Some abstract versions of Gödel's second incompleteness theorem based on non-classical logics

Lev Beklemishev and Daniyar Shamkanov
Steklov Mathematical Institute, Moscow
Moscow M.V. Lomonosov State University
National Research University Higher School of Economics, Moscow

February 19, 2016

To Albert Visser, a remarkable logician and a dear friend, whose papers and conversations are a source of constant inspiration

Abstract

We study abstract versions of Gödel's second incompleteness theorem and formulate generalizations of Löb's derivability conditions that work for logics weaker than the classical one. We isolate the role of contraction rule in Gödel's theorem and give a (toy) example of a system based on modal logic without contraction invalidating Gödel's argument.

1 Introduction

One of the topics that have been fascinating logicians over the years is Gödel's second incompleteness theorem (G2). Both mathematically and philosophically G2 is well known to be more problematic than his first incompleteness theorem (G1). G1 and Rosser's Theorem are well understood in the context of recursion theory. Abstract logic-free formulations have been given by Kleene [12] ('symmetric form'), Smullyan [20] ('representation systems') and others. Sometimes G2 is considered as a minor addition to G1, whose role is to exhibit a specific form of the sentence independent from a given theory, namely its consistency assertion. However, starting with the work of Kreisel, Orey, Feferman, and others, who provided various nontrivial uses of G2, it has been gradually understood that the two results are of a rather different nature and scope. G2 has more to do with the (modal-logical) properties of the provability predicate and the phenomenon of self-reference in sufficiently expressive systems. A satisfactory general mathematical context for G2, however, still seems to be lacking.

The main difficulties in G2 are due to the fact that we cannot easily delineate a class of formulas that 'mean' consistency. Thus, the most intuitively appealing formulation of G2 — *sufficiently strong consistent theories cannot prove their own consistency* — remains non-mathematical. For a concrete formal system, such as Peano arithmetic PA, one can usually write out a specific 'natural' formula $\mathrm{Con}_{\mathsf{PA}}$ and declare it to be the expression of consistency. This approach is rather common in mathematics but has several deficiencies: Firstly, it ties the statement to a very particular formula, coding mechanism etc., and provides no clue why this choice is better than the other ones. Secondly, instead of a general theorem working uniformly for a wide class of theories, we only obtain a specific statement for an individual theory such as PA. We do not know what is the natural consistency assertion for an arbitrary extension of PA. Thus, we have a problem with translating our informal intuition into strict mathematical terms.

The way to better understand G2 is through investigating its range and generalizations. A lucky circumstance is that G2 also holds for larger syntactically defined classes of consistency formulas, some of which are apparently intensionally correct (adequately express consistency), but some are not. Thus, it is still possible to formulate mathematical results in certain important aspects *more* (rather than less) general than the broad intuitive formulation of G2 above.

A universally accepted approach to general formulations of G2 appeared in the fundamental paper by Feferman [3] who showed, among other things, that G2 holds for all consistency assertions defined by Σ_1-numerations. Feferman deals with first-order theories T in the language containing that of PA and specified by recursively enumerable (r.e.) sets of axioms. Feferman assumes fixed some natural Gödel numbering of the syntax of T as well as some specific axiomatization of first order logic. A Σ_1-formula $\alpha(x)$ defining the set of Gödel numbers of axioms of T in the standard model of PA is called a Σ_1-*numeration* of T.[1] It determines the provability formula $\mathsf{Prov}_\alpha(x)$ and the corresponding consistency assertion Con_α. Feferman's statement of G2 is that for all consistent theories T given by Σ_1-numerations α and containing a sufficiently strong fragment of PA, the formula Con_α is unprovable in T.

This theorem is considerably more general than any specific instance of G2 for an individual theory T. However, it also presupposes quite a lot: first order logic and its axiomatization, Gödel numbering, the way formula Prov_α is built from α.

Exploring bounds to G2 leads to relaxing various assumptions involved in Feferman's statement:

- One can weaken the axioms of arithmetic (for a representative selection see Bezboruah–Shepherdson [2], Pudlák [14], Wilkie–Paris [24], Adamowicz–Zdanowski [1], Willard [25, 26]).
- One can consider theories modulo interpretability. This approach started with the work of Feferman [3]. In the recent years it lead to particularly attractive coding-free formulations of generalizations of G2 due to Harvey Friedman and Albert Visser (see [21, 22, 23]).
- One can weaken the requirements on the proof predicate aka derivability conditions (see Feferman [3], Löb [13], Jeroslow [10, 11]).
- One can weaken the logic.

[1] Feferman deals with the notion of r.e. formula rather than with the equivalent notion of Σ_1-formula more common today.

It is the latter two aspects, less studied in the literature, that we are going to comment on in this note. Firstly, let us briefly recall the history of derivability conditions.

Gödel [5] gave a sketch of a proof of G2 and a promise to provide full details in a subsequent publication. This promise has not been fulfilled, and a detailed proof of this theorem — for a system Z related to first-order arithmetic PA — only appeared in a monograph by Hilbert and Bernays [9]. In order to structure a rather lengthy proof Hilbert and Bernays formulated certain conditions on the proof predicate in Z, sufficient for the proof of G2. Later Martin Löb [13] gave an elegant form to these conditions by stating them fully in terms of the provability predicate $\mathrm{Pr}(x)$ and obtained an important strengthening of G2 known as Löb's Theorem. Essentially the same properties of the provability predicate were earlier noted by Gödel in his note [6], where he proposed to treat the provability predicate as a connective $\Box$ in modal logic, though the idea that these conditions constitute necessary requirements on a provability predicate most likely only appeared later. For the sake of brevity we call the Gödel–Hilbert–Bernays–Löb conditions simply *Löb's conditions* below.

A traditional proof of G2 (for arithmetical theories) consists of a derivation of G2 from the fixed point lemma using Löb's conditions (see e.g. [17]). An accurate justification of these conditions is technically not so easy, and a rare textbook provides enough details here, however see Smoryński [19] and Rautenberg [15] for readable expositions.

Löb's conditions are applicable to formal theories at least containing the connective of implication and closed under the *modus ponens* rule. Here we give more general abstract formulations of G2 which presuppose very little about logic. They are rather close in the spirit and the level of generality to the recursion-theoretic formulations of G1 due to Smullyan. When a good implication is added to the language one essentially obtains the familiar Löb's conditions. However, we show that Gödel's argument presupposes admissibility of the contraction rule restricted to $\Box$-formulas in the logic under consideration. Moreover, the uniqueness of Gödelian fixed point is based on the similarly restricted form of weakening.

In the last part of the paper we present a system invalidating a formalized version of G2. We consider a version of propositional modal logic K4 based on the contraction-free fragment of classical logic extended by fixed point operators (defined for any formulas modalized in the fixed point variables). By means of a cut-elimination theorem for this system we establish the failure of G2 and some other properties such as the infinity of the Gödelian and Henkinian fixed points.

2 Abstract provability structures

Definition 2.1. Let us call *an abstract consequence relation* a structure $S = (L_S, \leqslant_S, \top, \bot)$, where L_S is a set of *sentences* of S, $\leqslant_S$ is a transitive reflexive relation on L_S, $\top$ and $\bot$ are distinguished elements of L_S ('axiom' and 'contradiction'). A sentence $x \in L_S$ is called *provable in* S, if $\top \leqslant_S x$, and *refutable in*

S, if $x \leqslant_S \bot$. Sentences x, y are called *equivalent in* S, if $x \leqslant_S y$ and $y \leqslant_S x$. The equivalence of x and y will be denoted $x =_S y$.

The structure S represents syntactical (rather than semantical) data about the theory in question. In a typical case, for example, for arithmetical theories S, the relation $x \leqslant_S y$ denotes the provability of y from hypothesis x, whereas $\top$ and $\bot$ are some standard provable and refutable formulas, respectively, e.g., $0 = 0$ and $0 \neq 0$.

In concrete situations we can enrich this structure by additional data, for example, by the conjunction and the implication connectives. Notice that we do not assume either $\bot \leqslant_S x$ or $x \leqslant_S \top$, nor do we assume the existence of any logical connectives (such as negation) in S.

S is called *inconsistent* if $\top \leqslant_S \bot$, otherwise it is called *consistent*. By transitivity, if S is consistent then no sentence is both provable and refutable. S is called *complete* if every $x \in L_S$ is either provable or refutable. S is called *r.e.*, if L_S is recursive and $\leqslant_S$ is r.e. (as a binary relation). T is called an *extension* of S if $L_T = L_S$ and $\leqslant_S$ is contained in $\leqslant_T$.

Let P_S and R_S denote the sets of provable and of refutable sentences of S, respectively. If S is consistent and r.e., then P_S and R_S is a pair of disjoint r.e. sets. We say that S *separates pairs of disjoint r.e. sets* if for each such pair (A, B) there is a total computable function f such that

$$\forall n \in A\; f(n) \in P_S \text{ and } \forall n \in B\; f(n) \in R_S.$$

The following statement is a natural version of G1 and Rosser's theorem for abstract consequence relations (á la Kleene and Smullyan); we omit the standard proof.

Proposition 2.2. (i) *If S is r.e., consistent and complete, then both P_S and R_S are decidable.*

(ii) *If S is r.e. and separates disjoint pairs of r.e. sets, then every consistent extension of S is incomplete and undecidable.*

Next we introduce two operators $\Box, \boxtimes : L_S \to L_S$ representing provability and refutability predicates in S.

Definition 2.3. *Provability* and *refutability operators* for an abstract consequence relation S are functions $\Box, \boxtimes : L_S \to L_S$ satisfying the following conditions, for all $x, y \in L_S$:

C1. $x \leqslant_S y \;\Rightarrow\; \Box x \leqslant_S \Box y$, $\boxtimes y \leqslant \boxtimes x$.

C2. $\top \leqslant_S \boxtimes\bot$;

C3. $x \leqslant_S \Box y,\; x \leqslant_S \boxtimes y \;\Rightarrow x \leqslant_S \boxtimes\top$;

C4. $\boxtimes x \leqslant_S \Box \boxtimes x$.

The algebra $(L_S, \leqslant_S, \top, \bot, \Box, \boxtimes)$ is called an *abstract provability structure* (APS).

Intuitively, $\Box x$ is the sentence expressing the provability of a sentence x, whereas $\boxtimes x$ expresses its refutability in S. Condition C1 means that provability of y follows from provability of x whenever y is derivable from x; similarly, refutability of y implies refutability of x. Conditions C2 and C3 are axioms for contradiction: according to C2, refutability of $\bot$ is provable in S; according to C3, $\top$ is refutable if some sentence y is both provable and refutable. Finally, Condition 4 means that the refutability of x can be formally checked in S. It is an analogue of Löb's condition L2 (see below).

Note that we consider the refutability operator on a par with the provability operator, since we do not assume that the logic of S necessarily has a well-defined operation of negation, that is, we cannot always define $\boxtimes x$ as $\Box\neg x$.

Remark 2.4. It is rather natural to additionally require that $\Box\bot =_S \boxtimes\top$: refutability of $\top$ and provability of $\bot$ are expressed by the same statement $\top \leqslant_S \bot$. Yet, it is not, strictly speaking, needed in this very abstract context, and we take $\boxtimes\top$ as our default expression of inconsistency.

Definition 2.5. We say that an abstract provability structure *S has a Gödelian fixed point* if there is a sentence $p \in L_S$ such that $p =_S \boxtimes p$.

Notice that Gödel considered a dual sentence q expressing its own unprovability in S. R. Jeroslow [11] noticed that the sentence stating its own refutability allows to prove G2 under somewhat more general conditions than those of Löb. In our formalism the sentence q is not expressible, therefore we are using Jeroslow's idea.

A very abstract version of G2 can now be stated as follows.

Theorem 1. *Suppose an APS S has a Gödelian fixed point.*

(i) *If S is consistent, then $\boxtimes\top$ is irrefutable in S.*

(ii) $\boxtimes\boxtimes\top \leqslant_S \boxtimes\top$*, that is, Statement* (i) *is formalizable in S.*

Proof. Let $p =_S \boxtimes p$. First we prove Statement (ii) omitting the subscript $_S$ everywhere:

1. $\boxtimes p \leqslant \Box\boxtimes p \leqslant \Box p$ by C4 and C1;

2. $p = \boxtimes p \leqslant \boxtimes\top$ by C3 (since $\boxtimes p \leqslant \boxtimes p$);

3. $\boxtimes\boxtimes\top \leqslant \boxtimes p = p \leqslant \boxtimes\top$ by C1.

Proof of Statement (i): Assume $\boxtimes\top \leqslant \bot$. By the previous argument $p \leqslant \boxtimes\top$, hence $p \leqslant \bot$. By C1, $\boxtimes\bot \leqslant \boxtimes p = p \leqslant \bot$. Therefore, by C2, $\top \leqslant \boxtimes\bot \leqslant \bot$. $\Box$

The following statement shows that under some additional condition the Gödelian–Jeroslowian fixed point is unique modulo equivalence in S and coincides with the inconsistency assertion for S. Therefore, the existence of such a fixed point is not only sufficient but also necessary for the validity of (a formalized version of) G2. The additional condition is

C5. $x \leqslant_S \top$, for all $x \in L_S$.

Theorem 2. *Assume C5 holds for S. Then $p =_S \boxtimes\top$ for all Gödelian fixed points p and (if such a sentence exists)*

$$\boxtimes\boxtimes\top =_S \boxtimes\top.$$

Proof. We know that $p \leqslant \boxtimes\top$. Since $p = \boxtimes p \leqslant \top$ we obtain $\boxtimes\top \leqslant \boxtimes p = p$. Hence $p = \boxtimes\top$ and therefore $\boxtimes\boxtimes\top = \boxtimes\top$. □

3 Consequence relations with implication

Classical Löb's conditions emerge for APS with an implication. A decent implication can be defined for consequence relations representing derivability of a sentence from a (multi)set of assumptions. In other words, we now go to a more general but less symmetric format $\Gamma \vdash \varphi$, where Γ is a finite multiset and φ an element of a given set L. In order to avoid confusion we use the more standard notation $\vdash$ instead of $\leqslant$ and will follow the standard conventions of sequential proof format. In particular, Γ, φ denotes the result of adjoining $\varphi \in L_S$ to a multiset of sentences Γ, and Γ, Δ denotes the multiset union of Γ and Δ.[2]

Definition 3.1. *A consequence relation with an implication on L* is a structure $S = (L_S, \vdash, \to, \top, \bot)$ where $\vdash$ is a binary relation between finite multisets of elements of L_S and elements of L_S; $\to$ is a binary operation on L; $\top$ and $\bot$ are distinguished elements of L such that the following conditions hold:

I1. $\varphi \vdash \varphi$;

I2. if $\Gamma, \psi \vdash \varphi$ and $\Delta \vdash \psi$ then $\Gamma, \Delta \vdash \varphi$;

I3. $\Gamma, \varphi \vdash \psi \iff \Gamma \vdash \varphi \to \psi$;

I4. $\Gamma, \top \vdash \varphi \iff \Gamma \vdash \varphi$.

Notice that Conditions I1 and I2 generalize reflexivity and transitivity of $\leqslant$. Setting $\varphi \leqslant_S \psi$ as $\varphi \vdash \psi$ yields an abstract consequence relation in the sense of Definition 2.1. Condition I3 speaks for itself. Condition I4 conveniently stipulates that provability from the empty multiset of assumptions is the same as provability from $\top$. It also implies $\top \to \bot =_S \bot$.

Similarly to the implication one can consider consequence relations with other additional connectives of which we are mostly interested in conjunction.

Definition 3.2. *Conjunction* is a binary operator $\otimes : L_S^2 \to L_S$ satisfying

$$\Gamma, \varphi, \psi \vdash \theta \iff \Gamma, \varphi \otimes \psi \vdash \theta.$$

[2] Our strive for generality does not go as far as to consider lists of formulas rather than multisets.

If conjunction is available, then $\varphi_1, \dots, \varphi_n \vdash \psi$ holds in S if and only if $\varphi_1 \otimes \cdots \otimes \varphi_n \vdash \psi$. Hence, in the presence of conjunction in S the relation $\leqslant_S$ uniquely determines the corresponding multiset consequence relation $\Gamma \vdash \varphi$.

For consequence relations with an implication we can define negation $\neg\varphi$ by $\varphi \to \bot$. The following simple lemma shows that the implication respects the deductive equivalence relation in S and the negation satisfies the contraposition principle.

Lemma 3.3. (i) *If $\Gamma \vdash \varphi \to \psi$ and $\Delta \vdash \varphi$, then $\Gamma, \Delta \vdash \psi$;*

(ii) *$\varphi_1 =_S \varphi_2$ and $\psi_1 =_S \psi_2$ implies $(\varphi_1 \to \psi_1) =_S (\varphi_2 \to \psi_2)$;*

(iii) *$\Gamma, \varphi \vdash \psi$ implies $\Gamma, \neg\psi \vdash \neg\varphi$.*

Next we turn to the derivability conditions. Assume S is a consequence relation with an implication.

Definition 3.4. $\Box : L_S \to L_S$ satisfies Löb's derivability conditions for S if

L1. $\Box(\varphi \to \psi) \vdash \Box\varphi \to \Box\psi$;

L2. $\Box\varphi \vdash \Box\Box\varphi$;

L3. $\vdash \varphi$ implies $\vdash \Box\varphi$.

Lemma 3.5. *For any consequence relation with an implication the following statements are equivalent:*

(i) *$\Box$ satisfies Löb's conditions for S;*

(ii) *$\Box$ satisfies L2 and S is closed under the rule*

$$\frac{\Gamma \vdash \varphi}{\Box\Gamma \vdash \Box\varphi};$$

(iii) *S is closed under the rule*

$$\frac{\Gamma, \Box\Delta \vdash \varphi}{\Box\Gamma, \Box\Delta \vdash \Box\varphi}.$$

Remark 3.6. Notice that the last rule is formulated slightly differently from the more standard rule for modal logic K4:

$$\frac{\Gamma, \Box\Gamma \vdash \varphi}{\Box\Gamma \vdash \Box\varphi}.$$

The latter has a form of built-in contraction that we are not assuming here.

It is natural to define refutability $\boxtimes\varphi$ as provability of negation $\Box\neg\varphi$. Notice that since $\bot =_S \top \to \bot$ we have $\boxtimes\top =_S \Box\bot$, whenever L1 holds for $\Box$. However, as the example in Section 4 shows, this translation does not always yield an APS in the sense of Definition 2.3. To sort things out we need to consider two additional conditions on the consequence relation.

Definition 3.7. A consequence relation with an implication

- *satisfies contraction* if $\Gamma, \varphi, \varphi \vdash \psi$ implies $\Gamma, \varphi \vdash \psi$;
- *satisfies weakening* if $\Gamma \vdash \psi$ implies $\Gamma, \varphi \vdash \psi$, for any φ.

The first condition intuitively means that any hypothesis can be used several times in a derivation. Recall that for Girard's linear logic this condition is not met, however it is postulated, for example, for relevant logics. It turns out that a certain amount of contraction is essential for the proof of G2.

The second condition corresponds to the requirement $x \leqslant_S \top$ that was needed to guarantee that $\boxtimes\top$ is a Gödelian fixed point and that such a fixed point is unique.

For consequence relations with an implication we have the following proposition.

Proposition 3.8. *Suppose S satisfies contraction, $\Box : L_S \to L_S$ satisfies Löb's conditions for S and $\boxtimes\varphi := \Box(\varphi \to \bot)$. Then $(L_S, \leqslant_S, \Box, \boxtimes, \top, \bot)$ is an APS.*

Proof. By Lemma 3.3 $\varphi \vdash \psi$ implies $\neg\psi \vdash \neg\varphi$. This yields Conditions C1 and C4. Condition C2 obviously follows from Condition 1 for a good consequence relation. Let us prove C3. By Lemma 3.3(i) we have: $\varphi, \neg\varphi, \top \vdash \bot$. Hence, $\varphi, \neg\varphi \vdash \top \to \bot$, therefore $\Box\varphi, \Box\neg\varphi \vdash \Box\neg\top$ by Condition 1. The rules of transitivity and contraction imply that, if $\Gamma \vdash \Box\varphi$ and $\Gamma \vdash \Box\neg\varphi$, then $\Gamma \vdash \Box\neg\top$. $\Box$

Thus, from Proposition 3.8 we obtain the following expected corollary, parallel to Theorem 1, for consequence relations satisfying contraction.

Theorem 3. *Suppose S satisfies contraction and $\Box$ satisfies Löb's conditions for S. Then Theorem 1 holds for S.*

For an analogue of Theorem 2 on the uniqueness of a Gödelian fixed point we also need a weakening property.

Theorem 4. *Suppose S satisfies contraction and weakening and $\Box$ satisfies Löb's conditions for S. Then all Gödelian fixed points in S (if exist) are equivalent to $\boxtimes\top =_S \Box\bot$.*

Remark 3.9. As it turns out, contraction and weakening for S, though natural, are somewhat excessive requirements for the validity of Theorems 3 and 4. A consequence relation with an implication

- *satisfies $\Box$-contraction* if $\Gamma, \Box\varphi, \Box\varphi \vdash \psi$ implies $\Gamma, \Box\varphi \vdash \psi$;
- *satisfies $\Box$-weakening* if $\Gamma \vdash \varphi$ implies $\Gamma, \Box\psi \vdash \varphi$, for any ψ.

Conditions C3 and C5 of APS can also be weakened to

C3′. $\boxtimes x \leqslant_S \Box y,\ \boxtimes x \leqslant_S \boxtimes y \ \Rightarrow \boxtimes x \leqslant_S \boxtimes\top$;

C5′. $\boxtimes x \leqslant_S \top$.

With these modifications, the proofs of Theorems 1 and 2 stay the same, which in turn yields more general versions of Theorems 3 and 4 for consequence relations satisfying only $\Box$-contraction and $\Box$-weakening.

The property of $\Box$-contraction actually holds for some meaningful arithmetical systems lacking general contraction rule, for example, for a version of Peano arithmetic based on affine predicate logic considered by the second author of this paper (as yet, unpublished).

4 A non-Gödelian theory with fixed points

In view of Theorems 3 and 4 it is natural to ask whether the assumptions of $\Box$-contraction and $\Box$-weakening are substantial for these results. More specifically, two questions immediately present themselves:

1. Does there exist a consequence relation with an implication satisfying Löb's conditions for $\Box$ in which a Gödelian fixed point exists, but G2 fails? (The failure of G2 can be understood in two different senses — as a failure of its formalized version, and as a failure of its non-formalized version. Our example will show the failure of the formalized version.)

2. Do Gödelian fixed points in such a system S have to be unique, even if S satisfies weakening?

In this section we provide an example showing that the answer to the first question is positive and to the second one negative. Moreover, we formulate a system in which there are many more fixed points than are officially required for a proof of G2. Our system S is a version of modal logic K4 based on the multiplicative $\{\to, \otimes, \bot\}$ fragment of a classical logic without contraction. It also has a built-in fixed point operator where the expression $\mathsf{fp}\, x.A(x)$ denotes some fixed point of $A(x)$ for formulas A modalized in the variable x. Thus, one will be able to derive

$$\mathsf{fp}\, x.A(x) =_S A(\mathsf{fp}\, x.A(x)),$$

for each formula $A(x)$ modalized in x. Let us now turn to the exact definitions.

Consider the set of formulas Fm_0 given by the grammar:

$$A ::= p \mid x \mid \bot \mid (A \to A) \mid \Box A\ ,$$

where p stands for *atomic propositions* and x stands for *variables* (the alphabets of atomic propositions and variables are disjoint). We define the *set of formulas of* S by extending the set Fm_0 by a new constructor: if A is a formula and all free occurrences of x in A are within the scope of modal operators, then $\mathsf{fp}\, x.A$ is a formula, and $\mathsf{fp}\, x$ binds all free occurrences of x. A formula B is *closed* if it does not contain any free occurrences of variables. For a closed formula B, we denote by $A[B/\!/x]$ the result of replacing all free occurrences of x in A by B. We also put $\neg A := A \to \bot$, $\top := \neg\bot$ and $A \otimes B := \neg(A \to \neg B)$.

A *sequent* is an expression of the form $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite multisets of closed formulas. The sequent calculus S is defined in the standard way by the following initial sequents and inference rules:

$$\Gamma, A \Rightarrow A, \Delta \qquad \Gamma, \bot \Rightarrow \Delta$$

$$(\mathsf{fix_L})\ \frac{\Gamma, A[\mathsf{fp}\, x.\, A//x] \Rightarrow \Delta}{\Gamma, \mathsf{fp}\, x.\, A \Rightarrow \Delta} \qquad (\mathsf{fix_R})\ \frac{\Gamma \Rightarrow A[\mathsf{fp}\, x.\, A//x], \Delta}{\Gamma \Rightarrow \mathsf{fp}\, x.\, A, \Delta}$$

$$(\to_{\mathsf{L}})\ \frac{\Gamma, B \Rightarrow \Delta \qquad \Sigma \Rightarrow A, \Pi}{\Gamma, \Sigma, A \to B \Rightarrow \Pi, \Delta} \qquad (\to_{\mathsf{R}})\ \frac{\Gamma, A \Rightarrow B, \Delta}{\Gamma \Rightarrow A \to B, \Delta}$$

$$(\Box)\ \frac{\Sigma, \Box\Pi \Rightarrow A}{\Gamma, \Box\Sigma, \Box\Pi \Rightarrow \Box A, \Delta}\ .$$

Explicitly displayed formulas in the conclusions of the rules are called *principal formulas* of the corresponding inferences. In the rules $(\mathsf{fix_L})$, $(\mathsf{fix_R})$, $(\to_{\mathsf{L}})$ and $(\to_{\mathsf{R}})$, the elements of Γ, Δ, Σ and Π are called *side formulas*. In initial sequents and in applications of the rule $(\Box)$, the elements of Γ and Δ are *weakening formulas*. We call the elements of $\Box\Sigma$ and $\Box\Pi$ in the corresponding applications of $(\Box)$ *active formulas*. In addition, explicitly displayed formulas in initial sequents are called *axiomatic formulas*.

A *proof in* S is a finite tree whose nodes are marked by sequents and leaves are marked by initial sequents that is constructed according to the rules of the sequent calculus. A sequent $\Gamma \Rightarrow \Delta$ *is provable in* S if there is a proof with the root marked by $\Gamma \Rightarrow \Delta$.

We associate with S a consequence relation with an implication and conjunction in the usual way by letting $\Gamma \vdash_{\mathsf{S}} \varphi$ iff $\Gamma \Rightarrow \varphi$ is provable in S. The main thing we need to prove about S is the closure of S under the cut rule, which would show that $\Gamma \vdash_{\mathsf{S}} \varphi$ is indeed a well-defined consequence relation (see Theorem 5 below).

Since S is cut-free, the following propositions are easy to establish. Firstly, we obtain the failure of formalized G2.

Proposition 4.1. *The sequent* $\Box(\Box\bot \to \bot) \Rightarrow \Box\bot$ *is not provable in* S.

Recall that an inference rule is called admissible (for a given proof system) if, for every instance of the rule, the conclusion is provable whenever all premises are provable.

Proposition 4.2. *The Löb rule and the Henkin rule*

$$(\mathsf{L\ddot{o}b})\ \frac{\Box A \Rightarrow A}{\Rightarrow A} \qquad (\mathsf{Hen})\ \frac{\Box A \Rightarrow A \qquad A \Rightarrow \Box A}{\Rightarrow A}$$

are not admissible in S.

Proof. Consider the Henkin fixed point $\mathsf{fp}\, x.\, \Box x$. The sequent $\Rightarrow \mathsf{fp}\, x.\, \Box x$ is not provable in S. Hence, the Henkin rule is not admissible and so is the stronger Löb rule. ∎

Proposition 4.3. *There are infinitely many Henkinian and Gödelian fixed points in* S*.*

Proof. The routine of bound variables in S is such that the formulas $\mathsf{fp}\, x_i.\Box x_i$ for graphically distinct variables x_i are all inequivalent. (There is no rule of bound variables renaming and, in fact, it is easy to convince oneself that there are no cut-free proofs in S of the sequents $\mathsf{fp}\, x_i.\Box x_i \Rightarrow \mathsf{fp}\, x_j.\Box x_j$, for $i \neq j$.) The same holds for the Gödelian fixed points of S. ☐

5 Cut-admissibility for S

For a proof of the cut-admissibility theorem for S we need the following standard lemma. Let the *size* $\|\pi\|$ *of a proof* π be the number of nodes in π.

Lemma 5.1. *The weakening rule*

$$(\mathsf{weak})\,\frac{\Gamma \Rightarrow \Delta}{\Sigma, \Gamma \Rightarrow \Delta, \Pi}$$

is admissible for S*, and its conclusion has a proof of at most the same size as the premise.*

Theorem 5. *The cut rule*

$$(\mathsf{cut})\,\frac{\Gamma \Rightarrow \Delta, A \qquad A, \Sigma \Rightarrow \Pi}{\Gamma, \Sigma \Rightarrow \Pi, \Delta}\,,$$

is admissible for S*. Moreover, if* π_1 *and* π_2 *are proofs of the premises of* (cut)*, then the conclusion of* (cut) *has a proof with the size being less than* $\|\pi_1\| + \|\pi_2\|$*.*

Proof. Assume we have an inference

$$(\mathsf{cut})\,\frac{\begin{matrix}\pi_1 \\ \vdots \\ \Gamma \Rightarrow \Delta, A\end{matrix} \qquad \begin{matrix}\pi_2 \\ \vdots \\ A, \Sigma \Rightarrow \Pi\end{matrix}}{\Gamma, \Sigma \Rightarrow \Pi, \Delta}\,,$$

where π_1 and π_2 are proofs in S. We proof by induction on $\|\pi_1\| + \|\pi_2\|$ that for any formula A there exists a proof $\mathcal{E}_A(\pi_1, \pi_2)$ of $\Gamma, \Sigma \Rightarrow \Pi, \Delta$ with the size being less than $\|\pi_1\| + \|\pi_2\|$.

Consider the final inference in π_1. If the formula A is in a position of a weakening formula in it, then we erase A in π_1 and extend the sequent $\Gamma \Rightarrow \Delta$ to $\Gamma, \Sigma \Rightarrow \Pi, \Delta$ by adding new weakening formulas. This transformation of π_1 defines $\mathcal{E}_A(\pi_1, \pi_2)$. Moreover, we have $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\pi_1\| < \|\pi_1\| + \|\pi_2\|$.

Suppose the formula A is an axiomatic formula in the final inference of π_1. Then the proof π_1 consists of an initial sequent and the multiset Γ has the form Γ_0, A. We obtain $\mathcal{E}_A(\pi_1, \pi_2)$ by applying the admissible rule (weak):

$$(\mathsf{weak})\,\frac{\begin{matrix}\pi_2 \\ \vdots \\ A, \Sigma \Rightarrow \Pi\end{matrix}}{\Gamma_0, A, \Sigma \Rightarrow \Pi, \Delta}\,.$$

We have $\|\mathcal{E}_A(\pi_1, \pi_2)\| \leqslant \|\pi_2\| < \|\pi_1\| + \|\pi_2\|$.

Now suppose the formula A is a side formula. Then the final inference in π_1 can be $(\mathsf{fix}_\mathsf{L})$, $(\mathsf{fix}_\mathsf{R})$, $(\rightarrow_\mathsf{L})$ or $(\rightarrow_\mathsf{R})$.

In the case of $(\rightarrow_\mathsf{R})$, the proof π_1 has the form

$$\begin{array}{c} \pi_1' \\ \vdots \\ (\rightarrow_\mathsf{R})\,\dfrac{\Gamma, B \Rightarrow C, \Delta_0, A}{\Gamma \Rightarrow B \rightarrow C, \Delta_0, A} \end{array},$$

where $\Delta = B \rightarrow C, \Delta_0$. We define $\mathcal{E}_A(\pi_1, \pi_2)$ as

$$\begin{array}{c} \mathcal{E}_A(\pi_1', \pi_2) \\ \vdots \\ (\rightarrow_\mathsf{R})\,\dfrac{\Gamma, B, \Sigma \Rightarrow \Pi, C, \Delta_0}{\Gamma, \Sigma \Rightarrow \Pi, B \rightarrow C, \Delta_0} \end{array}.$$

The proof $\mathcal{E}_A(\pi_1', \pi_2)$ is defined by the induction hypothesis for π_1' and π_2. We also have $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_A(\pi_1', \pi_2)\| + 1 < \|\pi_1'\| + \|\pi_2\| + 1 = \|\pi_1\| + \|\pi_2\|$.

In the case of $(\mathsf{fix}_\mathsf{R})$, the proof π_1 has the form

$$\begin{array}{c} \pi_1' \\ \vdots \\ (\mathsf{fix}_\mathsf{R})\,\dfrac{\Gamma \Rightarrow B[\mathsf{fp}\, x.\, B//x]\Delta_0, A}{\Gamma \Rightarrow \mathsf{fp}\, x.\, B, \Delta_0, A} \end{array},$$

where $\Delta = \mathsf{fp}\, x.\, B, \Delta_0$. We define $\mathcal{E}_A(\pi_1, \pi_2)$ as

$$\begin{array}{c} \mathcal{E}_A(\pi_1', \pi_2) \\ \vdots \\ (\mathsf{fix}_\mathsf{R})\,\dfrac{\Gamma, \Sigma \Rightarrow \Pi, B[\mathsf{fp}\, x.\, B//p], \Delta_0}{\Gamma, \Sigma \Rightarrow \Pi, \mathsf{fp}\, x.\, B, \Delta_0} \end{array}.$$

The proof $\mathcal{E}_A(\pi_1', \pi_2)$ is defined by the induction hypothesis, and $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_A(\pi_1', \pi_2)\| + 1 < \|\pi_1'\| + \|\pi_2\| + 1 = \|\pi_1\| + \|\pi_2\|$.

The remaining cases of $(\rightarrow_\mathsf{L})$ and $(\mathsf{fix}_\mathsf{L})$ can be analyzed analogously, so we omit them.

Now consider the final inference in π_2. If the formula A is a weakening, an axiomatic or a side formula in it, then we can define $\mathcal{E}_A(\pi_1, \pi_2)$ in a similar way to the previous cases.

Suppose that the formula A is a principal or an active formula in the final inferences of π_1 and π_2. Then A has the form $\mathsf{fp}\, x.\, A_0$, $A_0 \rightarrow A_1$ or $\Box A_0$.

If $A = \Box A_0$, then π_2 has one of the two forms

$$\begin{array}{c} \pi_2' \\ \vdots \\ (\Box)\,\dfrac{A_0, \Sigma_1, \Box\Sigma_2 \Rightarrow D}{\Sigma_0, \Box A_0, \Box\Sigma_1, \Box\Sigma_2 \Rightarrow \Box D, \Pi_0} \end{array} \qquad \begin{array}{c} \pi_2' \\ \vdots \\ (\Box)\,\dfrac{\Sigma_1, \Box A_0, \Box\Sigma_2 \Rightarrow D}{\Sigma_0, \Box\Sigma_1, \Box A_0, \Box\Sigma_2 \Rightarrow \Box D, \Pi_0} \end{array},$$

where $\Sigma = \Sigma_0, \Box\Sigma_1, \Box\Sigma_2$ and $\Pi = \Box D, \Pi_0$. In addition, the proof π_1 has the form

$$(\Box)\,\dfrac{\begin{array}{c}\pi_1'\\ \vdots\\ \Gamma_1, \Box\Gamma_2 \Rightarrow A_0\end{array}}{\Gamma_0, \Box\Gamma_1, \Box\Gamma_2 \Rightarrow \Box A_0, \Delta}\,,$$

where $\Gamma = \Gamma_0, \Box\Gamma_1, \Box\Gamma_2$. If π_2 has the first form, then we define $\mathcal{E}_A(\pi_1, \pi_2)$ as

$$(\Box)\,\dfrac{\begin{array}{c}\mathcal{E}_{A_0}(\pi_1', \pi_2')\\ \vdots\\ \Gamma_1, \Box\Gamma_2, \Sigma_1, \Box\Sigma_2 \Rightarrow D\end{array}}{\Gamma_0, \Box\Gamma_1, \Box\Gamma_2, \Sigma_0, \Box\Sigma_1, \Box\Sigma_2 \Rightarrow \Box D, \Pi_0, \Delta}\,.$$

We have $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_{A_0}(\pi_1', \pi_2')\| + 1 < \|\pi_1'\| + \|\pi_2'\| + 1 < \|\pi_1\| + \|\pi_2\|$. If π_2 has the second form, then we define $\mathcal{E}_A(\pi_1, \pi_2)$ as

$$(\Box)\,\dfrac{\begin{array}{c}\mathcal{E}_A(f(\pi_1), \pi_2')\\ \vdots\\ \Box\Gamma_1, \Box\Gamma_2, \Sigma_1, \Box\Sigma_2 \Rightarrow D\end{array}}{\Gamma_0, \Box\Gamma_1, \Box\Gamma_2, \Sigma_0, \Box\Sigma_1, \Box\Sigma_2 \Rightarrow \Box D, \Pi_0, \Delta}\,,$$

where $f(\pi_1)$ is the proof obtained by erasing multisets Γ_0 and Δ from the conclusion of π_1. We have $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_A(f(\pi_1), \pi_2')\| + 1 < \|f(\pi_1)\| + \|\pi_2'\| + 1 = \|\pi_1\| + \|\pi_2\|$.

In the case of $A = \mathsf{fp}\, x.\, A_0$, the proofs π_1 and π_2 have the form

$$(\mathsf{fix_R})\,\dfrac{\begin{array}{c}\pi_1'\\ \vdots\\ \Gamma \Rightarrow \Delta, A_0[\mathsf{fp}\, x.\, A_0//x]\end{array}}{\Gamma \Rightarrow \Delta, \mathsf{fp}\, x.\, A_0} \qquad (\mathsf{fix_L})\,\dfrac{\begin{array}{c}\pi_2'\\ \vdots\\ A_0[\mathsf{fp}\, x.\, A_0//x], \Sigma \Rightarrow \Pi\end{array}}{\mathsf{fp}\, x.\, A_0, \Sigma \Rightarrow \Pi}\,.$$

We put $\mathcal{E}_A(\pi_1, \pi_2) = \mathcal{E}_{A_0[\mathsf{fp}\, x.\, A_0//x]}(\pi_1', \pi_2')$ and see that $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_{A_0[\mathsf{fp}\, x.\, A_0//x]}(\pi_1', \pi_2')\| < \|\pi_1'\| + \|\pi_2'\| < \|\pi_1\| + \|\pi_2\|$.

If $A = A_0 \to A_1$, then the proofs π_1 and π_2 have the form

$$(\to_{\mathsf{R}})\,\dfrac{\begin{array}{c}\pi_1'\\ \vdots\\ A_0, \Gamma \Rightarrow \Delta, A_1\end{array}}{\Gamma \Rightarrow \Delta, A_0 \to A_1} \qquad (\to_{\mathsf{L}})\,\dfrac{\begin{array}{c}\pi_2'\\ \vdots\\ A_1, \Sigma_1 \Rightarrow \Pi_1\end{array} \qquad \begin{array}{c}\pi_2''\\ \vdots\\ \Sigma_0 \Rightarrow \Pi_0, A_0\end{array}}{\Sigma_0, A_0 \to A_1, \Sigma_1 \Rightarrow \Pi_0, \Pi_1}\,,$$

where $\Sigma = \Sigma_0, \Sigma_1$ and $\Pi = \Pi_0, \Pi_1$. By the induction hypothesis, $\mathcal{E}_{A_0}(\pi_2'', \pi_1')$ is defined and $\|\mathcal{E}_{A_0}(\pi_2'', \pi_1')\| < \|\pi_2''\| + \|\pi_1'\|$. Since $\|\mathcal{E}_{A_0}(\pi_2'', \pi_1')\| + \|\pi_2'\| < \|\pi_2''\| + \|\pi_1'\| + \|\pi_2'\| < \|\pi_1\| + \|\pi_2\|$, then $\mathcal{E}_{A_1}(\mathcal{E}_{A_0}(\pi_2'', \pi_1'), \pi_2')$ is defined by the induction hypothesis. We put $\mathcal{E}_A(\pi_1, \pi_2) = \mathcal{E}_{A_1}(\mathcal{E}_{A_0}(\pi_2'', \pi_1'), \pi_2')$. In addition, we have $\|\mathcal{E}_A(\pi_1, \pi_2)\| = \|\mathcal{E}_{A_1}(\mathcal{E}_{A_0}(\pi_2'', \pi_1'), \pi_2')\| < \|\mathcal{E}_{A_0}(\pi_2'', \pi_1')\| + \|\pi_2'\| < \|\pi_1\| + \|\pi_2\|$. $\Box$

6 Conclusions and future work

The preliminary results presented in this paper indicate the following conclusions:

- Derivability conditions can be stated in a way not assuming much about logic. However,
- Gödel's argument presupposes a certain amount of contraction for the logic under consideration.

The role of contraction rule here is somewhat similar to its role in Liar-type paradoxes including Russell's paradox in set theory. Thus, Vyacheslav Grishin (see [7, 8]) pioneered the study of set theory with full comprehension based on a logic without contraction. He demonstrated that the pure comprehension scheme is consistent in this logic. He also showed, however, that the extensionality principle allows for this system to actually *prove* contraction even if there is no postulated contraction in the logic.

One can also consider systems of arithmetic based on contraction-free logic, see e.g. Restall [16, Chapter 11]. For one such system, considered by the second author of this paper, the rule of $\Box$-contraction is admissible, which according to our results still yields G2. Thus, we are still missing convincing examples of mathematical theories based on weak logics for which G2 would fail.

- For consequence relations with an implication and with $\Box$ satisfying Löb's conditions, the existence of appropriately many fixed points does not imply their uniqueness. Nor does it imply formalized versions of G2 and Löb's theorem $\Box(\Box\varphi \to \varphi) \vdash \Box\varphi$.

This shows that the move from *diagonalized algebras* in the sense of R. Magari, i.e., Boolean algebras with $\Box$ satisfying Löb's conditions and having fixed points, to *diagonalizable algebras* (modal algebras satisfying Löb's identity) is, in general, not possible for logics without contraction and weakening. See Smoryński [18, 19] for a nice exposition of the original setup.

- One can also show that the admissibility of Löb's rule does not, in general, imply a formalized version of G2.

A system S^* witnessing this property can be obtained by extending the notion of proof in the system S to possibly non-well-founded proof trees. Infinite proofs may arise because of the presence of the fixed point rules. For S^*, unlike S, one can show that Löb's rule is admissible. Yet, formalized G2 is still underivable. The analysis of S^* is based on another cut-admissibility theorem, which we postpone to a later publication.

We remark that the system S does not provide a counterexample to the non-formalized version of G2, since $\Rightarrow \neg\Box\bot$ is not provable. We believe that such a counterexample can be constructed by extending the language of S by an operator similar to ! from linear logic and adding to S a fixed point of the form $a = \Diamond !a$. However, a confirmation of this hypothesis is left for future work.

7 Acknowledgements

The authors would like to thank Johan van Benthem for useful comments and questions. This work is supported by the Russian Foundation for Basic Research, grant 15-01-09218a, and by the Presidential council for support of leading scientific schools.

References

[1] Z. Adamowicz and K. Zdanowski. Lower bounds for the provability of herbrand consistency in weak arithmetics. *Fundamenta Mathematicae*, 212(3):191–216, 2011.

[2] A. Bezboruah and J. C. Shepherdson. Gödel's second incompleteness theorem for Q. *The Journal of Symbolic Logic*, 41(2):503–512, 1976.

[3] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[4] S. Feferman, J.R. Dawson, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, editors. *Kurt Gödel Collected Works, Volume 1: Publications 1929–1936*. Oxford Univeristy Press, 1996.

[5] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.

[6] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalkuls. *Ergebnisse Math. Kolloq.*, 4:39–40, 1933. English translation in [4], pages 301–303.

[7] V.N. Grishin. On some non-standard logic and its application to set theory. In *Investigations on formalized languages and non-classical logics*, pages 135–171. Nauka, Moscow, 1974. In Russian.

[8] V.N. Grishin. Predicate and set-theoretic calculi based on logic without contractions. *Mathematics of the USSR-Izvestiya*, 18(1):41–59, 1982.

[9] D. Hilbert and P. Bernays. *Grundlagen der Mathematik, Vols. I and II, 2d ed.* Springer-Verlag, Berlin, 1968.

[10] R.G. Jeroslow. Consistency statements in formal theories. *Fundamenta Mathematicae*, 72:2–39, 1970.

[11] R.G. Jeroslow. Redundancies in the Hilbert–Bernays derivability conditions. *The Journal of Symbolic Logic*, 38(3):359–367, 1973.

[12] S.C. Kleene. A symmetric form of Gödel's theorem. *Indagationes Mathematicae*, 12:244–246, 1950.

[13] M.H. Löb. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic*, 20:115–118, 1955.

[14] P. Pudlák. Cuts, consistency statements and interpretations. *The Journal of Symbolic Logic*, 50:423–441, 1985.

[15] W. Rautenberg. *A Concise Introduction to Mathematical Logic*. Springer, second edition, 2006.

[16] G. Restall. *On Logics Without Contraction*. PhD thesis, The University of Queensland, 1994. http://consequently.org/papers/onlogics.pdf.

[17] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North Holland, Amsterdam, 1977.

[18] C. Smoryński. Fixed point algebras. *Bull. Amer. Math. Soc.*, 6(3):317–356, 1982.

[19] C. Smoryński. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.

[20] R.M. Smullyan. *Diagonalization and Self-Reference*. Oxford Logic Guides 27. Oxford University Press, 1994.

[21] A. Visser. Unprovability of small inconsistency. *Archive for Math. Logic*, 32:275–298, 1993.

[22] A. Visser. Can we make the Second Incompleteness Theorem coordinate free? *Journal of Logic and Computation*, 21(4):543–560, 2011.

[23] A. Visser. The Second Incompleteness Theorem and bounded interpretations. *Studia Logica*, 100(1–2):399–418, 2012.

[24] A. Wilkie and J. Paris. On the scheme of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, 35:261–302, 1987.

[25] D. Willard. Self-verifying systems, the incompleteness theorem and the tangibility reflection principle. *The Journal of Symbolic Logic*, 66:536–596, 2001.

[26] D. Willard. A generalization of the Second Incompleteness Theorem and some exceptions to it. *Annals of Pure and Applied Logic*, 141:472–496, 2006.