A Note on the Lengths of $\mathbf{G}_0$ Proofs

Stephen Cook

December 21, 2004

Here the system $\mathbf{G}_0$ is a sequent-based proof system for the Quantified Propositional Calculus (QPC) as defined in [1] and [3]. It is based on the system $\mathbf{G}$ defined by Krajicek and Pudlak [2], but it is restricted so that all cut formulas must be quantifier-free, and "target formulas" for $\exists$-right and $\forall$-left must also be quantifier-free.

The system $\mathbf{G}_0$ is a complete proof system for QPC formulas. We are interested in proving an upper bound on the number of lines in a $\mathbf{G}_0$ proof of a valid sequent S in terms of the length $|S|$ of S.

One straightforward approach is to proceed by induction on the total number of connectives $\wedge, \vee, \neg, \forall, \exists$ in S, since G_0 has left and right introduction rules for each of these connectives. This works fine except for formulas $\exists x\varphi(x)$ on the RHS (succedent) or formulas $\forall x\varphi(x)$ on the LHS (antecedent). In the case $\exists x\varphi(x)$, we assume that the sequent $\Gamma \to \Delta, \exists x\varphi(x)$ is valid, and the idea would be to find a quantifier-free formula A such that the sequent $\Gamma \to \Delta, \varphi(A)$ is valid. Then the original sequent can be derived from this by the $\exists$-right rule.

If $p_1 \dots p_n = \vec{p}$ is a list of the free variables in the original sequent then we can think of the formula $A(\vec{p})$ as specifying a Skolem function to replace the existentially quantified variable x in $\exists x\varphi(x)$. This function is a Boolean function of n variables, and in general it requires a formula $A(\vec{p})$ of size $O(2^n)$ to represent it. The formula $\varphi(A)$ could be much bigger than the formula $\exists x\varphi(x)$, but it has reduced quantifier complexity, so with a suitable induction hypothesis we may assume that $\Gamma \to \Delta, \varphi(A)$ has a G_0 proof of a certain length, hence derive a G_0 proof of $\Gamma \to \Delta, \exists x\varphi(x)$.

This argument can be carried out, but it leads to a larger upper bound on proof lengths than necessary. Instead, we use a different idea. Observe that if the sequent $\Gamma \to \Delta, \exists x\varphi(x)$ is valid and has no free variables, then one of the two shorter sequents $\Gamma \to \Delta, \varphi(\mathbf{T})$ or $\Gamma \to \Delta, \varphi(\mathbf{F})$ is also valid, and the original sequent can be derived from either of these by the $\exists$-right rule.

If $\Gamma \to \Delta, \exists x\varphi(x)$ does have free variables $p_1 \dots p_n$, then for each truth assignment τ to $\vec{p}$ we can add each p_i either to the left or right of $\Gamma \to \Delta, \exists x\varphi(x)$ to determine the truth assignment τ to $\vec{p}$. For example, the sequent

$$S \equiv\ p_1, p_2, \Gamma \to \Delta, \exists x\varphi(x), p_3$$

determines the truth assignment τ which assigns True to p_1 and p_2, and False to p_3. Thus every truth assignment except τ trivially satisfies S, but τ satisfies S iff τ satisfies the original sequent $\Gamma \to \Delta, \exists x\varphi(x)$.

This idea leads to the following definition.

Definition: We say that the free variables in a sequent S are *determined* if every free variable in S (also) occurs as a separate atomic formula (either in the succedent or antecedent).

Proposition: If the free variables in S are determined, then there is at most one truth assignment τ to these free variables that fails to satisfy S without regard to the nonatomic formulas of S.

We are now ready to state our main result. In counting lines in a $\mathbf{G}_0$ proof, we ignore lines resulting from the exchange and weakening rules.

Theorem: Let S be a valid sequent in the language of $\mathbf{G}_0$. Then S has a $\mathbf{G}_0$ proof with $O(2^{|S|+n})$ lines, not counting weakenings and exchanges. Here $|S|$ is the total number of occurrences of $\wedge, \vee, \neg, \forall, \exists$ in S, and n is the number of distinct free variables in S. Each line has length $O(\mathit{length}(S)$, and all cut formulas are atomic.

Note that if we counted weakenings and exchanges, the upper bound would increase by a factor of $\mathit{length}(S)$.

We can actually state a sharper result. Let $|S|_g$ be the number of occurrences of "general connectives" in S, and let $|S|_r$ be the number of occurrences of "restricted connectives" and $\neg$ in S. Here the general connectives are $\wedge$ and $\forall$, provided they occur positively on the RHS of S, or negatively on the LHS, and the general connectives are $\vee, \exists$ provided they occur negatively in the RHS or positively in the LHS. The restricted connectives are the dual; for example $\vee, \exists$ occurring positively on the RHS.

Lemma: If S is a valid sequent with its free variables determined, then S has a $\mathbf{G}_0$ proof with

$$O(|S|_r \cdot 2^{|S|_g})$$

lines (not counting weakenings and exchanges). Each line has length $O(\mathit{length}(S))$, and all cuts are atomic.

Corollary: Every valid sequent S has a proof with

$$O(|S|_r \cdot 2^{|S|_g+n})$$

lines, where n is the number of distinct free variables in S. The proof satisfies the conditions of the Lemma.

Notice that the Theorem follows from the Corollary.

Proof of Corollary: For each of the 2^n truth assignments τ to the free variables of S, let S^τ be the result of adding atoms to S to determine τ (i.e. true atoms on the LHS and false atoms on the RHS). Each S^τ is valid and determined, and so by the Lemma has a proof with $O(|S|_r \cdot 2^{|S|_g})$ lines. By combining these sequents with 2^{n-1} atomic cuts, S can be derived. $\square$

Proof of Lemma: We proceed by induction on the total number of connectives $\wedge, \vee, \neg, \forall, \exists$

in S. The base case is trivial: any valid sequent with no connectives is a subsequent of an axiom. (We do allow the constants $\mathbf{T}$ and $\mathbf{F}$, and the axioms $\rightarrow \mathbf{T}$ and $\mathbf{F} \rightarrow$.)

For the induction step, we have a case for each of the connectives $\wedge, \vee, \neg, \forall, \exists$. We consider the case of producing formulas in the RHS, since the case of the LHS is dual. The cases of $\neg$ and the restricted connective $\vee$ are easy. For example, if $\Gamma \rightarrow \Delta, \neg\varphi$ is determined and valid, then so is the shorter sequent $\Gamma, \varphi \rightarrow \Delta$. The induction hypothesis yields a proof of the latter, and the former can be derived by the $\neg$-right rule.

Now consider the case of the general connective $\wedge$. Suppose that S is $\Gamma \rightarrow \Delta, \varphi_1 \wedge \varphi_2$, and this sequent is determined and valid. Then each of the shorter sequents $\Gamma \rightarrow \Delta, \varphi_1$ and $\Gamma \rightarrow \Delta, \varphi_2$ is determined and valid, and the induction hypothesis applies. Combining these two proofs with the $\wedge$-right rule gives us S. The length of the resulting proof satisfies the claim in the Lemma, using the fact that

$$|\varphi_1 \wedge \varphi_2|_g = |\varphi_1|_g + |\varphi_2|_g + 1$$

Now suppose that S is $\Gamma \rightarrow \Delta, \forall x\varphi(x)$, and S is determined and valid. Then each of the two sequents $p, \Gamma \rightarrow \Delta, \varphi(p)$ and $\Gamma \rightarrow \Delta, \varphi(p), p$ is determined and valid, where p is a new variable. Hence each has a $\mathbf{G}_0$ proof with $O(|S|_r \cdot 2^{|S|_g - 1})$ lines. From these sequents, S can be derived with a cut on p and the $\forall$-right rule. The resulting proof has

$$O(|S|_r \cdot 2^{|S|_g}) + 2 = O(|S|_r \cdot 2^{|S|_g})$$

lines, as required.

Finally, suppose that S is $\Gamma \rightarrow \Delta, \exists x\varphi(x)$, and S is determined and valid. It follows from the **Proposition** that one of the two shorter sequents $\Gamma \rightarrow \Delta, \varphi(\mathbf{T})$ and $\Gamma \rightarrow \Delta, \varphi(\mathbf{F})$ is valid, and hence has a suitable $\mathbf{G}_0$ proof. Then S can be derived from either of these by the $\exists$-right rule. □

References

[1] Stephen Cook and Tsuyoshi Morioka. Quantified Propositional Calculus and a Second-Order Theory for NC^1. *Archive for Mathematical Logic*, 2004. to appear.

[2] Jan Krajíček and Pavel Pudlák. Quantified Propositional Calculi and Fragments of Bounded Ar ithmetic. *Zeitschrift f. Mathematickal Logik u. Grundlagen d. Mathemati k*, 36:29–46, 1990.

[3] Tsuyoshi Morioka. *Logical Approaches to the Complexity of Search Problems: Proof Complexity, Quantified Propositional Calculus, and Bounded Arithmetic.* PhD thesis.