(Deep) Induction Rules for GADTs

Patricia Johann  
johannp@appstate.edu  
Appalachian State University  
USA

Enrico Ghiorzi  
ghiorzie@appstate.edu  
Appalachian State University  
USA

defined inductively, and almost all non-trivial proofs of their properties are either proved by induction outright or rely on lemmas that are. Every time a new inductive type is declared in Coq an induction rule is automatically generated for it.

The inductive data types handled by Coq include (possibly mutually inductive) polynomial algebraic data types (ADTs), and the induction rules Coq generates for them are the expected ones for standard structural induction. However, as discussed in [13], it has long been understood that these rules are too weak to be genuinely useful for deep ADTs.1 The following data type of rose trees, here coded in Agda and defined in terms of the standard type List of lists (see Section 2), is a deep ADT:

data Rose : Set → Set where  
empty : ∀ {A : Set} → Rose A  
node : ∀ {A : Set} → A → List (Rose A) → Rose A

The induction rule Coq automatically generates for (the analogous Coq definition of) rose trees is

∀ (A : Set) (P : Rose A → Set) →  
P empty →  
(∀ (a : A) (ts : List (Rose A)) → P (node a ts)) →  
∀ (x : Rose A) → P x

Unfortunately, this is neither the induction rule we intuitively expect, nor is it expressive enough to prove even basic properties of rose trees that ought to be amenable to inductive proof. What is needed here is an enhanced notion of induction that, when specialized to rose trees, will propagate the predicate P through the outer list structure and to the rose trees sitting inside node’s list argument. More generally, this enhanced notion of induction should traverse all of the layers present in a data structure, propagating suitable predicates to all of the data it contains. With data types becoming ever more advanced, and with deeply structured types becoming increasingly ubiquitous in formalizations, such an enhanced notion of induction is essential if proof assistants are to be able to automatically generate genuinely useful induction rules for data types that go beyond traditional ADTs. These include not just deep ADTs, but also (truly2) nested types [3], generalized algebraic data types (GADTs) [4, 14, 22, 27], more richly indexed families [5],

1An ADT/nested type/GADT is deep if it is (possibly mutually inductively) defined in terms of other ADTs/nested types/GADTs (including, possibly, itself).

2A truly nested type is a nested type that is defined over itself. The data type Bush in Section 2 provides a concrete example.

Abstract

Deep data types are those that are constructed from other data types, including, possibly, themselves. In this case, they are said to be truly nested. Deep induction is an extension of structural induction that traverses all of the structure in a deep data type, propagating predicates on its primitive data throughout the entire structure. Deep induction can be used to prove properties of nested types, including truly nested types, that cannot be proved via structural induction. In this paper we show how to extend deep induction to GADTs that are not truly nested GADTs. This opens the way to incorporating automatic generation of (deep) induction rules for them into proof assistants. We also show that the techniques developed in this paper do not suffice for extending deep induction to truly nested GADTs, so more sophisticated techniques are needed to derive deep induction rules for them.

CCS Concepts: • Theory of computation → Semantics and reasoning: Categorical semantics;

Keywords: GADTs, induction, proof assistants

1 Introduction

Induction is one of the most important techniques available for working with advanced data types, so it is both inevitable and unsurprising that it plays an essential role in modern proof assistants. In the proof assistant Coq [24], for example, functions and predicates over advanced types are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9182-5/22/01...

https://doi.org/10.1145/3497775.3503680

ACM Reference Format:
Patricia Johann and Enrico Ghiorzi. 2022. (Deep) Induction Rules for GADTs. In Proceedings of the 11th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP ’22), January 17–18, 2022, Philadelphia, PA, USA. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3497775.3503680
and deep variants of all of these. A summary of the various classes of data types considered in this paper is given in Table 1.

| Data types         | Discussed in Sections | Examples               |
|--------------------|-----------------------|------------------------|
| ADTs               | 1                     | List, Rose             |
| Nested types       | 2                     | PTree                  |
| Truly nested types | 2                     | Bush                   |
| GADTs              | 3                     | Eq, Seq                |
| Truly nested GADTs | 3 and 6               | G in (12)              |

Table 1. Data types in this paper

Deep induction [13] is a generalization of structural induction that fits this bill exactly. Whereas structural induction rules induct over only the top-level structure of data, leaving any data internal to the top-level structure untouched, deep induction rules induct over all of the structured data present. The key idea is to parameterize induction rules not just over a predicate over the top-level data type being considered, but also over additional custom predicates on the types of primitive data they contain. These custom predicates are then lifted to predicates on any internal structures containing these data, and the resulting predicates on these internal structures are lifted to predicates on any internal structures containing structures at the previous level, and so on, until the internal structures at all levels of the data type definition, including the top level, have been so processed. Satisfaction of a predicate by the data at one level of a structure is then conditioned upon satisfaction of the appropriate predicates by all of the data at the preceding level.

Deep induction was shown in [13] to deliver induction rules appropriate to nested types, including ADTs. For the (deep) ADT of rose trees, for example, it gives the following genuinely useful induction rule:

\[
\forall (A : Set) \exists (P : Rose A \rightarrow Set) \forall (Q : A \rightarrow Set) \rightarrow P \emptyset \rightarrow
\forall (a : A) (ts : List (Rose A)) \rightarrow Q a \rightarrow (1)
\]

\[
\forall (x : Rose A) \rightarrow P (\text{node} a ts) \rightarrow
\forall (x : Rose A) \rightarrow P (\text{node} \ A q x \rightarrow P x)
\]

Here, List\(^a\) (resp., Rose\(^a\)) lifts its predicate argument P (resp., Q) on data of type Rose A (resp., A) to a predicate on data of type List (Rose A) (resp., Rose A) asserting that P (resp., Q) holds for every element of its list (resp., rose tree) argument.\(^3\) Deep induction was also shown in [13] to deliver the first-ever induction rules — structural or otherwise — for the Bush data type [3] and other truly nested types. Deep induction for ADTs and (truly) nested types is reviewed in Section 2.

This paper shows how to extend deep induction to proper GADTs, i.e., GADTs that are not nested types (and thus are not ADTs). Typical applications of such GADTs include generic programming, modeling programming languages via higher-order abstract syntax, maintaining invariants in data structures, and expressing constraints in embedded domain-specific languages. They have also been used to implement tagless interpreters [14, 18, 19] by trading the definition of a universal value domain for a direct specification of the property of being a value. Other applications are described in, e.g., [17, 20]. A constructor for a GADT G may, like a constructor for a nested type, take as arguments data whose types involve instances of G other than the one being defined. These can even include instances involving G itself. But if G is a proper GADT, then at least one of its constructors will also have a structured instance of G — albeit one not involving G itself — as its codomain. For example, the constructor pair for the GADT

\[
\text{data Seq : Set } \rightarrow \text{Set where}
\]

\[
\text{const : } \forall {A : \text{Set}} \rightarrow A \rightarrow \text{Seq A}
\]

\[
\text{pair : } \forall {A B : \text{Set}} \rightarrow \text{Seq A } \rightarrow \text{Seq B } \rightarrow \text{Seq } (A \times B)
\]

(2)

of sequences\(^4\) only constructs sequences of data whose types are pair-structured, rather than sequences of arbitrary type, as does const. If one or more of the data constructors for a GADT G return structured instances of G, then the GADT will have two distinct, but equally natural, semantics: a functorial semantics interpreting it as a left Kan extension [15], and a parametric semantics interpreting it as the interpretation of its Church encoding [1, 26]. As detailed in [10], a key difference in the two semantics is that the former views GADTs as their functorial completions [12], and thus as containing more data than just those expressible in syntax. By contrast, the latter views them as what might be called syntax-only GADTs. Fortunately, these two views of GADTs coincide for those GADTs that are ADTs or (other, including truly) nested types. However, both they and their attendant properties differ greatly for proper GADTs. In fact, the functorial and parametric semantics for proper GADTs are sufficiently disparate that, by contrast with the semantics customarily given for ADTs and nested types [2, 7, 11], it is not at all clear how to define a functorial parametric semantics for GADTs [10].

This observation seems, at first, to be a death knell for the prospect of extending deep induction to GADTs. Indeed, induction can be seen as unary parametricity, so GADTs...
viewed as their functorial completions do not obviously support induction rules. This makes sense intuitively: induction is a syntactic proof technique, so it may not be possible to use it to prove properties of those elements of a GADT’s functorial completion that are not expressible in syntax. All is not lost, however. As we show below, the syntax-only view of GADTs determined by their Church encodings does support induction rules — including deep induction rules — for GADTs. Indeed, this paper gives the first-ever deep induction rules for proper GADTs. But it actually delivers far more: it gives a general framework for deriving deep induction rules for GADTs that can be instantiated to particular GADTs of interest. This framework can serve as a basis for extending modern proof assistants’ automatic generation of structural induction rules for ADTs to automatic generation of deep induction rules for GADTs. In addition, as for ADTs and nested types, the structural induction rule for any GADT can be recovered from its deep induction rule by taking the custom predicates in its deep induction rule to be constantly True-valued (i.e., constantly \( \top \)-valued) predicates.

Significantly, deep induction rules for GADTs cannot be derived by somehow extending the approach of [13] to syntax-only GADTs. Indeed, the approach taken there makes crucial use of the functoriality of data types’ interpretations from [12], and functoriality is precisely what interpreting GADTs as the interpretations of their Church encodings fails to deliver; see [9] for a discussion of why Seq, e.g., is not functorial. Our approach is to instead first give a predicate lifting styled after those of [13], together with a (deep) induction rule, for the simplest — and arguably most important — GADT, namely the equality GADT (4). We then derive the deep induction rule for a more complex GADT G by i) using the equality GADT to represent \( G \) as its so-called Henry Ford encoding [4, 8, 16, 21, 22], and ii) using the predicate liftings for the equality GADT and the other GADTs appearing in the definition of \( G \) to appropriately thread the custom predicates for the primitive types appearing in \( G \) throughout \( G \)’s structure. This two-step process delivers deep induction rules for a very general class of GADTs. To illustrate, we introduce a series of increasingly complex GADTs as running examples in Section 3 and derive a deep induction rule for each of them in Section 4. In particular, we derive the deep induction rules for the equality data type in Section 4.1 and the Seq data type in (2) in Section 4.2. We present our general framework for deriving (deep) induction rules for GADTs in Section 5, and observe that the derivations in Section 4 are all instances of it. In Section 6 we show that, by contrast with truly nested types, which do have a functorial semantics, syntax-only GADTs’ lack of functoriality means that it is not clear how to extend induction — deep or otherwise — to truly nested GADTs, i.e., to proper GADTs whose recursive occurrences appear below themselves. This does not appear to be much of a restriction, however, since truly nested GADTs do not, to our knowledge, appear in practice or in the literature. Section 7 comprises a case study in using deep induction. All of the deep induction rules appearing in this paper have been derived by instantiating our general framework. Our Agda implementation of them is available at https://cs.appstate.edu/~johannp/CPP22Code.html.

Additional Related Work Various techniques for deriving induction rules for data types that go beyond ADTs have been studied. For example, Fu and Selinger [6] show, via examples, how to derive induction rules for arbitrary nested types. Unfortunately, however, their technique is rather ad hoc, so is unclear how to generalize it to nested types other than the specific ones studied there. Moreover, [6] actually derives induction rules for data types related to the original nested types rather than for the original nested types themselves, and it is unclear whether or not the derived rules are sufficiently expressive to prove all results about the original nested types that we would expect to be provable by induction. This latter point echoes the issue with Coq-derived induction rule for rose trees raised in Section 1, which has the unfortunate effect of forcing users to manually write induction (and other) rules for such types for use in that system. Tassi [23] derives induction rules for data type definitions in Coq using unary parametricity. His technique seems to be essentially equivalent to that of [12] for nested types, although he does not permit true nesting. More recently, Ullrich [25] has implemented a plugin in MetaCoq to generate induction rules for nested types. This plugin is also based on unary parametricity, and true nesting still is not permitted. As far as we know no attempts have been made to extend either implementation to truly nested types or to proper GADTs or their deep variants. Other systems, including Isabelle and Lean, also derive induction rules for data types that go beyond ADTs. But we know of no work other than that reported here that specifically addresses induction rules for the (deep) GADTs considered in this paper.

2 Deep Induction for ADTs and Nested Types
A structural induction rule for a data type allows us to prove that if a predicate holds for every element inductively produced by the data type’s constructors then it holds for every element of the data type. In this paper, we are interested in induction rules for proof-relevant predicates. A proof-relevant
As in Coq’s induction rule for rose trees, the data inside a structural induction rule for the list type is parameterized over a custom polymorphic predicates but otherwise follows the familiar deep induction rule for lists.

As in Coq’s induction rule for rose trees, the data inside a structure of type List is treated monolithically (i.e., is ignored) by this structural induction rule. By contrast, the deep induction rule for lists is

\[
\forall (A : \text{Set})((P : \text{List } A \to \text{Set}) \to \text{P nil} \to (\forall (a : A)(\text{as : List } A) \to \text{P as} \to \text{P (cons as)})) \to \forall (\text{as : List } A) \to \text{P as}
\]

As in Coq’s induction rule for rose trees, the data inside a structure of type List is treated monolithically (i.e., is ignored) by this structural induction rule. By contrast, the deep induction rule for lists is parameterized over a custom polymorphic predicates rather than the monomorphic predicates appearing in structural induction rules for ADTs.

Structural induction can be extended to nested types, such as the following type of perfect trees:

\[
\forall (A : \text{Set})((P : \text{List } A \to \text{Set}) \to \text{P nil} \to (\forall (a : A)(\text{as : List } A) \to \text{Q a} \to \text{P as} \to \text{P (cons as)})) \to \forall (\text{as : List } A) \to \text{Q as} \to \text{P as}
\]

Perfect trees can be thought of as lists constrained to have lengths that are powers of 2. In the above code, the constructor \text{pnode} uses data of type \text{PTree } (A \times A) to construct data of type \text{PTree } A. Thus, it is clear that the instances of \text{PTree} at various indices cannot be defined independently, and that the entire inductive family of types must therefore be defined at once. This intertwinedness of the instances of nested types is reflected in their structural induction rules, which, as explained in [13], must necessarily involve polymorphic predicates rather than the monomorphic predicates appearing in structural induction rules for ADTs. The structural induction rule for perfect trees, for example, is

\[
\forall (P : \forall (A : \text{Set}) \to \text{PTree } A \to \text{Set}) \to (\forall (A : \text{Set})(a : A) \to \text{P A (pleaf a)}) \to (\forall (A : \text{Set})(p : \text{PTree } (A \times A)) \to \text{P (A \times A) pp} \to \text{P A (pnode pp)}) \to \forall (A : \text{Set})(p : \text{PTree } A) \to \text{P A p}
\]

The deep induction rule for perfect trees similarly uses polymorphic predicates but otherwise follows the familiar pattern. It is given by the first expression in Figure 1. There, \text{Pair}^\wedge : \forall (A B : \text{Set}) \to (A \to \text{Set}) \to (B \to \text{Set}) \to A \times B \to Set lifts predicates \text{Qa} on data of type \text{A} and \text{Qb} on data of type \text{B} to a predicate on pairs of type \text{A} \times \text{B} in such a way that \text{Pair}^\wedge A B \text{QA Qb} (a, b) = \text{QA} a \times \text{QB} b. Similarly, \text{PTree}^\wedge : \forall (A : \text{Set}) \to (A \to \text{Set}) \to \text{PTree } A \to Set lifts a predicate on data of type \text{A} to a predicate on data of type \text{PTree } A asserting that \text{Q} holds for every element of type \text{A} contained in its perfect tree argument. A general definition of liftings for a robust class of GADTs including all those appearing in the literature is given in Section 5.

Using deep induction we can extend structural induction to truly nested types, i.e., to nested types whose recursive occurrences appear below themselves. The quintessential example of such a type is that of bushes:

\[
\begin{align*}
data \text{Bush} & : \text{Set} \to \text{Set} \\
\text{bnil} & : \forall (A : \text{Set}) \to \text{Bush } A \\
\text{bcons} & : \forall (A : \text{Set}) \to \text{A} \to \text{Bush } (\text{Bush } A) \to \text{Bush } A
\end{align*}
\]

Even defining a structural induction rule for bushes requires that we be able to lift the rule’s polymorphic predicate argument to Bush itself. This observation was, in fact, the original motivation for the development of deep induction in [13]. The deep induction rule for bushes is given by the second expression in Figure 1, where

\[
\begin{align*}
\text{Bush}^\wedge A & : \forall (A : \text{Set}) \to (A \to \text{Set}) \to \text{Bush } A \to \text{Set} \\
\text{Bush}^\wedge A \text{Q bn} & = \top \\
\text{Bush}^\wedge A \text{Q (bcons a bb)} & = (3) \text{Q a} \times \text{Bush}^\wedge A \text{Q} (\text{Bush } A) (\text{Bush}^\wedge A \text{Q} \text{bb})
\end{align*}
\]

We note that, as for ADTs, the structural induction rule for any (truly) nested type can be obtained as the special case of its deep induction rule in which the custom predicates are taken to be constantly \(\top\)-valued predicates. This instantiation ensures that the resulting induction rule only inspects the top-level structure of its argument, rather than the contents of that structure, which is exactly what structural induction should do.

Under some circumstances deep induction can be mimicked by hand-threading applications of structural induction through the layers of data comprising a deep data type. But this is not the case if, e.g., one or more of the custom predicates in the data type’s deep induction rule is not the characteristic function of an inductive data type.

### 3 (Deep) GADTs

While a data constructor for a nested type can take as arguments data whose types involve instances of that type at indices other than the one being defined, its return type must

---

To define truly nested types in Agda we must use the \text{NO_POSITIVITY_CHECK} flag, and to define functions over them we must use the \text{TERMINATING} flag. (Similar flags are required in Coq.) Although as programmers we know from the metatheory in [12] that Bush is well-defined and the functions we define over them terminate, the flags are necessary because Agda fails to infer these facts. Analogous comments apply at several places below.
still be at the (variable) type instance being defined. For example, each of pleaf and pnode returns an element of type PTree A regardless of the instances of PTree appearing in the types of its arguments. GADTs relax this restriction, allowing their data constructors both to take as arguments and return as results data whose types involve instances other than the one being defined. That is, GADTs’ constructors’ return type instances can, like that of pair in (2), be structured. For every GADT in this paper, we require that the instance of the return type for each of its data constructors is a polynomial in that constructor’s type arguments.

GADTs are used in precisely those situations in which different behaviors at different instances of data types are desired. This is achieved by allowing the programmer to give the type signatures of the GADT’s data constructors independently, and then using pattern matching to force the desired type refinement. For example, the equality GADT

\[
\text{data Equal : Set} \rightarrow \text{Set} \rightarrow \text{Set where} \\
\text{refl} : \forall \{A : \text{Set}\} \rightarrow \text{Equal} A A
\]

(4)
is parameterized by two type indices, but it is only possible to construct data elements of type Equal A B if A and B are instantiated at the same type. If the types A and B are syntactically identical then the type Equal A B contains the single data element refl. It contains no data elements otherwise.

The importance of the equality GADT lies in the fact that we can understand other GADTs in terms of it. For example, the GADT Seq from (2) comprises constant sequences of data of any type A and sequences obtained by pairing the data in two already existing sequences. This GADT can be rewritten as its Henry Ford encoding \([4, 8, 16, 21, 22]\), which makes critical use of the equality GADT, as follows:

\[
\text{data Seq : Set} \rightarrow \text{Set where} \\
\text{const} : \forall \{A : \text{Set}\} \rightarrow A \rightarrow \text{Seq} A \\
\text{pair} : \forall \{A : \text{Set}\} \rightarrow (B C : \text{Set}) \rightarrow \text{Equal} A (B \times C) \rightarrow \text{Seq} B \rightarrow \text{Seq} C \rightarrow \text{Seq} A
\]

(5)

Here, the requirement that pair produce data at an instance of Seq that is a product type is replaced with the requirement that pair produce data at an instance of Seq that is equal to a product type. As we will see in Section 4, this encoding in terms of the equality GADT is key to deriving deep induction rules for GADTs.

Although Seq does not at first glance appear to be a deep GADT, when written as its Henry Ford encoding, it, like all GADTs, can be regarded as “deep over Equal”. By contrast, the GADT LTerm in Figure 2, which is inspired by \([28]\), is inherently deep. It encodes terms of a simply typed lambda calculus. More robust variations on LTerm are, of course, possible, but this variation is rich enough to illustrate all essential aspects of deep GADTs — and later, in Section 4.3, their deep induction rules — while still being small enough to ensure clarity of exposition.

Types are either booleans, arrow types, or list types. They are represented by the Henry Ford GADT LType in Figure 2. Terms are either variables, abstractions, applications, or lists of terms. They are similarly represented by the Henry Ford GADT LTerm. The type parameter for LTerm tracks the types of simply typed lambda calculus terms. For example, LTerm A is the type of simply typed lambda terms of type A. Variables are tagged with their types by the data constructors var and abs, whose LType arguments ensure that their type tags are legal types. This ensures that all lambda terms produced by var, abs, app, and list are well-typed. We will revisit these GADTs in Sections 4 and 7.

4 (Deep) Induction for GADTs

The equality constraints engendered by GADTs’ data constructors makes deriving (deep) induction rules for them more involved than for ADTs and other nested types. Nevertheless, we show in this section how to do so. We first illustrate the key components of our approach by deriving deep induction rules for the three specific GADTs introduced in Section 3. Then, in Section 5, we abstract these to a general framework that can be applied to any GADT that is not a truly nested GADT. As hinted above, the predicate lifting for the equality GADT plays a central role in deriving both structural and deep induction rules for more general GADTs.

4.1 (Deep) Induction for Equal

To define the (deep) induction rule for any GADT G we first need to define a predicate lifting that maps a predicate on a type A to a predicate on G A. Such a predicate lifting

\[
\text{Equal}^\land : \forall \{A B : \text{Set}\} \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \text{Equal} A B \rightarrow \text{Set}
\]
data LType : Set → Set where
  bool : ∀(A : Set) → (A → Set) → (A → Set) → LType A
  arr : ∀(A : Set) → ∀(B C : Set) → Equal A (B → C) → LType B → LType C → LType A
  list : ∀(A : Set) → ∀(B : Set) → Equal A (List B) → LType B → LType A

data LTerm : Set → Set where
  var : ∀(A : Set) → String → LType A → LTerm A
  abs : ∀(A : Set) → ∀(B : Set) → Equal A (B → C) → String → LType B → LTerm C → LTerm A
  app : ∀(A : Set) → ∀(B : Set) → LTerm(B(A)) → LTerm C → LTerm A
  list : ∀(A : Set) → ∀(B : Set) → Equal A (List B) → List (LTerm B) → LTerm A

Figure 2. The LType and LTerm data types

for Equal is defined by

\[ \text{Equal}^\wedge A Q Q' \text{refl} = \forall(a : A) \rightarrow \text{Equal}(Q, a)(Q'a) \]

It does exactly what we expect: it takes two predicates on the same type as input and is inhabited if they are extensionally equal. Next, we need to associate with each data constructor \( c \) of \( G \) an *induction hypothesis* asserting that, if the custom predicate arguments to a predicate \( P \) on \( G \) can be lifted to \( G \) itself, then \( c \) respects \( P \), i.e., \( c \) constructs data satisfying the instance of \( P \) at those custom predicates. The following induction hypothesis \( \text{dIndRefl} \) is thus associated with the \( \text{refl} \) constructor for \( \text{Equal} \):

\[
\lambda(P : \forall(A B : Set) \rightarrow (A \rightarrow Set) \rightarrow (B \rightarrow Set) \rightarrow Equal A B \rightarrow Set) \rightarrow
\forall(C : Set)(Q Q' : C \rightarrow Set) \rightarrow Equal^\wedge C C Q Q' \text{refl} \rightarrow PC C Q Q' \text{refl}
\]

The deep induction rule for \( G \) now states that, if all of \( G \)’s data constructors respect a predicate \( P \), then \( P \) is satisfied by every element of \( G \) to which the custom predicate arguments to \( P \) can be successfully lifted. The deep induction rule for \( \text{Equal} \) is thus

\[
\forall(P : \forall(A B : Set) \rightarrow (A \rightarrow Set) \rightarrow (B \rightarrow Set) \rightarrow Equal A B \rightarrow Set) \rightarrow \text{dIndRefl P} \rightarrow
\forall(A B : Set)(Q A : A \rightarrow Set)(Q B : B \rightarrow Set)(e : Equal A B) \rightarrow
Equal^\wedge A B Q A Q B e \rightarrow PA B Q A Q B e
\]

(6)

To prove that this rule is sound we must provide a witness \( \text{dIndEqual} \) inhabiting the type in (6). By pattern matching, we need only consider the case where \( A = B \) and \( e = \text{refl} \), so we can define \( \text{dIndEqual} \) by

\[
\text{dIndEqual P crefl A A QA Q' A refl liftE = crefl A QA Q' A liftE}
\]

To recover \( \text{Equal} \)’s structural induction rule

\[
\forall(Q : \forall(A B : Set) \rightarrow Equal A B \rightarrow Set) \rightarrow
(\forall(C : Set) \rightarrow QC C C refl) \rightarrow
\forall(A B : Set)(e : Equal A B) \rightarrow QA B e
\]

we define a term \( \text{indEqual} \) of the type in (7) by

\[
\text{dIndEqual Q srefl A B refl = dIndEqual P srefl' A B K_A^\wedge K_B^\wedge \text{refl} liftE.}
\]

Here, \( P : \forall(A B : Set) \rightarrow (A \rightarrow Set) \rightarrow (B \rightarrow Set) \rightarrow Equal A B \rightarrow Set \) is defined by \( PA B QA Q B e = QA B e, K_A^\wedge \) and \( K_B^\wedge \) are the constantly \( \top \)-valued predicates on \( A \) and \( B \), respectively, \( \text{sliftE} : Equal^\wedge A B K_A^\wedge K_B^\wedge \text{refl} \) is defined by

\[
\text{sliftE a =refl} : Equal \top \top
\]

for every \( a : A \), and

\[
\text{srefl'} : \forall(C : Set)(Q Q' : C \rightarrow Set) \rightarrow
Equal^\wedge C C Q C Q' \text{refl} \rightarrow QC C C Q Q' \text{refl}
\]

is defined by \( \text{srefl'} C Q Q' \text{liftE'} = \text{srefl C}. \) The structural induction rule for any GADT \( G \) that is not truly nested can similarly be recovered from its deep induction rule by instantiating every custom predicate by the appropriate constantly \( \top \)-valued predicate.

4.2 (Deep) Induction for \( \text{Seq} \)

To derive the deep induction rule for the GADT \( \text{Seq} \) we use its Henry Ford encoding from (5). We first define its predicate lifting

\[
\text{Seq}^\wedge : \forall(A : Set) \rightarrow (A \rightarrow Set) \rightarrow \text{Seq} A \rightarrow Set
\]

as in Figure 3. There, \( a : A, QA : B \rightarrow Set, QC : C \rightarrow Set, e : \text{Equal A (B \times C)} \rightarrow Set, \) \( s : \text{Seq} B, sc : \text{Seq} C, \) \( \exists x : F x \) \( : A \rightarrow Set \). The lifting \( \text{Seq}^\wedge \) is derived as in Section 5. Next, let \( \text{dIndConst} \) and \( \text{dIndPair} \) be the induction hypotheses associated with the constructors \( \text{const} \) and \( \text{pair} \), respectively. These are given in Figure 3 as well. Then the deep induction rule for \( \text{Seq} \) is given in the last two lines of Figure 3.

To prove that this rule is sound we provide a witness \( \text{dIndSeq} \) inhabiting the type in the last two lines of Figure 3 by

\[
\text{dIndSeq P cconst cpair A QA (const a) liftA = cconst A QA a liftA}
\]

and

\[
\text{dIndSeq P cconst cpair A QA (pair B C e s B s C) (QA, QC, liftE, liftB, liftC)} =
\text{cpair A B C QA QS s B s C e liftE p B p C}
\]
We therefore have (Deep) Induction Rules for GADTs

\[
\begin{align*}
\text{Seq}^\land A Q_A \text{ (const a)} &= Q_A a \\
\text{Seq}^\land A Q_A \text{ (pair B C e } s_B s_C) &= \exists [Q_B] \exists [Q_C] \text{ Equal}^\land A (B \times C) Q_A (\text{Pair}^\land B C Q_B Q_C) e \times \text{Seq}^\land B Q_B s_B \times \text{Seq}^\land C Q_C s_C
\end{align*}
\]

\[
\begin{align*}
d\text{IndConst} &= \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq} A \rightarrow \text{Set}) \rightarrow \\
& \quad \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (a : A) \rightarrow Q_A a \rightarrow P A Q_A \text{ (const a)}
\end{align*}
\]

\[
\begin{align*}
d\text{IndPair} &= \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq} A \rightarrow \text{Set}) \rightarrow \\
& \quad \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (Q_C : C \rightarrow \text{Set}) \\
& \quad \forall (Q_B : \text{Seq} B) (Q_C : \text{Seq} C) (e : \text{Equal} A (B \times C)) \rightarrow \\
& \quad \text{Equal}^\land A (B \times C) Q_A (\text{Pair}^\land B C Q_B Q_C) e \rightarrow P B Q_B s_B \rightarrow \\
& \quad P C Q_C s_C \rightarrow P A Q_A \text{ (pair } B C e s_B s_C)
\end{align*}
\]

\[
\begin{align*}
\forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq} A \rightarrow \text{Set}) \rightarrow d\text{IndConst} P \rightarrow d\text{IndPair} P \rightarrow \\
\forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (s_A : \text{Seq} A) \rightarrow \text{Seq}^\land A Q_A s_A \rightarrow P A Q_A s_A
\end{align*}
\]

**Figure 3.** Deep induction rule for Seq

In the first clause, \(a : A, Q_A : A \rightarrow \text{Set}\), and liftA : Seq^\land A Q_A (const a) = Q_A a. In the second clause we also have for LTerm are defined in Figure 4 following the general framework in Section 5. There,

\[
\begin{align*}
Q_B & : B \rightarrow \text{Set} \\
Q_C & : C \rightarrow \text{Set} \\
e & : \text{Equal} A (B \times C) \\
s_B & : \text{Seq} B \\
s_C & : \text{Seq} C \\
liftE & : \text{Equal}^\land A (B \times C) Q_A (\text{Pair}^\land B C Q_B Q_C) e \\
liftB & : \text{Seq}^\land B Q_B s_B \\
liftC & : \text{Seq}^\land C Q_C s_C
\end{align*}
\]

Together these give that

\[
(\text{Q}_B, \text{Q}_C, \text{liftE}, \text{liftB}, \text{liftC}) : \text{Seq}^\land A Q_A \text{ (pair } B C e s_B s_C)
\]

We therefore have

\[
\begin{align*}
p_B &= d\text{IndSeq} P \text{ cconst cpair } B Q_B s_B \text{ liftB} : P B Q_B s_B \\
p_C &= d\text{IndSeq} P \text{ cconst cpair } C Q_C s_C \text{ liftC} : P C Q_C s_C
\end{align*}
\]

and \(K^\land_{\text{Bool}}\) is the constantly \(\top\)-valued predicate on \(\text{Bool}\) and \(\text{List}^\land\) is the predicate lifting for lists from (1). Also,

\[
\begin{align*}
e & : \text{Equal} A \text{ Bool} \text{ in the first clause,} \\
e & : \text{Equal} A (B \rightarrow C) \text{ in the second and fifth clauses,} \\
e & : \text{Equal} A (\text{List } B) \text{ in the third clause,} \\
e & : \text{Equal} A (\text{List } B) \text{ in the seventh clause.}
\end{align*}
\]

With these liftings in hand we can define the induction hypotheses dIndVar, dIndAbs, dIndApp, and dIndList associated with LTerms’s data constructors. These are given in Figure 5. The deep induction rule for LTerm is thus

\[
\begin{align*}
\forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}) \rightarrow \\
d\text{IndVar} P \rightarrow d\text{IndAbs} P \rightarrow d\text{IndApp} P \rightarrow d\text{IndList} P \rightarrow \\
\forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (t_A : \text{LTerm} A) \rightarrow \\
\text{LTerm}^\land A Q_A t_A \rightarrow P A Q_A t_A
\end{align*}
\]

(8)

To prove that this rule is sound we define a witness dIndLTerm inhabiting the type in (8) as in Figure 6. There,
Figure 4. Predicate liftings for LType and LTerm

\[ \text{dIndVar} = \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}) \rightarrow \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (s : \text{String}) (T_A : \text{LType} A) \rightarrow \text{LType}^\wedge A Q_A T_A \rightarrow P A Q_A \left( \text{var} s T_A \right) \]

\[ \text{dIndAbs} = \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}) \rightarrow \forall (A B C : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (Q_C : C \rightarrow \text{Set}) (e : \text{Equal} A (B \rightarrow C)) (s : \text{String}) \rightarrow (T_B : \text{LType} B) (T_C : \text{LTerm} C) \rightarrow \text{Equal}^\wedge A (B \rightarrow C) Q_A \left( \text{Arr}^\wedge B C Q_B Q_C \right) e \rightarrow \text{LType}^\wedge B Q_B T_B \rightarrow P C Q_C T_C \rightarrow P \text{A Q}_A \left( \text{abs} B C e s T_B T_C \right) \]

\[ \text{dIndApp} = \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}) \rightarrow \forall (A B : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (Q_{\text{BA}} : \text{LTerm} (B \rightarrow A)) (T_B : \text{LTerm} B) \rightarrow P (B \rightarrow A) \left( \text{Arr}^\wedge B A Q_B Q_A \right) t_{\text{BA}} A \rightarrow P B Q_B T_B \rightarrow P \text{A Q}_A \left( \text{app} B T_{\text{BA}} B A \right) \]

\[ \text{dIndList} = \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}) \rightarrow \forall (A B : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (e : \text{Equal} A (\text{List} B)) (T_s : \text{List} (\text{LTerm} B)) \rightarrow \text{Equal}^\wedge A (\text{List} B) Q_A \left( \text{List}^\wedge B Q_B \right) e \rightarrow \text{List}^\wedge (\text{LTerm} B) (P B Q_B) T_s \rightarrow P \text{A Q}_A \left( \text{List} B e t s \right) \]

where

\[ \text{PC} = \text{dIndLTerm} P \text{ cvar cabs capp cList} A Q_A \left( \text{var} s T_A \right) \text{ liftA} \]
\[ \text{PB} = \text{dIndLTerm} P \text{ cvar cabs capp cList} A Q_A \left( \text{abs} B C e s T_B T_C \right) \left( Q_B, Q_C, \text{ liftE, liftT_B, liftT_C} \right) \]
\[ \text{PBA} = \text{dIndLTerm} P \text{ cvar cabs capp cList} A Q_A \left( \text{app} B T_{\text{BA}} B A \right) \left( Q_B, \text{ liftBA}, \text{ liftTB} \right) \]
\[ \text{PList} = \text{liftListMap} (\text{LTerm} B) (\text{LTerm}^\wedge B Q_B) (P B Q_B) \text{ Pts ts liftList : List}^\wedge (\text{LTerm} B) (P B Q_B) \text{ ts} \]
\[ \text{Pts} = \text{dIndLTerm} P \text{ cvar cabs capp cList} B Q_B : \text{PredMap} (\text{LTerm} B) (\text{LTerm}^\wedge B Q_B) (P B Q_B) \]

Figure 5. Induction hypotheses for LTerm

\[ \text{lift}_{\text{ts}} = \text{LTerm}^\wedge B Q_B T_B \]
\[ \text{liftE'} = \text{Equal}^\wedge A (\text{List} B) Q_A \left( \text{List}^\wedge B Q_B \right) e \]
\[ \text{liftList} = \text{List}^\wedge (\text{LTerm} B) (\text{LTerm}^\wedge B Q_B) \]

Moreover, in the definition of \( p_{\text{ts}} \),

\[ \text{PredMap} : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Set} \]

is the type constructor producing the type of morphisms between predicates defined by

\[ \text{PredMap} A Q Q' = \forall (a : A) \rightarrow Q a \rightarrow Q' a \]

and

\[ \text{liftListMap} : \forall (A : \text{Set}) \rightarrow (Q Q' : A \rightarrow \text{Set}) \rightarrow \text{PredMap} A Q Q' \rightarrow \text{PredMap} (\text{List} A) (\text{List}^\wedge A Q) (\text{List}^\wedge A Q') \]
which takes a morphism $f$ of predicates and produces a morphism of lifted predicates, is defined by

$$\text{liftListMap} \ A \ Q \ Q' \ m \ n l \ t \ t = tt$$

(since $x : \text{List}^a A \ Q \ nil$ must necessarily be the sole inhabitant $tt$ of $\tau$), and by

$$\text{liftListMap} \ A \ Q \ Q' \ m (\text{cons} \ a') (y, x') = (m \ a \ y, \text{liftListMap} \ A \ Q \ Q' \ m' \ l' \ x')$$

(since $x : \text{List}^a A \ Q$ (cons a') must be of the form $x = (y, x')$ where $y : Q_a$ and $x' : \text{List}^a A \ Q'$).

### 5 The General Framework

We can generalize the approach in Section 4 to a general framework for deriving deep induction rules for GADTs that are not truly nested GADTs. We will treat GADTs of the form

$$\text{data} \ G : Set^\alpha \ \Rightarrow \ \text{Set}$$

$$c : \forall \{B : \text{Set}\} \rightarrow F \ G \ B \ \Rightarrow \ \text{G}(K \ B)$$

(9)

For brevity and clarity we indicate only one constructor $c$ in (9), even though a GADT can have any finite number of them, each with a type of the same form as $c$’s. In (9), $F$ and each $K$ in $\overline{K}$ are type constructors with signatures $(\text{Set}^\alpha \ \Rightarrow \ \text{Set}) \rightarrow \text{Set}$ and $(\text{Set}^\beta \ \Rightarrow \ \text{Set}) \rightarrow \text{Set}$, respectively. If $T$ is a type constructor with signature $\text{Set}^\gamma \ \Rightarrow \ \text{Set}$ then $T$ has arity $\gamma$. The overline notation denotes a finite list whose length is exactly the arity of the type constructor being applied to it. The number of type constructors in $\overline{K}$ (resp., $B$) is thus $\alpha$ (resp., $\beta$). In addition, the type constructor $F$ must be constructed inductively according to the following grammar:

$$F \ G \ B := F_1 \ G \ B \times F_2 \ G \ B \ | \ F_1 \ G \ B \ \Rightarrow \ F_2 \ G \ B \ | \ F_1 \ \Rightarrow \ F_2 \ G \ B \ | \ H \ \Rightarrow \ H \ (F_1 \ G \ B)$$

This grammar is subject to the following restrictions. In the third clause the type constructor $F_1$ does not contain $G$. In the fourth clause, none of the $\alpha$-many type constructors in $\overline{F_1}$ contains $G$. This prevents nesting, which would make it impossible to give an induction rule for $G$; see Section 6 below. In the fifth and sixth clauses, $H : \text{Set}^\gamma \ \Rightarrow \ \text{Set}$ is the syntactic reflection of some functor, and thus has an associated map function. It is worth noting that the fifth clause subsumes the cases in which $F \ G \ B$ is a closed type or one of the $B_i$, and that $H$ can be the data type constructor for any (truly) nested type. From the map function for $H$ we can also construct a map function

$$H^\wedge \text{Map} : \forall (A : \text{Set}) (Q Q') : A \ \Rightarrow \ \text{Set} \rightarrow$$

$$\text{PredMap} A Q Q' \rightarrow (H^\wedge \text{PredMap}) (H^\wedge \ A Q') (H^\wedge \ A Q')$$

(10)

for $H^\wedge$. A concrete way to define $H^\wedge \text{Map}$ is by induction on the structure of the type $H$, but we omit such details since they are not essential to the present discussion. A further requirement that applies to all of the type constructors appearing in the right-hand side of the above grammar, including those in $\overline{K}$, is that they must all admit predicate liftings. This is not an overly restrictive condition, though: all GADTs constructed from the above grammar admit predicate liftings. (The fact that the domain of an arrow type is independent of $G$ is crucial for this.) In particular, the lifting for each type constructor $H$ is constructed using its map function. A concrete way to define more general predicate liftings is, again, by induction on the structure of the types in a suitable calculus; this will ensure that the liftings satisfy the crucial property needed to derive deep induction rules, namely that of distributing over the type constructors. We do not give a general definition of predicate liftings here, since that would require us to first design a full type calculus, which is beyond the scope of the present paper. We can however, define liftings for the type constructor $F$ defined by the grammar on page 9 by

1. $F \ G \ B = F_1 \ G \ B \times F_2 \ G \ B$ then

$$F^\wedge \ G \ B \ P \ Q \ x = \text{Pair}^\wedge (F_1 \ G \ B) (F_2 \ G \ B) (F_1^\wedge \ G \ B \ P \ Q_\beta) (F_2^\wedge \ G \ B \ P \ Q_\beta)$$

2. $F \ G \ B = F_1 \ G \ B + F_2 \ G \ B$ then

$$F^\wedge \ G \ B \ P \ Q \ x = \text{Pair}^\wedge (F_1 \ G \ B) (F_2 \ G \ B) (F_1^\wedge \ G \ B \ P \ Q_\beta) (F_2^\wedge \ G \ B \ P \ Q_\beta)$$

3. If $F \ G \ B = F_1 \ G \ B \rightarrow F_2 \ G \ B$ then

$$F^\wedge \ G \ B \ P \ Q_\gamma x \ = \ \forall (z : F_1 \ G \ B) \rightarrow F_1^\wedge \ G \ B \ P \ Q_\beta z \rightarrow F_2^\wedge \ G \ B \ P \ Q_\beta (x z)$$

4. If $F \ G \ B = G (F_1 \ B)$ and $F_1$ does not contain $G$, then

$$F^\wedge \ G \ B \ P \ Q_\beta = P (F_1 \ B) (F_1^\wedge \ B \ Q_\beta)$$

for all $P : \forall (A : \text{Set}) \rightarrow (A \ \Rightarrow \ \text{Set}) \rightarrow G A \ \Rightarrow \ \text{Set}$. 

5. If $F \ G \ B = H (F_1 \ G \ B)$ and $H$ does not contain $G$, then

$$F^\wedge \ G \ B \ P \ Q_\beta = H^\wedge (F_1 \ G \ B) (F_1^\wedge \ G \ B \ P \ Q_\beta)$$

for all $P : \forall (A : \text{Set}) \rightarrow (A \ \Rightarrow \ \text{Set}) \rightarrow G A \ \Rightarrow \ \text{Set}$. 

We assume in the development below that $G$ is a unary type constructor, i.e., that $\alpha = 1$ in (9). Extending the argument to GADTs of arbitrary arity presents no difficulty other than heavier notation. In this case the type of $G$’s single data constructor $c$ can be rewritten as

$$c : \forall (B : \text{Set}) \rightarrow \text{Equal} \ A (K \ B) \rightarrow F \ G \ B \rightarrow G A$$

The predicate lifting $G^\wedge : \forall (A : \text{Set}) \rightarrow (A \ \Rightarrow \ \text{Set}) \rightarrow G A \ \Rightarrow \ \text{Set} \rightarrow G$ is therefore

$$G^\wedge \ A \ Q \ A \ (c B e x) =$$

$$\exists (Q_\alpha) \ \text{Equal}^\wedge \ A (K \ B) \ Q_\alpha (K^\wedge \ B \ Q_\beta) \ e \times F^\wedge \ G \ B \ G^\wedge \ Q_\beta x$$
where \( Q_A : A \rightarrow \text{Set} \), \( Q_B : B \rightarrow \text{Set} \), \( e : \text{Equal } A (K \overline{B}) \), and \( x : F G \overline{B} \). If we have predicate liftings

\[
F^\wedge : \forall (G : \text{Set}^x \rightarrow \text{Set}) (B : \text{Set}) \rightarrow \\
(\forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \rightarrow \\
(B \rightarrow \text{Set}) \rightarrow F G \overline{B} \rightarrow \text{Set}
\]

for \( F \) and

\[
K^\wedge : \forall (B : \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \overline{B} \rightarrow \text{Set}
\]

for \( K \), then the induction hypothesis \( \text{dlnDC} \) associated with \( c \) is

\[
dlnDC = \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \rightarrow \\
\forall (A : \text{Set}) (B : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) \\
(e : \text{Equal } A (K \overline{B}) (x : F G \overline{B}) \rightarrow \\
\text{Equal}^\wedge (K \overline{B}) Q_A (K^\wedge \overline{B} Q_B) e \rightarrow \\
F^\wedge G \overline{B} P Q_B x \rightarrow P A Q_A (c B e x)
\]

and the induction rule for \( G \) is

\[
\forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \rightarrow \\
dlnDC P \rightarrow \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (y : G A) \rightarrow \\
G^\wedge A Q_A y \rightarrow P A Q_A y
\]

(11)

To prove that this rule is sound we define a witness \( \text{dlnGD} \) inhabiting this type by

\[
dlnGD P cc A Q_A (c B e x) (Q_B, \text{liftE}, \text{liftF}) = \\
c c A B Q_A \overline{Q_B} e x \text{liftE} (p \times \text{liftF})
\]

Here,

\[
cc : \text{dlnDC P} \\
e : \text{Equal } A (K \overline{B}) \\
x : F G \overline{B} \\
Q_A : A \rightarrow \text{Set} \\
\text{liftE} : \text{Equal}^\wedge (K \overline{B}) Q_A (K^\wedge \overline{B} Q_B) e \\
\text{liftF} : F^\wedge G \overline{B} G^\wedge Q_B x
\]

and \( Q_B : B \rightarrow \text{Set} \), so

\[
(Q_B, \text{liftE}, \text{liftF}) : G^\wedge A Q_A (c B e x)
\]
as expected. Finally, the morphism of predicates

\[
p : \text{PredMap} (F G \overline{B})(F^\wedge G \overline{B} G^\wedge Q_B)(F^\wedge G \overline{B} P Q_B)
\]
is defined by structural induction on \( F \) as follows:

- If \( F G \overline{B} = F_1 G \overline{B} \times F_2 G \overline{B} \) then
  \[
  F^\wedge G \overline{B} P Q_B = \\
  \text{Pair}^\wedge (F_1 G \overline{B})(F_2 G \overline{B})(F_1^\wedge G \overline{B} P Q_B)(F_2^\wedge G \overline{B} P Q_B)
  \]
The induction hypothesis ensures morphisms of predicates

\[
p_1 : \text{PredMap} (F_1 G \overline{B})(F_1^\wedge G \overline{B} G^\wedge Q_B Q)(F_1^\wedge G \overline{B} P Q_B)
\]
and

\[
p_2 : \text{PredMap} (F_2 G \overline{B})(F_2^\wedge G \overline{B} G^\wedge Q_B Q)(F_2^\wedge G \overline{B} P Q_B)
\]

For \( x_1 : F_1 G \overline{B}, \text{liftF}_1 : F_1^\wedge G \overline{B} G^\wedge Q_B x_1, x_2 : F_2 G \overline{B} \) and \( \text{liftF}_2 : F_2^\wedge G \overline{B} G^\wedge Q_B x_2 \) we then define

\[
p (x_1, x_2) (\text{liftF}_1, \text{liftF}_2) = (p_1 x_1 \text{liftF}_1, p_2 x_2 \text{liftF}_2)
\]

- The case \( F G \overline{B} = F_1 G \overline{B} + F_2 G \overline{B} \) is analogous.
- If \( F G \overline{B} = F_1 \overline{B} \rightarrow F_2 \overline{B} \) then

\[
F^\wedge G \overline{B} P Q_B x = \\
\forall (z : F_1 \overline{B}) \rightarrow F_1^\wedge \overline{B} Q_B z \rightarrow F_2^\wedge G \overline{B} P Q_B (x z)
\]

where \( x : F G \overline{B} \). The induction hypothesis ensures a morphism of predicates

\[
p_2 : \text{PredMap} (F_2 G \overline{B})(F_2^\wedge G \overline{B} G^\wedge Q_B)(F_2^\wedge G \overline{B} P Q_B)
\]

We therefore define \( p x \cdot \text{liftF} : F^\wedge G \overline{B} P Q_B x \) where \( \text{liftF} : F^\wedge G \overline{B} G^\wedge Q_B x \), to be

\[
p x \cdot \text{liftF} z \cdot \text{liftF}_1 = p_2 (x z) (\text{liftF} z \cdot \text{liftF}_1)
\]

for \( z : F_2 \overline{B} \) and \( \text{liftF}_1 : F_1^\wedge \overline{B} Q_B z \). Note that \( F_1 \) not containing \( G \) is a necessary restriction since the proof relies on \( F^\wedge G \overline{B} G^\wedge Q_B x \) and \( F^\wedge G \overline{B} P Q_B x \) having the same domain \( F_1 \overline{B} Q_B z \).

- If \( F G \overline{B} = H (F_1 \overline{B}) \) and \( F_1 \) does not contain \( G \), then

\[
F^\wedge G \overline{B} P Q_B = H^\wedge \overline{B} Q_B
\]

for all \( P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set} \). We therefore define

\[
p : \text{PredMap} (H B)(H^\wedge B Q_B)(H^\wedge \overline{B} Q_B)
\]
to be the identity morphism on predicates.

- If \( F G \overline{B} = H (F_1 G \overline{B}) \) and \( H \) does not contain \( G \), then

\[
F^\wedge G \overline{B} P Q_B = H^\wedge (F_1 G \overline{B}) (F_1^\wedge G \overline{B} P Q_B)
\]

for all \( P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set} \). Since \( H \) is not a GADT, \( H^\wedge \) has a map function \( H^\wedge \text{Map} \) as in (10). The induction hypothesis ensures morphisms of predicates

\[
p_k : \text{PredMap} (F_k G \overline{B})(F_k^\wedge G \overline{B} G^\wedge Q_B)(F_k^\wedge G \overline{B} P Q_B)
\]

We therefore define

\[
p = H^\wedge \text{Map} (F_k G \overline{B})(F_k^\wedge G \overline{B} G^\wedge Q_B)(F_k^\wedge G \overline{B} P Q_B) p_k
\]

Observing that the above development essentially uses the equality GADT and its predicate lifting in the discrete category of types to extend the lifting in [15] — now specialized to the same category — to GADTs, we have established the following theorem:

**Theorem 5.1.** A GADT \( G \) of the form in (9) admits the deep induction rule in (11).
6 Truly Nested GADTs Need Not Admit Deep Induction Rules

In Sections 4 and 5 we derived deep induction rules for GADTs that are not truly nested GADTs. Since both (truly) nested types and GADTs without true nesting admit deep induction rules, we might expect truly nested GADTs to admit them as well. Unfortunately, however, the techniques developed in the previous sections do not extend to truly nested GADTs. Indeed, while the induction rule for a data type generally relies on (unary) parametricity of the model interpreting it, deep induction for a truly nested type or a truly nested GADT crucially relies on this interpretation being functorial. Whereas ADTs and nested types both admit functorial parametric semantics, proper GADTs admit parametric semantics but do not admit functorial semantics. In this section we show how the techniques developed in this paper for deriving deep induction rules go wrong for truly nested GADTs by analyzing the following very simple example:

\[
data \, G : \text{Set} \rightarrow \text{Set} \quad \text{where} \\
c : \forall (A : \text{Set}) \rightarrow G \, (G \, A) \rightarrow G \, (A \times A)
\]

We acknowledge that G is semantically equivalent to the empty data type, and thus has a trivial (deep) induction principle. We could, of course, consider a more realistic counterexample, but this would only add notational overhead for no gain in conceptual clarity. Indeed, we need only exhibit a single GADT whose deep induction rule cannot be obtained using the techniques of this paper, and for which more robust techniques will therefore be needed if deep induction rules are to be derived for them.

To see this, we first rewrite the constructor c’s type as

\[
c : \forall (B : \text{Set}) \rightarrow \text{Equal} \, (A \times B) \rightarrow G \, (G \, B) \rightarrow G \, A
\]

The predicate lifting

\[
G^\land : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G \, A \rightarrow \text{Set}
\]

for G is therefore

\[
G^\land \, A \, Q_A \, (c \, B \, e \, x) = \\
\exists \, [Q_B] \, \text{Equal}^\land \, (A \times B) \, Q_A \, (\text{Pair}^\land \, B \, Q_B \, Q_A) \, e \times \, G^\land \, (G \, B) \, (G^\land \, B \, Q_B) \, x
\]

where \( Q_A : A \rightarrow \text{Set} \), \( Q_B : B \rightarrow \text{Set} \), \( e : \text{Equal} \, (A \times B) \), and \( x : G \, (G \, B) \). The induction hypothesis \( \text{dIndC} \) for c is

\[
\lambda \, (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G \, A \rightarrow \text{Set}) \rightarrow \\
\forall \, (A \, B : \text{Set}) \, (Q_A : A \rightarrow \text{Set}) \, (Q_B : B \rightarrow \text{Set}) \\
\quad \quad (e : \text{Equal} \, (A \times B)) \rightarrow \\
\quad \quad \text{Equal}^\land \, (A \times B) \, Q_A \, (\text{Pair}^\land \, B \, Q_B \, Q_A) \, e \rightarrow \\
\quad \quad P \, (G \, B) \, (P \, B \, Q_B) \, x \rightarrow P \, A \, Q_A \, (c \, B \, e \, x)
\]

so the deep induction rule for G is

\[
\forall \, (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G \, A \rightarrow \text{Set}) \rightarrow \\
\text{dIndC} \, P \rightarrow \forall \, (A \, : \text{Set}) \, (Q : A \rightarrow \text{Set}) \, (y : G \, A) \\
\quad \quad G^\land \, A \, Q \, y \rightarrow P \, A \, Q \, y
\]

But if we now try to show that this rule is sound by constructing a witness \( \text{dIndG} \) inhabiting this type we run into problems. We can define

\[
\text{dIndG} \, P \, c \, c \, Q \, (c \, B \, e \, x) \, (Q', \, \text{liftE}, \, \text{liftG}) \\
= \quad c \, A \, B \, Q \, Q' \, e \, x \, \text{liftE} \, p
\]

where

\[
c : \text{dIndC} \, P \\
Q : A \rightarrow \text{Set} \\
Q' : B \rightarrow \text{Set} \\
e : \text{Equal} \, (A \times B) \\
x : G \, (G \, B) \\
\text{liftG} : G^\land \, (G \, B) \, (G^\land \, B \, Q') \, x \\
\text{liftE} : \text{Equal}^\land \, A \, (A \times B) \, Q \, (\text{Pair}^\land \, B \, B \, Q' \, Q') \, e
\]

but we still need to define \( p : (G \, B) \, (P \, B \, Q') \, x \). For this we can use the induction rule and let

\[
p = \text{dIndG} \, P \, c \, c \, (G \, B) \, (P \, B \, Q') \, x \, q
\]

but we still need to define

\[
q : G^\land \, (G \, B) \, (P \, B \, Q') \, x
\]

If we had the map function

\[
G^\land \, \text{Map} : \forall (A : \text{Set}) \, (Q, \, Q' : A \rightarrow \text{Set}) \rightarrow \\
\text{PredMap} \, A \, Q \, Q' \rightarrow \\
\text{PredMap} \, (G \, A) \, (G^\land \, A \, Q) \, (G^\land \, A \, Q')
\]

for \( G^\land \) then we could define

\[
q = G^\land \, \text{Map} \, (G \, B) \, (G^\land \, B \, Q') \, (P \, B \, Q') \, (\text{dIndG} \, P \, c \, c \, B \, Q' \, x \, \text{liftG})
\]

Unfortunately, however, we cannot define \( G^\land \, \text{Map} \). Indeed, its definition would have to be

\[
G^\land \, \text{Map} \, A \, Q \, Q' \, m \, (c \, B \, e \, x) \, (Q_A, \, \text{liftE}, \, \text{liftG}) \\
= (Q_A, \, \text{liftE}', \, \text{liftG}')
\]

for some

\[
Q_A : B \rightarrow \text{Set} \\
\text{liftE}' : \text{Equal}^\land \, A \, (A \times B) \, Q' \, (\text{Pair}^\land \, B \, Q' \, Q'_A) \, e \\
\text{liftG}' : G^\land \, (G \, B) \, (G^\land \, B \, Q'_A) \, x
\]

where

\[
Q : A \rightarrow \text{Set} \\
Q' : A \rightarrow \text{Set} \\
Q_A : B \rightarrow \text{Set} \\
m : \text{PredMap} \, A \, Q \, Q' \\
e : \text{Equal} \, (A \times B) \\
x : G \, (G \, B) \\
\text{liftE} : \text{Equal}^\land \, A \, (A \times B) \, Q \, (\text{Pair}^\land \, B \, Q_A \, Q_B) \, e \\
\text{liftG} : G^\land \, (G \, B) \, (G^\land \, B \, Q_B)
\]

That is, we would need to produce a proof \( \text{liftE}' \) of the (extensional) equality of the predicates \( Q' \) and \( \text{Pair}^\land \, B \, Q_A \, Q'_A \) from just a proof \( \text{liftE} \) of the (extensional) equality of the predicates \( Q \) and \( \text{Pair}^\land \, B \, Q_A \, Q_B \) and a morphism of predicates \( m \) from \( Q \) to \( Q' \). But this will not be possible in general: the facts that \( Q \) is equal to \( \text{Pair}^\land \, B \, Q_A \, Q_B \) and that there
is a morphism of predicates m from Q to Q′ do not guarantee that there exists a predicate Q′_m such that Q′ is equal to Pair^m B Q_m Q′_m.

At a deeper level, the fundamental issue is that the Equal type does not have functorial semantics [10], so that having morphisms A → A′ and B → B′ (for any type A, A′, B and B′) and a proof that A is equal to A′ does not provide a proof that B is equal to B′. And not being able to define

```
Equal^\wedge \map \colon \forall (A B \colon \text{Set})
  \quad (Q_A Q'_A \colon A \rightarrow \text{Set})
  \quad (Q_B Q'_B \colon B \rightarrow \text{Set}) \rightarrow
  \quad \text{PredMap} A Q_A Q'_A \rightarrow
  \quad \text{PredMap} B Q_B Q'_B \rightarrow
  \quad \text{PredMap} (\text{Equal} A B) (\text{Equal}^\wedge A B Q_A Q_B)
  \quad (\text{Equal}^\wedge A B Q'_A Q'_B)
```

of course makes it unclear how to define G^\wedge \map for more general G.

7 Case Study: Extracting Types of Lambda Terms

In this section, we use deep induction for the LTerm GADT from Figure 2 to infer the type from a lambda term. The following predicate either returns the type of its input lambda term if that type can be inferred or indicates that the type inference fails:

```
\text{GetType} \colon \forall (A \colon \text{Set}) \rightarrow \text{LTerm} A \rightarrow \text{Set}
\text{GetType} A t = \text{Maybe} (\text{LType} A)
```

Of course, since GetType is (trivially) defined by structural induction, we could perform type inference using hand-threaded applications of structural induction as observed at the end of Section 2. Nevertheless, the example as given nicely illustrates deep induction.

By construction every lambda term in LTerm is well-typed, but that (necessarily unique) type cannot always be inferred. The predicate GetType uses the standard Maybe data type to represent failure of type inference. It is defined by:

```
data Maybe : \text{Set} \rightarrow \text{Set} where
  \text{nothing} : \forall (A : \text{Set}) \rightarrow \text{Maybe} A
  \text{just} : \forall (A : \text{Set}) \rightarrow A \rightarrow \text{Maybe} A
```

We want to show that GetType A t is satisfied by every element t in LTerm A, i.e., we want to prove:

```
\text{groupBy} \colon \forall (A : \text{Set}) (t : \text{LTerm} A) \rightarrow \text{GetType} A t
```

This property can be proved with deep induction, which is used to apply the induction hypothesis to the individual terms in the list of terms that the data constructor list takes as an argument. Indeed, using the deep induction rule dIndLTerm from Section 4.3 we can define GetTypeProof by

```
\text{groupBy} A t = \text{dIndLTerm} P \text{CVar} \text{CApp} \text{CList} \text{Just} (\text{LTerm}^\wedge \text{KT} A t)
```

where t : LTerm A, P is the polymorphic predicate

```
\lambda (A : \text{Set}) (Q : A \rightarrow \text{Set}) (t : \text{LTerm} A) \rightarrow \text{Maybe} (\text{LType} A)
```

and K_T is the constantly T-valued predicate on A, and

```
\text{LTerm}^\wedge \text{KT} \colon \forall (A : \text{Set}) (t : \text{LTerm} A) \rightarrow \text{LTerm}^\wedge A K_T t
```

is a term, to be defined below, witnessing that K_T can be lifted to all terms. We also need the applications to P of each of the induction hypotheses from Section 4.3. These are given in Figure 7. In the first clause, CVar returns just T_A.

In the second clause, CApp returns nothing if its final argument is nothing and

```
\text{capp} A B C Q A B Q C e s T_B T_C \text{lift} E \text{lift} T_B \text{just} T_C
```

otherwise. In the third clause,

```
\text{clist} A B Q Q' e \text{nil} \text{lift} E \text{lift} ts = \text{nothing}
```

where lift E : Equal^\wedge A (List B) Q (List^\wedge B Q') e, and lift ts : List^\wedge (LTerm B) (GetType B) ts. When ts = cons t ts' the type of lift ts becomes

```
\text{List}^\wedge (\text{LTerm B}) (\text{GetType B}) (\text{cons} t \text{ts}')
\text{just} \text{List}^\wedge (\text{LTerm B}) (\text{Extract} B) (\text{GetType B}) \text{ts}'
\text{just} \text{List}^\wedge (\text{LTerm B}) (\text{Extract}^\wedge B) (\text{GetType B}) \text{ts}'
```

We pattern match on the first component of the pair to define

```
\text{clist} A B Q Q' e (\text{cons} t \text{ts}') \text{lift} E (\text{nothing}, \text{lift} ts') = \text{nothing}
\text{clist} A B Q Q' e (\text{cons} t \text{ts}') \text{lift} E (\text{just} T', \text{lift} ts') = \text{just} (\text{List} B e T')
```

Here e : Equal A (List B), T' : LType B, and

```
\text{lift} ts' : \text{List}^\wedge (\text{LTerm B}) (\text{GetType B}) \text{ts}'
```

To finish definingGetTypeProof we still need a proof

```
\text{LTerm}^\wedge \text{KT} \colon \forall (A : \text{Set}) (t : \text{LTerm} A) \rightarrow \text{LTerm}^\wedge A K_T t
```

Since LTerm^\wedge is defined in terms of LType^\wedge and Arr^\wedge, and since LType^\wedge is also defined in terms of List^\wedge, we need analogous functions LType^\wedge KT, Arr^\wedge KT and List^\wedge KT, respectively, for each of these liftings as well. We only give the definition of LTerm^\wedge KT here since LType^\wedge KT, Arr^\wedge KT, and List^\wedge KT are defined analogously. We have:

- If s : String and T : LType A we define

```
\text{LTerm}^\wedge KT A (\text{var} s T) = \text{LType}^\wedge KT A T
```
\(\forall (A : \text{Set}) \ (Q_A : A \rightarrow \text{Set}) \ (s : \text{String}) \ (T_A : L\text{Type} A) \rightarrow L\text{Type}^\Lambda A Q_A T_A \rightarrow \text{Maybe} (L\text{Type} A)\)

cabs : \(\forall (A B C : \text{Set}) \ (Q_A : A \rightarrow \text{Set}) \ (Q_B : B \rightarrow \text{Set}) \ (Q_C : C \rightarrow \text{Set})
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad …
The above techniques can be used to define a function
$$G^\wedge\text{KT} : \forall (A : \text{Set}) \times G^\wedge A K_T x \to G^\wedge A K_T x$$
for any GADT G as defined in Section 5. To prove a proof of $G^\wedge A K_T x$ for every term $x : G A$, we need to know that, if G has a constructor $c : H (F G B) \to G (K B)$, then H cannot construct a GADT so the generalization H^\wedge Map of listLiftMap in the final bullet point above is guaranteed to exist. We also need to know that the lifting of $K_T$ to types constructed by any nested type constructor F is extensionally equal to $K_T$ on the types it constructs. For example, we might need a proof that Pair^\wedge A B K_T K_T is equal to $K_T$ on $A \times B$. Given a pair $(a, b) : A \times B$, we have that
$$\text{Pair}^\wedge A B K_T K_T (a, b) = K_T a \times K_T b = \top \times \top$$
whereas $K_T (a, b) = \top$. While these types are not equal, they are clearly isomorphic. Similar isomorphisms between $F^\wedge A K_T$ and $K_T$ hold for all other nested type constructors F as well. These isomorphisms can either be proved on an as-needed basis or, since $F^\wedge A K_T = K_T$ is the unary analogue of the Identity Extension Lemma, be obtained at the meta-level as a consequence of unary parametricity. At the object level, our Agda code simply postulates each isomorphism needed since an Agda implementation of full parametricity for some relevant calculus is beyond the scope of the present paper.

8 Conclusion

This paper extends (deep) induction to GADTs that are not truly nested GADTs. It also shows that truly nested GADTs do not obviously admit (deep) induction rules. Our development is implemented in Agda, as is our case study from Section 7. Our development opens the way to incorporating automatic generation of (deep) induction rules for them into proof assistants.

Acknowledgments

This work was supported by National Science Foundation award 1906388.

References

[1] R. Atkey. 2012. Relational parametricity for higher kinds. In Computer Science Logic. 46–61.
[2] E. S. Bainbridge, P. Freyd, A. Scedrov, and P. J. Scott. 1990. Functorial polymorphism. Theoretical Computer Science 70(1) (1990), 35–64. https://doi.org/10.1016/0304-3975(90)90151-7
[3] R. Bird and L. Meertens. 1998. Nested datatypes. In Mathematics of Program Construction. 52–67. https://doi.org/10.1007/BFb0054258
[4] J. Cheney and R. Hinze. 2003. First-class phantom types. (2003). CU-CIS TR2003-1901, Cornell University.
[5] P. Dybjer. 1994. Inductive families. Formal Aspects of Computing 6(4) (1994), 440–465. https://doi.org/10.1007/BF01211308
[6] P. Fu and P. Selinger. 2018. Dependent typed folds for nested data types. (2018). https://arxiv.org/abs/1806.05230
[7] N. Ghani, P. Johann, F. Nordvall Forsberg, F. Orsanigo, and T. Revell. 2015. Bifibrational functorial semantics for parametric polymorphism. In Mathematical Foundations of Program Semantics. 165–181. https://doi.org/10.1007/978-3-319-18833-1_10
[8] R. Hinze. 2003. Fun with phantom types. In The Fun of Programming. 245–262.
[9] P. Johann and N. Ghani. 2008. Foundations for Structured Programming with GADTs. In Proceedings. Principles of Programming Languages. 297–308. https://doi.org/10.1109/POPL.2008.4328475
[10] P. Johann, E. Ghiorzi, and D. Jeffries. 2021. GADTs, functoriality, parametricity: Pick two. In Logical and Semantic Frameworks with Applications.
[11] P. Johann, E. Ghiorzi, and D. Jeffries. 2021. Parametricity for primitive nested types. In Foundations of Software Science and Computation Structures. 324–343. https://doi.org/10.1007/978-3-030-71995-1_17
[12] P. Johann and A. Polonsky. 2019. Higher-kind ed data types: Syntax and semantics. In Logic in Computer Science. 1–13. https://doi.org/10.1109/LICS.2019.8785657
[13] P. Johann and A. Polonsky. 2020. Deep induction: Induction rules for (truly) nested types. In Foundations of Software Science and Computation Structures. 339–358. https://doi.org/10.1007/978-3-030-54231-5_18
[14] S. Peyton Jones, D. Vytiniotis, S. Weirich, and G. Washburn. 2006. Simple unification-based type inference for GADTs. In International Conference on Functional Programming. 50–61. https://doi.org/10.1109/ICFP.2006.17
[15] S. Mac Lane. 1971. Categories for the Working Mathematician. Springer.
[16] C. McBride. 1999. Dependently Typed Programs and their Proofs. (1999). PhD thesis, University of Edinburgh.
[17] Y. Minsky. 2015. Why GADTs Matter for Performance. (2015). blogs.janestreet.com/why-gadts-matter-for-performance
[18] E. Pasalic and N. Linger. 2004. Meta-programming with typed object-language representations. In Generic Programming and Component Engineering. 156–167. https://doi.org/10.1007/978-3-540-30175-2_8
[19] F. Pottier and Y. Régis-Gianas. 2006. Stratified type inference for generalized algebraic data types. In Principles of Programming Languages. 232–244. https://doi.org/10.1101/111320.1113058
[20] D. Roundy. 2006. Implementing the darcs Patch Formalism ...and Verifying It. (2006). https://physics.oregonstate.edu/~roundy/talks/odedem
[21] T. Schrijvers, S. L. Peyton Jones, M. Sulzmann, and D. Vytiniotis. 2009. Complete and decidable type inference for GADTs. In International Conference on Functional Programming. 341–352. https://doi.org/10.1101/116368.1596599
[22] T. Sheard and E. Pasalic. 2004. Meta-programming with built-in type equality. In Workshop on Logical Frameworks and Meta-languages. 106–124.
[23] E. Tassi. 2019. Deriving proved equality tests in Coq-elpi: Stronger induction principles for containers in Coq. In Interactive Theorem Proving. 1–18. https://doi.org/10.4230/LIPIcs.CIPT.2019.23
[24] The Coq Development Team. 2020. The Coq Proof Assistant, version 8.11.0. (2020). https://doi.org/10.5281/zenodo.3744225
[25] M. Ullrich. 2020. Generating Induction Principles for Nested Induction Types in MetaCoq. (2020). PhD thesis, Saarland University.
[26] D. Vytiniotis and S. Weirich. 2010. Parametricity, type equality, and higher-order polymorphism. Journal of Functional Programming 20(2) (2010), 175–210. https://doi.org/10.1017/S0956796810000079
[27] H. Xi, C. Chen, and G. Chen. 2003. Guarded recursive datatype constructors. In Proceedings of Programming Languages. 224–235. https://doi.org/10.1101/604131.604150
[28] N. Zilberstein. 2015. CIS194 homepage. (2015). https://www.seas.upenn.edu/~cis194/spring15/lectures/11-stlc.html