Extensions of models of bounded arithmetic

Jan Krajíček

Charles University

JAF, September 2025

a correspondence

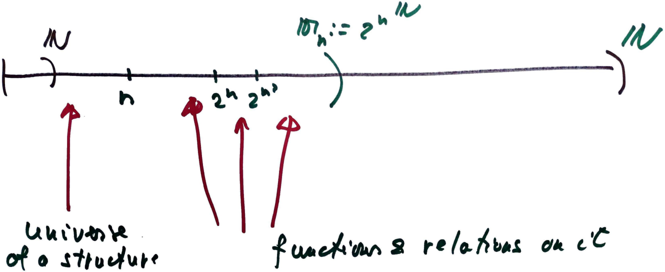
Subsets: $u \subseteq n \; (:= \{0, \ldots, n-1\})$

$$\Updownarrow$$

strings: $u = u_{n-1} \ldots u_1 u_0 \in \{0,1\}^n$

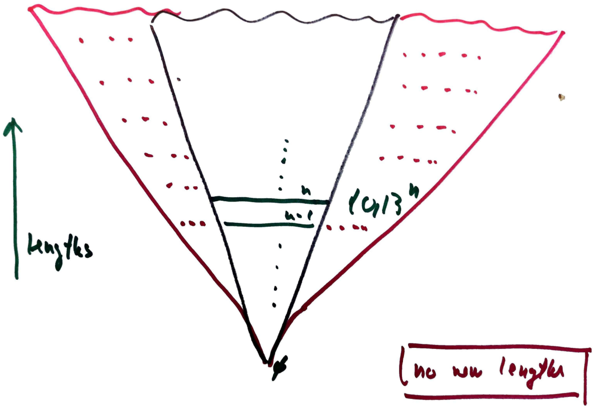
$$\Updownarrow$$

numbers: $2^n \leq u = 2^n + \sum_{i<n} u_i 2^i < 2^{n+1}$.



ℕ
n
2^n
universe of a structure
functions & relations

extensions



example properties

- $u < 2^{n^2}$ is a graph on n and $v < 2^n$ is a clique in it
- $u < 2^n$ is an input string for a p-time Turing machine A and $v < 2^{n^{O(1)}}$ is the computation of A on u
- $u < 2^{n^{O(1)}}$ is a CNF and $v < 2^{n^{O(1)}}$ is its satisfying assignment: $Sat(u, v)$
- $u < 2^{n^{O(1)}}$ is a CNF and $v < 2^{n^{O(1)}}$ is its resolution refutation

All these properties are p-time.

Language L_{PV}:
function (relation) symbols for all p-time clocked Turing machines computing functions (relations). In particular,

- $0, <, suc, \dots$
- $|u| \;:=\; \lceil log_2(u+1) \rceil$, for $u \neq 0$ $(and\ |0| = 0)$
- $i \in u \;\leftrightarrow\;$ the i-th bit of u is 1

Observation

All p-time properties are definable by open formulas and hence are absolute between $\mathbf{M} \subseteq \mathbf{M}'$.

More complex properties:

$$(*) \quad \forall y_1 < n \exists z_1 \forall y_2 < n \exists z_2 \ldots \; openfla(x, \overline{y}, \overline{z})$$

Ex.

u is an n-tuple of propositional formulas

$$(u)_0, \ldots, (u)_{n-1}$$

and

$$\forall i < n \; (u)_i \in SAT$$

i.e.

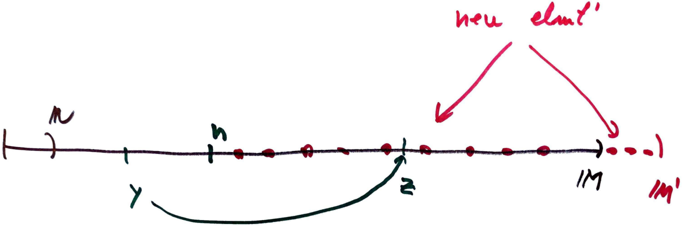
$$\forall i < n \exists v \; Sat((u)_i, v) \; .$$

Denote

$$[n]_{\mathbf{M}} \ := \ \{i \in \mathbf{M} \mid \mathbf{M} \models i < n\} \ .$$

Observation

Assume $\mathbf{M} \subseteq \mathbf{M}'$ while $[n]_{\mathbf{M}} = [n]_{\mathbf{M}'}$. Then all (*) properties are preserved from $\mathbf{M}$ to $\mathbf{M}'$



$\forall y < n \ \exists z \ \ldots\ldots$

a variant of syntactic form (*)

Def. (Buss'85)

Σ_1^b-formulas: when in prenex form all $\forall$ quant's are sharply bounded and all $\exists$ quant's are bounded.

(If the length of parameters is n then $\forall$ are bounded by $n^{O(1)}$ while $\exists$ by $2^{n^{O(1)}}$.)

Observation

Assume $\mathbf{M} \subseteq \mathbf{M}'$ and $Log(\mathbf{M}) = Log(\mathbf{M}')$ where

$$Log(\mathbf{M}) := \{|u| \mid u \in M\} .$$

Then all Σ_1^b-properties are preserved from $\mathbf{M}$ to $\mathbf{M}'$.

a digression: an alternative set-up

Non-standard finite structures in a *finite* language L:

- *universe*: n
- $\mathbf{A}$: an interpretation of L on n coded by an element of $\mathbf{M}$

Instead of extensions of models study

expansions $(\mathbf{A}, R)$

(coded by an element of $\mathbf{M}'$) of $\mathbf{A}$ by interpreting on n a new relation symbol R s.t. a theory in $L(R)$ is satisfied.

a digression: why b.arithmetic models

We could consider non-standard finite 2nd order structures with infinitely many relations:

$$\mathbf{A} := (n, \mathcal{X})$$

where $\mathcal{X}$ is a set of relations (or functions) on n and their expansions

$$\mathbf{A}' := (n, \mathcal{X}')$$

with $\mathcal{X} \subset \mathcal{X}'$ but a number of complications arise; for example, we need to consider properties of $\mathbf{A}'$ involving quantification over elements of $\mathcal{X}'$.

There is also a useful machinery around bounded arithmetic theories (correspondence to pps', propositional translations, witnessing theorems, etc.).

background def's

Def. (Cook-Reckhow'79)

A propositional proof system (abbr. pps) is a p-time function $P : \{0,1\}^* \to \{0,1\}^*$ such that $Rng(P) = TAUT$.

Def.

Theory T_{PV} is the true universal theory in L_{PV}.

The soundness of P: $Ref_P := \forall x, y\ (P(x) = y \to y \in TAUT)$ is in T_{PV}.

Def. (Buss'85)

Σ_1^b-LIND (Length IND):

$$[A(0) \wedge \forall y < |x|(A(y) \to A(y+1))] \to A(|x|)\ .$$

background result

Thm. (K.-Pudlák'90)

Assume

- $\mathbf{M} \models T_{PV} + \Sigma_1^b(PV) - LIND + $ *φ is a propositional formula*

and

- *φ has no proof in* $\mathbf{M}$ *in any pps P.*

Then there is an extension $\mathbf{M}' \supseteq \mathbf{M}$ s.t.

- $\mathbf{M}' \models T_{PV} + \Sigma_1^b(PV) - LIND$,
- $\mathbf{M}' \models \neg\varphi \in SAT$,
- $\mathbf{M}'$ *preserves all $\Sigma_1^b(PV)$-properties from* $\mathbf{M}$.

an extra property

Assuming:

- $\mathbf{M}$ *is countable,*
- $n \in Log(\mathbf{M})$ *s.t.* $\{n^k\}_{k \in \mathbf{N}}$*, are not cofinal in* $Log(\mathbf{M})$

then one can arrange that:

$$[n]_{\mathbf{M}} = [n]_{\mathbf{M}'} .$$

(I.e. no new lengths below n.)

a key question

Do we need $\Sigma_1^b(PV)$-LIND in these results? Specifically:

Problem

Assume $\mathbf{M}$ is countable and

- $\mathbf{M} \models T_{PV} + \varphi$ *is a propositional formula*,
- φ *has no proof in* $\mathbf{M}$ *in any pps* P.

Are there $\mathbf{M} \subseteq \mathbf{M}^* \subseteq \mathbf{M}'$ s.t.

- $\mathbf{M} \preceq \mathbf{M}^*$
  (preservation of $\Sigma_1^b(PV)$ or just (*) formulas would suffice),
- $\mathbf{M}' \models T_{PV} + \neg\varphi \in SAT$,
- $Log(\mathbf{M}^*) = Log(\mathbf{M}')$?

Remarks:

- $T_{PV} \nvdash \Sigma_1^b(PV) - LIND$ unless $NP \subseteq P/poly$ (K.-Pudlák-Takeuti '91).

- For the theorem only the collection scheme $BB\Sigma_1^b(PV)$ - a consequence of $\Sigma_1^b(PV) - LIND$ - suffices but $T_{PV} \nvdash BB\Sigma_1^b(PV)$ either unless factoring is not hard (Cook-Thapen '06).

search problems

Given a pps P consider a total search problem DD_P:

- *Input*: α, π where
  - $P(\pi) = \alpha$,
  - α is a disjoint disjunction $\dot{\bigvee}_i \alpha_i$ (no two disjuncts share an atom).
- *Task*: find i s.t. $\alpha_i \in TAUT$.

(Motivated by the theory of proof complexity generators.)

interactive comp's

Student-teacher computations:

- *Common input*: α, π.
- *Round 1*:
  - S proposes solution i_1,
  - T either approves or sends a counter-example: an assignment w_1 falsifying α_{i_1}.
- *Round 2*:
  - S proposes solution i_2 using also w_1,
  - T either approves or sends a counter-example: an assignment w_2 falsifying α_{i_2}.
- etc. (either until a solution is found or for a predetermined nb. of rounds).

(K.-Pudlák-Sgall'90 formalizing the notion underlying the KPT theorem.)

ST classes

Def.

$ST[\mathcal{F}, t(n)]$ is the class of total Σ_2^p search problems that are solvable on size n inputs in $t(n)$ rounds by a student from the algorithm class $\mathcal{F}$.

Σ_2^p *search problems*:

$$\exists y(|y| \leq |x|^{O(1)}) \forall z(|z| \leq |x|^{O(1)})\ A(x, y, z)$$

with A an open L_{PV}-formula with no other free var's than x.

Hypothesis (ST)

There is a *strong* pps P such that

$$DD_P \notin ST[FP, O(1)] .$$

Remarks:

- FP: the class of p-time alg's,
- strong pps: EF plus a p-time set of tautologies as extra axioms (any pps can be p-simulated by a strong one)
- I think (ST) holds for EF (and hence for all strong pps).

Fact (K.'11 and '20)

(ST) follows from the existence of one-way permutations.

a variant search problem

A variant of DD_P is search problem D_P:

- *Input*: a, α, π where
  - $P(\pi) = \alpha$,
  - $\alpha = \bigvee_i \alpha_i(p, q^i)$ (no two distinct tuples q^i, q^j share an atom),
  - a is a truth assignment to atoms in the tuple p.

- *Task*: find i s.t. $\alpha_i(a, q^i) \in TAUT$.

(Pich-Santhanam '21 considered the possibility that it is in $ST[FP, O(1)]$) for all strong pps P.)

Lemma

For all strong pps P:

$$D_P \in ST[FP, O(1)] \ \leftrightarrow \ DD_P \in ST[FP, O(1)] \ .$$

Theorem

Assume that the model-theoretic problem has the affirmative answer. Then:

$$(ST) \ \rightarrow \ NP \neq coNP \ .$$

Remark:

- (ST) is a *computational complexity* hardness hypothesis: p-time alg's cannot solve a specific task
- $NP \neq coNP$ is a *proof complexity* hardness statement: no pps is p-bounded

Feasible interpolation yields such a reduction for a variety of proof systems but none of them is strong.

prf1

Proof:

We shall assume both (ST) and $NP = coNP$ and derive - using the model-theoretic assumption - a contradiction.

P: a p-bounded pps that also witnesses (ST)

theory S in $L_{PV} \cup \{\alpha, \pi\}$:

- T_{PV},
- $P(\pi) = \alpha$,
- α is of the form $\dot{\bigvee}_{i<m} \alpha_i$,
- $\forall i < m \; (\neg \alpha_i) \in SAT$.

prf2

Claim 1

S is consistent.

Otherwise the KPT theorem would provide $k \geq 1$ and a p-time student S that solves DD_P in $\leq k$ rounds, contradicting (ST).

M: some model of S
(necessarily non-standard)

prf3

Let $c \geq 1$ be s.t. any tautology β has a P-proof of size $\leq |\beta|^c$. We shall abbreviate $[P(\sigma) = \beta \ \wedge \ |\sigma| \leq |\beta|^c]$ by

$$\sigma \ : \ P \ \vdash_* \ \beta \ .$$

The hypothesis $NP = coNP$ implies

Claim 2

For any pps Q: the universal closure of the formula

$$P(x) = \dot{\bigvee}_i (y)_i \ \rightarrow \ Q \ \not\vdash \ ||\forall i < m \ (z)_i : P \not\vdash_* (y)_i||$$

is true and hence in T_{PV}.

prf3 - fla explanation

If the lengths of y and z are a priori bounded we can translate the fla

$$\forall i < m\ (z)_i : P \nvdash_* (y)_i$$

into a propositional circuit:

$$||\forall i < m\ (z)_i : P \nvdash_* (y)_i||$$

of the form

$$\bigwedge_{i<m} \psi(\overline{q}^i, \overline{r}^i)$$

where

- $\overline{q} = (\overline{q}^i)_i$ and $\overline{r} = (\overline{r}^i)_i$ are tuples of atoms representing bits of y and z, resp.,
- $\psi(\overline{q}^i, \overline{r}^i)$ is a circuit expressing that $(z)_i : P \nvdash_* (y)_i$.

prf4

Substitute in the formula in Claim 2

$$x := \pi \quad \text{and} \quad y := \alpha$$

and let

$$\varphi(\overline{r}) \; := \; \bigwedge_{i<m} \psi(\alpha_i, \overline{r}^i)$$

(substitute bits of α for $\overline{q}$).

Claim3

Formula $\varphi(\overline{r})$ has no proof in **M** in any pps Q.

prf5

Now invoke the model-theoretic assumption: there are

$$\mathbf{M} \ \preceq \ \mathbf{M}^* \ \subseteq \ \mathbf{M}' \ \models \ T_{PV} + \neg\varphi(\sigma) = 1$$

for some assignment $\sigma \in \mathbf{M}'$ and so for some $i_0 < m$

$$\mathbf{M}' \models P((\sigma)_{i_0}) = \alpha_{i_0} \ .$$

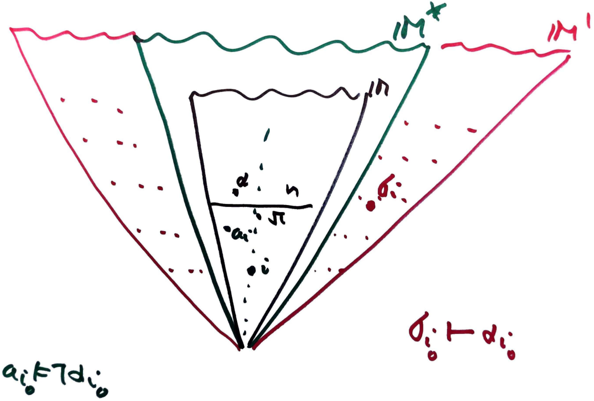
But $Log(\mathbf{M}^*) = Log(\mathbf{M}')$ and hence $i_0 \in Log(\mathbf{M}^*)$ too and thus

$$\mathbf{M}' \ \models \ \neg\alpha_{i_0} \in SAT \ .$$

That contradicts the soundness of P (axiom Ref_P in T_{PV}).

$\square$

a summary pic



A remark:

Using a more precise correspondence

$$\textit{pps}\ \ P\ \leftrightarrow\ \textit{theory}\ \ T_P$$

the proof yields that the model-theoretic assumption for T_P plus the hypothesis

$$DD_P \notin ST[FP, O(1)]$$

implies that P has no strong feasible disjunction property, i.e.

- *some* $\dot{\bigvee}_i \alpha_i$ *has a short P-proof while none of* α_i *does.*

This implies that P is not p-bounded.

Main reference:

- J.K., *On NP ∩ coNP proof complexity generators*, ArXiv 2506.20221v2

References to all other results I mentioned can be found there as well as pointers to a literature offering more background.