Proof-theoretic Conservative Extension for HOL with Ad-hoc Overloading

Arve Gengelbach Tjark Weber

Uppsala University, Uppsala, Sweden

ICTAC, December 2, 2020

UPPSALA
UNIVERSITET

Motivation

What's in a definition?
That is, what new theorems can be derived with a definition?

Sufficient criterion for when a formula
is also provable without some definitions

Motivation (Example)

Declared constant: d_α
$T = \{\mathsf{d}_{\alpha\ \mathsf{list}} \equiv \ldots\}$

What new theorems can be derived with a definition of $\mathsf{d}_{\mathsf{bool}}$?

Assume $\mathsf{d}_{\mathsf{bool}}$ not occurring in the definition of $\mathsf{d}_{\alpha\ \mathsf{list}}$.
Any formula without $\mathsf{d}_{\mathsf{bool}}$ is provable from T
iff the formula is provable from $T \cup \{\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{True}\}$.
$T \cup \{\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{False}\}$.

Higher-Order Logic (HOL)

- Typed λ-calculus with rank 1 polymorphism
- With built-in types $\rightarrow$, bool
  and a built-in constant $=_{\alpha\rightarrow\alpha\rightarrow\mathsf{bool}}$
- Theories of type and constant definitions
- Constants may be overloaded
  Example: $+_{\alpha\rightarrow\alpha\rightarrow\alpha}$ may be overloaded at
  $+_{\mathsf{real}\rightarrow\mathsf{real}\rightarrow\mathsf{real}}$ and $+_{\mathsf{nat}\rightarrow\mathsf{nat}\rightarrow\mathsf{nat}}$

Definition: Proof-theoretic Conservativity

Theory extension by definition(s) is *proof-theoretically conservative* if for all theories $T \subseteq T'$
and for all formulae φ (in the language of T) we have:

$$T' \vdash \varphi \quad \Leftrightarrow \quad T \vdash \varphi$$

Initial Example:
$T = \{\mathsf{d}_{\alpha\ \mathsf{list}} \equiv \ldots\}$ and $T' = T \cup \{\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{True}\}$
have the same languages.
Problem:
$T' \vdash \mathsf{d}_{\mathsf{bool}} = \mathsf{True}$ and $T \nvdash \mathsf{d}_{\mathsf{bool}} = \mathsf{True}$.

Result:
Proof-theoretic Conservativity done right

In HOL with ad-hoc overloading, definitions are conservative in the following sense:

If a formula φ is independent[1] of symbols defined in $T' \setminus T$ then

$$T' \vdash \varphi \quad \Leftrightarrow \quad T \vdash \varphi$$

[1] expressed by closure of definitional dependencies

Example: Proof-theoretic Conservativity done right

$T = \{\mathsf{d}_{\alpha\ \mathsf{list}} \equiv \ldots\}$ and $T' = T \cup \{\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{True}\}$.
Assume $\mathsf{d}_{\mathsf{bool}}$ not occurring in the definition of $\mathsf{d}_{\alpha\ \mathsf{list}}$.

- Any formula φ without $\mathsf{d}_{\mathsf{bool}}$ is independent of $\mathsf{d}_{\mathsf{bool}}$, thus $T' \vdash \varphi \quad \Leftrightarrow \quad T \vdash \varphi$
- $T' \vdash \mathsf{c}_{\alpha\ \mathsf{list}} = \mathsf{d}_{\alpha\ \mathsf{list}} \quad \Leftrightarrow \quad T \vdash \mathsf{c}_{\alpha\ \mathsf{list}} = \mathsf{d}_{\alpha\ \mathsf{list}}$
- But we can **not** prove
  $T' \vdash \mathsf{c}_{\mathsf{bool}} = \mathsf{d}_{\mathsf{bool}} \quad \Leftrightarrow \quad T \vdash \mathsf{c}_{\mathsf{bool}} = \mathsf{d}_{\mathsf{bool}}$

Proving Proof-theoretic Conservativity

- Generalise semantics
  Relaxed interpretation of function types: $[\tau \to \sigma] \subseteq [\tau] \to [\sigma]$
  Only interpret type-variable free constants and types.
- HOL with ad-hoc overloading is sound and complete (Andrews/Henkin)
- Prove model-theoretic conservativity to obtain proof-theoretic conservativity for theories of definitions

Relevance

- Foundation of Isabelle/HOL:
  Consistency; Definitions are abbreviations
- Practical: Ignore unrelated definitions in proof-search

Related Work [Kunčar and Popescu, POPL 2017]

- Any theory of definitions T' is a proof-theoretically conservative extension of the *initial* theory.
- *Meta-safety*: Definitions in a formula can be unfolded, resulting in a logically equivalent formula.
- We can recover this result (not meta-safety).

What's in a definition?

All the definitions that are neither implicitly nor explicitly relevant to a formula are irrelevant to the provabilty of the formula.

Dependency Relation

[Kunčar & Popescu 2015, Åman Pohjola & Gengelbach 2020]

Track a definition's dependencies (to disallow cyclic definitions).

- $u \equiv t$, and $v \in t^\bullet \cup t^\circ$ then $u \rightsquigarrow v$
  Example: $2\mathbb{N} \rightsquigarrow \mathbb{N}$, $2\mathbb{N} \rightsquigarrow \mathsf{even}_{\mathbb{N}\to\mathsf{bool}}$
  (from a definition $2\mathbb{N} \equiv \mathsf{even}_{\mathbb{N}\to\mathsf{bool}}$)
- $\mathsf{c}_\sigma \rightsquigarrow v$ for $v \in \sigma^\bullet$
  Example: $\mathsf{map}_{(\alpha\to\beta)\to\alpha\ \mathsf{list}\to\beta\ \mathsf{list}} \rightsquigarrow v$
  for $v \in \{\alpha,\ \beta,\ \alpha\ \mathsf{list},\ \beta\ \mathsf{list}\}$
- $(\sigma_1, \ldots, \sigma_n)k \rightsquigarrow \sigma_i$ for a type constructor k
  Reason/Example: $\tau \to \sigma \rightsquigarrow \tau$ and $\tau \to \sigma \rightsquigarrow \sigma$

Symbol-independent Fragment [Gengelbach & Weber 2017]

The U-independent fragment[2] is

$$F_U := \mathsf{Symb} \setminus \{x | \exists u \in U, \rho.\ x \rightsquigarrow^{\downarrow *} \rho(u)\}.$$

F_U contains all symbols that are not depending on any instance from U.

Typically, $U =$ set of symbols defined by definitions.

Example: $\mathsf{c}_\alpha \equiv \mathsf{d}_\alpha$, $\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{True}$

$\mathsf{c}_{\mathsf{bool}}, \mathsf{d}_{\mathsf{bool}} \notin F_{\{\mathsf{d}_{\mathsf{bool}}\}}$ by $\mathsf{c}_{\mathsf{bool}} \rightsquigarrow^{\downarrow *} \mathsf{d}_{\mathsf{bool}}$

$\mathsf{c}_\alpha\ \mathsf{list} \in F_{\{\mathsf{d}_{\mathsf{bool}}\}}$ by $\mathsf{c}_\alpha\ \mathsf{list} \not\rightsquigarrow^{\downarrow *} \mathsf{d}_{\mathsf{bool}}$

[2] $\cdot^\downarrow$ type-substitutive closure, $\cdot^*$ reflexive-transitive closure

Model-theoretic Conservative Extension

Let $T \subseteq T'$ be a *wellformed* definitional extension
and let $\mathcal{M}$ be a model of T.
There exists a model $\mathcal{M}'$ of T' such that $\mathcal{M}$ and $\mathcal{M}'$ interpret terms equally, that are independent of symbols defined in $T' \setminus T$.

Example: $T = \{\mathsf{c}_\alpha \equiv \mathsf{d}_\alpha\}$, $T' = T \cup \{\mathsf{d}_{\mathsf{bool}} \equiv \mathsf{True}\}$.
Any model $\mathcal{M}$ for T has an extension $\mathcal{M}'$ for T'.

$\mathsf{c}_{\mathsf{bool}}$ is dependent on $\mathsf{d}_{\mathsf{bool}}$, thus may have $\mathcal{M}(\mathsf{c}_{\mathsf{bool}}) \neq \mathcal{M}'(\mathsf{c}_{\mathsf{bool}})$.
But $\mathcal{M}(\mathsf{c}_{\alpha\ \mathsf{list}}) = \mathcal{M}'(\mathsf{c}_{\alpha\ \mathsf{list}})$.

Model-theoretic Conservativity implies Proof-theoretic Conservativity

Theorem: Let $T \subseteq T'$ be a *wellformed* definitional extension.
If φ independent from the symbols defined in $T' \setminus T$
and $T' \vdash \varphi$ then $T \vdash \varphi$.

Proof:
With completeness it suffices to prove: φ holds in all models of T.
For a model $\mathcal{M}$ of T model-theoretic conservativity gives a
model $\mathcal{M}'$ of T' such that $\mathcal{M}(\varphi) = \mathcal{M}'(\varphi)$.
From $T' \vdash \varphi$ soundness gives $\mathcal{M}'(\varphi) = \text{true}$, thus $\mathcal{M}(\varphi) = \text{true}$.

Lazy Ground Semantics [Åman Pohjola & Gengelbach 2020]

- φ_{bool} satisfied w.r.t. $[\cdot]$ iff
  for all ground type substitutions ρ
  and all variable assignments ξ_ρ: $[\varphi]_{\xi_\rho} = \mathsf{true}$
  - Earlier semantics: $[\rho(\varphi)]_{\xi_\rho} = \mathsf{true}$
    Problem: Term variables x_α and x_{bool} are distinct,
    but immediatly applying ρ equates these.
  - Lazy semantics
    For $\rho(\sigma)$ ground type, have $[\mathsf{c}_\sigma]_{\xi_\rho} = [\mathsf{c}_{\rho(\sigma)}]$ and
    $[x_\sigma]_{\xi_\rho} = \xi_\rho(x_\sigma)$ such that $\xi_\rho(x_\sigma) \in [\rho(\sigma)]$
- $\mathcal{M} \models T$ iff every $\varphi_{\mathsf{bool}} \in T$ is satisfied w.r.t. $\mathcal{M}$.