Алгебра и логика, 63, № 1 (2024), 89—99

DOI: 10.33048/alglog.2024.63.107 УДК 510.633

О ДЛИНЕ НЕВЫПОЛНИМОЙ ПОДФОРМУЛЫ

А. В. СЕЛИВЕРСТОВ

Введение

Пропозициональной конъюнктивной нормальной формой (КНФ) называется конъюнкция элементарных дизъюнкций, каждая из которых содержит лишь литералы. Литерал — пропозициональная переменная или отрицание переменной. Если в каждой элементарной дизъюнкции по k литералов, то такая КНФ называется k-КНФ. Через $\perp$ и $\top$ обозначены булевы константы: ложь и истина, соответственно.

Для двух чисел $\alpha < \beta$ множество α-or-β-in-SAT состоит из КНФ, для которых существует такая $(\perp, \top)$-оценка переменных, что в каждой элементарной дизъюнкции истинными оказались либо ровно α, либо ровно β литералов. Для $k < \alpha < \beta$ множество α-or-β-in-SAT не содержит ни одной k-КНФ. Для $k < \beta$ некоторая k-КНФ φ принадлежит множеству 1-or-β-in-SAT тогда и только тогда, когда φ принадлежит множеству 1-in-k-SAT, состоящему из k-КНФ, для которых существует такая $(\perp, \top)$-оценка переменных, что в каждой элементарной дизъюнкции ровно один литерал истинный. Некоторая 2-КНФ φ принадлежит множеству 1-or-2-in-SAT тогда и только тогда, когда φ выполнима. Некоторая 3-КНФ φ принадлежит множеству 1-or-2-in-SAT тогда и только тогда, когда φ принадлежит множеству NAE-3-SAT. Оно состоит из тех 3-КНФ, для которых существует такая $(\perp, \top)$-оценка переменных, что в каждой дизъюнкции некоторый литерал ложный и некоторый литерал истинный. Принадлежность 3-КНФ $\varphi(p_1, \ldots, p_n)$ к NAE-3-SAT равносильна выполнимости формулы

$\varphi(p_1,\dots,p_n) \wedge \varphi(\neg p_1,\dots,\neg p_n)$. Обе задачи распознавания принадлежности 3-КНФ к NAE-3-SAT и к 1-in-3-SAT, как известно, NP-полные. Длина формулы служит естественным параметром для оценки вычислительной сложности, см. [1, 2]. В [3] предложен алгоритм для проверки выполнимости КНФ, время работы которого на КНФ от n переменных с $m > n$ элементарными дизъюнкциями ограничено сверху функцией $O(2^{m-n}n^3)$. В частности, время работы полиномиально ограничено при фиксированной разности $m - n$, но экспоненциально быстро возрастает при $m > 2n$. Поэтому интересна возможность замены исходной 3-КНФ на подформулу. Известны некоторые результаты, ограничивающие возможность такой замены не только в худшем, но и в общем случае.

Фиксируем сколь угодно малое $\varepsilon > 0$. Почти все 2-КНФ от n переменных, у которых число элементарных дизъюнкций меньше порога $(1 - \varepsilon)n$, выполнимы. Напротив, невыполнимы почти все 2-КНФ от n переменных, у которых число элементарных дизъюнкций больше порога $(1+\varepsilon)n$, см. [4, 5]. Существование такого порога обычно называется фазовым переходом и на случайных выборках формул нескольких типов показано в [6].

Если каждая элементарная дизъюнкция содержит точно k литералов и каждая переменная входит точно в d элементарных дизъюнкций, то k-КНФ называется d-регулярной. Для любого достаточно большого числа k принадлежность случайной d-регулярной k-КНФ к множеству NAE-k-SAT испытывает с ростом d фазовый переход при некотором критическом значении d_k, которое зависит от k. С ростом числа переменных доля принадлежащих к NAE-k-SAT d-регулярных k-КНФ при $d < d_k$ стремится к единице, а при $d > d_k$ эта доля стремится к нулю, см. [7]. Похожий результат для d-регулярных k-КНФ, у которых при некоторой оценке истинны ровно два литерала каждой элементарной дизъюнкции, получен в [8].

В доказательствах мы заменяем КНФ на систему алгебраических уравнений, зависящих от вспомогательных переменных, по одной для каждой элементарной дизъюнкции. Так исходная задача о выполнимости КНФ сводится к задаче об инцидентности подпространства, заданного системой линейных уравнений, и множества точек с координатами из множества

$\{0,1\}$. Близкие задачи об инцидентности прямых и точек рассмотрены в работах [9,10]. С геометрической точки зрения удаление из КНФ элементарной дизъюнкции соответствует проекции на некоторое координатное подпространство. Эта проекция соответствует исключению вспомогательной переменной. Такой подход применим не только к КНФ, но и к другим формулам. Обычно (см. [11]) булевым формулам сопоставляют уравнения над полем из двух элементов, но аналогично можно рассматривать системы уравнений над другими полями. Решение системы уравнений, в котором каждая переменная принимает значения из множества $\{0,1\}$, называется $(0,1)$-решением. Поиск $(0,1)$-решения у системы линейных уравнений над полем рациональных чисел служит известной вычислительно трудной задачей, см. [12].

Далее следуют три раздела. Предварительные сведения, где собраны леммы из линейной алгебры, которые ниже используются для доказательства теорем. Результаты, где доказаны две теоремы. Обсуждение, где собраны примеры и замечания. Автор благодарен участникам конференции Мальцевские чтения (Новосибирск) за обсуждение этой работы.

§ 1. Предварительные сведения

Обозначим через K поле, содержащее хотя бы три элемента. Пусть система линейных уравнений от переменных $x_1, \ldots, x_n$ содержит больше одного уравнения и некоторое уравнение нетривиально зависит от переменной x_k. Новая система линейных уравнений получена из исходной системы исключением переменной x_k, если новая система не зависит от переменной x_k, а исходная система эквивалентна объединению новой системы и ровно одного уравнения (зависящего от x_k), равного линейной комбинации уравнений исходной системы.

Мы рассматриваем аффинное пространство над полем K с фиксированной системой декартовых координат. Ранг квадратной матрицы M связан с размерностью аффинной оболочки L точек, соответствующих столбцам матрицы. Если L содержит начало координат, то $\mathrm{rank}(M) = \dim(L)$, иначе $\mathrm{rank}(M) = \dim(L) + 1$.

Точка аффинного пространства, каждая координата которой равна нулю или единице, называется $(0,1)$-точкой. Замена координаты x_k на разность $1 - x_k$ переводит $(0,1)$-точку в другую $(0,1)$-точку. При этом размерность любого подпространства не меняется.

ЛЕММА 1. *Матрица M размера $n \times n$ над полем K такова, что каждый её элемент на главной диагонали отличен от нуля и единицы, а каждый элемент вне главной диагонали равен либо нулю, либо единице. Тогда ранг матрицы M не меньше числа $n/2$.*

ДОКАЗАТЕЛЬСТВО. Если $n = 2$, то ранг 2×2 матрицы M не меньше числа $n/2$, поскольку $\mathrm{rank}(M) \geqslant 1$. Предположим, что для некоторого $n \geqslant 3$ лемма уже доказана для всех рассматриваемых квадратных матриц меньшего порядка и рассмотрим $n \times n$ матрицу M.

Преобразование $x_k \to 1 - x_k$ означает замену всех элементов в k-ой строке матрицы. Применяя эти преобразования, можно перейти от матрицы M к матрице $\hat{M}$ того же типа, но в последнем столбце матрицы $\hat{M}$ все элементы, кроме элемента на главной диагонали, равны нулю. Матрица

$$\hat{M} = \left(\begin{array}{ccc|c} & & & 0 \\ & N & & \vdots \\ & & & 0 \\ \hline * & \cdots & * & c \end{array}\right)$$

для некоторого $c \notin \{0,1\}$. Размерности аффинных оболочек столбцов обеих матриц M и $\hat{M}$ совпадают. Поэтому выполнено $\mathrm{rank}(M) \geqslant \mathrm{rank}(\hat{M}) - 1$. Но если аффинная оболочка столбцов матрицы M содержит начало координат, то ранг матрицы M может быть меньше ранга матрицы $\hat{M}$.

Выполняя элементарные преобразования над столбцами матрицы $\hat{M}$, получим матрицу

$$\widetilde{M} = \left(\begin{array}{ccc|c} & & & 0 \\ & N & & \vdots \\ & & & 0 \\ \hline 0 & \cdots & 0 & c \end{array}\right)$$

того же ранга. У $\hat{M}$ и $\widetilde{M}$ могут отличаться лишь нижние строки. В нижней строке матрицы $\widetilde{M}$, кроме элемента на главной диагонали, остальные элементы равны нулю. Удаляя последний столбец и последнюю строку из матрицы $\widetilde{M}$, мы получим $(n-1)\times(n-1)$ матрицу N меньшего ранга. По предположению индукции, $\mathrm{rank}(N) \geqslant (n-1)/2$. Поэтому выполнено неравенство $\mathrm{rank}(\widetilde{M}) \geqslant (n+1)/2$.

Обозначим через L аффинную оболочку столбцов матрицы $\widetilde{M}$. Если L проходит через начало координат, то $\mathrm{rank}(\widetilde{M}) = \dim(L)$. В этом случае $\mathrm{rank}(M) \geqslant \dim(L) = \mathrm{rank}(\widetilde{M}) \geqslant (n+1)/2$.

Если же подпространство L не проходит через начало координат, то выполнены неравенства $\mathrm{rank}(M) \geqslant \mathrm{rank}(\widetilde{M}) - 1 = \mathrm{rank}(N)$. В этом случае аффинная оболочка столбцов матрицы N не проходит через начало координат. Вновь применим преобразования вида $x_k \to 1 - x_k$ к матрице N и получим матрицу

$$\widetilde{N} = \left(\begin{array}{ccc|c} & & & 0 \\ & W & & \vdots \\ & & & 0 \\ \hline * & \cdots & * & * \end{array}\right)$$

того же вида, но в последнем столбце матрицы $\widetilde{N}$ все элементы, кроме элемента на главной диагонали, равны нулю. Более того, поскольку аффинная оболочка столбцов матрицы N не проходит через начало координат, выполнено неравенство $\mathrm{rank}(N) \geqslant \mathrm{rank}(\widetilde{N})$. Удаляя из матрицы $\widetilde{N}$ последний столбец и последнюю строку, получим $(n-2)\times(n-2)$ матрицу W меньшего ранга. По предположению индукции, ранг матрицы W ограничен снизу: $\mathrm{rank}(W) \geqslant (n-2)/2$. Тогда $\mathrm{rank}(\widetilde{N}) = \mathrm{rank}(W) + 1 \geqslant n/2$. Наконец, $\mathrm{rank}(M) \geqslant \mathrm{rank}(N) \geqslant \mathrm{rank}(\widetilde{N}) \geqslant n/2$. Лемма 1 доказана.

ЛЕММА 2. *Дана $n\times n$ матрица M над полем K, в которой каждый элемент на главной диагонали отличен от нуля и от единицы, а вне главной диагонали равен либо нулю, либо единице. Если никакая $(0,1)$-точка не лежит в аффинной оболочке столбцов матрицы M, то ранг матрицы M не меньше числа $(n+1)/2$.*

ДОКАЗАТЕЛЬСТВО. Применение к строкам матрицы преобразований $x_k \to 1 - x_k$, как в доказательстве теоремы 1, позволяет перейти от матрицы M к матрице $\hat{M}$ того же типа, но в последнем столбце матрицы $\hat{M}$ все элементы, кроме элемента на главной диагонали, равны нулю. Матрица

$$\hat{M} = \left(\begin{array}{ccc|c} & & & 0 \\ & N & & \vdots \\ & & & 0 \\ \hline * & \cdots & * & c \end{array}\right)$$

для некоторого $c \notin \{0, 1\}$. Поскольку аффинная оболочка столбцов матрицы M не содержит никакой $(0, 1)$-точки, это верно и для матрицы $\hat{M}$. В частности, аффинная оболочка столбцов матрицы не содержит начала координат. Поэтому $\mathrm{rank}(M) = \mathrm{rank}(\hat{M}) = \mathrm{rank}(N) + 1$. По лемме 1, выполнено $\mathrm{rank}(N) \geqslant (n-1)/2$. Следовательно, $\mathrm{rank}(M) \geqslant (n+1)/2$. Лемма 2 доказана.

ЛЕММА 3. *Дана система из m линейно независимых линейных уравнений от n переменных над полем K. Если эта система не имеет $(0, 1)$-решения и выполнено неравенство $n < 2m - 1$, то существует такая переменная, что система, полученная исключением этой переменной, также не имеет $(0, 1)$-решения.*

ДОКАЗАТЕЛЬСТВО. Предположим, что система линейных уравнений не имеет $(0, 1)$-решения, но исключение любой переменной приводит к системе, которая имеет $(0, 1)$-решение. Тогда для любого индекса j исходная система имеет решение, в котором только значение j-й переменной не принадлежит к $\{0, 1\}$, а значения остальных переменных принадлежат к $\{0, 1\}$. Отождествим такое решения с j-м столбцом некоторой $n \times n$ матрицы. У этой матрицы ранг не выше выражения $n - m + 1$, на главной диагонали никакой элемент не принадлежит к $\{0, 1\}$, а все элементы вне главной диагонали принадлежат к $\{0, 1\}$. Согласно лемме 2, это невозможно при условии $n < 2m - 1$. Противоречие. Лемма 3 доказана.

§ 2. Результаты

ТЕОРЕМА 1. *Дана система из m линейных уравнений специального вида $y_j = \ell_j(x_1, \ldots, x_n)$ от $m+n$ переменных $y_1, \ldots, y_m, x_1, \ldots, x_n$, где через ℓ_j обозначены линейные функции над некоторым полем. Если эта система не имеет $(0,1)$-решения и выполнено неравенство $m > 2n + 2$, то в системе существует такое уравнение, что подсистема, полученная удалением этого уравнения, также не имеет $(0,1)$-решения.*

ДОКАЗАТЕЛЬСТВО. Удаление j-го уравнения эквивалентно исключению переменной y_j. С другой стороны, рассматриваемая система состоит из линейно независимых уравнений и определяет аффинное подпространство размерности n. Исключение любой переменной приводит к новой системе, содержащей $m-1$ или m линейно независимых уравнений. Но исключение переменной x_i может потребовать замену исходных уравнений на их линейные комбинации. По лемме 3, при условии $n+m-r < 2(m-r)-1$, можно исключить хотя бы r каких-то переменных, сохраняя отсутствие $(0,1)$-решения. Это условие эквивалетно неравенству $r < m-n-1$. Но если $n < r$, то можно исключить некоторую переменную y_j. Поэтому при $m > 2n+2$ из исходной системы можно исключить переменную y_j, следовательно, удалить j-е уравнение. Теорема доказана.

ТЕОРЕМА 2. *Дана пропозициональная КНФ $\varphi(p_1, \ldots, p_n)$ с m элементарными дизъюнкциями от n переменных. Если φ не принадлежит к α-or-β-in-SAT и выполнено условие $m > 2n+2$, то существует КНФ, которая не принадлежит к α-or-β-in-SAT и получена удалением из φ некоторой элементарной дизъюнкции.*

ДОКАЗАТЕЛЬСТВО. Определим по индукции функцию f, которая элементарной дизъюнкции сопоставляет псевдобулеву линейную функцию над полем рациональных чисел. Константам сопоставим числа: $f(\bot) = 0$ и $f(\top) = 1$. Пропозициональной переменной p_i сопоставим переменную $f(p_i) = x_i$. Отрицанию $\neg p_i$ сопоставим выражение $f(\neg p_i) = 1 - x_i$. Далее j-й элементарной дизъюнкции $\varphi_j = \ell_1 \vee \cdots \vee \ell_k$ сопоставим выражение $f(\ell_1) + \cdots + f(\ell_k) - \alpha - (\beta - \alpha) y_j$, где новая переменная y_j встречается лишь один раз. Далее конъюнкции элементарных дизъюнкций φ_j сопоставим систему линейных уравнений $f(\varphi_j) = 0$ для всех индексов $1 \leqslant j \leqslant m$. Эта система зависит от $m+n$ переменных. Все m уравнений линейно независимые, поскольку каждое зависит от своей вспомогательной переменной. Мы используем, что выполнено $\alpha \neq \beta$, следовательно, полученное уравнение нетривиально зависит от переменной y_j.

Каждое $(0,1)$-решение этой системы соответствует $(\bot, \top)$-оценке пропозициональных переменных, при которой в каждой элементарной дизъюнкции истинными оказались либо ровно α, либо ровно β литералов. И обратно, для такой $(\bot, \top)$-оценки пропозициональных переменных существует $(0,1)$-решение системы уравнений. Если $p_i = \bot$, то $x_i = 0$. Если $p_i = \top$, то $x_i = 1$. Если в j-й элементарной дизъюнкции выполнено ровно α литералов, то $y_j = 0$. Если выполнено ровно β литералов, то $y_j = 1$.

Согласно теореме 1, если эта система не имеет $(0,1)$-решения, то это свойство сохраняется после удаления некоторого уравнения, что соответствует удалению некоторой элементарной дизъюнкции из КНФ φ. Теорема доказана.

§ 3. Обсуждение

Полученная граница для числа элементарных дизъюнкций невыполнимой подформулы в 2-КНФ близка к оптимальной. Существует невыполнимая 2-КНФ с $m = 2n$ элементарными дизъюнкциями от n переменных, для которой выполнима подформула, полученная удалением любой элементарной дизъюнкции. Примером служит 2-КНФ

$$(\neg p_1 \vee p_2) \wedge (p_1 \vee \neg p_2) \wedge \cdots \wedge (\neg p_{n-1} \vee p_n) \wedge (p_{n-1} \vee \neg p_n) \wedge (p_n \vee p_1) \wedge (\neg p_n \vee \neg p_1),$$

где каждая переменная дважды входит позитивно и дважды — негативно. Эта 2-КНФ эквивалентна конъюнкции формул, выражающих эквивалентности переменных p_j и p_{j+1} для $j < n$, а также эквивалентность переменной p_n и отрицания переменной p_1. Поэтому она невыполнимая. Но удаление из этой 2-КНФ одной элементарной дизъюнкции соответствует замене некоторой эквивалентности на импликацию. Полученная формула выполнима на некоторой $(\bot, \top)$-оценке, при которой антецедент этой импликации ложный.

С другой стороны, граница для уменьшения числа элементарных дизъюнкций лежит в области, где почти любая 2-КНФ невыполнима [4, 5]. Это естественно: возможность удаления некоторых элементарных дизъюнкций из невыполнимой 2-КНФ не накладывает неожиданных дополнительных ограничений на невыполнимую подформулу. Также полученные результаты для других классов формул дают верхнюю оценку для границы фазового перехода, когда он существует.

Возможно обобщение на конъюнкции формул другого вида. Обозначим через $\mathrm{maj}(p_1, p_2, p_3)$ функцию голосования (majority), значение которой равно наиболее часто встретившейся среди $(\bot, \top)$-оценок пропозициональных переменных p_1, p_2 и p_3. Для литералов ℓ_{ij} конъюнкция $\bigwedge\limits_j \mathrm{maj}(\ell_{1j}, \ell_{2j}, \ell_{3j})$ выполнима тогда и только тогда, когда 3-КНФ $\bigwedge\limits_j (\ell_{1j} \vee \ell_{2j} \vee \ell_{3j})$ принадлежит множеству 2-or-3-in-SAT. Поэтому к формулам такого вида применима теорема 2.

По аналогии с доказательством теоремы 2, теорема 1 позволяет накладывать разные ограничения на различные элементарные дизъюнкции в составе КНФ. Например, можно рассматривать такие конъюнкции $\varphi \wedge \psi$ двух КНФ, что на некоторой $(\bot, \top)$-оценке в каждой элементарной дизъюнкции в φ истинно одно число литералов, а в ψ — другое число литералов.

Уменьшение длины формулы приводит к снижению некоторых оценок вычислительной сложности, как показано в работах [2,3]. Однако это не гарантирует сокращения времени работы некоторых алгоритмов. Так, применяя метод резолюций для проверки выполнимости КНФ, иногда можно на промежуточных шагах получить те элементарные дизъюнкции, которые ранее были удалены при переходе от исходной КНФ к подформуле меньшей длины. С другой стороны, основным результатом служит чистая теорема существования, не дающая быстрого алгоритма для поиска невыполнимой подформулы. Это согласуется с гипотезой о высокой вычислительной сложности задачи о выполнимости в худшем случае, высказываемой многими авторами [11, 12].

ЛИТЕРАТУРА

1. *W. Li, C. Xu, Y. Yang, J. Chen, J. Wang*, A refined branching algorithm for the maximum satisfiability problem, Algorithmica, **84** (2022), 982—1006; `https://doi.org/10.1007/s00453-022-00938-8`
2. *J. Peng, M. Xiao*, Further improvements for SAT in terms of formula length, Inform. Comput., **294** (2023), 105085; `https://doi.org/10.1016/j.ic.2023.105085`
3. *S. Szeider*, Minimal unsatisfiable formulas with bounded clause-variable difference are fixed-parameter tractable, J. Comput. System Sci., **69**, No. 4 (2004), 656—674; `https://doi.org/10.1016/j.jcss.2004.04.009`
4. *A. Goerdt*, A threshold for unsatisfiability, J. Comput. System Sci., **53**, No. 3 (1996), 469—486; `https://doi.org/10.1006/jcss.1996.0081`
5. *D. Achlioptas, A. Coja-Oghlan, M. Hahn-Klimroth, J. Lee, N. Müller, M. Penschuck, G. Zhou*, The number of satisfying assignments of random 2-SAT formulas, Random Structures Algorithms, **58**, No. 4 (2021), 609—647; `https://doi.org/10.1002/rsa.20993`
6. *J. Giráldez-Cruand, J. Levy*, Popularity-similarity random SAT formulas, Artificial Intelligence, **299** (2021), 103537; `https://doi.org/10.1016/j.artint.2021.103537`
7. *J. Ding, A. Sly, N. Sun*, Satisfiability threshold for random regular NAE-SAT, Comm. Math. Phys., **341** (2016), 435—489; `https://doi.org/10.1007/s00220-015-2492-8`
8. *G. Nie, D. Xu, X. Wang, X. Wang*, The phase transition analysis for the random regular exact 2-(d, k)-SAT problem, Symmetry, **13**, No. 7 (2021), 1231; `https://doi.org/10.3390/sym13071231`
9. *А. А. Бойков, А. В. Селиверстов*, О кубе и проекциях подпространства, Вестн. Удмуртского ун-та. Матем. Мех. Компьют. науки, **33**, № 3 (2023), 402—415; `https://doi.org/10.35634/vm230302`
10. *Н. Т. Когабаев*, О системах диофантовых уравнений над конечными конфигурациями, Сиб. матем. ж., **64**, № 2 (2023), 321—338.

11. *M. N. Vyalyi*, Testing the satisfiability of algebraic formulas over the field of two elements, Probl. Inf. Transm., **59**, No. 1 (2023), 57—62;
`https://doi.org/10.1134/S0032946023010052`

12. *A. Abboud, K. Bringmann, D. Hermelin, D. Shabtay*, Scheduling lower bounds via AND subset sum, J. Comput. System Sci., **127** (2022), 29—40;
`https://doi.org/10.1016/j.jcss.2022.01.005`

Поступило 15 декабря 2023 г.

Окончательный вариант 4 декабря 2024 г.

Адрес автора:

СЕЛИВЕРСТОВ Александр Владиславович, Ин-т проблем передачи информ. им. А. А. Харкевича РАН, г. Москва, РОССИЯ.
e-mail: SLVSTV@iitp.ru