CS257: Introduction to Automated Reasoning

First-Order Theories

Stanford University

CENTAUR

Outline

- First-order Theory
- Satisfiability modulo Theories
- Examples of First-order Theories

After-class readings:

- CC: Chapter 3
- (Optional) Barrett, Clark, and Cesare Tinelli. "Satisfiability modulo theories." Handbook of model checking. Springer, Cham, 2018. 305-343.

* Some of the slides today are contributed by Clark Barrett.

Motivations

Consider the signature $\Sigma = \left\langle \Sigma^S, \Sigma^F \right\rangle$ for a fragment of set theory presented last time:

$$\Sigma^S = \{E, S\} \qquad \Sigma^F = \{\varnothing, \in\} \qquad sort(\varnothing) = S \qquad sort(\in) = \langle E, S, \text{Bool} \rangle$$

Variable v_e has sort E and variable v_s has sort S

Consider the Σ-formula $\forall v_e.\neg(v_e \in \varnothing)$. Is the formula valid?
Now consider the formula $\forall v_e.(v_e \in \varnothing)$. Is the formula satisfiable?
In practice, we often only care about satisfiability and validity with respect to a limited class of interpretations.

First-order theories

A **theory** $\mathcal{T}$ is a pair $(\Sigma, \mathcal{S})$, where:

- Σ is a signature, which we recall from Lecture 4 consists of a set Σ^S of **sorts** and a set Σ^F of function symbols.
- $\mathcal{S}$ is a class (in the sense of set theory) of Σ-**structures**.

A theory limits interpretations of Σ-formulas to with the structures in $\mathcal{S}$.

Example: the Theory of Real Arithmetics: $\mathcal{T}_{RA}$

$\Sigma RA^S = \{\mathsf{R}\}$, $\Sigma RA^F = \{+, -, *, \leq, =_{\mathsf{R}}, q_i \text{ for each rational number constant } i\}$

$\mathcal{S}$ is the class of structures that interprets R as the set of real numbers, and the function symbols in the usual way.

$\mathcal{T}$-interpretations

Given two signatures Σ and Ω, and two set of variables X and Y, where $\Sigma \subseteq \Omega$ (i.e., $\Sigma^S \subseteq \Omega^S$ and $\Sigma^F \subseteq \Omega^F$) and $X \subseteq Y$

Let $\mathcal{I}$ be an Ω-interpretation over Y. A **reduct of $\mathcal{I}$** to (Σ, X), denoted $\mathcal{I}^{\Sigma,X}$, is a Σ-interpretation over X obtained from $\mathcal{I}$ by restricting it to interpret only the symbols in Σ and the variables in X

Given a theory $\mathcal{T} := (\Sigma, \mathcal{S})$, a **$\mathcal{T}$-interpretation** is any Ω-interpretation $\mathcal{I}$ for some $\Omega \supseteq \Sigma$ such that $I^{\Sigma,\varnothing} \in \mathcal{S}$

Example: Consider again $\mathcal{T}_{RA}$, where $\Sigma^S_{RA} = \{\mathrm{R}\}$, $\Sigma^F_{RA} = \{+, -, *, \leq, =_{\mathrm{R}}, q_i\}$, $\mathcal{S}$: $dom(\mathrm{R}) = \mathbb{R}$, function symbols interpreted in the usual way. Suppose we have a set of variables $v_0, v_1, \ldots$

Are the following interpretations $\mathcal{T}_{RA}$-interpretations?

- $dom(\mathrm{R})$ is the rational numbers, functions in Σ^F_{RA} interpreted in the usual way
- $dom(\mathrm{R}) = \mathbb{R}$, functions in Σ^F_{RA} interpreted in the usual way, and $v_i^{\mathcal{I}} = 0$
- $dom(\mathrm{R}) = \mathbb{R}$, functions in Σ^F_{RA} interpreted in the usual way, $\varnothing^{\mathcal{I}} = \{\}$, and $v_i^{\mathcal{I}} = 0$

Note: This definition allow us to consider the satisfiability in a theory $\mathcal{T} := (\Sigma, \mathcal{S})$ of formulas that contain sorts or function symbols not in Σ. These symbols are **uninterpreted**.

$\mathcal{T}$-satisfiability, $\mathcal{T}$-validity

Given a theory $\mathcal{T} := (\Sigma, \mathcal{S})$, a formula α is **satisfiable modulo $\mathcal{T}$**, or **$\mathcal{T}$-satisfiable**, if it is satisfied by some $\mathcal{T}$-interpretation $\mathcal{I}$.

A set Γ of Σ-formulas **$\mathcal{T}$-entails** an Σ-formula α, written $\Gamma \models_{\mathcal{T}} \alpha$, iff every $\mathcal{T}$-interpretation that satisfies all formulas in Γ satisfies α as well.

An Σ-formula ϕ is **$\mathcal{T}$-valid**, written $\models_{\mathcal{T}} \phi$, iff $\varnothing \models_{\mathcal{T}} \phi$.

Example: Are the following Σ_{RA}-formulas $\mathcal{T}$-valid/$\mathcal{T}$-satisfiable?

- $((v_0 + v_1 \leq 1) \wedge (v_0 - v_1 \leq 2))$
- $\forall v_0.((v_0 + v_1 \leq 1) \vee (-v_0 - v_1 \leq -1))$
- $\forall v_0.\forall v_1.((v_0 + v_1 \leq 1) \wedge (-v_0 \leq -1) \wedge (-v_1 \leq -1))$

Exercise

Given a theory $\mathcal{T} := (\Sigma, \mathcal{S})$, a formula α is **satisfiable modulo $\mathcal{T}$**, or **$\mathcal{T}$-satisfiable**, if it is satisfied by some $\mathcal{T}$-interpretation $\mathcal{I}$.
A set Γ of Σ-formulas **$\mathcal{T}$-entails** an Σ-formula α, written $\Gamma \models_{\mathcal{T}} \alpha$, iff every $\mathcal{T}$-interpretation that satisfies all formulas in Γ satisfies α as well.
An Σ-formula ϕ is **$\mathcal{T}$-valid**, written $\models_{\mathcal{T}} \phi$, iff $\varnothing \models_{\mathcal{T}} \phi$.

Are the following statements true?

- Is a $\mathcal{T}$-valid formula always $\mathcal{T}$-satisfiable?
- Is a valid Σ-formula always $\mathcal{T}$-valid?
- Is a $\mathcal{T}$-valid formula always valid?

Submit your answers to

`https://pollev.com/andreww095`

Exercise: alternative definition of theory

In Chapter 3 of CC, a theory is defined by a signature Σ and a set of Σ-sentences $\mathcal{A}$ called **axioms**. We refer to this definition as **theory***.

In particular, a formula α is $\mathcal{T}$-**valid*** iff every interpretation $\mathcal{I}$ that satisfies $\mathcal{A}$ also satisfies α.

Theory* is a special case in our earlier definition of theory:

- given a theory* $\mathcal{T}^*$ defined by Σ and $\mathcal{A}$, we define a theory $\mathcal{T} := (\mathcal{T}, \mathcal{S})$, where $\mathcal{S}$ is the class of structures that satisfies $\mathcal{A}$.
- By definition, a formula α is $\mathcal{T}$-valid* iff it is $\mathcal{T}$-valid.

However, $\mathcal{T}^*$ is not general enough, because not every class of Σ-models can be characterized by a set of axioms (e.g., integer arithmetic).

Completeness of theories

A theory $\mathcal{T}$ is **complete** iff for every sentence α, either α or $\neg\alpha$ is $\mathcal{T}$-valid.

Examples:

- for theory $\mathcal{T} =: (\Sigma, \mathcal{S})$ where $\mathcal{S}$ has only one element, $\mathcal{T}$ is complete. Why?
- the theory of field, $\mathcal{T}f := (\Sigma f, \mathcal{S}_f)$, is not complete. In this case, $\mathcal{S}_f$ contains all structures that satisfies the basic axioms of fields. In particular the following sentence is true in some field but false in others:

$$1 + 1 = 0$$

Decidability

Given a set of Σ-formulas Γ, we say Γ is a **decidable** set of formulas, if there exists a terminating algorithm, which given a Σ-formula α, returns "yes" if $\alpha \in \Gamma$ and "no" otherwise.

Given a theory $\mathcal{T} := \langle \Sigma, \mathcal{S} \rangle$, let Γ be the set of $\mathcal{T}$-valid Σ-formulas.
We say $\mathcal{T}$ is **decidable** if Γ is a decidable set.

A **fragment** of a theory $\mathcal{T}$ is a syntactically-restricted subset of formulas in $\mathcal{T}$.

The **quantifier-free** fragment of $\mathcal{T}$ are $\mathcal{T}$-valid formulas without quantifiers.

Theory of Uninterpreted Functions: $\mathcal{T}_{EUF}$

Given a signature Σ with equalities, the most unrestricted theory would include the class of all Σ-models.

This family of theories parameterized by the signature, is known as the theory of **Equality with Uninterpreted Functions (EUF)** or the **empty theory**, since it imposes no restrictions on its models.

Satisfiability modulo $\mathcal{T}_{EUF}$ is undecidable.

However, satisfiability of conjunctions of $\mathcal{T}_{EUF}$-literals (i.e., an atomic formula or its negation) is decidable in polynomial time with the congruence closure algorithm (covered later).

Example: $f(a) = a \wedge g(a) \neq g(f(a))$

Theory of Real Arithmetics: $\mathcal{T}_{RA}$

$\Sigma^S = \{\mathsf{R}\}$

Equality: Yes

$\Sigma^F = \{+, -, *, \leq, q_i$ for each rational number constant $i\}$

$\mathcal{S}$ is the class of structures that interprets R as the set of real numbers, and the functions in the usual way ($sort(q_i) = \langle \mathsf{R} \rangle$).

Satisfiability modulo $\mathcal{T}_{RA}$ is decidable (worst-case doubly-exponential)

But, restricted classes of Σ-formulas can be efficiently decided:

Quantifier-free linear real arithmetic (LRA): $*$ can only appear if at least one of the two operands is a rational constant.

Theory of Integer Arithmetics: $\mathcal{T}_{IA}$

Equality: Yes

$\Sigma^S = \{Z\}$

$\Sigma^F = \{+, -, *, \leq, c_i$ for each integer number constant $i\}$

$\mathcal{S}$ is the class of structures that interprets Z as the set of integers numbers, and the functions in the usual way.

Satisfiability modulo $\mathcal{T}_{IA}$ is undecidable.

Satisfiability of quantifier-free Σ-formulas modulo $\mathcal{T}_{IA}$ is undecidable.

Linear integer arithmetic (LIA) (i.e., Presburger arithmetic) is decidable.

Theory of Array with Extensionality: $\mathcal{T}_A$

$\Sigma^S = \{A, I, E\}$ (for array, indices, elements)

Equality: Yes

$\Sigma^F = \{\text{read}, \text{write}\}$
, where $sort(\text{read}) = \langle A, I, E\rangle$ and $sort(\text{write}) = \langle A, I, E, A\rangle$

Useful for modelling memories or array data structures.

Let a, i, and v be variables of sort A, I, E, respectively.

Example 1: $\text{read}(\text{write}(a, i, v), i) = v$

"The value stored at position i of an array a to which we write v to position i is v"

Intuitively, is this formula valid/satisfiable/unsatisfiable modulo $\mathcal{T}_A$?

Example 2: $(\text{read}(a, i) = \text{read}(a', i)) \rightarrow (a = a')$

Intuitively, is this formula valid/satisfiable/unsatisfiable modulo $\mathcal{T}_A$?

Theory of Array with Extensionality: $\mathcal{T}_A$

$\mathcal{S}$ is the class of structures that satisfy the following axioms:

1. $\forall a. \forall i, \forall v, \text{read}(\text{write}(a, i, v), i) = v$
2. $\forall a. \forall i. \forall i'. \forall v. (i \neq i' \rightarrow \text{read}(\text{write}(a, i, v), i') = read(a, i'))$
3. $\forall a. \forall a'. ((\forall i. \text{read}(a, i) = \text{read}(a', i)) \rightarrow a = a')$

Note: 3 can be omitted to obtain a theory without extensionality.

Satisfiability modulo $\mathcal{T}_A$ is undecidable.

But there are several decidable fragments (**next lecture**).