Transactions of the Institute for Informatics and Automation Problems of NAS RA.
Collected reports of participants of JAF-23.

DIOPHANTINE FLAVOR OF KOLMOGOROV COMPLEXITY

Yuri Matiyasevich

The paper consists of three parts. In the first part a short introduction is given to some main notions and results in the theory of Kolmogorov complexity. The second part is a similar introduction to some main results in the theory of Diophantine computations. The third part is devoted to the interplay between Kolmogorov complexity and Diophantine equations, in particular, a new Diophantine definition of Chaitin's Ω is given.

§1. SOME DEFINITIONS IN THE THEORY OF KOLMOGOROV COMPLEXITY

The notion nowadays widely known as *Kolmogorov complexity* was in fact first introduced by Ray J. Solomonoff [8] and then by Andrei N. Kolmogorov [4]. This Section gives only information required for reading Section 3, more complete information, including bibliographical data, can be found, for example, in [5].

The main underlying motivation of Kolmogorov (or descriptional) complexity is the desire to measure complexity of individual combinatorial objects. Without loss of generality, we can deal only with words in a fixed two-letter alphabet, say, $B = \{0, 1\}$. Informally, a word in this alphabet is "simple" if has has a short description. In this sense the 1-million-letter word described as

$$\textit{The first (after the point) 1000000 binary digits of the number } \pi \qquad (1)$$

is "simple".

Here there is some kind of cheating. On the one hand, the short word (1) is from an alphabet which has much more than two letters; we can agree to use only a binary alphabet for descriptions. On the other hand, a lot of knowledge is hidden behind (1). In fact, in order to be able to write down the required word with one million letters, one has

to know what the number π is and how its binary digits can be calculated. To take into account this hidden knowledge in the most general setting, one can introduce the notion of a *description mode*.

Formally, a description mode is just a binary relation D between words, i.e., $D \subseteq B^* \times B^*$. A word A is called a *description* of a word W (according to given description mode D) if $\langle A, W \rangle \in D$.

This terminology is used in papers on Kolmogorov complexity, but other names are used in different situations. One can say that D is a *method of archiving* or a *method of compressing* and respectively say that A is an *archive* of W or that A is a *compressed form* of W. All this terminology is used alternatively in this paper.

We shall always presuppose the following properties of any description mode D:

(i) Every word has at least one description.

(ii) $\langle A, W' \rangle \in D \,\&\, \langle A, W'' \rangle \in D \Rightarrow W' = W''$.

(iii) D is recursively enumerable .

These three conditions are quite natural. Condition (i) expresses the universality of compression method so we can put everything into an archived form. Condition (ii) guaranties the uniqueness of the decompression so we don't loose information during archiving. At last, condition (iii) implies that both compression and decompression can be performed on a computer.

Clearly, any relation having properties (I)–(III) defines a computable function Δ (decompression) from the set $\mathcal{P}_D = \{A : \exists W \, \langle A, W \rangle \in D\} \subseteq B^*$ onto the set B^*.

A possible additional condition of *prefix-free* description mode can be imposed:

(iv) The set $\mathcal{P}_D = \{A : \exists W \, \langle A, W \rangle \in D\}$ is prefix-free, i.e., no word from $\mathcal{P}_D$ is a prefix of another word from $\mathcal{P}_D$.

This condition looks less natural than conditions (i)–(iii). In fact, it inevitably results in less efficient compression because not every word can be used as archive. One justification of the use of this additional condition is as follows. We can imagine that A is a computer program which prints the word W (using no input). In typical programming languages all program starts and finishes by a kind of coupled **begin** and **end**, thus one

program cannot be proper beginning of another program and hence the condition (iv) is automatically fulfilled. But the real justification of the use of condition (iv) is the fact that one can prove more interesting theorems about prefix-free description modes than about general description modes.

The *complexity* $K_D(W)$ of a word W *with respect to a given description mode* D is defined as the length of the shortest compressed form of W:

$$K_D(W) = \min\{|A| : \langle A, W\rangle \in D\}$$

where $|A|$ is the length of the word A.

In this definition the complexity of a word depends not only on it but also on a description mode which is in a sense irrelevant to the word itself. We could try to use "the best" description mode. Formally, a (prefix-free) description mode D is called *optimal* if for every (prefix-free) description mode D' there is a number c such that for every word W

$$K_D(W) \leq K_{D'}(W) + c. \tag{2}$$

The first question to answer is: *does an optimal (prefix-free) description mode exist?* The positive answer to this question forms the basis for the whole theory.

Theorem (A.N.Kolmogorov[4]–R.J.Solomonoff[8]). There are optimal description modes.

Proof: Enumerate all description modes

$$\Delta_0, \Delta_1, \ldots, \Delta_m, \ldots$$

(viewed as partial functions from B^* into B^*) and define

$$D_{\text{opt}} = \{\langle 1^m 0 A, \Delta_m(A)\rangle : A \in B^* \,\&\, m \in \mathbb{N}\},$$

□

Optimal prefix-free description modes also exit, the proof (based on the same idea) is technically a bit more complicated.

Now we fix a particular optimal (prefix-free) description mode D_{opt} and define *Kolmogorov (prefix-free) entropy* of a word W as the complexity of the word W with respect to D_{opt} (the word "entropy" is used to emphasize that the complexity is measured with respect to fixed optimal description mode).

So now the complexity of a word is determined by the word itself but up to additive constant only. Thus any number can be the entropy of a given word, and meaningful results about Kolmogorov entropy should involve either infinitely many words or *infinite words*, the complexity of which can be characterized by the growth of (prefix-free) entropy of their initial finite fragments.

If an infinite word is decidable (i.e., the set of positions of "1" is a decidable set), then the entropy of its initial fragment of length l grows essentially as $\log(l)$. For example, the verbal description of a 1-milliard-letter word

The first (after the point) 1000000000 binary digits of the number π (3)

is longer than the description of a 1-million-letter word (1) by 3 letters only.

The condition of the decidability can be weakened. If an infinite word is recursively enumerable (i.e., the set of positions of "1" is recursively enumerable), then the entropy of its initial fragment of length l still grows still essentially as $O(\log(l))$. Namely, a description of such a fragment can consists of the couple $\langle m_0, m_1\rangle$ where m_0 and m_1 are the numbers of "0" and "1" respectively ($m_0 + m_1 = l$).

The entropy of any word cannot be essentially greater than its length (because we can use the word itself as its archived form; in the case of prefixed-free entropy additional $\log(l)$ bits would be sufficient for the property (iv)).

If the entropy of initial fragments of length l of an infinite word W grows as l (up to an additive constant), then this word W can be considered as completely random sequence of bits. Gregory J. Chaitin [2] introduced a particular method of constructing an infinite word of so big prefix-free complexity. This word can be defined as the sequence of binary digits of some real number Ω. This number can be interpreted as the probability that a randomly selected Turing machine stops (for a suitable probability distribution on on Turning machines). Of course, digits of this number do not form a recursively enumerable set. However, there is an effectively computable increasing sequence $\Omega_1, \Omega_2, \ldots, \ldots$ of rational numbers which converges to Ω.

§2. MAIN RESULTS IN DIOPHANTINE COMPUTATIONS

A *Diophantine equation* is an equation of the form

$$P(x_1, \ldots, x_m) = 0, \tag{4}$$

where P is a polynomial with integer coefficients and *the unknowns* $x_1, \ldots, x_m$ are supposed to be natural numbers.

A *family* of Diophantine equations is an equation of the form

$$P(a_1; \ldots, a_n, x_1, \ldots, x_m) = 0,$$

where P is a polynomial with integer coefficients, the variables of which are split into two groups:

- the *parameters* $a_1, \ldots, a_n$
- the *unknowns* $x_1, \ldots, x_m$

Consider the set $\mathfrak{M}$ such that

$$\langle a_1, \ldots, a_n \rangle \in \mathfrak{M} \Longleftrightarrow \exists x_1 \ldots x_m \{P(a_1, \ldots, a_n, x_1, \ldots, x_m) = 0\}. \tag{5}$$

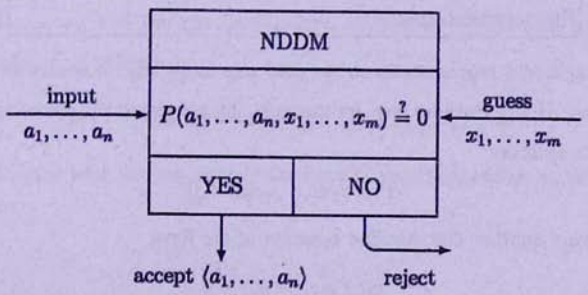
Sets which can be defined in this way are called *Diophantine*, the above equivalence being called *Diophantine representation* of the set $\mathfrak{M}$. We have:

Trivial fact. Every Diophantine set is recursively enumerable.

Main result (DPRM-theorem). Every recursively enumerable set is Diophantine.

The latter result is known as *DPRM-theorem* after Martin Davis, Hilary Putnam, Julia Robinson, and the present author; full proof of the DPRM-theorem can be found nowadays in many books, for example, in [6].

Leonard Adleman and Kenneth Manders [1] restated the DPRM-theorem by introducing the notion of **Non-Deterministic Diophantine Machine** (NDDM for short). A NDDM is specified by a parametric Diophantine equation () and works as follows: on input $a_1, \ldots, a_n$ it guesses the numbers $x_1, \ldots, x_m$ and checks (); if the equality holds, the n-tuple $\langle a_1, \ldots, a_n \rangle$ is accepted.



DPRM-theorem says that NDDMs are as powerful as, say, non-deterministic Turing machines, that is every set recognizable by a Turing machine is recognized by some NDDM, and, of course, vise versa.

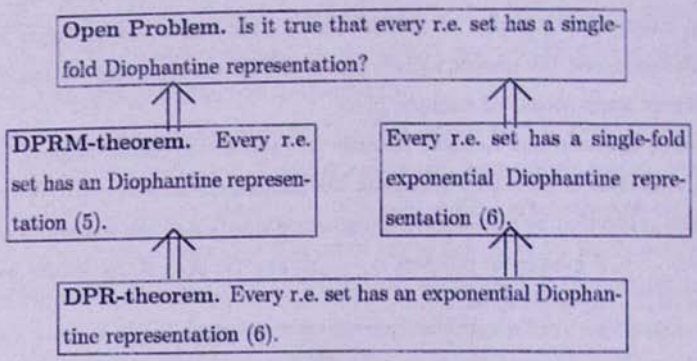
DPRM-theorem was obtained as an improvement of an earlier DPR-theorem [3] stating that every r.e. set has an *exponential Diophantine representation*

$$(a_1, \ldots, a_n) \in \mathfrak{M} \Longleftrightarrow \exists x_1 \ldots x_m \{E(a_1, \ldots, a_n, x_1, \ldots, x_m) = 0\}$$

where E is an *exponential polynomial*, i.e, an expression constructed with addition, subtraction, multiplication and exponentiation.

Besides DPRM-theorem, DPR-theorem was improved in another directions, and this improvements turned out to be vital for application in Kolmogorov complexity. Namely, we can demand that the representation is *single-fold* which means that the x's, if they exist, are unique for every fixed values of the a's (a proof of this improvement of the DPR-theorem can be found, for example, in [6]).

The question whether these two improvements of DPR-theorem could be combined remains open.



Besides single-fold representations, we need one more way to specialize Diophantine representations. Hilary Putnam (see, for example, [6]) suggested the following trick: *Given a Diophantine equation*

$$T(p, q, x_1, \ldots, x_m) = 0$$

we can construct another Diophantine equation of the form

$$S(p, y_1, \ldots, y_n) = q$$

such that the former equation has a solution in $x_1, \ldots, x_m$ for the same values of the parameters p and q for which the latter equation has a solutions in $y_1, \ldots, y_n$.

This trick works provided that the parameter q assumes only non-negative integer values.

§3. DIOPHANTINE FLAVOR OF KOLMOGOROV COMPLEXITY

Without loss of generality, we can replace the two-letter alphabet $B = \{0,1\}$ by another two-letter alphabet $C = \{1,2\}$. There is a natural one-to-one correspondence between the words from C^* and natural numbers: we can treat words in the alphabet C as numbers written in positional notation with standard weights of digits 1, 2, 4, 8, ..., but with non-standard digits 1 and 2, that is, number 0 is the empty word, number 1 is the word "1", number 2 is the word "2", number 3 is the word "11", number 4 is the word "12", number 5 is the word "21", and so on.

Kolmogorov complexity is machine-independent notion (in condition (iii) we demanded only that a description mode should be a r.e. relation without specifying particular tool for its enumeration). In particular, now we can use the NDDM for the definition of a r.e. relation and give

Definition P. A *Diophantine description* mode is a polynomial P such that

(i) for all w the Diophantine equation

$$P(a, w, x_1, \ldots, x_k) = 0 \tag{6}$$

has a solution in natural $a, x_1, \ldots, x_k$;

(ii) $P(a, w', x_1', \ldots, x_k') = P(a, w'', x_1'', \ldots, x_k'') = 0 \Rightarrow w' = w''$.

A number a is called *compressed form* of a number w as soon as the equation (6) has a solution in $x_1, \ldots, x_k$, and the the *Diophantine complexity* $D_P(w)$ of a number w with respect to such a P is defined by

$$D_P(w) = \min\{|a| : \exists x_1 \ldots x_k P(a, w, x_1, \ldots, x_k) = 0\}. \tag{7}$$

Using Putnam's trick, we can specify the form of the Diophantine equation (6) either as

$$Q(a, x_1, \ldots, x_k) = w \tag{8}$$

or

$$R(w, x_1, \ldots, x_k) = a \tag{9}$$

for some polynomials Q and R. Respectively, we can give two other definitions.

Definition Q. A Diophantine description mode is a polynomial Q such that

(i) for all w the Diophantine equation (8) has a solution in natural $a, x_1, \ldots, x_k$;

(ii) $Q(a, x'_1, \ldots, x'_k) \geq 0 \,\&\, Q(a, x''_1, \ldots, x''_k) \geq 0 \Rightarrow$

$$Q(a, x'_1, \ldots, x'_k) = Q(a, x''_1, \ldots, x''_k).$$

Respectively, the *Diophantine complexity* $D_Q(w)$ of a number w with respect to such a P is defined by

$$D_Q(w) = \min\{|a| : \exists x_1 \ldots x_k Q(a, x_1, \ldots, x_k) = w\}. \tag{10}$$

Definition R. A Diophantine description mode is a polynomial R such that

(i) for all w the inequality $R(w, x_1, \ldots, x_k) \geq 0$ has a solution in natural $x_1, \ldots, x_k$;

(ii) $R(w', x'_1, \ldots, x'_k) = R(w'', x''_1, \ldots, x''_k) \geq 0 \Rightarrow w' = w''$.

Respectively, the *Diophantine complexity* $D_R(w)$ of a number w with respect to such a P is defined by

$$D_R(w) = \min\{|R(w, x_1, \ldots, x_k)| : R(w, x_1, \ldots, x_k) \geq 0\}. \tag{11}$$

Definitions Q and R show that either decompression or compression can be done simply by calculation of the value of some polynomial having one explicit argument (a or w) and several "hidden arguments" $(x_1, \ldots, x_k)$, the only restriction on the latter consisting in the non-negativity of the value of the polynomial.

A Diophantine description mode T (in the sense any of the above three definitions) is *optimal* if for every Diophantine description mode T' there is a number c such that

$$D_T(w) \leq D_{T'}(w) + c.$$

Theorem. There are optimal Diophantine description modes.

Of course, this theorem immediately follows from the existence of optimal description modes and the DPRM-theorem. However, a direct proof can be given based on a *universal Diophantine equation* $U(a, w, k, y_1, \ldots, y_n) = 0$ having the following property: *for every polynomial* $P(a, w, x_1, \ldots, x_m)$ *there exist a number* k_P *such that*

$$\exists y_1 \ldots y_n U(a, w, k_P, y_1, \ldots, y_n) = 0 \Leftrightarrow \exists x_1 \ldots x_m P(a, x_1, \ldots, x_m) = 0. \qquad (12)$$

Fixing particular polynomial, we get *Diophantine entropy* $E(w)$ of a number w as its Diophantine complexity with respect to this polynomial. Since the existence of a universal Diophantine equation can be proved by purely number-theoretical methods (see [6] for such a proof) so the whole theory of Diophantine entropy can be developed in the framework of Number theory.

The above considerations were nothing else but straightforward translation of the main definitions and results in Kolmogorov complexity into the language of Number theory. The question is whether we can obtain some results more specific for Diophantine equations. This is in fact possible, and results of such spirit are given below.

Theorem 1. For every optimal Diophantine description mode $R(w, x_1, \ldots, x_k)$ there exist a polynomial $R_{pos}(w, x_0, \ldots, x_k)$ such that

$$\min\{R(w, x_1, \ldots, x_k) : R(w, x_1, \ldots, x_k) \geq 0\} = \min\{R_{pos}(w, x_0, \ldots, x_k)\} \qquad (13)$$

and hence

$$D_R(w) = \min\{|R_{pos}(w, x_0, \ldots, x_k)|\}. \qquad (14)$$

Note that condition (13) implies that polynomial $R_{pos}(w, x_0, \ldots, x_m)$ assumes only non-negative values, and every natural number is a value of this polynomial, that is, it satisfies condition (i) from Definition R. However, we cannot assert that condition (ii) is satisfied as well.

Proof: Consider the trivial Diophantine description mode $a = w$. Since R is optimal,

$$\min\{|R(w, x_1, \ldots, x_m)| : R(w, x_1, \ldots, x_m) \geq 0\} = D_R(w) \leq |w| + c \qquad (15)$$

for some constant c and hence

$$\min\{R(w, x_1, \ldots, x_m) : R(w, x_1, \ldots, x_m) \geq 0\} \leq c_1 w + c_0 \qquad (16)$$

for some positive constants c_1 and c_0. Put

$$R_{\mathrm{pos}}(w, x_0, \ldots, x_k) = (x_0 + 1)\big(1 + (c_1 w + c_0)(R(w, x_1, \ldots, x_k) - x_0)^2\big) - 1. \qquad (17)$$

If $x_0 = R(w, x_1, \ldots, x_k)$, then $R_{\mathrm{pos}}(w, x_0, \ldots, x_k) = R(w, x_1, \ldots, x_k)$ and hence

$$\min\{R(w, x_1, \ldots, x_k) : R(w, x_1, \ldots, x_k) \geq 0\} \geq \min\{R_{\mathrm{pos}}(w, x_0, \ldots, x_k)\}. \qquad (18)$$

On the other hand, if $x_0 \neq R(w, x_1, \ldots, x_k)$, then $R_{\mathrm{pos}}(w, x_0, \ldots, x_k) \geq c_1 w + c_0$ and together with (16) we get

$$\min\{R(w, x_1, \ldots, x_k) : R(w, x_1, \ldots, x_k) \leq 0\} \leq \min\{R_{\mathrm{pos}}(w, x_0, \ldots, x_k)\}. \qquad (19)$$

The two inequalities (18) and (19) imply the desired equality (15). □

It is not clear whether this theorem can be extended to non-optimal Diophantine description modes.

G. J. Chaitin gave several definitions of his unpredictable number Ω, and one of them was connected with Diophantine equations. Of course, the binary digits of this number don't produce a Diophantine set (it was indicated above that all r.e. set have very low Kolmogorov entropy), so something more involved than solvability of Diophantine equations should be used for a definition of Ω. That is why instead of the question "Has given equation a solution?" Chaitin considered more difficult question "Has given equation infinitely many solutions". Also, needing by technical reasons single-fold representations, he was forced to deal with exponential Diophantine equations rather than with ordinary Diophantine equations.

Theorem (G. J. Chaitin [2]). *There is exists an exponential Diophantine equation $C(k, z_1, \ldots, z_m) = 0$ such that the k-th digit of Ω is equal to 1 if and only if the equation has infinitely many solutions in $z_1, \ldots, z_m$.*

For a long time this remained an isolated fact showing that in Number theory there are quite chaotic objects. Recently a similar result was obtained by Toby Ord and Tien D. Kieu. They considered exponential Diophantine equations always having only finitely many solutions and asked the question "Is the number of solutions odd?".

Theorem (T. Ord and T. D. Kieu [7]). *There is exists an exponential Diophantine equation $O(k, z_1, \ldots, z_m) = K(k, z_1, \ldots, z_m)$ which for every value of k has only finitely many solutions*

in $z_1, \ldots, z_m$ and the k-th digit of Ω is equal to 1 if and only if the equation has odd number of solutions.

It turned out that this results can be generalized to the case of many other questions.

Theorem 2. Let $\mathfrak{C}$ be any infinite decidable set with infinite complement. Then we can construct an exponential Diophantine equation $M(k, z_1, \ldots, z_m) = 0$ such that for every k the equation has finitely many solutions in $z_1, \ldots, z_m$ and the k-th digit of Ω is equal to 1 if and only if the number of solutions of the equation belongs to the set $\mathfrak{C}$.

Proof: We define a computable sequences of numbers in the following way. Let $a_0 = 0$; let a_{2k+1} be least element from the set of $\mathfrak{C}$ which is greater than a_{2k}; at last, for a positive k let a_{2k} be least element from the complement of the set $\mathfrak{C}$ which is greater than a_{2k-1}.

Let $\Omega_1, \Omega_2, \ldots, \ldots$ be (constructed by Chaitin) computable increasing sequence of rational numbers which converges to Ω. Let $\omega_{m,k}$ be the k-th binary digit of Ω_m. Let $\kappa_{0,k} = 0$, $\kappa_{1,k} = \omega_{1,k}$ and $\kappa_{m,k} = \kappa_{m-1,k} + (\omega_{m-1,k} - \omega_{m,k})^2$ for $m > 1$. The sequence $\kappa_0, \kappa_1, \ldots$ is computable, is increasing, and has some limiting value κ_k (due to the convergence of the sequence $\Omega_0, \Omega_1, \ldots$). Depending on whether the k-th digit of Ω is equal to 1 or to 0, the number κ_k is odd or even and respectively the number a_{κ_k} belongs to $\mathfrak{C}$ or to its complement.

The relation between k, m and t expressed by formula

$$a_{\kappa_{m,k}} \leq t < a_{\kappa_{m+1,k}} \tag{20}$$

is decidable and hence it has a single-fold exponential Diophantine representation

$$\kappa_{m,k} \leq t < \kappa_{m+1,k} \Leftrightarrow \exists z_3 \ldots z_m \{M(k, m, t, z_3, \ldots, z_m) = 0\}. \tag{21}$$

The exponential polynomial $M(k, x_1, x_2, x_3, \ldots, x_m)$ constructed in this way has the required properties. It has exactly a_{κ_k} solutions, namely, for the value of z_2 we can take any of the numbers 0,1, ... , $a_{\kappa_k} - 1$, the value of z_1 is determined in a unique way from the condition $a_{\kappa_{z_1,k}} \leq z_2 < a_{\kappa_{z_1+1,k}}$, and the choice of the remaining unknowns is unique thanks to single-foldness of the representation (21). □

Open Problem. Could we improve any of the above mentioned definitions of Ω via the number of solutions of exponential Diophantine equations to similar definitions via the number of solutions of genuine (i.e., without exponentiation) Diophantine equations?

REFERENCES

[1] L. Adleman, K. Manders. Diophantine complexity. In: *17th Annual Symposium on Foundations of Computer Science*, pages 81–88, Houston, Texas, 25-26 October 1976. IEEE.

[2] G. J. Chaitin, *Algorithmic Information Theory*, Vol. 1 of Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, Cambridge, 1987.

[3] M. Davis, H. Putnam, J. Robinson, The decision problem for exponential diophantine equations, *Ann. of Math.* (2) 74 (1961) 425–436.

[4] A. N. Kolmogorov, Three approaches to the definition of the concept "quantity of information", *Problemy Peredači Informacii* 1 (vyp. 1) (1965) 3–11.

[5] M. Li, P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*, 2nd Edition, Graduate Texts in Computer Science, Springer-Verlag, New York, 1997.

[6] Yu. Matiyasevich. *Desyataya Problema Gilberta.* Moscow, Fizmatlit, 1993. English translation: Hilbert's tenth problem. MIT Press, 1993. French translation: Le dixième problème de Hilbert, Masson, 1995. URL: http://logic.pdmi.ras.ru/ yumat/H10Pbook, mirrored at http://www.informatik.uni-stuttgart.de/ifi/ti/personen/Matiyasevich/H10Pbook.

[7] T. Ord and T. D. Kieu. On the existence of a new family of Diophantine equations for Ω. *Fundam. Inform.* 56, No.3, 273-284 (2003).

[8] R. J. Solomonoff, A formal theory of inductive inference. I, *Information and Control* 7 (1964) 1–22.

16 September 2005

Steklov Institute of Mathematics at St-Petersburg

E-mail: yumat@pdmi.ras.ru