in \mathbb{N}^*$. It will now be shown in three parts that for any sufficiently large range $\varepsilon \in \mathbb{N}^*$, the total number of natural numbers of the form $nk \leq \varepsilon$ is *greater than* the total number of natural numbers of the form $np \leq \varepsilon$. By doing so, it would be concludable that $|\mathbb{N}^* \setminus \mathbb{P}| \neq |\mathbb{C}|$, for any sufficiently large range $\varepsilon \in \mathbb{N}^*$ and that the set $\mathbb{N}^* \setminus \mathbb{P}$ contains more elements than $\mathbb{C}$, up to that range $\varepsilon \in \mathbb{N}^*$. Which would then be used to form a contradiction.

Part 1: In this part, we will calculate the total number of natural numbers in $\mathbb{C}$ up to a given range ε. For that, notice that $\mathbb{C}$ is the set of all natural numbers of the form np, where $n \in \mathbb{N}^*$ and $p \in \{p_1, p_2, p_3, \cdots, p_k\}$. Therefore, for the ith prime $p_i \in \{p_1, p_2, \cdots, p_i, \cdots, p_k\}$, the natural numbers of the form np_i, where $n \in \mathbb{N}^*$, (that is, multiples of p_i) are as follows,

$$2p_i, 3p_i, 4p_i, 5p_i \cdots$$

Now, the number of natural numbers of the form np_i, where $n \in \mathbb{N}^*$, less than or equal to a finite given range $\varepsilon \in \mathbb{N}^*$ yields,

$$2p_i, 3p_i, 4p_i, \cdots, c_i p_i \leq \varepsilon$$

here, c_i is the largest integer such that $c_i p_i \leq \varepsilon$. This gives: $c_i \leq \dfrac{\varepsilon}{p_i}$. Since $n \geq 2$ in all natural numbers of the form np, therefore $(c_i - 1)$ represents a count for the total number of natural numbers of the form np_i less than or equal to the given range ε. Now, to count the total number of natural numbers of the *general form* np less than or equal to the given range ε, the following is done.

$$\text{for } p_1: \quad 2p_1, 3p_1, 4p_1, \cdots, c_1 p_1 \leq \varepsilon \implies \qquad c_1 \leq \frac{\varepsilon}{p_1} \quad \therefore (c_1 - 1) < c_1 \leq \frac{\varepsilon}{p_1}$$

$$
\begin{aligned}
&\text{for } p_2: \quad 2p_2, 3p_2, 4p_2, \cdots, c_2 p_2 \le \varepsilon \implies && c_2 \le \frac{\varepsilon}{p_2} \quad \therefore (c_2 - 1) < c_2 \le \frac{\varepsilon}{p_2} \\
&\text{for } p_3: \quad 2p_3, 3p_3, 4p_3, \cdots, c_3 p_3 \le \varepsilon \implies && c_3 \le \frac{\varepsilon}{p_3} \quad \therefore (c_3 - 1) < c_3 \le \frac{\varepsilon}{p_3} \\
&\quad \vdots && \vdots \\
&\text{for } p_m: \quad 2p_m, 3p_m, 4p_m, \cdots, c_m p_m \le \varepsilon \implies && c_m \le \frac{\varepsilon}{p_m} \quad \therefore (c_m - 1) < c_m \le \frac{\varepsilon}{p_m}
\end{aligned}
$$

where, $p_m \in \{p_1, p_2, \cdots, p_m, \cdots, p_k\}$ is the mth prime number such that $p_m < \varepsilon$.

This implies that the total number of natural numbers of the form np_1, np_2, np_3, $\cdots$, np_m less than or equal to the given range ε are $(c_1 - 1), (c_2 - 1), (c_3 - 1), \cdots, (c_m - 1)$, respectively. Summing these quantities together results in the total number of natural numbers of the *general form* np, less than or equal to ε. Let the total number be α.

Now, note that the abovementioned largest integer c_i is always greater than 1, for all $i \in \{1, 2, \cdots, m, \cdots, k\}$, due to the fact that each n in the natural number of the form np is an element of $\mathbb{N}^*$ (i.e., greater than 1). As a result, $\alpha = (c_1 - 1) + (c_2 - 1) + (c_3 - 1) + \cdots + (c_m - 1) > 1$, as shown below.

$$
\begin{aligned}
1 &< (c_1 - 1) + (c_2 - 1) + (c_3 - 1) + \cdots + (c_m - 1) = \alpha \\
&< c_1 + c_2 + c_3 + \cdots + c_m \\
&\le \frac{\varepsilon}{p_1} + \frac{\varepsilon}{p_2} + \frac{\varepsilon}{p_3} + \cdots + \frac{\varepsilon}{p_m} \\
&= \varepsilon \left(\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_m} \right) \\
\therefore 1 &< \alpha < \varepsilon \left(\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_m} \right) \qquad (1)
\end{aligned}
$$

where, α represents the total number of natural numbers of the form np, where $n \in \mathbb{N}^*$ and $p \in \mathbb{P}$, less than or equal to $\varepsilon \in \mathbb{N}^*$.

Part 2: Similar to the part above, in this part, we will calculate the total number of natural numbers in $\mathbb{N}^* \setminus \mathbb{P}$ up to a given range ε. For that, notice that the set $\mathbb{N}^* \setminus \mathbb{P}$ contains all natural numbers of the form nk, where $n, k \in \mathbb{N}^*$. Therefore, for the ith natural number $k_i \in \{k_1, k_2, \cdots, k_i, \cdots\} = \mathbb{N}^*$, the natural numbers of the form nk_i, where $n \in \mathbb{N}^*$, (that is, multiples of k_i) are as follows,

$$2k_i, 3k_i, 4k_i, 5k_i \cdots$$

Now, the total number of natural numbers of the form nk_i less than or equal to a given range $\varepsilon \in \mathbb{N}^*$ yields,

$$2k_i, 3k_i, 4k_i, \cdots, d_i k_i \le \varepsilon$$

here, d_i is the largest integer such that $d_i k_i \le \varepsilon$. This gives: $d_i \le \frac{\varepsilon}{k_i}$. Since $n \ge 2$ in all natural numbers of the form nk, therefore $(d_i - 1)$ represents a count for natural numbers of the form nk_i less than or equal to the given range ε. Now, to count the number of natural numbers of the *general form* nk less than or equal to the given range ε, the following is done.

$$
\begin{aligned}
&\text{for } k_1: \quad 2k_1, 3k_1, 4k_1, \cdots, d_1 k_1 \le \varepsilon \implies && d_1 \le \frac{\varepsilon}{k_1} \quad \therefore (d_1 - 1) < d_1 \le \frac{\varepsilon}{k_1} \\
&\text{for } k_2: \quad 2k_2, 3k_2, 4k_2, \cdots, d_2 k_2 \le \varepsilon \implies && d_2 \le \frac{\varepsilon}{k_2} \quad \therefore (d_2 - 1) < d_2 \le \frac{\varepsilon}{k_2}
\end{aligned}
$$

 doi:10.20944/preprints202406.1296.v2

$$\begin{aligned}
&\text{for } k_3: \quad 2k_3, 3k_3, 4k_3, \cdots, d_3k_3 \leq \varepsilon \implies && d_3 \leq \frac{\varepsilon}{k_3} \quad \therefore (d_3-1) < d_3 \leq \frac{\varepsilon}{k_3} \\
&\quad \vdots && \vdots \\
&\text{for } k_j: \quad 2k_j, 3k_j, 4k_j, \cdots, d_jk_j \leq \varepsilon \implies && d_j \leq \frac{\varepsilon}{k_j} \quad \therefore (d_j-1) < d_j \leq \frac{\varepsilon}{k_j}
\end{aligned}$$

where, $k_j \in \{k_1, k_2, \cdots, k_j, \cdots\}$ is the jth natural number in the set $\mathbb{N}^*$ such that $k_j < \varepsilon$.

This implies that the total number of natural numbers of the form $nk_1, nk_2, nk_3, \cdots, nk_j$ less than or equal to the given range ε are $(d_1-1), (d_2-1), (d_3-1), \cdots, (d_j-1)$, respectively. Summing these quantities together results in the total number of natural numbers of the *general form* nk, less than or equal to ε. Let the total number be β.

Now, note that the abovementioned largest number d_i is always greater than 1, for all $i \in \{1, 2, 3, \cdots, j, \cdots\}$, due to the fact that each n in natural numbers of the form nk is an element of $\mathbb{N}^*$ (i.e., greater than 1). As a result, $\beta = (d_1-1) + (d_2-1) + (d_3-1) + \cdots + \left(d_j-1\right) > 1$, as shown below.

$$\begin{aligned}
1 &< (d_1-1) + (d_2-1) + (d_3-1) + \cdots + \left(d_j-1\right) = \beta \\
&< d_1 + d_2 + d_3 + \cdots + d_j \\
&\leq \frac{\varepsilon}{k_1} + \frac{\varepsilon}{k_2} + \frac{\varepsilon}{k_3} + \cdots + \frac{\varepsilon}{k_j} \\
&= \varepsilon\left(\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}\right) \\
\therefore 1 < \beta &< \varepsilon\left(\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}\right) \qquad (2)
\end{aligned}$$

where, β represents the total number of natural numbers of the form nk, where $n, k \in \mathbb{N}^*$, less than or equal to $\varepsilon \in \mathbb{N}^*$

Part 3: In this part, we compare equation (2) with (1) to show that the total number of natural numbers in the set $\mathbb{N}^* \setminus \mathbb{P}$ (that is, β) up to the given range ε is *greater than* that of in the set $\mathbb{C}$ (that is, α) up to that same range ε, provided that the given range ε is sufficiently large. That is,

$$\begin{aligned}
\frac{1}{1} < \frac{\beta}{\alpha} &< \frac{\varepsilon\left(\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}\right)}{\varepsilon\left(\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_m}\right)} \\
1 < \frac{\beta}{\alpha} &< \frac{\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}}{\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_m}} \qquad (3)
\end{aligned}$$

Now, observe that since the set of all prime numbers $\mathbb{P}$ is assumed to be finite, therefore, the sum $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_m}$ in the denominator of equation (3) can never exceed the maximum sum $\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_m} + \cdots + \frac{1}{p_k}$ to diverge[3] to infinity for however large range ε is considered, due to the assumption that $p_k = \max\{\mathbb{P}\}$. On the other hand, since $\mathbb{N}^*$ is considered infinite, therefore, the

[3] since the sum of the reciprocal of primes diverges

 doi:10.20944/preprints202406.1296.v2

sum $\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}$ in the numerator of equation (3) can diverge[4] to infinity if we allow the range ε to do so. Therefore, for any sufficiently large *finite* range $\varepsilon' > \varepsilon$ the following condition is met,

$$\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_m} + \cdots + \frac{1}{p_k} < \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}$$

The condition above holds for any sufficiently large ε' due to the fact that the number of terms in the summations in both the numerator and denominator of equation (3) was derived based on the given range.

This implies, for any sufficiently large finite range ε', that,

$$\frac{\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} + \cdots + \frac{1}{k_j}}{\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots + \frac{1}{p_k}} = L > 1$$

As a result, from equation (3), we have the following.

$$\begin{aligned} 1 < \frac{\beta}{\alpha} &< L \\ \alpha < \beta &< \alpha L \\ \therefore \alpha &< \beta \end{aligned}$$

Conclusion: The analysis above suggests that for any sufficiently large range $\varepsilon' \in \mathbb{N}^*$, the total number of natural numbers of the form $np \leq \varepsilon'$ (i.e., α) is *less than* that of the form $nk \leq \varepsilon'$ (i.e., β). In other words, the total number of natural numbers in $\mathbb{N}^* \setminus \mathbb{P}$ (i.e., β) is *greater than* the total number of natural numbers in $\mathbb{C}$ (i.e., α), up to any sufficiently large range $\varepsilon' \in \mathbb{N}^*$, even though both sets have an infinite number of natural numbers when no range ε is considered.

Contradiction: Based on the conclusion above, for any sufficiently large range ε', since $\alpha < \beta$ for ε', therefore $\exists \delta \in \mathbb{N}^* \setminus \mathbb{P}$ such that $\delta \notin \mathbb{C}$. Since $\delta \notin \mathbb{C}$, therefore, $\delta \neq np$, $\forall p \in \mathbb{P}$, $\forall n \in \mathbb{N}^*$, where $p, n < \delta$. However, if $\delta \neq np$, then, $\forall p \in \mathbb{P}$, $\frac{\delta}{p} \neq n$, where $n \in \mathbb{N}^*$. This implies that $\forall p \in \mathbb{P}$, $p \nmid \delta$, and therefore—by the definition of a prime number—$\delta \in \mathbb{P}$. However, this contradicts the premise that $\delta \in \mathbb{N}^* \setminus \mathbb{P}$.

Should δ be in $\mathbb{C}$ for even large range $\varepsilon'_1 > \varepsilon'$, then, since $\alpha < \beta$ also for ε'_1 (because $\alpha < \beta$ for ε' and $\varepsilon'_1 > \varepsilon'$), therefore $\exists \delta_1 \in \mathbb{N}^* \setminus \mathbb{P}$ such that $\delta_1 \notin \mathbb{C}$ and the rest follows a similar contradiction.

As a result, the argument can be repeated endlessly for every new chosen sufficiently large range $\varepsilon' \in \mathbb{N}^*$, proving the infinitude of the prime numbers each time by contradiction. This completes the proof. $\square$

References

1. Narkiewicz, W. *The Development of Prime Number Theory: From Euclid to Hardy and Littlewood*; Springer Berlin: Heidelberg, 2000.
2. Fitzpatrick, R.; Heiberg, J. Euclid's Elements, 2007.
3. Euclid, Johan Ludvig Heiberg, S.T.L.H. *The Thirteen Books of Euclid's Elements*; The University Press: University of Minnesota, 1908.
4. G. H. Hardy, C.P.S. *A Mathematician's Apology*; Cambridge University Press: Cambridge, 1967.

[4] since the sum of the reciprocal of natural numbers diverges

 doi:10.20944/preprints202406.1296.v2

5. Martin Aigner, G.M.Z. *Proofs from THE BOOK*; Springer Berlin: Heidelberg, 2018.
6. Dickson, L.E. *History of the Theory of Numbers: Divisibility and Primality*; Number v. 1, Dover Publications: Washington, D.C., 2005.
7. Meštrović, R. Euclid's theorem on the infinitude of primes: a historical survey of its proofs (300 B.C.–2022) and another new proof **2023**. [arXiv:math.HO/1202.3670].
8. Ribenboim, P. The Little Book of Bigger Primes, 2013.