

PROOF COMPLEXITIES ON A CLASS OF BALANCED FORMULAS
IN SOME PROPOSITIONAL SYSTEMS

A. A. CHUBARYAN *

Chair of Discrete Mathematics and Theoretical Informatics, YSU, Armenia

In this paper four proof complexity characteristics for some class of balanced tautologies are investigated in two proof systems of propositional logic. One of the considered systems is based on determinative disjunctive normal form, the other on the generalization of splitting method. The optimal upper and lower bounds by logarithmic scale for all main proof complexity characteristics of considered tautologies are obtained in both systems.

<https://doi.org/10.46991/PYSU:A/2022.56.2.058>

MSC2010: Primary: 03F20; Secondary: 03F07.

Keywords: balanced tautologies, elimination system, generalized splitting system, proof complexity characteristics.

Introduction. Propositional proof complexity has its origin in the seminal paper by Cook and Reckhow [1]. It provides a path for approaching the P vs. NP problem: proving super-polynomial lower bounds to all propositional proof systems is equivalent to showing that NP is different from $coNP$ and, therefore, P is different from NP . It is well known that the exponential lower bounds for proof sizes of some sets of tautologies are obtained in many systems, but for some most natural calculi, in particular for Frege systems, the question about polynomial bounded sizes is still open. In many papers, some specific sets of tautologies are introduced, and it is shown that the question about polynomial bounded sizes for Frege proofs of all tautologies is reduced to an analogous question for a set of specific tautologies. In particular, Lutz Strasburger introduced in [2] the notion of balanced formulas and showed, that if there are polynomial bounded Frege proofs for the set of balanced tautologies, then the Frege systems should have a polynomial-size $p(n)$ proof for every tautology of size n . In this paper the main proof complexity characteristics (lines, size, space and width) for the balanced formulas $QHQ_n = \bigvee_{0 \leq i \leq n} \bigwedge_{1 \leq j \leq n} \left[\bigvee_{1 \leq k \leq i} \bar{q}_{i,j,k} \vee \bigvee_{i < k \leq n} q_{k,j,i+1} \right]$ are investigated in the system E [3], which is based on the determinative disjunctive normal, and in the system GS [4], which is based on the generalization of splitting

* E-mail: achubaryan@ysu.am

method. While the system E is polynomial equivalent with of well-known resolution system, cut-free sequent and cut-free Frege systems [3], the place of the system GS in the hierarchy of the propositional proof systems [1] is still unknown. Moreover, the comparison of the two main proof complexity characteristics (*lines* and *sizes*) for two classes of formulas in the system GS and Frege systems is considered in [5, 6]. It is shown that for one class of considered formulas the bounds in the system GS are better, than in Frege systems and vice versa for the second class. For all above considered cases the investigation of proof size for formulas QHQ_n in the system E and especially in the system GS are very important.

In this paper we obtain optimal upper and lower bounds by logarithmic scale for all main proof complexity characteristics of considered tautologies in both systems.

Preliminaries. We will use the current concepts of the unit Boolean cube (E^n), a propositional formula, conjunct, disjunctive normal form (DNF), a proof system for propositional logic, and proof complexity. The language of considered systems contains the propositional variables, logical connectives \neg , $\&$, \vee , and parentheses $(,)$. Note that some parentheses can be omitted in generally accepted cases. Following the usual terminology we call the variables and negated variables *literals*. The conjunct K (clause) can be represented simply as a set of literals (no conjunct contains a variable and its negation simultaneously). In [3] the following notions were introduced.

We call a *replacement-rule* each of the following trivial identities for a propositional formula ψ :

$$\begin{aligned} 0 \& \psi = 0, & \psi \& 0 = 0, & 1 \& \psi = \psi, & \psi \& 1 = \psi, \\ 0 \vee \psi = \psi, & \psi \vee 0 = \psi, & 1 \vee \psi = 1, & \psi \vee 1 = 1, \\ \bar{0} = 1, & \bar{1} = 0, & \overline{\bar{\psi}} = \psi. \end{aligned}$$

Application of a replacement-rule to some word requires the replacing of its sub words, having the form of the left-hand side of identity by the right-hand side.

Let φ be a propositional formula, $P = \{p_1, p_2, \dots, p_n\}$ be the set of all variables of φ , and $P' = \{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$ ($1 \leq m \leq n$) be some subset of P .

Definition 1. Given $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_m\} \subset E^m$, the conjunct $K^\sigma = \{p_{i_1}^{\sigma_1}, p_{i_2}^{\sigma_2}, \dots, p_{i_m}^{\sigma_m}\}$ is called φ -1-determinative (φ -0-determinative), if assigning σ_j ($1 \leq j \leq m$) to each p_{i_j} and successively using replacement-rules we obtain the value of φ (1 or 0) independently of the values of the remaining variables.

φ -1-determinative conjunct and φ -0-determinative conjunct are called also φ -determinative or determinative for φ .

Definition 2. 1-determinative for φ conjunct $K^\sigma = \{p_{i_1}^{\sigma_1}, p_{i_2}^{\sigma_2}, \dots, p_{i_m}^{\sigma_m}\}$ is called minimal determinative, if no subset of K^σ is determinative for φ .

Definition 3. $DNF D = \{K_1, K_2, \dots, K_j\}$ is called determinative DNF ($dDNF$) for φ , if $\varphi = D$ and every conjunct K_i ($1 \leq i \leq j$) is 1-determinative for φ .

The Proof System E . The axioms of E are not fixed, but for every formula φ each conjunct from some $dDNF$ of φ can be considered as an axiom. The *elimination rule* (e -rule) infers $K' \cup K''$ from clauses $K' \cup \{p\}$ and $K' \cup \{\bar{p}\}$, where K' and K'' are clauses and p is a variable.

The proof in E is a finite sequence of clauses such that every clause in the sequence is one of the axioms of E , or is inferred from earlier clauses in the sequence by e -rule. $DNF D = \{K_1, K_2, \dots, K_l\}$ is called full (tautology), if using e -rule the empty conjunction (\emptyset) can be proved from the axioms $\{K_1, K_2, \dots, K_l\}$.

The Proof System GS . Let φ be some formula and p be some of its variable. Results of splitting method of formula φ by variable p (splinted variable) are the formulas $\varphi[p^\delta]$ for every δ from the set $\{0, 1\}$, which are obtained from φ by assigning δ to each occurrence of p and successively using replacement-rules. The generalization of splitting method allow as associate with every formula φ some tree with root, nodes of which are labeled by formulas and edges, labeled by literals. The root is labeled by itself formula φ . If some node is labeled by formula v and α is some its variable, then both edges, which going out from this node, are labeled by one of literals α^δ for every δ from the set $\{0, 1\}$, and each of 2 “sons” of this node is labeled by corresponding formula $v[\alpha]^\delta$. Each of the tree’s leafs is labeled with some constant from the set $\{0, 1\}$. The tree, which is constructed for formula φ by described method, we will call *splitting tree* (s.t.) of φ . It is obvious, that changing the order of splinted variables in a given formula φ , we can obtain the different splitting trees of φ .

The GS proof system can be defined as follows: for every formula φ must be constructed some s.t., and if all tree’s leafs are labeled by the value 1, then formula φ is tautology and, therefore, we can consider the pointed constant 1 as axiom, and for every formula v , which is label of some s.t. node, and p is its splinted variable, then the Figure $v[p^0], v[p^1] \vdash v$ can be considered as some inference rule, hence every above described s.t. can be considered as some proof of φ in the system GS . Note, that if we consider splitting method for formulas given in DNF , then GS system is the well-known system Analytic Tableaux.

Proof Complexity Measures. In the theory of proof complexity two main characteristics of the proof are: t -complexity, defined as the number of proof steps (lines) and l -complexity, defined as total number of proof symbols (size). Now we consider two measures (space and width) also. s -Complexity (space), informal defined as maximum of minimal number of symbols on blackboard needed to verify all steps in the proof and w -complexity (width), defined as the maximum of widths of proof formulas. Follow to [7], we give the formal definitions of the mentioned proof complexity measures.

If a proof in the system Φ is a sequence of lines L (lines, for example, are conjuncts in E , formulas in GS), where each line is an axiom, or is derived from previous lines by one of a finite set of allowed inference rules, then a Φ -configuration is a set of such lines. A sequence of Φ -configurations $\{D_0, D_1, \dots, D_r\}$ is said to be Φ -derivation, if D_0 is empty set and for all t ($1 \leq t \leq r$), the set D_t is obtained from D_{t-1} by one of the following derivation steps:

Axiom Download. $D_t = D_{t-1} \cup \{L_A\}$, where L_A is an axiom of Φ .

Inference. $D_t = D_{t-1} \cup \{L\}$ for some L inferred by one of the inference rules for Φ from a set of assumptions, belonging to D_{t-1} .

Erasure. $D_t \subset D_{t-1}$.

A Φ -proof of a tautology φ is a Φ -derivation $\{D_0, D_1, \dots, D_r\}$ such that D_0 is empty and $\tilde{\varphi} \in D_r$, where $\tilde{\varphi}$ is empty conjunct in E and $\tilde{\varphi}$ is φ in GS .

By $|\varphi|$ we denote the size of a formula φ , defined as the number of all logical signs entries. It is obvious that the full size of a formula, which is understood to be the number of all symbols, is bounded by some linear function in $|\varphi|$.

The *size* (l) of a Φ -derivation is a sum of the sizes of all lines in a derivation, where lines that are derived multiple times are counted with repetitions. The *lines* (t) of a Φ -derivation is the number of axioms downloads and inference steps in it. The *space* (s) of a Φ -derivation is the maximal space of a configuration in a derivation, where the space of a configuration is the total number of logical signs in a configuration, counted with repetitions. The *width* (w) of a Φ -derivation is the size of the widest line in a derivation.

Let Φ be a proof system and φ be a tautology. We denote by $t_\varphi^\Phi (l_\varphi^\Phi, s_\varphi^\Phi, w_\varphi^\Phi)$ the minimal possible value of t -complexity (l -complexity, s -complexity, w -complexity) for all proofs of tautology φ in Φ . Further we denote proof complexities of formulas QHQ_n in the system Φ by $t^\Phi(n) (l^\Phi(n), s^\Phi(n), w^\Phi(n))$.

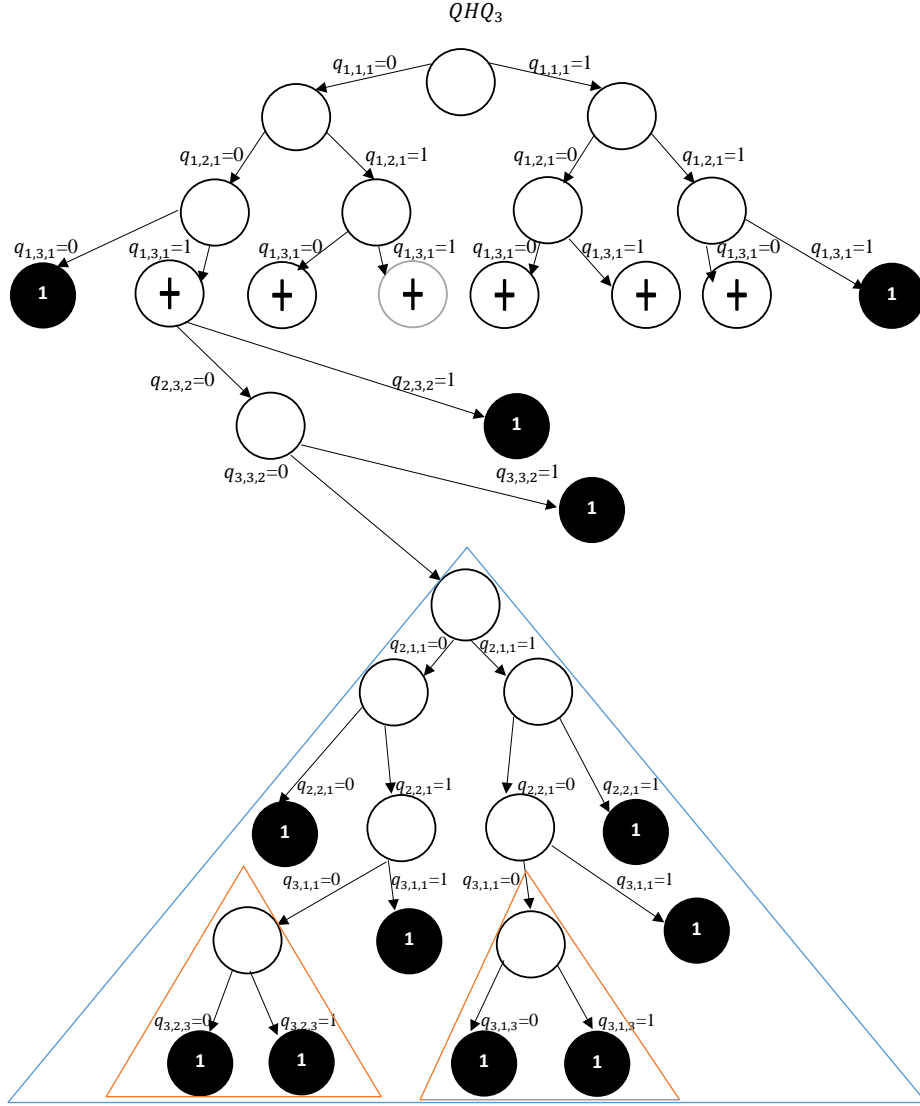
Balanced Formulas. A formula A is balanced, if every propositional variable occurring in A , occurs exactly twice, once positive and once negative. The tautologies

$QHQ_n = \bigvee_{0 \leq i \leq n} \bigwedge_{1 \leq j \leq n} \left[\bigvee_{1 \leq k \leq i} \bar{q}_{i,j,k} \vee \bigvee_{i < k \leq n} q_{k,j,i+1} \right]$ are balanced. At first, we give some quantitative characteristics for these formulas. It is not difficult to see that $|QHQ_n| = \frac{3n^2(n+1)}{2} - 1$. Let for all $n \geq 1$ and $0 \leq i \leq n$, $1 \leq j \leq n$ $Q_{i,j}$ be the formula $Q_{i,j} = \bigvee_{1 \leq k \leq i} \bar{q}_{i,j,k} \vee \bigvee_{i < k \leq n} q_{k,j,i+1}$, then

$$QHQ_n = \bigvee_{0 \leq i \leq n} (Q_{i1} \wedge Q_{i2} \wedge \dots \wedge Q_{ij} \wedge \dots \wedge Q_{i(n-1)} \wedge Q_{in}),$$

and H_i be the formula $(Q_{i1} \wedge Q_{i2} \wedge \dots \wedge Q_{in}) = H_i$, then $QHQ_n = H_0 \vee H_1 \vee \dots \vee H_n$. It is obvious, that minimal numbers of literals in 1-determinative conjunct for every $Q_{i,j}$ is 1, and minimal numbers of literals in 1-determinative conjunct for every H_i , therefore for QHQ_n is n .

Proof Complexities Bounds of QHQ_n in the System GS . Here we describe some algorithm for construction s.t. for QHQ_n . At first, we choose as splinted some variable from Q_{01} . As result, we will have on the first level 2 different formulas without of first splinted variable. Then, choosing as splinted some variable from Q_{02} for both formulas on next stage, we will have 4 different formulas without of second splinted variable and etc. in same way. The further description can be easier, if we consider it on the example of s.t. for QHQ_3 , which is given below (see Figure).

Splitting tree for QHQ_3 .

After splitting by one variable from every Q_{0j} ($1 \leq j \leq n$) on the n -th level, we obtain two nodes, labeled by 1 and $\left(\sum_{i=0}^n 2^i - 2\right)$ different formulas labeled by +, which have the same size and “the same construction”. From each of such formulas after $n - 2$ splitting we can obtain s.t. for QHQ_{n-1} . So, for $t^{GS}(n)$ we can write the following recurrent relation

$$\begin{aligned} t^{GS}(n) &\leq \left(\left(\sum_{i=0}^n 2^i \right) - 2 + (2^n - 2) \left((n-2) + (t^{GS}(n-1) - 1) \right) \right) + 1 \leq \\ &\leq (2^n - 2)t^{GS}(n-1) + 4 + n2^n \leq \dots \leq 2^{\frac{1}{2}(n+1)(n+2)}. \end{aligned}$$

For the lower bound of $t^{GS}(n)$ we must note, that occurs of different variables in formulae QHQ_n is “the same” and, therefore, the turn of splendid variables is immaterial, consequently in every s.t. of QHQ_n only the number of nodes, labeled by +, must be no less than $(2^n - 2)(2^{n-1} - 2) \dots (2^3 - 2)(2^2 - 2)$. Taking into consideration that $(2^n - 2) \geq 2^{n-1}$ for $n \geq 2$, we obtain

$$t^{GS}(n) \geq 2^{n-1} 2^{n-2} \dots 2^1 \geq 2^{\frac{1}{2}(n-1)n}.$$

So, for sufficiently large n $\log_2 t^{GS}(n) = \Omega(n^2)$ and $\log_2 t^{GS}(n) = O(n^2)$.

For the bounds of proof size we use the trivial relations $t^\varphi(\varphi) \leq l^\varphi(\varphi) \leq t^\varphi(\varphi) |A|$, where A is the largest formula in the proof. As the formula QHQ_n is the largest in it s.t., we obtain

$$l^{GS}(n) \leq \frac{3n^2(n+1)}{2} 2^{\frac{1}{2}(n+1)(n+2)} \quad \text{and} \quad l^{GS}(n) \geq 2^{\frac{1}{2}(n-1)n}.$$

So, for sufficiently large n $\log_2 l^{GS}(n) = \Omega(n^2)$ and $\log_2 l^{GS}(n) = O(n^2)$.

It is not difficult to prove, that the longest branch in suggested s.t. for QHQ_n has n^2 nodes, and in every tree-like k -depth proof the maximum of minimal number of lines on the blackboard needed to verify all steps in the proof is $k + 2$, therefore $s^{GS}(n) \leq (n^2 + 2) \frac{3n^2(n+1)}{2}$. Taking into consideration that the formula QHQ_n and the results of it splitting by one variable must be in each s.t., we have $s^{GS}(n) \geq \frac{3n^2(n+1)}{2} - 1 + n(n-1)2 - 2 = \frac{5n^2(n-1)}{2} - 2$.

So, it is proved that $\log_2 s^{GS}(n) = \Omega(\log_2 n)$ and $\log_2 s^{GS}(n) = O(\log_2 n)$.

It is obvious that $w^{GS}(n) = \frac{3n^2(n+1)}{2} - 1$ and for sufficiently large n $\log_2 w^{GS}(n) = \theta(\log_2 n)$.

So, it is proved the following

Theorem 1. *Let n be sufficiently large, then for lines, size, space, and width complexities of sequences QHQ_n in the system GS the following holds:*

$$\begin{aligned} \log_2 t^{GS}(n) &= \theta(n^2), \quad \log_2 l^{GS}(n) = \theta(n^2), \\ \log_2 s^{GS}(n) &= \theta(\log_2 n) \quad \text{and} \quad \log_2 w^{GS}(n) = \theta(\log_2 n). \end{aligned}$$

Proof Complexities Bounds of QHQ_n in the System E. Here we describe some trivial algorithm for transformation s.t. for QHQ_n into proof in E . Note that the set of all literals for every branch of s.t., going from the root to the node, labeled by 1, is 1-determinative conjunct for QHQ_n and the set of all such conjuncts is $dDNF$ for QHQ_n , therefore, if we turn over the s.t., and label every node with 1 by corresponding 1-determinative conjunct, the root with empty conjunct, the other nodes – with the set of literals on the edges of branch, going from the root to this node, we obtain the tree of proof for QHQ_n in the system E . As the formula QHQ_n is balanced, then all conjunct, labeled to nodes of constructed tree are different from each other, therefore, we have

$$t^E(n) \leq 2^{\frac{1}{2}(n+1)(n+2)} \quad \text{and} \quad t^E(n) \geq 2^{\frac{1}{2}(n-1)n}.$$

So, for sufficiently large n $\log_2 t^E(n) = \Omega(n^2)$ and $\log_2 t^E(n) = O(n^2)$.

It is not difficult to prove, that the largest branch in above constructed s.t. is n^2 , therefore, the maximal number of literals in 1-determinative conjunct for QHQ_n is also n^2 , and hence the maximal size of formulas labeled to leafs of tree proof in E is no more, than $2n^2 - 1$. So, we have

$$l^E(n) \leq 2^{\frac{1}{2}(n+1)(n+2)}(2n^2 - 1) \quad \text{and} \quad l^E(n) \geq 2^{\frac{1}{2}(n-1)n}.$$

So, for sufficiently large n we have $\log_2 l^E(n) = \Omega(n^2)$ and $\log_2 l^E(n) = O(n^2)$.

For the *width* we have trivial bounds $n^2 - 1 \leq w^E(n) \leq 2n^2 - 1$.

So, for sufficiently large n $\log_2 w^E(n) = \Omega(\log_2 n)$ and $\log_2 w^E(n) = O(\log_2 n)$.

As for the *space* it is proved in [8] that for the sequence of tautologies $\varphi(k)$ in k variables $s_\varphi^E = O(k^2)$, therefore, since the number of different variables in QHQ_n is $\frac{n^2(n+1)}{2}$, for sufficiently large n we have $s^E(n) = O(n^6)$. Using the fact that at least two determinative conjunct with the result of its e -rule must be in every E -proof, we obtain $s^E(n) \geq 2(n^2 - 1) + n^2 - 2 \geq 3n^2 - 4$.

So, we have $\log_2 s^E(n) = \Omega(\log_2 n)$ and $\log_2 s^E(n) = O(\log_2 n)$.

So, it is proved the following

Theorem 2. *Let n be sufficiently large, then for lines, size, space, and width complexities of sequences QHQ_n in the system E the following holds:*

$$\begin{aligned} \log_2 t^E(n) &= \theta(n^2), \quad \log_2 l^E(n) = \theta(n^2), \\ \log_2 s^E(n) &= \theta(\log_2 n) \quad \text{and} \quad \log_2 w^E(n) = \theta(\log_2 n). \end{aligned}$$

I am grateful to my students A. Balyan and H. Azizyan for very helpful remarks in the process of QHQ_n derivation in the system GS .

Received 25.04.2022

Reviewed 23.06.2022

Accepted 01.07.2022

REFERENCES

1. Cook S.A., Reckhow A.R. The Relative Efficiency of Propositional Proof Systems. *Journal of Symbolic Logic* **44** (1979), 36–50.
<https://doi.org/10.2307/2273702>
2. Strasburger L. Extension without Cut. *Annals of Pure and Applied Logic* **163** (2012), 1995–2007.
3. Chubaryan A. Relative Efficiency of Some Proof Systems for Classical Propositional Logic. *Proceedings of NASA RA* **37** (2002); *Journal of CMA (AAS)* **37** (2002), 71–84.
4. Chubaryan An., Chubaryan Arm. Bounds of Some Proof Complexity Characteristics in the System of Splitting Generalization. *Otechestv. Nauka v Epokhu Izmeneniy* **10** (2015), 11–14 (in Russian).
5. Chubaryan A., Hovhannisyan S., Gasparyan G. About Some Properties of a Propositional System of Generalized Splittings. *Vestnik RAU* **2** (2019), 34–42 (in Russian).

6. Chubaryan A., Hovhannisyan S., Gasparyan G. *Comparison of Two Propositional Proof Systems by Lines and by Sizes*, *ASL, ESM*. Logic Colloquium-2021. Book of Abstracts. Poznan (2021), 166 p.
7. Filmus Y., Lauria M., Nordstrom J., Thapen N., Ron-Zewi N. *Space Complexity in Polynomial Calculus*. 2012 IEEE Conference on Computational Complexity (CCC) (2012), 334–344.
8. Chubaryan A., Mnatsakanyan A. On the Bounds of the Main Proof Measures in Some Propositional Proof Systems. *Scholar Journal of Phis. Math. and Stat.* **1** (2014), 111–117.

Ա. Ա. ՉՈՒԲԱՐՅԱՆ

ԱՍՈՒՅԹԱՅԻՆ ՆԱԵՎԻ ՈՐՈՇ ՆԱՄԱԿԱՐԳԵՐՈՒՄ ԲԱՐԱՆՍԱԿՈՐՎԱԾ ԲԱՆԱԶԵՎԵՐԻ ՄԻ ԴԱՄԻ ԱՐՏԱԾՈՒՄՆԵՐԻ ԲԱՐԴՈՒԹՅՈՒՆՆԵՐԻ ՎԵՐԱԲԵՐՅԱԼ

Սույն աշխատանքում ասույթային հաշվի երկու համակարգերում՝ ընթանալից արտահանված փրոհման մեթոդի վրա հիմնված համակարգում և որոշիչ դիզյունկարիվ նորմալ ձևի վրա հիմնված համակարգում ուսումնասիրվել են բալանսավորված բանաձևերի մի դասի համար արտածումների բարդությունների չորս բնութագրիչների արժեքները: Բոլոր ուսումնասիրված մեծությունների համար ստացվել են լոգարիթմական սանդղակով միևնույն կարգի վերին և ստորին գնահատականներ:

А. А. ЧУБАРИАН

О СЛОЖНОСТЯХ ВЫВОДОВ ОДНОГО КЛАССА БАЛАНСИРОВАННЫХ ФОРМУЛ В НЕКОТОРЫХ ПРОПОЗИЦИОНАЛЬНЫХ СИСТЕМАХ

В настоящей работе для одного класса балансированных формул исследованы четыре сложностные характеристики выводов в двух пропозициональных системах: в системе, основанной на обобщенном методе расщепления, и в системе, основанной на определяющей дизъюнктивной нормальной форме. Для всех исследуемых величин получены одинаковые по порядку (по логарифмической шкале) верхние и нижние оценки.