Estratto da

R. Ferro e A. Zanardo (a cura di), *Atti degli incontri di logica matematica* Volume 3, Siena 8-11 gennaio 1985, Padova 24-27 ottobre 1985, Siena 2-5 aprile 1986.

Disponibile in rete su http://www.ailalogica.it

SOME RECENT DECIDABILITY RESULTS IN SET THEORY

D. CANTONE[1], A. FERRO[2]
[1]Courant Institute
[2]Università di Catania

0. Introduction

Most mathematical theories can be expressed in set theory. Hence a theorem prover for set theory could be used as an *all-purpose* theorem prover or as a general proof-checker. In this sense, the discovery of a result analogous to Herbrand's theorem for predicate calculus, but applying to set theory, would be of great importance (see [Om]).

Herbrand's Theorem. *There exists an automatic procedure P, which, given an arbitrary formula of predicate calculus, produces an infinite sequence $g_1, g_2, g_3, \ldots$ of propositional formulas such that q is a theorem if and only if at least one of the g_i's is unsatisfiable.*

It can be conjectured that the relationship of set theory to the class of unquantified formulas in the language including the operator $\cup$ (binary union), $\cap$ (binary intersection), $\setminus$ (set difference), $\{\cdot\}$ (singleton), *pow* (powerset), *Un* (general union), η (choice function), X (cartesian product) and the predicates $=$ (equality), $\in$ (membership) is analogous to the relationship of predicate calculus to propositional calculus. A preliminary step in finding such an analogue of the Herbrand's theorem might therefore consist in proving the decidability of the theory sketched above.

We will review some partial results in this direction. More precisely we will deal with the langage $=, \in, \cup, \cap, \setminus$ extended either by *pow* or *Un*. Partial decidability results for this language have been found in [BF] and [F]. Here we describe a technique which allows the general case to be solved (see [CFS1], [CFS2] and [C]).

1. Multilevel syllogistic and its extensions

The language MLS is composed using variables $x, y, z, \ldots$, operators $\cup$, $\cap$, $\setminus$, predicates $\in$, $=$. The MLS theory is then the set of unquantified formulas which can be built up from the above constituents by observing the usual syntactic rules.

Normalizing formulas, it is immediate to see that we can limit ourselves to considering only those formulas Q of MLS which are conjunctions of literals of the following types:

($\in$) $x \in y$
($\notin$) $x \notin y$

(=) $x = y \cup z$, $x = y \setminus z$.

The following concepts play a crucial rôle in the solution of the satisfiability problem for MLS and its extensions.

Definition. A boolean function α defined on the variables occurring in Q is called a *place of* Q if $\alpha(x) = \alpha(y) \vee \alpha(z)$ (resp. $\alpha(x) = \alpha(y) \& \neg\alpha(z)$) whenever the literal $x = y \cup z$ (resp. $x = y \setminus z$) occurs in Q. If in addition for a place α of Q and a variable x in Q $\alpha(y) = 1$ (resp. $\alpha(y) = 0$) whenever $x \in y$ (resp. $x \notin y$) occurs in Q, α is said a *place of* Q *at the varibale* x.

Since the number of possible places of Q is finite, the following theorem establishes the decidability of MLS.

Theorem ([FOS]). Q is satisfiable if and only if there exist

(i) a set Γ of places of Q,

(ii) a correspondence $x \mapsto \alpha_x$, where x ranges over the variables of Q (after identification of variables equivalent with respect to the equivalence relation $\sim_\Gamma$, defined by $x \sim_\Gamma y$ iff $\alpha(x) = \alpha(y)$ for all $\alpha \in \Gamma$), and α_x is a place at x,

(iii) an ordering $<$ of variables in Q such that if $x \leq y$ then $\alpha_y(x) = 0$. ■

It can be shown that when conditions (i)–(iii) of the above theorem are satisfied, sets $\overline{\alpha}$ can be effectively associated with the places α in Γ in such a way that the assignment $Mx = \bigcup_{\alpha(x)=1} \overline{\alpha}$ satisfies Q.

Strenghtning such construction, it is possible to take into account even cardinality constraints and the elementary arithmetic of cardinal numbers without loosing decidability. This gives in particular the decidability of the satisfiability problem of MLS extended by the singleton operator (see [FOS]).

2. MLS extended by the general union operator

The theory MLSU obtained from MLS by allowing an unrestricted number of occurrences of the general union operator Un has been shown to have a solvable satisfiability problem [CFS1]. Here we give a brief description of the techniques used in the proof.

Let Q be a conjunction of literals of the form (=), ($\in$), ($\notin$), and (Un), $u = Un(y)$.

In proving the decidability of MLSU decidable conditions are given on places of Q in such a way that a procedure of the type

$ INIZIALIZATION $
For every place α *let*
$\overline{\alpha} \leftarrow \overline{\alpha}^{(0)}$
end for;
Stabilize;
$ END INIZIALIZATION $
$ $\in$-PHASE $
Following the order $<$ *of variables* x *in* Q *do*
$\overline{\alpha}_x \leftarrow \overline{\alpha}_x \cup \{Mx\}$; , $ where $Mx = \bigcup_{\alpha(x)=1} \overline{\alpha}$ $
Stabilize;
end do;
$ END $\in$-PHASE $

can produce a model of Q if any exists.

To this end it turns useful the following definition of Ugraph.

Definition. Given a conjunction Q and a set Γ of places of Q as above, the Ugraph G of Q, Γ is the graph whose set of nodes is Γ, plus one additional node Ω, and whose edges are as follows:

(i) A directed edge connects α to Ω if and only if $\alpha(y) = 0$ for every variable y for which $u = Un(y)$ is in Q. (Intuitively this means that clauses $u = Un(y)$ of Q tells us nothing about the set $Un(\overline{\alpha})$, which allows the proper initialization to start with such places).

(ii) Otherwise, a directed edge connects the place α to the place β if and only if $\beta(u) = 1$ for all clauses $u = Un(y)$ such that $\alpha(y) = 1$. (Intuitively, the nodes β such that (α, β) is an edge of G represent all the sets $\overline{\beta}$ in which elements of $Un(\overline{\alpha})$ can appear).

Three kinds of nodes can be distinguished in the Ugraph. Those from which there is a directed path which reaches Ω are called *safe*. A node is *trapped* if every sufficiently long path from it eventually reaches a node from which no edge branches off (*null* node). Finally a node which is neither safe nor tropped is called *cyclic*. Intuitively trapped places are those places α whose associated set $\overline{\alpha}$ must satisfy severe rank restrictions. In particular it can be seen that such places can be assigned only sets having rank at most one more than the maximum length of a largest path forward from each of them to a null node. Therefore only a finite number of possible choices must be checked. On the other hand it turns out that each nontrapped place α can be assigned an infinite set $\overline{\alpha}$.

A rough description of the first initialization phase is as follows (for simplicity we only consider the case in which no trapped place exists.) First of all infinitely many individuals are associated with every place α of Q such that (α, Ω) is an edge of the Ugraph G. Then any safe place can iteratively be given an infinite supply of elements by drawing elements from its descendants and forming their singletons.

The same technique can also be used to initialize cyclic places, once we observe that the null node $\alpha_\emptyset$ must lie on a cycle which can be given elements by successive formation of singletons of the emptyset $\emptyset$ (which is assigned to $\overline{\alpha}_\emptyset$), and that, by the regularity axiom of set theory, the null node must be rechable along edges of G from every other node. This observation, which in substance is a condition for the satisfiability of Q, guarantees that proper inizialization can be accomplished successfully. Once this phase is completed, all literals of type (=) are correctly modeled; however for literals $u = Un\,(y)$ in Q all we can say is that $Un\left(\bigcup_{\alpha(y)=1} \overline{\alpha}\right) \subseteq \bigcup_{\alpha(u)=1} \overline{\alpha}$.

To get equalities in place of these inclusions, the following stabilization phase is then performed. For each element p which has been put into $\overline{\alpha}$ and for every clause $u = Un\,(y)$ such that $\alpha(u) = 1$ (i.e., intuitively $\overline{\alpha} \subseteq Mu$), an element A is found such that after inserting A into $\overline{\alpha}$ no inclusion of the type above is disrupted. Then the pair $\{p, A\}$ is inserted in a place β such that (β, α) is an edge of the Ugraph G.

We refrain from starting the conditions which guarantee that such a stabilization phase can actually take place, since they are quite involved. The interested reader can find a complete description of them in [CFS1].

An analogous stabilization is then performed in the $\in$-phase.

3. MLS extended by the powerset operator

Let Q be a conjunction of literals of the type $(=)$, $(\in)$, $(\notin)$, and (pow), $p = pow(q)$. Let $s_1, s_2, \dots, s_n$ be nonempty disjoint sets. Then we have

$$pow(s_1 \cup s_2 \cup \cdots \cup s_n) = \bigcup_{\emptyset \neq A \subseteq \{s_1,\dots,s_n\}} pow^*(A),$$

where $pow^*(A)$ stands for the set of those subsets of $\bigcup_{s \in A} s$ which have nonempty intersection with every element of A.

Hence, if $p = pow(q)$ is a powerset clause in Q, and $\alpha_1, \dots, \alpha_l$ are places of Q such that $\alpha_1(q) = \cdots = \alpha_l(q) = 1$, there must exist places $\beta_1, \dots, \beta_k$ with $\beta_1(p) = \cdots = \beta_k(p) = 1$ and such that elements of $pow^*(\overline{\alpha}_1, \dots, \overline{\alpha}_l)$ can lie in $\overline{\beta}_1 \cup \overline{\beta}_2 \cup \cdots \cup \overline{\beta}_k$. This relationship is called a *P-edge*, $\{\alpha_1, \dots, \alpha_l\}$ is a P-node and $\beta_1, \dots, \beta_k$ are *targets* of $\{\alpha_1, \dots, \alpha_l\}$. If a place is not the target of any P-node, then it is called *initial*. (Intuitively, initial places are those places which are not constrained by powerset clauses. It is reasonable to start initializaion from these places.)

A first condition for Q to be satisfiable follows from the consideration that if $s = pow(t)$, then $u \in s$ if and only if $u \subseteq t$. In terms of places this translates as follows: "if $p = pow(q)$ occurs in Q, then $\alpha_x(p) = 1$ if and only if for every place α such that $\alpha(x) = 1$, we have $\alpha(q) = 1$". This condition assures that during the $\in$-phase, insertion of Mx into $\overline{\alpha}_x$ will not disrupt any inclusion of the type

$$\bigcup_{\alpha(p)=1} \overline{\alpha} \subseteq pow\Big(\bigcup_{\alpha(q)=1} \overline{\alpha}\Big),$$

for any powerset clause $p = pow(q)$ in Q.

To force equalities in place of the above inclusions the stabilization assignments

$$\overline{\beta} \leftarrow pow^*(\overline{\alpha}_1, \dots, \overline{\alpha}_l) \setminus \bigcup_{\{\alpha_1,\dots,\alpha_l\} \to \gamma} \overline{\gamma}$$

(where β is a "special" target of $\{\alpha_1, \dots, \alpha_l\}$) will take place each time a new variable x is processed in the ε-phase.

The initialization phase can be described as follows: since initial places are not restricted by any powerset clause, we can initialize them freely using a sufficiently large numbers of individuals. Moreover the empty set can be assigned to the place $\alpha_\emptyset$ (we are assuming that $\emptyset$ is a variable in Q designating the empty set). At this point proliferation of elements can start. This will continue until each place has been assigned at least one element. More specifically, for each P-node $\{\alpha_1, \dots, \alpha_l\}$, with $\overline{\alpha}_1, \dots, \overline{\alpha}_l$ nonemtpy, elements in $pow^*(\overline{\alpha}_1, \dots, \overline{\alpha}_l) \setminus \bigcup_{\{\alpha_1,\dots,\alpha_l\} \to \gamma} \overline{\gamma}$ are opportunely distributed among all its targets γ. [CFS2] states conditions which ensure that the initialization and subsequent stabilization phases can execute properly.

Note, also that it can be proved that if m is the number of distinct variables in Q, then Q is satisfiable if and only if Q has a model of rank at most $2^{2^m + m + 2} + 1$.

The same result can be proved also by allowing the singleton operator to appear together with the powerset operator. However in this case initialization and stabilization

must be carried out simultaneously (see [C]).

Acknowledgements.
Authors acknowledge support from the Italian C.N.R. and ENIDATA–BOLOGNA.

Bibliography

[FOS] FERRO A., OMODEO E.G. and SCHWARTZ J.T., "Decision procedures for elementary sublanguages of set theory. I. Multi-level syllogistic and some extensions", Comm. Pure Appl. Math., 33, (1980).

[BF] BREBAN M. and FERRO A., "Decision procedures for elementary sublanguages of set theory. III. Restricted classes of formulas involving the powerset operator and the general set union operator", Advances in Appl. Math., 5 (1984).

[F] FERRO A., "A note on the decidability of MLS extended by the powerset operator", Comm. Pure Appl. Math., 38, 1985.

[CFS1] CANTONE D., FERRO A., SCHWARTZ J.T., "Decision procedures for elementary sublanguages of set theory. V. Multilevel syllogistic extended by the general Unione Operator", Journal of Comp. Syst. Sci. (to appear).

[CFS2] CANTONE D., FERRO A., SCWARTZ J.T., "Decision procedures for elementary sublanguages of Set Theory. VI. Multilevel syllogistic extended by the powerset operator", Comm. Pure Appl. Math., 38, 1985.

[C] CANTONE D., "A decision procedure for a class of unquantified formulae of set theory involving the powerset and singleton operators", Ph. D. Dissertation, New York University, 1986.

[Om] OMODEO E.G., "Decidability and proof procedures for set theory with a choice operator", Ph. D. Dissertation, New York University, 1984.