Geometric Reasoning in Lean: from Algebraic Structures to Presheaves

Kenji Maillard[1] and Yiming Xu[2]

[1] Inria, Nantes, France
[2] Ludwig-Maximilian Universität, München, Germany

Abstract

Algebraic theories such as semigroups, monoids, rings or Heyting algebras can easily be described in dependent type theories, such as that of Lean, by packing together sorts, operations and equations. Abstractly, these theories should be interpretable in many different settings beyond bare types. Leveraging semantics of geometric logic in presheaf categories, i.e. categories of "varying sets", we explore the potential of interpreting such algebraic theories in these extended settings.

Motivation & Goals By packing together types, operations and (equational) proofs, dependent type theories provide direct representation of algebraic structures. For instance, the notion of semigroup is defined in Lean's mathlib library [2] as (a variant of)

```
class Semigroup (G : Type u) where
  * : G -> G -> G
  /-- Multiplication is associative -/
  protected mul_assoc : forall a b c : G, a * b * c = a * (b * c)
```

Algebraic structures also admit an alternative approach within dependent type theories by presenting them through equational theories. The benefits of this latter indirect point of view is that we can consider models of such theories in settings beyond that of bare types as carriers of sorts. In our work, we leverage geometric theories to represent such algebraic structures and use their well-studied categories of models in presheaf categories, e.g. categories of families of sets parametrized by some base category. Following [1], we formalize a notion of geometric theories $\mathcal{T}$ in Lean [3] and construct the category $\mathcal{T}\text{-Mod}(\mathcal{C})$ of $\mathcal{T}$-models, using the notation of [1, Definition 1.2.12], in any presheaf category $\text{Func}(\mathcal{C}^{\text{op}}, \mathcal{S}et)$, functorially with respect to $\mathcal{C}$. We then explore the correspondence between the direct and indirect presentations of algebraic structures, with the hope of being able to transfer direct proofs about the former to constructions applicable on models in any presheaf category. Our Lean formalization is available at https://github.com/kyoDralliam/model-theory-topos.

Geometric Logic, Geometric Theories With an eye towards interpretations in presheaf categories, we choose to work with geometric logic, a slight extension of the $\wedge, \vee, =, \exists$ fragment of first-order logic with arbitrarily large disjunctions $\bigvee_{i \in I} \phi_i$. This logic is both expressive enough to represent many theories of interest, in particular any equational, algebraic theory, and admits interpretations in large classes of categories, notably in any (pre)sheaf topos. We define a deep embedding of single-sorted geometric logic in Lean parametrized by a signature collecting operations and predicates of finitary arity. The formalization employs well-scoped syntax and parametrizes geometric formulas with a small universe $(U, (I_k)_{k \in U})$ to deal with arbitrary disjunction $\bigvee_{i \in I} \phi_i$. A geometric theory $\mathcal{T}$ consists of a signature $\Sigma_{\mathcal{T}}$ together with a collection of axioms $(\phi_i \vdash_{\vec{x}_i} \psi_i)_i$ provided by sequents of geometric formulas over $\Sigma_{\mathcal{T}}$.

Interpreting into Presheaf Categories We formalize the notion of model of a geometric theory $\mathcal{T}$ in presheaves over a base category $\mathcal{C}$. Concretely, it consists of a collection of models in $\mathcal{S}et$ indexed by the objects of $\mathcal{C}$ equipped with structure-preserving maps induced by the morphisms of $\mathcal{C}$. This formalization choice goes much beyond the specific case of $\mathcal{S}et$ but uses that most constructs are lifted pointwise from their counterpart $\mathcal{S}et$, a simplification that wouldn't be achievable in an arbitrary (Grothendieck) topos or a geometric category.

We establish the soundness of our interpretation of geometric logic, obtaining the following theorem by a straightforward induction:

```
theorem soundness {T : theory} (M:Mod T D) {n : RenCtx} (phi psi: fml T.sig n)
    (h:Hilbert.proof phi psi): InterpPsh.Str.model M.str (sequent.mk _ phi psi)
```

In words, given a model $\mathtt{M} \in \mathcal{T}\text{-Mod}(\mathcal{D})$, any proof $h : \phi \vdash_{x_1,\ldots,x_n} \psi$ induces a factorization $[\![\phi]\!]^{\mathtt{M}} \rightarrowtail [\![\psi]\!]^{\mathtt{M}}$ of the interpretation of formulas ϕ, ψ as subobjects of $\mathtt{M}^n$.

Rather than working with a single presheaf category each time in isolation, we establish the interaction between them: to transport interpretations of geometric formulas between presheaf categories, we formalize the construction of a functor $\mathcal{T}\text{-Mod}(\mathcal{D}) \to \mathcal{T}\text{-Mod}(\mathcal{C})$ from a functor $F : C \to D$ via pullback.

Connecting to Lean's Algebraic Structures We now turn to the interplay between Lean's algebraic structures (i.e. dependently typed records) and their encoding as geometric theories. We experiment with the simple examples of semigroups and monoids. For instance, the signature of semigroups consists of a single binary operation and no predicates, while the corresponding theory adds the associativity axiom as follows:

```
def semigroup_sig : monosig where
  ops := Unit
  arity_ops := fun \_ => 2
  preds := Empty
  arity_preds := Empty.rec

abbrev mul (t1 t2: tm semigroup_sig n)
    : tm semigroup_sig n :=
 .op () (fun i => [t1 , t2][i])
```

```
def assoc : sequent semigroup_sig where
  ctx := 3
  premise := .true
  concl := fml.eq (mul (.var 0) (mul (.var 1)
    (.var 2))) (mul (mul (.var 0) (.var 1))
    (.var 2))

def semigroup_thy : theory where
  sig := semigroup_sig
  axioms := [ assoc ]
```

As a sanity check, we prove that the category of models of semigroups over the trivial category with a single object is equivalent to Mathlib's category of semigroups.

Beyond relating algebraic structures with geometric theories, we also want to transport proofs. An interesting aspect of that process, is that even simple proofs about standard objects can turn into interesting properties once evaluated on non-trivial presheaf categories. We plan to investigate the specific case of monoids, building upon existing Lean proofs. We hope this example can inform us about the general recipe for extracting geometric proofs from proof scripts for other algebraic structures.

References

[1] Peter T. Johnstone. *Sketches of an Elephant: Volume 2*. Oxford University Press UK, Oxford, England, 2002.

[2] The mathlib Community. The lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, CPP 2020, page 367–381, New York, NY, USA, 2020. Association for Computing Machinery.

[3] Leonardo de Moura and Sebastian Ullrich. The lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28: 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings*, page 625–635, Berlin, Heidelberg, 2021. Springer-Verlag.