МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 510.52

О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ЭКЗИСТЕНЦИАЛЬНЫХ ТЕОРИЙ[1]

А. Н. Рыбалов

Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

Многие задачи о конечных графах и конечных полях могут быть сформулированы на языке универсальной алгебраической геометрии, в рамках которой эти объекты рассматриваются как алгебраические системы в заданном языке. Алгебраическая геометрия над этими объектами тесным образом связана со свойствами экзистенциальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий. В данной работе изучается вычислительная сложность экзистенциальных теорий алгебраических систем конечного предикатного языка с равенством. Известно, что для любой алгебраической системы с более чем одноэлементным основным множеством эта теория является NP-трудной (NP-полной, если основное множество конечно). Поэтому возникает вопрос о генерической сложности экзистенциальной теории алгебраической системы конечного предикатного языка. Доказывается, что при условиях $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$ для распознавания этой теории не существует полиномиального сильно генерического алгоритма.

Ключевые слова: *генерическая сложность, конечная алгебраическая система, экзистенциальная теория.*

DOI 10.17223/20710410/49/9

ON GENERIC COMPLEXITY OF THE EXISTENTIAL THEORIES

A. N. Rybalov

Sobolev Institute of Mathematics, Omsk, Russia

E-mail: alexander.rybalov@gmail.com

Many problems about finite graphs and finite fields can be formulated in the universal algebraic geometry, where these objects are considered as algebraic structures in the given language. Algebraic geometry over such objects is closely related to properties of existential theories. From a practical point of view, the most important are questions about decidability and computational complexity of these theories. In this paper we study the computational complexity of existential theories of algebraic structures of finite predicate language. It is known that the existential theory of any algebraic structure with more than one element is NP-hard. We prove that under the conditions $\mathrm{P} \neq \mathrm{NP}$ and $\mathrm{P} = \mathrm{BPP}$, for this theories there is no polynomial strongly generic

[1]Работа поддержана грантом РНФ № 19-11-00209.

algorithm. To prove this theorem we use the method of generic amplification, which allows to construct generically undecidable problems from the problems undecidable in the classical sense. The main ingredient of this method is a technique of cloning, which unites inputs of the problem together in the large enough sets of equivalent inputs. Equivalence is understood in the sense that the problem is solved similarly for them.

Keywords: *generic complexity, finite algebraic structure, existential theory.*

Введение

В XX веке, в связи с бурным развитием компьютерной техники и прикладной математики, на первый план вышли исследования различных конечных комбинаторных и алгебраических объектов. Например, конечные графы находят многочисленные приложения при решении практических задач, связанных с сетями, маршрутами, классификацией объектов и т. д. Другой пример — это конечные поля, без которых немыслимы современная криптография и теория помехоустойчивой передачи информации. Классическими подходами к изучению конечных объектов являются алгебраический и комбинаторный. Новый подход — логический и теоретико-модельный — родился внутри так называемой универсальной алгебраической геометрии [1]. В рамках этого подхода изучаемые объекты рассматриваются как алгебраические системы в заданном языке (сигнатуре). Многие практически важные задачи о конечных объектах можно формулировать как задачи, связанные с решением систем уравнений над соответствующими алгебраическими системами, что приводит к необходимости развития алгебраической геометрии. Алгебраическая геометрия тесным образом связана со свойствами экзистенциальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий.

Генерический подход к алгоритмическим проблемам предложен в [2]. В рамках этого подхода алгоритмическая проблема рассматривается не на всём множестве входов, а на некотором подмножестве «почти всех» входов. Такие входы образуют так называемое генерическое множество. Понятие «почти все» формализуется введением естественной меры на множестве входных данных. С точки зрения практики алгоритмы, решающие быстро проблему на генерическом множестве, так же хороши, как и быстрые алгоритмы для всех входов. Классическим примером такого алгоритма является симплекс-метод — он за полиномиальное время решает задачу линейного программирования для большинства входных данных, но имеет экспоненциальную сложность в худшем случае. Более того, может так оказаться, что проблема трудноразрешима или вообще неразрешима в классическом смысле, но легкоразрешима на генерическом множестве.

В данной работе изучается генерическая сложность экзистенциальных теорий алгебраических систем конечного предикатного языка с равенством. Для любой алгебраической системы с более чем одноэлементным основным множеством к этой теории может быть полиномиально сведена проблема выполнимости булевых формул [3]. Таким образом, проблема распознавания любой такой теории является NP-трудной (NP-полной, если основное множество конечно) и, при условии $\mathrm{P} \neq \mathrm{NP}$, не существует полиномиального алгоритма, распознающего её для всех входов. В работе доказывается, что при условиях $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$ для распознавания этой теории не существует полиномиального сильно генерического алгоритма. Класс BPP состоит из проблем, разрешимых за полиномиальное время на вероятностных машинах Тьюрин-

га. В [4] доказано, что равенство P = BPP следует из весьма правдоподобных гипотез о вычислительной сложности некоторых трудных проблем.

1. Предварительные сведения

Напомним некоторые определения из математической логики и теории моделей [1]. *Сигнатурой* называется множество σ, состоящее из предикатных, функциональных и константных символов. *Алгебраическая система* сигнатуры σ есть набор $\mathfrak{A} = \langle A, \sigma \rangle$, где A — непустое основное множество, причём каждому предикатному символу сигнатуры σ сопоставлен некоторый предикат на множестве A, каждому функциональному символу из σ — функция на множестве A со значениями в множестве A, а каждому константному символу из σ — элемент из A.

Формула логики первого порядка сигнатуры σ, в которой каждая переменная связана некоторым квантором, называется *предложением*. Тот факт, что предложение Φ сигнатуры σ истинно в алгебраической системе $\mathfrak{A} = \langle A, \sigma \rangle$, обозначается как $\mathfrak{A} \models \Phi$. Предложение Φ сигнатуры σ называется *экзистенциальным* (или $\exists$-предложением), если оно имеет вид

$$\Phi = \exists x_1 \ldots \exists x_n \varphi(x_1, \ldots, x_n),$$

где φ — бескванторная формула сигнатуры σ. *Элементарной теорией* алгебраической системы $\mathfrak{A}$ называется множество $Th(\mathfrak{A})$ всех предложений сигнатуры σ, истинных в $\mathfrak{A}$. Множество всех $\exists$-предложений теории $Th(\mathfrak{A})$ называется *экзистенциальной* теорией $Th_{\exists}(\mathfrak{A})$ алгебраической системы $\mathfrak{A}$.

Будем использовать числа Каталана C_n, которые определяются следующим образом:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Здесь $\binom{2n}{n}$ — соответствующий биномиальный коэффициент. В дальнейшем понадобится следующее утверждение о числах Каталана.

Лемма 1. Для $n > m$ имеет место

$$\frac{C_{n-m}}{C_n} > \frac{1}{4^m}.$$

Доказательство. Оценим отношение чисел Каталана:

$$\frac{C_{n-m}}{C_n} = \frac{n+1}{n-m+1}\,\frac{\binom{2(n-m)}{n-m}}{\binom{2n}{n}} = \frac{n+1}{n-m+1}\,\frac{\dfrac{(2(n-m))!}{(n-m)!(n-m)!}}{\dfrac{(2n)!}{n!n!}} >$$

$$> \frac{(2(n-m))!n!n!}{(n-m)!(n-m)!(2n)!} = \frac{n!}{(n-m)!}\,\frac{2(n-m)\ldots(n-m+1)}{2n\ldots(n+1)} =$$

$$= \frac{n(n-1)\ldots(n-m+1)2(n-m)\ldots(n-m+1)}{2n\ldots(n+1)} =$$

$$= \frac{(n\ldots(n-m+1))^2}{2n\ldots(2(n-m)+1)} > \Big(\frac{(n-1)\ldots(n-m)}{2(n-1)\ldots(2n-2m)}\Big)^2 > \frac{1}{2^{2m}} = \frac{1}{4^m}.$$

Лемма 1 доказана. ■

2. Генерические алгоритмы

Пусть I — некоторое множество входов. Для подмножества $S \subseteq I$ определим последовательность

$$\rho_n(S) = \frac{|S \cap I_n|}{|I_n|},\ n = 1, 2, 3, \ldots,$$

где I_n — множество входов размера n. Заметим, что $\rho_n(S)$ — это вероятность попасть в S при случайной и равновероятной генерации входов из I_n. *Асимптотической плотностью* S назовём предел

$$\rho(S) = \overline{\lim_{n\to\infty}} \rho_n(S).$$

Множество S называется *генерическим*, если $\rho(S) = 1$, и *пренебрежимым*, если $\rho(S) = 0$. Следуя [2], назовём множество S *сильно пренебрежимым*, если последовательность $\rho_n(S)$ экспоненциально быстро сходится к 0, т. е. существуют константы σ, $0 < \sigma < 1$, и $C > 0$, такие, что для любого n

$$\rho_n(S) < C\sigma^n.$$

Теперь S называется *сильно генерическим*, если его дополнение $I \setminus S$ сильно пренебрежимо.

Алгоритм $\mathcal{A}$ с множеством входов I и множеством выходов $J \cup \{?\}$ $(? \notin J)$ называется (*сильно*) *генерическим*, если

1) $\mathcal{A}$ останавливается на всех входах из I;
2) множество $\{x \in I : \mathcal{A}(x) \neq ?\}$ является (сильно) генерическим.

Генерический алгоритм $\mathcal{A}$ вычисляет функцию $f : I \to J$, если $(\mathcal{A}(x) = y \in J) \Rightarrow$ $\Rightarrow (f(x) = y)$ для всех $x \in I$. Ситуация $\mathcal{A}(x) = ?$ означает, что $\mathcal{A}$ не может вычислить функцию f на аргументе x. Но условие 2 гарантирует, что $\mathcal{A}$ корректно вычисляет f на почти всех входах (входах из генерического множества).

3. Представление экзистенциальных предложений

Рассмотрим естественное представление экзистенциальных предложений с помощью двоичных деревьев. Это представление, с одной стороны, настолько же компактно, как и стандартное представление строками символов (с точностью до линейного множителя). С другой стороны, оно удобно для различного рода подсчётов. Кроме того, достаточно просто написать компьютерную программу для случайной генерации предложений, заданных с помощью этого представления.

Зафиксируем конечную предикатную сигнатуру

$$\sigma = \{P_1^{(a_1)}, \ldots, P_k^{(a_k)}, c_1, \ldots, c_l\},$$

где P_i — предикаты (включая двуместный предикат равенства); c_i — константы. Положим

$$A = \max_{i=1,\ldots,k} \{a_i\}.$$

Пусть экзистенциальное предложение Φ сигнатуры σ имеет вид

$$\Phi = \exists x_1 \ldots \exists x_t \phi,$$

где ϕ — бескванторная формула, полученная с помощью конъюнкций и дизъюнкций из атомарных формул вида $P_i(x_1, \ldots, x_{a_i})$ или их отрицаний. Структуру формулы ϕ естественно представлять в виде бинарного дерева T_ϕ, такого, что внутренние вершины T_ϕ

помечены символами $\vee$ и $\wedge$, а листья T_ϕ — простыми атомарными формулами или их отрицаниями. Если T_ϕ имеет n листьев, то не более An переменных могут встретиться в T_ϕ, поэтому в дальнейшем будем полагать, что все переменные T_ϕ лежат в множестве $\{x_1, \ldots, x_{An}\}$. Под размером предложения Φ будем понимать число листьев в дереве T_ϕ. Для упрощения подсчётов считается, что предложение Φ размера n зависит от всех переменных $x_1, \ldots, x_{An}$ и все эти переменные связаны кванторами.

Обозначим через $\mathcal{F}$ множество экзистенциальных предложений, представленных описанным способом.

Лемма 2. Число экзистенциальных предложений размера n есть

$$|\mathcal{F}_n| = 2^{n-1} C_{n-1} \Big(2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^n. \tag{1}$$

Доказательство. Любое предложение из $\mathcal{F}$ размера n состоит из кванторной приставки и бинарного дерева с n листьями и $n-1$ внутренними вершинами. Известно (см., например, [5]), что существует C_{n-1} неразмеченных бинарных деревьев с n листьями. Каждая внутренняя вершина такого дерева может быть помечена символами $\vee$ или $\wedge$, поэтому есть всего 2^{n-1} таких разметок. Каждый из n листьев может быть помечен одним из a_i-местных предикатов P_i, $i = 1, \ldots, k$, либо его отрицанием, в которые можно подставлять на каждое из a_i мест либо одну из переменных x_j, $j = 1, \ldots, An$, либо константу c_j, $j = 1, \ldots, l$. Таким образом, получаем формулу (1). ■

Для любого экзистенциального предложения $\Phi = \exists x_1 \ldots \exists x_t \phi$ рассмотрим множество предложений

$$eq(\Phi) = \left\{ \exists x_1 \ldots \exists x_{2n} \big(\phi \vee ((x_1 \neq x_1) \wedge \psi) \big) \right\},$$

где $n > t$; ψ — произвольная бескванторная формула от переменных $x_1, \ldots, x_n$. Легко видеть, что все предложения из $eq(\Phi)$ эквивалентны Φ в том смысле, что они истинны или ложны одновременно с Φ.

Лемма 3. Для любого экзистенциального предложения Φ имеет место

$$\frac{|eq(\Phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} > \frac{1}{\big(16k(An + l)^A\big)^{m+2}}$$

для любого $n > m + 2$, где m — размер предложения Φ.

Доказательство. Рассмотрим любое предложение из множества $eq(\Phi)$ размера $n > m + 2$. Заметим, что число вершин в дереве T_ψ из этого предложения равно $n - m - 2$. Теперь аналогично тому, как это делалось в доказательстве леммы 2, можно подсчитать значение

$$|eq(\Phi) \cap \mathcal{F}_n| = 2^{n-m-3} C_{n-m-3} \Big(2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n-m-2}.$$

В результате получим

$$\frac{|eq(\Phi) \cap \mathcal{F}_n|}{|\mathcal{F}_n|} = \frac{2^{n-m-3} C_{n-m-3} \Big(2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n-m-2}}{2^{n-1} C_{n-1} \Big(2 \sum_{i=1}^{k} (An + l)^{a_i} \Big)^{n}} =$$

$$= \frac{1}{2^{m+2}\Big(2\sum\limits_{i=1}^{k}(An+l)^{a_i}\Big)^{m+2}}\ \frac{C_{n-m-3}}{C_{n-1}} >$$

$$> \frac{1}{2^{m+2}\Big(2\sum\limits_{i=1}^{k}(An+l)^{a_i}\Big)^{m+2}}\ \frac{1}{4^{m+2}} > \frac{1}{\big(16k(An+l)^{A}\big)^{m+2}}.$$

Здесь использована лемма 1 для оценки отношения чисел Каталана. ■

4. Основной результат

Теорема 1. Пусть $\mathfrak{A}$ — алгебраическая система конечной предикатной сигнатуры с более чем одним элементом в основном множестве. Если существует сильно генерический полиномиальный алгоритм, распознающий $Th_{\exists}(\mathfrak{A})$, то существует вероятностный полиномиальный алгоритм, распознающий $Th_{\exists}(\mathfrak{A})$ на всём множестве предложений.

Доказательство. Допустим, что существует сильно генерический полиномиальный алгоритм $\mathcal{A}$, распознающий $Th_{\exists}(\mathfrak{A})$. Тогда множество

$$G(\mathcal{A}) = \{\Phi \in \mathcal{F} : \mathcal{A}(\Phi) \neq ?\}$$

является сильно генерическим. Построим вероятностный полиномиальный алгоритм $\mathcal{B}$, определяющий истинность любого предложения Φ. На предложении Φ размера n алгоритм $\mathcal{B}$ работает следующим образом:

1) Генерирует случайное предложение Ψ из множества $eq(\Phi)$ размера n^2.
2) Вычисляет $\mathcal{A}(\Psi)$.
3) Если $\mathcal{A}(\Psi) \neq ?$, то определяет истинность Φ.
4) Если $\mathcal{A}(\Psi) = ?$, то выдаёт ответ «НЕТ».

Заметим, что алгоритм выдает правильный ответ на шаге 3, а на шаге 4 может выдать неправильный ответ. Нужно доказать, что вероятность того, что ответ выдаётся на шаге 4, меньше $1/2$.

Вероятность того, что случайное предложение вида Ψ из $eq(\Phi)_{n^2}$ не попадёт в $G(\mathcal{A})$, не больше

$$\frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|eq(\Phi)_{n^2}|} = \frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|\mathcal{F}_{n^2}|}\ \frac{|\mathcal{F}_{n^2}|}{|eq(\Phi)_{n^2}|}.$$

Так как $G(\mathcal{A})$ сильно генерическое, то существует константа $\alpha > 0$, такая, что

$$\frac{|(\mathcal{F} \setminus G(\mathcal{A}))_{n^2}|}{|\mathcal{F}_{n^2}|} < \frac{1}{2^{\alpha n^2}}$$

для любого n. С другой стороны, по лемме 3

$$\frac{|\mathcal{F}_{n^2}|}{|eq(\Phi)_{n^2}|} < \big(16k(An^2+l)^{A}\big)^{n+2}.$$

Поэтому искомая вероятность не больше

$$\frac{\big(16k(An^2+l)^{A}\big)^{n+2}}{2^{\alpha n^2}} = \frac{2^{(n+2)\log\left(16k(An^2+l)^{A}\right)}}{2^{\alpha n^2}}$$

и при больших n меньше $1/2$. Это означает, что вероятность выдачи ответа на шаге 4 меньше $1/2$.

Полиномиальность описанного алгоритма следует из существования процедуры генерации двоичного дерева размера N за время, полиномиально ограниченное от N. Эта процедура описана в [6]. ■

Непосредственно из теоремы 1 следует

Теорема 2. Пусть $\mathfrak{A}$ — алгебраическая система конечной предикатной сигнатуры с более чем одним элементом в основном множестве. Если $\mathrm{P} \neq \mathrm{NP}$ и $\mathrm{P} = \mathrm{BPP}$, то не существует сильно генерического полиномиального алгоритма, распознающего $Th_{\exists}(\mathfrak{A})$.

Автор выражает благодарность рецензенту за полезные замечания и предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Даниярова Э.Ю., Мясников А.Г., Ремесленников В.Н.* Алгебраическая геометрия над алгебраическими системами. Новосибирск: СО РАН, 2016. 288 с.
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks // J. Algebra. 2003. V. 264. No. 2. P. 665–694.
3. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 419 с.
4. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma // Proc. 29th STOC. El Paso: ACM, 1997. P. 220–229.
5. *Кнут Д.* Искусство программирования. М.: Вильямс, 2010. 720 с.
6. *Рыбалов А. Н.* О генерической сложности проблемы общезначимости булевых формул // Прикладная дискретная математика. 2016. № 2(32). С. 119–126.

REFERENCES

1. *Daniyarova E. Y., Myasnikov A. G., and Remeslennikov V. N.* Algebraicheskaya geometriya nad algebraicheskimi sistemami [Algebraic Geometry over Algebraic Structures]. Novosibirsk, SB RAS Publ., 2016. 288 p (in Russian).
2. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks. J. Algebra, 2003, vol. 264, no. 2, pp. 665–694.
3. *Garey M. and Johnson D.* Computers and Intractability. N. Y., Freeman & Co, 1979. 340 p.
4. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma. Proc. 29th STOC, El Paso, ACM, 1997, pp. 220–229.
5. *Knuth D. E.* The Art of Computer Programming. Reading, Massachusetts, Addison-Wesley, 1997.
6. *Rybalov A.* O genericheskoy slozhnosti problemy obshcheznachimosti bulevykh formul [On generic complexity of the validity problem for Boolean formulas]. Prikladnaya Diskretnaya Matematika, 2016, no. 2(32), pp. 119–126. (in Russian)