Methods of Witnessing in Bounded Arithmetic

Dimitrios Tsintsilidas

March 2024

1 "Restricting the model" Method

1.1 Herbrand's Theorem

Herbrand's theorem is the most basic result about witnessing existential quantifiers in logically valid first-order formulas of a certain syntactic form ([Her30]).

The result applies to universal theories, theories all of whose axioms are universal sentences of the form $\forall z B(z)$, where B is quantifier-free.

Theorem 1.1. *Let L be a first-order language and let T be a universal L-theory. Suppose $A(x,y)$ is a quantifier-free formula and $T \vdash \forall x \exists y A(x,y)$. Then there are $k \geqslant 1$ and L-terms $t_i(x)$ with $1 \leqslant i \leqslant k$, such that T proves*

$$\bigvee_{1 \leqslant i \leqslant k} A(x, t_i(x)). \tag{1.1}$$

Proof. There are two main steps in this proof method: "compactness theorem" and "restricting the model".

At first, assume that the theory T does not prove any formula in the form of (1.1). We extend T to a new theory T' by adding as axioms all instances of $\neg A$, which means all the sentences of the form $\neg A(x, t(x))$ for every term $t(x)$ in the language L. Every finite subset of the new theory T' has a model, since T is consistent with finite instances of $\neg A$, thus by compactness theorem T' has also a model. We denote that initial model by M.

At the next step, we consider the set S of all L-terms $u(x)$ and define the equivalence relation

$$u \sim v \text{ if and only if } u(m) = v(m) \text{ for all } m \in M.$$

This relation defines $\sim$-equivalence classes $[u]$ for $u \in S$. In this way, we make a new model $N := S/\sim$, which is interpreted as

$$N \models R([u_1], ..., [u_n]) \text{ if and only if } M \models R(u_1(x), ..., u_n(x)),$$

where R is a predicate symbol and $u_1, ..., u_n \in S$. By induction, one can show that for all quantifier-free L-formulas $C(z_1, ..., z_n)$, the same is true:

$$N \models C([u_1], ..., [u_n]) \text{ if and only if } M \models C(u_1(x), ..., u_n(x)),$$

which immediately means that N is also a model of T', since it is a universal theory. However, T' as an extension of T proves $\forall x \exists y A(x,y)$, while for the term $u(x) = x$ and the element $[u] \in N$

$$N \models \neg \exists y A([u], y),$$

because otherwise if $N \models A([u],[t])$, then $M \models A(x, t(x))$, which is not true by the axioms of T'. Therefore, $N \not\models \forall x \exists y A(x,y)$, thus getting a contradiction to the initial hypothesis. □

Remark. The proof presented here is due to [Kra19]. The initial theorem and proof are more general and this version is just a corollary.

A simpler proof would choose at the second step a cut of M with elements generated by some $e \in M$ and all the elements of the form $t(e)$ for $t \in S$. This new model M_e is a substructure of M, thus it satisfies all the axioms of T' by universality, but it is easy to see that it does not satisfy $\forall x \exists y A(x,y)$. This argument is presented in detail in the next proof.

1.2 Parikh's Theorem

Parikh's theorem is first presented in [Par71]. The proof here follows [Kra95] or [Kra19].

Theorem 1.2. *Let $A(x, y)$ be a Δ_0-formula (bounded formula) and T be one of the theories S_2^i, T_2^i, $I\Delta_0$. Assume that*

$$T \vdash \forall x \exists y A(x, y).$$

Then there is a term $t(x)$ in the language of the theory T, such that

$$T \vdash \forall x \exists y \leqslant t(x) A(x, y).$$

Proof. The proof follows the same argument as above. At first, we assume that for any term $t(x)$, T does not prove the sentence

$$\forall x \exists y \leqslant t(x) A(x, y).$$

We extend the language of the theory by adding a new constant e and the theory by adding the axioms:

1. $e > s_n$, for each numeral s_n, for $n \geqslant 1$,
2. $\neg \exists y \leqslant t(e) A(e, y)$, for each term $t(x)$.

We denote the new theory by T'. From the hypothesis, we get that any finite subset of T' has a model, thus T' has a model. Let M be this model. M is a non-standard model because of the axioms $e > s_n$ (and this is the definition of a non-standard model).

Next, we restrict the model M by a cut. A cut in M is any $I \subseteq M$ such that

- I contains all the M-interpretations of constants from the language of the theory and is closed under all functions,
- I is closed downwards, so that if $a < b$ and $b \in I$, then $a \in I$, for all $a, b \in M$.

We denote the cut we use here by I_e and is defined as

$$I_e := \{a \in M \mid \text{there exists a term } t(x) \text{ such that } M \models a \leqslant t(e)\}.$$

This cut is also referred as the Skolem closure of e in M ([KS06]).

By the axioms $\neg \exists y \leqslant t(e) A(e, y)$ and the definition of I_e, it is easy to see that

$$I_e \models \neg \exists y A(e, y).$$

Therefore, if I_e is also a model of T', we have a contradiction, since T' proves $\forall x \exists y A(x, y)$, hence the initial hypothesis was not correct. The axioms we added to T' are easily satisfied by I_e, since they are satisfied by M. It suffices to show that I_e is a model of T.

Firstly, by induction on the number of quantifiers and logical connectives, we can prove that for all bounded formulas $B(z_1, ..., z_n)$ and any $a_1, ..., a_n \in I_e$

$$I_e \models B(a_1, ..., a_n) \text{ if and only if } M \models B(a_1, ..., a_n).$$

The only thing we need to observe for this is that due to the downward closure of cuts, the bounded quantifiers are defined in the same way for M and I_e.

From the fact above, we can simulate the induction of the theory T in I_e and as a result $I_e \models T$ and the proof is complete. □

1.3 KPT Theorem

The KPT theorem is named after its authors in [KPT91] and it extends the result of Herbrand's theorem. The second proof in the original paper follows a similar model-theoretic argument as above. Below we denote by $\square^p_{i+1}$ the class of functions computable by a polynomial-time machine with access to an oracle from the class Σ^p_i (alternatively $FP^{\Sigma^p_i}$).

Theorem 1.3. *Let $i \geqslant 1$ and $\phi(a, x, y)$ an $\exists\Pi^b_i$-formula. If*

$$T^i_2 \vdash \exists x \forall y \phi(a, x, y),$$

then there are $\square^p_{i+1}$-functions $f_1(a), f_2(a, b_1), ..., f_k(a, b_1, ..., b_{k-1})$ with the free variables shown, such that

$$T^i_2 \vdash \phi(a, f_1(a), b_1) \vee \phi(a, f_2(a, b_1), b_2) \vee \cdots \vee \phi(a, f_k(a, b_1, ..., b_{k-1}), b_k).$$

The same holds for PV_{i+1} replacing T^i_2, for $i \geqslant 0$.

Proof. For the proof we need a lemma first. For a more systematic demonstration of these facts, see also [HP17] Chapter V, 4.20-4.26. Actually for point 1, it is proven there that M^* is elementary to M with respect to the closure of Σ^b_1 formulas.

Lemma 1.4. *Suppose M is a model of T^i_2, for $i \geqslant 1$, or of PV_1 for $i = 0$. Let M^* be a subset of M closed under all standard $\square^p_{i+1}$-functions definable in M with parameters from M^*. Then*

1. *M^* is a Σ^b_i-elementary substructure of M (every Σ^b_i formula with parameters from M^* is true in M^* if and only if it is true in M),*
2. *$M^* \models T^i_2$ for $i \geqslant 1$, or $M^* \models PV_1$ for $i = 0$.*

Proof. For (1), if we have a Σ^b_i formula ϕ, we can transform it to a quantifier-free by Skolemization (see Section 2). The new Skolem functions can be defined in T^i_2 by the maximum value that satisfy the respective existential quantifier, since T^i_2 proves the Σ^b_i–MAX principle.

In detail, if we want the Skolemization of a formula $\exists y \leqslant t(x)\psi(x, y)$, then the Skolem function $f(x)$, such that $\exists y \leqslant t(x)\psi(x, y) \equiv \psi(x, f(x))$ can be defined by the formula

$$A(x, y) := (y = 0 \vee \psi(x, y)) \wedge \forall z \leqslant t(x), y < z \rightarrow \neg\psi(x, z).$$

If this formula is written starting with a bounded existential quantifier, then it needs at most $n+1$ bounded quantifiers, using the fact that the initial formula is at most Σ^b_i. For the function $f(x)$ to be definable in T^i_2, we need:

1. For all $x \in M$, $A(x, f(x))$ is true,
2. $T^i_2 \vdash \forall x \exists y A(x, y)$,
3. $T^i_2 \vdash \forall x \forall y \forall z (A(x, y) \wedge A(x, z)) \rightarrow y = z$.

The first one is obvious from the definition and the other two can be proved using the fact that

$$T^i_2 \vdash \psi(x, 0) \rightarrow \exists y \leqslant t(x) \forall z \leqslant t(x), \psi(x, y) \wedge (y < z \rightarrow \neg\psi(x, z),$$

which is an instance of the Σ^b_i–MAX principle.

This means that the Skolem functions are all Σ^b_{i+1}-definable in T^i_2, which by [Bus90] means that they are in $\square^p_{i+1}$. However, M^* is closed under $\square^p_{i+1}$-functions, which means that the Skolemization of ϕ has the same truth value in M and M^*.

For (2), we want to show that the induction scheme Σ^b_i–IND holds in M^*, or equivalently that for any Σ^b_i-formula $\phi(x)$ and any $b \in M^*$ we have:

$$M^* \models \neg\phi(0) \vee \phi(b) \vee (\exists x < b, \phi(x) \wedge \neg\phi(x + 1)).$$

Assume that we have $M^* \models \phi(0) \wedge \neg\phi(b)$ in M^*. From (1), we get that $M \models \phi(0) \wedge \neg\phi(b)$, as well. The PV_{i+1}-axiom is satisfied by M:

$$(\phi(0) \wedge \neg\phi(b) \wedge h(b,b) = (x,y)) \rightarrow (x+1 = y \wedge \phi(x) \wedge \neg\phi(y)).$$

The function $h(b,u)$ is defined in [KPT91] and it is in the class $\square^p_{i+1}$, as it is the projection on the first component of $h(b,b)$. It is easy to prove that $(h(b,b))_1 < b$ and also from the axiom above, if $x = (h(b,b))_1$, then

$$M \models \phi(x) \wedge \neg\phi(x+1).$$

However, $x \in M^*$, since $b \in M^*$ and M^* is closed under $\square^p_{i+1}$-functions. This means that from (1), we get

$$M^* \models \exists x < b, \phi(x) \wedge \neg\phi(x+1),$$

and the proof is complete. □

Back to the original proof, assume for the sake of contradiction that for no k and $f_1, ..., f_k \in \square^p_{i+1}$, T^i_2 proves the required disjunction. Then, we take an enumeration $f_1, f_2, ...$ of all $\square^p_{i+1}$-functions with the properties:

1. The jth function f_j depends on $\leqslant j$ arguments.
2. Each $\square^p_{i+1}$-function occurs in the list infinitely many times.

We consider new constants $c, d_1, d_2, ...$ and we define the new theory

$$T^i_2 + \neg\phi(c, f_1(c), d_1) + \cdots + \neg\phi(c, f_j(c, d_1, ..., d_{j-1}), d_j) + \cdots$$

which is consistent (it has a model) by the compactness theorem, since any finite subset of it is consistent by the hypothesis.

We take a model M of the new theory and define a substructure $M^* \subseteq M$ as

$$M^* := \{f_1(c), f_2(c, d_1), ...\},$$

where $c, d_1, d_2, ...$ are the interpretations of the constants in M. All the projection functions are $\square^p_{i+1}$-definable, which means that $\{c, d_1, d_2, ...\} \subseteq M^*$, and each $\square^p_{i+1}$-function occurs infinitely many times on the list, so M^* is closed under the $\square^p_{i+1}$-functions. These two properties (which are similar to the two properties of cuts) enable us to use Lemma 1.4.

Therefore, $M^* \models T^i_2$ and $M^* \prec_{\Sigma^b_i} M$. However, we also have:

$$M^* \models \forall x \exists y \neg\phi(c, x, y),$$

by getting $y = d_j \in M^*$ for every $x = f_j(c, d_1, ..., d_{j-1})$. This contradicts the original hypothesis that $T^i_2 \vdash \exists x \forall y \phi(c, x, y)$, hence the theorem is proved. □

1.4 Buss's Theorem

Buss's theorem is proved with model theory in [HP17] *(more details soon)*

1.5 A model of S^i_2

[Zam96], [Kra95] *(more details soon)*

2 Herbrandization

Skolemization and Herbrandization are two techniques for reducing quantifier alternations by introducing new functions.

Let $\phi(x) = \exists y \psi(x, y)$ be a formula with x its free variable. The Skolem function for ϕ is a new function represented by a function symbol f_ϕ and has the defining axiom:

$$Sk - def(f_\phi): \quad \forall x \forall y(\psi(x, y) \to \psi(x, f_\phi(x))).$$

The important fact is that this axiom implies:

$$\forall x(\exists y \psi(x, y) \leftrightarrow \psi(x, f_\phi(x))).$$

The process of Skolemization can be continued in two different ways: inside out and outside in. In the second case, we eliminate only the existential quantifiers by starting by the outermost quantifier and defining a Skolem function only if it is existential.

We care for the first case, which is applied in the proof of Lemma 1.4. If the initial formula is in prenex form, we start from the first (innermost) quantifier and define new Skolem functions, as above if the subformula is of the form $\exists \psi$, or to $\exists \neg \psi$ if the subformula is of the form $\forall \psi \equiv \neg \exists \neg \psi$.

The dual notion of Skolemization is Herbrandization. Let $\phi(x) = \forall y \psi(x, y)$ be a formula with x its free variable. The Herbrand function for ϕ is a new function represented by a function symbol h_ϕ and has the defining axiom:

$$Her - def(h_\phi): \quad \forall x \forall y(\neg\psi(x, y) \to \neg\psi(x, h_\phi(x))).$$

The important fact is that this axiom implies:

$$\forall x(\forall y \psi(x, y) \leftrightarrow \psi(x, h_\phi(x))).$$

However, for our purpose, we do not need the Herbrand axiom. By introducing a new function symbol h_ϕ, we have the desired equivalence $\forall y \psi(x, y) \leftrightarrow \psi(x, h_\phi(x))$. If $\forall y \psi(x, y)$ is true (in a model), then $\psi(x, h_\phi(x))$ must also be true, and if $\forall y \psi(x, y)$ is falsified, then there must be a counterexample and the value of $h_\phi(x)$ can be one of these counterexamples.

2.1 KPT Theorem

This is another proof of Theorem 1.3 as presented in [Kra19]. The first proof in the original paper ([KPT91]) follows the same argument. We slightly change the statement using universal theories. Then, Theorem 1.3 is implied, since PV_{i+1} is a universal theory conservative to T^i_2 and its terms are exactly the $\square^p_{i+1}$-functions. Also, the formula $\phi(a, x, y) \in \exists \Pi^b_{i-1}$ (it is not clear how to do Π^b_i) can be transformed to a quantifier-free formula by Parikh's theorem (?) and Skolemization, since Skolem functions are $\square^p_{i+1}$ (see the proof of Lemma 1.4).

Theorem 2.1. *Let T be a universal theory in the language L and $\phi(a, x, y)$ a quantifier-free formula. If*

$$T \vdash \forall x \exists y \forall z \phi(x, y, z),$$

then there are L-terms $t_1(x), t_2(x, z_1), ..., t_k(x, z_1, ..., z_{k-1})$, such that

$$T \vdash \phi(x, t_1(x), z_1) \vee \phi(x, t_2(x, z_1), z_2) \vee \cdots \vee \phi(x, t_k(x, z_1, ..., z_{k-1}), z_k).$$

Proof. We want to apply Herbrandization to the formula $\forall x \exists y \forall z \phi(x, y, z)$. We define the Herbrand function $h(x, y)$ and we consider T as a theory in the new language $L \cup \{h\}$. As we have seen, $\forall x \exists y \forall z \phi(x, y, z)$ is equivalent with $\forall x \exists y \phi(x, y, h(x, y))$. This means that

$$T \vdash \forall x \exists y \phi(x, y, h(x, y)),$$

and now we can apply Herbrand's theorem. As a result, we get that T proves

$$\bigvee_{1 \leqslant i \leqslant k} \phi(x, t'_i(x), h(x, t'_i(x))).$$

We now have to transform this disjunction. We start by finding a subterm s in the disjunction that starts with h and has the maximum size among all such terms. It must be one of the terms sitting at position z (otherwise, if it is a subterm of some t'_i, the term $h(x, t'_i(x))$ has greater size). Assume without loss of generality that it is $h(x, t'_k(x))$. We can replace this term by a new variable z_k without changing the validity of the disjunction, as we have not fixed an interpretation of h. From maximality of the size, this term does not occur in any term t'_i with $i \leqslant k$. We can continue this process by replacing the next maximum size subterm that starts with h by z_{k-1}. This term either occurs in t'_k, so we have just simplified it, or in some of the terms in the position of z, which means that we continue as above. By repeating this process as long as there is no occurrence of the symbol h in the disjunction, we have transformed it to the desired form. □

Remark. KPT theorem can also be derived as a corollary of generalized Herbrand's Theorem as mentioned in [Bus94].

2.2 S^i_2-KP Theorem

The next theorem is a KPT analogue for the theories S^i_2. It was proven in [Kra92] and [Pud92] at the same time, so we name it S^i_2-KP Theorem. The proof here follows the technique used above and it is the proof from [Kra92] (also in [Kra95]).

Theorem 2.2. *Let $\phi(a, x, y)$ be a Σ^b_i-formula and assume that*

$$S^i_2 \vdash \exists x \forall y \leqslant a\ \phi(a, x, y).$$

Then there is a function $g(a)$ such that

$$\mathbb{N} \models \forall n \forall y \leqslant n\ \phi(n, g(n), y),$$

where g is computable by a $\square^p_i$-algorithm (student) which may ask for any (polynomial) number of counterexamples to $\forall y \leqslant a \phi(a, b, y)$. (This class can be denoted by $FP^{\Sigma^p_{i-1}}[wit, poly]$.)

Proof. We apply Herbrandization to the initial formula for the y-quantifier. We add the new symbol f for the Herbrand function and we get the formula

$$\exists x\ f(a, x) \leqslant a \to \phi(a, x, y).$$

From the hypothesis, we get that

$$S^i_2(f) \vdash \exists x\ f(a, x) \leqslant a \to \phi(a, x, f(a, x)),$$

where $S^i_2(f)$ is the theory S^i_2 but in the new language $L \cup \{f\}$ and with axiom $f(a, x) \leqslant a$.

Now we apply Buss's theorem to the relativized S^i_2, thus getting a function F from the class $\square^p_i(f)$, which means computable by a polynomial time Turing machine querying a Σ^p_{i-1} oracle and also values of f. As a result, this function F satisfies

$$f(a, F(a, f)) \leqslant a \to \phi(a, F(a, f), f(a, F(a, f))).$$

As we have seen, the Herbrand function f is some counterexample function, but we need to specify it. So, we substitute it in the computation of $F(a, f)$ by a particular function f^* defined by

$$f^*(a, b) := \begin{cases} \min c \leqslant a \text{ s.t. } \neg\phi(a, b, c) & \text{if it exists} \\ a + 1 & \text{if } \forall y \leqslant a\ \phi(a, b, y). \end{cases}$$

With this definition the formula

$$f^*(a, b) \leqslant a \to \phi(a, b, f^*(a, b))$$

implies

$$\forall y \leqslant a\ \phi(a, b, y)$$

which by $b = F(a, f^*)$ means that

$$\forall y \leqslant a\ \phi(a, F(a, f^*), y),$$

which is the desired result.

If the quantifier $\forall y$ was not argument, we can have the same result, but the algorithm computing F would have a running time polynomial in $|a| + \sum_i |f^*(a, u_i)|$, instead of polynomial in just $|a|$. □

Remark. What is the difference between these two proofs? Both come from Herbrandization of similar results; Herbrand's theorem and Buss's theorem. **I think** the only difference is that Herbrand's theorem uses terms instead of functions, which enables us to transform counterexamples to variables and insert them in the term for the variable in the second position (the one with the existential quantifier). So, the difference is in the languages:
We do not know how many times the counterexample oracle function f is used in the latter case, but in the formal expression of a term t in the former case, we know that the counterexample Herbrand function h can appear only a constant number of times (because the term has a fixed formal expression).

2.3 Wilkie's Witnessing Theorem

[Kra95] 7.3.7

3 Sequent Calculus

Witnessing theorems that are proved by harnessing the sequent calculus proofs and cut elimination:

- Buss's main theorem (Σ_i^b-consequences of S_2^i) [Bus85]
- Buss's proofs of Herbrand's [Bus94] and Parikh's [Bus85] theorems
- Pudlak's proof of S_2^i-KP Theorem [Pud92]
- Krajicek's witnessing of Σ_{i+1}^b-consequences of S_2^i [Kra93]

The last two proofs are improvements on the main theorem of Buss. The last one is the only one which does not have a model-theoretic proof yet.

Remark. The core of Buss's theorem uses only bounded formulas. It transforms a formula $\forall\exists\phi$, where ϕ is bounded, by removing $\forall$ thus making the formula open and changing $\exists$ to bounded by Parikh's theorem. This may indicate that this method does not work well with more quantifiers.

References

[Bus85] Samuel R Buss. *Bounded arithmetic*. Princeton University, 1985.

[Bus90] Samuel R Buss. Axiomatizations and conservation results for fragments of bounded arithmetic. In *Logic and Computation: Proceedings of a Workshop Held at Carnegie Mellon University, June 30-July 2, 1987*, volume 106, page 57. American Mathematical Soc., 1990.

[Bus94] Samuel R Buss. On Herbrand's theorem. In *International Workshop on Logic and Computational Complexity*, pages 195–209. Springer, 1994.

[Her30] Jacques Herbrand. *Recherches sur la théorie de la démonstration*. PhD Thesis, 1930.

[HP17] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. Perspectives in Logic. Cambridge University Press, 2017.

[KPT91] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti. Bounded arithmetic and the polynomial hierarchy. *Annals of pure and applied logic*, 52(1-2), 1991.

[Kra92] Jan Krajíček. No counter-example interpretation and interactive computation. In *Logic from Computer Science: Proceedings of a Workshop held November 13–17, 1989*, pages 287–293. Springer, 1992.

[Kra93] Jan Krajíček. Fragments of bounded arithmetic and bounded query classes. *Transactions of the American Mathematical Society*, 338(2):587–598, 1993.

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

[Kra19] Jan Krajíček. *Proof Complexity*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2019.

[KS06] Roman Kossak and James Schmerl. *The Structure of Models of Peano Arithmetic*. Oxford University Press, 06 2006.

[Par71] Rohit Parikh. Existence and feasibility in arithmetic. *The journal of symbolic logic*, 36(3):494–508, 1971.

[Pud92] Pavel Pudlák. Some relations between subsystems of arithmetic and complexity of computations. In *Logic from Computer Science: Proceedings of a Workshop held November 13–17, 1989*, pages 499–519. Springer, 1992.

[Zam96] Domenico Zambella. Notes on polynomially bounded arithmetic. *The Journal of Symbolic Logic*, 61(3):942–966, 1996.