PDMI PREPRINT — 17/1998

UPPER BOUND ON THE HEIGHT OF TERMS IN PROOFS WITH CUTS

Boris Konev[1]

Russian Academy of Sciences
Steklov Institute of Mathematics
St. Petersburg Department
27 Fontanka, St. Petersburg, 191011, Russia
E-mail: konev@pdmi.ras.ru

August 30, 1998

Abstract

We describe an upper bound on the heights of terms occurring in a most general unifier of a system of pairs of terms that contains unknowns of two types. An unknown belongs to the first type if all occurrences of this unknown have the same depth; we call such unknown an unknown of the cut type. Unknowns of the second type (unknowns of not the cut type) are unknowns that have arbitrary occurrences. We bound from above the heights of terms occurring in a most general unifier in terms of the number of unknowns of not the cut type and of the height of the system. This bound yields an upper bound on the sizes of proofs in the Gentzen sequent calculus LK. Namely, we show that one can transform a proof $\mathcal{D}$ in LK by substituting some free terms in places of variables in such a way that the heights of terms occurring in the proof may be bounded from above by $\mathrm{ar}\,[\mathcal{D}]^{h_1} \cdot q^-[\mathcal{D}] \cdot h_0$, where $\mathrm{ar}\,[\mathcal{D}]$ is the maximal arity of function symbols occurring in $\mathcal{D}$, h_1 is the maximal height of terms occurring in analysis of the cut rules, h_0 is the maximal height of terms occurring in S and in side formulas of applications of the cut rule, and q^- is the number of analysis of applications of the rules $\to\exists$, $\forall\to$ in $\mathcal{D}$.

[1]Supported by INTAS-RFBR, Project 95-0095

There are two basic ways of measuring the complexity of proofs: to count the number of proof lines and to count the total size of the proof. This study is related to the second approach. Our goal is to estimate the sizes of terms occurring in proofs in the Gentzen-style calculi. For cut-free proofs such bounds were studied by Krajíček and Pudlák [3]. Since Orevkov [4] and Statman [12] independently showed that the height of deduction obtained by eliminating all cuts cannot be bounded from above by a Kalmar elementary function of the length of the original deduction, the application of cut-elimination theorem does not provide any achievements. Related problem has been considered in [8]. A technique of transforming proofs in the Gentzen-style calculus that was described there does it in such a way that the heights of proper terms occurring in the cut rule applications and applications of the rules $\rightarrow \exists$ and $\forall \rightarrow$ were bounded from above in terms of the scheme of the proof and its last sequent. However, there was a restriction that the depths of closed variables occurring in formulas of the cut rule applications had to be zeros. In this paper we surmount this restriction.

The results are based on a reduction to the unification problem. Reductions to unification problems are widely used in investigations of proof structure in predicate calculi and in bounds for such proofs. The reduction to the unification problem is implicit in Parikh's paper [9], it was later developed by Farmer [1, 2], by Krajíček and Pudlák, and by Orevkov [6, 5].

The unification problem for a system Γ of pairs of terms $\{(T_i, \Theta_i)\}$ is a problem to find a substitution of terms for the free variables such that $\sigma(T_i) = \sigma(\Theta_i)$ (we write substitution application on the left). Such a substitution is called a *unifier* for a system Γ. A *most general unifier (mgu)* for Γ is a unifier σ such that any unifier σ' for Γ can be decomposed into $\sigma' = \lambda \circ \sigma$ for some substitution λ. Robinson [11] showed that if a system Γ can be unified there is a most general unifier for it.

Since the unification problem arises in this paper in connection with predicate calculus, we distinguish (object) variables and *unknowns* (metavariables for terms). Henceforth we use the term unknown when referring to the unification problem.

1 Upper bound on the heights of terms in the most general unifier

We use the representation of terms by directed acyclic graphs (DAG) developed by Paterson and Wegman ([10]). In such representation equal subterms are represented by a single subgraph. Given a unifier of a system of pairs of terms one can construct an equivalence relation on the vertices of DAG declaring two vertices to be equivalent if and only if the unifier maps the corresponding subterms into equal terms. In [10] a *valid* equivalence relation has been introduced. Given a valid equivalence relation on vertices of DAGs, one can construct a unifier and to a minimal valid equivalence relation on vertices corresponds a most general unifier corresponds of the system ([10], Lemma 1). We use this correspondence between unifiers and valid equivalences in the rest of the paper.

Suppose a system of pairs of terms $\Gamma = \{(T_i, \Theta_i)\}$ has a most general unifier

σ. Consider a graph $G(\Gamma)$ representing terms of Γ and a minimal valid equivalence relation Σ on the vertices of $G(\Gamma)$. We define a graph $G'(\Gamma)$ whose vertices are the equivalence classes of vertices of the graph G, an edge leads from a vertex C_1 to a vertex C_2 if and only if the class C_1 contains a vertex x_1, the class C_2 contains a vertex x_2 and there is an edge $x_1 \to x_2$ in the graph $G(\Gamma)$. Clearly the graph $G'(\Gamma)$ represents all the terms from the result of application of σ to the terms T_i, Θ_i. If a vertex of G' (i.e., the equivalence class) contains an unknown t we say that the vertex is *labeled* by t. One vertex may be labeled by a variety of unknowns.

Let u be an occurrence of a subterm t' in a term t. We denote the *depth* of the occurrence u in the term t by $h_u[t]$. The maximum of depths of all occurrences of unknowns and function constants to the term t is the *height* $h[t]$ of the term. The height of a most general unifier of a system of terms $\Gamma = \{(T_i, \Theta_i)\}$ is the maximum of the heights of $\sigma(T_i)$, $\sigma(\Theta_i)$. We call an unknown x an unknown of *the cut type* if depths of all occurrences of x in Γ are equal, and an unknown of *not the cut type* otherwise.

Lemma 1 *Let $\Gamma = \{(T_i, \Theta_i)\}$ be a system of pairs of terms. Then there is such a system $\Gamma^* = \{(T_i^*, \Theta_i^*)\}$ that $h[\Gamma^*] = h[\Gamma]$, unknowns of the system Γ^* are unknowns of Γ, most general unifiers of both systems map common unknowns into equal terms and for all unknowns of the cut type of a system Γ^* holds the follows. If a vertex u is labeled by an unknown x there is such a vertex v labeled by an unknown of not the cut type that there is a path in $G'(\Gamma^*)$ from u to v whose length is not greater than $h_x[\Gamma^*]$*

Proof. We introduce a notion of a trace to a subterm in a term. Let T be a term, and t its subterm. A *trace* in the term T to an occurrence u of the term t is a sequence of pairs $\{(f_{i_1}, n_1), (f_{i_2}, n_2), \ldots, (f_{i_l}, n_l)\}$ where f_{i_k} is a N_k-ary function symbol and n_k is the number of an argument of f that leads to the occurrence u of t. The length of the trace is the length of the sequence. Let T be a term and $\mathcal{P}$ be a sequence of pairs of function symbols and integers. We say that $\mathcal{P}$ *determines* a subterm t of T and its occurrence if a trace in T to this occurrence of t is equal to $\mathcal{P}$.

We construct the system Γ^* inductively depending on two parameters—the depth of unknowns of the cut type and on the number N such that the following condition $(*)$ holds.

- $(*)$ For any unknown x of the cut type whose depth is less then N, if a vertex u of the graph $G'[\Gamma]$ is labeled by x then there is such a vertex v labeled by an unknown of not the cut type that there is a path in $G'[\Gamma]$ beginning from v and ending in u of the length at most $h_x[\Gamma]$

We begin our construction from the system Γ. Let N be the minimum of depths of all occurrences of unknowns of the cut type. Then the condition $(*)$ holds since the depths of all unknowns of the cut type are greater than or equal to N.

Let x be such an unknown of the cut type that for any unknown y of the cut type if $h_y[\Gamma] \geq N$ then $h_y[\Gamma] \geq h_x[\Gamma]$. Let u denote an occurrence of x in one of the terms

T_i, Θ_i, say in T_k. Consider the pair (T_k, Θ_k) and a trace $\mathcal{P}$ to the occurrence u in the term T. As the system Γ has a most general unifier there is such a subterm of the term Θ_k that either the trace to this subterm is prefix of $\mathcal{P}$ or $\mathcal{P}$ is prefix of this path. One of the follows holds.

1. The trace in the term Θ_k is shorter than $\mathcal{P}$ and it terminates by an unknown z of *not* the cut type.

2. The trace in the term Θ_k is shorter than $\mathcal{P}$ and it terminates by an unknown w of the cut type. (This condition may not hold for the first system of our construction, that is, for Γ).

3. The length of the trace in Θ_k is greater than or equal to the length of $\mathcal{P}$. Then $\mathcal{P}$ determines a subterm t_u ant its occurrence in Θ_k.

In the first case there is a path in the graph $G'[\Gamma]$ ending in x whose length is not greater than $h_x[\Gamma]$ and that begins from a vertex labeled by an unknown of not the cut type.
In the second case since the depth of the unknown w is less than the depth of x the condition $(*)$ holds for w. That means that there is an unknown of not the cut type z and a path in $G'[\Gamma]$ of the length at most $h_w[\Gamma]$ from a vertex labeled by z to a vertex labeled by w. Hence, there is a path in the graph $G'[\Gamma]$ from z to x of the length at most $h_w[\Gamma]$ plus the difference of depths of w and x that is $h_x[\Gamma]$.
If the third case holds for all occurrences of x in Γ we replace all occurrences of the unknown x by the term t_u and denote the resulting system by Γ'. Since for all occurrences u of x in Γ, terms t_u belongs to one equivalence class, a unifier of Γ' unifiers the system Γ as well. As the difference of depths for the unknown x is zero, $h[\Gamma'] = h[\Gamma]$. The system Γ' has one unknown of the cut type less then the system Γ. We denote by Γ^* the last system of our construction. □

Lemma 2 *Suppose that a system of pairs of terms $\Gamma = \{(T_i, \Theta_i)\}$ contains n unknowns of not the cut type. Let $h_0 = h[\Gamma]$ and h_1 be the maximal depth of unknowns of the cut type. Then the height of a most general unifier for Γ is at most $(\mathrm{ar}\,[\Gamma]^{h_1} + 1) \cdot n \cdot h_0$ where $\mathrm{ar}\,[\Gamma]$ is the maximal arity of function symbols of Γ.*

Proof. We use a bound on height of a most general unifier by Orevkov [8]:

$$h[\sigma] \leq N \cdot h_0$$

where N is the total number of unknowns of the system Γ. We construct a system Γ^* using Lemma 1 and estimate the total number of unknowns of the system Γ^*. Consider a graph $G'[\Gamma^*]$. By Lemma 1, for any unknown of the cut type there is such a vertex u of $G'[\Gamma^*]$ labeled by an unknown of not the cut type that the length of the path from u to x is not greater than $h[\Gamma^*]$. Thus we can estimate the number of unknowns of the cut type of Γ^* by counting the number of vertices of $G'[\Gamma^*]$ that are at the distance of at most h_1 from a vertex labeled by an unknown of not the cut type. There can be at most $\mathrm{ar}\,[\Gamma]^{h_1}$ such unknowns. Hence, the total number of unknowns of Γ^* does not exceed $\mathrm{ar}\,[\Gamma]^{h_0} \cdot n + n$. It follows the bound from the condition of the lemma. $\square$

2 Upper bounds on the height of terms in proofs

We transform proofs using its scheme, by which we mean a sequence of analysis of the applications of rules and axioms. In the analysis of a logical axiom we indicate the index numbers of the occurrences of the principal formula of the axiom in the succedent and the antecedent, in the analysis of the application of a rule we indicate the index number of the rule, the index numbers of the premises and the index numbers of the occurrences of all the formulas which play an active role in the application of the rule. Orevkov [7] and Krajíček and Pudlák [3] showed that the problem of finding a proof of a sequent given its scheme is unsolvable. In [8] a notion of *deduction scheme with fixed cut types* were introduced. To any analysis of an application of the cut rule a formula is added in such a way that the added formulas have no free variables in common. We call the added formula a *type* of the analysis of the cut rule. Let S be a sequent, $t_1, \ldots, t_n$ a list of pairwise distinct object variables, U a deduction scheme with fixed cut types, $\mathcal{D}$ a proof. We say that $\mathcal{D}$ is a $(t_1, \ldots, t_n)$-*deduction of S in accordance with U*, if the following conditions hold.

1. $\mathcal{D}$ is a deduction in accordance with the scheme obtained from U by deleting all formulas.

2. The type of the analysis of any application of the cut rule in $\mathcal{D}$ is transformable into its side formula by substituting terms in place of free variables.

3. There are terms $T_1, \ldots, T_n$ such that T_i is free for t_i in S for all i, and the last sequent of $\mathcal{D}$ is $[S]^{t_1,\ldots,t_n}_{T_1,\ldots,T_n}$.

A deduction scheme is correct if there exists a deduction in accordance with this scheme.

A proof is called a *universal* $(t_1, \ldots, t_n)$-deduction of a sequent S in accordance with a scheme U if $\mathcal{D}$ is a $(t_1, \ldots, t_n)$-deduction of S in accordance with U and any $(t_1, \ldots, t_n)$-deduction of S in accordance with U can be obtained by substituting terms for free occurrences of variables.

Consider the unification problem for a system of pairs of terms *with constraints* (CT-systems for short). Terms of such system are constructed from *unknowns* (metavariables for terms) $t_1, \ldots, t_n$, function symbols $f_1, \ldots, f_m$ and object variables $x_1, \ldots, x_p$. The system Ξ itself consists of two parts $\Xi = \Gamma \cup \Pi$, the first is an ordinary unification problem

$$\Gamma = \{(T_i, \Theta_i)\}_{i=1}^N$$

and the second is a set of constraints of the following type

$$\Pi = \{x_i \notin R_i\}_{i=1}^M$$

where T_i,Θ_i,R_i are terms and x_i are object variables. We call a mapping σ a *solution* of a system $\{\Gamma;\Pi\}$ if its unifiers Γ and $\sigma(R_i)$ do not contain x_i for all i: $1 \leq i \leq M$. We have following trivial lemma.

Lemma 3 *If there exists a unifier for Γ which satisfies a set of conditions Π, then any most general unifier for Γ satisfies the conditions too.*

We construct a CT-system $\Xi\{U;S;t_1,\dots,t_n\}$ given a $(t_1,\dots,t_t)$-proof of a sequent S by induction on $h(U)$ similarly to the construction of such a system in [8]. In fact we only change the original construction for axioms. Let $\Pi\{S;t_1,\dots,t_n\}$ denote the list of all expressions $x \notin t_j$, $1 \leq j \leq n$, where x is an object variable occurring in a quantifier complex governing some free occurrence of t_i in S.

If $h(U) = 0$, the proof consists of a single axiom, the only sequent in the proof is $\Delta_1, A, \Delta_2, \to \Delta_3, A, \Delta_4$ for a principal formula A. The analysis of an axiom contains indices of two principal occurrences of A in the axiom, say n_1 and n_2. Let A_1 denote a formula whose occurrence in S has index n_1 and A_2 the formula whose occurrence in S has index n_2. It is not hard to see that A_1 and A_2 can be transformed by replacing certain free unknowns with terms into A.

We construct a set Γ of pairs of terms Γ by induction on the construction of the formula A as follows. If A is atomic,

$$A_1 = P(T_1,\dots,T_s), \qquad A_2 = P(\Theta_1,\dots,\Theta_s)$$

where P is a s-ary predicate symbol, we consider a set of pairs of terms $\Gamma = \{(T_i,\Theta_i)\}_{i=1}^{s}$. If $A = A' \odot A''$ where $\odot$ is one of &, $\vee$, $\supset$, we construct two sets Γ' for A_1', A_2' and Γ'' for A_1'', A_2'' inductively and let $\Gamma = \Gamma' \cup \Gamma''$. If $A = Qx\ A'$ where Q is one of the quantifiers or $A = \neg A'$ we construct a set of pairs of terms Γ' for A' and let $\Gamma = \Gamma'$. Finally we define $\Xi\{U;S;t_1,\dots,t_n\} = \{\Gamma, \Pi\{S;t_1,\dots,t_n\}\}$.

For $h(U) > 0$ the proof ends with the application of an l-premise rule L ($1 \leq l \leq 2$). Let U_i denote the deduction schema occurring in U above the i-th premise of the last analysis. Suppose that L introduces a propositional connective. Let $S_1,\dots S_l$ be the sequents from which S is obtained by L. We define CT-system $\Xi\{U;S;t_1,\dots,t_n\}$ to be the union of CT-systems $\Xi\{U_1;S_1;t_1,\dots,t_n\},\dots,\ \Xi\{U_l;S_l;t_1,\dots,t_n\}$ and list $\Pi\{S;t_1,\dots,t_n\}$, first renaming the unknowns for terms in these CT-systems so that they will have no common unknowns other than $t_1,\dots,t_n$.

Suppose that L is the cut rule. Then it has two premises. Let A be the type of the last analysis in U and V a list of all free variables of A. Let S_1 and S_2 be sequents from which S follows by the cut rule with A as a side formula. We define CT-system $\Xi\{U;S;t_1,\dots,t_n\}$ to be the union of $\Xi\{U_1;S_1;t_1,\dots,t_n,V\}$ and $\Xi\{U_2;S_2;t_1,\dots,t_n,V\}$, with variables for terms renamed so that they have no common unknowns in the list $t_1,\dots,t_n,V$.

Suppose that L is the rule $\to \forall$ or $\exists \to$. Then $l = 1$. Let b be a variable not occurring in S and distinct from $t_1,\dots,t_n$. Let S_1 be a sequent from which S follows by L with b as proper variable. As $\Xi\{U;S;t_1,\dots,t_n\}$ we take the CT-system

$$\left\{\begin{array}{l} \Xi\{U_1;S_1;t_1,\dots,t_n\} \\ \Pi\{S;t_1,\dots,t_n\} \\ b \notin t_{i_1},\dots,b \notin t_{i_r} \end{array}\right.$$

where $t_{i_1}, \dots t_{i_r}$ is the list of all variables $t_1, \dots, t_n$ that occur free in S.

Suppose that L is the rule $\forall \rightarrow$ or $\rightarrow \exists$. Then $l = 1$. Let t be a variable not occurring in S and distinct from $t_1, \dots, t_n$. Let S_1 be a sequent from which S follows by L with t as proper variable. As $\Xi\{U; S; t_1, \dots, t_n\}$ we take the union of $\Xi\{U_1; S_1; t_1, \dots, t_n, t\}$ and the list of constraints $\Pi\{S; t_1, \dots, t_n\}$.

Lemma 4 *For any correct deduction scheme U in* LK *with fixed cut types, any sequent S, if there exists a $(t_1, \dots, t_n)$-deduction of S in accordance with U such that, for any term T that is a proper term of $\mathcal{D}$, $\rightarrow \exists$, $\forall \rightarrow$, cut, the following inequality holds:*

$$h[T] \leq (\mathrm{ar}[\mathcal{D}]^{h_1}) \cdot (q^- + n) \cdot h_0,$$

where $\mathrm{ar}[\mathcal{D}]$ *is the maximal arity of function symbols occurring in S and in the analysis of the cut rules, h_1 is the maximal height of terms occurring in analysis of the cut rules, h_0 is the maximal height of terms occurring in S and in side formulas of applications of the cut rule, and q^- is the number of analysis of applications of the rules $\rightarrow \exists$, $\forall \rightarrow$ in the scheme U.*

Proof. It is not hard to see by induction on $h[U]$ that a proof $\mathcal{D}$ in LK is a $(t_1, \dots, t_n)$-deduction of a sequent S in accordance with a schema U if and only if the list of proper terms of $\mathcal{D}$ and of proper terms of applications of the cut rules, $\rightarrow \exists$, $\forall \rightarrow$, is a solution of $\Xi\{U; S; t_1, \dots, t_n\}$ and universal solution generates a universal $(t_1, \dots, t_n)$-deduction of S in accordance with U.

While we construct a system Ξ we enlarge the set of n unknowns $t_1, \dots, t_n$ when we consider the cut rule and the rules $\forall \rightarrow$ and $\rightarrow \exists$. The differences of depths of unknowns introduced by the cut rule are zeros. The total number of unknowns of not the cut type is $n + q^-$. Thus we can apply Lemma 2 to bound the heights of a universal decision of the system Ξ. □

Theorem 1 *Let $\mathcal{D}$ be a proof of a sequent S in* LK*. Then, by replacing certain free terms by variables, we can transform $\mathcal{D}$ into a proof $\mathcal{D}^*$ of S in* LK*, such that the height of the proper terms of applications of the rules $\rightarrow \exists$ and $\forall \rightarrow$ in $\mathcal{D}^*$ and the height of the terms occurring free in the side formulas of applications of the cut rule in $\mathcal{D}^*$ is at most*

$$\mathrm{ar}[\mathcal{D}]^{h_1} \cdot q^-[\mathcal{D}] \cdot h_0$$

Proof. The proof follows [8]. Let $\mathcal{D}$ be a proof in LK and U the deduction scheme of $\mathcal{D}$. For any formula A consider a formula A^* such that A^* does not contain 0-place function symbols, any object variable has at most one free occurrence in A^*, if a term t occurs free in A^*, then t is a variable and A^* can be transformed into A by replacing free variables by terms. If A is a side formula of the cut rule application, we add a formula A^* to the analysis, renaming, if necessary, free variables in the types of the analysis. Denote the resulting deduction scheme with fixed cut types by U^*. It is clear that $\mathcal{D}$ is a ()-deduction of S in accordance with U^*. □

References

[1] W. M. Farmer. *Length of proofs and unification theory.* PhD dissertation, Univ. of Wisconsin—Madison, 1984.

[2] W. M. Farmer. A unification-theoretic method for investigating the k-provability problem. *Annals of Pure and Applied Logic*, 51:173–214, 1991.

[3] J. Krajíček and P. Pudlák. The number of proof lines and size of proofs in first logic. *Arch. Math. Logic*, 27:69–84, 1988.

[4] V. P. Orevkov. Lower bounds on the increase in complexity of deductions in cut elimination. *Zap. Nauchn. Sem. Leningrad Otdel. Mat. Inst. Steklov. (LOMI)*, 88:137–162, 1979.

[5] V. P. Orevkov. Applications of cut elimination to estimates for lengths of proofs. *Dokl. Akad. Nauk SSSR*, 296(3):539–542, 1987. English transl. in Soviet Math Dokl. **36** (1988).

[6] V. P. Orevkov. Reconstruction of a proof from its scheme. *Dokl. Akad. Nauk SSSR*, 293(2):313–316, 1987. English transl. in Soviet Math Dokl. **35** (1987).

[7] V. P. Orevkov. Proof schemata in axiomatic Hilbert-type theories. *Zap. Nauchn. Sem. Leningrad Otdel. Mat. Inst. Steklov. (LOMI)*, 176:132–146, 1988. English transl. in J. Soviet Math **55** (191), no.3.

[8] V. P. Orevkov. *Complexity of Proofs and Their Transformations an Axiomatic Theories.* Translations of Mathematical Monographs. American Mathematical Society, 1991.

[9] R. J. Parikh. Some results on the length of proofs. *Trans. Amer. Math. Soc.*, 117:29–36, 1973.

[10] M. S. Paterson and M. N. Wegman. Linear unification. *Journal of computer and system science*, 16:168–167, 1978.

[11] J. A. Robinson. A machine oriented logic based on the resolution principle. *J. Assoc. Comput. Mach*, 12:23–41, 1965.

[12] R. Statman. Bounds for proof-search and speed-up in the predicate calculus. *Ann. Math. Logic*, 15(3):225–287, 1978.