On algebraic structure of the set of prime numbers

by: Ramin Zahedi*

The set of prime numbers has been analyzed, based on their algebraic and arithmetical structure. Here by obtaining a sort of linear formula for the set of prime numbers, they are redefined and identified; under a systematic procedure it has been shown that the set of prime numbers is combinations (unions and intersections) of some subsets of natural numbers, with more primary structures. In fact generally, the logical essence of obtained formula for prime numbers is similar to formula \(2n - 1\) for odd numbers, and so on. Subsequently, using obtained formula we can define all composite numbers. Finally specified examples for obtained formula are presented.

Keywords: Prime numbers, Prime numbers Formula, Primality; (factorization, cryptography, fractality, thermodynamic, non-linear dynamics and chaos, complexity theory, computer programming, quantum computing, network security). MSC2000: 11A41, 11N05, 11B25, 11A51.

1 Introduction:

The formal definition of a prime number is as follows (using the division method):
"An integer \(p > 1\) is a prime if the only positive divisors of \(p\) are 1 and \(p\) itself."

Examples of prime numbers include: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37...; there are infinitely prime numbers [1, 2, 37].

The prime number detection and generation has been of great interest for mathematicians all over the world for over two centuries. Prime numbers lie at the core of some of the oldest and most perplexing questions in mathematics. Evenly divisible only by themselves and 1, they are the building blocks of integers. In recent decades, prime numbers have emerged from their starring roles in mathematical research, by becoming prized commodities - as elements in a cryptographic scheme widely used to keep digital messages secret [3].

Senior Max Planck Institute mathematician Don Zagier, in his article discusses prime numbers [4]: "...Despite prime numbers simple definition and role as the building blocks of the natural numbers, the prime numbers... grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout" (Havil, (2003) 171, [5]).

Today there are many applications for primes in many scientific fields such as computer science, engineering, security science, physics and chemistry, etc. [6-24, 27, 33-36].

In addition, there are dozens of algorithms in computer science that depend heavily on prime numbers- hashing schemes, sorting schemes, and so on. Indirectly, as a result of studying nonlinear dynamics and chaos, Polish physicist Marek Wolf has discovered at least two instances of fractality within the distribution of prime numbers [6-8].

B.L. Julia has reinterpreted the (pure mathematical) Riemann zeta function as a (thermodynamic) partition function by defining an abstract numerical 'gas' using the prime numbers [10].

*zahedi@let.hokudai.ac.jp, zahedi@logic.let.hokudai.ac.jp, zahedi.r@gmail.com, Logic Group, Hokkaido University, Japan.
Prime numbers are a fundamental ingredient in public-key cryptography, be it in schemes based on the hardness of factoring (e.g. RSA), of the discrete logarithm problem, or of other computational problems. Generating appropriate prime numbers is a basic, security-sensitive cryptographic operation [25-27, 38].

Complexity theory is a field in theoretical computer science, which attempts to quantify the difficulty of computational tasks and tends to aim at generality while doing so. "Complexity" is measured by various natural computing resources, such as the amount of memory needed, communication bandwidth, time of execution, etc. By analyzing several candidate algorithms for a problem, a most efficient one can be easily identified; for the problem of determining the primality of an integer, the resource that could be examined, is the time of execution [13]. There are some sorts of formulas for prime numbers; the most of these formulas have been constructed and formulated by using the floor functions (in the field of real numbers) [1, 2, 28-31, 37].

The logical nature of formula for the set of prime numbers that has been obtained here (formula (21)), is similar to formula 2n - 1 for the set of odd numbers, and so on. All these kind of algebraic formulas (also as definitions) only contain operators of the ring of integers: multiplication, addition and subtraction.

In fact, linear formula (21) is obtained in the same and unique process that formula 2n - 1 is formulated for the set of odd numbers. However above formal definition (using the division method, where the division operator is not an operator of the ring of integers) of the set of prime numbers seems simple, formula (21) shows clearly how this set of numbers is not a simple set.

2 A new formulation - definition:

Suppose \( p_1, p_2, \ldots, p_r \) are given prime numbers where \( p_i \) is the \( i \)th prime number. It follows from the definition that in a given range of \( (p_r, p_r^2+1) \) any number that is not divisible by any of \( p_1, p_2, \ldots, p_r \) is a prime number, thus let \( H_r \) be a set of natural numbers \( N \) excluding the set of all positive multiples of \( p_1, p_2, \ldots, p_r \).

That is:

\[
H_r = \{ s \mid s \in N, \text{ and } s \text{ is not divisible by } p_1, p_2, \ldots, p_r \}
\]  

(1)

Let \( E_1 \) be the set of natural numbers \( N \) excluding the set of all multipliers of first prime number \( p_1 = 2 \), define \( E_{11} \) as:

\[
E_{11} = \{ m_{11} \mid m_{11} = p_1 x_1 - h_1, x_1 \in N \}
\]  

(2)

where \( h_1 = 1 \), we get \( E_1 = E_{11} \), and let \( E_2 \) be the set of natural numbers \( N \) excluding the set of all multipliers of the second prime number \( p_2 = 3 \), define \( E_{22} \) and \( E_{21} \) as:

\[
E_{21} = \{ m_{21} \mid m_{21} = p_2 x_2 - 2, x_2 \in N \},
\]

\[
E_{22} = \{ m_{22} \mid m_{22} = p_2 x_2 - 2, x_2 \in N \}
\]
hence
\[ E_2 = E_{21} \cup E_{22} = \{ m_2 \mid m_2 = p_2x_2 - h_2, x_2 \in \mathbb{N} \} \] (3)

where \( h_2 = 1, 2 \). Similarly let \( E_i \) be the set of natural numbers \( \mathbb{N} \) excluding the set of all multipliers of the \( i \)th prime number, define \( E_{i1}, E_{i2}, ..., E_{i(p_i-1)} \):

\[ E_{i1} = \{ m_{i1} \mid m_{i1} = p_ix_i - 1, x_i \in \mathbb{N} \} , \]
\[ E_{i2} = \{ m_{i2} \mid m_{i2} = p_ix_i - 2, x_i \in \mathbb{N} \} , \]
\[ \ldots \]
\[ E_{i(p_i-1)} = \{ m_{i(p_i-1)} \mid m_{i(p_i-1)} = p_ix_i - (p_i - 1), x_i \in \mathbb{N} \} \]

then
\[ E_i = E_{i1} \cup E_{i2} \cup \ldots \cup E_{i(p_i-1)} \]

which is equivalent to
\[ E_i = \{ m_i \mid m_i = p_ix_i - h_i, x_i \in \mathbb{N} \} \] (4)

where \( h_i = 1, 2, 3, ..., p_i - 1 \). It follows from the definitions above that for any set \( E_{ij} \)
\[ E_{ij} \cap E_{ik} = \emptyset \] (5)

for \( j \neq k \) and \( i = 1, 2, ..., r \) and \( j, k = 1, 2, ..., p_i - 1 \).

It follows from (4) and (1) that
\[ H_r = E_1 \cap E_2 \cap E_3 \ldots \cap E_r \] (6)

The system of linear equations obtained from (4) and (6) define the set \( H_r \) in natural numbers
\[ (H)_r = p_1x_1 - h_1 = p_2x_2 - h_2 = p_3x_3 - h_3 = \ldots = p_rx_r - h_r \] (7)

where \( (H)_r \) is the general formula for \( H_r \). The linear equations in (7) can be re-written as
\[
\begin{align*}
p_1x_1 - h_1 &= p_2x_2 - h_2 \\
p_1x_1 - h_1 &= p_3x_3 - h_3 \\
p_1x_1 - h_1 &= p_4x_4 - h_4 \\
&\quad \vdots \\
p_1x_1 - h_1 &= p_rx_r - h_r
\end{align*}
\] (8)
for \( n_1 = 1, 2, 3, ..., p_i - 1 \) and \( i = 1, 2, ..., r \).

Consider a simple linear equation in the set of integer numbers

\[ ax - by = c \quad (9) \]

where \( x \) and \( y \) are the unknown variable and \((a, b) = 1, a > 0, b > 0\).

Equation (9) has infinite number of positive and negative integer solutions and in general

\[ x = cx' + bt, y = cy' + at \quad (10) \]

where \( x' \), \( y' \) are given solutions (these given solutions always exist) of \( ax' - by' = 1 \) and \( t \) can take any integer value \([2, 32]\). Using formula (10) for the first equation in (8) we get

\[ p_2x_2 - p_1x_1 = h_2 - h_1 = h_2 - 1 \]

with the general solution of:

\[ x_2 = (h_2 - 1)x'_2 + p_1t_1, x_1 = (h_2 - 1)x'_2 + p_2t_1 \quad (11) \]

where \( x'_2 \), \( x'_2 \) are given solutions for \( p_2x'_2 - p_1x'_1 = 1 \) and \( t_1 \) is any integer value.

Using formula (11) and the second equation of (8) we get

\[ p_3x_3 - p_1p_2t_1 = (h_2 - 1)p_1x'_1 + h_3 - 1 \quad (12) \]

with the general solution of:

\[ x_3 = [h_3 + (h_2 - 1)p_1x'_1 - 1]x'_3 + p_1p_2t_2, \\
    t_1 = [h_3 + (h_2 - 1)p_1x'_1 - 1]x'_3 + p_1p_2t_2 \quad (13) \]

where \( x'_3 \) and \( t'_1 \) are a given solution for \( p_1x'_3 - p_1p_2t'_1 = 1 \) and \( t_2 \) is any integer value.

Using (11), (13) and the third equation of (8) we obtain

\[ p_4x_4 - p_1p_2p_3t_2 = [(h_4 - 1) + (h_3 - 1)p_1p_2t'_1 + (h_2 - 1)p_1x'_1p_3x'_3] \quad (14) \]

The general solution of (14) is

\[ x_4 = x'_4[(h_4 - 1) + (h_3 - 1)p_1p_2t'_1 + (h_2 - 1)p_1x'_1p_3x'_3] + p_1p_2p_3t_3, \\
    t_2 = t'_2[(h_4 - 1) + (h_3 - 1)p_1p_2t'_1 + (h_2 - 1)p_1x'_1p_3x'_3] + p_4t_3 \quad (15) \]

Continuing this solution process the following general solutions are obtained:

\[ x_1 = x'_1[(h_1 - 1) + (h_{i-1} - 1)t_{i-3} \prod_{l=1}^{i-2} p_l + \sum_{j=2}^{i-2} [(h_j - 1)t_{j-2} \prod_{l=1}^{j-1} p_l \prod_{q=j+1}^{i-1} p_q x'_q]] + t_{i-1} \prod_{j=1}^{i-1} p_j, \]

where

\[ x' \text{ and } t' \]

are given solutions (these given solutions always exist) of \( ax' = by' = 1 \) and \( t \) can take any integer value \([2, 32]\).
\[ t_{i-2} = \hat{t}_{i-2}[(h_i-1)+(h_{i-1}-1)\bar{t}_{i-2} \ \prod_{l=1}^{i-2} p_l + \sum_{j=2}^{i-2} [(h_j-1)\bar{t}_{j-2} \ \prod_{l=1}^{j-1} p_l \ \prod_{q=j+1}^{r} p_q \bar{x}_q]] + t_{i-1}p_i \]

for \( j = 2, 3, 4, ..., i - 2; i = 4, 5, 6, ..., r; \) and \( \hat{t}_0 = \hat{x}'_1. \) Using (16) the values of \( x_r \) and \( t_{r-2} \) can be obtained as:

\[ x_r = \hat{x}'_r[(h_r-1)+(h_{r-1}-1)\bar{t}_{r-3} \ \prod_{l=1}^{r-2} p_l + \sum_{j=2}^{r-2} [(h_j-1)\bar{t}_{j-2} \ \prod_{l=1}^{j-1} p_l \ \prod_{q=j+1}^{r} p_q \bar{x}_q]] + t_{r-1} \ \prod_{j=1}^{r-1} p_j \]

\[ t_{r-2} = \hat{t}_{r-2}[(h_r-1)+(h_{r-1}-1)\bar{t}_{r-3} \ \prod_{l=1}^{r-2} p_l + \sum_{j=2}^{r-2} [(h_j-1)\bar{t}_{j-2} \ \prod_{l=1}^{j-1} p_l \ \prod_{q=j+1}^{r} p_q \bar{x}_q]] + t_{r-1}p_r \]

where \( \hat{x}'_i \) and \( \hat{t}_{i-2} \) are given solutions of:

\[ p_i x'_i - t'_{i-2} \ \prod_{k=1}^{i-1} p_k = 1 \]  

(19)

It is clear that given solutions \( \hat{x}'_i \) and \( \hat{t}_{i-2} \) always exist, as equation (19) is an especial case of equation (9).

Note, in (17) and (18) \( t_{r-1} \) is a free integer variable. Using (17) and (18) the variable \( t_i \) can be re-written in terms of \( t_{r-1} \). Furthermore, using (16) the variable \( x_i \) can be re-written in terms of \( t_{r-1} \), and general solutions of (19), and \( h_i \) and \( p_i \). Using these, \( x_1 \) can be obtained:

\[ x_1 = t_{r-1} \ \prod_{l=2}^{r} p_l + (h_r-1)\bar{t}_{r-2} \ \prod_{l=2}^{r-1} p_l + (h_{r-1}-1)\bar{t}_{r-3}p_r \ \bar{x}'_r \ \prod_{l=2}^{r-2} p_l + \sum_{j=2}^{r-2} [(h_j-1)\bar{t}_{j-2} \ \prod_{l=1}^{j-1} p_l \ \prod_{q=j+1}^{r} p_q \bar{x}_q]] \]

(20)

Using (20) and (7), the general formula for set \( H_r \) (for \( r > 2 \)) can be formulated as follows:

\[ (H)_r = p_1 x_1 - 1 = t_{r-1} \ \prod_{l=1}^{r} p_l + (h_r-1)\bar{t}_{r-2} \ \prod_{l=1}^{r-1} p_l + \sum_{j=2}^{r-1} [(h_j-1)\bar{t}_{j-2} \ \prod_{l=1}^{j-1} p_l \ \prod_{q=j+1}^{r} p_q \bar{x}_q]] - 1 \]

(21)

where \( r = 3, 4, 5, 6, ..., j = 2, 3, 4, ..., r - 1; \) and \( h_j = 1, 2, 3, 4, ..., p_j - 1; \) and parameter \( t_{r-1} \) is a free integer variable, and \( \hat{t}_0 = \hat{x}'_1. \)

Following the definition of set \( H_r \), the integer values of formula (21) are primes in the range \( (p_r, p_{r+1}^2) \) or range \( [p_{r+1}, p_{r+2}^2). \) We remember that all terms in (21) are made up of prime numbers \( p_1, p_2, p_3, ..., p_r. \) Also we must note it is clear that in range \( (p_r, p_{r+1}^2), \) we have at least \( p_{r+1} \) as a prime number. Furthermore, \( x'_1 \) and \( t'_{i-2} \) in equation (19) do not have unique values and hence formula (21) can be written in different but equivalent cases. As example, let \( r = 2, \) using formula (11) and (7) we have:

\[ (H)_2 = p_1 x_1 - 1 = 6t_1 + 2h_2 - 3 \]

(22)
prime numbers in range (3, 25) can be obtained by (22). For \( r = 3 \), using formula (21) (we use formula (21) for \( r = 3, 4, 5, 6, 7, \ldots \)) we have
\[
(H)_3 = 30t_2 - 6h_3 - 10h_2 + 15 \tag{23}
\]
formula (23) gives all primes in range (5, 49). For \( r = 4 \) we may get
\[
(H)_4 = 210t_4 + 90h_4 + 2184h_3 + 4550h_2 - 6825 \tag{24}
\]
formula (24) defines all primes in range (7, 121). For \( r = 5 \) we have
\[
(H)_5 = 2310t_4 - 210h_5 - 18810h_4 + 114114h_3 + 190190h_2 - 285285 \tag{25}
\]
formula (25) defines all primes in range (11, 169). Similar formulas can be derived to obtain other primes in the proceeding ranges.

From the definition of set \( H_r \), it is clear that the integer values of formula (21) gives all primes and 2nd numbers, in range \( [p_r^2 + 1, p_r^3 + 1) \); \( kth \) number is a number which, except itself and 1, is divisible by \( k \) number and only \( k \) number of primes. Similarly using formula (21), all primes and also all composite numbers i.e. 2nd, 3rd, \ldots, \( kth \) numbers can be define in the range \( [p_r^k + 1, p_r^{k+1}) \).

In addition using some theorems such as Bertrand’s postulate (that states for every \( n > 1 \), there is always at least one prime \( p \) such that \( n < p < 2n \)), the action ranges of (21) can be expanded for larger ranges. Here by Bertrand’s postulate, it is easy to show that all primes and composite numbers i.e. 2nd and 3rd and \ldots \( kth \) numbers can be defined by (21), in range \( [p_r^k + s, p_r^{k+1}) \), if \( p_r + 1 > 2^{(k+1)(s-1)} \) and all \( (k+1)th \) numbers i.e.:
\[
p_{r+i_1}, p_{r+i_2}, \ldots, p_{r+i_{k+1}}
\]
are set aside from this range; where \( k \geq 2, s \geq 1, i_j = 1, 2, 3, \ldots, s \) and \( j = 1, 2, 3, \ldots, k \).

As one of the perspectives of application of formula (21), we may point to the Integer Factorisation Problem, which is a basic discussion in cryptography and security sciences. For study of this problem, that considers the prime factors of natural numbers, we can put equally the given number(s) to value of formula (21) and study the obtained equation(s).

Also formula (21), specially follow to its linear structure, can give us some new ways to study prime numbers from geometrical points of view.

We believe formula (21) is a basic formula for the set of prime numbers (simply as formula 2n -1 is a basic formula for the set of odd numbers), and doubtless it can be useful in many fields where prime numbers are used and applied [6-24, 27, 33-36].

3 Conclusion:

As you could see in various stages of the article, in fact we assumed that prime numbers: \( p_1, p_2, \ldots, p_r \) are given, and then through a systematic method and process the prime numbers in range of \( (p_r, p_r^{2+1}) \) or range \( [p_{r+1}, p_{r+1}^2) \), by formula (21) has been obtained. This formula is linear, and factors of this formula only depend on
The process can also be used for obtaining the next and larger ranges continually. Finally, we could specify the set of prime numbers and define and identify them. Based on the structure of formula (21) for the set of prime numbers, we show here that the prime numbers are the result of combination of some subsets of natural numbers with more primary structure. Subsequently, using formula (21) we define composite numbers. It should be emphasized again that the logical nature of formula (21) for the set of prime numbers is similar to formula \(2n - 1\) for the set of odd numbers and so on, as it was obtained from the same (and unique) process which this formula is formulated for the set odd numbers. All these kind of algebraic formulas (also as definitions) only contain operators of the ring of integers: multiplication, addition and subtraction. For more clarity, we may simply and correctly compare formula \(2n – 1\) for (positive) odd numbers with formula (21) for prime numbers as follows:

“An odd number is a positive integer that is not divisible by 2” (using the division method), or
“An odd number is an integer value of algebraic linear formula \(2n – 1\) in the range \([1, +\infty)\)”;

“A prime number is a positive integer (> 1) that is not divisible by any number except 1 and itself” (using the division method), or
“A prime number is an integer value of algebraic linear formula (21) in the range \([p_r, p_{r+1})\) or range \([p_{r+1}, p_{r+2})\), where \(r =3, 4, 5, 6, \ldots, +\infty\)”.

Thus formula (21) could not only be used in theoretical and logical studies of natural numbers but also especially it could be used for practical applications of prime numbers. In addition it could be used for study of some mathematical problems that are closed to prime numbers, such as Riemann hypothesis.

*
(take there is same linear algebraic formulas for \(r =1, 2\); see previous page).

References:

[14]- http://www.maths.ex.ac.uk/~mwatkins/zeta/unusual.htm
[16]- Joseph F. Lawler Jr., “Identifying prime numbers with a DNA computer,” Johns Hopkins University, Paul Ehrlich Research Award (research project), (2002). (See: http://www.jhu.edu/~gazette/2002/08apr02/08young.html)
[20]- http://www.wolframscience.com/preview/nks_pages/?NKS0640.gif (Stephen Wolfram explains how primes can be computed (at least theoretically) using cellular automata).
[30]- http://mathworld.wolfram.com/PrimeNumber.html
[31]- http://mathworld.wolfram.com/PrimeFormulas.html