Notes on Sets and Multisets

Consider a theory T_{set} of sets with signature

$$\Sigma_{\mathsf{set}} : \{\in,\ \subseteq,\ =,\ \subset,\ \cap,\ \cup,\ \setminus\}\ ,$$

where symbols are intended as follows:

- $e \in s$: e is a member of s;
- $s \subseteq t$: s is a subset of t;
- $s = t$: s and t are equal;
- $s \subset t$: s is a *strict* subset of t;
- $s \cap t$ is the intersection of s and t;
- $s \cup t$ is the union of s and t;
- $s \setminus t$, the set difference of s and t, is the set that includes all elements of s that are not members of t.

The language is redundant; for example, $\subset$ is definable from $\subseteq$ and $=$.

Let us encode an arbitrary Σ_{set}-formula as a Σ_{E}-formula (or a Σ_{A}-formula). To do so, simply consider the atoms:

- $e \in s$: let $s(\cdot)$ be a unary predicate; then replace $e \in s$ by $s(e)$;
- $s \subseteq t$: $\forall e.\ e \in s\ \rightarrow\ e \in t$, or in other words, $\forall e.\ s(e)\ \rightarrow\ t(e)$;
- $s = t$: $\forall e.\ s(e)\ \leftrightarrow\ t(e)$;
- $s \subset t$: $s \subseteq t\ \wedge\ s \neq t$;
- $u = s \cap t$: $\forall e.\ u(e)\ \leftrightarrow\ s(e) \wedge t(e)$;
- $u = s \cup t$: $\forall e.\ u(e)\ \leftrightarrow\ s(e) \vee t(e)$;
- $u = s \setminus t$: $\forall e.\ u(e)\ \leftrightarrow\ s(e) \wedge \neg t(e)$.

Atoms with complex terms can be written more simply via "flattening" (as in the Nelson-Oppen procedure); for example, write $s \cap (t \cap u)$ as $s \cap w \wedge w = t \cap u$. Then the encodability of an arbitrary Σ_{set}-formula into a Σ_{E}-formula follows by structural induction.

Satisfiability of the quantifier-free fragment of T_{set} is decidable: simply apply the decision procedure for T_{A} to the new formula.

Consider a theory T_{mset} of multisets with signature

$$\Sigma_{\mathsf{mset}} : \{C,\ \leq,\ =,\ <,\ \uplus,\ \cap,\ -\}\ .$$

Multisets can have multiple occurrences of elements, are the symbols are intended as follows:

- $C(s, e)$: the number of occurrences (the "count") of e in s;
- $s \leq t$: the count of each element of s is bounded by its count in t;
- $s = t$: element counts are the same in s and t;
- $s < t$: the count of each element of s is bounded by its count in t, and some element has a lower count;
- $s \uplus t$ is the multiset union, whose counts are the element-wise sums of counts in s and t;

- $s \cap t$ is the multiset intersection, whose counts are the element-wise minima of counts in s and t;
- $s - t$ is the multiset difference, whose counts are the element-wise maxima of 0 and the difference of counts in s and t.

Let us encode an arbitrary Σ_{mset}-formula as a $(\Sigma_{\mathsf{E}} \cup \Sigma_{\mathbb{Z}})$-formula (or a $(\Sigma_{\mathsf{A}} \cup \Sigma_{\mathbb{Z}})$-formula). A multiset is modeled by an uninterpreted function whose range is the nonnegative integers. Now consider the atoms:

- $C(s, e)$: let s be a unary function whose range is $\mathbb{N}$; then replace $C(s, e)$ by $s(e)$ and conjoin $\forall e.\ C(s, e) \geq 0$ to the formula;
- $s \leq t$: $\forall e.\ s(e) \leq t(e)$;
- $s = t$: $\forall e.\ s(e) = t(e)$;
- $s < t$: $s \leq t\ \wedge\ s \neq t$;
- $u = s \uplus t$: $\forall e.\ u(e) = s(e) + t(e)$;
- $u = s \cap t$: $\forall e.\ (s(e) < t(e) \wedge u(e) = s(e))\ \vee\ (s(e) \geq t(e) \wedge u(e) = t(e))$;
- $u = s - t$: $\forall e.\ (s(e) < t(e) \wedge u(e) = 0)\ \vee\ (s(e) \geq t(e) \wedge u(e) = s(e) - t(e))$.

As before, encodability follows by structural induction.

Exercise: Consider augmenting Σ_{mset} with set operations and a function *set* that maps a multiset m to a set s, where $e \in s$ iff $C(m, e) > 0$. Describe how to encode formulae of the augmented signature into $T_{\mathsf{E}} \cup T_{\mathbb{Z}}$.