

# ESSENTIALLY $\Sigma_1$ FORMULAE IN $\Sigma\mathcal{L}$

JACOB VOSMAER

ABSTRACT. The essentially  $\Sigma_1$  formulae of  $\Sigma\mathcal{L}$  are exactly those which are provably equivalent to a disjunction of conjunctions of  $\Box$  and  $\Sigma_1$  formulae.

## 1. INTRODUCTION AND PRELIMINARIES

**1.1.** Fix some r.e. theory of arithmetic  $T$  extending PA. Let  $A$  be a formula of  $\mathcal{L}_{\Sigma\mathcal{L}}$ . We say  $A$  is *essentially  $\Sigma_1$  with respect to  $T$*  if for every arithmetical realization  $*$ :  $\mathcal{L}_{\Sigma\mathcal{L}} \rightarrow \mathcal{L}_T$ ,  $A^*$  is ( $T$ -equivalent to) a  $\Sigma_1$  formula of  $T$ . In standard provability logic GL, the essentially  $\Sigma_1$  formulae are exactly those which are equivalent to a disjunction of boxed formulae. Below we will prove an analogous result for the logic  $\Sigma\mathcal{L}$ . In §1.2 we will motivate the classification of essentially  $\Sigma_1$  formulae. In §1.3 we briefly touch upon  $\Sigma\mathcal{L}$  and  $\Sigma\text{ILM}$ , the modal logics we are concerned with, before we prove the theorem proper in §2. Final remarks, historical notes and acknowledgements can be found in §3.

**1.2.** One way to think of the classification of the essentially  $\Sigma_1$  formulae of a modal logic  $L$  (with language  $\mathcal{L}$ ) with respect to an arithmetical theory  $T$  is to liken it to arithmetical completeness: if

$$\{A \in \mathcal{L} \mid \forall * (T \vdash A^*)\} = \{A \in \mathcal{L} \mid \vdash_L A\},$$

we say that  $L$  is arithmetically complete (and sound) with respect to  $T$ . On the left hand side we have a set of (modal) formulae characterized arithmetically and on the right hand side we have a set of formulae characterized modally. Similarly, we may arithmetically define a set

$$\mathcal{S} := \{A \in \mathcal{L} \mid \forall * (A^* \text{ is } (T\text{-equivalent to) a } \Sigma_1 \text{ formula of } T)\},$$

and wonder if we can define it modally. This question was first asked by Guaspari for  $L=R$  and  $T=PA$  in [5] (he also conjectured what the modal counterpart would be) and solved by Visser for  $L=GL$  in [8]. The question has later been answered for several other logics (including  $R$ ), see §3 for a brief overview.

**1.3.**  $\Sigma\mathcal{L}$  is an extension of GL, and the language of  $\Sigma\mathcal{L}$  is that of GL extended with a unary modality  $\Sigma_1$ . The intended arithmetical interpretation of the modal formula  $\Sigma_1 A$  is ‘the interpretation of  $A$  is  $T$ -equivalent to a  $\Sigma_1$  formula’. It happens that for our present purposes, we are not at all interested in the modal semantics of  $\Sigma\mathcal{L}$ , but those of its bigger sibling  $\Sigma\text{ILM}$  instead.  $\Sigma\text{ILM}$  frames are  $\text{ILM}$ -frames with and extended forcing relation. An  $\text{ILM}$  frame is a triple  $\langle W, R, S \rangle$ , where  $\langle W, R \rangle$  is a Kripke-frame and  $S$  is a ternary relation; we usually write  $uS_w v$  for  $(w, u, v) \in S$ , treating  $S$  as a collection of binary relations indexed by the set of worlds. Moreover, abusing language we will sometimes write  $S$  when we mean the binary relation  $\bigcup_{w \in W} S_w$ . The forcing relation for fomulae of the form  $\Sigma_1 A$ , which is the main novelty, is

$$M, w \Vdash \Sigma_1 A \text{ if } \forall u, v, w' \text{ s.t. } w(R \cup S)^* w' \text{ and } uS_w v, M, u \Vdash A \Rightarrow M, v \Vdash A.$$

The reason we presently introduce  $\Sigma\text{ILM}$  is that it has been proven to be arithmetically complete with respect to (r.e. theories extending) PA, with  $A \triangleright B$  being interpreted as ‘PA + B is  $\Pi_1$ -conservative over PA + A’; in other words,  $\Sigma\text{ILM}$  is (contains) the logic of  $\Pi_1$ -conservativity over PA. We will make use of this below by negating  $\Sigma_1$  formulae, thus making them  $\Pi_1$ .

The idea to extend GL with an operator for  $\Sigma_1$ -ness is due to Japaridze, who introduced a logic of provability extended with (among other things) modalities for  $\Sigma_n$  formulae for all  $n \in \mathbb{N}$  (see [1]). For a real exposition about  $\Sigma\text{L}$  and  $\Sigma\text{ILM}$  we refer the reader to Goris ([3]), where one will also find both modal and arithmetical soundness and completeness results. For an introduction to GL and ILM, one may consult Japaridze and De Jongh’s [2].

## 2. THE THEOREM PROPER

We define

$$\mathbf{C} := \left\{ \bigwedge_{0 \leq i < n} B_i \mid n \in \mathbb{N}, \text{ and either } B_i \equiv \Box D_i \text{ or } B_i \equiv \Sigma_1 D_i \text{ for some } D_i \in \mathcal{L}_{\Sigma\text{L}} \right\},$$

and  $\mathbf{C}_\vee := \{ \bigvee_{0 \leq i < n} C_i \mid n \in \mathbb{N}, C_i \in \mathbf{C} \}$ . Observe that not only is  $\mathbf{C}_\vee$  closed under disjunctions, but (up to equivalence) also under conjunctions: suppose

$$(A \wedge B) \vee (C \wedge D), (E \wedge F) \vee (G \wedge H) \in \mathbf{C}_\vee,$$

then by distributivity of  $\wedge$  over  $\vee$  and associativity of  $\wedge$ , their conjunction is equivalent to

$$(A \wedge B \wedge E \wedge F) \vee (A \wedge B \wedge G \wedge H) \vee (C \wedge D \wedge E \wedge F) \vee (C \wedge D \wedge G \wedge H),$$

which is again a member of  $\mathbf{C}_\vee$ . Additionally, we have  $\top, \perp \in \mathbf{C}_\vee$ .

**Theorem 2.1.** *Let  $A$  be a formula of  $\mathcal{L}_{\Sigma\text{L}}$ . Then  $A$  is essentially  $\Sigma_1$  w.r.t.  $T$  iff there exists  $\tilde{A} \in \mathbf{C}_\vee$  with  $\vdash_{\Sigma\text{L}} A \leftrightarrow \tilde{A}$ .*

We will break the proof of this main theorem down into several lemmata.

**Lemma 2.2.** *If  $A \in \mathbf{C}_\vee$ , then  $A$  is essentially  $\Sigma_1$  w.r.t.  $T$ .*

*Proof.* Let  $*$ :  $\mathcal{L}_{\Sigma\text{L}} \rightarrow \mathcal{L}_T$  be an arithmetical realization. First of all, both  $(\Box B)^*$  and  $(\Sigma_1 B)^*$  are  $\Sigma_1$  for any  $\Sigma\text{L}$  formula  $B$ . Secondly, a conjunction of  $\Sigma_1$  formulae is again  $\Sigma_1$ , so any element of  $\mathbf{C}$  is  $\Sigma_1$ . Since a disjunction of  $\Sigma_1$  formulae is also  $\Sigma_1$ , we conclude that any element of  $\mathbf{C}_\vee$  has a  $\Sigma_1$  interpretation. Since  $*$  was arbitrary, any element of  $\mathbf{C}_\vee$  must be essentially  $\Sigma_1$  with respect to  $T$ .  $\square$

The right to left direction of Theorem 2.1 follows from Lemma 2.2. We prove the other direction by contraposition. First, we show that if  $A$  does not have the desired shape, we can find two  $\Sigma\text{ILM}$ -maximal consistent sets, one containing  $A$ , the other containing  $\neg A$ , with a  $\subseteq_{\Box, \Sigma}$ -relation between them (see below). Next, we use these MCSs to create a  $\Sigma\text{ILM}$  model invalidating a special  $\Sigma\text{ILM}$  formula (containing  $A$ ). Finally, we show that if this special formula is not a theorem of  $\Sigma\text{ILM}$  (which it cannot be, by modal soundness of  $\Sigma\text{ILM}$ ), then  $A$  cannot be essentially  $\Sigma_1$ .

The lemma below was originally Lemma 7.9 (‘the  $\Sigma$ -Lemma’) in [4]; our proof extends that of Goris and Joosten. We use the following notation: if  $X, Y$  are sets of formulas we say that  $X \subseteq_{\Box, \Sigma_1} Y$  if  $\Box B \in X \Rightarrow \Box B \in Y$  and  $\Sigma_1 B \in X \Rightarrow \Sigma_1 B \in Y$  for all  $B \in \mathcal{L}_{\Sigma\text{L}}$ .

**Lemma 2.3.** *Let  $A$  be a  $\Sigma\text{L}$  formula such that for no  $\tilde{A} \in \mathbf{C}_\vee$  we have  $\vdash_{\Sigma\text{ILM}} A \leftrightarrow \tilde{A}$ . Then there exist  $\Sigma\text{ILM}$ -maximal consistent sets  $\Gamma_0, \Gamma_1$  such that  $A \in \Gamma_0 \subseteq_{\Box, \Sigma_1} \Gamma_1 \ni \neg A$ .*

*Proof.* Assume that  $A$  is not equivalent to any member of  $\mathcal{C}_\vee$ . First, we define

$$\mathcal{C}_{\text{con}} := \{Y \subseteq \mathcal{C}_\vee \mid \{\neg A\} + Y \text{ is } \Sigma\text{ILM}\text{-consistent and maximally such}\}.$$

(Note that if  $\mathcal{C}_{\text{con}}$  is empty, then  $\vdash_{\Sigma\text{ILM}} A \leftrightarrow \top$ , contradicting our assumption about  $A$ .) A useful property of elements  $Y$  of  $\mathcal{C}_{\text{con}}$  is that

$$(2.1) \quad B \vee C \in Y \text{ implies } B \in Y \text{ or } C \in Y.$$

For if  $Y \in \mathcal{C}_{\text{con}}$  and  $B \vee C \in Y$ , then  $B, C \in \mathcal{C}_\vee$ . If neither  $B$  nor  $C$  were consistent with  $\{\neg A\} + Y$ , then we would have  $\neg A + Y \vdash_{\Sigma\text{ILM}} \neg B \wedge \neg C$  and  $\{\neg A\} + Y$  would be inconsistent. So either  $B$  or  $C$  must be consistent with  $\{\neg A\} + Y$ , whence by  $Y$ 's maximality,  $B \in Y$  or  $C \in Y$ . We conclude that (2.1) holds.

**Claim.** For some  $Y \in \mathcal{C}_{\text{con}}$ , the set  $\{A\} + \{\neg\sigma \mid \sigma \in \mathcal{C}_\vee \setminus Y\}$  is consistent.

From the assumption that the claim is false we will derive a contradiction to our initial assumption about  $A$ . If the claim is false, then for each  $Y \in \mathcal{C}_{\text{con}}$  there is some finite  $Y^{\text{fin}} \subseteq \mathcal{C}_\vee \setminus Y$  such that  $\{A\} + \{\neg\sigma \mid \sigma \in Y^{\text{fin}}\}$  is inconsistent. Therefore,

$$(2.2) \quad \text{for each } Y \in \mathcal{C}_{\text{con}}, \text{ we have } \vdash_{\Sigma\text{ILM}} A \rightarrow \bigvee_{\sigma \in Y^{\text{fin}}} \sigma.$$

Next, we will show that

$$(2.3) \quad \{\neg A\} + \left\{ \bigvee_{\sigma \in Y^{\text{fin}}} \sigma \mid Y \in \mathcal{C}_{\text{con}} \right\} \text{ is inconsistent.}$$

For if not, then that fact must be witnessed by some  $S \in \mathcal{C}_{\text{con}}$  such that  $\{\bigvee_{\sigma \in Y^{\text{fin}}} \sigma \mid Y \in \mathcal{C}_{\text{con}}\} \subseteq S$ . Now we are in a case of fatal self-reference, because this means that in particular  $\bigvee_{\sigma \in S^{\text{fin}}} \sigma \in S$ , so by (2.1), we have  $\sigma \in S$  for some  $\sigma \in S^{\text{fin}}$ , contradicting the fact that  $S^{\text{fin}} \subseteq \mathcal{C}_\vee \setminus S$ . We conclude that (2.3) holds.

There must be some finite  $\mathcal{C}_{\text{con}}^{\text{fin}} \subseteq \mathcal{C}_{\text{con}}$  witnessing this situation, so we get

$$\vdash_{\Sigma\text{ILM}} \left( \bigwedge_{Y \in \mathcal{C}_{\text{con}}^{\text{fin}}} \bigvee_{\sigma \in Y^{\text{fin}}} \sigma \right) \rightarrow A.$$

As a consequence of (2.2), we also get

$$\vdash_{\Sigma\text{ILM}} A \rightarrow \left( \bigwedge_{Y \in \mathcal{C}_{\text{con}}^{\text{fin}}} \bigvee_{\sigma \in Y^{\text{fin}}} \sigma \right).$$

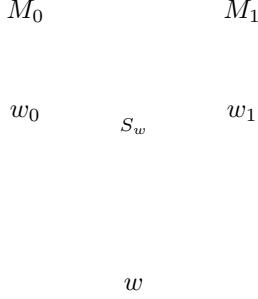
Combining the above two results we get

$$\vdash_{\Sigma\text{ILM}} A \leftrightarrow \left( \bigwedge_{Y \in \mathcal{C}_{\text{con}}^{\text{fin}}} \bigvee_{\sigma \in Y^{\text{fin}}} \sigma \right).$$

Because  $\mathcal{C}_\vee$  is closed under disjunctions, for every  $Y \in \mathcal{C}_{\text{con}}^{\text{fin}}$  we have  $\bigvee_{\sigma \in Y^{\text{fin}}} \sigma \in \mathcal{C}_\vee$ . Because  $\mathcal{C}_\vee$  is also closed under conjunctions, we conclude that  $A$  is equivalent to a member of  $\mathcal{C}_\vee$ , contradicting the initial assumption of the lemma. Therefore, the claim must be true.

If we take  $Y$  to be a set witnessing the truth of the claim above, we find that both  $\{A\} + \{\neg\sigma \mid \sigma \in \mathcal{C}_\vee \setminus Y\}$  and  $\{\neg A\} + Y$  are consistent, so by (an appropriately modified version of) Lindenbaum's Lemma, there exist MCS's  $\Gamma_0$  and  $\Gamma_1$  extending  $\{A\} + \{\neg\sigma \mid \sigma \in \mathcal{C}_\vee \setminus Y\}$  and  $\{\neg A\} + Y$ , respectively. Also, if  $B \equiv \Box C$  or  $B \equiv \Sigma_1 C$  for some  $C \in \mathcal{L}_{\Sigma\mathcal{L}}$  and  $B \notin \Gamma_1$ , then  $B \in \mathcal{C}_\vee \setminus Y$ , whence  $\neg B \in \Gamma_0$ , so by consistency of  $\Gamma_0$ ,  $B \notin \Gamma_0$ . It follows that  $\Gamma_0 \subseteq_{\Box, \Sigma_1} \Gamma_1$ .  $\square$

Before we proceed, we would like to point out a few subtleties. First of all, because  $\Sigma\text{ILM}$  is conservative over  $\Sigma\mathcal{L}$  (Theorem 4.11 in [3]), it follows that if there is no  $\hat{A} \in \mathcal{C}_\vee$  such that  $\vdash_{\Sigma\mathcal{L}} A \leftrightarrow \hat{A}$ , then there can be no  $\hat{A} \in \mathcal{C}_\vee$  for which  $\vdash_{\Sigma\text{ILM}} A \leftrightarrow \hat{A}$ . Secondly, although we have two  $\Sigma\text{ILM}$ -MCS's  $\Gamma_0$  and  $\Gamma_1$  that are

FIGURE 1. Our new model  $M$ 

$\subseteq_{\square, \Sigma_1}$ -related, this relation itself only pertains to  $\Sigma\text{L}$  formulae. We continue with the main argument.

Let  $\text{Sub}(A)$  denote the (finite) set of all subformulas of  $A$ , and define  $\neg X := \{B, \neg B \mid B \in X\}$ . We now want to construct models based on  $\Gamma_0$  and  $\Gamma_1$ . Because of the compactness failure of  $\text{GL}$ , we will reduce  $\Gamma_i$  to  $\Gamma_i^A := \Gamma_i \cap \neg \text{Sub}(A)$  for  $i = 0, 1$ . If it were the case that  $\vdash_{\Sigma\text{ILM}} \neg \bigwedge \Gamma_i^A$ , then  $\Gamma_i^A$  would be inconsistent which is impossible, since  $\Gamma_i^A \subseteq \Gamma_i$  which is an MCS. Therefore we conclude that  $\not\vdash_{\Sigma\text{ILM}} \neg \bigwedge \Gamma_i^A$ , so that by completeness of  $\Sigma\text{ILM}$  (Theorem 3.13 of [3]), there exists a model  $M_i$  with root  $w_i$  such that  $M_i, w_i \Vdash \bigwedge \Gamma_i^A$ . Since  $A \in \Gamma_0^A$  and  $\neg A \in \Gamma_1^A$ , this gives us  $M_0, w_0 \Vdash A$  and  $M_1, w_1 \Vdash \neg A$ . Additionally we may require that for certain fresh variables  $p$  and  $q$ , we have  $\llbracket p \rrbracket^{M_0} = \{w_0\}$ ,  $\llbracket q \rrbracket^{M_1} = \{w_1\}$  and  $\llbracket p \rrbracket^{M_1} = \llbracket q \rrbracket^{M_0} = \emptyset$  ( $p$  and  $q$  mark  $w_0$  and  $w_1$  respectively, so to say).

**Lemma 2.4.** *Fix a formula  $A$  of  $\mathcal{L}_{\Sigma\text{L}}$ . Suppose that there exist  $\Sigma\text{ILM}$  models  $M_0, M_1$  are (with roots  $w_0, w_1$ ) such that  $M_0, w_0 \Vdash A$ ,  $M_1, w_1 \Vdash \neg A$ ,  $\llbracket p \rrbracket^{M_0} = \{w_0\}$ ,  $\llbracket q \rrbracket^{M_1} = \{w_1\}$  and  $\llbracket p \rrbracket^{M_1} = \llbracket q \rrbracket^{M_0} = \emptyset$  for certain propositional variables  $p, q$ . Moreover, assume that  $\{C \in \neg \text{Sub}(A) \mid M_0, w_0 \Vdash C\} \subseteq_{\square, \Sigma_1} \{C \in \neg \text{Sub}(A) \mid M_1, w_1 \Vdash C\}$ . Then  $\not\vdash_{\Sigma\text{ILM}} p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A$ .*

*Proof.* The idea of our proof is to glue  $M_0 = \langle W_0, R_0, S_0 \rangle$  and  $M_1 = \langle W_1, R_1, S_1 \rangle$  together into a  $\Sigma\text{ILM}$  model  $M = \langle W, R, S \rangle$  with root  $w$ , so that  $M, w \not\vdash p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A$ , which proves the lemma by the soundness of  $\Sigma\text{ILM}$ .

$M$  will be the disjoint union of  $M_0$  and  $M_1$  with a new root  $w$  below  $w_0$  and  $w_1$ . To that end, we extend the relation  $R$  with  $wRw_0$  and  $wRw_1$  to attach the new root. To accommodate an  $S_w$ -arrow we will add in a minute, we also add  $w_0R_s$  for all  $s \in W_1 \setminus \{w_1\}$  (see Figure 1). With these new links in place, we add the  $R$ -links required to keep  $R$  transitive. This ensures that  $\langle W, R \rangle$  is a  $\text{GL}$ -frame. Now as promised, we add the new connection  $w_0S_w w_1$ . To turn our contraption into a  $\Sigma\text{ILM}$ -frame, we must also add  $sS_w t$  for all  $s, t$  such that  $wRsRt$  and  $uS_{w_0}v$  for all  $u, v$  such that  $uS_{w_1}v$ . After all this, we close  $S$  off under reflexive steps. This completes the construction of our new  $\Sigma\text{ILM}$  model  $M$ . It is important to remember that not only have we added  $R, S_w$  and  $S_{w_0}$ -links, but indirectly we have also added  $(R \cup S)^*$ -links (which are not directly visible, since this structure resides not in the frame but in the forcing relation). So, if we look closely, we see that no new  $(R \cup S)^*$ -links have been added in the part of  $M$  that is a copy of  $M_1$ : this we will need below.

We now want to show that  $M, w_0 \Vdash A$  and  $M, w_1 \Vdash \neg A$ . Since  $w_1$  does not see any new worlds (not even via the new  $(R \cup S)^*$ -relation, as discussed above), we must have that for all  $B \in \neg \text{Sub}(A)$ ,  $M_1, w_1 \Vdash B$  implies  $M, w_1 \Vdash B$  (so in particular,  $M, w_1 \Vdash \neg A$ ). For  $w_0$ , we will show by induction on formula construction that for all  $B \in \neg \text{Sub}(A)$  and all  $s \in W_0$ , we have  $M, s \Vdash B$  iff  $M_0, s \Vdash B$  (it will then follow that  $M, w_0 \Vdash A$ ). The case for propositional variables and Boolean connectives is immediate, so we will turn to the case that  $B \equiv \Box C$ . If  $s \neq w_0$ , then  $s$  sees the same worlds in both  $M$  and  $M_0$ , and it follows that  $M, s \Vdash \Box C$  iff  $M_0, s \Vdash \Box C$ . Now suppose that  $s = w_0$  and  $M_0, w_0 \Vdash \Box C$ . Then if  $w_0 R t$  (in  $M$ ), either  $t \in W_0$  or  $t \in W_1$ . In the former case, we know by induction hypothesis that  $M, s \Vdash C$ . In the latter case, we use the fact that  $M_1, w_1 \Vdash \Box C$  (remember the last condition of the lemma) whence  $M, w_1 \Vdash \Box C$ , so since it must be that  $w_1 R t$ , we also get  $M, t \Vdash C$ , so since  $t$  was arbitrary,  $M, w_0 \Vdash \Box C$ . Conversely if  $M, w_0 \Vdash \Box C$ , then for all  $t$  s.t.  $w_0 R t$  we have  $M, t \Vdash C$ , so in particular for all  $t \in W_0 \setminus \{w_0\}$ , whence  $M_0, w_0 \Vdash \Box C$ . Finally if  $B \equiv \Sigma_1 C$  and  $s \neq w_0$ , we again immediately get  $M, s \Vdash \Sigma_1 C$  iff  $M_0, s \Vdash \Sigma_1 C$ . Now suppose that  $s = w_0$ , and assume that  $M_0, w_0 \Vdash \Sigma_1 C$ ,  $w_0(R \cup S)^* w'$  and  $u S_w v$  with  $M, u \Vdash C$ . If  $w' \in W_0 \setminus \{w_0\}$ , then  $u, v \in W_0$ . Since  $M_0, v \Vdash C$ , by induction hypothesis we get  $M, v \Vdash C$ . If  $w' = w_0$ , then either  $u, v \in W_0$  or  $u, v \in W_1$ . In the former case it again follows from  $M_0, w_0 \Vdash \Sigma_1 C$  that  $M, v \Vdash C$ . In the latter case, we need the last condition of the lemma again, which gives us  $M_1, w_1 \Vdash \Sigma_1 C$ , so also  $M, w_1 \Vdash \Sigma_1 C$ . Since the connection  $u S_w v$  can only be in place because  $u S_{w_1} v$ ,  $M, v \Vdash C$  follows from  $M, w_1 \Vdash \Sigma_1 C$ . If  $w' \in W_1$ , then  $w_1(R \cup S)^* w'$ , so it again follows from  $M, w_1 \Vdash \Sigma_1 C$  that  $M, v \Vdash C$ . Conversely, if  $M, w_0 \Vdash \Sigma_1 C$  then it follows that  $M_0, w_0 \Vdash \Sigma_1 C$  as above with  $\Box C$ . This completes our induction, and as a consequence we get  $M, w_0 \Vdash A$ .

Now we have  $M, w \Vdash p \triangleright q$  (since  $p$  is only true at  $w_0$  and  $w_0 S_w w_1 \Vdash q$ ). However,  $M, w \not\Vdash p \wedge A \triangleright q \wedge A$ , which is witnessed by  $w_0$ : we have  $w R w_0$  and  $M, w_0 \Vdash p \wedge A$ . Towards a contradiction, suppose that there is  $t$  such that  $w_0 S_w t$  and  $M, t \Vdash q \wedge A$ . Since  $\llbracket q \rrbracket^M = \{w_1\}$ , it follows that  $t = w_1$ , but  $M, w_1 \Vdash \neg A$ . We conclude that for no  $t$ ,  $w_0 S_w t \Vdash q \wedge A$ , whence  $M, w \not\Vdash p \wedge A \triangleright q \wedge A$ , whence  $M, w \not\Vdash p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A$ . The lemma follows.  $\square$

**Lemma 2.5.** *Let  $A$  be a  $\Sigma L$  formula. If  $\not\Vdash_{\Sigma ILM} p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A$  for certain propositional letters  $p, q$ , then  $A$  is not essentially  $\Sigma_1$  w.r.t.  $T$ .*

*Proof.* We reason by contraposition. Assume that  $A$  is essentially  $\Sigma_1$  with respect to  $T$ . Since  $T$  extends PA, we know that  $\Sigma ILM$  is the logic of  $\Pi_1$ -conservativity over  $T$  (by Theorem 4.3 of [3]). Let  $*$  be an arbitrary arithmetical realization. Reasoning in  $T$ , we will argue that  $(p \triangleright q)^*$  implies  $(p \wedge A \triangleright q \wedge A)^*$ . Assume  $(p \triangleright q)^*$ , i.e. that  $T + q^*$  is  $\Pi_1$ -conservative over  $T + p^*$ . We want to show that  $T + q^* + A^*$  is  $\Pi_1$ -conservative over  $T + p^* + A^*$ . Let  $\pi$  be an arbitrary  $\Pi_1$  formula of  $T$ , then from  $T + p^* + A^* \vdash \pi$  it follows that  $T + p^* \vdash A^* \rightarrow \pi$ . Because  $A$  is essentially  $\Sigma_1$  by assumption,  $A^* \rightarrow \pi$  is  $\Pi_1$ , whence by  $\Pi_1$ -conservativity of  $T + q^*$  over  $T + p^*$ ,  $T + q^* \vdash A^* \rightarrow \pi$ , so  $T + q^* + A^* \vdash \pi$ . We conclude that  $(q \wedge A)^*$  is  $\Pi_1$ -conservative over  $(p \wedge A)^*$ . Since we reasoned in  $T$ , we have  $T + (p \triangleright q)^* \vdash (p \wedge A \triangleright q \wedge A)^*$ , and hence  $T \vdash (p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A)^*$ . Now because  $*$  was arbitrary, by arithmetical completeness we may conclude that  $\vdash_{\Sigma ILM} p \triangleright q \rightarrow p \wedge A \triangleright q \wedge A$ , concluding our proof.  $\square$

This concludes the proof of (the left to right direction of, and hence also the full) Theorem 2.1.

### 3. NOTES AND ACKNOWLEDGEMENTS

**3.1.** As mentioned in §1.2, the first classification of the essentially  $\Sigma_1$  formulae of GL is found in [8]. De Jongh and Pianigiani ([6]) classified them for GL and R (R is GL extended with a binary modality for witness comparison). Goris and Joosten ([4]) later classified the essentially  $\Sigma_1$  formulae of ILM. Our proof of Theorem 2.1 is structured like that in [6]. The difference is that presently the two countermodels are constructed out of MCS's obtained by the  $\Sigma$ -Lemma of [4]. We believe that the condition that  $T$  contains PA may be weakened; the only thing we need (for Lemma 2.5) is that  $\Sigma$ ILM is the logic of  $\Pi_1$ -conservativity over  $T$ . In [3] this is proved for superarithmetical<sup>1</sup>  $T$ , so presently we have also stuck to  $T$  extending PA, even though  $\text{I}\Sigma_1$  will probably do, or maybe even less. Additionally, we are optimistic about the possibility to extend the method Goris and Joosten use to classify the essentially  $\Sigma_1$  formulae of ILM to a method classifying those of  $\Sigma$ ILM.

**3.2.** This note was written to aid the author in earning his Master's degree in the Logic Programme at the University of Amsterdam. The author would like to thank his project supervisors Dick de Jongh and Joost Joosten, without the help and advice of either of whom the present note would contain considerably more mistakes than it does now, if it had been written in the first place. Additionally, he thanks Evan Goris, who pointed out some serious oversights in what was meant to be the final version of this note.

### REFERENCES

- [1] G. Japaridze: *The logic of the arithmetical hierarchy*, Annals of Pure and Applied Logic 66, pp. 89–112 (1994).
- [2] Giorgi Japaridze and Dick de Jongh: 'The Logic of Provability', in *Handbook of Proof Theory* (ed. S.R. Buss), 1998, Elsevier.
- [3] Evan Goris: *Extending ILM with an operator for  $\Sigma_1$ -ness*, 2003, ILLC Publications PP-2003-17, Amsterdam.
- [4] Evan Goris and Joost J. Joosten: *Modal Matters in Interpretability Logics*, 2004, Logic Group Preprint Series 226, Utrecht University.
- [5] D. Guaspari: 'Sentences Implying their own Provability', Journal of Symbolic Logic 48, pp. 777–789 (1983).
- [6] Dick de Jongh and Duccio Pianigiani: 'Solution to a problem of David Guaspari', pp. 246–254 in *Logic at Work* (ed. Ewa Orłowska), 1999, Physica-Verlag, Heidelberg.
- [7] Joost J. Joosten: *Interpretability formalized*, 2004, Quaestiones Infnitae, Publications of the Department of Philosophy, Utrecht University.
- [8] A. Visser: 'Notes on Bimodal Provability Logic', in *Dirk van Dalen Festschrift* (eds. H. Barendregt, M. Bezem, J.W. Klop), 1993, Quaestiones Infnitae, Publications of the Department of Philosophy, Utrecht University.

*E-mail address:* Jacob Vosmaer, student uva nl

---

<sup>1</sup>Strictly speaking, Goris assumes  $T = \text{PA}$ . However, his results still hold for r.e. superarithmetical  $T$ .