IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 14, Issue 1 Ver. II (Jan. - Feb. 2018), PP 01-05
www.iosrjournals.org

Proof of the Riemann hypothesis

Vadim N. Romanov
Saint-Petersburg, Russia
Corresponding Author: Vadim N.Romanov

Abstract: *The article investigates the behavior of the ζ-function by numerical methods. On this basis, a proof of the Riemann hypothesis is given.*
Keywords: *number theory, natural numbers, ζ-function, Riemann hypothesis*

Date of Submission: 18-01-2018 Date of acceptance: 02-02-2018

I. Introduction. Analysis of general relations

A large number of papers have been devoted to the study of the properties of the Riemann ζ-function [1 − 5]. Here and below $z = x + iy$, where x and y are real numbers. In [3] the values of the function are shown in steps of 0.1 in the interval of interest to us $0 \le x \le 1$, as well as the values of its roots for $y < 100$. The accuracy of the calculations is not indicated, which reduces the value of the data presented. The main unsolved problem is the proof of the Riemann hypothesis, which consists in the assertion that all the zeros of the ζ-function in the strip $0 \le x \le 1$ are on a line $x = 1/2$. By now, it has been proved that there are an infinite number of zeros on the line $x = 1/2$, and in addition there are no zeros on the ends of the interval. This paper is devoted to the study of characteristic points of the ζ-function, discussion of the reasons for the validity of the Riemann hypothesis and its proof. We use for the zeta function a representation valid for $x > 0$ [5]:

$$(1-2^{1-z})\varsigma(z) = \sum_{N=1}^{\infty}(-1)^{N+1}N^{-z} \equiv S(z). \tag{1}$$

The first factor on the left-hand side of (1) can be discarded, since for $0 < x < 1$ it does not vanish. Thus, the problem reduces to investigating the zeros of the sum of the series, that is, function $S(z)$. Using the exponential representation, we obtain

$$S(z) = u + iv = \sum_{N=1}^{\infty}(-1)^{N+1}\cdot\exp(-x\ln N + 2k\pi y)\cdot\exp[-i(y\ln N + 2k\pi x)]. \tag{2}$$

The common factor $\exp(2k\pi y)$ can be ignored, and to simplify the calculations we use the main value of the logarithm, assuming $k = 0$ that in our case it does not reduce the generality. Then (2) can be written as the system of two equations

$$u = F_1(x,y) = \sum_{N=1}^{\infty}(-1)^{N+1}A(x,N)B_1(y,N) = 0, \tag{3a}$$

$$v = F_2(x,y) = \sum_{N=1}^{\infty}(-1)^{N+1}A(x,N)B_2(y,N) = 0, \tag{3b}$$

where $A(x,N) = 1/N^x$, $B_1(y,N) = \cos(y\ln N)$, $B_2(y,N) = \sin(y\ln N)$. We are interested in the joint zeros of the functions F_1 and F_2. It is clear that $F_1(x, 0) = S(x)$, $F_2(x, 0) = 0$. These values can be considered as initial conditions in the problem of finding zeros. We carried out direct calculations of the quantity $F_1(x, 0) = S(x)$ by (1), which was replaced by a finite segment of the series $S_n(x)$, in the range $0.4 \le x \le 0.96$. In this case, n was chosen so that the computation error was acceptable. Calculations for $x < 0.4$ were not carried out, because they require a large expenditure of computer time and are not needed for our purposes. The results of the calculations are given in table 1.

Table 1 The values of the sum $S(x)$ and the ζ-function $\zeta(x)$ for rational x

x	$S(x)$	Error	Number of terms of the series	$1/(1-2^{1-x})$	$\zeta(x)$
0,40	0,58328087	0,003906	1048550	-1,93904960	-1,13101054
0,45	0,59415255	0,001995	1000000	-2,15477445	-1,28026473

0,49	0,60238832	0,001148	1000000	-2,35821139	-1,42055898
0,50	0,60439864	0,001	1000000	-2,41421356	-1,45914740
0,51	0,60639269	0,000871	1000000	-2,47252484	-1,49932098
0,55	0,61411205	0,000734	500000	-2,73193995	-1,67771723
0,60	0,62365499	0,000473	350000	-3,12981296	-1,95192346
0,65	0,63300955	0,000249	350000	-3,64218282	-2,30553649
0,70	0,64214318	0,000132	350000	-4,32629967	-2,77810384
0,75	0,65108093	0,0000695	350000	-5,28521351	-3,44110174
0,80	0,65983631	0,0000366	350000	-6,725023959	-4,43741499
0,85	0,66841681	0,0000196	350000	-9,126629718	-6,10039272
0,90	0,67682681	0,0000102	350000	-13,93272617	-9,43004268
0,95	0,68506912	0,0000054	350000	-28,35678887	-19,4263605
0,96	0,68669761	0,0000048	350000	-35,56968648	-24,4256200

Consider how the functions F_1 and F_2 change. Expressions (3a), (3b) represent the sum of cosines and sinuses, respectively, with decreasing period, bounded by the values of the amplitude A (x, N) and parity indicators that change sign depending on the parity of the number N. The amplitude A (x, N) decreases with increase in N (at constant x) and with an increase in x (at constant N). In this case, F_1 and F_2 take positive and negative values, depending on whether the sets of which sign predominate (plus or minus). With increasing y, the period of the functions B_1 and B_2 decreases from $2\pi / \ln 2$ to $2\pi / \ln N_0$, where N_0 is the number of terms in the series considered in (3a), (3b). Consider how y must vary to ensure periodicity, if $\ln N$ is fixed. For N = 2, the period is $\Delta y = 2\pi$ / ln2 = 9.06; for $N = 10^6$ it is $\Delta y = 2\pi$ / 6ln10 = 0.455. Let us now consider how the period and the distance between the zeros of the functions B_1 (B_2) vary for a fixed y (the number of terms of the series is finite). At y = 0,1 ... 0,2 zeros are absent, since $y \ln N_0 < \pi$. At y = 0.3 there is one zero, since $y \ln N_0$ only slightly exceeds π. For y = 1, the function B_2 has five zeros, namely: for N_0 = 1, 23, 535, 12391, 286751 (the first value is exact and the others are approximate), which is caused by a very slow change in the logarithm function with increasing N. The zeros of the function B_1 at the value of y = 1 are shifted by π / 2 relative to the zeros of B_2 and fall on the maxima and minima of the function B_2. The function B_1 has 4 zero values: for N_0 = 5, 111, 2576, 59610 (the values are approximated). These regularities are also valid for the functions F_1 and F_2, namely, for small values of y, long-period components predominate, and the distance between the zeros of these functions is significant, and for large y the short-period components play a major role, and the distance between the zeros becomes small.

II. The study of the characteristic points of the ζ-function

The behavior of the functions F_1 and F_2 was studied in the range $0.4 \le x \le 0.96$ for different values of y from 0.1 to 10^6. For each fixed x, the function F_1 (x, y) varies periodically with increasing y. For y = 0, it is S (x) (see table 1), then it increases, reaches a positive maximum, decreases, passes through zero, reaches a minimum, increases again, passes through 0, reaches a maximum, etc. The function F_2 (x, y) also varies periodically. For y = 0, it is 0, then decreases, reaches a minimum, then increases, passes through 0, reaches a maximum, decreases, passes through 0, and so on. For each fixed y, as x increases from 0.4 to 0.96, the functions F_1 (x, y) and F_2 (x, y) change monotonically over the entire range of x or one of them has a weak maximum (minimum) and the other changes monotonically over the entire range with the change sufficiently slow (smooth). It is not possible to give all these data in view of the large volume, so we list the characteristic cases representing a complete group of possibilities. Both functions F_1 (x, y) and F_2 (x, y) are positive and simultaneously decrease or one function decreases and the other increases or one has a weak maximum (minimum) and the other decreases; both functions are negative and simultaneously increase (their absolute values decrease); one of the functions is positive and the other is negative: both functions increase or the negative increases and the positive decreases or the positive has a weak minimum and the negative increases; both functions or one of them change sign, changing monotonically or one of them has a weak minimum (maximum) and the other monotonically increases (decreases). The cases listed repeat periodically with increasing y. The analysis shows that the functions F_1 (x, y), F_2 (x, y) for a given y can intersect no more than once with increasing x and only at the joint (common) zero. The joint zeros correspond to those values of y in a sufficiently small neighborhood of which both functions change sign in the same or opposite directions (and the monotonic behavior remains with increasing x), and the sign change occurs near (in the neighborhood) value of x at which there is a joint zero, i.e. in our case, near 1/2. We note that a change in the sign of only one of the functions can be observed when this function is approached to zero or at a significant distance from the common zero in the variable y. In the remaining cases listed above, the values of the functions F_1 (x, y) and F_2 (x, y) are considerably different over the entire range of variation of x. The position of the characteristic points (maxima, minima and zeros) of the functions F_1 and F_2 was studied, when y increases, for x in the range from 0.4 to 0.96. The results show that the values of y, corresponding to the characteristic points of the functions, vary monotonically over the entire range of variation of x or have a weak maximum (minimum) with a change that is within narrow limits. Therefore, it is always possible to establish a correspondence between points, choosing a step sufficiently small, for example, Δx = 0.01 ... 0.05. Difficulties can arise at very large values of y due to a decrease in the period of the change of functions, but here too, acting sequentially and choosing a step small enough, it is always possible to determine the displacement of the corresponding point. In table 2, as an example, the results of calculating the positions of some maxima of the functions F_1 and F_2 are shown when x varies from 0.4 to 0.96 (maxima with different behavior are chosen). Table 3 gives the results of the study of the alternation of zeros of functions (maxima and minima are not specified so as not to increase the volume of the table) for x = 0.49, 0.50, 0.51, which allows us to understand the existing regularities and draw conclusions. In table 3, the values of y for the same zeros for x = 0.75 are also given for comparison.

Table 2 The change in the positions of the maxima of the functions F_1 and F_2 depending on x

$\boldsymbol{x}$	0,4	0,45	0,49	0,5	0,51	0,55
$\boldsymbol{y_{11}}$	11,952	11,9495	11,948	11,947	11,94696	11,944
$\boldsymbol{F_1(y_{11})}$	2,81603	2,70660	2,62483	2,60504	2,58551	2,50951
$\boldsymbol{y_{12}}$	100000,0979	100000,0955	100000,093	100000,092	100000,091	100000,088
$\boldsymbol{F_1(y_{12})}$	19,50806	14,16015	11,18868	10,58045	10,01742	8,14856
$\boldsymbol{y_{22}}$	100000,291	100000,296	100000,300	100000,301	100000,302	100000,307
$\boldsymbol{F_2(y_{22})}$	18,54156	13,37520	10,49040	9,89778	9,34836	7,51734
$\boldsymbol{x}$	0,6	0,65	0,7	0,75	0,8	0,85
$\boldsymbol{y_{11}}$	11,9469	11,9472	11,948	11,9493	11,951	11,953
$\boldsymbol{F_1(y_{11})}$	2,42069	2,33705	2,25862	2,18506	2,11607	2,05135
$\boldsymbol{y_{12}}$	100000,082	100000,074	100000,064	100000,0523	100000,0375	100000,02
$\boldsymbol{F_1(y_{12})}$	6,47099	5,29306	4,45088	3,83727	3,38077	3,03402
$\boldsymbol{y_{22}}$	100000,314	100000,322	100000,330	100000,3403	100000,351	100000,364
$\boldsymbol{F_2(y_{22})}$	5,85785	4,67857	3,82338	3,19004	2,71101	2,34109
$\boldsymbol{x}$	0,9	0,95	0,96			
$\boldsymbol{y_{11}}$	11,955	11,958	11,9581			
$\boldsymbol{F_1(y_{11})}$	1,99062	1,93364	1,92267			
$\boldsymbol{y_{12}}$	99999,9995	99999,976	99999,972			
$\boldsymbol{F_1(y_{12})}$	2,76489	2,55134	2,51386			
$\boldsymbol{y_{22}}$	100000,377	100000,392	100000,395			
$\boldsymbol{F_2(y_{22})}$	2,04968	1,81573	1,77445			

Note. The values of y_{11} correspond to the second maximum of F_1; y_{12} and y_{22} correspond to the maxima of F_1 and F_2, respectively, for large values of y.

Table 3 The change in the position of the zeros of the functions F_1 and F_2, depending on x

F_1, F_2	x=0,49	x=0,5	x=0,51	x=0,75
	y	y	y	y
F_2=0	0	0	0	0
F_1=0 F_2=0	8,2413 5,4094	8,2482 5,4066	8,2553 5,4037	8,4735 5,3371
F_1=0 F_2=0	10,0799 9,1129	10,0722 9,1118	10,0645 9,1107	9,83 9,0867
F_1=0 F_2=0	13,9805 12,0339	14,0563 12,0351	zero is missing (the preceding minimum is positive) 12,0363	zero is missing 12,0692
F_1=0 F_2=0	14,2140 14,1340	**14,1383** **14,1346**	zero is missing 14,1352	zero is missing 14,1547
F_1=0 F_2=0	17,3244 15,8890	17,3295 15,8869	17,3348 15,8848	17,5031 15,8273
F_1=0 F_2=0	18,7353 18,0771	18,7311 18,07795	18,7268 18,0788	18,5798 18,1016
F_1=0 F_2=0	zero is missing 19,9574	- 19,9614	- 19,9654	- 20,0984
F_1=0 F_2=0	21,0025 21,0266	**21,02196** **21,02180**	21,0437 21,0169	zero is missing 20,8616
F_1=0 F_2=0	21,4860 22,9751	21,4650 22,9748	21,4416 22,9744	zero is missing 22,9648
F_1=0 F_2=0	24,4143 zero is missing	24,4293 -	24,4452 -	zero is missing -
F_1=0 F_2=0	25,0237 25,0032	**25,01121** **25,01087**	24,9978 25,0187	zero is missing zero is missing
…	…	…	…	…

*F_1=0 F_2=0	98,8295 98,8701	**98,8311** **98,8322**	98,8327 zero is missing	98,7941 zero is missing
…	…	…	…	…
F_1=0 F_2=0	99999,6797 99999,7042	**99999,7009** **99999,7009**	zero is missing 99999,6975	zero is missing zero is missing
F_1=0 F_2=0	100000,3466 100000,0596	100000,3504 100000,0569	100000,3544 100000,0540	100000,6009 99999,9222
…	…	…	…	…
F_1=0 F_2=0	999997,9784 999997,9605	**999997,9682** **999997,9683**	999997,9540 999997,9778	zero is missing zero is missing

Note. The zeros of each function are arranged in order of increasing y; the values of y are given with rounding. Bold-face type denotes joint (common) zeros. In the line with the * sign, there is a weak maximum y = 98.8402 at x = 0.6 (step Δx = 0.05).

Now we consider the behavior of the solutions of system (3a), (3b), i.e. joint zeros of the functions F_1 and F_2. First, we prove that the functions F_1 and F_2 have infinitely many zeros for each x. We choose an arbitrary rational x from the interval 0.4 ... 0.96. The functions F_1 and F_2 are periodic, and their characteristic points are repeated with increasing y, while the period decreases (see above). We divide the infinite interval $0 < y < \infty$ into segments containing one negative and one positive value of the function F_1 (respectively F_2), which follow one another. Between them, the function F_1 (F_2) takes a zero value. The value of the segments decreases when $y \to \infty$, and there are infinitely many such segments, which proves the assertion. Since F_1 and F_2 are continuous functions in x and y, this conclusion is valid for any x in the interval $0 < x < 1$. For $x = 1/2$, some of the zeros of F_1 and F_2 coincide, forming a set of solutions of system (3a), (3b). To better understand what is happening, we use a graphical representation on the plane. Let x is the radius of the circle, and $| y |$ is the angle reckoned from zero in the positive direction. Then the zeros of the functions F_1 and F_2 will correspond to points on the circle of a given radius, and there will be infinitely many of them. For $x = 1/2$, the common zeros of F_1 and F_2 correspond to "double" points on the circle. If $x \neq 1/2$, namely, $x = 1/2 \pm \alpha$, where $0 < \alpha < 1/2$, then the double points split. The above analysis (see table 3) shows that in this case two points are appeared: one corresponds to zero F_1, and the second to zero F_2, and the splitting occurs in opposite directions (in the value of y) from the double point. The magnitude of the splitting depends on $| x - 1/2 |$, and the signs alternate by turns depending on y. With the removal of x from the value 1/2, the points are increasingly divided (table 3). In some cases, one of the zeros may be absent, i.e. as a result of the splitting, only one point appears, corresponding to zero of one of the functions, which depends on the ratio of the shift of the function caused by the change of x and the minimum value of the function preceding the double point. In this case, the preceding minimum for a given $x \neq 1/2$ is positive or, more generally, the signs of the neighboring minimum and maximum are the same (the positive minimum is between two positive maxima or the negative maximum is between two negative minima). If for some $x_1 \neq 1/2$ this "anomaly" is observed for the first time, then for all $x > x_1$ it remains. The number of anomalies increases with increasing x (and y). Since in our calculations the infinite series was replaced by a finite segment of the series, then the results of table 3 can be influenced by the following factors: the approximation error, the different rate of change, the different rate of convergence of the functions F_1 and F_2 near zero, and also the path through which we approach zero (from the one side or from different sides with maintenance of the sign). But these errors are small compared to the useful effect and do not affect the analysis results in principle, therefore the double points are reliably identified by the coincidence of two digits after the comma in the value of y corresponding to the double point. As can be seen from table 3, the difference between the zeros after the splitting is significantly greater than the error. We show that there are infinitely many double points on the line $x = 1/2$. The proof is carried out as in the previous case. Note that for the appearance of a double point, it is necessary that F_1 and F_2 change synchronously in the same or opposite directions. Such cases, as our analysis showed, are repeated periodically with increasing y from 0 to ∞. The period of repetition of double points depends on the range of values of y and the initial conditions, i.e. of the values $F_1(x, 0)$ and $F_2(x, 0)$. We divide the interval $0 < y < \infty$ into segments, so that each function has a pair of positive and negative values closest to each other, and the functions vary synchronously, i.e. with approximately the same sensitivity factor in y. The length of the segments will decrease when $y \to \infty$, and there will be infinitely many such segments, which proves the assertion. It is clear that for a finite y the number of double points is always less than the number of zeros of each of the functions F_1, F_2, since their appearance is associated with more strong constraints. We have the following *upper bound estimate*: the number of double points is less than $y \ln N_0 / 2\pi$. What can be said about the reasons for the appearance of double points? They are due to the symmetry properties of the functions. As can be seen from (3a), (3b), F_1 is an even function and F_2 is an odd function of y, therefore the set of zeros of these functions is invariant with respect to the substitution of y by $-y$, which corresponds to the reflection symmetry about the x axis (for fixed x, that is, in the one-dimensional case, this symmetry can be likened to symmetry with respect to time inversion). The second type of symmetry is the central symmetry with respect to the point $x = 1/2$, i.e. to the center of the strip, which determines the invariance of the set of zeros with respect to the substitution $x \to 1-x$ (with the simultaneous replacement of y by $-y$). For $x = 1/2$, both types of symmetry coincide, so double points appear for some values of y. If $x_1 \neq 1/2$, then the second type of symmetry disappears, and the double point splits.

III. The proof of the Riemann hypothesis and discussion of results

To prove the Riemann hypothesis, we use the functional relation for the ζ-function [5]. We write it in the form

$$C(z)\varsigma(z) = D(z)\varsigma(1-z), \tag{4}$$

where $C(z)$, $D(z)$ are functions of z. It follows from (4) that if $z_0 = 1/2 + iy_0$ is the zero of the ζ-function, then $1 - z_0 = 1/2 - iy_0$ is also a zero of this function, which agrees with the symmetry of the functions F_1, F_2 (see above). Here it is assumed that $C(z_0)$ does not become infinite, and $D(z_0)$ does not vanish; these conditions are satisfied in our case. Suppose now that there is a zero of function for $\tilde{z} = (1/2-\alpha) + i\,\tilde{y}$, where $0 < \alpha < 1/2$. Then it follows from (4) that zero is also $1-\tilde{z} = (1/2+\alpha) - i\,\tilde{y}$ (if the condition specified above is satisfied for $\tilde{z}$). But since the functions F_1 and F_2 with increasing x for the same value of y (axial symmetry remains for any x) vary monotonically and can intersect only once, the second double point cannot arise (see the analysis above) that proves the validity of the Riemann hypothesis. Otherwise, we could argue that equations (3a), (3b) always have two joint solutions (for the same y), as soon as they have one solution. And this is impossible. In other words, for $x \neq 1/2$ two types of symmetry required for the appearance of a double point cannot be realized simultaneously: the axial symmetry (reflection symmetry) and the central symmetry. To better understand the meaning of the Riemann hypothesis, we use the following analogy. There is an atomic system. It is required to study its spectrum in the strip $0 < x < 1$, i.e. to determine the levels of the stationary energy of the system. As we know, this problem reduces to the solution of the wave equation (the Schrödinger equation). Suppose that the wave function of the system ψ is approximated (for fixed x) by the weighted sum of the basis functions B_1, B_2 and has the form (3a), (3b). These basis functions belong to two "non-combining" sets, so $\psi = F_1$ corresponds to the cosine-like state of the system, and $\psi = F_2$ to the sine-like state. Therefore, the problem of finding energy levels (eigenvalues) reduces to a system of two equations separately for two non-combining sets. In general, the solutions for these sets are different; however, the functions can have common knots (zeros). Their position is completely determined by the symmetry properties of functions F_1, F_2, i.e. admissible transformations necessary for the appearance of common zeros. In our case, as noted above, this is the reflection symmetry and the central symmetry. For $x = 1/2$, we have a "degenerate" state, since both types of symmetry coincide. If $x \neq 1/2$, degeneracy is removed, since only the symmetry of reflection remains.

IV. Conclusion

In conclusion, let us consider how the number of common zeros (double points) p depends on y. The upper bound estimate is obtained above. To obtain the lower bound estimate, calculate the average period of appearance of such zeros in the interval $0 < y < 100$ for $x = 1/2$. It is approximately equal to 3.09, i.e. it is close to π. Taking into account the results of the previous analysis, we can write down the following lower bound estimate for the number of common zeros (double points):

$$p > 1 + [(y - y_0)/\pi], \tag{5}$$

where $y_0 = 14{,}13$ corresponds to the first common zero; [·] is the integer part of a number; $y = 10^2 k$, $k = 1, 2, 3$ .... It is easy to verify that the estimate (5) is satisfied for $y = 100$; in addition, it agrees with the estimates for large values of y obtained by more complicated methods [5].

References

[1]. Gradshtein I. S., Ryzhik I. M. Tables of integrals, sums, series and products. Moscow, Publ. house "Nauka", 1962 (in Russian).
[2]. Handbook of mathematical functions. Ed. by M. Abramowitz and I. A. Stegun. Moscow, Publ. house "Nauka", 1979 (in Russian).
[3]. Janke E., Emde F., Losch F. Special functions. Moscow, Publ. house "Nauka", 1968 (in Russian).
[4]. Romanov V.N. Investigation of the fundamental problems of number theory. Publishing house "Asterion", 2015 (in Russian).
[5]. Whittaker E. T., Watson G. N. A course of modern analysis. Moscow, Publ. house "Nauka", 1963 (in Russian).

Vadim N. Romanov,"Proof of the Riemann hypothesis." IOSR Journal of Mathematics (IOSR-JM) 14.1 (2018): 01-05.