Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 339 (2004) 15–20
COMPTES RENDUS MATHEMATIQUE

Number Theory

Empirical estimates of the average orders of orbits period lengths in Euler groups

Francesca Aicardi

Sistiana Mare 56 pr, 34019 Trieste, Italy

Received 15 September 2003; accepted after revision 24 February 2004

Available online 18 May 2004

Presented by Vladimir Arnold

Abstract

The averaged growth rate of period's length of the geometrical progressions $\{q^t \bmod n,\ t = 0, 1, \ldots\}$ for increasing n is empirically estimated for different values of q. The experimental results, obtained for n up to 10^6, allow us to conjecture that the average order of period's length is $C\frac{n}{\ln(n)}$, where constant C depends on q. ***To cite this article: F. Aicardi, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

Résumé

Estimations empiriques des ordres moyens des longueurs des périodes d'orbites dans les groupes d'Euler. On donne une estimation expérimentale du taux moyen de croissance de la longueur de la période des progressions géométriques $\{q^t \bmod n,\ t = 0, 1, \ldots\}$ pour n croissant, pour des valeurs différentes de q. Les résultats empiriques, obtenus pour n jusqu'à 10^6, permettent de conjecturer que l'ordre moyen de la longueur de la période est $C\frac{n}{\ln(n)}$, où la constante C dépend de q. ***Pour citer cet article : F. Aicardi, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

Version française abrégée

Orbites dans les Groupes d'Euler

Récemment Arnold a mis en évidence toute une série de problèmes non résolus concernant les groupes des residues modulo n qui sont premiers à n [1–4]. Arnold a appelé ces groupes *groupes d'Euler* et les a notés par $\Gamma(n)$.

La fonction d'Euler φ associe à chaque entier naturel n l'ordre du groupe $\Gamma(n)$.

Les valeurs de la fonction φ varient de façon très irrégulière. Toutefois, un théorème classique remontant au moins à Dirichlet [6] dit que l'ordre moyen[1] de φ est $\frac{6}{\pi^2}n$, c'est-à-dire que

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \varphi(k)}{n^2} = \frac{3}{\pi^2}.$$

E-mail address: aicardi@sissa.it (F. Aicardi).

1631-073X/$ – see front matter © 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2004.02.021

La multiplication de tous les $\varphi(n)$ éléments du groupe d'Euler par un élément q fixé de ce groupe définit une permutation (notée par $(q*)$) de l'ensemble fini $\Gamma(n)$. Un théorème d'Euler dit que, pour chaque $q \in \Gamma(n)$, la permutation $(q*)$ contient exactement N_q cycles de la même longueur T_q, ainsi que

$$T_q N_q = \varphi(n).$$

L'ordre de la permutation $(q*)$ est alors égal à la longueur de la période T_q de la progression géométrique $\{q^t \bmod n\}$.

Comme les valeurs de la fonction φ, les valeurs $T_q(n)$ et $N_q(n)$ varient très irrégulièrement.

Il n'y a pas de résultats, comme le théorème de Dirichlet, sur les ordres moyens ni pour les fonctions T_q, ni pour les fonctions N_q. Je donne dans la prochaine partie des résultats expérimentaux sur ces asymptotiques.

Résultats expérimentaux

Définition 0.1. Les intégrales approchées $\mathrm{I}[T_q](n)$ et $\mathrm{I}[N_q](n)$ sont définis par

$$\mathrm{I}[T_q](n) = r(q) \sum_{k=q+1}^{n} T_q(k), \qquad \mathrm{I}[N_q](n) = r(q) \sum_{k=q+1}^{n} N_q(k), \tag{1}$$

où les k sont premiers à q et le facteur $r(q) \equiv \frac{q}{\varphi(q)}$ prend en compte le fait que les fonctions $\mathrm{I}[T_q]$ et $\mathrm{I}[N_q]$ ne sont pas définies pour tout n.

A première vue les functions $\mathrm{I}[T_q]$ et $\mathrm{I}[N_q]$ croissent comme des puissances de n (voir Figs. 1 et 2). Les graphiques en échelle bi-logarithmique des fonctions $\mathrm{I}[T_q]$, pour des valeurs différentes de q, sont approximativement des droites parallèles. La même chose est vraie pour les fonctions $\mathrm{I}[N_q]$. Toutefois, on observe que certains graphiques (par exemple ceux de $\mathrm{I}[N_2]$ et de $\mathrm{I}[N_3]$) se croisent. Pour avoir un argument expérimental en faveur de l'existence d'exposants universels il faut faire une étude plus soignée.

Nous calculons, pour des valeurs différentes de q, les intégrales approchées $\mathrm{I}[T_q](n)$ et $\mathrm{I}[N_q]$ selon (1) jusqu'à une valeur $\bar{n}$ de n.

Résultats pour $\mathrm{I}[T_q](n)$. Les calculs numériques pour $\bar{n} \leqslant 10^6$ montrent que les fonctions $\mathrm{I}[T_q]$ sont mieux approchées par des fonctions du type $\gamma \frac{n^\delta}{\ln(n)}$ que par des fonctions du type βn^α. On applique la méthode de régression linéaire aux graphiques bi-logarithmiques des valeurs $\mathrm{I}[T_q](n)\ln(n)$ en fonction de n. En effet, ces graphiques sont encore bien approchés par des droites parallèles.

Dans ce cas nous avons supposé que les ordres moyens des fonctions T_q sont du type $C\frac{n^D}{\ln(n)}$, approchés par les fonctions $c\frac{n^d}{\ln(n)}$, où $d = \delta - 1$ et $c = \gamma \exp(\delta)$, convergent, pour $\bar{n} \to \infty$, respectivement vers D et vers C.

Les valeurs des paramètres d et c atteintes pour $\bar{n} = 10^6$ se trouvent dans le Tableau 3.

Les données empiriques sont compatibles avec l'ordre moyen suivant pour T_q :

$$T_q(n) \sim c(q)\frac{n}{\ln n}.$$

Le coefficient $c(q)$ est approché, pour les valeurs dans le Tableau 1, par

$$c(q) = 2(\varphi(q))^{-1/8}.$$

Résultats pour $\mathrm{I}[N_q](n)$. L'intervalle de $\bar{n}$ considéré ne nous permet pas d'observer une convergence du taux moyen de croissance de N_q vers un exposant universel. Nous pouvons seulement dire que, si tel exposant existe, il est vraisemblablement plus petit que $2/5$ et plus grand que $3/8$.

1. Orbits in Euler groups

Recently, Arnold singled out a series of unsolved problems concerning the multiplicative groups $\Gamma(n)$ ($n \in \mathbb{N}$) of the residues modulo n which are relatively prime to n [1–4]. Arnold has called these groups *Euler Groups*.

The Euler function φ associates to every integer n the order of group $\Gamma(n)$.

The values of function φ vary in a quite irregular way. However, a theorem dating back at least to Dirichlet [6] states that the average order[1] of φ is $\frac{6}{\pi^2}n$, i.e., that

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n}\varphi(k)}{n^2}=\frac{3}{\pi^2}.$$

The multiplication of all $\varphi(n)$ elements of the Euler group by a fixed element q of this group defines a permutation (denoted by $(q*)$) of the finite set $\Gamma(n)$. A theorem by Euler states that, for every $q\in\Gamma(n)$, permutation $(q*)$ contains exactly N_q cycles of the same length T_q, so that

$$T_qN_q=\varphi(n).$$

The order of permutation $(q*)$ is thus equal to the length of period T_q of orbit $\{q^t \bmod n\}$.

Like the values of function φ, values $T_q(n)$ and $N_q(n)$ vary in a very irregular way.

There are no results, like Dirichlet's theorem, on the average orders neither of functions T_q nor of functions N_q. In the next sections I present some empirical estimates of such average orders.

2. First observations

Definition 2.1. The *approximate integrals* $\mathrm{I}[T_q](n)$ *and* $\mathrm{I}[N_q](n)$ are defined by

$$\mathrm{I}[T_q](n)=r(q)\sum_{k=q+1}^{n}T_q(k),\qquad \mathrm{I}[N_q](n)=r(q)\sum_{k=q+1}^{n}N_q(k),\tag{2}$$

where k is relatively prime to q and the factor $r(q)\equiv\frac{q}{\varphi(q)}$ takes into account the fact that T_q and N_q are defined not for all values of n.

At first glance functions $\mathrm{I}[T_q]$ and $\mathrm{I}[N_q]$ grow as powers of n (see Figs. 1 and 2). The graphs in bi-logarithmic scale of functions $\mathrm{I}[T_q]$, for different values of q, become, for increasing n, approximately parallel straight lines.

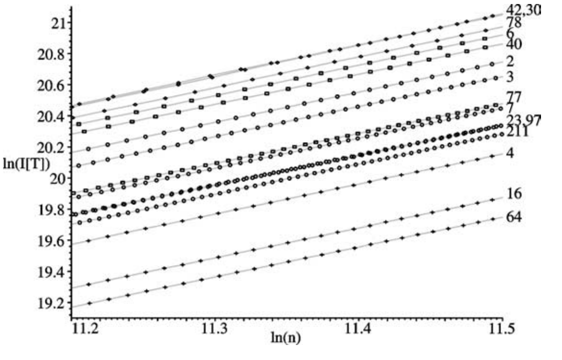

Fig. 1. Graphs of $\mathrm{I}[T_q]$, for $8\times10^4<n<10^5$. The q values are shown near the graphs right ends. *Circles* are used for q prime, *boxes* for $q=$ product of 2 primes, *rhombi* for $q=$ product of 3 primes, *crosses* for $q=$ power of 2.

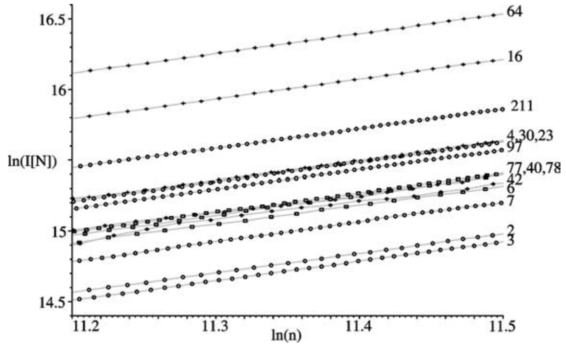

Fig. 2. Graphs of $\mathrm{I}[N_q]$, for $8\times10^4<n<10^5$. The q values are shown near the graphs right ends. *Circles* are used for q primes, *boxes* for $q=$ product of 2 primes, *rhombi* for $q=$ product of 3 primes, *crosses* for $q=$ power of 2.

[1] Note that the expression "weak asymptotics" is the literal translation of the Russian expression for "average order", introduced in [5].

The same is true for functions $\mathrm{I}[N_q]$. This should suggest that the asymptotic averaged growth rates of both families T_q and N_q are universal – i.e., independent of q. This was in fact the Arnold conjecture.

Remark 1. The bi-logarithmic plots of $\mathrm{I}[T_q]$ for different values of primes q, ordered from top to bottom, correspond to increasing values of primes q. Graphs of functions $\mathrm{I}[T_q]$ (for primes q) seem to be disposed in the opposite order but there are some exceptions. For example, the graph of $\mathrm{I}[T_2]$ is higher than that of $\mathrm{I}[T_3]$ (see Fig. 2).

Remark 2. Note that the 'worse straight lines' in Fig. 2 correspond to values of q with 3 distinct prime factors.

3. Estimate of the average orders of $T_q(n)$

Let us suppose that for large n the periods length grows in mean as

$$T_q(n) \sim An^B, \tag{3}$$

where the values of coefficient A and of exponent B depend on q. This means that we suppose that, for $n \to \infty$, the following limit does exist, being equal to

$$\lim_{n\to\infty} \frac{\mathrm{I}[T_q](n)}{n^{B+1}} = \frac{A}{B+1}.$$

We calculate, for different values of q, the approximate integral $\mathrm{I}[T_q](n)$ according to (2) for $n < \bar{n}$. It is a function defined for all values of n relatively prime to q in the interval $(q, \ldots, \bar{n})$. We find the straight line fitting its bi-logarithmic graph by the least square regression method. From parameters α and β of this straight line, satisfying:

$$\ln\bigl(\mathrm{I}[T_q](n)\bigr) \approx \beta \ln(n) + \alpha,$$

we obtain the approximate values b and a of exponent B and of coefficient A of the supposed average orders (3) of T_q:

$$b = \beta - 1, \qquad a = \beta \exp(\alpha).$$

Of course, b and a are functions of $\bar{n}$, which will converge, for $\bar{n} \to \infty$, to B and to A respectively, if it is true that the average orders of functions T_q are given by (3).

Computer experiments show that the values of parameters a and b strongly vary while $\bar{n}$ increases up to a value n_t after which their behaviour becomes more regular. This explains why the values of coefficient a in the case $q = 2$ calculated by Arnold and by me (for different values of $\bar{n}$, both much smaller than n_t), were considerably different (see [4]).

The length n_t of the transient interval depends on q. If q is prime, n_t increases for increasing values of q; if q is not prime, then already for small values of q, n_t goes beyond the values of $\bar{n}$ attained in a reasonable computation time.

For this reason I present here the result of the study of the cases $q = 2, 3, 5, 7, 11, 23, 43$ for values of $\bar{n} < 10^6$.

Tables 1 and 2 show the values attained by b and by a for some values of $\bar{n}$ among the one hundred values considered ($\bar{n} = 10^4 m$, $m = 1, \ldots, 100$).

The main remarks coming from the observation of the plots of the values attained by b and by a versus $\bar{n}$ are the following: firstly, the difference between the values of b for different q decreases for increasing $\bar{n}$; secondly, the values of exponent b for $\bar{n} > 10^5$ appear to be eventually increasing, and the values of coefficient a eventually decreasing, for all values of q.

Analogous computer experiments were executed to approximate functions $\mathrm{I}[T_q]$ by functions of type $\frac{n^\delta}{\ln(n)}$ simply applying the linear regression method to the bi-logarithmic plots of $\mathrm{I}[T_q](n)\ln(n)$ versus n. In fact, these graphs, too, are fitted by parallel straight lines.

Table 1
Values of exponent b

$\bar{n} \Downarrow q \Rightarrow$	2	3	5	7	11	23	43
10^5	0.874	0.887	0.887	0.890	0.880	0.914	0.915
2.5×10^5	0.917	0.921	0.921	0.921	0.921	0.923	0.922
5×10^5	0.924	0.925	0.926	0.93	0.926	0.927	0.929
7.5×10^5	0.928	0.929	0.929	0.931	0.929	0.930	0.931
10^6	0.930	0.931	0.931	0.932	0.931	0.931	0.932

Table 2
Values of coefficient a

$\bar{n} \Downarrow q \Rightarrow$	2	3	5	7	11	23	43
10^5	0.37	0.60	0.55	0.46	0.50	0.36	0.35
2.5×10^5	0.26	0.45	0.4	0.37	0.35	0.33	0.32
5×10^5	0.25	0.43	0.38	0.35	0.34	0.31	0.30
7.5×10^5	0.24	0.42	0.37	0.34	0.33	0.30	0.29
10^6	0.23	0.41	0.36	0.33	0.32	0.30	0.28

Table 3
Values of c and d for $\bar{n} = 10^6$

q	2	3	5	7	11	23	43
d	1.016	1.018	1.017	1.018	1.017	1.018	1.018
c	2.01	1.80	1.60	1.46	1.42	1.32	1.26

Table 4
Values of b and a at $\bar{n} = 10^6$

q	2	3	5	7	11	23	43
b	0.401	0.380	0.387	0.389	0.380	0.402	0.391
a	0.46	0.49	0.57	0.65	0.70	0.90	1.0

In this case we have supposed that the asymptotics of $T_q(n)$ is

$$T_q(n) \sim C \frac{n^D}{\ln(n)} \tag{4}$$

and it is approximated by $c\frac{n^d}{\ln(n)}$, where the exponent $d = \delta - 1$ and the coefficient $c = \gamma \exp(\delta)$ converge, for $\bar{n} \to \infty$, to D and C respectively.

In the interval $7 \times 10^5 < \bar{n} < 10^6$ the estimated values of parameters c and d, for every q, are approximately constant (i.e., their variations are smaller than the values of their standard deviation[2]). Moreover, the values of parameter d attained for different q differ by an amount less than 2×10^{-3}, which is less than the value of the standard deviation $\sigma(d)$, for all q.

The values of d and c attained at $\bar{n} = 10^6$ are listed in Table 3.

Remark 3. For $q = 43$, the data of Table 3 are reached at $\bar{n} = 2 \times 10^6$. The values attained at $\bar{n} = 10^6$ are $d = 1.019$ and $c = 1.25$.

The empirical estimates of d and c show thus that an asymptotical law of type (4) is more convenable than the power law (3) for the average orders of $T_q(n)$. Moreover, they suggest the following average orders for $T_q(n)$:

$$T_q(n) \sim c(q) \frac{n}{\ln n}. \tag{5}$$

The coefficient $c(q)$ is roughly approximated, for the values in Table 1, by

$$c(q) = 2\big(\varphi(q)\big)^{-1/8}.$$

Remark 4. The difference between the estimated value of exponent d and 1, i.e., the difference between the value of δ and 2, is considerably higher than the uncertainty of δ. This fact could be due to the influence on the approximate integral $\mathrm{I}[T_q](n)$ of the nonleading terms of its asymptotical expression. Moreover, the fact that this difference is positive implies that the asymptotics (5) should be attained from below. For example, trying to approximate

[2] The standard deviation σ of exponent d is equal to the standard deviation of coefficient δ, i.e., $\sigma = \sqrt{\sum(y_n - \bar{y})^2 / \sum(x_n - \bar{x})^2}$, where $x_n = \ln(n)$, $y_n = \ln(\mathrm{I}[T_q](n)\ln(n))$ and $\bar{x}$, $\bar{y}$ are their mean values (sums are taken over all n relatively prime to q).

function $\frac{n^2}{\ln(n)} - 2\frac{n^2}{(\ln(n))^2}$ by a function of type $\gamma \frac{n^\delta}{\ln(n)}$, one finds the value of δ varying exactly from 2.021 to 2.018, for $\bar{n}$ increasing from 5×10^5 to 10^6.

4. Empirical results on the average orders of $N_q(n)$

We proceed as for the study of T_q, i.e., we suppose that the average orders of functions N_q are

$$N_q(n) \sim An^B,$$

where the values of coefficient A and of exponent B depend on q.

Also in this case we denote by a and b the empirical estimates of A and B. Their values, calculated from values $\mathrm{I}[N_q](n)$ for $n < \bar{n}$, strongly vary (and oscillate) while $\bar{n}$ is growing.

For the same reasons explained above, still holding for N_q, I present the result of the study of cases $q = 2, 3, 5, 7, 11, 23, 43$ for increasing $\bar{n}$ till 10^6.

Table 4 shows the values of b and a reached at $\bar{n} = 10^6$.

Whereas the behaviour of exponent b becomes more regular for $\bar{n} > 2 \times 10^5$, we do not obtain, at $\bar{n} = 10^6$, a convergence of the values of b, for different q, to a unique value (i.e., to values inside an interval less than the standard deviation). The difference between the minimal and the maximal values of b is greater than 10 times the estimated value of the standard deviation ($\sigma(b)$) of parameter b, being, for all considered values of q, $\sigma(b) \approx 2 \times 10^{-3}$.

Note that, despite of the nonmonotonic sequence of the values of $b(q)$ in Table 2, the values of coefficient a are increasing with q. They grow approximately as $\frac{2}{5}q^{1/4}$.

Remark 5. The straight lines approximating the graphs of $\mathrm{I}[N_2]$ and $\mathrm{I}[N_3]$, for $\bar{n} = 10^6$, actually intersect at $n \approx 10^3$.

Moreover, for some values of q (precisely, $q = 2, 3, 7, 11$), the value of b is slightly increasing and for other ones ($q = 5, 23, 43$) is slightly decreasing for increasing $\bar{n} > 5 \times 10^5$.

For these reasons, we can conclude that in the interval of $\bar{n}$ we had considered we do not observe a convergence of the averaged growth rate of $N_q(n)$ to a universal exponent. We can only infer that, if such an exponent exists, it is likely smaller than $2/5$ and larger than $3/8$.

Acknowledgements

The present work was done and this note was written by the kind request of V.I. Arnold. I would like to thank Alessandra Potrich and Bruno Caprile (IRST, Trento) for their assistance in software's problems.

References

[1] V.I. Arnold, Euler Groups and Arithmetics of Geometric Progressions, MCMME, Moscow, 2003, 40 p.
[2] V.I. Arnold, Fermat–Euler dynamical system and statistics of the geometric progressions, Funct. Anal. Appl. 37 (1) (2002) 1–20.
[3] V.I. Arnold, Ergodic arithmetical properties of the dynamics of geometric progressions, Moscow Math. J. (2003).
[4] V.I. Arnold, Topology and statistics of arithmetic and algebraic formulae, Cahier de Ceremade, avril 2003;
V.I. Arnold, Russian Math. Surveys 58 (4) (2003).
[5] V.I. Arnold, Weak asymptotics of the solutions numbers of Diophantine problems, Funct. Anal. Appl. 37 (3) (1999) 65–66.
[6] P.G. Dirichlet, Abhand. Ak. Wiss., Berlin (Math.), 1849, pp. 78–81; Werke, II, pp. 60–64.