

ABSTRACTION BOUNDARIES AND SPEC DRIVEN DEVELOPMENT IN PURE MATHEMATICS

JOHAN COMMELIN AND ADAM TOPAZ

ABSTRACT. In this article we discuss how *abstraction boundaries* can help tame complexity in mathematical research with the help of an interactive theorem prover. While many of the ideas we present here have been used implicitly by mathematicians for some time, we argue that the use of an interactive theorem prover introduces additional qualitative benefits in the implementation of these ideas.

1. INTRODUCTION

Modern research in pure mathematics has a clear tendency toward increasing complexity. New striking mathematical results may involve complex proof techniques, deep mastery of a subfield within mathematics, or nontrivial input from several areas of mathematics. All of these play a role in increasing the inherent complexity of a piece of modern mathematics.

In many respects, such increasing complexity is an indication of *progress* in pure mathematics. However, with such complexity comes a significant increase in *cognitive load* for both readers and authors. It is now routine to see significant new papers in pure mathematics with one hundred pages or more. Similarly, the refereeing process of a complex mathematical result regularly takes multiple years. This state of affairs is clearly not sustainable.

Mathematicians have long been using the concept of abstraction in order to tame such complexities. For example, they often break down proofs into more manageable steps, such as lemmas, propositions, etc. They also frequently introduce definitions that capture a curated selection of properties satisfied by the objects they wish to study. More generally, it is often helpful to create “abstraction boundaries”, which separate the use of a mathematical object from its actual implementation. Staying within the boundaries of a given abstraction can help reduce cognitive load to some extent, but inherent complexities may nevertheless remain.

In this article we discuss how the use of abstraction boundaries, in conjunction with an *interactive theorem prover* (ITP), can facilitate complex interdisciplinary mathematical collaboration. We argue that the implementation of these ideas within an ITP can have significant qualitative benefits in further reducing cognitive load for both authors and readers. The application of this methodology

Received by the editors July 31, 2023.

2020 *Mathematics Subject Classification*. Primary 68V30, 68V20, 00A35.

The first author was supported by the Deutsche Forschungsgemeinschaft (DFG) grant CO 2587/1-1.

The second author was supported by NSERC discovery grant RGPIN-2019-04762.

was a key ingredient in the success of the *Liquid Tensor Experiment* (LTE), and we use examples from this project to illustrate our primary points.

Here is a brief outline of the paper. In §2, we review some aspects of condensed mathematics and LTE, focusing primarily on the topics which are relevant for this paper. In §3, we give a more precise description of what we mean by an *abstraction boundary* along with a discussion on sources of complexity within mathematics. In §4, we highlight some of the qualitative benefits offered by the use of abstraction boundaries alongside an ITP, while illustrating *spec driven development* with several examples related to LTE. The issue of aligning formal and informal mathematics using abductive reasoning is addressed in §5. Finally, we summarize and conclude our discussion in §6.

2. AN OVERVIEW OF THE LIQUID TENSOR EXPERIMENT

The *Liquid Tensor Experiment* (LTE) began with a challenge posed by Scholze [14], whose ultimate goal was to formally verify the proof of the *main theorem of liquid vector spaces*:

Theorem A (Clausen and Scholze). *Let $0 < p' < p \leq 1$ be real numbers and let S be a profinite set. Let V be a p -Banach space, considered as a condensed abelian group. Write $\mathcal{M}_{p'}(S)$ for the space of real-valued p' -measures on S , also considered as a condensed abelian group. Then $\mathrm{Ext}_{\mathrm{Cond}(\mathrm{Ab})}^i(\mathcal{M}_{p'}(S), V) = 0$ for all $i \geq 1$.*

This theorem is of utmost foundational importance; it is the technical core of the assertion that the real numbers, endowed with the collection of so-called *p -liquid vector spaces*, forms an *analytic ring* in the sense of Clausen–Scholze [2, 10, 11]. In [14], Scholze mentions that this might be his “most important theorem to date.” In this section we will only give a brief overview and motivation for the objects involved in this theorem along with a general summary of the experiment itself, as will be relevant for the remainder of this paper. We refer the reader to [14] and [11, Theorem 9.1] for more details.

2.1. Condensed and liquid mathematics. Condensed mathematics is a new foundational system which replaces topological spaces with objects called *condensed sets*, which are set-valued sheaves on the category Profinite of profinite sets with respect to the coherent Grothendieck topology. The power of this new theory shines particularly in algebraic contexts, whereas the classical approaches using point-set topology have several defects. For example, the category of topological abelian groups fails (quite badly) to be an abelian category, while the category of condensed abelian groups, defined either as sheaves of abelian groups on Profinite or equivalently as the category of abelian group objects in condensed sets, is an exceptionally nice abelian category. Condensed mathematics and its applications are still under development by Clausen and Scholze et al. [2, 10, 11], while a similar approach, albeit with a different overall focus, was introduced by Barwick and Haine in [1] under the name *pyknotic sets*.¹

Any topological space (resp., group, ring, module, etc.) can be considered as a condensed set (resp., group, ring, module, etc.) via a Yoneda-like construction. One may thus speak about condensed modules over condensed rings, which are the

¹The actual definition of a condensed set involves cardinality bounds, which we have omitted from the discussion, while the only distinction between condensed and pyknotic objects is in such set-theoretic issues.

condensed incarnation of topological modules over topological rings. But in order to capture the building blocks of *analytic geometry*, one needs a notion of *completeness* for such modules that behaves well algebraically. This is exactly what the notion of an *analytic ring* entails: it is a condensed ring with a notion of “completeness” for condensed modules which behaves well in a suitable *derived* sense.

Examples of analytic rings include \mathbb{Z}^\blacksquare , whose underlying condensed ring is \mathbb{Z} endowed the discrete topology, where the notion of completeness for modules is modeled on $\mathbb{Z}[S]^\blacksquare := \varprojlim_i \mathbb{Z}[S_i]$ for profinite sets $S = \varprojlim_i S_i$ expressed as cofiltered limits of finite sets S_i . A similar construction makes \mathbb{Z}_p , with the p -adic topology, into an analytic ring denoted $\mathbb{Z}_p^\blacksquare$. The field \mathbb{Q}_p can be considered as an analytic ring as well, essentially by base-changing from \mathbb{Z}_p .

The prototypical complete \mathbb{Q}_p -module in the last case turns out to be $\mathcal{M}(S, \mathbb{Q}_p)$, the space of (bounded) \mathbb{Q}_p -valued measures on a profinite set S . However, the analogous assertion *fails* if one replaces \mathbb{Q}_p with \mathbb{R} . The theory of *p -liquid vector spaces* solves this issue, by replacing the prototype $\mathcal{M}(S, \mathbb{R})$ by a *smaller* space $\mathcal{M}_{<p}(S, \mathbb{R})$.

Theorem A eventually translates to the assertion that \mathbb{R} , endowed with the collection of such p -liquid vector spaces, forms an analytic ring, thus providing a family of analytic rings parameterized by $0 < p \leq 1$, whose underlying condensed ring is \mathbb{R} . This result shows real and complex geometry can be studied in the context of analytic spaces alongside objects from algebraic and nonarchimedean geometry, see [2, 11]. At the same time, classical objects from functional analysis, such as Banach spaces (or more generally, p -Banach spaces for $0 < p \leq 1$), can fit in this framework as well, and thus many facets of functional analysis can now be studied *algebraically* with the necessary analysis essentially encapsulated in Theorem A. This work has already resulted in generalizations and new *algebraic* proofs of classical results from algebraic geometry which were originally analytic in nature, such as Serre’s GAGA. Many additional applications will surely come in the future.

At this point, it should be quite clear that the proof of Theorem A involves significant amounts of category theory, topos theory, and homological algebra, as well as analysis and topology. Surprisingly, a key step in the proof (see [12, 13]) brings combinatorics and (discrete) convex geometry into the picture as well.

2.2. Verifying the theorem on a computer. The LTE was completed on July 14, 2022 [4] with a complete formal verification of Theorem A using the Lean proof assistant [5] and its mathematics library `Mathlib` [3]. The formalization project—which was led by Commelin and Topaz—was a collaboration involving about a dozen mathematicians and several computer scientists. A detailed exposition of the project itself will appear elsewhere.

To conclude this section, we would like to highlight again that the proof of Theorem A involves many disparate subfields of pure mathematics, and at the same time, many relatively complicated mathematical objects. This significant source of complexity in the proof would have posed serious challenges in a traditional refereeing process. Indeed, this is one of the main reasons Scholze sought a formalization of the proof of Theorem A as explained in [14]. In [12], he writes:

The Lean Proof Assistant was really that: An assistant in navigating through the thick jungle that this proof is. Really, one key problem I had when I was trying to find this proof was that I was

essentially unable to keep all the objects in my “RAM”, and I think the same problem occurs when trying to read the proof.

The Lean ITP helped wrangle this complex proof, and, as we explain below, abstraction boundaries played a key role in the process.

3. MANAGING COMPLEXITY WITH ABSTRACTION BOUNDARIES

Complexity in mathematical research projects can come from various sources. We distinguish between two kinds. The first is *inherent complexity*, by which we mean the raw complexity of the Platonic ideal of a piece of mathematics. While this notion is hard to quantify, an approximate measure of inherent complexity could be the amount of time it takes a generic mathematician to understand the proof or definition in question starting from scratch.

The second form of complexity we call *accidental complexity*. This encompasses all sources of complexity that are not essential to the mathematics but rather imperfections that could be removed given enough time and energy.

Both inherent and accidental complexity contribute to the total complexity of a piece of mathematics. It is this total complexity which can overwhelm both authors and readers.

3.1. Inherent complexity. The main source of inherent complexity for a piece of mathematics comes from the scope of its prerequisites. This can be measured along the two axes of *breadth* and *depth*: a piece of mathematics may combine ingredients from many different areas of mathematics, or it may be a pinnacle within some subject, or a combination of both.

Note that the complexity of a theorem statement can be unrelated to the complexity of its proof. Fermat’s last theorem is an excellent example in this regard. The statement can be understood by anyone who understands the usual operations on natural numbers, but understanding of the proof requires mastery of several subfields of mathematics. Conversely, an elementary property of a complicated definition may have a simple proof. In this case, the *statement* has a high inherent complexity, but the proof does not.

Moreover, two different approaches to a proof of a given theorem may differ in their inherent complexities, even if both are written in an ideal manner. Thus, inherent complexity does not measure the minimal complexity of any possible proof or argument, but rather the complexity of a particular argument or proof strategy. An excellent proof strategy that is executed poorly may have low inherent complexity and high accidental complexity (see §3.2). Vice versa, a poorly chosen proof strategy that is executed perfectly may have high inherent complexity and low accidental complexity.

Complexity may also arise from the applications of proof *techniques* as opposed to applications of theorems. By this we are referring to proofs that apply variations of a common technique, but where it is unreasonable to capture all applicable variations in single general theorem statement. This happens frequently in parts of analysis, probability theory, and combinatorics, to name some examples. A specific example, provided to us by Peter Nelson, is “connectivity reduction,” where one proves theorems about combinatorial objects such as graphs, matroids, hypergraphs, etc., by reducing to certain “connected” objects. There are many variations

in the objects involved, such as the notion of connectivity, the way connected objects are glued together, etc. While such connectivity reductions are common, it does not seem reasonable to unify them in a precise theorem statement.

While it may appear as if this is more a form of accidental complexity, we argue that these proof techniques should be rather viewed as inherent complexity of the subfield of mathematics in which the proof takes place. Although it might be possible to rephrase such a proof technique as an actual theorem for the purpose of a given proof or mathematical argument, the theorem will not be sufficiently broadly applicable: only the proof strategy generalizes.

3.2. Accidental complexity. We illustrate accidental complexity by some well-known examples that occur as part of the reading and writing of mathematical texts.

Referring to the proof of a lemma, instead of its statement. This may occur in the rather common situation where the proof of Lemma X.Y.Z in reference [AB] actually proves something stronger than what the statement of Lemma X.Y.Z claims. In this case we might write something like, “We now see that our quasi-perfect gadget G is strongly separated by the proof of [Lemma X.Y.Z, AB]. Indeed, it is not hard to check that the proof shows G is not only separated, but even *strongly* separated”.

Using modifications of source material. Sometimes in the mathematical writing process we need a verbatim copy of some source material modulo a handful of completely algorithmic modifications. A mild example of this might be: “In the following proof, we will rely on a slightly modified version of the material in section 5 of [CD], but with the condition that the field k has characteristic 0 replaced by the condition that k is perfect.” This can often occur in combination with the preceding point, where a modification of a proof of some result applies while the statement itself does not.

Unwritten assumptions. Sometimes statements of theorems or lemmas can omit assumptions. In the good cases, these assumptions were listed at the beginning of the section (“in this section we assume that G is infinite”) or at the beginning of the paper in a section on notation. But there might also be assumptions that are common practice within certain communities (“all rings are commutative”) which might be hard to discover for the uninitiated reader.

Unclear definitions or terminology. A related point is that it may be hard to understand definitions or terminology. For example, definitions may change over time. In the early literature on scheme theory, all schemes were separated. In Whitney’s excellent book on geometric integration theory [17, p.15], a function is called smooth if it is C^1 . Such changes occur naturally over time, but can lead to significant confusion when studying older works, and even to a second degree when contemporary sources quote older works without properly warning about the change in terminology.

Keeping track of side conditions. When a lemma is applied, we verify the main assumptions of the lemma. But it would be burdensome and distracting to verify all the “trivial” side conditions (that some number is positive, or less than ϵ , or that some map is continuous, etc). At the same time, requiring the reader to keep track of these conditions implicitly increases the cognitive load.

Editing a paper. Making changes to a paper in its final stages can be tricky, because of all the possible ramifications downstream. Moving a lemma from section 2

to section 3 might require changing the statement, because the standing assumptions on X differ between the two sections. And yet it is all too easy to make such changes in a careless manner, possibly even shortly before publication and after a refereeing process took place.

All of the examples above, among others, are well-known pitfalls in mathematical writing, and are often warned against (see [7, 8, 15]). Nevertheless, these issues are still commonplace in the mathematical literature, increasing the accidental complexity of the work.

3.3. What is an abstraction boundary? An *abstraction boundary* is a formal separation between the implementation details of a concept from its extrinsic properties and behaviour. This involves introducing a *specification* (or *interface*) which describes how the object interacts with the outside environment. By using such specifications, one can work with the concepts they capture without relying on any actual implementation details.

This concept is prevalent in software engineering, with prominent examples including C header files, public methods in object-oriented programming, and type-classes in functional programming. In mathematics, abstraction boundaries play a fundamental role as well, with the mathematical concept of a *definition* being the principal example. For example, the definition of a “group” allows us to develop group theory, and apply results from it in the most varying circumstances, instead of only working with particular examples such as permutation groups and matrix groups. In this context, the specification is provided by the group axioms. Another example of abstraction boundaries in mathematics includes “black box” theorems, such as Cauchy’s residue theorem, the existence of a Néron model, the law of large numbers, Theorem A, etc. Such theorems can be effectively applied in particular situations without requiring an understanding of the argument.

3.4. Abstraction boundaries in interactive theorem provers. Interactive theorem provers (ITPs) enforce certain abstraction boundaries rigidly. While this can be restraining at times, it also relieves the user from cognitive load arising from several sources of complexity.

ITPs differ in the features that they offer, and certainly in the implementation details of those features. But most of this section applies to all modern ITPs, even though we use the Lean theorem prover as a running example. Lean is the system that we used for LTE, and it is the ITP that we are most familiar with.

The most well-known feature offered by ITPs is *proof checking*, often as an instance of *type checking*. This means that the ITP ensures that lemmas and theorems are always applied correctly, and that all side conditions are satisfied. Certain side conditions that are amenable to automation, such as continuity proofs, functoriality proofs, etc., can often be handled in a way that is transparent to the user.

Another feature that most ITPs offer is *proof irrelevance*.² This notion can be seen as an analogue of the common mathematical practice of treating certain theorem statements as black boxes. Essentially, this means that after an ITP

²ITPs built on a foundation of homotopy-type-theoretic nature will often not have proof irrelevance for *explicit reasons* arising from the foundational theory. The particular reasons for this choice are not relevant to this paper, and a discussion of homotopy type theory would lead us too far astray.

finishes checking the proof of a lemma or theorem, it remembers only one bit of information: that the proof is valid. Any constructions performed within the proof become inaccessible as soon as the proof is complete. If it is necessary to refer to such a construction later on, it must be factored out of the proof into a standalone definition. In doing so, the construction becomes an object in its own right, that can be named, referred to, and reasoned about. Conversely, if a construction is local to a proof, it implicitly comes with the strict promise that it will not be used outside of the proof.

4. SPEC DRIVEN DEVELOPMENT

Spec driven development is an approach for “doing mathematics” which aims to both use and capture the abstraction boundaries of the mathematical objects involved. For our purposes, by “doing mathematics” we mean the process of making mathematical objects such as definitions, constructions, etc., and proving theorems about these objects. As we explain below, the use of an ITP in conjunction with spec driven development provides substantial benefits in taming the various sources of complexity, particularly in complex and collaborative projects.

While this approach arose naturally during the LTE, it is quite commonplace, although *implicit*, in usual mathematical practice. For example, an informal proof may assume some properties up front with the promise of a justification in a later section. The main difference here is that organizing a proof in this way usually increases the accidental complexity and hence the cognitive load. For example, the reader may feel inclined to check that there is no circular reasoning in the final argument. The use of an ITP completely eliminates this source of complexity, and, as explained below, even *encourages* such uses of mathematical debt.

Spec driven development can be summarized as the following recursive process:

- (1) Isolate a desired mathematical definition, construction, theorem, etc., as a “target”.
- (2) When appropriate, isolate an initial *specification* (“spec” for short) of the target.
- (3) Break down the object and its specification into parts with lower complexity.
- (4) Repeat the above, with each new part acting as the target.

Furthermore, at each step, definitions and/or theorems, including the targets and/or their specs, may be *refactored* as necessary. Refactoring refers to the process of restructuring the implementation of an existing object without changing its external behaviour. As long as these implementation details and external properties of the object are properly separated by an abstraction boundary, other parts of the environment will not be affected by such a restructuring.

Spec driven development is also deeply intertwined with the notion of an abstraction boundary discussed in §3. In broad terms, the relationship between the two can be broken up into two interconnected categories:

- (1) Spec driven development uses abstraction boundaries in order to control complexity.
- (2) New abstraction boundaries arise naturally in the process of spec driven development.

Of course, these two categories are far from disjoint. For instance, as item (2) produces abstraction boundaries, they naturally feed into item (1). At the same time there are situations where spec driven development can straddle both categories.

In the following subsections we discuss the idea behind spec driven development in more detail by going through a series of examples based on constructions that were necessary in the LTE. We will present a few snippets of Lean code, although experience with the language or any other ITP will not be necessary. The precise meaning of the code we present and the mathematics involved will not be important, and are only meant to illustrate the implementation of spec driven development in an ITP. The only important concept needed from Lean’s syntax is that of a `sorry`; this is a keyword which should be thought of as a *placeholder* that needs to be filled in later. A `sorry` can be used as a placeholder for both *data* and *proofs*, which is crucial for this approach.³

4.1. Using and extending abstraction boundaries. The simplest use-case of spec driven development is in using previously established abstraction boundaries in order to reduce complexity for the overall objective at hand. As an example, consider the assertion that the category of condensed abelian groups is an abelian category, a property that is used in numerous places in LTE. In Lean, this might be initially formulated as follows:

```
instance : abelian (Condensed Ab) := sorry
```

In this example, the `sorry` appearing should be considered as our initial *target*: it is a unit of mathematical debt that must be accounted for at some point in the future. At this point, all collaborators can immediately start using the assertion that the category `Condensed Ab` is abelian, even before providing a proof of this fact. For example, we can now speak about kernels and cokernels of morphisms, exact sequences, etc. The abstraction boundary here is the notion of an *abelian category*.

There are instances where we wish to extend the boundary in some way, and at the basic level, this can be accomplished by adding additional *spec lemmas/theorems*. For instance, in the example above we may want to use the fact that morphisms of condensed abelian groups are morphisms of presheaves on `Profinite` taking values in `Ab`, and that sums of such morphisms are computed “object-wise”. In practice, this assertion would be true by definition, but since the actual definition has yet to be provided, we cannot formally rely on this. Instead, we add this property with a *spec lemma*, which becomes an additional target that can again be used immediately:

```
lemma val_app_add {F G : Condensed Ab} {X : Profiniteop}
  (f g : F → G) : (f + g).val.app X = f.val.app X + g.val.app X :=
  sorry
```

In the code above, if `f` is a morphism of condensed abelian groups, then `f.val.app X` denotes the corresponding morphism on sections over the profinite set `X`. Note that if we *use* the targets above at this point, then we are *forced* to remain within their associated abstraction boundary. We can’t rely on an implementation detail because there is no implementation yet!

In order to start repaying the existing debt, we can break down each target and its associated spec into smaller parts, and repeat the process until the individual

³In type-theoretic terms, a `sorry` can be used as a placeholder for an arbitrary *term* of an arbitrary type, including propositions.

targets can be handled directly. The initial target and its spec may thus become a collection of several parts, each of which becomes an additional target. In our example, recall that an abelian category is a preadditive category satisfying some additional conditions. After we do this for the abelian category instance in the example above, we are left with the following which summarizes all the current targets:

```
instance : preadditive (Condensed Ab) := sorry

instance : abelian (Condensed Ab) :=
{ normal_mono_of_mono := sorry,
  normal_epi_of_epi := sorry,
  has_finite_products := sorry,
  has_kernels := sorry,
  has_cokernels := sorry,
  ..(infer_instance : preadditive (Condensed Ab)) }

lemma add_val_app {F G : Condensed Ab} {X : Profiniteop}
  (f g : F → G) : (f + g).val.app X = f.val.app X + g.val.app X :=
sorry
```

At each step, a contributor may fill in some target(s) by replacing the `sorry` with an actual construction/proof. For example, one contributor may construct the preadditive structure on `Condensed Ab` in such a way that the spec lemma `add_val_app` above is true by definition, resulting in a reduction of debt:

```
instance : preadditive (Condensed Ab) := /- actual proof omitted -/

lemma add_val_app {F G : Condensed Ab} {X : Profiniteop}
  (f g : F → G) : (f + g).val.app X = f.val.app X + g.val.app X :=
rfl
```

At this point, the remaining debt in our running example consists of `normal_mono_of_mono`, `normal_epi_of_epi`, etc., which appear as components of the proof that `Condensed Ab` is an abelian category. These targets may again be handled directly, or further broken down as necessary.

4.2. Related targets and specs. In practice, the targets appearing in spec driven development may depend on others in nontrivial ways. To illustrate this, let us consider a more interesting example that came up during the LTE. In order to state Theorem A formally, we must be able to talk about Ext-groups, and a natural context for this construction is using an abelian category with enough projectives. For each natural number n , $\text{Ext}^n(-, -)$ is a bifunctor taking values in abelian groups, which is contravariant in the first variable. We may thus consider the following as our target:

```
-- The arrow '=>' is used to denote functors.
def Ext {A : Type*} [category A] [abelian A] [enough_projectives A] (n : ℕ
) :
  Aop ⇒ A ⇒ Ab :=
sorry
```

In this case, it makes sense to immediately introduce specifications describing the intended behavior of this definition. For example, we may want to ensure that Ext^0 is naturally isomorphic to `Hom`, and to speak about long exact sequences of

Ext-groups, say in the first variable. This can all be accomplished by adding the following:

```
-- `preadditive_yoneda.flip` is the bifunctor `X ↦ (Y ↦ Hom(X,Y))`,
-- where `Hom(X,Y)` is considered as an abelian group.
def Ext_zero_iso_Hom {A : Type*} [category A]
  [abelian A] [enough_projectives A] :
  (Ext 0 : Aop ⇒ A ⇒ Ab) ≅ preadditive_yoneda.flip :=
sorry

def d {A : Type*} [category A] [abelian A]
  [enough_projectives A] (n : ℕ) (X Y : A) :
  ((Ext n).obj (op X)).obj Y → ((Ext (n+1)).obj (op X)).obj Y :=
sorry

lemma Ext_LES {A : Type*} [category A]
  [abelian A] [enough_projectives A] (n : ℕ)
  (X1 X2 X3 Y : A) (f : X1 → X2) (g : X2 → X3)
  (ses : short_exact f g) :
  exact_seq Ab
  [((Ext n).map g.op).app Y, ((Ext n).map f.op).app Y, d n X1 Y] :=
sorry
```

The definitions `Ext_zero_iso_Hom`, `d` and spec lemma `Ext_LES` are additional targets, while `Ext_LES` relates to *all three* targets `Ext`, `Ext_zero_iso_Hom` and `d`.

As additional targets and/or preexisting definitions/theorems are introduced into the context, the situation can quickly become unwieldy. It is at this stage where the use of an ITP really shines. Essentially, the ITP keeps track of all the connections between the targets and existing definitions/theorems while ensuring consistency *at all times*. One enormous qualitative benefit of using an ITP with this approach is that a contributor working on a given target only needs to keep track of the immediately relevant portions of this dependency graph of targets/definitions/theorems, thereby substantially reducing cognitive load. This approach also facilitates collaboration, as it is no longer necessary to remember (or even understand) the context of the project as a whole when working on individual targets. In collaborative projects involving various mathematical subfields, each contributor may thus choose to only work on targets within their own area(s) of expertise.

4.3. Refactoring. Another major benefit of using abstraction boundaries, which is highlighted in the process of spec driven development, is that *refactoring* becomes easier in most cases. For instance, in the example from §4.1, the actual *implementation* of the fact that `Condensed Ab` is an abelian category is irrelevant, as long as one only needs results which hold true in an arbitrary abelian category. Changing the implementation details will therefore not interfere with the validity of such results.

Continuing with the example of Ext groups from §4.2, in LTE there was a point where we had a working definition of `Ext`, similar to the one described in §4.2, which had to be refactored to accommodate Ext-groups of (bounded above) *complexes* as opposed to just objects. In other words, at one point we introduced targets resembling the following:

```
def bounded_above_complex.Ext {A : Type*} [category A]
  [abelian A] [enough_projectives A] (n : ℤ) :
  (bounded_above_complex A)op ⇒ (bounded_above_complex A) ⇒ Ab :=
```

sorry

```
def bounded_above_complex.single_zero {A : Type*} [category A] [abelian A] :
  A ⇒ bounded_above_complex A :=
```

sorry

and *redefined* the `Ext` from §4.2 using these roughly as follows:

```
def Ext {A : Type*} [category A]
  [abelian A] [enough_projectives A] (n : ℕ) :
  Aop ⇒ A ⇒ Ab :=
(bounded_above_complex.single_zero A).op >>>
  (bounded_above_complex.single_zero A >>>
    (bounded_homotopy_category.Ext n).flip).flip
```

By relying on the given abstraction boundaries, as one is essentially *required* to do in the context of spec driven development, the process of resolving errors arising as a consequence of such a refactor becomes significantly easier. While this change in `Ext` may indeed cause certain other targets/specs/definitions/theorems to fail, this failure will be *localized* to those constructions and proofs that depend on the *implementation*. Crucially, *the ITP will catch all such errors!* Any results elsewhere in the project that rely on the abstraction boundary in question, such as proofs that only rely on the lemma `Ext_LES` from §4.2, should still work without any further modification.

5. ALIGNING THE FORMAL WITH THE INFORMAL

Every mathematician has to make an ongoing effort to align internal mental models of mathematics with the pen-and-paper representations found in the literature. This alignment goes both ways: mental models have to be updated and adjusted, and expositions of mathematics can be streamlined and improved. This alignment is an integral part of the creative process. It happens at the blackboard, during seminar talks, at the writing table, or during a walk in the park. Furthermore, it cannot be captured in formal rules: one cannot *prove* that a mental model aligns well with a pen-and-paper representation of some piece of mathematics.

5.1. The alignment problem. With the use of ITPs, this alignment issue is extended in a new direction: besides mental models and the pen-and-paper representations, there are the additional *digital* representations. To illustrate the alignment problem, consider the statement of Theorem A as it was formalized in LTE.

```
variables (p' p : ℝ≥0) [fact (0 < p')] [fact (p' < p)] [fact (p ≤ 1)]

theorem liquid_tensor_experiment (S : Profinite.{0}) (V : pBanach.{0} p) :
  ∀ i > 0, Ext i (M_{p'} S) V ≅ 0 :=
/- proof omitted -/
```

At a superficial level, this code snippet shows similarity to the statement of Theorem A. However, there is a possibility that the definition of the symbols `Profinite`, `pBanach`, `Ext`, and so forth, do not mean what they should mean according to established mathematical tradition. In other words: do their formal definitions align with the pen-and-paper representations and/or with the mental model of anyone looking at the source code of LTE? If the contributors to LTE were evil, the symbol `Ext` might be defined as the zero functor, trivializing the statement in the code snippet above. More innocently, there may be some subtle mistake in the definition of `Ext`

that implies the statement for completely uninteresting reasons. Even this is quite rare in practice, as a formal proof written by an honest human mathematician still relies on the mathematician’s mental models of the objects involved. Nevertheless, the alignment issue still needs to be addressed in order to efficiently *communicate* the results of a formalization to the greater mathematical community.

5.2. Abductive reasoning. Although there is no formal system for verifying alignment of different representations of mathematics, we argue that *abductive reasoning*, along with purposefully designed abstraction boundaries, can be fruitfully applied to provide convincing evidence of alignment. Abductive reasoning is the technical term for a mode of reasoning that seeks to explain a set of facts or observations in the simplest way possible, and is also called *inference to the best explanation*. Unlike with deductive reasoning, an abduction is not a formal logical consequence of its premises. The method is concisely captured by the colloquial “Duck Test”:

If it looks like a duck, swims like a duck,
and quacks like a duck, then it probably is a duck.

We provide a curated selection of examples and lemmas, somewhat analogous to “test suites” in software developments, that exhibit the ways in which a particular formal definition can be used. This is the input that allows others to abductively conclude that these formal symbols align with their mental models of the corresponding mathematical notions. In the terminology of §4, this input would be a carefully crafted “spec” for the formal representation of a mathematical concept.

The crucial aspect of such a spec is that it can be much shorter and easier to read than the actual definitions themselves. Indeed, typically the definitions will recursively unwind to a long list of prerequisite definitions.⁴ Furthermore, definitions might be made using general constructions that the reader is unfamiliar with.

For LTE we provided such specs for all the objects occurring in the main theorem of liquid vector spaces. We collected these specs in the subfolder `examples/` of the project; see [16] for a detailed discussion. For example, we show that the real numbers in Lean are a conditionally complete linearly ordered field, that $\text{Ext}^1(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$, and that p -Banach spaces V can be given a norm that satisfies $\|\lambda v\| = |\lambda|^p \|v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$. Such examples are a *reality check*, which collectively are meant to provide a large degree of confidence that the formal representations of mathematical concepts align with corresponding mental models and pen-and-paper representations.

Abductive reasoning is not unique to LTE or the use of ITPs, and many forms are often used in the mathematical practice at large. For example, the following common methods of establishing and/or strengthening belief in a mathematical argument or proof can be considered as instances of abductive reasoning:

- the ingredients and/or ideas seem strong enough to prove the claim;
- the claims hold up in well-chosen examples or numerical experiments;
- the proof is accepted by the community, including the experts in the field.

⁴The main result of LTE depends recursively on several thousand definitions. Part of these definitions are made in the project itself, but the majority is imported from Lean’s mathematical library `mathlib`.

With such methods, a reader may obtain some level of trust in the validity of a piece of mathematics without necessarily digging into every detail of the exposition. The reader must balance their desired level of trust with the amount of effort they are willing to exert.

The main differences between such forms of abductive reasoning and the strategy we put forth in LTE is in the scope to which it is applied. While the three examples above focus on deciding whether or not to believe the *proofs* of certain results, in the case of formalized mathematics, the focus shifts to merely the validity of the *definitions* involved. Thus, in formal projects such as LTE, the “surface area” to which abductive reasoning applies is comparatively small and carefully delineated. Once the reader is confident that the definitions align with their internal mental model, which can be done with abductive reasoning as discussed above, they may choose how far to dig into the associated proofs while remaining confident that the ITP ensures correctness. Overall, this has the effect of *separating* the exposition of the material from the issue of trusting its correctness.

6. CONCLUSION

In §2 we discussed two kinds of sources of complexity in mathematics: *accidental* and *inherent*. We also saw that ITPs can manage and remove accidental complexity, essentially *by default*. On the other hand, in §3 we saw that spec driven development and abstraction boundaries, in conjunction with an ITP, can also help in taming *inherent* complexity. As discussed in §3, a fundamental feature of spec driven development is the concept of *mathematical debt* which can be effectively tracked using an ITP. While this debt has to be repaid at some point, it is fundamentally *nonblocking* for the project as a whole.

The targets arising in spec driven development, along with other mathematical objects in question, are frequently interconnected in highly nontrivial ways. With the use of an ITP, contributors only need to keep track of the objects which are *locally* relevant, thereby reducing the cognitive load significantly. At the same time, these targets can be recursively broken into smaller components until they can be handled directly. This is particularly useful in collaborations as this allows for the work to be easily distributed among contributors with potentially varying areas of expertise.

Another key property of spec driven development is that reliance on abstraction boundaries becomes effectively *required* with this approach. As a byproduct, refactors generally become much easier. As the ITP ensures consistency within the project at all times, it also keeps track of the changes that must be made as a consequence of such refactors.

We feel that there is a lot of potential to develop tools to help facilitate spec driven development in pure mathematics, in conjunction with an ITP. Above we discussed how an ITP “keeps track” of the complicated dependency graph of objects (including targets) within a mathematical project. Right now, this is essentially done in the process of *type-checking*. However, a tool that does this in a more *explicit* manner, by identifying precisely the various objects associated with a given piece of mathematics within an ITP, would be even more beneficial as it would tell contributors *precisely* what they need to focus on when working on individual

targets. Patrick Massot’s blueprint software [9], which was used in LTE,⁵ can be seen an initial approximation to this idea.

Another tool that would be useful in projects such as LTE is a system that would keep track of the “weights” of targets. Essentially, we envision that each `sorry` in a given project would be tagged with a “weight” that is meant to act as a quantitative approximation for its complexity. Progress in the project as a whole can then be tracked by computing the total remaining weight. As the history of projects such as LTE is stored using a version control system (e.g., `git`), the evolution of the weights of targets (along with other metadata) could be mined and studied, and potentially used in machine learning applications.

In this article, we have painted a very positive picture about the current use of ITPs in pure mathematics. The formal verification of significant results, such as Hales’s proof of the Kepler conjecture [6] and the Liquid Tensor Experiment, have clearly demonstrated that ITPs are valuable tools for the verification of hard proofs. To summarize, we discussed how the use of abstraction boundaries and spec driven development within the rigid framework of an ITP helps

- facilitate collaboration amongst mathematicians with different expertise;
- manage accidental complexity and thus reduce cognitive load;
- provide a guarantee of formal correctness of proofs, thereby reducing the amount of material that needs to be trusted.

We envision that these methods can also provide similar benefits in original research.

However, we acknowledge that there are still a number of high costs that come with the use of ITPs and related tools that pose a barrier to wider adoption within mathematics. For example, they currently have a steep learning curve, and the difference between an informal piece of mathematical exposition and its formalization is still too large. Nevertheless, ITPs and their formalized mathematics libraries have seen tremendous improvements in the last few years, while the successes in the field have helped accelerate progress. We are also optimistic that advances in metaprogramming, artificial intelligence, and type theory itself, will help resolve some of the remaining issues in the near future.

ACKNOWLEDGMENTS

We would like to warmly thank all of the contributors to the Liquid Tensor Experiment for the amazing collaboration and excellent discussions. We thank the Lean community at large for providing the technological and social environments which made this work possible. Many thanks to the referees of this article and the editors of this volume for their comments which helped improve this paper in various ways. Finally, we would like to thank Akshay Venkatesh for proposing such an exciting topic for his Fields Medal Symposium.

REFERENCES

- [1] Clark Barwick and Peter Haine, *Pyknotic objects, I. Basic notions*, Preprint, [arXiv:1904.09966](https://arxiv.org/abs/1904.09966), (2019).
- [2] Dustin Clausen and Peter Scholze, *Condensed Mathematics and Complex Geometry*, (2022), <https://people.mpim-bonn.mpg.de/scholze/Complex.pdf>.

⁵<https://leanprover-community.github.io/liquid/>

- [3] The Mathlib Community, *The Lean mathematical library*, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (New York, NY, USA), CPP 2020, Association for Computing Machinery, 2020, pp. 367–381.
- [4] The Mathlib Community, *Completion of the Liquid Tensor Experiment*, Blog Post, 2022, <https://leanprover-community.github.io/blog/posts/lte-final/>.
- [5] Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer, *The lean theorem prover (system description)*, Automated deduction—CADE 25, Lecture Notes in Comput. Sci., vol. 9195, Springer, Cham, 2015, pp. 378–388, DOI 10.1007/978-3-319-21401-6_26. MR3446905
- [6] Thomas Hales, Mark Adams, Gertrud Bauer, Tat Dat Dang, John Harrison, Le Truong Hoang, Cezary Kaliszyk, Victor Magron, Sean McLaughlin, Tat Thang Nguyen, Quang Truong Nguyen, Tobias Nipkow, Steven Obua, Joseph Pleso, Jason Rute, Alexey Solovyev, Thi Hoai An Ta, Nam Trung Tran, Thi Diep Trieu, Josef Urban, Ky Vu, and Roland Zumkeller, *A formal proof of the Kepler conjecture*, Forum Math. Pi **5** (2017), e2, 29, DOI 10.1017/fmp.2017.1. MR3659768
- [7] P. R. Halmos, *How to write mathematics*, Enseign. Math. (2) **16** (1970), 123–152. MR277319
- [8] G. W. Bond, *Eulogy: Donald E. Knuth* (Latin), Bull. London Math. Soc. **21** (1989), no. 1, 110–112, DOI 10.1112/blms/21.1.110. With an English paraphrase. MR967799
- [9] Patrick Massot, Git Repository, <https://github.com/PatrickMassot/leanblueprint>.
- [10] Peter Scholze, *Lecture notes on Condensed Mathematics*, (2019), <https://www.math.uni-bonn.de/people/scholze/Condensed.pdf>.
- [11] Peter Scholze, *Lectures on Analytic Geometry*, (2019), <https://www.math.uni-bonn.de/people/scholze/Analytic.pdf>.
- [12] Peter Scholze, *Half a year of the Liquid Tensor Experiment: Amazing developments*, Blog Post, 2021, <https://xenaproject.wordpress.com/2021/06/05/half-a-year-of-the-liquid-tensor-experiment-amazing-developments/>.
- [13] Peter Scholze, *Nonconvexity and discretization*, MathOverflow, 2021, <https://mathoverflow.net/q/386796>.
- [14] Peter Scholze, *Liquid tensor experiment*, Exp. Math. **31** (2022), no. 2, 349–354, DOI 10.1080/10586458.2021.1926016. MR4458116
- [15] Jean-Pierre Serre, *How to write mathematics badly*, 2007, <https://www.youtube.com/watch?v=ECQyFzzBH1o>.
- [16] Adam Topaz, *Definitions in the liquid tensor experiment*, Blog Post, 2022, <https://leanprover-community.github.io/blog/posts/lte-examples/>.
- [17] Hassler Whitney, *Geometric integration theory*, Princeton University Press, Princeton, NJ, 1957. MR87148

MATHEMATISCH INSTITUT, UTRECHT UNIVERSITY, BUDAPESTLAAN 6, 3584CD UTRECHT, NEDERLAND

Email address: j.m.commelin@uu.nl

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF ALBERTA, 632 CENTRAL ACADEMIC BUILDING, EDMONTON AB T6G 2G1, CANADA

Email address: topaz@ualberta.ca