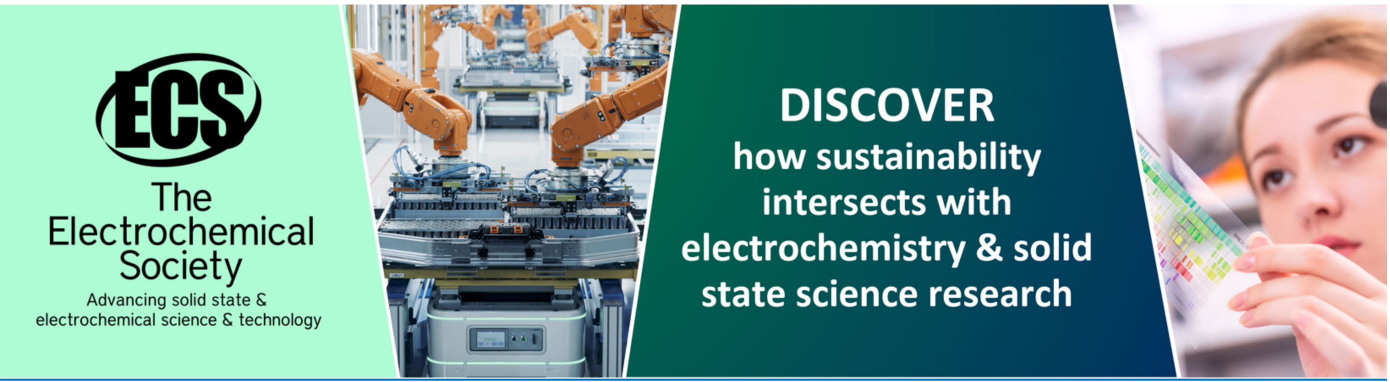
IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

A method for data analysis in algebraic structures

To cite this article: Aleksei Shlepkin *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **919** 052047



A method for data analysis in algebraic structures

Aleksei Shlepkin[1], Irina Sabodakh[1] and Konstantin Filippov[2]

[1]Siberian federal university, 79 Svobodny pr., 660041 Krasnoyarsk, Russia
[2]Krasnoyarsk state agricultural university, 90 Mira pr. ,660049, Krasnoyarsk, Russia

E-mail: shlyopkin@mail.ru

Abstract. Computational group theory investigates the arithmetic properties of algebraic structures. These properties may be utilized in development of algorithms for data analysis and machine learning. This direction is associated with the design, analysis of algorithms and data structures for calculating various characteristics for (mostly finite) groups. The area is interesting for investigation of important from various points of view groups that are important, the data about which cannot be obtained by manually. The present work was carried out in line with computer calculations in groups. Based on the concept of the group growth function, the concept of the group density function is introduced. The growth and density functions of finite simple non-Abelian groups of small orders are constructed. The question of recognizability of finite simple non-Abelian groups by the density function is considered. The problem of effective storage of finite group elements in computer memory using the AVL tree is investigated. Based on the AVL tree, a fast algorithm for enumerating all elements of a finite group is developed. With the computer calculations, the recognizability of two finite simple non-Abelian groups by the group density function is proved.

1. Introduction

One of the main goals of structuring the data is the possibility to apply various strategies and algorithms to establish deep connections and equivalence between the studied objects. It should noted that the speed of the data processing algorithm plays one the most important role in big data. The speed of data analysis directly depends on the structure in which the data is presented, and on how this structure is stored in the computer's memory. In this paper, the objects of study are the sets endowed with an algebraic structure, represented by finite simple non-Abelian groups, as well as their arithmetic properties. These objects are of interest from the point of view of big data and machine learning, since they may be utilized in development of black box algorithms.

In group theory, the properties of a group are called "arithmetic" if they are deter-mined by its numerical parameters: the order of the group and the set of its simple divisors, the orders of the elements, the orders of the subgroups, the degrees of irre-ducible representations, the orders of conjugacy classes, etc. Due to development of high performance computing, the fundamental tasks of studying the arithmetic properties of finite groups and characterizing finite groups using arithmetic parameters have occupied a special place in group theory. Note that the study of the arithmetic properties of groups is of interest for applications, e.g. the orders of group elements determine the parameters necessary for the development of black-box algorithms. On the other hand, the development of algorithms for working with arithmetic parameters of a group is of interest for group theory itself. For example, Vasiliev and

Buturlakin [1] developed an algorithm, which builds in polynomial time the groups with a given set of element orders, or showing that such groups do not exist.

A profound direction in characterizing finite groups by their arithmetic parameters is recognition of a finite group by spectrum and prime graph. A group G is called recognizable (by spectrum) if it is determined by its spectrum up to isomorphism. Recognizability of the finite group by spectrum has been studied by many scientists for three decades: V. Shi, V.D. Mazurov, A.V. Vasiliev, M.A. Grechkoseeva and other mathematicians. To date, it is known that almost all finite simple groups, with a few exceptions, are recognizable by spectrum [2,3] and, in general, the question of recognizing finite simple non-Abelian groups by spectrum is close to completion. The already well-established direction in the studies on the recognition of finite groups by spectrum is closely connected with a new promising approach for research in the recognition of finite groups by the prime graph (Grünberg – Kegel graph). It is easy to see that a group recognized by the Grünberg – Kegel graph will be recognizable by spectrum, while the converse is not true. In 2006, A.V. Zavarnitsin [4] established the recognition according to the Grünberg – Kegel graph of the C5(7) group and the Ri 2C5(4) groups. It is also worth noting that Thompson's hypothesis, which was confirmed in the work of I.B. Gorshkov [5,6] essentially claims that a finite simple non-Abelian group can be recognized by the orders of the classes of conjugate ele-ments. The common to these areas is the idea of unambiguously determining a finite simple group by some of its arithmetic parameters.

Obviously, the concept of the group growth function discussed above is essentially arithmetic. Concerning the said above, it seems interesting to look at the concept of the growth function from the point of view of finite simple non-Abelian groups and the recognition problem of finite simple non-Abelian groups by their arithmetic para-meters. As an arithmetic property, which allows us to determine the finite group up to isomorphism, we use the concepts of an independent generator system and the group density functions introduced below.

2. Methods

In this work, we use standard methods for studying finite simple non-Abelian groups (local analysis of the structure of an Abelian group, analysis of a subgroup structure [7]) as well as methods for processing data using computer calculations. The choice of the research methods is determined by the nature of the object and the items to study. In this case, the object of the study is the finite simple non-Abelian groups PSL(3,4) and A(8), and the item is the growth and density functions of the finite sim-ple non-Abelian groups PSL(3,4), A8 and the sporadic Mathieu group M11. As the methods for local structure analysis of a finite simple non-Abelian group, a permutation representation of the studied groups is used. The methods of computer calculations are the following: data storage of the permutations in computer memory, as well as the GAP computer algebra system for determining generating pairs of a group.

3. Results

Let F(n) be a function of a natural argument, F(n) can take only integer positive values. Let n be the length of an element of a group in a given system of generators. If n is supplied to the input of the function F(n), and the function outputs the number of elements representable as an irreducible product of n generators of a given group, then we call F(n) density function of the group.

Further we will consider only density functions constructed for systems of genera-ors of two elements.

Let G be a finite simple non-Abelian group of fixed order. We say that G is recognizable by the set of density functions in the class of finite simple non-Abelian groups if there is no finite simple non-Abelian group of the same order with identical density functions.

The main result of this work is the following theorem:

Theorem. Groups A8 and L(3,4) are recognizable by their set of density functions in class of all finite simple groups.

Moreover, all density functions for finite simple groups A8, M11, L(3,4) were build.

IOP Conf. Series: Materials Science and Engineering **919** (2020) 052047 doi:10.1088/1757-899X/919/5/052047

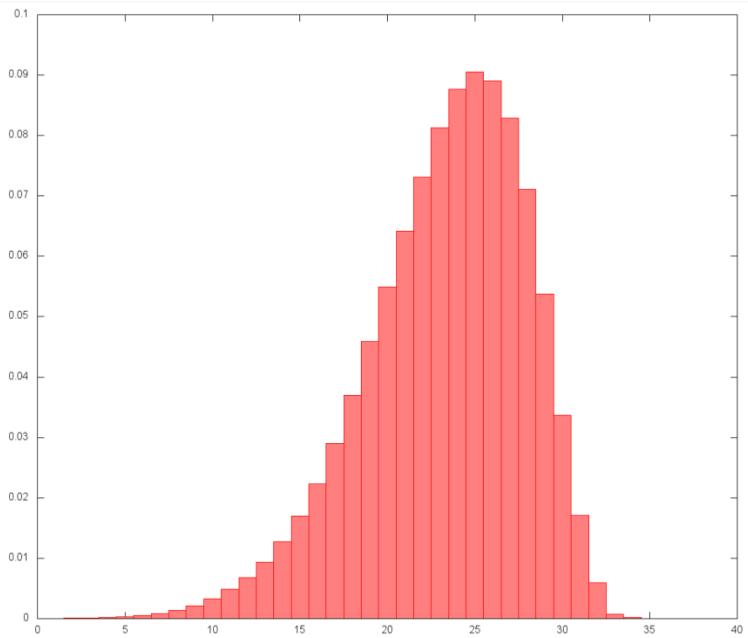


Figure 1. Example of density function of A8.

4. Discussion

The proof of this theorem is obtained via computer calculations. Below we give a brief description of the algorithm used.

4.1. Search for all possible density functions of the A8 group

Using the permutation representation of the group, we construct an array consisting of all kinds of pairs of elements. Next, using the GAP computer algebra system, we check which of these pairs will generate the entire group. Having received the set of generator pairs, we construct the density function for each pair of generators.

4.2. Data format

The words are stored as unsigned 128-bit integers. The existing processor does not support this type of data in hardware, however, the compiler provides an effective software implementation and allows one to perform operations on 128-bit numbers just as on numbers of lower bit depth.

The permutations are stored as unsigned 64-bit integers. Each permutation element is stored in nibble. Thus, each element can take 16 different values from 0 to 15, and 16 substitution elements can be saved in the entire 64-bit number. For large values of n, the storage is quite efficient, and the comparison and assignment operations are supported by the processor in hardware and are performed in 1 clock cycle. Moreover, on such representations of permutations, an order relation is introduced, the comparison operation "<" which is also performed by the processor in hardware. Any permutation will be represented by a nonzero number, so zero can be used to indicate uninitialized variables or to indicate errors. For reasons of minimizing the number of bitwise operations, words and substitutions are stored in reverse order.

4.3. Search for all possible density functions of PSL(3,4)

To find the density functions for the PSL(3,4) group, its representation in the form of permutations of length 21 was used. Such permutations cannot be stored as 64-bit integers, as was in the case of A8. Therefore, to find the density functions of PSL(3,4), a completely new program was written using a tree-like data structure on the set of permutations. The depth of such a tree is exactly equal to the length of the permutation, and the number of descendants of the tree node is less than or equal to the length of the permutation. Thus, each permutation uniquely determines the path to a tree leaf. To perform Operation 1, it is enough to store one bit in each leaf of the tree, indicating whether the path to this leaf was previously traveled.

In practice, the representation of the group in the form of permutations is available before the start of the calculations, so it is possible to construct the indicated tree in memory, eliminating the cost of allocating dynamic memory when calculating each new density function. After calculating the density function (i.e. enumerating all the permutations that make up the group), all leaves of the tree turn out to be in the same state. Therefore, before calculating the next density function, instead of zeroing the tree leaves in $O(n)$ time, where n is the order of the group, you can invert the pro-gram logic in constant time.

4.4. Comparison of density functions for PSL(3,4) and A8

We compare pairwise density functions one by one for the A8 and PSL(3,4) groups. As a result, we do not find coinciding density functions, which proves the theorem. The table below shows the statistics of calculations. The last row of the table shows the possible frequencies of appearance of the same density function (it is impossible to indicate the frequencies of all the distributions within the table). Also note that all sorts of density functions of the sporadic Mathieu group M11 (in all generator sys-tems consisting of two elements) were constructed, since it has the smallest order among sporadic groups.

5. Conclusion

The result of this work is a developed and implemented algorithm for storing permutation group elements using AVL trees. This algorithm ensures the rational use of computer memory and provides a faster search for permutation group elements in comparison with the standard representation methods, which is confirmed both theoretically (link) and practically in our work. In our case, the search time also turns out to be logarithmic (depending on the number of objects). Having large data sets this gives a significant gain, since without using the AVL tree the search speed will be linear [8]. With large data sets this gives a significant gain, since without using the AVL tree the search speed will be linear. Moreover, in this paper a new prospective view of the recognition problem of finite simple non-Abelian groups by their arithmetic parameters was proposed. The density functions of finite simple non-Abelian groups with minimal orders are constructed. For groups A8 and PSL (3,4), their recognizability by the density function was established. The algorithms for effective storage of group elements in computer memory are developed. In fact, the results of this work open up a new direction in the study of the problem of recognizability of finite simple non-Abelian groups, moreover, the obtained methods for the representation of permutation group elements in computer memory can be used in other sections of group theory and computational group theory, as well as in the applications. With the development of this direction, it is planned in the future to obtain similar results for finite simple non-Abelian groups of higher orders.

Acknowledgments

The results of this work were obtained using a complex of high-performance computing at the Siberian Federal University. The research was supported by the Russian Science Foundation (project no. 18-71-10007).

References

[1] Buturlakin A A and Vasil'ev A V 2019 The graph of atomic divisors and recognition of finite simple groups *Journal of Algebra* **537** 478-502 doi:10.1016/j.jalgebra.2019.07.013

[2] Grechkoseeva M A 2017 On orders of elements of finite almost simple groups with linear or unitary socle *J. Group Theory. Ahead of Print* doi: 10.1515/jgth-2017-0010

[3] Grechkoseeva M A 2017 On orders of elements of finite almost simple groups with linear or unitary socle *J. Group Theory. Ahead of Print* doi: 10.1515/jgth-2017-0010
[4] Zavarnitsine A V 2006 Recognition of finite groups by the prime graph *Algebra and Logic* **45**
[5] Gorshkov I B 2019 On Thompson's conjecture for finite simple groups *Communications in Algebra* doi: 10.1080/00927872.2019.1612424
[6] Gorshkov I B and Maslova N V 2018 Finite almost simple groups whose Grundberg-Kegel graphs coincide with Gruenberg-Kegel graphs of solvable groups *Algebra and Logic* **57(2)** doi: 10.1007/s10469-018-9484-7Stth
[7] Conway J H, Curtis R T, Norton S P, Parker R A and Wilson R A 1985 *Atlas of finite groups* (Oxford : Clarendon Press)
[8] Cormen T H, Leizerson C E, Rivest R L and Stein C 2010 *Introduction to algorithms* (Massachusetts: The MIT Press)