Rapport d'activité
pour la session de automne 2016
(mi-vague, 5 dernieres semestres)

Nom : SHEN Alexander (ШЕНЬ Александр)
Grade : DR2
Numero d'agent : 00024457
Section : 6
Délégation : Languedoc-Roussillon
Département scientifique : Institut des sciences informatiques et de leurs interactions
Délgation régionale : Languedoc-Roussillon

Intitulé de l'unité : Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM)
Code unité : UMR5506
Directeur : Philippe POIGNET
Adresse électronique du directeur : `Philippe.Poignet@lirmm.fr`

Adresse :

CC 477 - 161 Rue Ada
34095 MONTPELLIER CEDEX 5
France

email : `alexander.shen@lirmm.fr`
Homepage : `http://www.lirmm.fr/~ashen`.

A.1 – CIRRICULUM VITÆ

Date et lieu de naissance : 31/12/1958, Moscou, Russie
Nationalité : Russe
Titre : Doctorat (Ph. D.) de 1985, L'Université d'etat de Moscou Lomonossov. Thèse: "Algorithmic versions of entropy" (sous la direction de V.A. Uspensky).
Formation : L'Université d'etat de Moscou Lomonossov, faculté de la mècanique et des mathématiques:

- préparant de la thèse: 1979–1982
- etudiant: 1974–1979

Cursus :

- directeur de recherche (DR2) du CNRS, France:
  - de 09/2011: LIRMM (UMR 5506), Montpellier
  - 10/2005–08/2011: LIF de Marseille (UMR 6166)
- 1982–2005: chercheur de l'Institute des problemes de la transmission de l'information (Moscou); en congé depuis 2005
- Enseignement:
  - 91th Moscow Math School (1978–1980);
  - 57th Moscow Math School (1982–1986, with V.A. Ginzburg, L.S. Levitov, 1988–2008, with A. Vaintrob, M. Finkelberg, A. Romashchenko, V. Dotsenko, I. Vyugin a.o.)
  - depuis 1979: undergraduate seminar at the Moscow State University (with V.A. Uspensky, A.L. Semenov, N.K. Vereshchagin)
  - depuis 1982: Kolmogorov seminar at the Moscow State University (started by A.N. Kolmogorov, with A.L. Semenov, N.K. Vereshchagin, A. Romashchenko)
  - 1991, printemps: Instructor of Applied Mathematics (MIT)
- Lecturer at Moscow State University and Independent University of Moscow
- 1985–1987: a member of a team responsible for the Computer Literacy course in the USSR high school (under A.P. Ershov guidance)

- Visiting scientist at LABRI (Bordeaux), Bonn University (several times), Rutgers University (working with I.M. Gelfand on mathematics textbooks for high school students), Ecole Normale Superiore de Lyon (1998,1999), Universitè de Provence (2000–2005), Penn State Univerisity (2010).
- STINT invited professor at Uppsala University, Sweden (2000–2002, courses on Kolmogorov Complexity and algorithmic problem solving)
- Member of the Council of the Mathematics College of Independent University of Moscow (1993–2003)
- Member of the Council of the Computer Science Deparment, National Research University High School of Economics, Moscow (2015–now).

A.2 – RECHERCHE SCIENTIFIQUE

Thèmes de recherche : complexité de Kolmogorov, théorie d'information algorithmique et methodes combinatoires, sequences aleatoires, statistique algorithmique, theorie de calculabilite.

Activité de recherche

The definition of Kolmogorov complexity of a string as the minimal length of a program that produces this string looks like something trivial, and it is hard to believe at first that a serious mathematical theory can arise out of it. Still it is the case, and this theory continues to find unexpected connections with different fields of mathematics and computer science. Some of these connections were investigated. Before describing the contents of the corresponding paper, let us recall the basic definitions, just to fix the notation. (A short introduction is published as [AC2], a textbook exposition can be found in [T6].)

To show the definition from a slightly different angle, let us rephrase it as follows. Let $D(x, y)$ be a binary relation in the set of binary strings; here we call these relations *description modes*. If $D(x, y)$, we say that x is a *description* of y (for the mode D). For a given D we define the complexity of y as the minimal length of its description, i.e.,

$$C_D(y) = \min\{l(x) \,:\, D(x, y)\}$$

Here $l(x)$ stands for the length of the string x. Of course, the definition looks stupid: if $D(x, y)$ is true for all x and y, then everything is a description of everything, and complexity of all strings is 0. To make it reasonable, we should

impose some restrictions on D. For plain Kolmogorov complexity, we consider the class the descriptions that are graphs of functions, i.e.,

$$D(x, y) \wedge D(x, z) \Rightarrow y = z.$$

There is another version of complexity, for which the restriction in more strict (prefix-stability):

$$(x \text{ is a prefix of } x') \wedge D(x, y) \wedge D(x', z) \Rightarrow y = z.$$

We say that description mode D is *better* (=*not worse*) than another description mode D' if

$$C_D(x) \leq C_{D'}(x) + c$$

for some c and all x. Kolmogorov–Solomonov theorem says that among all enumerable description modes there exists an optimal one (better than any other). We fix some optimal mode D, denote C_D by C and call it *plain complexity* function. Similar procedure for the class of prefix-stable enumerable description modes gives us *prefix complexity* function K.

Generic algorithms

A classical undecidable problem in the computation theory is the halting problem: given some program p, find out whether it terminates or not. (We consider programs without inputs for simplicity.) There is no algorithm that solves this problem, i.e., always produces the right answer. What if we require that the algorithm solves this problem on most inputs, not always? Is it possible, for example, that for some algorithm the fraction of n-bit inputs where the answer is incorrect or the computation diverges, tends to 0 as $n \to \infty$? This question was asked and negative answer were obtained in several different settings: such an approximately correct algorithm (the names "generic algorithm" and "coarse algorithm" are used for different special cases of this notion) does not exist. It turns out, as shown in [AR1], that the most technically simple and "conceptual" proof of these results (and some their generalizations) uses Kolmogorov complexity.

Consider *the number of terminating programs of length at most n*. It is an $(n+1)$-bit number H_n. The crucial observation: this number is maximally complex:

$$C(H_n) = n + O(1).$$

The proof sketch: if H_n is simple, then, knowing it, we can find all the terminating programs; from that list one can obtain the string of complexity greater than n. (Here we assume that programming language is used as a description mode: each program p is a description of its output.) So the amount of information (complexity) in H_n cannot be significantly less than n.

Now we explain why the high complexity guarantees the non-existence of an approximation algorithm with few errors. For simplicity we consider the case of total algorithms. Then this almost-everywhere correct algorithm can be used to find a good approximation for H_n (if we use to decide whether a program terminates or not). And then H_n cannot have high complexity, since to specify H_n it is enough to specify its deviation from the approximation.

This is the simplest version of the statement, but the same argument can be used to prove other results (see [AR1] for details). Let us mention one of them: *the sequence $\#H_n/2^n$ has no limit and every limit point is random in the sense of algorithmic randomness theory*. Here we see how the notions from algorithmic randomness (the notion of random point can be defined in terms of prefix complexity) appear in the problem where the statement has nothing to do with complexity or randomness.

Logic

The first (and very impressive) connection between logic and Kolmogorov complexity theory was the famous Chaitin's version of Gödel incompleteness theorem. Gödel theorem says that there exist true non-provable statements; Chaitin's version of this theorem gives a specific class of statements; most of them are true but not provable. These are statements of the form

$$C(x) > n,$$

where x is some specific string (constant, literal) where n is some specific number. Chaitin has proved by a nice simple argument that these statements can be provable only for n that do not exceed some constant (depending on the theory and on the choice of complexity function). In other words, if n is sufficiently large, then this statement may be true or false but never would be provable. On the other hand, for each n only $O(2^n)$ objects could have complexity at most n, so for almost all x such a statement is true.

In [AR4] we perform a more detailed classification and start with the following question. Chaitin tells us that most statements of the form $C(x) > n$ are not provable. So we can add such a statement, randomly chosen, as an axiom and with high probability get a theory where more statements are provable while still all provable statements are true. Note that we get different theories by adding different statements (for some fixed value of n and different values of x). How these theories are related? It turns out that there are some special values of x that make this theory maximally strong (technically, it proves all the true statements about complexities up to $n - O(1)$), but for most values of x this does not happen. There are many questions than can be asked about these theories. For example, they could look as some practical way to do mathematics if the existing axioms

are not enough: toss a coin million times, and you can be practically sure that the resulting bit string x has complexity at least 900 000. So we get a new axiom that is safe from practical viewpoint; may be, in this way we can prove some theorems that are not provable in the original theory? We show that the answer is no; more precisely, for a given conjecture that is not provable without additional axioms, the chances for it to become provable are about the same as the chances for the theory to become inconsistent. However, this approach could shorten the proofs, unless some popular assumption from complexity theory is false (unless PSPACE=NP).

Foundations of statistics

One of the main motivations for the Kolmogorov complexity theory is the foundation of statistics. There is a basic philosophical question here: why, seeing the sequence

$$x = 01010101010101010101$$

we reject the statistical model of fair coin tossing while seeing some other sequences of the same length we accept it. The answer suggested by the complexity theory is that this x has a simple description (has small complexity) while most of the strings of the same length do not (they have complexity close to length).

As the article [PE1] (requested by the editors) shows, there is some significant interest to this approach even in the philosophy community. But from the mathematical point of view we need to go farther and explain what is a good statistical model (probability distribution) for an experimental data string x. This is a topic of algorithmic statistics. A short survey without proof is given in [AC3] while a much more detailed account with proofs can be found in [TR1]. For the historical account see also the last chapter of [T6]. There are several approaches to measure the "quality" of a statistical models; these ideas go back to Kolmogorov who considered these definition in 1970s (but published almost nothing). One of them is based on the *randomness deficiency*: a finite set A containing some string x has randomness deficiency

$$\log \#A - C(x|A)$$

that measures the quality (the smaller the deficiency, the better the model). Here we use the conditional complexity where A is a parameter of a description mode. The other is based on *two-part descriptions*. Each element $x \in A$ has a two-part description: first we describe A using $C(A)$ bits, and then we specify the ordinal number of x in A using $\log \#A$ bits. The *optimality deficiency* shows how much longer is this two-part description compared to the optimal description for x that has length $C(x)$. In other words, we consider the difference

$$C(A) + \log \#A - C(x).$$

These two approaches are not equivalent (the optimality deficiency could be much larger than the randomness deficiency). However, now classical result of Vereshchagin–Vitanyi (2002) shows that if we consider the trade-off between the complexity and the quality of the description, then we get the same (with logarithmic precision) curves. In [TR1] it is explained in details (including another result of Vereshchagin–Vitanyi from 2008 for the restricted class of models), as well as the description of these curves in terms of ordinal number in the enumeration of objects of bounded complexity. Also it is shown how all this is related to time-bounded complexity if we use the busy beaver function as a time bound (a result of Bauwens).

There is one more question related to good statistical models. Generally, they may exist or not — but assuming they exist, how can we find them? Is there some learning algorithm that reads a sequence and gets a statistical model? This question is studied in [AC5] in the framework of infinite sequences and computable measures on the space of infinite sequence. It is shown, that (contrary to some claims) such an algorithm does not exist even for reasonably restricted classes of measures (for example, the class of all Bernoulli measures).

Probability theory

Not only statistics but also probability theory is closely related to complexity. One may define a random sequence ω with respect to some computable measure P on the Cantor space requiring that every prefix x has prefix complexity at least $-\log P(x) - O(1)$, where $P(x)$ is the probability of the event "P-distributed random variable starts with x".

There are two technical reports [TR2 TR3] that deal with this notion. In [TR 2] we consider the question of conditional probabilities. For the discrete case (finitely many values) the definition of conditional probability is straightforward: $P(A|B) = P(A \cap B)/P(B)$. However, for the case when $P(B) = 0$ (for example, when B is a singleton $\{\beta\}$) this definition does not work. It was a big step in probability theory when people understood that one can use Radon–Nikodym derivative to define this conditional probability *for almost all* β. The algorithmic randomness theory notes that this can be done *for algorithmically random* β. This was noted by Takahashi who also proved some natural results about pair randomness (but proofs were not clear). In [TR2] cleaner proofs are given, and also an example showing that the general result cannot be proven without additional assumptions.

In [TR3] another question is studied. Assume that we have a box that combines a random bits generator that generates some sequence α with output distribution P (on the Cantor space), and some algorithm T that transforms the output bits. Then the sequence $\beta = T(\alpha)$ is distributed according to the image

distribution $T(P)$. Assume that somebody claims that such an experimental device produced a sequence β. When this claim is plausible? Answer 1: when β is random with respect to the image distribution. Answer 2: when $\beta = T(\alpha)$ for some α that is random with respect to P. There is an old folklore result that says that these two answers are equivalent if T is a computable mapping defined almost everywhere. We generalize this claim to its natural generality by using the notion of *layerwise computable mapping* introduced by Hoyrup and Rojas.

Topological arguments

Let me also briefly mention a short paper [RI2] when simple topological arguments were used to prove the existence of strings with given complexities. It is not the statement that is interesting here (though it improves some older result by Muchnik and Vyugin who used a complicated game proof), but the tool: Sperner's lemma.

Games and expository writings

Here we speak not about general game theory but about using game approach to prove results about Kolmogorov complexity. In [RI3] this approach is used to provide simple proofs for two results: an improvement of an old result of Gács saying that $C(C(x)|x)$ can be maximal ($\log n + O(1)$) for strings of length n (Gács has a slightly weaker statement with a rather involved proof), and Miller's result about strings that have maximal *plain* complexity among n-bit strings, but not maximal *prefix* complexity among them.

Also we wanted to give a simplified exposition of results that are available in different sources with often incompatible definition. The main project of this type was the textbook [T6] that finally is translated to English by the authors and accepted for publication by American Mathematical Society.

However, there are more special topics in Kolmogorov complexity theory that also miss an easy-to-read clean exposition. Of course, this is a matter of taste but we tried hard to provide something readable. In [AC2] we present a rather recent but classical result (equivalence of the notions of K-trivial, K-low and MLR-low sequence) that usually is proven by some argument called 'decanter method'. It is not formally defined what this method is, this is just a name, but we tried to present the proof in a game form that somehow does not have this 'decanter' flavour (but it is essentially the same proof, just simplified).

A short tutorial for beginners was also published as [AC2]

La place de votre recherche dans celle de votre unité

I am a member of the ESCAPE team of LIRMM; randomness and computability theory are central research topics of this team, and it was strengthened when Laurent Bienvenu has arrived. As it is clear from the publication list, I am collaborating with my LIRMM colleagues, Laurent Bienvenu and Andrei Romashchenko, with our former visitors (Bruno Bauwens, N. Vereshchagin). Damien Desfontains was LIRMM stagiere. Together with A. Romashenko we were directing the summer internship for three students (two from ENS Lyon and one from ENS Cachan).

We obtained ANR RaCAF grant; the grant team consists mainly from ESCAPE researchers. Some my visits was also supported by the French–Russian cooperation program (and visits to Moscow were supported by HSE).

Mobilités

I was regularly visiting Kolmogorov seminar, Computer Science Department of HSE, Poncelet laboratory in Moscou, and maintain working relations with students there. Paper [RI1] was partly written during the visit to Moscow supported by the HSE. In 2015 I visited the randomness program at Singapore IMS, and in 2016 I participated in the program in IHP, including a conference at CIRM, and Oleron program on tilings where I gave some introductory talk. I was also an invited speaker in 2016 (Leicester University, UK, and EQINOCS conference in IRIF, Paris).

Liste de publications scientifiques 2014–2016

Publication dans une revue internationale [RI]

1.

2. ,

3.

4.

Publications dans les actes d'une conférence [AC]

([AC2] and [AC3] are in a Festschrift collection published by Springer;
it is not a conference proceedings but reviewed in a similar way;
[AC4] is a preliminary version of [RI1])

1.

2.

3.

4.

5.

Arxiv et Technical Reports [TR]

Here all the technical reports are listed, including preliminary versions of papers published: TR4 (AR1), TR5 (AC2), TR6 (AC3),

1.

2.

3.

4.

5.

6.

Textbooks [T]

(Books 1-5 are already sent to the printer, but are labeled as 2017 for marketing reasons. Book 6 is accepted for publication by AMS, but not published yet.)

1.

2.

3.

4.

5.

6. ?

Popular expositions etc. [PE]

1.

2.

3.

4.

5.

A.3 – ENSEIGNEMENT, FORMATION ET DIFFUSION DE LA CULTURE SCIENTIFIQUE

Enseignement : A student from Moscow, Mikhail Andreev, whom I co-advised, is now finishing his PhD thesis in Moscow (his main official advisor is N. Vereshchagin). His paper (submitted to CiE2016 conference) about the comparison of fast growing "busy-beaver-type" functions based on the plain and prefix complexities, got a best student paper award at this conference. (Disclaimer: I was in the program committee, but was not involved in this decision due to the conflict of interests). Another student in Moscow who visited LIRMM to work with me, got his paper accepted by STACS2016. In summer 2016, a stagiere from Lyon, Guilhem Marion, worked under my supervision.

Organisation : I served as a program committee member for CiE2016 and for several workshops (2015, 2016). As a member of the Council of the Computer Science Department of the National Research University High School of Economics I am responsible for the curriculum on discrete mathematics.

In 2016 a new ANR RaCAF project started where, I hope, the group of researchers interested in randomness and participating in NAFIT grant will continue their collaboration.

Many of these topics are covered in the book on Kolmogorov complexity and algorithmic randomness that was published in Russian version. Currently an updated English edition is in preparation. The first part of it is a systematic exposition of basic results of the field (which, I believe, are not explained well in the existing texts); the second part is the exposition of some new results where our group was involved.

Vulgarisation : I prepared an online course (Basic Tools of Theoretical Computer Science, in Russian) with video lectures, test exercises and problems, while visiting Saint-Petersbourg (Russia) Computer Science center. It was available at stepic.org; several dozen people earned the certificate (while much more, up to several thousands, started trying it; it was freely available for all). I wrote an encyclopedia article [PE1] for Routledge handbook of philosophy of information.

In 2015 and 2016 I gave some popular talks for high school students visiting Moscow. New editions of textbooks [T1–T5] were prepared. A textbook about high school geometry (theory presented as a sequence of problems and their solutions) was translated by the American Mathematical Society and published in 2016 [PE2]. The new editions of some popular expositions [PE3–PE6] were also prepared. Recently I was contacted about the Polish translation of the book [T2], the translation is finished but the publication is not agreed yet.

B – OBJECTIFS

There are several lines of research that I hope to follow:

1. In algorithmic statistics, the main question is whether it has some more practical (bounded resources? restricted models?) version. This question look very difficult (first of all, because of the unsolved problems in computational complexity like NP=P), but still something could be done. Also it would be interesting to find whether there is some connection between logical results mentioned above, and algorithmic statistics. The two specific problems that seem interesting here are: (1) which x lead to powerful axioms and which are not? (2) Imagine we have a weaker bound $C(x) > m$ where m is not the full complexity (say, $C(x)/2$). This axiom is also non-provable, but what is its power?

2. The relation between theoretical results about pseudorandomness and practical attempts to get pseudorandom sequences starting from weakly random sources.

3. The Kolmogorov complexity theory is two-faced: first, it is a part of very abstract recursion theory, second, it is closely related to combinatoric results, and an interesting interplay happens here. For example, it seems that some types of isoperimetric inequalities for Boolean cube could be used to settle an open question about increasing Hausdorff dimension of a sequence by changing a small fraction of bits.