Proving soundness for the quantified propositional calculus $\mathbf{G}_i^\star$

Phuong Nguyen *
University of Montreal

November 3, 2011

Abstract

The links between Buss's hierarchy $\mathbf{S}_2$ and Krajíček–Pudlák's quantified propositional proof system $\mathbf{G}$ were shown by Steven Perron to be stronger than what had been known before. Perron's theorem states that $\mathbf{S}_2^i$ proves the soundness of the treelike system $\mathbf{G}_i^\star$ with respect to $\mathbf{\Sigma}_{i+1}^q$ tautologies. Previously this was known only for $\mathbf{\Sigma}_i^q$ tautologies. The theorem is proved by first proving a kind of Herbrand theorem for $\mathbf{G}_i^\star$. Perron statement of the latter theorem contains a subtle error which does not affect the overall correctness of his main result. To fix this error turns out to require some new subsystems of $\mathbf{G}$ that we call $\mathbf{G}_{\mathcal{B},i}^\star$: for each i, $\mathbf{G}_{\mathcal{B},i}^\star$ is the subsystem where the cut formulas are allowed to be boolean combinations of $\mathbf{\Sigma}_i^q$ and $\mathbf{\Pi}_i^q$ formulas (instead of being in $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$ as it is the case for $\mathbf{G}_i^\star$). In this paper we discuss the Herbrand theorems for these systems, and observe that when the boolean combinations in question are just conjunctions or disjunctions (of arbitrary arity), then $\mathbf{G}_{\mathcal{B},i}^\star$ is p-equivalent to $\mathbf{G}_i^\star$ for proving $\mathbf{\Sigma}_{i+1}^q$ tautologies. This observation follows from a slight extension of Perron's theorem. However, such an equivalence is not known if the boolean combinations have depth two or more. Finally, and independently, we present a "faithful" formalization of truth definition in $\mathbf{V}^i$ for a restricted class of quantified boolean formulas.

1 Introduction

The sequent calculus $\mathbf{G}$ and its subsystems $\mathbf{G}_i$, $\mathbf{G}_i^\star$, developed in [KP90, KT92, Kra95], are proof systems for quantified propositional formulas. $\mathbf{G}_i$ and $\mathbf{G}_i^\star$ are associated with the first-order theories $\mathbf{TV}^i$ and $\mathbf{V}^i$ (two-sorted versions of Buss's theories $\mathbf{T}_2^i$ and $\mathbf{S}_2^i$ [Bus86]): first-order proofs are translated, by the Paris-Wilkie translation, to families of polynomial-size proofs in the associated subsystems, and the soundness of the subsystems is provable in the associated

*This work is supported by an NSERC postdoctoral fellowship carried out at McGill University. Email: pnguyen@cs.toronto.edu

theories. In this paper we discuss some issues related to proving the soundness of these systems.

In particular, it is known that the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_i^q$ formulas is provable in $\mathbf{V}^i$ [KT92, Kra95] (see [CN10, Theorem X.2.17] for the outline of a different proof). Originally $\mathbf{G}_i$ and $\mathbf{G}_i^\star$ are defined to prove only tautologies in $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$; under the revised definition of $\mathbf{G}_i$ and $\mathbf{G}_i^\star$ by Cook and Morioka [CM05] we can now talk about the soundness of these systems with respect to proving formulas of higher complexity, e.g., $\mathbf{\Sigma}_{i+1}^q$, $\mathbf{\Sigma}_{i+2}^q$, etc. The provability of these systems' soundness then becomes more interesting, and this has been investigated by Steven Perron in [Per08a, Per08b]. He proves, in particular, that the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_{i+1}^q$ formulas is provable in $\mathbf{V}^i$. This implies that $\mathbf{V}^i$ can be axiomatized using an axiom formulating the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_{i+1}^q$ formulas, over the base theory $\mathbf{V}^1$ [Per08a] [Per08b, Chapter 5]. By combining this with a simple, but perhaps surprising, result [Per08a, Lemma 7.1] (see also [CN10, Theorem X.2.27]) he also shows that proving the soundness of $\mathbf{G}_i^\star$, or even of cut-free $\mathbf{G}^\star$, with respect to $\mathbf{\Sigma}_j^q$ formulas, where $j \geq i+2$, already requires $\mathbf{V}^{j-1}$.

Perron gave a detailed proof of the special case where the soundness is stated for prenex formulas in [Per08a, Theorem 6.1] (reproduced in [Per08b, Theorem 5.1.1]). The general case is proved as Theorem 5.1.2 of [Per08b]. The current author's original motivation in this work is to understand the latter proof. Along the way, we identify some new subsystems of $\mathbf{G}$ that are required to correctly state Perron's Herbrand theorem for $\mathbf{G}_i^\star$ which is needed for his main result: for each i, $\mathbf{G}_{\mathcal{B},i}$ is the subsystem of $\mathbf{G}$ where cut formulas are in $\mathcal{B}_i$, the set of boolean combinations of $\mathbf{\Sigma}_i^q$ and $\mathbf{\Pi}_i^q$ formulas ($\mathcal{B}$ stands for boolean combination). It follows from several results in the area that $\mathbf{G}_{\mathcal{B},i}$ and $\mathbf{G}_i$ are p-equivalent for proving $\mathbf{\Sigma}_i^q$ formulas (or even $\mathcal{B}_i$ formulas). On the other hand, the situation for $\mathbf{G}_{\mathcal{B},i}^\star$ and $\mathbf{G}_i^\star$ is much less clear, and we conjecture that $\mathbf{G}_{\mathcal{B},i}^\star$ is p-equivalent to $\mathbf{G}_i^\star$ (for proving $\mathbf{\Sigma}_i^q$ formulas). We observe that this holds for depth-1 $\mathbf{G}_{\mathcal{B},i}^\star$, the subsystem of $\mathbf{G}_{\mathcal{B},i}^\star$ where the cut formulas are disjunctions or conjunctions of $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$ formulas, i.e., where the boolean combinations have depth 1. Our observation follows from a slight extension of Perron's theorem which is proved by extending the Herbrand theorem for $\mathbf{G}_i^\star$. We will also prove a Herbrand theorem for $\mathbf{G}_{\mathcal{B},i}^\star$, but this does not allow us to settle the conjecture.

Perron's technically involved proof of the general case [Per08b, Theorem 5.1.2] contains a (nonessential) gap which can be fixed, for example, using our notion of *direct variable* (see Definition A.2 and the paragraph above it). For completeness, we will reproduce, with necessary modifications, Perron's proof in the appendix.

The usual way of formalizing truth that we know of suffers from a subtle problem which arises when provability in theories not containing the Replacement axioms is concerned.[1] Informally, the problem is in the formalization of

[1] I am indebted to an anonymous referee who points this problem out and also gives the example that I use here.

truth definitions for non-prenex formulas. Consider, for example, formalizing truth definition for a given $\mathbf{\Sigma}_1^q$ formula. This is usually done by stating collectively the existence of all existentially quantified variables. But this is the same as implicitly assuming the Replacement axioms for $\mathbf{\Sigma}_1^B$ formulas, and it might be problematic somewhere else if we work in a theory that does not prove these axioms. This fortunately does not affect Perron's theorem (note that for $\mathbf{V}^1$ Perron's theorem is about the Σ_2^q formulas, so the soundness principle is $\mathbf{\Sigma}_2^B$, and that as far as we know $\mathbf{V}^1$ does not prove the Replacement axioms for this class.) Nevertheless, we think that the problem is worth discussing, and we will give a short discussion in Section 2.3.

1.1 Proving soundness of $\mathbf{G}_i^\star$ for $\mathbf{\Sigma}_{i+1}^q$ formulas

To prove the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_{i+1}^q$ formulas is to prove that any $\mathbf{\Sigma}_{i+1}^q$ formula A that has a $\mathbf{G}_i^\star$-proof is valid. Let π denote the proof of A. One way is to extract from π the witnessing values for the outermost existential variables in A. Perron's idea is to do so by first transforming π into a proof π' where there are explicit defining formulas for the variables to be witnessed. This proof involves extension variables, similar to the extended Frege system. (In fact, such a transformation generalizes the one used for showing that extended Frege p-simulates $\mathbf{G}_1^\star$.) The transformation gives a so-called Herbrand theorem for $\mathbf{G}_i^\star$. Perron then proceeds to show how to compute the witnesses from the sequent given by the Herbrand theorem. Furthermore, in order to show that the principle is provable in $\mathbf{V}^i$, these steps need to be formalized in $\mathbf{V}^i$.

Both steps are technically involved. The first step gives a kind of KPT theorem for quantified propositional proof systems. The original KPT theorem, named after the authors of [KPT91], is a special form of Herbrand's theorem and is stated for first-order theories of bounded arithmetic. (For a survey on Herbrand's theorem, see [Bus95a].) The KPT theorem states, for example, that for any $\mathbf{\Sigma}_2^B$ theorem $\exists Y \forall Z \varphi(X, Y, Z)$ of a universal theory $\mathcal{T}$ (here implicitly X is universally quantified), there are a constant k and k terms

$$T_1(X), T_2(X, Z_1), \ldots, T_k(X, Z_1, \ldots, Z_{k-1}),$$

such that $\mathcal{T}$ proves

$$\forall X \forall Z_1 \forall Z_2 \ldots \forall Z_k \bigvee_{i=1}^{k} \varphi(X, T_i(X, Z_1, \ldots, Z_{i-1}), Z_i).$$

These terms can be used in an interactive procedure to compute strings Y that satisfy the existential quantifier in the formula. In essence, a proof-theoretic way to prove this is to display explicitly the terms that have been used to derive the (outermost blocks of) existentially quantified variables in the proven formula; this enables us to compute values that satisfy these quantifiers.

Perron states a version of Herbrand's theorem for the proof systems $\mathbf{G}_i^\star$ and gives a proof-theoretic proof for it. The situation for propositional proof systems

is slightly different from first-order logic. First of all, there are no functions in propositional logic. Therefore the "witnessing" terms are given in the form of *extension* variables; each of these variables is given a defining formula. Another difference is that for first-order logic the Herbrand theorem is stated for universal theories, hence any quantifier introduction rule is applied to a subformula of the formula being proved. On the other hand, for $\mathbf{G}_i^\star$ where $i \geq 1$ the quantifier introduction rules can be applied to cut formulas (and there is no equivalent notion of universal theories).

Moreover, unlike the first-order setting where the KPT theorem gives a disjunction of constant size, for $\mathbf{G}_i^\star$ we must use a complicated notion of *expansion* of the original formula φ. In the simple case where φ is in prenex form, this expansion is just a disjunction of instances of φ (where the outermost quantifiers have been replaced by extension variables). When we are given a family of $\mathbf{G}_i^\star$ proofs π, the number of disjuncts in such disjunctions can grow with the size of π.

The statement of the Herbrand theorem for $\mathbf{G}_i^\star$ in [Per08a, Theorem 3.16] contains a subtle error (but this does not affect other results in the paper). In particular, to state the theorem correctly requires the systems $\mathbf{G}_{\mathcal{B},i}^\star$ mentioned earlier. To our knowledge, these have not been studied before.

The first step as outlined above can indeed be formalized in $\mathbf{VPV}$, because there is a polytime algorithm that transforms a given proof into the form required by the Herbrand theorem. On the other hand, formalizing the second step in $\mathbf{V}^i$ requires formalizing a complicated student–teacher computation. This computation is easy to describe when the reflection principle is stated for prenex formulas. However, for general formulas this becomes more technical, because of the structure of the expansion that results from the Herbrand theorem.

To see the second step, consider for example the case $i = 1$: here we need to argue in $\mathbf{V}^1$ that some witnesses exist that satisfy a $\mathbf{\Pi}_1^q$ formula. In other words, we need to prove in $\mathbf{V}^1$ the existence of a string that satisfies some $\mathbf{\Pi}_1^B$ condition. This is done by formalizing a kind of student–teacher game where there are two players: a student and a teacher. The student tries to learn a string that satisfies the $\mathbf{\Pi}_1^B$ condition by proposing candidates for the string, and the teacher helps by providing some counter-examples, if possible, to the candidates proposed by the student. The game is played in rounds. In the first round, the student announces a candidate for the witnesses, and the teacher responds by either admitting that the witnesses are indeed valid, or providing a counter-example showing that the $\mathbf{\Pi}_1^B$ condition is violated. In the next round, using the counter-example the students announces the next candidate, and the teacher responds as before, and so on. The game is played in at most polynomially many rounds. If in some round the student obtains a correct value then the computation stops and outputs the witness that the student gets. Otherwise the computation fails to produce an output. Here the student can compute polytime functions; in general, he can compute $\mathbf{FP}^{\Sigma_{i-1}^P}$ functions, i.e., functions computable by polytime, oracled Turing machines that

can query oracles in the level Σ^P_{i-1} of the polynomial hierarchy. The teacher is always all-powerful.

Basically, in order to compute the next set of candidate, the student needs to use the expansion, which is obtained in the first step, of the $\mathbf{\Sigma}^q_{i+1}$ tautology that is being proved. He also needs to refer to a dependence relation between the eigenvariables and extension variables. The complicated structure of the expansion together with the dependence relation render the argument rather nontrivial.

Organization. In Section 2 we define **G** and its subsystems, and present a "faithful" formalization of truth definitions for the class of what we call *well structured* formulas. In Section 3 we give the corrected statement of the Herbrand theorem for $\mathbf{G}^\star_i$. In Section 4 we prove the Herbrand theorem for the new system $\mathbf{G}^\star_{\mathcal{B},i}$, and show that depth-1 $\mathbf{G}^\star_{\mathcal{B},i}$ is p-equivalent to $\mathbf{G}^\star_i$ with respect to proving $\mathbf{\Sigma}^q_{i+1}$ formulas. Section 5 concludes with some open questions. The proof of Perron's theorem, that $\mathbf{V}^i$ proves the soundness of $\mathbf{G}^\star_i$ w.r.t. $\mathbf{\Sigma}^q_{i+1}$ formulas, is presented in Appendix A.

2 Preliminaries

2.1 G and its subsystems

We refer to [CN10, Section VII.3] for definitions of quantified propositional formulas (or just formulas) and the sets $\mathbf{\Sigma}^q_i$, $\mathbf{\Pi}^q_i$ of formulas. For $i \geq 1$, a $\mathbf{\Sigma}^q_i$ formula is said to be *proper* if it does not belong to $\mathbf{\Pi}^q_i$. Similarly for proper $\mathbf{\Pi}^q_i$ formulas. Also, we write $\mathcal{B}_i$ for the boolean closure of $\mathbf{\Sigma}^q_i$ and $\mathbf{\Pi}^q_i$.

The proof system **G** extends Gentzen's propositional sequent calculus **PK** to include introduction rules for the quantifiers. The axioms of **G** are

$$\longrightarrow \top; \qquad \bot \longrightarrow \ ; \qquad A \longrightarrow A$$

where A is any atomic formula. See [CN10, Section II.1.1] for the rules of **PK**. Here we will use the "multiplicative" form for the binary rules ($\vee$-left, $\wedge$-right and the cut rules):

$$\frac{B, \Gamma_1 \longrightarrow \Delta_1 \qquad C, \Gamma_2 \longrightarrow \Delta_2}{B \vee C, \Gamma_1, \Gamma_2 \longrightarrow \Delta_1, \Delta_2} \vee\text{-left} \qquad \frac{\Gamma_1 \longrightarrow B, \Delta_1 \qquad \Gamma_2 \longrightarrow C, \Delta_2}{\Gamma_1, \Gamma_2 \longrightarrow B \wedge C, \Delta_1, \Delta_2} \wedge\text{-right}$$

$$\frac{D, \Gamma_1 \longrightarrow \Delta_1 \qquad \Gamma_2 \longrightarrow D, \Delta_2}{\Gamma_1, \Gamma_2 \longrightarrow \Delta_1, \Delta_2} \text{ cut}$$

The reason is to avoid implicit contraction. (In subsequent sections we do not have to transform the given proofs into multiplicative form. Formally, we treat the cases $\vee$-left and $\wedge$-right as we do there followed by a series of contractions.)

The quantifier-introduction rules are as follows:

$$\forall\text{-left: } \frac{A(B), \Gamma \longrightarrow \Delta}{\forall x A(x), \Gamma \longrightarrow \Delta} \qquad \forall\text{-right: } \frac{\Gamma \longrightarrow \Delta, A(p)}{\Gamma \longrightarrow \Delta, \forall x A(x)}$$

$$\exists\text{-left: } \frac{A(p), \Gamma \longrightarrow \Delta}{\exists x A(x), \Gamma \longrightarrow \Delta} \qquad \exists\text{-right: } \frac{\Gamma \longrightarrow \Delta, A(B)}{\Gamma \longrightarrow \Delta, \exists x A(x)}$$

In the rules $\forall$-right and $\exists$-left, p is a free variable called an *eigenvariable* that must not occur in the bottom sequent. For the rules $\forall$-left and $\exists$-right, $A(B)$ is the result of substituting B for all free occurrences of x in $A(x)$. The formula B is called the *target* formula and may be any quantifier-free formula (with no bound variables).

Definition 2.1. *Let π be a proof of a sequent $\mathcal{S}$. A free variable appearing in $\mathcal{S}$ is called a* parameter variable *of π. We say that π is in* free variable normal form *if (i) no free variable is completely eliminated from any sequent in π by any rule except possibly $\forall$-right and $\exists$-left, and in these case the eigenvariable which is eliminated is not a parameter variable, and (ii) every nonparameter variable in π is used exactly once as an eigenvariable.*

A proof can be viewed as a directed graph whose nodes are labeled by sequents in the proof, and there is an edge from $\mathcal{S}_1$ to $\mathcal{S}_2$ if the former is used in the inference resulting in the latter. The treelike subsystem of a proof system $\mathcal{P}$ is always denoted by $\mathcal{P}^\star$. The systems $\mathbf{G}_i$, $\mathbf{G}_{\mathcal{B},i}$ are subsystems of $\mathbf{G}$ defined by restricting the set of cut formulas:

Definition 2.2 (The systems $\mathbf{G}_i$ and $\mathbf{G}_{\mathcal{B},i}$)**.** *For $i \geq 0$ $\mathbf{G}_i$ is the subsystem of $\mathbf{G}$ where all cut formulas are in $\mathbf{\Sigma}^q_i \cup \mathbf{\Pi}^q_i$, and $\mathbf{G}_{\mathcal{B},i}$ is the subsystem of $\mathbf{G}$ where all cut formulas are in $\mathcal{B}_i$.*

A (quantified) propositional proof system $\mathcal{P}_2$ is said to p-simulate $\mathcal{P}_1$ if there is a polytime function F such that for any $\mathcal{P}_1$-proof π of a tautology, $F(\pi)$ is a $\mathcal{P}_2$-proof of the same tautology. $\mathcal{P}_1$ and $\mathcal{P}_2$ are said to be p-equivalent if they p-simulate each other.

It was shown recently [JN10] that $\mathbf{G}_i$ is p-equivalent to $\hat{\mathbf{G}}_i$, the subsystem of $\mathbf{G}_i$ where all cut formulas are prenex $\mathbf{\Sigma}^q_i$ formulas. For treelike systems $\mathbf{G}^\star_i$ and $\hat{\mathbf{G}}^\star_i$ this had been proved by a different method by Morioka [Mor05] (see also [CN10, Theorem VII.4.7]).

Note that $\mathbf{G}_{\mathcal{B},0}$ is the same as $\mathbf{G}_0$, and $\mathbf{G}^\star_{\mathcal{B},0}$ is the same as $\mathbf{G}^\star_0$. For $i \geq 0$, the proof that $\mathbf{G}^\star_{i+1}$ p-simulates $\mathbf{G}_i$ for $\mathbf{\Sigma}^q_i$ and $\mathbf{\Pi}^q_i$ formulas, (see e.g. the proof given in [CN10, Theorem VII.4.3]) actually shows that $\mathbf{G}^\star_{i+1}$ p-simulates $\mathbf{G}_{\mathcal{B},i}$ for $\mathcal{B}_i$ formulas. Perron shows that for $i > 0$, $\mathbf{G}_i$ p-simulates $\mathbf{G}^\star_{i+1}$ for all formulas [Per08b, Theorem 4.2.3] (see also [CN10, Theorem VII.4.8]). As a result, we have:

Corollary 2.3. *For $i \geq 1$, $\mathbf{G}_i$, $\mathbf{G}_{\mathcal{B},i}$ and $\mathbf{G}^\star_{i+1}$ are p-equivalent for formulas in $\mathcal{B}_i$. For other formulas, $\mathbf{G}_{\mathcal{B},i}$ p-simulates $\mathbf{G}_i$, which in turns p-simulates $\mathbf{G}^\star_{i+1}$.*

2.2 Two-sorted theories of bounded arithmetic

We use Zambella's two-sorted setting for theories of bounded arithmetic. The theories $\mathbf{V}^i$ and $\mathbf{TV}^i$ are the two-sorted versions of Buss's theories $\mathbf{S}^i_2$ and $\mathbf{T}^i_2$,

respectively. We will also use **VPV**, the two-sorted version of Cook's theory **PV**. Note that **VPV** has a symbol for every (two-sorted) polytime function, and is a conservative extension of $\mathbf{TV}^0$. Other theories that we use are $\mathbf{VTC}^0$ and $\mathbf{VNC}^1$. Note also that in our setting, there are two kinds of induction axioms: number induction (or just induction) and string induction. Similarly, the maximization and minimization principles also have two versions, one for numbers and one for strings. See [CN10, Chapters V, VIII and IX] for more details.

Recall that two-sorted formula in $\mathbf{\Sigma}_i^B$ and $\mathbf{\Pi}_i^B$ have all their string quantifiers precede the number quantifiers and boolean connectives. Here we will use also the larger sets $\mathbf{g\Sigma}_i^B$ and $\mathbf{g\Pi}_i^B$ (see [CN10, Definition VI.3.1]), which correspond respectively to $\mathbf{\Sigma}_i^b$ and $\mathbf{\Pi}_i^b$ in single-sorted setting. Thus, in $\mathbf{g\Sigma}_i^B$ formulas the bounded number quantifiers can mix with bounded string quantifiers. Note that $\mathbf{V}^i$ proves the comprehension and number induction axioms for $\mathbf{g\Sigma}_i^B$ formulas (see [CN10, Corollary VI.3.8]).

For Perron's theorem (Theorem 2.4) we use the usual formalization of truth definition, for example, as in [Per08a] or [CN10, Chapter X]. However, in the next subsection we will discuss the notion of faithful formalization which might be of independent interest.

2.3 Faithful formalization of truth definitions

This subsection is independent of the rest of the paper.

The way truth definitions are formalized in [Per08a], or similarly in [CN10, Section X.2.1], is rather nonintuitive for non-prenex formulas, as illustrated by the following example.[2] The reader is referred to [CN10, Section X.2] for notations.

Consider a $\mathbf{g\Sigma}_2^B$ formulas $\varphi(a)$ of the form:

$$\varphi(a) \equiv \forall x \le a \exists X \le a \forall Y \le a \psi(x, X, Y)$$

where $\psi(x, X, Y)$ is a $\mathbf{\Sigma}_0^B$ formula. Let A_n denote the propositional translation $\varphi(a)[n]$:

$$A_n \;\equiv\; \bigwedge_{i=0}^{n} \exists p_0 \exists p_1 \ldots \exists p_{n-2} \bigvee_{k=0}^{n} B_{n,k}(p_0, p_1, \ldots, p_{k-2})$$

Here

$$B_{n,k}(p_0, p_1, \ldots, p_{k-2}) \equiv \forall q_0 \forall q_1 \ldots \forall q_{n-2} \bigwedge_{j=0}^{n} C_{n,k,j}(p_0, p_1, \ldots, p_{k-2}, q_0, q_1, \ldots, q_{j-2})$$

is the translation $\forall Y \le a \psi(x, X, Y)[n; k]$; note that $B_{n,k}$ is $\mathbf{\Pi}_1^q$.

Now, suppose further that there is a model of $\mathbf{V}^1$ where $\varphi(a)$ holds, but

$$\exists Z \forall x \le a \forall Y \le a \psi(x, Z^{[x]}, Y) \tag{1}$$

[2] This example is suggested by the referee.

does not hold. This is possible for some φ under the hypothesis that $\mathbf{V}^1 \nvdash \mathbf{\Pi}_1^B$-**REPL**. Intuitively, in this model, A_n should be true, because it is the translation of a true first-order formula. However, according to the usual formalization of truth definition, it is not. This is because to say that A_n is true is the same as to say that the formula (1) is true.

To resolve this issue we need a formalization in bounded arithmetic of truth definitions which is "faithful", in the sense that it does not collect at once, say, all outermost existentially quantified variables of a non-prenex formula. Here we are able to do this for only the class of *well structured formulas*, which is defined inductively as follows.

A quantifier-free formula is by default well structured. A $\mathbf{\Sigma}_{i+1}^q$ formula F is well structured if it is a boolean combination of formulas of the forms

$$\exists x_1 \exists x_2 \ldots \exists x_m A(\vec{x}) \quad \text{or} \quad \forall y_1 \forall y_2 \ldots \forall y_m B(\vec{y})$$

where $A \in \mathbf{\Pi}_i^q$ and $B \in \mathbf{\Sigma}_i^q$, and both A, B are well structured. For example, the following $\mathbf{\Sigma}_1^q$ formula is *not* well structured:

$$\exists x_1 (\exists x_2 B(x_1, x_2) \wedge \exists x_3 C(x_1, x_3))$$

Obviously, if $F \in \mathbf{\Sigma}_{i+1}^q$, then the occurrences of formulas of the first form in F must be positive, i.e., they are inside the scope of an even number of $\neg$'s, while the occurrences of formulas of the second form must be negative. A well structured $\mathbf{\Pi}_{i+1}^q$ formula is defined similarly. We denote by $w\mathbf{\Sigma}_i^q$ and $w\mathbf{\Pi}_i^q$ the sets of well structured $\mathbf{\Sigma}_i^q$ and well structured $\mathbf{\Pi}_i^q$ formulas, respectively.

Notice that well structured formulas include the propositional translations of all formulas in $\cup_{i \geq 0} \mathbf{\Sigma}_i^B$ (see [CN10, Chapter VII]), while other formulas, e.g., those in $\mathbf{g\Sigma}_1^B - \mathbf{\Sigma}_1^B$ may not translate into well structured propositional formulas.

Recall a suitable encoding of (quantified) propositional formulas in our two-sorted vocabulary, such as in [CN10, Section X.1]. Note that valid encodings can be recognized in $\mathbf{TC}^0$, and the same holds for well structured formulas. In addition, propositional translations of bounded arithmetic formulas can also be computed in $\mathbf{TC}^0$. The encoding of a propositional formula A is often denoted by $\widehat{A}$.

We use the formulas $Z \models_0^\Sigma X$ and $Z \models_0^\Pi X$ from [CN10, Lemma X.2.1]. Informally, $Z \models_0^\Sigma X$ (resp. $Z \models_0^\Pi X$) is a $\mathbf{\Sigma}_1^B$ (resp. $\mathbf{\Pi}_1^B$) formula stating that the truth assignment specified by Z satisfies the formula encoded by X. Let $Z \models_{w\mathbf{\Sigma}_0^q} X$ be $Z \models_0^\Sigma X$, and $Z \models_{w\mathbf{\Pi}_0^q} X$ be $Z \models_0^\Pi X$. Then for $i \geq 0$ the formulas $Z \models_{w\mathbf{\Sigma}_{i+1}^q} X$ and $Z \models_{w\mathbf{\Pi}_{i+1}^q} X$ are defined inductively; below we will informally describe the inductive definition.

Suppose that we are given a formula $A \in w\mathbf{\Sigma}_{i+1}^q$. (We do not display the free variables in A.) Then, by definition, A has the form $B(C_0, C_1, \ldots, C_m)$, where $B(p_0, p_1, \ldots, p_m)$ is a quantifier-free formula, and each C_k $(0 \leq k \leq m)$ is of one of the forms

$$\exists x_0 \exists x_1 \ldots \exists x_n D(\vec{x}) \quad \text{or} \quad \forall y_0 \forall y_1 \ldots \forall y_n F(\vec{y})$$

where $D \in w\mathbf{\Pi}^q_i$, and $F \in w\mathbf{\Sigma}^q_i$. (The former kind must appear positively in A, and the latter negatively.) Without loss of generality, suppose that all C_k's are of the former kind, and thus appear positively in A. Then $(Z \models_{w\mathbf{\Sigma}^q_{i+1}} \widehat{A})$ states that there exists a setting of $p_0, p_1, \ldots, p_m$ that satisfies $B(\vec{p})$, and such that for all $0 \leq k \leq m$, $p_k = 1$ implies that C_k is true. Thus, essentially $(Z \models_{w\mathbf{\Sigma}^q_{i+1}} \widehat{A})$ is equivalent to

$$\exists Y \big((Z, Y \models^{\Sigma}_0 \widehat{B(\vec{p})}) \wedge \forall k \leq m,\, Y(k) \supset (Z \models_{w\mathbf{\Pi}^q_i} \widehat{C_k}) \big)$$

Here $(Z, Y \models^{\Sigma}_0 \widehat{B(\vec{p})})$ is the $\mathbf{\Sigma}^B_1$ formula stating that the truth assignment specified by Z and Y satisfies $B(\vec{p})$.

The formal definition of $(Z \models_{w\mathbf{\Sigma}^q_{i+1}} \widehat{A})$ requires additional $\exists$- and $\forall$-string quantifiers to extract $\widehat{B}$ and $\widehat{C_k}$ from $\widehat{A}$. In particular, in order to extract $\widehat{B}$ from $\widehat{A}$ we need another $\exists$-string quantifier, while if $i \geq 1$ then to extract $\widehat{C_k}$ we need an additional $\forall$-string quantifier, and if $i = 0$ then we need an additional $\exists$-string quantifiers to extract $\widehat{C_k}$. We omit the details here.

The formula $(Z \models_{w\mathbf{\Pi}^q_{i+1}} \widehat{A})$ is defined similarly: basically, it is equivalent to

$$\forall Y \big((\forall k \leq m, (Z \models_{w\mathbf{\Sigma}^q_i} \widehat{C_k}) \supset Y(k)) \supset (Z, Y \models^{\Sigma}_0 \widehat{B(\vec{p})}) \big)$$

It can be seen that for $i \geq 1$, $(Z \models_{w\mathbf{\Sigma}^q_i} X)$ and $(Z \models_{w\mathbf{\Pi}^q_i} X)$ are respectively in $\mathbf{g\Sigma}^B_i$ and $\mathbf{g\Pi}^B_i$. Because $\mathbf{V}^i$ proves number induction for $\mathbf{g\Sigma}^B_i$ and $\mathbf{g\Pi}^B_i$ formulas (see [CN10, Corollary VI.3.8]), it follows that known provability of the Reflection principles (see the next subsection) whose proofs use number induction, such as $\mathbf{V}^i \vdash \mathbf{\Pi}^q_i\text{-}RFN_{\mathbf{G}_{i-1}}$ (see [CN10, Theorem X.2.17]), continues to hold for the class of well structured formulas.

Finally, note that for any well structured formula $\mathbf{\Sigma}^q_i$ formula A,

$$\mathbf{VNC}^1 \vdash (Z \models_{\mathbf{\Sigma}^q_i} \widehat{A}) \supset (Z \models_{w\mathbf{\Sigma}^q_i} \widehat{A})$$

where $(Z \models_{\mathbf{\Sigma}^q_i} \widehat{A})$ is the formalization from [CN10, Section X.2]. Here we need $\mathbf{VNC}^1$ for arguing about the correctness of the algorithm for the Formula Evaluation problem.

2.4 The Reflection principle

The Reflection principle for a proof system $\mathcal{P}$ states that $\mathcal{P}$ is sound, i.e., that any formula that has a $\mathcal{P}$-proof is valid. This is usually formulated for sets of formulas of certain complexity, such as $\mathbf{\Sigma}^q_i$. From now on we use the usual formalization of truth definitions, see, for example, [CN10, Section X.2].

For a set Φ of formulas the Reflection principle for $\mathcal{P}$ with respect to proving formulas in Φ, denoted by $\Phi\text{-}RFN_{\mathcal{P}}$, is the statement saying that if A is a formula in Φ and A has a $\mathcal{P}$-proof π, then A is valid. For $i \geq 1$, $\mathbf{\Sigma}^q_i\text{-}RFN_{\mathcal{P}}$ is a $\forall \mathbf{g\Sigma}^B_i$ sentence, while $\mathbf{\Sigma}^q_0\text{-}RFN_{\mathcal{P}}$ is a $\forall \mathbf{\Sigma}^B_1$ sentence, and for systems $\mathcal{P}$ of interest here $\mathbf{\Pi}^q_0\text{-}RFN_{\mathcal{P}}$ is equivalent to the *consistency* statement of $\mathcal{P}$.

The provability of the Reflection principles for a propositional proof system $\mathcal{P}$ can be used to prove p-simulation for $\mathcal{P}$. For example, with respect to $\mathbf{\Sigma}^q_i$ formulas $\mathbf{G}^\star_i$ is the strongest system whose Reflection principle is provable in $\mathbf{V}^i$. Moreover, some principle for $\mathcal{P}$ can be used to derive all axioms of the first-order theory associated with $\mathcal{P}$. As a result, it is known, e.g., that $\mathbf{TV}^0$ does not prove $\mathbf{\Sigma}^q_2$-$RFN_{\mathbf{G}^\star_1}$ unless the polytime hierarchy collapses. This is because $\mathbf{\Sigma}^q_2$-$RFN_{\mathbf{G}^\star_1}$ together with $\mathbf{TV}^0$ prove all axioms of $\mathbf{V}^1$, so if $\mathbf{TV}^0$ proves $\mathbf{\Sigma}^q_2$-$RFN_{\mathbf{G}^\star_1}$, then $\mathbf{TV}^0 = \mathbf{V}^1$, and this implies that the polynomial time hierarchy $\mathbf{PH}$ provably (in $\mathbf{TV}^0$) collapses to the boolean closure of $\mathbf{NP}$ [KPT91, Bus95b, Zam96, Jeř09]. So it is natural to ask whether or not the next theory in the $\mathbf{V}^i$ hierarchy proves $\mathbf{\Sigma}^q_2$-$RFN_{\mathbf{G}^\star_1}$, and Steven Perron answers this question affirmatively:

Theorem 2.4 (Perron)**.** *For $i \geq 1$, $\mathbf{V}^i \vdash \mathbf{\Sigma}^q_{i+1}$-$RFN_{\mathbf{G}^\star_i}$.*

As mentioned in the introduction, the proof of this theorem requires first the Herbrand theorem for $\mathbf{G}^\star_i$, which we present in the next section. Using this, a student–teacher algorithm can be constructed to compute the witness for the $\mathbf{\Sigma}^q_{i+1}$ tautologies being proved in $\mathbf{G}^\star_i$. The algorithm will be given in Appendix A.

3 The Herbrand theorem for $\mathbf{G}^\star_i$

The Herbrand theorem for $\mathbf{G}^\star_i$ is given in Theorem 3.4 below. The proof of this theorem will be almost a reproduction of Perron's proof of his Theorem 3.16 in [Per08a], although our analysis contains more cases. In particular, in the proof of Theorem 3.4, Cases 3c, 4d, 4e (and their dual cases) are places where we correct the complexity of the cut formulas (to be in $\mathcal{B}_{i-1}$ rather than $\mathbf{\Sigma}^q_{i-1} \cup \mathbf{\Pi}^q_{i-1}$). Perron's statement of the Herbrand theorem for $\mathbf{G}^\star_i$ also includes an ordering of the variables; in Subsection 3.2 we will analyze such an ordering in more detail and prove several important properties that will be needed for the student–teacher computation.

Understanding the Herbrand theorem for $\mathbf{G}^\star_i$ is important for designing the algorithm for the second step of proving Perrons' theorem, which will be presented in the appendix. In the next section we will give a Herbrand theorem for $\mathbf{G}^\star_{\mathcal{B},i}$ which will require the same case analysis as in the current section.

Let π be a $\mathbf{G}^\star_i$-proof of a $\mathbf{\Sigma}^q_{i+1}$ formula A. Recall that our ultimate goal is to compute some witnessing values for the (outermost) existentially bound variables in A. Basically, the role of the Herbrand theorem is to display explicitly the target formulas that are used to derive the variables. These target formulas can be used to compute the desired witnesses.

The statement of the theorem involves the notions of *extension* variables and *expansion* of a formula. To explain these notions, the best way is perhaps to see

where we need them. First, consider the $\exists$-right rule:

$$\frac{\mathcal{S}_1}{\mathcal{S}} \quad = \quad \frac{\Gamma \longrightarrow D(F), \Delta}{\Gamma \longrightarrow \exists x D(x), \Delta}$$

Here F is the target formula that is used to introduce x. We will focus on the case where D has high complexity, for example in $\mathbf{G}_i^\star$ proofs we will focus on the case where the formula D does *not* belong to $\mathbf{\Sigma}_i^q$. Then $\exists x D(x)$ cannot be a cut formula, and hence must be part of the final formula. In this case a new variable e (called an extension variable) is introduced, the formula $\exists x D(x)$ in $\mathcal{S}$ is replaced by $D(e)$, and we add a "definition" $e \leftrightarrow F$ for e to the antecedent of $\mathcal{S}$. Thus, extension variables play the role of the witnessing terms in first-order logic. (Extension variables are used for defining extension Frege proof systems [CR79], and the idea of using extension variables for computing witnessing for existential variables has been used to show that extended Frege can p-simulate $\mathbf{G}_1^\star$.)

The above changes lead to several issues that need to be handled, and the notion of an expansion of a formula deals with the issue involving the contraction rule. Consider a contraction right, where the formula involved is the formula $\exists x D(x)$ above, e.g.,

$$\frac{\mathcal{S}_1}{\mathcal{S}} \quad = \quad \frac{\Gamma \longrightarrow \exists x D(x), \exists x D(x), \Delta}{\Gamma \longrightarrow \exists x D(x), \Delta}$$

Now the two copies of $\exists x D(x)$ in $\mathcal{S}_1$ have been replaced by $D(e_1)$ and $D(e_2)$, for two distinct extension variables e_1 and e_2. It is not possible to identify e_2 with e_1 so that the contraction rule can be applied (to the new $\mathcal{S}_1$), because e_1 and e_2 have different definitions. In this case the formula $\exists x D(x)$ in $\mathcal{S}$ is replaced with the disjunction $D(e_1) \vee D(e_2)$. In other words, we modify $\mathcal{S}$ by substituting the extension variables e_1, e_2 for the bound variables x and x' in $\exists x D(x) \vee \exists x' D(x')$ and removing the corresponding quantifiers.

For simplicity we will be working only with formulas where the quantifiers are not in the scope of any $\neg$. Without this assumption, the notions defined below need to distinguish between positive and negative occurrences of the quantifiers.

In general, we want to replace certain subformulas B of a given formula A by $B \vee B'$, where B' is the same as B except for the bound variables which are renamed. In the following definition we require that such subformulas B must *not* belong to a certain set of formulas Φ.

Definition 3.1. *For a class Φ of formulas (e.g., $\mathbf{\Sigma}_i^q$ or $\mathbf{\Pi}_i^q$) and any formula A, a* Φ-expansion *of A is any formula that can be obtained by finitely many applications of the following operation on A: take a subformula B such that $B \notin \Phi$, and replace B by $B \vee B'$, where B' is obtained from B by renaming every bound variable.*

We write i-expansion instead of $(\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q)$-expansion.

We are mostly concerned with $\mathcal{B}_i$-expansion. Note that a $\mathcal{B}_i$-expansion of a formula A is less "deformed" than its i-expansions in the sense that $\mathcal{B}_i$ subformulas of A are intact in the former while they can be modified in the latter. For example, suppose that

$$A \;=\; \exists x_1(\exists x_2 B(x_1, x_2) \wedge \forall y C(x_1, y))$$

where B and C are quantifier-free. Consider

$$\begin{aligned}
A_1 &= \big(\exists x_1(\exists x_2 B(x_1, x_2) \wedge \forall y_1 C(x_1, y_1))\big) \vee \big(\exists x_3(\exists x_4 B(x_3, x_4) \wedge \forall y_2 C(x_3, y_2))\big) \\
A_2 &= \exists x_1\big(\big(\exists x_2 B(x_1, x_2) \wedge \forall y_1 C(x_1, y_1)\big) \vee \big(\exists x_3 B(x_1, x_3) \wedge \forall y_2 C(x_1, y_2)\big)\big) \\
A_3 &= \exists x_1((\exists x_2 B(x_1, x_2) \vee \exists x_3 B(x_1, x_3)) \wedge (\forall y_1 C(x_1, y_1) \vee \forall y_2 C(x_1, y_2))).
\end{aligned}$$

Here A_1 is a $\mathcal{B}_1$-expansions of A (and hence also an 1-expansion of A). A_2 is a 1-expansion, but it is not a $\mathcal{B}_1$-expansion of A, because the subformula that is duplicated is $\mathcal{B}_1$. Finally, A_3 is a 0-expansion of A.

Another issue arising from keeping the target formulas F around (as explained above for the $\exists$-right rule) is the fact that the conditions for eigenvariables required for the $\exists$-left and $\forall$-right rules are no longer satisfied. In fact, the $\forall$-right inferences that introduce outermost universal quantifiers of proper $\mathbf{\Pi}_i^q$ formulas will be skipped; in effect, these quantifiers are replaced by the eigenvariables that introduce them. In general, transformations of formulas in the given proof is given in the following definition.

Definition 3.2 ((Φ, Q, E)-instance)**.** *For set Φ of formulas (e.g., $\Phi = \mathbf{\Sigma}_i^q$ or $\Phi = \mathbf{\Pi}_i^q$ for some $i \geq 0$) and sets E and Q of distinct variables, a (Φ, Q, E)-instance of a formula A is a formula obtained from A by applying the following operations on its subformulas whenever possible, starting from the outermost quantifier:*

- *for a subformula $\exists x B(x)$ such that $\exists x B(x) \notin \Phi$, replacing x by a distinct variable e in E, and removing the quantifier,*
- *for a subformula $\forall y B(y)$ such that $\forall y B(y) \notin \Phi$, replacing y by a distinct variable q in Q, and removing the quantifier.*

When $\Phi = \mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$ we write (i, Q, E)-instance for (Φ, Q, E)-instance. Note that an (i, Q, E)-instance of a formula is always a formula in $\mathcal{B}_i$. Also, for $\Phi = \mathbf{\Pi}_i^q$ or $\Phi = \mathbf{\Sigma}_i^q$, a (Φ, Q, E)-instance of a formula is always a formula in Φ. Finally, note that some variables in Q, E may not be used for replacing bound variables in A.

For example, let $Q = \{q_1, q_2, \ldots\}$ and $E = \{e_1, e_2, \ldots\}$, then the following formula is a $(1, Q, E)$-instance of the formula A_1 above:

$$(\exists x_2 B(e_1, x_2) \wedge \forall y_1 C(e_1, y_1)) \vee (\exists x_4 B(e_3, x_4) \wedge \forall y_2 C(e_3, y_2)).$$

The next formula is a $(0, Q, E)$-instance of A_2:

$$(B(e_1, e_2) \wedge C(e_1, q_1)) \vee (B(e_1, e_3) \wedge C(e_1, q_2)).$$

The following formula is a $(\mathbf{\Pi}^q_1, Q, E)$-instance of A_3:

$$(B(e_1, e_2) \vee B(e_1, e_3)) \wedge (\forall y_1 C(e_1, y_1) \vee \forall y_1 C(e_1, y_2)).$$

Definition 3.3. *Let E, K be disjoint sets of variables, $E = \{e_1, e_2, \ldots, e_m\}$. Let Φ be a set of formulas. We say that a cedent*

$$e_1 \leftrightarrow F_1, e_2 \leftrightarrow F_2, \ldots, e_m \leftrightarrow F_m$$

is a Φ-cedent defining E in terms of K if each F_t is a formula belonging to Φ whose free variables are among $e_1, e_2, \ldots, e_{t-1}$ and variables in K. (In particular, the free variables of F_1 are from K.)

Notice that a $\mathcal{B}_i$-cedent can be turned into a $\mathbf{\Sigma}^q_i$-cedent (or $\mathbf{\Pi}^q_i$-cedent) by using new extension variables. For example, if e has a $\mathcal{B}_1$-definition

$$e \leftrightarrow (A \wedge B) \vee C \tag{2}$$

where A is $\mathbf{\Sigma}^q_1$ and B, C are $\mathbf{\Pi}^q_1$. Then we can introduce three new extension variables e_1, e_2 and e_3 with $\mathbf{\Sigma}^q_1$ definition as follows:

$$e_1 \leftrightarrow A, \quad e_2 \leftrightarrow B', \quad e_3 \leftrightarrow C', \quad e \leftrightarrow (e_1 \wedge \neg e_2) \vee \neg e_3 \tag{3}$$

where B' and C' are obtained from $\neg B$ and $\neg C$ respectively by using De Morgan's laws to push $\neg$ inside the scope of all quantifiers. In most cases, the proof using e with definition in (2) can be replaced by a proof using e_1, e_2, e_3, e with definitions in (3).

Theorem 3.4 (The Herbrand theorem for $\mathbf{G}^\star_i$)**.** *Let $i \geq 1$. Let π be a $\mathbf{G}^\star_i$ proof of a formula A. Then there is a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof π' of a sequent*

$$\Lambda \longrightarrow A' \tag{4}$$

where A' is an $(i-1, Q, E)$-instance of a $\mathcal{B}_i$-expansion $A^\star$ of A, for some sets Q, E of new distinct variables, and Λ is a $\mathbf{\Sigma}^q_{i-1}$-cedent defining E in terms of Q and the free variables in A. Moreover, π' is provably in $\mathbf{VPV}$ computable by a polytime function.

Although Theorem 3.2.16 of [Per08b] states that π' is a $\mathbf{G}^\star_{i-1}$ proof, the construction there actually gives a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof. This is because the cut formula in Case 3 on pages 50–51 of [Per08b] is in $\mathcal{B}_{i-1}$.

Perron only states that A' is an $(i-1, Q, E)$-instance of an $(i-1)$-expansion of A. So our statement here is slightly stronger, because a $\mathcal{B}_i$-expansion is also an $(i-1)$-expansion. He also provides a total order on the variables in $E \cup Q$. By examining the proof of Theorem 3.4 given below we will point out that a partial order exists and we will analyze this order in Subsection 3.2. This partial order can be made total in the way stated in [Per08b]. However, it is easier to describe the algorithm in Subsection A.3 and argue for its correctness using the partial order rather than using the total order of [Per08b].

To prove the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_{i+1}^q$ tautologies, the transformation we obtain in Theorem 3.4 seems to be optimal. For example, we cannot make π' a $\mathbf{G}_{i-1}^\star$ proof, or give lower complexity defining formulas for extension variables, or make $A^\star$ a $\mathbf{\Sigma}_{i+1}^q$ expansion of A. However, to prove the soundness of $\mathbf{G}_i^\star$ with respect to $\mathbf{\Sigma}_i^q$ formulas we can improve this transformation a little, see Theorem 4.7.

Proof of Theorem 3.4. Recall that, for simplicity, we assume that in A there are no quantifiers in the scope of any $\neg$. Without this assumption we have to consider separately positive and negative occurrences of $\exists$ and $\forall$. By this assumption, no quantified ancestors of A appear in the antecedents of the proof. In addition, we will regard each cedent as a multi-set of formulas as opposed to an ordered list of formulas (so effectively we do not mention the weakening and exchange rules). We will assume furthermore that all cut formulas in π are prenex $\mathbf{\Sigma}_i^q$ formulas. (This is possible by Morioka's theorem [Mor05], see also [CN10, Theorem VII.4.7].) Our last assumption is that π is in free variable normal form. This is because π can be transformed in polytime into such a form. In other words, each eigenvariable in π is used exactly once in a quantifier introduction rule. We denote by $Q_\mathcal{S}$ the set of eigenvariables of the $\forall$-right and $\exists$-left inferences in the subproof ending in sequent $\mathcal{S}$.

Let $\mathcal{S}$ be any sequent in π. Then $\mathcal{S}$ is of the form

$$\Gamma, \Upsilon \longrightarrow \Delta, \Omega$$

where Γ and Δ consist of all ancestors of a cut formula in $\mathcal{S}$, and Υ, Ω consist of all ancestors of the final formula. Note that all quantified formulas in Γ and Δ are prenex $\mathbf{\Sigma}_i^q$. Also, by our assumptions (that quantifiers do not appear in the scope of $\neg$), all formulas in Υ are quantifier-free. We will prove by induction on $\mathcal{S}$ that there is a sequent $\mathcal{S}'$ of the form

$$\Lambda, \Gamma', \Upsilon \longrightarrow \Delta', \Omega' \tag{5}$$

with a polynomial size $\mathbf{G}_{\mathcal{B},i-1}^\star$ proof and a set $E_\mathcal{S}$ of extension variables, such that

(A1) Γ' is obtained from Γ by replacing each proper $\mathbf{\Sigma}_i^q$ formula $\exists \vec{y} D(\vec{y})$ by $D(\vec{q})$, for $\vec{q} \in Q_\mathcal{S}$,

(A2) Δ' is obtained from Δ by replacing each proper $\mathbf{\Sigma}_i^q$ formula $\exists \vec{x} D(\vec{x})$ by $D(\vec{e})$, for $\vec{e} \in E_\mathcal{S}$,

(A3) Ω' is obtained from Ω by replacing each formula B by an $(i-1, Q_\mathcal{S}, E_\mathcal{S})$-instance of a $\mathcal{B}_i$-expansion of B,

(A4) Λ is a $\mathbf{\Sigma}_{i-1}^q$-extension cedent defining $E_\mathcal{S}$ in terms of $Q_\mathcal{S}$ and the free variables in $\mathcal{S}$,

(A5) each variable in $Q_\mathcal{S} \cup E_\mathcal{S}$ appears in at most one formula in Γ', Δ', Ω'.

When $\mathcal{S}$ is the final sequent in π:

$$\mathcal{S} \;=\; \longrightarrow A$$

(here Γ, Υ and Δ are empty) the new sequent $\mathcal{S}'$ is what is required by the theorem.

For the base case $\mathcal{S}$ is an axiom. We can take $\mathcal{S}' = \mathcal{S}$ and there is nothing to do. For the induction step, we consider the cases depending on how $\mathcal{S}$ is derived from previous sequents. In each case below we can give an appropriate numbering of the extension variables so that they satisfy the condition of Definition 3.3 (i.e., each e_t is defined using only $e_1, e_2, \ldots, e_{t-1}$). However to show that such a numbering is correct we need a rigorous argument, using the fact that the dependence relation (Definition 3.6) is a partial order on the variables in $Q_{\mathcal{S}} \cup E_{\mathcal{S}}$. We will not discuss the numbering here, but in Subsection 3.2 we will show various properties of the dependence relation that are useful for the correctness of the algorithm presented in Subsection A.3.

For the case where $\mathcal{S}$ is derived by the $\vee$, $\wedge$ or $\neg$ rules, the same rule can be applied to obtain $\mathcal{S}'$ (and it is clear what $\mathcal{S}'$ should be). So we focus on other cases. We will describe what $\mathcal{S}'$ is in each case, and leave it to the reader to verify that it satisfies the five items listed above.

In Cases 2a, 4c and 4d below, although the defining formulas for the new extension variables are $\mathcal{B}_{i-1}$, by the remark following Definition 3.3 (adapted to each case appropriately) the $\mathcal{B}_{i-1}$-cedent can be turned into a $\mathbf{\Sigma}^q_{i-1}$-cedent without requiring new cut formulas.

Case 1: $\mathcal{S}$ is derived by a cut. Consider the interesting case where the cut formula is a proper (prenex) $\mathbf{\Sigma}^q_i$ formula:

$$\frac{\mathcal{S}_1 \qquad \mathcal{S}_2}{\mathcal{S}} \;=\; \frac{\exists \vec{z} D(\vec{z}), \Gamma_1, \Upsilon_1 \longrightarrow \Delta_1, \Omega_1 \qquad \Gamma_2, \Upsilon_2 \longrightarrow \exists \vec{z} D(\vec{z}), \Delta_2, \Omega_2}{\Gamma_1, \Gamma_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta_1, \Delta_2, \Omega_1, \Omega_2}$$

Here D is $\mathbf{\Pi}^q_{i-1}$. By the induction hypothesis, we have $\mathbf{G}^{\star}_{\mathcal{B},i-1}$ proofs of the sequents

$$\begin{aligned} \mathcal{S}'_1 &= \Lambda_1(\vec{q}), D(\vec{q}), \Gamma'_1, \Upsilon_1 \longrightarrow \Delta'_1, \Omega'_1 \\ \mathcal{S}'_2 &= \Lambda_2(\vec{e}), \Gamma'_2, \Upsilon_2 \longrightarrow D(\vec{e}), \Delta'_2, \Omega'_2 \end{aligned}$$

Note that $\vec{q}$ do not appear in the proof of $\mathcal{S}_2$ and hence do not appear also in the proof of $\mathcal{S}'_2$. We modify $\mathcal{S}'_1$ and its proof by replacing $\vec{q}$ by $\vec{e}$, and obtain the following sequent by a cut on $D(\vec{e})$:

$$\mathcal{S}' \;=\; \Lambda_2(\vec{e}), \Lambda_1(\vec{e}), \Gamma'_1, \Gamma'_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta'_1, \Delta'_2, \Omega'_1, \Omega'_2$$

This is the sequent we need for $\mathcal{S}$.

Case 2a: $\mathcal{S}$ is inferred by $\forall$-right where the principal formula is $\mathbf{\Pi}^q_{i-1}$:

$$\frac{\mathcal{S}_1}{\mathcal{S}} \;=\; \frac{\Gamma, \Upsilon \longrightarrow D(q), \Delta, \Omega}{\Gamma, \Upsilon \longrightarrow \forall y D(y), \Delta, \Omega} \tag{6}$$

where $\forall y D(y)$ is $\mathbf{\Pi}^q_{i-1}$. By the induction hypothesis, there is a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof of

$$\mathcal{S}'_1 \quad = \quad \Lambda(q), \Gamma', \Upsilon \longrightarrow D(q), \Delta', \Omega' \tag{7}$$

Although q does not appear in Γ', Υ, Δ' or Ω' because it is used as an eigenvariable in (6) and hence cannot appear in Γ, Υ, Δ or Ω, it can appear in $\Lambda(q)$, so we cannot apply the $\forall$-right rule to $\mathcal{S}'_1$. Perron handles this by showing [Per08b, Lemma 3.1.3] that for a new variable e there is a $\mathbf{G}^\star_{i-1}$ proof of

$$e \leftrightarrow D(\bot), D(e) \longrightarrow \forall y D(y)$$

and replacing q by e so that $\mathcal{S}'_1$ becomes

$$\Lambda(e), \Gamma', \Upsilon \longrightarrow D(e), \Delta', \Omega'$$

The sequent

$$\mathcal{S}' \quad = \quad e \leftrightarrow D(\bot), \Lambda(e), \Gamma', \Upsilon \longrightarrow \forall y D(y), \Delta', \Omega'$$

is derived from the two sequents above by a cut on $D(e)$.

Case 2b: $\mathcal{S}$ is inferred by $\forall$-right rule for other formulas. That is, $\mathcal{S}$ is derived as in (6) but here $\forall y D(y)$ is not $\mathbf{\Pi}^q_{i-1}$. Therefore $\forall y D(y)$ must be an ancestor of the final formula A. By the induction hypothesis there is a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof of a sequent $\mathcal{S}'_1$ of the form (7). We simply take $\mathcal{S}' = \mathcal{S}'_1$.

Case 3a: $\mathcal{S}$ is derived by the $\exists$-right rule where the principal formula is $\mathbf{\Sigma}^q_{i-1}$:

$$\frac{\mathcal{S}_1}{\mathcal{S}} \quad = \quad \frac{\Gamma, \Upsilon \longrightarrow D(F), \Delta, \Omega}{\Gamma, \Upsilon \longrightarrow \exists x D(x), \Delta, \Omega} \tag{8}$$

where $\exists x D(x)$ is $\mathbf{\Sigma}^q_{i-1}$. The target formula F is quantifier-free and does not contain any variables in $Q_{\mathcal{S}_1} \cup E_{\mathcal{S}_1}$ (although some free variables in F may later be used as eigenvariables). By the induction hypothesis, there is a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof of a sequent $\mathcal{S}'_1$ of the form

$$\mathcal{S}'_1 \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow D(F), \Delta', \Omega'$$

We define

$$\mathcal{S}' \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow \exists x D(x), \Delta', \Omega'$$

It can be obtained from $\mathcal{S}'_1$ by $\exists$-right.

Cases 3b and 3c below are handled in the same way, but we treat them separately because we will refer to them individually later.

Case 3b: $\mathcal{S}$ is derived by the $\exists$-right rule where the principal formula is a proper (prenex) $\mathbf{\Sigma}^q_i$ ancestor of a cut formula. That is, $\mathcal{S}$ is derived as in (8), but here $\exists x D(x) = \exists \vec{x} C(\vec{x}, F)$ where C is a proper $\mathbf{\Pi}^q_{i-1}$ formula. We have

$$\mathcal{S}'_1 \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow C(\vec{e}, F), \Delta', \Omega'$$

Define

$$\mathcal{S}' \quad = \quad e \leftrightarrow F, \Lambda, \Gamma', \Upsilon \longrightarrow C(\vec{e}, e), \Delta', \Omega'$$

for a new extension variable e. The following sequent has a cut-free $\mathbf{G}^\star$ proof of polynomial size:

$$e \leftrightarrow F, C(\vec{e}, F) \longrightarrow C(\vec{e}, e)$$

(See for example [CN10, Lemma VII.4.11].) $\mathcal{S}'$ can be obtained from $\mathcal{S}_1'$ and the above sequent by a cut on $C(\vec{e}, F)$. (Here we cut a $\mathbf{\Pi}^q_{i-1}$ formula.)

Case 3c: $\mathcal{S}$ is derived by $\exists$-right where the principal formula is a non-$\mathbf{\Sigma}^q_{i-1}$ ancestor of the final formula. Thus

$$\mathcal{S}_1' \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow D'(F), \Delta', \Omega'$$

where $D'(F)$ is an $(i-1, Q_{\mathcal{S}_1}, E)$-instance of a $\mathcal{B}_i$-expansion of $D(F)$. In particular, $D'(F)$ is in $\mathcal{B}_{i-1}$.

Define

$$\mathcal{S}' \quad = \quad e \leftrightarrow F, \Lambda, \Gamma', \Upsilon \longrightarrow D'(e), \Delta', \Omega'$$

for a new extension variable e. As in the previous case, the following sequent has a cut-free $\mathbf{G}^\star$ proof of polynomial size:

$$e \leftrightarrow F, D'(F) \longrightarrow D'(e)$$

Therefore $\mathcal{S}'$ can be obtained from $\mathcal{S}_1'$ and the above sequent by a cut on $D'(F)$. (Here we cut a formula in $\mathcal{B}_{i-1}$.)

Case 4a: $\mathcal{S}$ is obtained by contraction-right on a formula in $\mathcal{B}_{i-1}$:

$$\frac{\mathcal{S}_1}{\mathcal{S}} \quad = \quad \frac{\Gamma, \Upsilon \longrightarrow D, D, \Delta, \Omega}{\Gamma, \Upsilon \longrightarrow D, \Delta, \Omega} \tag{9}$$

where $D \in \mathcal{B}_{i-1}$. Here D remains the same in $\mathcal{S}_1'$, so $\mathcal{S}_1'$ has the form

$$\Lambda, \Gamma', \Upsilon \longrightarrow D, D, \Delta', \Omega'$$

Define

$$\mathcal{S}' = \Lambda, \Gamma', \Upsilon \longrightarrow D, \Delta', \Omega'$$

It can be derived from $\mathcal{S}_1'$ by contraction-right.

Cases 4b and 4c are actually subcases of 4d, but we treat them separately to make 4d easier to follow.

Case 4b: $\mathcal{S}$ is obtained by contraction-right on a proper $\mathbf{\Pi}^q_i$ formula. So $\mathcal{S}$ is obtained from $\mathcal{S}_1$ as in (9), but here D is a proper $\mathbf{\Pi}^q_i$ formula. For simplicity, assume that D is a prenex formula, i.e.,

$$D = \forall \vec{y} B(\vec{y})$$

where B is a proper $\mathbf{\Sigma}^q_{i-1}$ formula. Because D is $\mathbf{\Pi}^q_i$, the only $\mathcal{B}_i$-expansion of D is D itself. Therefore $\mathcal{S}'_1$ has the form

$$\Lambda(\overrightarrow{q^1}, \overrightarrow{q^2}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}), B(\overrightarrow{q^2}), \Delta', \Omega'$$

We modify $\mathcal{S}'_1$ and its proof by replacing $\overrightarrow{q^2}$ by $\overrightarrow{q^1}$; from this we can obtain the sequent $\mathcal{S}'$ defined below by a contraction-right:

$$\mathcal{S}' \quad = \quad \Lambda(\overrightarrow{q^1}, \overrightarrow{q^1}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}), \Delta', \Omega'$$

Case 4c: $\mathcal{S}$ is obtained by contraction-right on a proper $\mathbf{\Sigma}^q_i$ formula. That is, $\mathcal{S}$ is derived as in (9), but here D is a proper $\mathbf{\Sigma}^q_i$ formula. For simplicity assume that D is a prenex formula, i.e.,

$$D = \exists \vec{x} C(\vec{x})$$

where C is a proper prenex $\mathbf{\Pi}^q_{i-1}$ formula. As above, D is the only $\mathcal{B}_i$-expansion of itself, so by the induction hypothesis, we have a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof of a sequent

$$\mathcal{S}'_1 \quad = \quad \Lambda(\overrightarrow{e^1}, \overrightarrow{e^2}), \Gamma', \Upsilon \longrightarrow C(\overrightarrow{e^1}), C(\overrightarrow{e^2}), \Delta', \Omega'$$

Intuitively, there are two sets $\overrightarrow{e^1}$ and $\overrightarrow{e^2}$ of witnesses for $\exists \vec{x} C(\vec{x})$. We compute one set of witnesses $\overrightarrow{e^3}$ by defining

$$e^3_i \leftrightarrow \big((C(\overrightarrow{e^1}) \wedge e^1_i)) \vee (\neg C(\overrightarrow{e^1}) \wedge e^2_i)\big)$$

Let F_i denote $(C(\overrightarrow{e^1}) \wedge e^1_i)) \vee (\neg C(\overrightarrow{e^1}) \wedge e^2_i)$. We define the sequent $\mathcal{S}'$ to be (here ℓ is the length of $\overrightarrow{e^1}$):

$$\Lambda(\overrightarrow{e^1}, \overrightarrow{e^2}), e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, \Gamma', \Upsilon \longrightarrow C(\overrightarrow{e^3}), \Delta', \Omega'$$

Note that here the definition of the extension variables $\overrightarrow{e^3}$ are $\mathcal{B}_{i-1}$; these can be turned into $\mathbf{\Sigma}^q_{i-1}$ definitions as in the remark following Definition 3.3. The sequent $\mathcal{S}'$ can be obtained from $\mathcal{S}'_1$ and the following sequents

$$e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, C(\overrightarrow{e^1}) \longrightarrow C(\overrightarrow{e^3}) \tag{10}$$

$$e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, C(\overrightarrow{e^2}) \longrightarrow C(\overrightarrow{e^1}), C(\overrightarrow{e^3}) \tag{11}$$

These two sequents can be derived in $\mathbf{G}^\star_{i-1}$, e.g., see [CN10, Lemma VII.4.11]. Here we need to use the cut rule on $\mathbf{\Pi}^q_{i-1}$ formulas.

Case 4d: $\mathcal{S}$ is obtained by contraction-right on other $\mathcal{B}_i$ formulas. This case can be seen as a combination of Cases 4b and 4c. Although this case is more complicated, it can be handled in the same way as before. Suppose for simplicity that

$$D = \forall \vec{y} B(\vec{y}) \wedge \exists \vec{x} C(\vec{x})$$

where B is a proper $\mathbf{\Sigma}^q_{i-1}$ formula, and C is a proper $\mathbf{\Pi}^q_{i-1}$ formula. Because D is in $\mathcal{B}_i$, it is the only $\mathcal{B}_i$-expansion of itself. So $\mathcal{S}'_1$ has the form

$$\Lambda(\overrightarrow{q^1}, \overrightarrow{q^2}, \overrightarrow{e^1}, \overrightarrow{e^2}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1}), B(\overrightarrow{q^2}) \wedge C(\overrightarrow{e^2}), \Delta', \Omega'$$

We proceed first as in Case 4b, rename the $\overrightarrow{q^2}$ so that they become $\overrightarrow{q^1}$. Thus $\mathcal{S}'_1$ becomes

$$\Lambda(\overrightarrow{q^1}, \overrightarrow{q^1}, \overrightarrow{e^1}, \overrightarrow{e^2}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1}), B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^2}), \Delta', \Omega'$$

Now we proceed as in Case 4c, i.e., we introduce new extension variables $\overrightarrow{e^3}$ with definitions $e^3_i \leftrightarrow F_i$ where

$$F_j = \big(C(\overrightarrow{e^1}) \wedge e^1_j\big) \vee \big(\neg C(\overrightarrow{e^1}) \wedge e^2_j\big)$$

(As remarked earlier, although the defining formulas F_j are in $\mathcal{B}_{i-1}$ we can introduce new extension variables and make all defining formulas $\mathbf{\Sigma}^q_{i-1}$.) As in Case 4c we can prove sequents (10) and (11). From these we can derive

$$e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1}) \longrightarrow B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^3})$$
$$e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^2}) \longrightarrow B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1}), B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^3})$$

(In general we need to prove this by induction on the structure of D.) Then we can obtain the following sequent $\mathcal{S}'$:

$$\Lambda(\overrightarrow{q^1}, \overrightarrow{q^1}, \overrightarrow{e^1}, \overrightarrow{e^2}), e^3_1 \leftrightarrow F_1, \ldots, e^3_\ell \leftrightarrow F_\ell, \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^3}), \Delta', \Omega'$$

(here ℓ is the length of $\overrightarrow{e^1}$). Note that in this case to derive the sequent $\mathcal{S}'$ we need to cut on $\mathcal{B}_{i-1}$ formulas (e.g., in the example above we need to cut the formulas $B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1})$ and $B(\overrightarrow{q^2}) \wedge C(\overrightarrow{e^2})$).

Note also that we could have defined F_j as

$$F_j = \big((B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1})) \wedge e^1_j\big) \vee \big(\neg(B(\overrightarrow{q^1}) \wedge C(\overrightarrow{e^1})) \wedge e^2_j\big)$$

However, this creates undesirable dependence of $\overrightarrow{e^3}$ on $\overrightarrow{q^1}$ which is problematic, e.g., for the proof of Theorem 4.4.

Case 4e: $\mathcal{S}$ is obtained by contraction-right on other formulas. Here $\mathcal{S}$ is obtained as in (9), but D is not in $\mathcal{B}_i$. By the induction hypothesis, $\mathcal{S}_1$ can be transformed to a sequent $\mathcal{S}'_1$ of the form

$$\Lambda, \Gamma', \Upsilon \longrightarrow D'_1, D'_2, \Delta', \Omega'$$

where D'_1 and D'_2 are $(i-1, Q_{\mathcal{S}_1}, E_{\mathcal{S}_1})$-instances of some $\mathcal{B}_i$-expansions of D. We define

$$\mathcal{S}' \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow D'_1 \vee D'_2, \Delta', \Omega'$$

It is derived from $\mathcal{S}'_1$ by $\vee$-right. Because D is non-$\mathcal{B}_i$ and because each $(Q_{\mathcal{S}_1} \cup E_{\mathcal{S}_1})$-variable appears in at most one formula D'_1, D'_2 (condition A5), $D'_1 \vee D'_2$ is also a $(i-1, Q_{\mathcal{S}_1}, E_{\mathcal{S}_1})$-instance of a $\mathcal{B}_i$-expansion of D.

(Note that although we can handle Case 4e as in Case 4d and still have a valid proof, doing so will create unwanted dependencies between variables. Dependencies between variables are discussed in Section 3.2.)

Other cases are handled similarly to the cases above. In particular, the case $\exists$-left, where the principal formula is $\mathbf{\Sigma}^q_{i-1}$, is handled as in Case 2a; if the principal formula is in $\mathbf{\Sigma}^q_i - \mathbf{\Pi}^q_{i-1}$ we proceed as in Case 2b, even though here the principal formula must be an ancestor of a cut formula (and not a subformula of A). The case $\forall$-left is handled as in Case 3a. (Note that by our assumptions, here the principal formula must be in $\mathbf{\Pi}^q_{i-1}$.) The case contraction-left on $\mathcal{B}_{i-1}$ formulas is similar to Case 4a, and contraction-left on proper $\mathbf{\Sigma}^q_i$ formulas is similar to Case 4b. (Again, by our assumptions, there are no contraction-left inferences on other formulas.)

Finally, in each case there is at most an additional factor increase in size that is polynomial in the size of a sequent in the original proof. Furthermore, in each case the transformation can be done by a polytime function and is indeed formalizable in **VPV**. As a result, the proof π' is provably in **VPV** computable by a polytime function. □

3.1 A special case

In this paper we focus on quantified propositional proofs of general formulas. However, we will show here that the Herbrand theorem for $\mathbf{G}^\star_i$ above can be improved slightly if, for example, the formula being proved is prenex. The improvements include a slightly lower complexity of the cut formulas (so we have a $\mathbf{G}^\star_{i-1}$ proof π', instead of a $\mathbf{G}^\star_{\mathcal{B},i-1}$ proof) and a simpler form of the succedent of the final sequent (we have a sequence of instances of A instead of an instance of some expansion of A).

Theorem 3.5. *Let $i \geq 1$ and π be a $\mathbf{G}^\star_i$ proof of a prenex formula $A(\vec{p})$. Then there is a $\mathbf{G}^\star_{i-1}$ proof π' of a sequent*

$$\Lambda \longrightarrow A_1, A_2, \ldots, A_t$$

*where each A_j is an $(i-1, Q, E)$-instance of A. Here $Q = Q_\pi$ and E is a set of new distinct variables, and Λ is a $\mathbf{\Sigma}^q_{i-1}$-cedent defining E in terms of Q and the free variables in A. Moreover, π' is provably in **VPV** computable by a polytime function.*

The theorem can be extended a little to allow for slightly more general formulas A. More specifically, if $i \geq 3$ the theorem holds for formulas $A(\vec{p})$ of the form

$$Q_1 \overrightarrow{x^1} Q_2 \overrightarrow{x^2} \ldots Q_j \overrightarrow{x^j} B(\overrightarrow{x^1}, \overrightarrow{x^2}, \ldots, \overrightarrow{x^j}, \vec{p})$$

where B is a formula in $\mathcal{B}_{i-2}$, and $Q_1, Q_2, \ldots, Q_j$ are either $\exists$ or $\forall$.

Proof of Theorem 3.5. The proof is similar to the proof of Theorem 3.4. We will prove by induction on sequent $\mathcal{S}$ of the proof π that there is a $\mathbf{G}^\star_{i-1}$ proof π' of a sequent $\mathcal{S}'$ as in (5), but here (A3) is replaced by (A3') and (A5) is replaced by (A5'):

(A3') Ω' is obtained from Ω by replacing each non-$(\mathbf{\Sigma}^q_{i-1} \cup \mathbf{\Pi}^q_{i-1})$ ancestor B of A by a list L_B of $(i-1, Q_\mathcal{S}, E_\mathcal{S})$-instances of B,

(A5') each variable in $Q_\mathcal{S} \cup E_\mathcal{S}$ appears in at most one formula in Γ', Δ', or in at most one list L_B for some $B \in \Omega$ (but not both).

The proof is the same as before except for some modifications. We focus on the cases where there are changes to be made. These consist of cases where a cut on a $\mathcal{B}_{i-1}$ formula is required, and cases involving a non-$(\mathbf{\Sigma}^q_{i-1} \cup \mathbf{\Pi}^q_{i-1})$ ancestor of A. Note that Case 4d does not happen here, because all $\mathcal{B}_i$ formulas that appear in π are prenex and are either proper $\mathbf{\Pi}^q_i$ or proper $\mathbf{\Sigma}^q_i$. There remain Cases 3c and 4e.

In Case 3c:

$$\frac{\mathcal{S}_1}{\mathcal{S}} \quad = \quad \frac{\Gamma, \Upsilon \longrightarrow \Delta, \Omega, D(F)}{\Gamma, \Upsilon \longrightarrow \Delta, \Omega, \exists x D(x)}$$

where $\exists x D(x)$ is a non-$\mathbf{\Sigma}^q_{i-1}$ ancestor of the final formula. Here we have

$$\mathcal{S}'_1 \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow \Delta', L_{D(F)}, \Omega'$$

where $L_{D(F)}$ is a list of the form

$$D'(\overrightarrow{q^1}, \overrightarrow{e^1}, F), D'(\overrightarrow{q^2}, \overrightarrow{e^2}, F), \ldots, D'(\overrightarrow{q^t}, \overrightarrow{e^t}, F)$$

These are $(i-1, Q_{\mathcal{S}_1}, E_{\mathcal{S}_1})$-instances of $D(F)$, so in particular D' is either $\mathbf{\Sigma}^q_{i-1}$ or $\mathbf{\Pi}^q_{i-1}$ (depending on the formula A). Define $L_{\exists x D(x)}$ to be

$$D'(\overrightarrow{q^1}, \overrightarrow{e^1}, e), D'(\overrightarrow{q^2}, \overrightarrow{e^2}, e), \ldots, D'(\overrightarrow{q^t}, \overrightarrow{e^t}, e)$$

for some new extension variable e. Define

$$\mathcal{S}' \quad = \quad e \leftrightarrow F, \Lambda, \Gamma', \Upsilon \longrightarrow \Delta', L_{\exists x D(x)}, \Omega'$$

Then $\mathcal{S}'$ can be obtained from $\mathcal{S}'_1$ by combining with the following sequents by repeated applications of the cut rule (the cut formulas are in $\mathbf{\Sigma}^q_{i-1} \cup \mathbf{\Pi}^q_{i-1}$):

$$e \leftrightarrow F, D'(\overrightarrow{q^j}, \overrightarrow{e^j}, F) \longrightarrow D'(\overrightarrow{q^j}, \overrightarrow{e^j}, e).$$

As before, these sequents have polysize cut-free $\mathbf{G}^\star$ proofs.

Case 4e does not require the cut rule. Here the two copies of D give two lists L_1 and L_2 of instances of D. We define

$$\mathcal{S}' \quad = \quad \Lambda, \Gamma', \Upsilon \longrightarrow L, \Delta', \Omega'$$

where L is the concatenation of L_1 and L_2. □

3.2 Partial orders on the variables

An important property of the $\mathbf{G}^{\star}_{\mathcal{B},i-1}$ proof π' given in Theorem 3.4 is the dependence relation between extension variables and other variables. For example, in the student–teacher argument (Appendix A), the value of an extension variable will be adjusted once the counter-examples are given for the Q-variables that it depends on. We will look at this relation now. We will need to verify, for example, that an E-variable does not depend on any other variable appearing inside its scope (see formal definitions below). The proofs of the results in this subsection usually require rigorously examining the cases considered in the proof of Theorem 3.4.

The dependence relation defined in Definition 3.6 is also used by Perron (denoted by $\prec$ in his notation). Note that in our notation (Definition 3.12), $e_1 \prec e_2$ is a weaker than "e_2 depends on some Q-variables in the scope of e_1," see Lemma 3.14.

Definition 3.6. *Consider the proof π' constructed in the proof of Theorem 3.4. Define recursively in polytime a* dependence *relation between the variables in $E \cup Q$ as follows. We say that an extension variable $e \in E$, with defining formula F,* depends *on another variable v in $E \cup Q$ if (i) F contains v, or (ii) F contains another extension variable that depends on v.*

Lemma 3.7 (Provable in **VPV**)**.** *The dependence relation is a partial order and is computable in polytime.*

Proof. The proof is straightforward by examining the proof of Theorem 3.4. To show that the dependence relation is a partial order, we show that it does not contain cycle. We look at the following cases where the dependence relation is updated:

- Case 1: some Q-variables $\vec{q}$ are renamed to be extension variables $\vec{e}$. The variables in the subproof of $\mathcal{S}'_1$ that depend on $\vec{q}$ will now depend on $\vec{e}$. No cycle is produced because the subproof of $\mathcal{S}'_1$ and $\mathcal{S}'_2$ have disjoint sets of Q-variables as well as extension variables.
- Cases 2a, 3b, 3c, 4c: new extension variables are introduced. Here the only new dependence is of the new variables on the existing ones.
- Case 4b: $\overrightarrow{q^2}$ are renamed to $\overrightarrow{q^1}$, but the Q-variables do not depend on other variables, so this renaming does not create cycles.
- Case 4d: some Q-variables $\overrightarrow{q^2}$ are renamed $\overrightarrow{q^1}$, and new extension variables are introduced. The renaming of Q-variables does not create cycle because Q-variables do not depend on other variables. The introduction of new extension variables is as in Cases 2a, 3b, 3c, 4c above.

The fact that the dependence relation is computable in polytime is straightforward. □

Definition 3.8. *Consider the formulas $A^\star$ and A' given in Theorem 3.4, and let v be a variable in $E \cup Q$. Suppose that v replaces x in $\exists x B(x)$ or y in $\forall y B(y)$ (these are subformulas of $A^\star$), then the* scope *of v, denoted by B_v, is the formula in A' corresponding to $B(x)$ (respectively $B(y)$) in $A^\star$. Also, denote by π_v the subproof of π' that is rooted at the sequent where v is introduced.*

Note that B_v contains all variables in the scope of v, and it may contain some variables whose scope contains v.

The following lemma is as expected from the fact that a bound variable does not depend on other bound variables that appear in its scope.

Lemma 3.9 (Provable in **VPV**)**.** *Let $v \in E \cup Q$ and $e \in E$ both appear in A'. Suppose that v is in the scope of e. Then e does not depend on v.*

Proof. In Lemma 3.10 below we prove that e does not depend on any other extension variable appearing in A'. So it remains to prove that e does not depend on Q-variables that appear in its scope. We prove by induction on the sequent $\mathcal{S}$ in π that for the corresponding sequent $\mathcal{S}'$ in π', for any extension variable e that appears in the succedent of $\mathcal{S}'$, e does not depend on any Q-variable that appears in its scope.

For the induction step we go through the cases in the proof of Theorem 3.4 to see how the dependence relation is updated. Case 3c is the only case where an extension variable e with some Q-variable in its scope is introduced. In this case the definition of e is the target formula F. As a target formula, F does not contain any bound variable, so it does not contain any variable in $Q_{\mathcal{S}_1}$ and $E_{\mathcal{S}_1}$. Therefore e does not depend on any variable in $Q_{\mathcal{S}_1} \cup E_{\mathcal{S}_1}$. Also, there is no change in the dependence relation for other variables.

For other cases we verify that the dependences between existing variables are not updated in such a way that makes an e depend on a q in its scope. In Case 1 the only new dependences are between $E_{\mathcal{S}_1}$ and variables in $E_{\mathcal{S}_2} \cup Q_{\mathcal{S}_2}$. Clearly the latter do not appear in the scope of the former. In Cases 2a, 3b and 4c new extension variables are introduced but they do not change the dependence relation on the existing variables. In Case 4b some variables that depend on $\overrightarrow{q^2}$ now depend on $\overrightarrow{q^1}$. However $\overrightarrow{q^1}$ do not appear inside the scope of any extension variable. (Note that each variable in $Q_{\mathcal{S}} \cup E_{\mathcal{S}}$ appears in at most one formula in Γ', Δ', Ω'.) Case 4d is similar. In other cases the dependence relation does not change. □

We observe furthermore that:

Lemma 3.10 (Provable in **VPV**)**.** *For any two extension variables e_1, e_2 that appear in A', e_1 does not depend on e_2.*

Proof. We prove by induction on a sequent $\mathcal{S}$ of π that there is no dependence between any two extension variables in $E_{\mathcal{S}}$ that appear in the succedent of $\mathcal{S}'$.

The base case ($\mathcal{S}$ is an axiom) is obvious. For the induction step, consider first the cases where some new extension variables are introduced to the succedent: Cases 3b, 3c, 4c and 4d. In Cases 3b and 3c, the defining formula F is

the target formula and hence cannot contain any variable in $Q_{\mathcal{S}_1} \cup E_{\mathcal{S}_1}$, because target formulas do not contain bound variables. In Cases 4c (and hence also 4d), the defining formulas for the new extension variables $\overrightarrow{e^3}$ contain only extension variables $\overrightarrow{e^1}$ and $\overrightarrow{e^2}$. These appear in the succedent of the sequent $\mathcal{S}'_1$ and are moved to the Λ part in $\mathcal{S}'$. (Note that these variables appear in exactly one formula, i.e., either $C(\overrightarrow{e^1})$ or $C(\overrightarrow{e^2})$, in the succedent of $\mathcal{S}'_1$.) This proves the induction step for Cases 3b, 3c, 4c and 4d.

Now consider Case 1. Here some extension variables in $E_{\mathcal{S}_1}$ may become dependent on $\vec{e}$. However, $\vec{e}$ do not appear in the succedent of $\mathcal{S}'$, and by the induction hypothesis for $\mathcal{S}_2$ $\vec{e}$ do not depend on other extension variables in the succedent of $\mathcal{S}'_2$. Therefore the conclusion for the induction step follows.

Other cases do not change the existing dependence between extension variables in $\mathcal{S}'$. □

Although the above lemma shows that the dependence relation is not applied directly to any two extension variables that appear in A', these variables may depend on each other in an indirect way. For example, suppose that the expansion $A^\star$ of A is

$$A^\star \equiv \exists x_1 \forall y_1 B(x_1, y_1) \vee \exists x_2 \forall y_2 B(x_2, y_2)$$

and

$$A' \equiv B(e_1, q_1) \vee B(e_2, q_2)$$

Then e_2 may depend on q_1. In the student–teacher argument for this case the student finds (and fixes) a value for e_1 first. Now if a counter-example is given for q_1, then this gives a value for e_2 which will be used for the next round. In other words, e_2 depends indirectly on e_1.

We will introduce a binary relation $\prec$ which we can use to check whether e_2 depends indirectly on e_1 as discussed above. First, we need some terminologies. (See also Definition 3.18 below.)

Definition 3.11. *Let u, v be two distinct variables in $E \cup Q$. The* joint *of two subproofs π_u, π_v is defined to be their least common superproof. The two subproofs are said to* meet *at the root of their joint. In addition, they are said to* meet *at an inference rule (such as $\wedge$-left, etc) if this is the very rule that is used to derive the root of their joint.*

Definition 3.12. *Let e_1 and e_2 be two E-variables that appear in A'. We say that $e_1 \prec e_2$ (also $e_2 \succ e_1$) if*

- *either π_{e_1} contains π_{e_2} as a subproof, or*
- *π_{e_1} and π_{e_2} meet at a cut where the cut formula is a proper $\mathbf{\Sigma}^q_i$ formula as in Case 1, and π_{e_1} is on the right branch (subproof of $\mathcal{S}'_2$) and π_{e_2} is on the left branch (subproof of $\mathcal{S}'_1$).*

Observe that this definition only applies to extension variables that are introduced as in Case 3c. (Observe also that all extension variables that might remain in the succedents of π' are introduced in Cases 3b, 3c, 4c and 4d.)

Lemma 3.13 (Provable in **VPV**). *The $\prec$-relation is transitive.*

Proof. The proof is straightforward. □

Lemma 3.14. *Let e_1, e_2 be two E-variables appearing in A'.*

(a) Suppose that e_2 depends on some Q-variable in the scope of e_1, then $e_1 \prec e_2$.

(b) Suppose that $e_1 \prec e_2$, then e_1 does not depend on any Q-variable in the scope of e_2.

Proof. (a) Suppose that e_2 depends on a Q-variable q in the scope of e_1. First we show that π_{e_1} is not a subproof of π_{e_2}. Suppose for the contrary that π_{e_2} contains π_{e_1}. Note that e_1 and e_2 must be introduced as in Case 3c (see the remark after Definition 3.12). Now the defining formula F for e_2 (as in Case 3c) does not contain any variable in $Q_{\mathcal{S}}$, and in fact at the time when e_2 is introduced it does not depend on any variable in $Q_{\mathcal{S}}$. The only way the dependence relation between e_2 and q can change (so that e_2 becomes dependent on q) is for q to be involved in contraction as in Cases 4b or 4d. However in $\mathcal{S}$ the formula containing q is already not in $\mathcal{B}_i$ (it is at least $\mathbf{\Sigma}^q_{i+1}$) so these cases do not apply.

Now suppose that π_{e_1} and π_{e_2} are disjoint. Let π_0 denote their least common superproof. The last inference in π_0 must be a binary rule. If the last inference in π_0 is not a cut, then in π_0 e_2 does no depend on q, and as in the previous paragraph it remains so throughout the proof. The same argument applies if that inference is a cut but π_{e_1} belongs to the left branch and π_{e_2} belongs to the right branch, because as in Case 1 only variables in the left branch become dependent on variables on the right branch, and so e_2 can never become dependent on q.

(b) The argument is similar to (a). First, suppose that π_{e_1} contains π_{e_2}. Then at the time e_1 is introduced as in Case 3c, its defining formula does not contain any variable in $Q_{\mathcal{S}} \cup E_{\mathcal{S}}$. Now we prove by induction on subsequent sequents of the original proof that the dependence relation between e_1 and the Q-variables in the scope of does not change.

The case where π_{e_1} is joined with π_{e_2} by a cut is similar. □

Lemma 3.15 (Provable in **VPV**). *Let e_1, e_2 be two extension variables that appear in A'. If e_2 appears in the scope of e_1, then $e_1 \prec e_2$.*

Proof. The proof is straightforward by examining the cases where e_1 is introduced (Cases 3b, 3c, 4c, 4d). □

Theorem 3.16 (Provable in **VPV**). *The relation $\prec$ is a partial order on the extension variables and is computable in polytime.*

Proof. The fact that $\prec$ is computable in polytime follows from the facts that the transformation of π is done in polytime. It remains to show that $\prec$ is a partial order. Suppose for a contradiction that there are extension variables $e_1, e_2, \ldots, e_k$ so that

$$e_1 \succ e_2 \succ \ldots \succ e_k \succ e_1$$

We can prove by induction on $1 \leq i \leq k$ that π_{e_i} either contains π_{e_1} or is to the right of π_{e_1}. In particular, π_{e_k} either contains π_{e_1} or is to the right of it. But this contradicts the fact that $e_k \succ e_1$. □

Corollary 3.17 (Provable in **VPV**)**.** *There are $\prec$-minimal variables in E, and every such a variable does not depend on any Q-variable.*

Proof. Follows from Theorem 3.16 and Lemma 3.14 (b). □

The next definition is useful for stating an important connection between the dependence relation and the structure of A'. See also Definition 3.11.

Definition 3.18. *Let u, v be two distinct variables in $E \cup Q$. The least super-formula of B_u and B_v is called the* joint *of u and v. We say that u and v* meet *at an $\wedge$-gate (resp. an $\vee$-gate) if their joint is a conjunction (resp. disjunction).*

For example, suppose that $A^\star$ is

$$\exists x_1 \big(\exists x_2 \forall y_1 C(x_1, x_2, y_1) \vee \exists x_3 \forall y_2 C(x_1, x_3, y_2)\big)$$

and suppose that

$$A' = C(e_1, e_2, q_1) \vee C(e_1, e_3, q_2)$$

Then $B_{e_1} = A'$, $B_{e_2} = C(e_1, e_2, q_1)$, $B_{e_3} = C(e_1, e_3, q_2)$, and e_2, e_3 meet at an $\vee$-gate. Generally, observe that the only variables in $E \cup Q$ that can appear in B_v are those whose scope contains v, and those which are contained in v's scope.

Lemma 3.19. *Let e_1, e_2 be two extension variables appearing ing A'. Suppose that $e_1 \prec e_2$, then they meet at an $\vee$-gate.*

Proof. By definition, $e_1 \prec e_2$ implies that the two subproofs π_{e_1} and π_{e_2} meet at a quantifier introduction or a cut. In particular, in the sequent where they meet, the two corresponding scopes, namely B_{e_1} and B_{e_2}, appear as separate formulas in the succedent. Therefore the joint of e_1 and e_2 must be formed by an $\vee$-right inference, which implies that they meet at an $\vee$-gate. □

4 The Herbrand theorem for $\mathbf{G}^\star_{\mathcal{B},i}$

First, we show that it suffices to work with some simplified version of $\mathbf{G}^\star_{\mathcal{B},i}$. This is defined using the following notion. A *monotone boolean combination* (or just *monotone combination*) of some formulas is a boolean combination of the formulas that does not involve $\neg$. For example, a monotone combination

of atoms is a monotone formula. The next theorem allows us to simplify $\mathbf{G}^\star_{\mathcal{B},i}$ proofs so that all cut formulas are monotone combinations of prenex $\mathbf{\Sigma}^q_i$ and $\mathbf{\Pi}^q_i$ formulas.

Theorem 4.1. *Let $\hat{\mathbf{G}}^\star_{\mathcal{B},i}$ be $\mathbf{G}^\star_{\mathcal{B},i}$ with cut formulas restricted to monotone boolean combinations of prenex formulas in $\mathbf{\Sigma}^q_i \cup \mathbf{\Pi}^q_i$. Then $\hat{\mathbf{G}}^\star_{\mathcal{B},i}$ p-simulates $\mathbf{G}^\star_{\mathcal{B},i}$.*

Proof. This theorem is proved in the same way as a theorem of Morioka [Mor05, Theorem 5.13], see also [CN10, Theorem VII.4.7]. The idea is first to walk down the proof and show, for example, that a cut with cut formula of the form $\neg(B \wedge C)$ can be replaced by a cut whose cut formula is $(\neg B \vee \neg C)$, etc. This gives a proof where all cut formulas are monotone combinations of $\mathbf{\Sigma}^q_i$ and $\mathbf{\Pi}^q_i$ formulas. A similar idea can be used to show that such cuts can be further replaced by cuts where all $\mathbf{\Sigma}^q_i$ and $\mathbf{\Pi}^q_i$ subformulas of the cut formulas are prenex. □

Now we prove the Herbrand theorem for $\mathbf{G}^\star_{\mathcal{B},i}$ similar to Theorem 3.4. It should be compared to Theorem 3.4 where i is replaced by $i+1$.

Theorem 4.2 (The Herbrand theorem for $\mathbf{G}^\star_{\mathcal{B},i}$, provable in $\mathbf{VPV}$)**.** *Let $i \geq 1$. There is a polytime function that given a $\mathbf{G}^\star_{\mathcal{B},i}$ proof π of a formula A outputs a $\mathbf{G}^\star_i$ proof π' of a sequent of the form*

$$\Lambda \longrightarrow A' \tag{12}$$

where A' is an $(\mathbf{\Pi}^q_i, Q, E)$-instance of a $\mathcal{B}_{i+1}$-expansion $A^\star$ of A; Q, E are some sets of new distinct variables; and Λ is a $\mathbf{\Sigma}^q_i$-cedent defining E in terms of Q and the free variables in A. Moreover, the dependence and $\prec$ relations, defined exactly as in Subsection 3.2, satisfy the properties proved there.

We modify the proof of Theorem 3.4 to obtain a proof of the current theorem. We are not able to reduce the complexity of the cedent Λ, e.g., to $\mathbf{\Sigma}^q_{i-1}$. If we could do this, we would be able to show that $\mathbf{V}^i$ proves the soundness of $\mathbf{G}^\star_{\mathcal{B},i}$ with respect to $\mathbf{\Sigma}^q_i$ formulas, and this would mean that $\mathbf{G}^\star_i$ p-simulates $\mathbf{G}^\star_{\mathcal{B},i}$. Informally, the main difficulty is that here we need to keep intact all $\mathbf{\Pi}^q_i$ formulas in the proof. This is because of the way that we handle the cut rule (Case 1). Consequently, in Case 2a below we need to use extension variables with $\mathbf{\Pi}^q_i$ (or $\mathbf{\Sigma}^q_i$) definitions.

Consider, for example, an instance of the cut rule in a $\mathbf{G}^\star_{\mathcal{B},1}$ proof π:

$$\frac{\mathcal{S}_1 \qquad \mathcal{S}_2}{\mathcal{S}} = \frac{F, \Gamma_1 \longrightarrow \Delta_1 \qquad \Gamma_2 \longrightarrow F, \Delta_2}{\Gamma_1, \Gamma_2 \longrightarrow \Delta_1, \Delta_2}$$

Suppose that F is a $\mathcal{B}_1$ formula of the form

$$\big(\exists x_1 C_1(x_1) \wedge \forall y_1 D_1(y_1)\big) \vee \big(\exists x_2 C_2(x_2) \wedge \forall y_2 D_2(y_2)\big)$$

where C_1, C_2, D_1, D_2 are quantifier free. Suppose also that we try to transform π by replacing the copy of F in $\mathcal{S}_1$ by an instance

$$(C_1(q_1) \wedge D_1(e_1)) \vee (C_2(q_2) \wedge D_2(e_2))$$

and replacing the copy in $\mathcal{S}_2$ by

$$(C_1(e_1') \wedge D_1(q_1')) \vee (C_2(e_2') \wedge D_2(q_2'))$$

Now, it may happen that e_1 depends on q_1, and e_2 depends on q_2, while e_1' depends on q_2' and e_2' depends on q_1'. As a result, we cannot unify these two copies by replacing q_t by e_t' and q_t' by e_t (for $t = 1, 2$), because this creates a cyclic dependence in the following order: $e_1', e_2, e_2', e_1, e_1'$.

Before proving Theorem 4.2, observe that in the example given above we need an F which is a depth-2 boolean combination of $\mathbf{\Sigma}_1^q$ and $\mathbf{\Pi}_1^q$ formulas. This is because the assumption on the dependences between e_t and q_t and between e_t' and q_t' imposes certain structure on F which forces it to have depth at least 2. (see also Lemma 3.19). In fact, we can show that depth-1 $\mathbf{G}^\star_{\mathcal{B},i}$ is p-equivalent to $\mathbf{G}_i^\star$ w.r.t. proving $\mathbf{\Sigma}_{i+1}^q$ tautologies. First we need a formal definition.

Definition 4.3. *For $d \geq 0$,* depth-d $\mathbf{G}_{\mathcal{B},i}$ *(or just d-$\mathbf{G}_{\mathcal{B},i}$) is the subsystem of $\mathbf{G}_{\mathcal{B},i}$ where all cut formulas are depth-d monotone combinations of formulas from $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$.*

In particular, 0-$\mathbf{G}^\star_{\mathcal{B},i}$ is $\mathbf{G}_i^\star$, while 1-$\mathbf{G}^\star_{\mathcal{B},i}$ is the subsystem of $\mathbf{G}^\star_{\mathcal{B},i}$ where all cut formulas are conjunctions or disjunctions of $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$ formulas.

Theorem 4.4 (Herbrand theorem for 1-$\mathbf{G}^\star_{\mathcal{B},i}$)**.** *Theorem 3.4 continues to hold even if π is a 1-$\mathbf{G}^\star_{\mathcal{B},i}$ proof. Moreover, the dependence and $\prec$ relations on $Q \cup E$ continue to satisfy the properties listed in Subsection 3.2.*

Corollary 4.5. *For $i \geq 1$ the theory $\mathbf{V}^i$ proves the Reflection principle for 1-$\mathbf{G}^\star_{\mathcal{B},i}$ with respect to $\mathbf{\Sigma}_{i+1}^q$ formulas. With respect to proving $\mathbf{\Sigma}_{i+1}^q$ formulas, 1-$\mathbf{G}^\star_{\mathcal{B},i}$ and $\mathbf{G}_i^\star$ are p-equivalent.*

Proof. The fact that $\mathbf{V}^i$ proves the reflection principle for 1-$\mathbf{G}^\star_{\mathcal{B},i}$ with respect to $\mathbf{\Sigma}_{i+1}^q$ tautologies is exactly the same as for the proof of Perron's theorem; see also the appendix. It then follows by a standard argument that $\mathbf{G}_i^\star$ p-simulates 1-$\mathbf{G}^\star_{\mathcal{B},i}$ for $\mathbf{\Sigma}_{i+1}^q$ formulas. □

Now we prove the theorems.

Proof of Theorem 4.2. We will modify the transformation of π given in the proof of Theorem 3.4 by keeping intact all $\mathbf{\Pi}_i^q$ formulas. (Previously, these are replaced by their $\mathbf{\Pi}_{i-1}^q$ instances.) This leads to changes in Cases 1, 2b, and the cases for contraction rule.

Recall Definition 3.2 for the notion of $(\mathbf{\Pi}_i^q, Q, E)$-instances of a formula. In short, a $(\mathbf{\Pi}_i^q, Q, E)$-instance of a $\mathbf{\Sigma}_{i+1}^q$ formula B is a $\mathbf{\Pi}_i^q$ formula obtained from B by replacing the outermost $\exists$-quantifiers by extension variables from E.

We will replace each formula in the succedent of $\mathcal{S}$ by a $(\mathbf{\Pi}_i^q, Q_\mathcal{S}, E_\mathcal{S})$-instance of a $\mathcal{B}_{i+1}$-expansion of it. Note that for an (ancestor of a) cut formula B the $\mathcal{B}_{i+1}$-expansion of B is B itself. For a formula in the antecedent, the roles of $\forall$ and $\exists$ switch, so we will replace an ancestor of a cut formula by a

$(\mathbf{\Pi}_i^q, E_\mathcal{S}, Q_\mathcal{S})$-instance of it. Note the swap of $Q_\mathcal{S}$ and $E_\mathcal{S}$. To avoid confusion, we will call an (Φ, E, Q)-instance a *negative* (Φ, Q, E)-instance.

Formally, by Theorem 4.1 we can assume that all cut formulas in π are monotone combinations of prenex $\mathbf{\Sigma}_i^q$ and $\mathbf{\Pi}_i^q$ formulas. As in the proof of Theorem 3.4, we will assume also that π is in free variable normal form. Finally, for simplicity we assume that in A no quantifier appears in the scope of any $\neg$.

Note that by our assumptions the quantified formulas that appear in the antecedents of π must be ancestors of cut formulas and must be monotone combinations of prenex formulas in $\mathbf{\Sigma}_i^q \cup \mathbf{\Pi}_i^q$.

We will prove by induction on sequents of π that for each $\mathcal{S}$ as above there is a sequent $\mathcal{S}'$ that has a polynomial size $\mathbf{G}_i^\star$ proof. Suppose that $\mathcal{S}$ is of the form

$$\Gamma, \Upsilon \longrightarrow \Delta, \Omega$$

where Γ and Δ consist of all ancestors of cut formulas in $\mathcal{S}$, and Υ, Ω consist of all ancestors of the final formula A. Then $\mathcal{S}'$ is of the following form:

$$\Lambda, \Gamma', \Upsilon \longrightarrow \Delta', \Omega'$$

where for some set $E_\mathcal{S}$ of new distinct variables (and recall that $Q_\mathcal{S}$ is the set of eigenvariables that have been used to introduce quantifiers in the subproof of $\mathcal{S}$):

(B1) Γ' is obtained from Γ by replacing each formula by a negative $(\mathbf{\Pi}_i^q, Q_\mathcal{S}, E_\mathcal{S})$-instance of it,

(B2) Δ' is obtained from Δ by replacing each formula by a $(\mathbf{\Pi}_i^q, Q_\mathcal{S}, E_\mathcal{S})$-instance of it,

(B3) Ω' is obtained from Ω by replacing each formula by a $(\mathbf{\Pi}_i^q, Q_\mathcal{S}, E_\mathcal{S})$-instance of a $\mathcal{B}_{i+1}$-expansion of it,

(B4) Λ is a $\mathbf{\Sigma}_i^q$-cedent defining $E_\mathcal{S}$ in terms of $Q_\mathcal{S}$ and the free variables in $\mathcal{S}$,

(B5) each variable in $Q_\mathcal{S} \cup E_\mathcal{S}$ appears in at most one formula in Γ', Δ', Ω'.

The fact that the dependence and $\prec$ relations satisfy the properties listed in Subsection 3.2 is proved as before. Indeed, the case by case analysis is exactly the same. Thus, in each case considered in the induction step below an appropriate numbering of the extension variables (so that the condition of Definition 3.3 is satisfied) can be obtained using the dependence relation.

Now we present the inductive argument. The base case is straightforward. For the induction step, almost all cases are the same as in the proof of Theorem 3.4; the only cases that require modification are the cases for the cut, $\forall$-right, and contraction rules.

Case 1: $\mathcal{S}$ is derived by a cut:

$$\frac{\mathcal{S}_1 \qquad \mathcal{S}_2}{\mathcal{S}} = \frac{D, \Gamma_1, \Upsilon_1 \longrightarrow \Delta_1, \Omega_1 \qquad \Gamma_2, \Upsilon_2 \longrightarrow D, \Delta_2, \Omega_2}{\Gamma_1, \Gamma_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta_1, \Delta_2, \Omega_1, \Omega_2} \tag{13}$$

For simplicity, assume that D has the form

$$\exists \vec{x} B(\vec{x}) \wedge \forall \vec{y} C(\vec{y})$$

where B is proper prenex $\mathbf{\Pi}^q_{i-1}$, and C is proper prenex $\mathbf{\Sigma}^q_{i-1}$.

By the induction hypothesis we have polynomial size $\mathbf{G}^\star_i$ proofs of:

$$\begin{aligned} \mathcal{S}'_1 &= B(\vec{q}) \wedge \forall \vec{y} C(\vec{y}), \Lambda_1(\vec{q}), \Gamma'_1, \Upsilon_1 \longrightarrow \Delta'_1, \Omega'_1 \\ \mathcal{S}'_2 &= \Lambda_2(\vec{e}), \Gamma'_2, \Upsilon_2 \longrightarrow B(\vec{e}) \wedge \forall \vec{y} C(\vec{y}), \Delta'_2, \Omega'_2 \end{aligned}$$

In order to unify the two instances of the D we will replace $\vec{q}$ by $\vec{e}$. Then by a cut (on the $\mathbf{\Pi}^q_i$ formula $B(\vec{e}) \wedge \forall \vec{y} C(\vec{y})$) we obtain proofs of the following sequent:

$$\mathcal{S}' = \Lambda_1(\vec{e}), \Lambda_2(\vec{e}), \Gamma'_1, \Gamma'_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta'_1, \Delta'_2, \Omega'_1, \Omega'_2$$

The $\forall$-right rule is handled by the following cases:

Case 2a: $\mathcal{S}$ is inferred by $\forall$-right where the principal formula is a $\mathbf{\Pi}^q_i$ formula. This is handled exactly as in Case 2a in the proof of Theorem 3.4, except that here the formula $D(e)$ is $\mathbf{\Pi}^q_i$, and the defining formula $D(\bot)$ for the new extension variable e is also $\mathbf{\Pi}^q_i$.

Case 2b: $\mathcal{S}$ is derived from $\mathcal{S}_1$ by $\forall$-right where the principal formula is a non-$\mathbf{\Pi}^q_i$ ancestor of the final formula A. Here we simply take $\mathcal{S}' = \mathcal{S}'_1$.

The contraction-right rule is handled as follows.

Case 4a: Contraction right on a $\mathbf{\Pi}^q_i$ formula. This case is handled in the same way as Case 4a in the proof of Theorem 3.4. (No cut nor extension variable is required.)

Case 4b: Contraction right on a $\mathbf{\Pi}^q_{i+1}$ formula. Suppose that $\mathcal{S}$ is derived from $\mathcal{S}_1$:

$$\frac{\mathcal{S}_1}{\mathcal{S}} = \frac{\Gamma, \Upsilon \longrightarrow \forall \vec{y} \exists \vec{z} B(\vec{y}, \vec{z}), \forall \vec{y} \exists \vec{z} B(\vec{y}, \vec{z}), \Delta, \Omega}{\Gamma, \Upsilon \longrightarrow \forall \vec{y} \exists \vec{z} B(\vec{y}, \vec{z}), \Delta, \Omega}$$

Here by the induction hypothesis $\mathcal{S}'_1$ has the form

$$\Lambda(\overrightarrow{e^1}, \overrightarrow{e^2}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}, \overrightarrow{e^1}), B(\overrightarrow{q^2}, \overrightarrow{e^2}), \Delta', \Omega'$$

We proceed as in Case 4d in the proof of Theorem 3.4, first rename $\overrightarrow{q^2}$ to $\overrightarrow{q^1}$, then introduce new extension variables $\overrightarrow{e^3}$ that take the value of $\overrightarrow{e^1}$ or $\overrightarrow{e^2}$ depending on $B(\overrightarrow{q^1}, \overrightarrow{e^1})$. Details are left to the reader. Note that here $\overrightarrow{e^3}$ have $\mathbf{\Sigma}^q_{i-1}$ definitions, and we need to cut $\mathbf{\Pi}^q_{i-1}$ formulas.

Case 4c: Contraction right on a $\mathbf{\Sigma}^q_{i+1}$ formula. This is handled as Case 4c in the proof of Theorem 3.4. Here we need to use extension variables with $\mathbf{\Sigma}^q_i$ definitions, and cut on $\mathbf{\Pi}^q_i$ formulas.

Case 4d: Contraction right on other $\mathcal{B}_{i+1}$ formulas. This is similar to Case 4c above. Here we introduce new extension variables with $\mathbf{\Sigma}^q_i$ definitions, and cut on $\mathbf{\Pi}^q_i$ formulas.

Case 4e: Contraction right on other formulas. This is the same as Case 4e in the proof of Theorem 3.4. □

Proof of Theorem 4.4. Suppose that $\mathcal{S}$ is of the form

$$\Gamma, \Upsilon \longrightarrow \Delta, \Omega$$

where Γ and Δ consist of all ancestors of cut formulas in $\mathcal{S}$, and Υ, Ω consist of all ancestors of the final formula A. Then we transform $\mathcal{S}$ into a sequent $\mathcal{S}'$ of the following form:

$$\Lambda, \Gamma', \Upsilon \longrightarrow \Delta', \Omega'$$

where, for some set $E_{\mathcal{S}}$ of new distinct variables,

(C1) Γ' is obtained from Γ by replacing each formula by a negative $(i-1, Q_{\mathcal{S}}, E_{\mathcal{S}})$-instance of it,

(C2) Δ' is obtained from Δ by replacing each formula by a $(i-1, Q_{\mathcal{S}}, E_{\mathcal{S}})$-instance of it,

(C3) Ω' is obtained from Ω by replacing each formula by a $(i-1, Q_{\mathcal{S}}, E_{\mathcal{S}})$-instance of a $\mathcal{B}_i$-expansion of it,

(C4) Λ is a $\mathbf{\Sigma}^q_{i-1}$-cedent defining $E_{\mathcal{S}}$ in terms of $Q_{\mathcal{S}}$ and the free variables in $\mathcal{S}$,

(C5) each variable in $Q_{\mathcal{S}} \cup E_{\mathcal{S}}$ appears in at most one formula in Γ', Δ', Ω'.

We proceed exactly as in the proof of Theorem 3.4, except for Case 1 and Case 4c. Case 4c now handles also ancestors of cut formulas and is dealt with just as before. For Case 1, we consider the cut rule where the cut formula is a depth-1 combination of $\mathbf{\Sigma}^q_i \cup \mathbf{\Pi}^q_i$ formulas:

$$\frac{\mathcal{S}_1 \qquad \mathcal{S}_2}{\mathcal{S}} \quad = \quad \frac{F, \Gamma_1, \Upsilon_1 \longrightarrow \Delta_1, \Omega_1 \qquad \Gamma_2, \Upsilon_2 \longrightarrow F, \Delta_2, \Omega_2}{\Gamma_1, \Gamma_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta_1, \Delta_2, \Omega_1, \Omega_2}$$

Suppose without loss of generality that F is of the form

$$\exists \vec{x} B(\vec{x}) \wedge \forall \vec{y} C(\vec{y})$$

By the induction hypothesis, there are $\mathbf{G}^{\star}_{\mathcal{B},i-1}$ proofs of

$$\begin{aligned} \mathcal{S}'_1 &= \Lambda_1(\overrightarrow{e}, \overrightarrow{q}), B(\overrightarrow{q}) \wedge C(\overrightarrow{e}), \Gamma'_1, \Upsilon_1 \longrightarrow \Delta'_1, \Omega'_1 \\ \mathcal{S}'_2 &= \Lambda_2(\overrightarrow{e'}, \overrightarrow{q'}), \Gamma'_2, \Upsilon_2 \longrightarrow B(\overrightarrow{e'}) \wedge C(\overrightarrow{q'}), \Delta'_2, \Omega'_2 \end{aligned}$$

We replace $\overrightarrow{q}$ by $\overrightarrow{e'}$ and $\overrightarrow{q'}$ by $\overrightarrow{e}$, then combine the two sequents by a cut (on the $\mathcal{B}_{i-1}$ formula $B(\overrightarrow{e'}) \wedge C(\overrightarrow{e})$) to obtain the desired sequent $\mathcal{S}'$:

$$\mathcal{S}' \quad = \quad \Lambda_2(\overrightarrow{e}, \overrightarrow{e'}), \Lambda_1(\overrightarrow{e'}, \overrightarrow{e}), \Gamma'_1, \Gamma'_2, \Upsilon_1, \Upsilon_2 \longrightarrow \Delta'_1, \Delta'_2, \Omega'_1, \Omega'_2$$

To verify that the dependence and $\prec$ relations still satisfy the properties proved in Subsection 3.2, the main task is to verify that the dependence relation does not become cyclic. We need an additional lemma which says that no cycle is created because of the cut rule; for simplicity we will state it for the above scenario:

Lemma 4.6. *For the proof of $\mathcal{S}_2'$ in the example considered in Case 1 above, $\overrightarrow{e'}$ do not depend on $\overrightarrow{q'}$.*

Note that if F was a disjunction, then we would have $\overrightarrow{e}$ not depending on $\overrightarrow{q}$.

Proof. We exploit the fact that F is a conjunction. At the time when $B(\overrightarrow{e'}) \wedge C(\overrightarrow{q'})$ is formed (by the $\wedge$-right rule), $\overrightarrow{e'}$ and $\overrightarrow{q'}$ belong to two disjoint proofs, so there is no dependence between them. It can be verified that this remains so by looking at each of the cases. □

This completes the proof of Theorem 4.4. □

Theorem 4.2 actually follows from the following "strengthening" of Theorem 3.4, using $i+1$ in place of i. Theorem 4.7, however, is not useful in proving Theorem 2.4, because there the extension variables have $\mathbf{\Sigma}_i^q$ definitions, which makes it impossible to formalize in $\mathbf{V}^i$ the student–teacher computation as described in Section A.3.

Theorem 4.7. *Let $i \geq 1$. There is a polytime function that given a $\mathbf{G}_i^\star$ proof π of a formula A outputs a $\mathbf{G}_{i-1}^\star$ proof π' of a sequent of the form*

$$\Lambda \longrightarrow A'$$

where A' is an $(\mathbf{\Pi}_{i-1}^q, Q, E)$-instance of a $\mathcal{B}_i$-expansion $A^\star$ of A; Q, E are some sets of new distinct variables; and Λ is a $\mathbf{\Sigma}_i^q$-cedent defining E in terms of Q and the free variables in A. Moreover, the dependence and $\prec$ relations, defined exactly as in Subsection 3.2, satisfy the properties proved there.

Proof. The theorem is proved by slightly modifying the proof of Theorem 3.4. We prove by induction sequents $\mathcal{S}$ of π that a sequent $\mathcal{S}'$ of the form (5) has a $\mathbf{G}_{i-1}^\star$ proof, where here condition (A3) is replaced by

(A3") Ω' is obtained from Ω by replacing each formula B by a $(\mathbf{\Pi}_{i-1}^q, Q_\mathcal{S}, E_\mathcal{S})$-instance of a $\mathcal{B}_i$-expansion of B.

The proof is exactly the same as before, except for the following cases. In Case 3c the formula $D'(F)$ is now a $\mathbf{\Pi}_{i-1}^q$ formula, so the cut formula is $\mathbf{\Pi}_{i-1}^q$.

In Case 4b, if $i \geq 2$ then D has the form

$$\forall \vec{y} \exists \vec{z} B(\vec{y}, \vec{z})$$

where B is a proper $\mathbf{\Pi}^q_{i-2}$ formula. Therefore $\mathcal{S}'_1$ has the form

$$\Lambda(\overrightarrow{q^1}, \overrightarrow{q^2}, \overrightarrow{e^1}, \overrightarrow{e^2}), \Gamma', \Upsilon \longrightarrow B(\overrightarrow{q^1}, \overrightarrow{e^1}), B(\overrightarrow{q^2}, \overrightarrow{e^2}), \Delta', \Omega'$$

We handle this case as in Case 4c, by renaming $\overrightarrow{q^2}$ to $\overrightarrow{q^1}$ and introducing new extension variables $\overrightarrow{e^3}$ with definitions

$$e^3_j \leftrightarrow \big((B(\overrightarrow{q^1}, \overrightarrow{e^1}) \wedge e^1_j) \vee (\neg B(\overrightarrow{q^1}, \overrightarrow{e^1}) \wedge e^2_j)\big)$$

Here the cut formulas are $\mathcal{B}_{i-2}$. Note that $\overrightarrow{e^3}$ now depend on $\overrightarrow{q^1}$, but this should not be a problem since $\overrightarrow{e^3}$ are in the scope of $\overrightarrow{q^1}$.

Case 4d is as before, taking in to account the changes in Case 4b above. The cut formulas are $\mathbf{\Pi}^q_{i-1}$. □

5 Some open problems

We mention here several questions regarding the systems $\mathbf{G}_i$, $\mathbf{G}^\star_i$ and the new systems $\mathbf{G}^\star_{\mathcal{B},i}$. Most of these questions grow out of curiosity and answering them may not have significant implications. Nevertheless we find them intriguing.

First, although $\mathbf{V}^{i+1}$ is $\mathbf{\Sigma}^B_{i+1}$ conservative over $\mathbf{TV}^i$, we do not know whether $\mathbf{G}^\star_{i+1}$ p-simulates $\mathbf{G}_i$ with respect to $\mathbf{\Sigma}^q_{i+1}$ formulas. Note that $\mathbf{G}_i$ p-simulates $\mathbf{G}^\star_{i+1}$ with respect to all formulas, and $\mathbf{G}^\star_{i+1}$ p-simulates $\mathbf{G}_i$ with respect to $\mathcal{B}_i$ formulas. Similarly, we do not know whether $\mathbf{G}_i$ p-simulates $\mathbf{G}_{\mathcal{B},i}$ with respect to formulas outside $\mathcal{B}_i$.

Also, in the definition of $\mathbf{G}_i$ the target formulas are required to be quantifier-free. We know that with respect to proving $\mathbf{\Sigma}^q_i \cup \mathbf{\Pi}^q_i$ formulas $\mathbf{G}_i$ and $\mathbf{G}^\star_i$ can be defined such that all target formulas are constants ($\bot$ or $\top$). However we do not know whether with respect to proving other formulas, such as $\mathbf{\Sigma}^q_{i+1}$, they can be so defined without superpolynomial blow-up in the size of the proofs.

Putting a proof in free variable normal form is important in some arguments, for example, in the proof that extended Frege p-simulates $\mathbf{G}^\star_1$. Although any treelike proofs can be put in this form, we do not know whether the same is true for daglike proofs. In particular, we do not even know whether a $\mathbf{G}_i$ proof of a $\mathbf{\Sigma}^q_i$ formula can be put into free variable normal form with only a polynomial increase in size.

In this paper we introduce $\mathbf{G}^\star_{\mathcal{B},i}$ but we are not able to prove that $\mathbf{G}^\star_i$ p-simulates $\mathbf{G}^\star_{\mathcal{B},i}$, even with respect to $\mathbf{\Sigma}^q_i$ formulas. The facts that $\mathbf{G}_i$ p-simulates $\mathbf{G}_{\mathcal{B},i}$ with respect to $\mathcal{B}_i$ formulas, and that $\mathbf{G}^\star_i$ p-simulates 1-$\mathbf{G}^\star_{\mathcal{B},i}$ with respect to $\mathbf{\Sigma}^q_{i+1}$ formulas strongly suggest that this is indeed the case. However, the Herbrand theorem for $\mathbf{G}^\star_{\mathcal{B},i}$ that we give does not allow us to prove such a simulation. As a result, we do not know the complexity of the Witnessing problem for $\mathbf{G}^\star_{\mathcal{B},i}$, except for the trivial upper bound provided by the Witnessing problem for $\mathbf{G}^\star_{i+1}$.

Finally, we have not found a first-order theory that translates to $\mathbf{G}^\star_{\mathcal{B},i}$ in the same way that $\mathbf{V}^i$ translates to $\mathbf{G}^\star_i$.

Acknowledgment: I would like to thank Steve Cook and Pavel Pudlák for the conversations and comments. I am indebted to an anonymous referee who provides the example given in Section 2.3, and whose various constructive comments and suggestions help to improve the clarity of the paper throughout. I also benefit from conversations with Emil Jeřábek and Leszek Kołodziejczyk.

References

[Bus86] Samuel Buss. *Bounded Arithmetic*. Bibliopolis, Naples, 1986.

[Bus95a] Samuel Buss. On herbrand's theorem. In *Logic and Computational Complexity*, pages 195–209. Springer–Verlag, 1995. Lecture Notes in Computer Science #960.

[Bus95b] Samuel Buss. Relating the Bounded Arithmetic and Polynomial-Time Hierarchies. *Annals of Pure and Applied Logic*, 75:67–77, 1995.

[CM05] Stephen Cook and Tsuyoshi Morioka. Quantified Propositional Calculus and a Second-Order Theory for $\mathbf{NC}^1$. *Archive for Mathematical Logic*, 44(6):711–749, 2005.

[CN10] Stephen Cook and Phuong Nguyen. *Logical Foundations of Proof Complexity*. ASL Perspectives in Logic Series. Cambridge University Press, 2010.

[CR79] Stephen Cook and Robert Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44(1):36–50, 1979.

[Jeř09] Emil Jeřábek. Approximate counting by hashing in bounded arithmetic. *Journal of Symbolic Logic*, 74(3):829–860, 2009.

[JN10] Emil Jeřábek and Phuong Nguyen. Simulation of $\mathbf{G}_i$ with prenex cuts. unpublished, 2010.

[KP90] Jan Krajíček and Pavel Pudlák. Quantified Propositional Calculi and Fragments of Bounded Arithmetic. *Zeitschrift f. Mathematickal Logik u. Grundlagen d. Mathematik*, 36:29–46, 1990.

[KPT91] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti. Bounded Arithmetic and the Polynomial Hierarchy. *Annals of Pure and Applied Logic*, 52:143–153, 1991.

[Kra95] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory*. Cambridge University Press, 1995.

[KT92] Jan Krajíček and Gaisi Takeuti. On induction-free provability. *Annals of Mathematics and Artificial Intelligence*, 6:107–125, 1992.

[Mor05] Tsuyoshi Morioka. *Logical Approaches to the Complexity of Search Problems: Proof Complexity, Quantified Propositional Calculus, and Bounded Arithmetic*. PhD thesis, University of Toronto, 2005.

[Per08a] Steven Perron. Examining fragments of the quantified propositional calculus. *Journal of Symbolic Logic*, 73(3):1051–1080, 2008.

[Per08b] Steven Perron. *Power of Non-Uniformity in Proof Complexity*. PhD thesis, University of Toronto, 2008.

[Zam96] Domenico Zambella. Notes on Polynomially Bounded Arithmetic. *Journal of Symbolic Logic*, 61(3):942–966, 1996.

A Proving soundness of $\mathbf{G}_i^\star$ in $\mathbf{V}^i$

We give an algorithm that computes the values of the outermost existential variables in the $\mathbf{\Sigma}_{i+1}^q$ tautologies given their $\mathbf{G}_i^\star$ proofs. As a warming-up, we will first consider the soundness of proofs of $\mathbf{\Sigma}_i^q$ formulas (Subsection A.1) and of prenex $\mathbf{\Sigma}_{i+1}^q$ formulas (Subsection A.2). There are relatively simple algorithms for these cases. The case of $\mathbf{\Sigma}_{i+1}^q$ formulas, where all difficulties lie, is discussed in Subsection A.3.

Finally, in order to show that the soundness principles are theorems of $\mathbf{V}^i$, we need to show that the algorithms can be formalized and proved correct in $\mathbf{V}^i$. We refer to [Per08a] and [Per08b, Chapter 5] for details. Essentially, we need the fact that the (number) maximization principle for $\mathbf{\Sigma}_i^B$ formulas are provable in $\mathbf{V}^i$.

A.1 Soundness for proving $\mathbf{\Sigma}_i^q$ formulas

In this subsection we give a proof of a well known fact that the soundness of $\mathbf{G}_i^\star$ with respect to proving $\mathbf{\Sigma}_i^q$ tautologies is provable in $\mathbf{V}^i$. We follow the outline given in [CN10, Theorem X.2.17]. Here we benefit from the transformation given in Theorem 3.4.

Theorem A.1. *For $i \geq 1$ the theory $\mathbf{V}^i$ proves $\mathbf{\Sigma}_i^q$-$RFN_{\mathbf{G}_i^\star}$.*

Proof. Let π be a $\mathbf{G}_i^\star$ proof of a $\mathbf{\Sigma}_i^q$ formula A. We need to argue in $\mathbf{V}^i$ that A is valid. Reasoning in $\mathbf{V}^i$ as follows. By Theorem 3.4 we can transform π into a $\mathbf{G}_{\mathcal{B},i-1}^\star$ proof π' of a sequent of the form

$$\Lambda \longrightarrow A'$$

as in (4). Here because A is $\mathbf{\Sigma}_i^q$, the only $\mathcal{B}_i$-expansion of A is A itself. Therefore A' is an $(i-1, Q, E)$-instance of A.

Now we use the fact that we can prove the soundness of $\mathbf{G}_{\mathcal{B},i-1}^\star$ with respect to proving $\mathcal{B}_{i-1}$ sequents. This follows from the fact that $\mathbf{V}^i$ proves the $\mathbf{\Sigma}_{i-1}^q$-$RFN_{\mathbf{G}_{\mathcal{B},i-1}}$ principle. The latter is already proved by the same proof of [Kra95, Theorem 9.3.16] (see also [CN10, Theorem X.2.17]). Thus we can conclude that the above sequent is valid. We can now compute recursively the values for all extension variables using their definitions given in Λ. These values make Λ true, so they must also make A' true. Therefore we obtain some witnessing values for the outermost existential variables in A as required. □

A.2 Soundness for proving prenex $\mathbf{\Sigma}_{i+1}^q$ formulas

The next simple case of Theorem 2.4 is for proving prenex $\mathbf{\Sigma}_{i+1}^q$ tautologies. Unlike the previous case, here we already need to provide a student–teacher computation, but this computation is very much easier to describe than the computation for the general case of arbitrary $\mathbf{\Sigma}_{i+1}^q$ formulas. For simplicity, we

will consider the case $i = 1$. The argument for $i > 1$ is similar. Thus, suppose that we have a $\mathbf{G}_1^\star$-proof π of a prenex $\mathbf{\Sigma}_2^q$ formula $A(\vec{p})$ of the form

$$A(\vec{p}) \equiv \exists \vec{x} \forall \vec{y} B(\vec{p}, \vec{x}, \vec{y})$$

where B is quantifier-free. We need to argue in $\mathbf{V}^1$ that $A(\vec{p})$ is valid.

By Theorem 3.5 there is a proof π' of some sequent of the form:

$$\Lambda \longrightarrow B(\vec{p}, \overrightarrow{e^1}, \overrightarrow{q^1}), \ldots, B(\vec{p}, \overrightarrow{e^t}, \overrightarrow{q^t}) \tag{14}$$

for some $t \geq 1$, where Λ is a 0-cedent defining the extension variables in some set E (that contains all $\overrightarrow{e^j}$) in terms of Q_π and $\vec{p}$. In $\mathbf{V}^1$ we know that this sequent is valid, because $\mathbf{V}^1$ proves the RFN principle for $\mathbf{G}_1^\star$ w.r.t. $\mathbf{\Sigma}_1^q$ formulas (all formulas in (14) are quantifier-free).

By Theorem 3.16 we can assume that the $\overrightarrow{e^j}$ have been ordered in such a way that $\overrightarrow{e^j}$ do not depend on any Q-variables among $\overrightarrow{q^{j'}}$, for $j' > j$. In particular, by combining with Lemma 3.9, $\overrightarrow{e^1}$ do not depend on any $\overrightarrow{q^j}$, and $\overrightarrow{e^2}$ may depend only on $\overrightarrow{q^1}$, etc.

This facilitates a simple student–teacher computation of a witness for $\vec{x}$ that makes $\forall \vec{y} B(\vec{p}, \vec{x}, \vec{y})$ true. The student starts by initializing all the Q-variables to $\bot$ and evaluating the E-variables accordingly. These make Λ true. The student sends the values for $\overrightarrow{e^1}$ as a candidate to the teacher. Now the counter-example $\overrightarrow{q^1}$ provided by the teacher (if there is any) makes $B(\vec{p}, \overrightarrow{e^1}, \overrightarrow{q^1})$ false, but does not change the value of $\overrightarrow{e^1}$. The student updates the values of other E-variables, and sends $\overrightarrow{e^2}$ as new candidate to the teacher, etc. The computation must stop in at most t rounds, because in the t-th round Λ is true while all $B(\vec{p}, \overrightarrow{e^j}, \overrightarrow{q^j})$ evaluate to $\bot$, for $j < t$, so the teacher cannot provide counter-example $\overrightarrow{q^t}$, because otherwise the sequent (14) is not valid. Therefore the candidate provided by the teacher must be correct.

The above reasoning can be carried out in $\mathbf{V}^1$, using the principle for maximizing the length of a string that satisfies a $\mathbf{\Sigma}_1^B$ formula. This can be done in the same way as described in [Per08a, Theorem 6.1] (reproduced in [Per08b, Theorem 5.1.1]).

A.3 Soundness for proving $\mathbf{\Sigma}_{i+1}^q$ formulas

Now we prove Theorem 2.4. As above we will consider the case $i = 1$. Let π be a $\mathbf{G}_1^\star$ proof of a $\mathbf{\Sigma}_2^q$ formula A. Recall our assumption that the quantifiers in A do not appear inside the scope of any $\neg$. Let π', $A^\star$ and A' be as in the Herbrand theorem for $\mathbf{G}_1^\star$ (Theorem 3.4); so, in particular, $A^\star$ is a $\mathcal{B}_1$-expansion of A and A' is the $(0, Q, E)$-instance of $A^\star$. To show that A is valid, we will show that $A^\star$ is valid, and then use the fact that the equivalence $A \leftrightarrow A^\star$ is provable in $\mathbf{V}^1$ [Per08a, Lemma 6.3].

The fact that $A^\star$ is valid will be proved by exhibiting a student–teacher algorithm that, given π', computes the witnesses for the existentially quantified variables in $A^\star$. To show that this fact is a theorem of $\mathbf{V}^1$ we will prove the correctness of the algorithm in $\mathbf{V}^1$. The idea of the algorithm is almost as before. Here the student will compute in each round the values for all E-variables by setting all Q-variables to $\perp$ in the first round and using the counter-examples from the teacher for Q-variables in subsequent rounds. For the computation to stop, the student needs to make progress in each round. This is achieved by fixing at least one E-variable in each round.

We will now consider the first round. Basically, we need to show that if the teacher is able to give counter-examples, then there must be a formula B_e (recall Definition 3.8) that is evaluated to FALSE under the setting of Q-variables given by the current counter–examples. Furthermore, all extension variables appearing in B_e do not depend on any Q variables, so that the value of B_e will not be changed later.

It is necessary that such an e does not depend on any Q-variable, but this is not sufficient, because the formula B_e may contain other E-variables: these are variables which are in the scope of e, or variables whose scope contains e. In particular, the extension variables whose scope are inside the scope of e may depend on some Q variables. For this reason, we will in fact need to argue that there is a direct variable e (see Definition A.2 below) that satisfies the above conditions. The student will then fix e and all variables whose scope contains e.

Definition A.2. *An E-variable e that appears in A' is called a* direct *(extension) variable if it replaces x in $\exists x D(x)$ for some proper $\mathbf{\Pi}^q_i$ formula D. An E-variable e that replaces x in $\exists x D(x)$ for a proper $\mathbf{\Sigma}^q_{i+1}$ formula D is said to be* indirect*.*

Consider, for instance, the following example,

$$A \;=\; \exists x_1\big(\exists x_2\exists x_3\forall y_1 B(x_1,x_2,x_3,y_1)\wedge\forall y_2 C(x_1,y_2)\wedge\exists x_4 D(x_1,x_4)\big)$$

(we do not display the free variables in A). Suppose that $A^\star$ is

$$\begin{aligned}\exists x_1^1\big((\exists x_2^1\exists x_3^1\forall y_1^1 B(x_1^1,x_2^1,x_3^1,y_1^1)\vee\exists x_2^2\exists x_3^2\forall y_1^2 B(x_1^1,x_2^2,x_3^2,y_1^2))\\ \wedge\forall y_2^1 C(x_1^1,y_2^1)\wedge\exists x_4^1 D(x_1^1,x_4^1)\big)\vee\\ \exists x_1^2\big(\exists x_2^3\exists x_3^3\forall y_1^3 B(x_1^2,x_2^3,x_3^3,y_1^3)\wedge\forall y_2^2 C(x_1^2,y_2^2)\wedge\exists x_4^2 D(x_1^2,x_3^2)\big)\end{aligned}$$

So A' has the form

$$\begin{aligned}\big((B(e_1^1,e_2^1,e_3^1,q_1^1)\vee B(e_1^1,e_2^2,e_3^2,q_1^2))\wedge C(e_1^1,q_2^1)\wedge D(e_1^1,e_4^1)\big)\vee\\ \big(B(e_1^2,e_2^3,e_3^3,q_1^3)\wedge C(e_1^2,q_2^2)\wedge D(e_1^2,e_4^2)\big)\end{aligned}$$

(Here each variable x_j is duplicated multiple times to $x_j^1,x_j^2,\ldots,x_j^{t_j}$.) In this case, e_1^1 is indirect, while e_3^1 is direct.

Now we make some assumptions to simplify the discussion. First, we will treat the E-variables that have the same non-empty set of Q-variables in their

scope together as a single block. In the example above, e_2^1, e_3^1 form one block, e_2^2, e_3^2 form another, e_1^1 is itself a block, etc. We will simply assume that each such block consists of exactly one variable. (So this allows us to define the $\prec$ on blocks.) In other words, we assume that no two E-variables in A' have the same non-empty set of Q-variables in their scope.

Similarly we can define blocks of Q-variables: two Q-variables q_1, q_2 belong to the same block if for every E-variable e either they both belong to the scope of e, or they both do not. Also, we can assume that each Q-block consists of exactly one variable.

Note that the E-variables in A' whose scope does not contain any Q-variable are neither direct nor indirect. However, we will simply assume that these are direct by assigning to each of them a new, distinct "dummy" Q-variable. We will assume in addition that every Q-variable appears in the scope of some direct E-variable, by introducing new, distinct "dummy" direct E-variable if necessary. Thus, for example, in the formula $A^\star$ above, $\forall y_2^1 C(x_1^1, y_2^1)$ becomes $\exists x_C^1 \forall y_2^1 C(x_1^1, y_2^1)$ (here x_C^1 is the new variable introduced only for this subformula). Similarly, $\exists x_4^2 D(x_1^2, x_3^2)$ becomes $\exists x_4^2 \forall y_D^2 (x_1^2, x_3^2)$.

Note that the corresponding new E- and Q-variables (e.g., e_C^1, q_D^2) do not appear in A', nor in the proof π'. However, for a technical reason, if e is a new direct variable which is introduced because of a Q-variable q, then we will let π_e and B_e be the same as π_q and B_q, respectively. As a result, although the new "dummy" variables do not participate in the dependence relation, the new direct variables do participate in the $\prec$ relation by this convention. It can be verified that the results in Section 3.2 (regarding $\prec$) remain true even when the new variables are taken into account.

Lemma A.3. *Suppose that e is a direct variable which is $\prec$-minimal among all direct variables. Then e and all extension variable e' whose scope contains e do not depend on any Q-variable in A'.*

Proof. Because every Q-variable appears inside the scope of some (direct) extension variable, by Lemma 3.14 (a) e does not depend on any Q-variable that appears in A'. Now let e' be any extension variable whose scope contains e. By Lemma 3.15, $e' \prec e$. The variable e' cannot depend on any Q-variable in A', because such a variable q belongs to the scope of some direct variable e'', and if e' depends on q, then this implies that $e'' \prec e'$, and hence $e'' \prec e$ (by Lemma 3.13), a contradiction to the choice of e. □

A.3.1 The first round

Now the first round of the student–teacher computation is as follows. The student sets all Q-variables to $\bot$, and then compute the E-variables using their defining axioms in proper order. He sends the values of E-variables (that appear in A') to the teacher as candidates for the existentially quantified variables in $A^\star$. If these are in fact valid witnesses, then the computation halts and the student has succeeded. Otherwise, the teacher sends a setting of Q-variables

that falsifies A'. The following theorem is crucial for arguing that in this case the student makes some progress.

Theorem A.4 (Provable in **VPV**)**.** *There is some direct variable e such that e is $\prec$-minimal among all direct variables, and that B_e is false under the counter-example provided by the teacher.*

The following corollary follows from the above theorem and lemma.

Corollary A.5. *There is some direct variable e such that (i) e and all extension variables whose scope contains e do not depend on any Q-variable, and (ii) B_e is false under the counter-example provided by the teacher.*

Proof of Theorem A.4. We will use the usual tree structure of A', i.e., the root of the tree is labelled with A', and each inner node is labelled with a subformula of A' so that if a node is labelled with, e.g., $B = C \vee D$, then it has two children which are labelled with C and D, etc. We will prune the tree so that all leaves are labelled with subformulas of the form B_e, where e is an direct variables. Note that the nodes on a path from A' to a leave B_e are labelled with superformulas of B_e. Note also that originally there can be some root–leaf paths in A' whose nodes are not labelled by any subformula of the form B_e; these paths are pruned, because the values of the formulas labelling their nodes stay the same in all rounds.

Lemma A.6. *For each node B in the tree A' there is a path from B to some leaf which is a descendant of B so that all formulas on the path evaluate to the same value as B.*

Proof. The lemma is true because of our assumption that no quantifier appears inside the scope of $\neg$. □

The next definition is important for the following discussion.

Definition A.7. *We say that a subformula B of A' is* involved in making A' false *if B as well as all its superformulas in A' are false. In other words, all subformulas on the path from B to A' evaluate to FALSE.*

Notice that when all Q-variables are $\bot$ (as the student sets them at the beginning of the round), and E-variables are as computed by the student, then A' evaluates to TRUE. (And this is provable in $\mathbf{V}^1$, because A' is a quantifier-free formula that has a $\mathbf{G}_0^\star$ proof, namely π'.) So the reason why A' becomes false under the counter-example from the teacher is that at least one B_q formula as well as all of its superformulas become false because q gets a new value. By our assumption that every Q-variable is inside the scope of some direct variable, it follows that there must be some direct variable e such that B_e is involved in making A' false. Let E_1 be the set of all such direct variables e.

We prove the theorem by contradiction. So assume that for all direct variables e such that B_e evaluates to FALSE, there is some direct variable $e' \prec e$. For each $e \in E_1$ let e' be the direct variable that satisfies the following conditions:

(E1) $e' \prec e$,

(E2) e' is $\prec$-minimal among all direct variables satisfying (E1),

(E3) e' is lexicographically first among all direct variables satisfying (E1) and (E2).

(Here we fix in advance an arbitrary lexicographical ordering on the extension variables.) Under the hypothesis that the theorem is false, these two conditions force $B_{e'}$ to be true. Also, by Lemma 3.19 e and e' meet at an $\vee$-gate, i.e., the joint of B_e and $B_{e'}$ is a disjunction.

Now from e' we compute another direct variable $e'' \in E_1$ as follows. Starting from $B_{e'}$, walk towards the root A', and stop at the last formula which is true. Such formula must exist because $B_{e'}$ is itself true. Call this formula C, then the immediate superformula of C must have the form $C \wedge D$. Note that this must happen before we meet the joint of $B_{e'}$ and B_e. (See Figure 1.) Now D must evaluate to FALSE, and we deterministically take e'' to be a direct variable that is

(E4) involved in making A' false,

(E5) $\prec$-minimal among all direct variables satisfying (E4),

(E6) lexicographically first among all direct variables satisfying (E4) and (E5).

The fact that such an e'' exists which satisfies (i) follows from our choice of C and from Lemma A.6. The relationship between B_e, $B_{e'}$ and $B_{e''}$ is roughly as depicted in Figure 1. (In particular, it can be shown that $e'' \neq e$.)

Observe that the subproof $\pi_{e'}$ and $\pi_{e''}$ meet at the $\wedge$-right rule that introduces the formula $C \wedge D$, and that their joint contains both of them. On the other hand, because $e' \prec e$, by definition, either π_e is a subproof of $\pi_{e'}$, or π_e and $\pi_{e'}$ meet at a cut (whose cut formula is proper $\mathbf{\Sigma}_1^q$) and π_e is on the left branch while $\pi_{e'}$ is on the right branch. Consequently, either π_e and $\pi_{e''}$ meet at the $\wedge$-right rule that introduces $C \wedge D$, or they meet at a cut rule where π_e is on the left branch and $\pi_{e''}$ is on the right branch. Moreover, in the latter case, π_e and $\pi_{e'}$ meet at that same place.

Now we consider a sequence of variables from E_1 as follows. Start with an $e_1 \in E_1$, then obtain e_1', e_1'', C_1, D_1 as above. Let $e_2 = e_1''$, and obtain e_2', e_2'', C_2, D_2 as above. Let $e_3 = e_2''$, etc. This sequence must repeat itself, so without loss of generality, suppose that we obtain a sequence of the form

$$e_1, e_2, \ldots, e_m, e_{m+1} = e_1$$

where all e_j, $1 \leq j \leq m$ are distinct. We have $m \geq 2$, because it can be shown that $e_2 = e_1'' \neq e_1$. We obtain a contradiction by proving the following:

Claim: The joints of the pairs

$$(\pi_{e_1}, \pi_{e_2}), (\pi_{e_2}, \pi_{e_3}), \ldots, (\pi_{e_m}, \pi_{e_1})$$

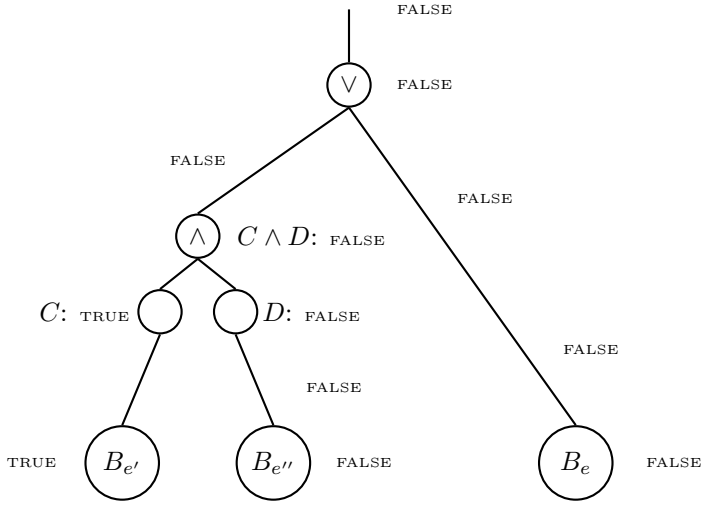

Figure 1: e, e' and e''. B_e and $B_{e''}$ are *involved in making* A' *false.*

are pairwise distinct.

This gives a contradiction, because $\pi_{e_1}, \pi_{e_2}, \ldots, \pi_{e_m}$ are subproofs of a tree-like proof.

To prove the claim, recall that π_{e_j} and $\pi_{e_{j+1}}$ meet either at the $\wedge$-right inference that introduce $C_j \wedge D_j$, or at the cut where π_{e_j} and $\pi_{e'_j}$ meet. (Here $j+1$ is taken modulo m.) Now, suppose that for some j, t such that $1 \leq j < t \leq m$, the pairs $(\pi_{e_j}, \pi_{e_{j+1}})$ and $(\pi_{e_t}, \pi_{e_{t+1}})$ have the same joint. (Again, $t+1$ stands for $(t+1 \mod m)$.) There are two cases to consider, and to prove the claim we will show that they are both impossible.

The first case is where the meeting places of the pairs $(\pi_{e_j}, \pi_{e_{j+1}})$ and $(\pi_{e_t}, \pi_{e_{t+1}})$ are the same $\wedge$-right inference that introduces $C_j \wedge D_j$. This means that either $C_j = C_t$ and $D_j = D_t$, or $C_j = D_t$ and $D_j = C_t$. Notice that if $D_j = D_t$, then because of the way $e_{j+1} = e''_j$ and $e_{t+1} = e''_t$ are computed (see conditions (E4,E5,E6)), we must have $e_{j+1} = e_{t+1}$. This, however, violates the condition that $e_1, e_2, \ldots, e_m$ are distinct. On the other hand, because C_j is evaluated to TRUE and D_j is evaluated to FALSE, we cannot have $C_j = D_t$.

The second case is where π_{e_j} and $\pi_{e_{j+1}}$, as well as π_{e_t} and $\pi_{e_{t+1}}$, meet at a cut (of a proper $\mathbf{\Sigma}^q_1$ formula), and π_{e_j} and π_{e_t} are on the left branch, while $\pi_{e_{j+1}}$ and $\pi_{e_{t+1}}$ are on the right branch. In this case, π_{e_j} and $\pi_{e'_j}$, as well as π_{e_t} and $\pi_{e'_t}$, meet at the same inference. It implies that we have $e_{t'} \prec e_j$. Now the algorithm that computes e'_j (from e_j) implies that e'_j is lexicographically before e'_t. (See conditions (E1,E2,E3).) A symmetric argument shows that e'_t is

lexicographically before e'_j. Thus e'_j and e'_t are the same. But this implies that $e_{j+1} = e_{t+1}$, a contradiction. □

A.3.2 Subsequent rounds

In the next round the student takes a direct variable e as in Theorem A.4, and fixes the value of e and all other E-variables whose scope contains e as he has computed in the previous round. He also fixes the values of all Q-variables in B_e as given by the counter-examples from the teacher. This makes B_e false, and we say that B_e is *fixed to false*. Basically, the student makes progress as more subformulas of A' are fixed to false, because A' is true (provable in $\mathbf{V}^1$), which implies that not all subformulas of A' can be fixed to false. This definition extends inductively to other subformulas in a natural way: if D_1 is fixed to false, then $D_1 \wedge D_2$ is also fixed to false, and if both D_1, D_2 are fixed to false, then so is $D_1 \vee D_2$. In addition, if D is fixed to false, then we also say that all of its subformulas are fixed to false. (Thus a subformula is "fixed to false" here if it is either "fixed to false" or "irrelevant" as in [Per08b, Proof of Theorem 5.1.2].)

For example, suppose that

$$\begin{aligned} A^\star &= \bigvee_{k=0}^{n} \exists x_k \bigwedge_{j=0}^{m_k} \exists x_{k,j} \forall y_{k,j} C_{k,j}(x_k, x_{k,j}, y_{k,j}) \\ A' &= \bigvee_{k=0}^{n} \bigwedge_{j=0}^{m_k} C_{k,j}(e_k, e_{k,j}, q_{k,j}) \end{aligned}$$

Suppose that $e_{0,0}$ is a $\prec$-minimal among all direct variables $e_{k,j}$, and that in the first round the counter-example $q_{0,0}$ falsifies $C_{0,0}(e_0, e_{0,0}, q_{0,0})$. Then $C_{0,0}$ is fixed to false, and hence $\bigwedge_{j=0}^{m_0} C_{0,j}(e_0, e_{0,j}, q_{0,j})$ as well as all $C_{0,j}(e_0, e_{0,j}, q_{0,j})$ are fixed to false. Thus, after this round, e_0, $e_{0,j}$ and $q_{0,j}$ (for $0 \leq j \leq m_0$) are fixed by the students.

Now the second round (and all subsequent rounds) are exactly the same as the first round, except that we have to take into account the fact that several variables and subformulas of A' are fixed to false.

Theorem A.8 (Provable in $\mathbf{V}^1$)**.** *The student succeeds in finding witnesses for $A^\star$ at most $|Q_E|$ many rounds.*

Proof. In each round the student is able to fix at least one new direct variable, and hence falsify at least one new subformula of A'. Because A' is true, the computation must halts in at most $|Q_E|$ many rounds.

The fact that the theorem is provable in $\mathbf{V}^1$ follows from the fact that we can formalize the whole computation by a $\mathbf{\Sigma}^B_1$ formula $\varphi(m)$ which says that the computation lasts for (at least) m rounds. Then the number $\mathbf{\Sigma}^B_1$ maximization principle (provable in $\mathbf{V}^1$) gives us the maximal number of rounds that can happen. At this point we obtain the witnesses for A'. For details of the formalization, see [Per08a]. □