Diagonalizable algebras and the length of proofs

Gunnar Adamsson

Magisteruppsats logik
November 2011

Handledare: Fredrik Engström

Institutionen för filosofi, lingvistik och vetenskapsteori
Göteborgs universitet

Diagonalizable algebras and the length of proofs

Gunnar Adamsson

Abstract. We rederive a theorem of Shavrukov [17] on the diagonalizable algebras of PA and ZF using an approach that differs ever-so slightly from the original. This gives a somewhat stronger result (which was implicit in the original proof) that we, coupled with a more careful analysis of the Parikh speed-up phenomenon, put to use by giving examples of how different choices of provability predicates for a given theory T can lead to non-isomorphic algebras. We also show that, by a minor tweak, the injectivity assumption can be dropped and the result extended to arbitrary epimorphisms.

A *diagonalizable algebra* $(\mathfrak{A}, \Box)$ is a boolean algebra $\mathfrak{A}$ equipped with an additional unary operator $\Box$ subject to the following conditions:

$$\Box\top = \top$$
$$\Box(A \to B) \leq \Box A \to \Box B$$
$$\Box A \leq \Box\Box A$$
$$\Box(\Box A \to A) = \Box A$$

'$A \to B$' abbreviates, as usual, '$A^c \cup B$'

A particularly prominent class of examples (actually the *raison d'être* of the whole theory of diagonalizable algebras) consists of the *diagonalizable Lindenbaum algebras*: Given an axiomatized theory T, containing (or interpreting) a large enough fragment of arithmetic to comfortably reason about its own syntax, we obtain a diagonalizable algebra $\mathfrak{D}_\mathrm{T}$ by letting the operator Pr_T act on the ordinary (i.e. boolean) Lindenbaum algebra of T; Pr_T being a suitable formalization of the notion '*... is theorem of T*' (the existence of such formalization(s)[1] is guaranteed by T being able to reason about its syntax).

The diagonalizable nature of $\mathfrak{D}_\mathrm{T}$ is clear: Reading $\Box$ as Pr_T, the first three conditions on $\Box$ are simply the Bernays-Löb derivability conditions (and we assume the formalization Pr_T to be natural enough to satisfy them), whereas the last one comes out as the formalized version of Löb's theorem.

The equational class of diagonalizable algebras (as defined above) was introduced by Magari [11] as an abstract algebraic setting for investigating the self-referential phenomenon discovered by Gödel. This study revolves around the equational theories of diagonalizable algebras and is more conveniently placed in a slightly different setting: that of *provability logic.* The modal logic **GL** (for Gödel and Löb) is the propositional modal logic with the following axioms

I wish to thank Volodya Shavrukov and my supervisor Fredrik Engström for proof-reading, criticism and a large number of suggestions that have much helped to improve this paper.

[1] Of course, neither the theorems of T, nor any particular axiomatization of them, determines Pr_T. In this respect our notation $\mathfrak{D}_\mathrm{T}$ is potentially ambiguous, by corollary 21 below it is (not just potentially but) actually so.

(the second one, as well as its algebraic counterpart above, turns out to be redundant):

$$\Box(A \to B) \to (\Box A \to \Box B)$$
$$\Box A \to \Box\Box A$$
$$\Box(\Box A \to A) \to \Box A$$

The similarity to the conditions defining a diagonalizable algebra is readily apparent. In fact, let $x_0, x_1, \ldots$ be the variables of the theory of diagonalizable algebras and $p_0, p_1, \ldots$ be propositional letters, let $f \colon \mathbb{N} \to \mathbb{N}$ be an arbitrary injective function and define a translation t by induction:

$$x_i^t = p_{f(i)},$$
$$(A \cup B)^t = A^t \vee B^t,$$
$$(A^c)^t = \neg A^t,$$
$$(\Box A)^t = \Box A^t.$$

The axioms defining a diagonalizable algebra are then seen to differ only in inessential respects from the axioms of **GL**, hence an identity $P(x_0, \ldots, x_n) = \top$ holds in every diagonalizable algebra iff $P^t(p_0, \ldots, p_n)$ is a theorem of **GL**. Finding the logical notation more manageable, we will usually prefer it to the algebraic one; consequently we write '$\vdash \Box(\Box A \to A) \to \Box A$' instead of '$\Box(\Box A \to A) \to \Box A = \top$' and similarly. We hope that the occasional mixture of notations will not confuse the reader.

The equational theories[2] of individual $\mathfrak{D}_{\mathrm{T}}$'s were completely classified in the wake of the fundamental work of Solovay [15]: for Σ_1-sound T the $\mathfrak{D}_{\mathrm{T}}$'s all enjoy the same equational theory, moreover: $\mathfrak{D}_{\mathrm{T}}$ is *functionally free* in the class of diagonalizable algebras in the sense that any identity true in $\mathfrak{D}_{\mathrm{T}}$ holds true in *every* diagonalizable algebra. Hence we obtain a simple and decidable axiomatization of the identities of $\mathfrak{D}_{\mathrm{T}}$; they coincide (modulo the translation above) with the theorems of the modal logic **GL**. For Σ_1-ill theories the identities are obtained by adding $\Box^n \bot = \top$, where n is the least number for which $T \vdash \Box^n \bot$ (this n is known as the *credibility extent of* T, if no such n exists T is said to be of *infinite credibility extent*). Hence, the equational theories of all theories of infinite credibility extent are, by the preceding, seen to be the same (namely **GL**). Following along these lines we can obtain one further piece of information: the sentence '$\forall x\,(\Box x = \top \to x = \top)$' holds in $\mathfrak{D}_{\mathrm{T}}$ iff T is Σ_1-sound; so $\mathfrak{D}_{\mathrm{S}} \not\cong \mathfrak{D}_{\mathrm{T}}$ for a Σ_1-sound S and Σ_1-ill T (even though T may be of infinite credibility extent and, hence, share its equational theory with S).

Also, the proper (purely boolean) Lindenbaum algebras of arithmetical theories have long been known to be isomorphic: They are atomless (by Rosser's theorem) and countable (trivial); by a theorem of Tarski they are all isomorphic (to the free boolean algebra on $\aleph_0$ generators).

By a theorem of Shavrukov [16, Proposition 11.9] $\mathfrak{D}_{\mathrm{S}}$ is embeddable into $\mathfrak{D}_{\mathrm{T}}$ for any Σ_1-sound S and T, from the discussion above one might be tempted to suppose that this result could be strengthened to show that, for any Σ_1-sound theories S and T, $\mathfrak{D}_{\mathrm{S}}$ and $\mathfrak{D}_{\mathrm{T}}$ are isomorphic, i.e. that there would essentially be only *one* Σ_1-sound $\mathfrak{D}_{\mathrm{T}}$.

[2] Using the translations t above, the equational theory of $\mathfrak{D}_{\mathrm{T}}$ is seen to coincide with the set $\{P^t \colon \forall t\,(T \vdash P^t)\}$, known as the *provability logic of* T.

This conjecture is refuted by another theorem of Shavrukov, showing that $\mathfrak{D}_{\mathrm{PA}}$ and $\mathfrak{D}_{\mathrm{ZF}}$ are not isomorphic. One object of this paper is to prove this theorem, in doing so we will follow the original proof in [17] closely, departing slightly in certain places to obtain a strengthened result that is used in the proof of corollary 21. The main line of the proof of theorem 17, however, is just as in [17] and the proof runs almost identical to Shavrukov's in many places. Most of the lemmas and auxiliary results, except the ones dealing with the functions $\mathscr{R}_{\mathrm{T}}$, $\mathscr{P}_{\mathrm{T}}$ and $\mathscr{D}_{\mathrm{T}}$, are taken over from [17].

Before launching into the details of the proof some words about the assumptions made on the theories to which our proof applies: in all what follows Σ_1-ill theories are entirely discarded and all theories are assumed Σ_1-sound, furthermore all theories are supposed to be consistent, axiomatized extensions of $\mathrm{I}\Delta_0$+exp, always equipped with a fixed elementary proof predicate.

We introduce the following notation:

Definition 1. Given two explicit Σ_1-formulae $\sigma = \exists x(\delta(x))$ and $\varsigma = \exists x(\delta'(x))$ the *witness comparison* formulae are defined by:

$$\sigma \trianglelefteq \varsigma = \exists x(\delta(x) \wedge \forall y < x(\neg\delta'(y)))$$
$$\sigma \triangleleft \varsigma = \exists x(\delta(x) \wedge \forall y \leq x(\neg\delta'(y)))$$

Furtermore, exponents are used to denote iteration ($\Box^1 = \Box$ and $\Box^{n+1} = \Box\Box^n$), both of various operators like $\Box$ but also of functional composition ($f^{n+1} = f \circ f^n$), $T \vdash \varphi$ indicates that there is a proof of φ from T and $T \vdash_n \varphi$ that there is such a proof containing at most n symbols. The arithmetical sentence expressing that $T \vdash \varphi$ is denoted by $\Box_{\mathrm{T}}\varphi$ (or, when confusion is not likely, just $\Box\varphi$) and the one expressing that $T \vdash_n \varphi$ by $\Box_{\mathrm{T},n}\,\varphi$. Let $\delta(x)$ be as above, we then write more generally $\Box_{\mathrm{T},\delta}\,\varphi$ to indicate that for some m, $T \vdash_m \varphi$ and $\forall y < m(\neg\delta(y))$. Finally, $\varphi(n)\!\downarrow$ means that the function φ is defined for the argument n.

Since we will be concerned with the *length* of proofs we need to ensure that our coding of syntax is reasonably effective, this means that the standard (ineffective) coding of numerals as '$1 + 1 + \cdots + 1$' (or something similar) is not usable. Instead we use a dyadic coding as follows (cf. [9, p. 318]):

$$\bar{0} = \bar{0}$$
$$\bar{1} = S(\bar{0})$$
$$\overline{2n} = (\bar{1} + \bar{1}) * (\bar{n})$$
$$\overline{2n+1} = (\bar{1} + \bar{1}) * (\bar{n}) + \bar{1}$$

This has the important consequence (*not* true for the ineffective coding of numerals above) that if $T \vdash \forall x\,(\varphi(x))$, then $T \vdash_n \varphi(\bar{n})$ holds for all large enough n. This is because the length of the numeral $\bar{n}$ is logarithmic in n and $n - \log(n)$ goes to ∞ with n.

1. The length of proofs

The incompleteness phenomenae discovered by Gödel have further consequences than just the existence of undecidable sentences, by taking also the *length* of (the shortest) proofs into account we arrive at a finer classification than the traditional decidable/undecidable dichotomy; for example, a sentence of manageable length might be provable, but only with proofs of astronomical length or certain sentences might have short proofs in *some* (naturally occurring) theories but only unmanageably long ones in others; such phenomenae carry the technical name *speed-up*.

The first general speed-up result was given by Gödel himself shortly after the publication of his incompleteness theorem:

Theorem 1 (Gödel [7]). *For any total recursive function f and any theory T, there is a sentence φ such that $T \vdash \varphi$, but $T \not\vdash_{f(\bar{\varphi})} \varphi$*

Proof. Take φ such that $T \vdash \varphi \leftrightarrow \neg\Box_{\mathrm{T},f(\bar{\varphi})}\varphi$.
First assume $T \not\vdash \varphi$, but then we have $T \vdash \neg\mathrm{Prf}_{\mathrm{T}}(\bar{n}, \bar{\varphi})$ for any n, so *a fortiori* $T \vdash \neg\Box_{\mathrm{T},f(\bar{\varphi})}\varphi$ and hence $T \vdash \varphi$.
Now assume $T \vdash_{f(\bar{\varphi})} \varphi$, then $T \vdash \Box_{\mathrm{T},f(\bar{\varphi})}\varphi$, hence $T \vdash \neg\varphi$; so T is inconsistent, contrary to assumption. ⊣

In his proof of the theorem mentioned above, Shavrukov used a similar result of Parikh, essentially this result is the observation that in the proof of theorem 1 just given, we proved that $T \vdash \varphi$ in just two lines and that this simple proof can be formalized in T. The result is a short T-proof of $\Box_{\mathrm{T}}\varphi$. Now, a Σ_1-sound T proves $\Box_{\mathrm{T}}\varphi$ iff it proves φ, so a proof of $\Box_{\mathrm{T}}\varphi$ could be considered just as good as a proof of φ itself. Formally this is represented by adding an additional rule to T, the Parikh rule: $\Box_{\mathrm{T}}\varphi/\varphi$. This rule adds no new theorems to T, but by Parikh's result it drastically shortens (certain) proofs. For more on the Parikh rule (and other related rules) cf. [8].

First we should note that our sketch in the paragraph above is not completely correct: If $T \vdash \Box_{\mathrm{T}}\varphi$, a proof of φ could be found by simply searching through all proofs in T; it follows that the length of the proof of φ is bounded by a recursive function of the length of the proof of $\Box_{\mathrm{T}}\varphi$, so Parikh's result cannot be valid for all recursive functions. The reason is easily seen: theorem 1 assumes the function f to be total, to formalize the proof in T we need to prove this as well in T, so f must be *provably recursive* in T. We can now state the proper version of Parikh's theorem:

Theorem 2 (Parikh [14]). *Let g be provably recursive in T, there are k and φ such that $T \vdash_k \Box_{\mathrm{T}}\varphi$, but $T \not\vdash_{g(k)} \varphi$.*

Parikh actually proved more than this, and much more could be proved by his methods (cf. [4, 5]). Since we will require some of this additional information later on, we postpone a detailed proof of Parikh's theorem until section 3.

2. Shavrukov's theorem

The main theorem of this paper makes heavy use of the finer details of Parikh speed-up, hence we begin by introducing some terminology:

Definition 2. Let $2^x_0 = x$ and $2^x_{n+1} = 2^{2^x_n}$. Call a function f *multi-exponentially bounded* if there is a k such that for all x, $f(x) < 2^x_k$.

Remark. It is argued in [9, III.3.13] that natural transformations involving syntax are multi-exponentially bounded, hence if $\bar{\varphi}$ is the gödelnumber of some sentence φ and a reasonably short proof (compared to $\bar{\varphi}$) of φ is given, the size of the gödelnumber of this proof is multi-exponentially bounded in $\bar{\varphi}$.

Definition 3. For a fixed set $V \subseteq \omega$ and functions f and g, we write $f \preceq_V g$ if there is a function h of negligibly small growth s.t. $f(n) \leq h \circ g(n)$ for any $n \in V$. If also $g \preceq_V f$ we write $f \approx_V g$, otherwise $f \prec_V g$. When $V = \omega$ we drop the subscripts and write $f \preceq g$, $f \approx g$ etc.

Remark. The 'negligibly small' in the definition needs an explanation; we will deal with theories T and certain functions g exceeding any T-provably recursive function f on some infinite set V, in this situation we want to say that g *grows much faster than* f *on* V ($f \prec_V g$). With this in mind we take the (Kalmár) elementary functions as our class of functions of *negligibly small growth*; these functions are provably recursive in $\mathrm{I}\Delta_0$+exp and hence in all theories under consideration. By another theorem of Parikh [9, V.1.4] any such function is multi-exponentially bounded, so by the remark above syntactical transformations are bounded by elementary functions. This is implicit in many of the arguments below.

Definition 4. If $g(n) \leq f(n)$ holds for only finitely many n, we say that g *dominates* f and write $f \ll g$.

Remark. Observe that $f \ll g$ implies $f \preceq g$ since only a finite number of n can satisfy $g(n) \leq f(n)$ and *any* function that reduces to the identity function except for a finite number of exceptions is elementary.

Definition 5. Given n, there is an m s.t. $T \vdash_m \psi$ whenever $T \vdash_n \Box_{\mathrm{T}}\psi$ (for any ψ); let $\mathscr{P}_{\mathrm{T}}$ be the (obviously recursive) function that takes n to the smallest such m and call it the *Parikh speed-up over* T. Theorem 2 implies that, for any T-provably recursive function f, $\mathscr{P}_{\mathrm{T}}$ must exceed f for infinitely many values. Further define two related functions as follows:

$$\mathscr{R}_{\mathrm{T}}(n) = \mu x[\forall\, \delta \in \Delta_0(T \vdash_n \exists x\, \delta(x) \Rightarrow \exists y \leq x(\delta(y)))],$$
$$\mathscr{D}_{\mathrm{T}}(n) = \mu x[T \vdash_n \Box_{\mathrm{T}}\varphi \vee \Box_{\mathrm{T}}\psi \Rightarrow T \vdash_x \varphi \text{ or } T \vdash_x \psi].$$

Lemma 3.

(a) $\forall x\, (\mathscr{P}_{\mathrm{T}}(x) \leq \mathscr{R}_{\mathrm{T}}(x))$.

(b) there is a constant c such that $\forall x\, (\mathscr{D}_{\mathrm{T}}(x) \leq \mathscr{R}_{\mathrm{T}}(cx))$.

Proof. (a) is immediate since $\Box_{\mathrm{T}}\psi \in \Sigma_1$ and the condition for a jump of $\mathscr{P}_{\mathrm{T}}$ is a special case of that for $\mathscr{R}_{\mathrm{T}}$.

(b) is almost as easy, we just have to pass from $\exists x\ \mathrm{Prf}(x, \varphi) \vee \exists y\ \mathrm{Prf}(y, \psi)$ to $\exists x\, (\mathrm{Prf}(x, \varphi) \vee \mathrm{Prf}(x, \psi)$; clearly this involves at most a linear increase in the number of symbols. ⊣

The function $\mathscr{R}_{\mathrm{T}}$ grows very fast, in fact:

Lemma 4. $\mathscr{R}_{\mathrm{T}}$ *dominates every T-provably recursive function f.*

Proof. By assumption $T \vdash_k \forall x\,(f(x){\downarrow})$ for some k, so for n comfortably larger than k we will then $T \vdash_n f(\bar{n}){\downarrow}$ and by the definition of $\mathscr{R}_\mathrm{T}$ it follows that $f(n) \leq \mathscr{R}_\mathrm{T}(n)$. ⊣

The essential step in the proof of Parikh's theorem is:

Proposition 5. *There is a constant d such that $\forall x\,(\mathscr{R}_\mathrm{T}(x) \leq \mathscr{P}_\mathrm{T}(dx))$.*

The proof is given in section 3.

Corollary 6. $f \ll \mathscr{P}_\mathrm{T}$.

Proof. Let $n = dm + r$ be an arbitrary sufficiently large number and apply lemma 4 to the function $f(dx + r)$, giving $f(dx + r) \leq \mathscr{R}_\mathrm{T}(x) \leq \mathscr{P}_\mathrm{T}(dx + r)$ by the proposition and the monotonicity of $\mathscr{P}_\mathrm{T}$. Putting $x = m$ gives that $f(n) \leq \mathscr{P}_\mathrm{T}(n)$. ⊣

Corollary 7. *For some constant e there holds $\forall x\,(\mathscr{D}_\mathrm{T}(x) \leq \mathscr{P}_\mathrm{T}(ex))$.*

Proof. $\mathscr{D}_\mathrm{T}(x) \leq \mathscr{R}_\mathrm{T}(cx) \leq \mathscr{P}_\mathrm{T}(dcx)$ by lemma 3 and 5. ⊣

Remark. It is clear that $\mathscr{P}_\mathrm{T}$ (so consequently by the above, $\mathscr{R}_\mathrm{T}$ and $\mathscr{D}_\mathrm{T}$ as well) depends strongly on the theory T: for example $\mathscr{P}_\mathrm{ZF}$ must grow much faster than $\mathscr{P}_\mathrm{PA}$. This is because ZF proves uniform Σ_1-reflection for PA (i.e. $ZF \vdash \forall \sigma \in \Sigma_1\,(\Box_\mathrm{PA}\sigma \rightarrow \sigma)$), and so in particular it proves $\forall \varphi\,(\Box_\mathrm{PA}\Box_\mathrm{PA}\varphi \rightarrow \Box_\mathrm{PA}\varphi)$ and one sees that $\mathscr{P}_\mathrm{PA}$ is ZF-provably recursive. It is this difference in growth that lies at the heart of Shavrukov's proof of his theorem.

By the remark just given and corollary 6 we have:

Corollary 8. $\mathscr{P}_\mathrm{PA} \ll \mathscr{P}_\mathrm{ZF}$.

$\mathscr{R}_\mathrm{T}$ is not provably recursive in T but is, in a sense, rather close to being so. Specifically we have the following result:

Proposition 9. *For all k, $\mathrm{I}\Delta_0{+}\mathrm{exp} \vdash \forall x\, \Box^k_\mathrm{T}(\mathscr{R}^k_\mathrm{T}(x){\downarrow})$.*

Proof. First assume $k = 1$ and set

$$\varphi(x) = \exists p < 2^x_j\,[\mathrm{Prf}_\mathrm{T}(p, \ulcorner \exists z\, \forall y \leq x\,(\mathrm{Prf}_\mathrm{T}(y, \ulcorner \exists v\, \delta(v) \urcorner) \rightarrow \exists w \leq z\, \delta(w)) \urcorner)],$$

the constant j is to be chosen large enough to make φ true. It will be sufficient to prove $\forall x\,\varphi(x)$, this we do by induction (since the provability predicate Prf_T is assumed elementary this induction can be performed inside $\mathrm{I}\Delta_0{+}\mathrm{exp}$). Reason in T:

> Assume $\forall y < x\,(\varphi(y))$, we can further assume
>
> $$\mathrm{Prf}_\mathrm{T}(x, \ulcorner \exists t\,(\delta(t)) \urcorner),$$
>
> for some $\delta \in \Delta_0$ (otherwise $\varphi(x)$ follows trivially) and that δ is chosen maximal so that $\mathscr{R}_\mathrm{T}(x){\downarrow} \trianglelefteq \delta$. By the induction assumption there is $p' < 2^{(x-1)}_j$ such that:
>
> $$\mathrm{Prf}_\mathrm{T}(p', \exists z'\, \forall y < x\,(\mathrm{Prf}_\mathrm{T}(y, \ulcorner \exists v\, \delta(v) \urcorner) \rightarrow \exists w \leq z'\, \delta(w))).$$
>
> It is then clear that we can take the desired $z = \max(t, z')$ and obtain a proof of this fact with a multi-exponential bound as stated.

The general case follows easily from the first, we reason in T:

> By the above we have
>
> (1) $\forall x\,(\Box_{\mathrm{T}}\mathscr{R}_{\mathrm{T}}(x){\downarrow})$
>
> and
>
> (2) $\forall x\,(\mathscr{R}_{\mathrm{T}}(x){\downarrow}\!\rightarrow \Box_{\mathrm{T}}(\mathscr{R}_{\mathrm{T}}(\mathscr{R}_{\mathrm{T}}(x)){\downarrow}))$
>
> so, from (1) and (2)
>
> (3) $\forall x\,(\Box_{\mathrm{T}}\Box_{\mathrm{T}}\mathscr{R}_{\mathrm{T}}(\mathscr{R}_{\mathrm{T}}(x){\downarrow}))$

iterating these steps we prove $\forall x\,(\Box^k_{\mathrm{T}}\mathscr{R}^k_{\mathrm{T}}(x){\downarrow})$ for all k. $\dashv$

We have already remarked that a jump of $\mathscr{P}_{\mathrm{T}}$ entails a jump of $\mathscr{R}_{\mathrm{T}}$ as well, so we have:

Corollary 10. *For all k,* $\mathrm{I}\Delta_0{+}\mathrm{exp} \vdash \forall x\,\Box^k_{\mathrm{T}}(\mathscr{P}^k_{\mathrm{T}}(x){\downarrow})$.

We now come to our main theorem which, assuming the existence of an isomorphism from $\mathfrak{D}_{\mathrm{T}}$ to $\mathfrak{D}_{\mathrm{S}}$, gives a bound on $\mathscr{P}_{\mathrm{T}}$ in terms of $\mathscr{P}_{\mathrm{S}}$. The following diagram might prove useful for the reader to keep in mind during the course of the proof:

$$\begin{tikzcd}
\Delta^T_A \arrow[r, "\mathscr{P}_{\mathrm{T}}"] \arrow[d, "e"'] & \Delta^T_B \\
\Delta^S_\alpha \arrow[r, "\mathscr{P}_{\mathrm{S}}"] & \Delta^S_\beta \arrow[u, "e^{-1}"']
\end{tikzcd}$$

$\mathscr{P}_{\mathrm{T}}$ is calculated via a detour (possible because $e\colon \mathfrak{D}_{\mathrm{T}} \to \mathfrak{D}_{\mathrm{S}}$ is an isomorphism) through the theory S, the main difficulty of the proof lies in choosing the sentences α and B so that no dramatic increases in proof lengths can occur while moving vertically in the diagram.

We assume that a standard gödelnumbering of Turing machines is given and write φ_k for the function computed by the kth Turing machine (k will be referred to as a *φ-index for φ_k*). $\varphi_k(n)$ will carry the double meaning of, on the one hand the *value* of φ_k on n, and on the other the computation executed by φ_k on n.

We will employ a notion of the *length* of computations that differs somewhat from the usual ones:

$$\Phi_k(n) = k + n + \textit{the number of steps in the computation of } \phi_k(n).$$

In the proof we will make use of another, non-standard representation of 0-1-valued recursive functions, that we now set up. First define the sequence of sentences $\{\ast^n_{\mathrm{T}}\}_{n\in\omega}$ (we set $\Diamond = \neg\Box\neg$):

$$\ast^n_{\mathrm{T}} = \Box^{n+1}\bot \wedge \Diamond^n\top$$

and associate to every sentence λ the (recursive) function $\delta^{\mathrm{T}}_{\lambda}$:

$$\delta^{\mathrm{T}}_{\lambda}(n) = \begin{cases} 0 & \text{if } T \vdash \circledast^n_{\mathrm{T}} \to \lambda, \\ 1 & \text{if } T \vdash \circledast^n_{\mathrm{T}} \to \neg\lambda, \\ \text{undefined} & \text{if } T + \circledast^n_{\mathrm{T}} \text{ does not decide } \lambda. \end{cases}$$

We will need a notion of length of computations for this representation as well, we define:

$$\Delta^{\mathrm{T}}_{\lambda}(n) = \mu x(T \vdash_x \circledast^n_{\mathrm{T}} \to \lambda \text{ or } T \vdash_x \circledast^n_{\mathrm{T}} \to \neg\lambda).$$

The two different notions of length will be connected by lemma 14 below.

Definition 6. A recursive function f is called *cumulative* if there is a φ-index $\bar{f}$ for f s.t. $\Phi_{\bar{f}} \preceq f$.

Cumulativity is inteded to ensure that the growth rate of f is faithfully mirrored by the time needed to compute it. Note that, for trivial reasons, we always have $f \preceq \Phi_{\bar{f}}$ for any index $\bar{f}$ (in the computation of $\varphi_{\bar{f}}(n)$ it takes $f(n)$ steps just to write the output to the tape).

We will need a couple of easy lemmas on cumulative functions:

Lemma 11. *The functions $\Delta^{\mathrm{T}}_{\lambda}$ and $\mathscr{P}_{\mathrm{T}}$ are cumulative.*

Proof. This is straightforward: $\Delta^{\mathrm{T}}_{\lambda}(n)$ is computed by a Turing machine that searches through all T-proofs in order, looking for a (dis)proof of λ from $\circledast^n_{\mathrm{T}}$. Let d φ-index such a machine, since the number of proofs of gödelnumber $\leq n$ is bounded by an elementary function in n it is then clear that $\Phi_d \preceq \Delta^{\mathrm{T}}_{\lambda}$. $\mathscr{P}_{\mathrm{T}}$ is cumulative for the same reasons. ⊣

Lemma 12. *Every elementary function is dominated by a monotone elementary cumulative function.*

Proof. It has already been pointed out that every elementary function is multi-exponentially bounded and hence dominated by one of the functions $\{\lambda x.2^x_n\}$. These functions are clearly monotone and easily seen to be cumulative. ⊣

Lemma 13. *Let f and g be cumulative, then the composition $f \circ g$ is cumulative as well.*

Proof (not really). This is hardly surprising but tedious; by the cumulativity of f and g there are φ-indices k and i for computing them s.t. $\Phi_k \preceq f$ and $\Phi_i \preceq g$, $f \circ g$ is then computed by the Turing machine that results from simply stringing φ_k and φ_i together. Let j φ-index the resulting machine, to show cumulativity one has but to verify that there is an elementary bound on Φ_j in terms of $f \circ g$; this is the tedious part that we omit. (Details can be found in [17].) ⊣

Lemma 14. *To every φ-index k for a 0-1-valued (partial) recursive function there corresponds a sentence λ with the properties*

$$\varphi_k \equiv \delta^{\mathrm{T}}_{\lambda} \quad \text{and} \quad \Delta^{\mathrm{T}}_{\lambda} \preceq \Phi_k.$$

Conversely, to every λ there corresponds a φ-index k with the properties

$$\delta^{\mathrm{T}}_{\lambda} \equiv \varphi_k \quad \text{and} \quad \Phi_k \preceq \Delta^{\mathrm{T}}_{\lambda}.$$

A λ corresponding to φ_k as in the lemma will be called a δ^{T}-index for φ_k. The proof is given in section 3.

Lemma 15. *If f is a cumulative function and dom $f = V$ there is a recursive 0-1-valued function g with dom $g = V$ and an index $\bar{g}$ for g such that whenever $\varphi_i \equiv_{V \cap dom\, \varphi_i} g$, there holds:*

$$\Phi_{\bar{g}} \approx_V f \preceq_{V \cap dom\, \varphi_i} \Phi_i.$$

Proof (sketch). The following fact about the complexity measure Φ will be of crucial importance in the proof: there can be only a finite number of pairs (k, m) for which there is a chance of $\Phi_k(m) = n$. Also, since f is cumulative there is an index $\bar{f}$ for f such that $f \approx \Phi_{\bar{f}}$ holds and we can forget about f itself and work with $\Phi_{\bar{f}}$.

Remembering this, one sees that the required function g can be obtained as follows: g successively computes the values $\Phi_{\bar{f}}(n)$, looking for an index j such that $\Phi_j(n) \leq \Phi_{\bar{f}}(n)$ and

$$\forall x\, ((\Phi_j(x) \leq \Phi_{\bar{f}}(x) < \Phi_{\bar{f}}(n)) \rightarrow \varphi_j(x) = g(x)).$$

When encountering such a j, g immediately decides that

$$g(n) = \begin{cases} 0 & \text{if } \varphi_j(n) > 0, \\ 1 & \text{if } \varphi_j(n) = 0. \end{cases}$$

ensuring that $g \neq \varphi_j$. In this way g cannot equal any φ_k with the property that $\Phi_k(n) < \Phi_{\bar{f}}(n)$ holds for infinitely many n. This settles one part of the lemma, since if $\varphi_i \equiv_{V \cap dom\, \varphi_i} g$, then $\Phi_i(n) < \Phi_{\bar{f}}(n)$ can hold for only finitely many n and trivially $f \preceq_{V \cap dom\, \varphi_i} \Phi_i$ (since $\Phi_{\bar{f}} \approx_V f$).

Finally, let $\bar{g}$ be the φ-index of the Turing machine that computes g by the method just outlined; it does so by first computing $\Phi_{\bar{f}}$ and then carrying out some simple bookkeeping, thus one sees that $\Phi_{\bar{g}} \preceq_V f$ as claimed. ⊣

Lemma 15 is a special case of Blum's *compression theorem* (cf. [3] for the full story), a complete proof of our case is in [17].

Lemma 16. *There is a partial recursive 0-1-valued function h with dom $h = V$ and an index $\bar{h}$ for h s.t. $\Phi_{\bar{h}} \preceq_V \Phi_k$, whenever $\varphi_k \equiv_V h$.*

Proof. By lemma 15 it is enough to show that there is a cumulative function with domain V, by lemma 11 and 14 there is. ⊣

We are now ready to begin the proof of our main theorem. To this end, fix a pair of theories S and T and assume that $e\colon \mathfrak{D}_{\mathrm{T}} \rightarrow \mathfrak{D}_{\mathrm{S}}$ is an isomorphism between their diagonalizable algebras. Next, fix a nonrecursive r.e. set X (this set will remain fixed throughout the whole proof). Now pick a function h with dom $h = X$ and a φ-index $\bar{h}$ as in lemma 16 and a δ^{S}-index α corresponding to $\bar{h}$ by lemma 14. Then let $A = e^{-1}(\alpha)$. Since e is an isomorphism it must send $\ast^n_{\mathrm{T}}$ to $\ast^n_{\mathrm{S}}$ and we have

$$\delta^{\mathrm{S}}_{\alpha} \equiv_X \delta^{\mathrm{T}}_A \equiv_X h$$

so, by lemma 14 and 16 and some φ-index k for h

$$\Delta^{\mathrm{S}}_{\alpha} \preceq_X \Phi_{\bar{h}} \preceq_X \Phi_k \preceq_X \Delta^{\mathrm{T}}_{A}$$

consequently

$$\Delta^{\mathrm{S}}_{\alpha} \leq_X p \circ \Delta^{\mathrm{T}}_{A},$$

for some elementary function p which we, by lemma 12, can assume cumulative. Since p was chosen cumulative, the function $\mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A}$ is, by lemma 13, cumulative as well. Hence lemma 15 provides a φ-index $\bar{f}$ for a partial recursive 0-1-valued function f such that

$$\Phi_{\bar{f}} \approx_X \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A} \preceq_X \Phi_i,$$

whenever φ_i is an extension of f.

We now need a δ^T-index for an extension of f, to ensure that the formalization below goes through we do not, however, use lemma 14 but rather a specially crafted sentence B.[3]

Let s be another elementary function such that $\Phi_{\bar{f}} \leq_X s \circ \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A}$, let $B(x)$ be the formula

$$\Phi_{\bar{f}}(x){\downarrow} \trianglelefteq s \circ \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A}(x){\downarrow} \to f(x) = 0$$

and finally let B be

$$\forall x\, (\ast^{x}_{\mathrm{T}} \to B(x)).$$

We now show that B gives the desired δ^T-index:

$$\delta^{\mathrm{T}}_{B} \approx_X f.$$

To see this, assume $n \in X$, then $T \vdash \ast^{n}_{\mathrm{T}} \to A$ or $T \vdash \ast^{n}_{\mathrm{T}} \to \neg A$ and, by the choice of f, $\Phi_{\bar{f}}(n) \leq s \circ \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A}(n)$ as well. Since these simple facts are also provable in T, $B(n)$ reduces to

$$f(n) = 0.$$

From this one derives

$$\begin{aligned} T \vdash \ast^{n}_{\mathrm{T}} &\to \forall x\, (\ast^{x}_{\mathrm{T}} \leftrightarrow x = n) \\ &\to \forall x\, (\ast^{x}_{\mathrm{T}} \to B(x)) \leftrightarrow (\ast^{n}_{\mathrm{T}} \to B(n)) \\ &\to B \leftrightarrow B(n) \\ &\to B \leftrightarrow f(n) = 0 \end{aligned}$$

and so, $\delta^{\mathrm{T}}_{B} \approx_X f$. Moreover, a slightly weakened form of this argument can be formalized in T to give

$$\begin{aligned} T \vdash \forall x\, (\delta^{\mathrm{T}}_{A}(x){\downarrow} &\to \Box_{\mathrm{T}}(\ast^{x}_{\mathrm{T}} \to A) \vee \Box_{\mathrm{T}}(\ast^{x}_{\mathrm{T}} \to \neg A) \\ &\to \Box_{\mathrm{T}}(s \circ \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A}(x){\downarrow})), \quad \text{by corollary 10} \\ &\to \Box_{\mathrm{T}}(\Box_{\mathrm{T}} B(x) \vee \Box_{\mathrm{T}} \neg B(x))) \\ &\to \Box_{\mathrm{T}}(\Box_{\mathrm{T}}(\ast^{x}_{\mathrm{T}} \to B) \vee \Box_{\mathrm{T}}(\ast^{x}_{\mathrm{T}} \to \neg B))) \end{aligned}$$

[3] A note of caution for readers of [17]: the sentence B was there defined in terms of $\mathscr{P}_{\mathrm{S}}$ rather than $\mathscr{P}_{\mathrm{T}}$. This discrepancy is behind the occurance of the extra box in our formulas below (as well as the need for a result like corollary 10).

in particular, for each n

$$T \vdash \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to A) \vee \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \neg A) \to \Box_{\mathrm{T}}(\Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to B) \vee \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \neg B)).$$

Now set $\beta = e(B)$, like before e preserves the relevant structure and we have

$$\delta^{\mathrm{S}}_{\beta} \approx \delta^{\mathrm{T}}_{B} \approx_X f$$

and so, by lemma 14 and the choice of f

$$\Phi_{\bar{f}} \preceq_X \Delta^{\mathrm{S}}_{\beta}.$$

By the same isomorphism we must also have

$$S \vdash \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \alpha) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\alpha) \to \Box_{\mathrm{S}}(\Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \beta) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\beta)),$$

hence, since S is assumed axiomatizable, there is a recursive total function j such that

$$S \vdash_{j(n)} \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \alpha) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\alpha) \to \Box_{\mathrm{S}}(\Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \beta) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\beta)).$$

Since j is total, the set

$$Y = \{n \in X \colon j(n) \leq \Delta^{\mathrm{S}}_{\alpha}(n)\}$$

is infinite, for otherwise the domain of $\Delta^{\mathrm{S}}_{\alpha}$, that is X, would be recursive. We now focus our attention on Y. Clearly there is a function $l \preceq_X \Delta^{\mathrm{S}}_{\alpha}$ such that

$$S \vdash_{l(n)} \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \alpha) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\alpha),$$

since constructing a proof of $\Box\varphi$ from that of φ is quite an easy task. It follows that for all $n \in X$ and some partial recursive $\mathscr{C}$ such that

$$\mathscr{C} \preceq_X \max(j, l) \preceq_Y \Delta^{\mathrm{S}}_{\alpha}$$

there holds

$$S \vdash_{\mathscr{C}(n)} \Box_{\mathrm{S}}(\Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \beta) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\beta)),$$

and so

$$S \vdash_{\mathscr{P}_{\mathrm{S}} \circ \mathscr{C}(n)} \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \beta) \vee \Box_{\mathrm{S}}(\ast^n_{\mathrm{S}} \to \neg\beta).$$

By the definition of the function $\mathscr{D}_{\mathrm{S}}$ we have:

$$S \vdash_{\mathscr{D}_{\mathrm{S}} \circ \mathscr{P}_{\mathrm{S}} \circ \mathscr{C}(n)} \ast^n_{\mathrm{S}} \to \beta \text{ or } S \vdash_{\mathscr{D}_{\mathrm{S}} \circ \mathscr{P}_{\mathrm{S}} \circ \mathscr{C}(n)} \ast^n_{\mathrm{S}} \to \neg\beta.$$

Now fix an elementary function q

$$\mathscr{C} \leq_Y q \circ \Delta^{\mathrm{S}}_{\alpha},$$

by corollary 7, $\mathscr{D}_{\mathrm{T}} \preceq \mathscr{P}^2_{\mathrm{S}}$ ($\mathscr{P}_{\mathrm{S}}$ will certainly dominate the linear function $x \mapsto ex$), so there there holds:

$$\Delta^{\mathrm{S}}_{\beta} \preceq_Y \mathscr{P}^2_{\mathrm{S}} \circ \mathscr{P}_{\mathrm{S}} \circ \mathscr{C} \preceq_Y \mathscr{P}^3_{\mathrm{S}} \circ q \circ \Delta^{\mathrm{S}}_{\alpha} \; \preceq_Y \mathscr{P}^4_{\mathrm{S}} \circ \Delta^{\mathrm{S}}_{\alpha},$$

by the fact that $\mathscr{P}_{\mathrm{S}}$ is monotone and $\mathscr{P}_{\mathrm{S}} \succ a$ for any elementary a. But also

$$\mathscr{P}_{\mathrm{T}} \circ \Delta^{\mathrm{S}}_{\alpha} \preceq_X \mathscr{P}_{\mathrm{T}} \circ p \circ \Delta^{\mathrm{T}}_{A} \preceq_X \Phi_{\bar{f}} \preceq_X \Delta^{\mathrm{S}}_{\beta},$$

by lemma 14 and the choice of $\bar{f}$. Putting everything together we get

$$\mathscr{P}_{\mathrm{T}} \circ \Delta^{\mathrm{S}}_{\alpha} \preceq_Y \mathscr{P}^{4}_{\mathrm{S}} \circ \Delta^{\mathrm{S}}_{\alpha}.$$

Now look at the set $Z = \{\Delta^{\mathrm{S}}_{\alpha}(y) \colon y \in Y\}$. Z must be infinite as well (by obvious properties of $\Delta^{\mathrm{S}}_{\alpha}$). So finally

$$\mathscr{P}_{\mathrm{T}} \preceq_Z \mathscr{P}^{4}_{\mathrm{S}}.$$

We sum up our results in the following

Theorem 17. *Assume that $\mathfrak{D}_{\mathrm{T}}$ is isomorphic to $\mathfrak{D}_{\mathrm{S}}$, there is then an elementary function t such that $\mathscr{P}_{\mathrm{T}}(n) \leq t \circ \mathscr{P}^{4}_{\mathrm{S}}(n)$ for infinitely many n.*

It has already been pointed out that $\mathscr{P}_{\mathrm{PA}} \ll \mathscr{P}_{\mathrm{ZF}}$ because $\mathscr{P}_{\mathrm{PA}}$ is ZF-provably recursive, but ZF must then prove $\mathscr{P}^{5}_{\mathrm{PA}}$ total as well and consequently $\mathscr{P}^{5}_{\mathrm{PA}} \ll \mathscr{P}_{\mathrm{ZF}}$. The following is then an immediate corollary to the theorem:

Corollary 18 (Shavrukov [17])**.** *$\mathfrak{D}_{\mathrm{PA}}$ is not isomorphic to $\mathfrak{D}_{\mathrm{ZF}}$.*

Proof. An isomorphism would, by the theorem, imply that there is an elementary function b such that $b \circ \mathscr{P}^{4}_{\mathrm{PA}}$ exceeds $\mathscr{P}_{\mathrm{ZF}}$ for infinitely many values. This contradicts $\mathscr{P}^{5}_{\mathrm{PA}} \ll \mathscr{P}_{\mathrm{ZF}}$. ⊣

Further results will follow once we have conducted a more careful analysis of the dependence of $\mathscr{P}_{\mathrm{T}}$ on T which we begin now.

3. Parikh speed-up

First of all we should fill the gap in our proof of Parikh's theorem:

Proof (of proposition 5). Fix x and pick δ so that

(4) $\quad T \vdash_x \exists z\,(\delta(z))$

and $\mathscr{R}_{\mathrm{T}}(x){\downarrow} \trianglelefteq \delta$. Now find θ such that $T \vdash \theta \leftrightarrow \neg\Box_{\mathrm{T},\delta}\,\theta$. Clearly, $T \vdash_{\mathscr{R}_{\mathrm{T}}(x)} \theta$ implies $T \vdash \neg\theta$, hence

(5) $\quad T \nvdash_{\mathscr{R}_{\mathrm{T}}(x)} \theta$.

On the other hand: $T \vdash \neg\Box_{\mathrm{T}}\theta \to \Box_{\mathrm{T}}\theta$ by (4) and so $T \vdash \Box_{\mathrm{T}}\theta$. A moment's thought shows that the length of the proof of $\Box_{\mathrm{T}}\theta$ is linear in x, say $T \vdash_{dx} \Box_{\mathrm{T}}\theta$, hence $T \vdash_{\mathscr{P}_{\mathrm{T}}(dx)} \theta$ and $\mathscr{R}_{\mathrm{T}}(x) < \mathscr{P}_{\mathrm{T}}(dx)$ follows by (5). ⊣

Proof (of lemma 14). The existence of the required φ-index corresponding to λ is nothing but a restatement of the fact that $\Delta^{\mathrm{T}}_{\lambda}$ is cumulative. Turning to the converse construction we choose $\lambda(n)$ provably equivalent to

$$\big[\Box_{\mathrm{T}}(\ast^{n}_{\mathrm{T}} \to \neg\lambda(n)) \vee \phi_k(n) = 0\big] \lhd \big[\Box_{\mathrm{T}}(\ast^{n}_{\mathrm{T}} \to \lambda(n)) \vee \phi_k(n) = 1\big].$$

First observe that T could not possibly refute $\ast^n_{\mathrm{T}}$, indeed

$$T \vdash \neg\ast^n_{\mathrm{T}} \leftrightarrow (\Diamond^n\top \to \Diamond^{n+1}\top)$$

and a proof of $\neg\ast^n_{\mathrm{T}}$ would contradict the second incompleteness theorem. Now assume

$$(6) \quad T \vdash \ast^n_{\mathrm{T}} \to \lambda(n) \quad \text{or} \quad T \vdash \ast^n_{\mathrm{T}} \to \neg\lambda(n)$$

say the first case holds and $\Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \lambda(n)) \trianglelefteq \phi_k(n) = 0$, it is easy to see that T then ends up refuting $\ast^n_{\mathrm{T}}$. The other case similarly leads to the conclusion that $\phi_k(n) = 1$ must hold, hence (6) implies:

$$\phi_k(n) = 0 \quad \text{resp.} \quad \phi_k(n) = 1.$$

Conversely, if $\phi_k(n) = 0$, then $\Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \neg\lambda(n))$ must be false and, consequently, $\lambda(n)$ be true. Similarly, $\phi_k(n) = 1$ implies $\neg\lambda(n)$ and the following holds:

$$T \vdash \ast^n_{\mathrm{T}} \to \lambda(n) \iff \phi_k(n) = 0$$
$$T \vdash \ast^n_{\mathrm{T}} \to \neg\lambda(n) \iff \phi_k(n) = 1$$

Now put $\lambda = \forall x\,(\ast^x_{\mathrm{T}} \to \lambda(x))$, just as in the earlier construction of the sentence B we have that $\ast^x_{\mathrm{T}} \wedge \ast^y_{\mathrm{T}} \to x = y$ implies $\ast^x_{\mathrm{T}} \to \lambda \leftrightarrow \lambda(x)$ and thus

$$T \vdash \ast^n_{\mathrm{T}} \to \lambda \iff \phi_k(n) = 0$$
$$T \vdash \ast^n_{\mathrm{T}} \to \neg\lambda \iff \phi_k(n) = 1$$

This settles everything claimed in the lemma except for the assertion that λ is decided with a proof whose length is elementary in $\Phi_k(n)$, but this is easily seen from a simple analysis of the proof outlined. ⊣

The precise growth of $\mathscr{P}_{\mathrm{T}}$ depends, however, not only on T but on the particular arithmetization of the provability predicate of T used in its definition. Specifically, $\Box^n_{\mathrm{T}}$ is a standard proof predicate of T (in the sense of satisfying the Bernays-Löb conditions) as long as $\Box_{\mathrm{T}}$ is one (by straightforward induction on n). Write $\mathscr{P}_{\Box^n_{\mathrm{T}}}$ for the (ambiguously denoted) $\mathscr{P}_{\mathrm{T}}$ based on $\Box^n_{\mathrm{T}}$ and we have:

Proposition 19. $\mathscr{P}^n_{\Box_{\mathrm{T}}} \ll \mathscr{P}_{\Box^{n+1}_{\mathrm{T}}}$.

Proof. Find a sentence $\xi(\bar{m})$ such that

$$T \vdash \xi(\bar{m}) \leftrightarrow \neg\Box_{\mathrm{T},\mathscr{P}^n_{\Box_{\mathrm{T}}}(\bar{m})}\,\xi(\bar{m})$$

Reason in T:

Corollary 10 gives

$$(7) \quad \Box^n\mathscr{P}^n_{\Box_{\mathrm{T}}}(\bar{m})\!\downarrow$$

Now observe that $\mathscr{P}^n_{\Box_{\mathrm{T}}}(\bar{m})\!\downarrow\to \Box\xi(\bar{m})$, applying necessiation and distribution n times gives:

$$\Box^n(\mathscr{P}^n_{\Box_{\mathrm{T}}}(\bar{m})\!\downarrow) \to \Box^{n+1}\xi(\bar{m})$$

Hence, by (7)

$$\Box^{n+1}\xi(\bar{m}).$$

So $T \vdash \Box^{n+1}\xi(\bar{m})$ and $\xi(\bar{m})$ is true. For a contradiction assume that, for infinitely many m, $\mathscr{P}_{\Box_\mathrm{T}^{n+1}}(m) \leq \mathscr{P}^n_{\Box_\mathrm{T}}(m)$. Since the length of the proof of $\Box^{n+1}\xi(\bar{m})$ just given depends only on the length of $\bar{m}$, $T \vdash_m \Box^{n+1}\xi(\bar{m})$ will hold once m is large enough, according to the definition of $\mathscr{P}_{\Box_\mathrm{T}^{n+1}}$ we then have $T \vdash_{\mathscr{P}_{\Box_\mathrm{T}^{n+1}}(m)} \xi(\bar{m})$, by assumption this implies

$$T \vdash_{\mathscr{P}^n_{\Box_\mathrm{T}}(m)} \xi(\bar{m}),$$

but this contradicts the truth of $\xi(\bar{m})$. $\dashv$

Corollary 20. $\mathscr{P}^n_{\Box_\mathrm{T}} \preceq \mathscr{P}_{\Box_\mathrm{T}^{n+1}}$.

This kind of situation (viz. the dependence of a metamathematical result on a particular (class of) arithmetical predicate(s) used for the representation of the concepts involved) was first analyzed in detail by Feferman in [6]. Results displaying this sort of dependence Feferman called *intensional* (Gödel's second incompleteness theorem being the main example) and those free of such dependencies (requiring only *numerical correctness* of the representations used) *extensional* (Gödel's first incompleteness theorem being the main example). The question of intensionality for diagonalizable Lindenbaum algebras (i.e. of the dependence of (the isomorphism type of) $\mathfrak{D}_\mathrm{T}$ on the particular proof predicate Pr_T used in its definition) has been raised here and there in the literature (a recent warning was given in [20, p. 78]) but has, to the author's knowledge, never been answered. We offer the following result, establishing the intensional nature of the construction:

Corollary 21. *Let $\mathfrak{L}_\mathrm{T}$ be the ordinary, boolean, Lindenbaum algebra of T, we can then construct two diagonalizable algebras $(\mathfrak{L}_\mathrm{T}, \Box_\mathrm{T})$ and $(\mathfrak{L}_\mathrm{T}, \Box^6_\mathrm{T})$. These two algebras are not isomorphic.*

Proof. Assume otherwise, by theorem 17

$$\mathscr{P}_{\Box^6_\mathrm{T}} \preceq_Z \mathscr{P}^4_{\Box_\mathrm{T}},$$

but by corollary 20

$$\mathscr{P}^5_{\Box_\mathrm{T}} \preceq \mathscr{P}_{\Box^6_\mathrm{T}},$$

implying

$$\mathscr{P}^5_{\Box_\mathrm{T}} \preceq_Z \mathscr{P}^4_{\Box_\mathrm{T}}.$$

But there would then be an elementary function exceeding $\mathscr{P}_{\Box_\mathrm{T}}$ for infinitely many values, contradicting theorem 6. $\dashv$

Obviously, the same reasoning can be extended to give an infinite number of pairwise non-isomorphic $\mathfrak{D}_\mathrm{T}$'s for any theory T and, somewhat less trivially, the sequence of proof predicates $\{\Box^n_\mathrm{T}\}_{n\in\omega}$ can, following Marongiu [12], be extended into the transfinite, associating with each recursive ordinal α a standard proof predicate $\Box^\alpha_\mathrm{T}$. The result is a sequence $\{\Box^\alpha_\mathrm{T}\}_{0<\alpha<\omega_1^{CK}}$ with the following properties:

(i) $\Box^1_\mathrm{T} = \Box_\mathrm{T}$

(ii) $\Box^{\alpha+1}_\mathrm{T} = \Box_\mathrm{T}\Box^\alpha_\mathrm{T}$.

The proof of proposition 19 extends to this sequence and hence $\mathscr{P}_{\Box^\alpha_{\mathrm{T}}} \prec \mathscr{P}_{\Box^{\alpha+1}_{\mathrm{T}}}$. Since the sequence runs through all the recursive ordinals we can in fact find, given an arbitrary total recursive function f, an ordinal (notation for) γ such that $f \preceq \mathscr{P}_{\Box^\gamma_{\mathrm{T}}}$ (proof of this is deferred). Applying this to the function $\mathscr{P}^7_{\mathrm{T}}$ we see that the following holds (all theories are still assumed Σ_1-sound):

Theorem 22. *For any theory T, any fixed provability predicate Pr_{T} of T and any theory S, there is a standard provability predicate $\mathrm{Pr}'_{\mathrm{S}}$ for S such that $\mathfrak{D}_{\mathrm{T}} \not\cong \mathfrak{D}'_{\mathrm{S}}$.*

4. The case of epimorphisms

We would like to extend theorem 17 to the less restrictive case where the homomorphism $e\colon \mathfrak{D}_{\mathrm{T}} \to \mathfrak{D}_{\mathrm{S}}$ is only assumed to be surjective (that is: an *epimorphism*). Our strategy is to simply carry the proof given over to the new setting, so we begin by finding the sentence α corresponding to the function h as before, since e is onto we have some A such that $e(A) = \alpha$, at this point however a problem emerges: since e is no longer assumed injective we cannot be sure that $\delta^{\mathrm{T}}_A \equiv_X \delta^{\mathrm{S}}_\alpha$ as before, only that $\delta^{\mathrm{T}}_A \subseteq \delta^{\mathrm{S}}_\alpha$.

This necessiates a slightly more elaborate construction of the sentence B than before. Set dom $\delta^{\mathrm{T}}_A = X'$, by lemma 12 and 16 we can find a monotone elementary function p such that $\Delta^{\mathrm{S}}_\alpha \leq_{X'} p \circ \Delta^{\mathrm{T}}_A$. Define a new function by:

$$\begin{aligned} q(n) = \mu x(S \vdash_x \ast^n_{\mathrm{S}} \to \alpha \vee S \vdash_x \ast^n_{\mathrm{S}} \to \neg\alpha \\ \vee \exists y\, (x = p(y) \wedge T \vdash_y \ast^n_{\mathrm{T}} \to A \vee T \vdash_y \ast^n_{\mathrm{T}} \to \neg A)) \end{aligned}$$

By construction $\Delta^{\mathrm{S}}_\alpha \leq_X q$ and dom $q = X$.

Now apply lemma 15 to the function $\mathscr{P}_{\mathrm{T}} \circ q$, as before fix a particular elementary function s such that $\Phi_{\bar f} \leq_X s \circ \mathscr{P}_{\mathrm{T}} \circ q$, then define

$$B(x) = \Phi_{\bar f}(x) \trianglelefteq s \circ \mathscr{P}_{\mathrm{T}} \circ q(x) \to f(x) = 0$$

and

$$B = \forall x\, (\ast^x_{\mathrm{T}} \to B(x)).$$

We can now proceed as before and show $\delta^{\mathrm{T}}_B \equiv_X f$ (since if $x \in X$ we have, provably in T, that $\Phi_{\bar f}(x) \leq s \circ \mathscr{P}_{\mathrm{T}} \circ q(x)$ and $B(x)$ reduces to $f(x) = 0$). By the construction of q we have $T \vdash \forall x\, (\delta^{\mathrm{T}}_A(x){\downarrow} \to q(x){\downarrow})$ and the formalization as well can be carried through:

$$T \vdash \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to A) \vee \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \neg A) \to \Box_{\mathrm{T}}(\Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to B) \vee \Box_{\mathrm{T}}(\ast^n_{\mathrm{T}} \to \neg B)).$$

By going through the exact same steps as in the original proof we then arrive at:

$$\Delta^{\mathrm{S}}_\beta \preceq_Y \mathscr{P}^6_{\mathrm{S}} \circ \Delta^{\mathrm{S}}_\alpha.$$

Since $\Delta^{\mathrm{S}}_\alpha \leq_X q$ it follows (by the monotonicity of $\mathscr{P}_{\mathrm{T}}$) that $\mathscr{P}_{\mathrm{T}} \circ \Delta^{\mathrm{S}}_\alpha \preceq_X \mathscr{P}_{\mathrm{T}} \circ q$, but also $\mathscr{P}_{\mathrm{T}} \circ q \preceq_X \Delta^{\mathrm{S}}_\beta$, by the construction of B. Hence

$$\mathscr{P}_{\mathrm{T}} \circ \Delta^{\mathrm{S}}_\alpha \preceq_X \Delta^{\mathrm{S}}_\beta,$$

and we get the desired

$$\mathscr{P}_{\mathrm{T}} \circ \Delta^{\mathrm{S}}_{\alpha} \preceq_Y \mathscr{P}^6_{\mathrm{S}} \circ \Delta^{\mathrm{S}}_{\alpha}.$$

This accomplishes the extension and we can replace the assumed isomorphism $e\colon \mathfrak{D}_{\mathrm{T}} \to \mathfrak{D}_{\mathrm{S}}$ in corollary 18 and theorem 22 by an epimorphism.

5. Related results and open problems

Beklemishev [2, §2, Lemma 7.b] established the intensional nature of diagonalizable algebras of Σ_1-*ill* theories, in fact, not even the equational theory is stable under change of provability predicate for such a theory. Specifically he showed how to construct a proof predicate of any desired (finite or infinite) credibility extent.

In our proof of theorem 17 sharper bounds were routinely sacrificed for simplification of derivations and the inequality stated is by no means optimal.Hence, sharper bounds are certainly possible and by persuing this line one might hope to replace the rather arbitrary $\Box^6$ of corollary 21 with the more natural $\Box\Box$.

The generalization of our results to epimorphisms prompts the question whether positive examples of epimorphisms between diagonalizable algebras can be found. Sadly, we have not made much progress on this question. Apart from the isomorphisms constructed in [18] we are only aware of one example:

Definition 7. A sentence φ is called a *self-prover (with respect to T)* if $T \vdash \varphi \to \Box_{\mathrm{T}}\varphi$.

Remark. Any Σ_1-sentence is a self-prover (over any theory T) by virtue of provable Σ_1-completeness. Furthermore, it is easy to see that $\psi \wedge \Box_{\mathrm{T}}\psi$ is a self-prover over T for any ψ and that any self-prover is (equivalent to one) of this form. It can be shown that the self-provers are by no means exhausted by the Σ_1-sentences, in fact Kent has shown that there are self-provers of arbitrarily high complexity ([10], cf. also [9, III.4.68]).

The following observation is mentioned in [1, p. 571]:

Theorem 23. *Let T be arbitrary and ψ a self-prover over T, there is then a canonical epimorphism $p\colon \mathfrak{D}_{\mathrm{T}} \to \mathfrak{D}_{\mathrm{T}+\psi}$.*

Proof. p is simply the projection that maps (the equivalence class of) a sentence φ to its equivalance class over $T + \psi$. This map is trivially seen to be a *boolean* homomorphism for any sentence ψ, to show that it is a diagonalizable homomorphism thus amounts to check that it also commutes with boxes. This reduces to checking that $T \vdash \psi \to (\Box_{\mathrm{T}}\varphi \leftrightarrow \Box_{\mathrm{T}}(\psi \to \varphi))$, left-to-right of the equivalance is a trivial matter, the converse follows from $\Box_{\mathrm{T}}\psi \to (\Box_{\mathrm{T}}(\psi \to \varphi) \to \Box_{\mathrm{T}}\varphi)$, since ψ is a self-prover. p is obviously surjective. ⊣

Unfortunately, theorem 23 comes far from providing an interesting converse to our result; $T + \psi$ is never Σ_1-sound unless ψ is provable and we cannot use Theorem 17 to conclude that no epimorphism exists in the other direction (in fact, we do not know if this is true).[4]

[4] Shavrukov points out to me that a *recursive* homomorphism (like the ones constructed in [16, 18, 21]) from the algebra of a Σ_1-ill theory T to that of a Σ_1-sound would furnish a recursive separation of the sets of provable, respectively refutable, Δ_1-sentences in T, contradicting [20, Lemma 3.4(a)].

Let RFN_{Σ_1} denote the sentence '$\forall\sigma \in \Sigma_1(\Box_{\mathrm{T}}\sigma \to \mathrm{True}_{\Sigma_1}(\sigma))$', then (the strengthened) Theorem 17 can be applied like in Corollary 18 (provided T is strong enough, say containing $\mathrm{I}\Sigma_1$) to show that $\mathfrak{D}_{\mathrm{T}}$ cannot be a homomorphic image of $\mathfrak{D}_{\mathrm{T}+\,\mathrm{RFN}_{\Sigma_1}}$. It seems to be of some interest to know whether there is a projection in the opposite direction:

Question. Is $\mathfrak{D}_{\mathrm{T}+\,\mathrm{RFN}_{\Sigma_1}}$ a homomorphic image of $\mathfrak{D}_{\mathrm{T}}$?

Unfortunately, we are not able to answer this question (nor offer a plausible conjecture). Constructing epimorphisms between algebras of Σ_1-sound theories seems very difficult, perhaps even as difficult as constructing isomorphisms.

The strengthening of Shavrukov's theorem at least has the following consequence: the collection of epimorphic images of a diagonalizable algebra $\mathfrak{D}_{\mathrm{T}}$ is *not* independent of T as was seen to be the case for the collection of subalgebras in the works of Shavrukov [16] and Zambella [21]; $\mathfrak{D}_{\mathrm{PA}}$ is trivially an epimorphic image of itself, but not of $\mathfrak{D}_{\mathrm{ZF}}$. This suggests the following

Problem. Give a characterization of the images of a diagonalizable Lindenbaum algebra under homomorphisms, similar to the Shavrukov-Zambella result on subalgebras.

There is a connection between this problem and more algebraic matters, Montagna has shown the following (essentially an algebraic version of the (uniform) Solovay completeness theorem):

Proposition 24 (Montagna [13]). *$\mathfrak{F}_\omega$, the free diagonalizable algebra on $\aleph_0$ generators, is a subalgebra of $\mathfrak{D}_{\mathrm{T}}$.*

Any homomorphic image of $\mathfrak{D}_{\mathrm{T}}$ is obviously countably generated and so has, for general algebraic reasons, a representation as a projection of $\mathfrak{F}_\omega$. Fix an embedding $i\colon \mathfrak{F}_\omega \to \mathfrak{D}_{\mathrm{T}}$ and a projection $p\colon \mathfrak{F}_\omega \to \mathfrak{D}$ onto some diagonalizable algebra $\mathfrak{D}$.

Problem. When can p be lifted to a homomorphism of $\mathfrak{D}_{\mathrm{T}}$? i.e. when is there a homomorphism ψ making the following diagram commute?

$$\begin{tikzcd}
\mathfrak{D}_{\mathrm{T}} \arrow[dr, "\psi"] & \\
\mathfrak{F}_\omega \arrow[u, "\mathrm{i}"] \arrow[r, "\mathrm{p}"] & \mathfrak{D}
\end{tikzcd}$$

Such a lifting would be a sufficient criterion for $\mathfrak{D}$ being a homomorphic image of $\mathfrak{D}_{\mathrm{T}}$, is it also a necessary one?

In [19] Shavrukov showed how to construct an interpretation of arithmetic into $\mathfrak{D}_{\mathrm{T}}$ and used this fact to strengthen corollary 18 to:

Theorem 25 ([19, Theorem 2.11]). *$\mathfrak{D}_{\mathrm{PA}}$ and $\mathfrak{D}_{\mathrm{ZF}}$ are not elementarily equivalent.*

It is immediate that this strengthening applies to our corollary 21 as well. (Of course, corollary 18 holds for many other pairs of theories besides PA and ZF, we have avoided going into the technical details of the exact scope of the result; this is discussed more fully in [17].)

The methods employed in this paper deal with the rate of growth of provably recursive functions or, which amounts to essentially the same, the (un)provability of Π_2-sentences. (It is easily seen that corollary 18 holds, for strong enough theories, as soon as T proves the Π_2-sentence '$\forall\varphi\,\Box_S\Box_S\varphi \to \Box_S\varphi$'.)

It is natural to ask about the effect on $\mathfrak{D}_T$ of instead adding unprovable Π_1-sentences to T; almost nothing seems to be known about this problem, but one naturally wonders about the following

Question. Are the diagonalizable algebras of T and $T + \mathrm{Con}_T$ isomorphic? Is $\mathfrak{D}_T$ a homomorphic image of $\mathfrak{D}_{T+\mathrm{Con}_T}$? Or vice versa?

Note that, in light of corollary 21, it cannot be ruled out that the answers depend (not only on T but) on the particular choice of the sentence Con_T.

References

[1] G. D'Agostino, *Topological Structure of Diagonalizable Algebras and Corresponding Logical Properties of Theories*, Notre Dame J. Formal Logic, **35(4)** (1994), pp. 563–572.

[2] L. D. Beklemishev, *On the classification of propositional provability logics* (russian), Izv. Akad. Nauk SSSR Ser. Mat., **53(5)** (1989), pp. 915–943; translation in Math. USSR Izvestiya **35(2)** (1990), pp. 247–275.

[3] M. Blum, *A machine-independent theory of the complexity of recursive functions*, Jour. Ass. Comp. Mach., **14(2)** (1967), pp. 322–336.

[4] D. de Jongh, F. Montagna, *Provable fixed points*, Zeit. f. math. Logik und Grundlagen d. Math., **34** (1988), pp. 229–250.

[5] D. de Jongh, F. Montagna, *Much shorter proofs*, Zeit. f. math. Logik und Grundlagen d. Math., **35** (1989), pp. 247–260.

[6] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fundamenta mathematicae, **49** (1960), pp. 35–92.

[7] K. Gödel, *Über die Länge von Beweisen*, Ergb. eines math. Kolloq., **7** (1936), pp. 23–24.

[8] P. Hájek, F. Montagna, P. Pudlák, *Abbreviating proofs using metamathematical rules*; in *Arithmetic, Proof theory and Computational complexity*, P. Clote, J. Krajíček eds., Oxford Univ. Press (1992), pp. 197–221.

[9] P. Hájek, P. Pudlák, *Metamathematics of first-order arithmetic*, Springer (1993).

[10] C. F. Kent, *The Relation of A to* $\mathrm{Prov}\ulcorner A\urcorner$ *in the Lindenbaum Sentence Algebra*, Journal of symbolic logic, **38** (1973), pp. 295–298.

[11] R. Magari, *The diagonalizable algebras (the algebraization of the theories which express Theor, II)*, Boll. Un. Mat. Ital., **4(12)** suppl. fasc. 3 (1975), pp. 117–125.

[12] G. Marongiu, *Sequenze dei predicati aritmetici di tipo Theor*, Boll. Un. Mat. Ital., **4(9)** (1974), pp. 361–375.

[13] F. Montagna, *On the diagonalizable algebra of Peano arithmetic*, Boll. Un. Mat. Ital., **5(16)** (1979), pp. 795–812.

[14] R. Parikh, *Existence and feasibility in arithmetic*, Journal of symbolic logic, **36** (1971), pp. 494–508.

[15] R. M. Solovay, *Provability interpretations of modal logic*, Israel journal of mathematics, **25(3)** (1976), pp. 287–304.

[16] V. Yu. Shavrukov, *Subalgebras of diagonalizable algebras of theories containing arithmetic*, Dissertationes mathematicae, **323** (1993).

[17] V. Yu. Shavrukov, *A note on the diagonalizable algebras of PA and ZF*, Annals of pure and applied logic, **61** (1993), pp. 161–173.

[18] V. Yu. Shavrukov, *Isomorphisms of diagonalizable algebras*, Theoria **63(3)** (1997), pp. 210–221.

[19] V. Yu. Shavrukov, *Undecidability in diagonalizable algebras*, Journal of symbolic logic, **62(1)** (1997), pp. 79–116.

[20] V. Yu. Shavrukov, *Effectively inseparable boolean algebras in lattices of sentences*, Archive for mathematical logic, **49** (2010), pp. 69–89.

[21] D. Zambella, *Shavrukov's Theorem on the Subalgebras of Diagonalizable Algebras for Theories Containing* IΔ_0+exp, Notre Dame J. Formal Logic **35(1)** (1994), pp. 147–157.