

(Deep) Induction for GADTs

Patricia Johann¹ Enrico Ghiorzi² Daniel Jeffries³

*Department of Computer Science
Appalachian State University*

Abstract

Abstract goes here.

Keywords: Please list keywords from your paper here, separated by commas.

1 Introduction

Introduction goes here.

2 Deep induction for ADTs and nested types

2.1 Syntax of ADTs and nested types

(Polynomial) algebraic data types (ADTs), both built-in and user-defined, have long been at the core of functional languages such as Haskell, ML, Agda, Epigram, and Idris. ADTs are used extensively in functional programming to structure computations, to express invariants of the data over which computations are defined, and to ensure the type safety of programs specifying those computations. ADTs include unindexed types, such as the type of natural numbers, and types indexed over other types, such as the quintessential example of an ADT, the type of lists (here coded in Agda)

$$\begin{aligned}
 \text{data List} &: \text{Set} \rightarrow \text{Set} \text{ where} \\
 \text{nil} &: \text{List A} \\
 \text{cons} &: \text{A} \rightarrow \text{List A} \rightarrow \text{List A}
 \end{aligned}
 \tag{1}$$

Notice that all occurrences of `List` in the above encoding are instantiated at the same index `A`. Thus, the instances of `List` at various indices are defined independently from one another. That is a defining feature of ADTs: an ADT defines a *family of inductive types*, one for each index type.

Over time, there has been a notable trend toward data types whose non-regular indexing can capture invariants and other sophisticated properties that can be used for program verification and other applications.

¹ Email: johannp@appstate.edu

² Email: ghiorzie@appstate.edu

³ Email: jeffriesd@appstate.edu

A simple example of such a type is given by Bird and Meertens' [2] prototypical *nested type*

$$\begin{aligned} \text{data PTree} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{pLeaf} &: A \rightarrow \text{PTree } A \\ \text{pNode} &: \text{PTree } (A \times A) \rightarrow \text{PTree } A \end{aligned} \tag{2}$$

of perfect trees, which can be thought of as constraining lists to have lengths that are powers of 2. In the above code, the constructor `PNode` uses data of type `PNode (A × A)` to construct data of type `PNode A`. Thus, it is clear that the instantiations of `PNode` at various indices cannot be defined independently, so that the entire family of types must actually be defined at once. A nested type thus defines not a family of inductive types, but rather an *inductive family of types*.

Nested types include simple nested types, like perfect trees, none of whose recursive occurrences occur below another type constructor, and *truly* nested types, such as the nested type

$$\begin{aligned} \text{data Bush} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{bNil} &: \text{Bush } A \\ \text{bCons} &: A \rightarrow \text{Bush } (\text{Bush } A) \rightarrow \text{Bush } A \end{aligned} \tag{3}$$

of bushes, whose recursive occurrences appear below their own type constructors. Note that, while the constructors of a nested type can contain occurrences of the type instantiated at any index, the return types of its constructors still have to be the same type instance of the type being defined. In other words, all constructors of `PTree A` have to return an element of type `PTree A`, and all constructors of `Bush A` have to return an element of type `Bush A`.

2.2 Induction principles for ADTs and nested types

An induction principle for a data type allows to prove that a predicate holds for every element of that data type, provided that it holds for every element inductively produced by the type's constructors. In this paper, we are interested in induction principles for proof-relevant predicates. A proof-relevant predicate on a type $A : \text{Set}$ is a function $A \rightarrow \text{Set}$ mapping each $x : A$ to the set of proofs that the predicate holds for x . For example, the induction principle for `List` is

$$\forall (A : \text{Set}) (P : \text{List } A \rightarrow \text{Set}) \rightarrow P \text{ Nil} \rightarrow (\forall (a : A) (ls : \text{List } A) \rightarrow P \text{ ls} \rightarrow P (\text{Cons } a \text{ ls})) \rightarrow \forall (ls : \text{List } A) \rightarrow P \text{ ls}$$

Note that the data inside a structure of type `List` is treated monolithically (i.e., ignored) by this induction rule. Indeed, the induction rule inducts over only the top-level structures of data types, leaving any data internal to the top-level structure untouched. Since this kind of induction principle is only concerned with the structure of the type, and unconcerned with the contained data, we will then refer to it as *structural induction*.

We can extend such a structural induction principle to some nested types, such as `PTree`. The only difference from the induction principle for ADTs is that, since a nested type is defined as a whole inductive family of types at once, its induction rule has to necessarily involve a polymorphic predicate. Thus, the induction rule for `PTree` is

$$\begin{aligned} &\forall (P : \forall (A : \text{Set}) \rightarrow \text{PTree } A \rightarrow \text{Set}) \rightarrow (\forall (A : \text{Set}) (a : A) \rightarrow P A (\text{PLeaf } a)) \\ &\rightarrow (\forall (A : \text{Set}) (\text{pt} : \text{PTree } (A \times A)) \rightarrow P (A \times A) \text{ pt} \rightarrow P A (\text{PNode pt})) \rightarrow \forall (A : \text{Set}) (\text{pt} : \text{PTree } A) \rightarrow P A \text{ pt} \end{aligned}$$

Structural induction principles cannot be extended to truly nested types, such as `Bush`. Instead, for such data types it is necessary to use a *deep induction* principle [6]. Such a principle, unlike structural induction, inducts over all of the structured data present, by traversing not just the outer structure with a predicate P , but also each data element contained in the data type with a custom predicate Q . This additional predicate is lifted to predicates on any internal structure 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.

For example, the deep induction rule for `Bush` is

$$\begin{aligned} & \forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Bush } A \rightarrow \text{Set}) \rightarrow (\forall (A : \text{Set}) \rightarrow P \text{ A BNil}) \\ & \rightarrow (\forall (A : \text{Set})(Q : A \rightarrow \text{Set})(a : A)(bb : \text{Bush } (\text{Bush } A)) \rightarrow Q a \rightarrow P (\text{Bush } A) (\text{Bush}^\wedge A Q) bb \rightarrow P A Q (\text{BCons } a \text{ } bb)) \\ & \rightarrow \forall (A : \text{Set})(Q : A \rightarrow \text{Set})(b : \text{Bush } A) \rightarrow \text{Bush}^\wedge A Q b \rightarrow P A Q b \end{aligned}$$

where $\text{Bush}^\wedge : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Bush } A \rightarrow \text{Set}$ lifts a predicate Q on data of type A to a predicate on data of type $\text{Bush } A$ asserting that Q holds for every element of type A contained in its argument bush. It is defined as

$$\begin{aligned} \text{Bush}^\wedge A Q \text{ BNil} &= \top \\ \text{Bush}^\wedge A Q (\text{BCons } a \text{ } bb) &= Q a \times \text{Bush}^\wedge (\text{Bush } A) (\text{Bush}^\wedge A Q) bb \end{aligned}$$

where $Q : A \rightarrow \text{Set}$ and $bb : \text{Bush } (\text{Bush } A)$.

Despite deep induction being motivated by the need to produce an induction principle for truly nested types, it can equally be applied to all other ADTs and nested types. For example, the deep induction principle for `List` is

$$\begin{aligned} & \forall (A : \text{Set})(P : \text{List } A \rightarrow \text{Set})(Q : A \rightarrow \text{Set}) \rightarrow P \text{ Nil} \rightarrow (\forall (a : A)(ls : \text{List } A) \rightarrow Q a \rightarrow P ls \rightarrow P (\text{Cons } a \text{ } ls)) \\ & \rightarrow \forall (ls : \text{List } A) \rightarrow \text{List}^\wedge A Q ls \rightarrow P ls \end{aligned}$$

where $\text{List}^\wedge : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{List } A \rightarrow \text{Set}$ lifts a predicate Q on data of type A to a predicate on data of type $\text{List } A$ asserting that Q holds for every element of its argument list.

As a further example, the deep induction rule for `PTree` is

$$\begin{aligned} & \forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{PTree } A \rightarrow \text{Set}) \rightarrow (\forall (A : \text{Set})(Q : A \rightarrow \text{Set})(a : A) \rightarrow Q a \rightarrow P A Q (\text{PLeaf } a)) \\ & \rightarrow (\forall (A : \text{Set})(Q : A \rightarrow \text{Set})(a : \text{PTree } (A \times A)) \rightarrow P (A \times A) (\text{Pair}^\wedge A A Q Q) a \rightarrow P A Q (\text{PNode } a)) \\ & \rightarrow \forall (A : \text{Set})(Q : A \rightarrow \text{Set})(a : \text{PTree } A) \rightarrow \text{PTree}^\wedge A Q a \rightarrow P A Q a \end{aligned}$$

where $\text{Pair}^\wedge : \forall (A : \text{Set})(B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow A \times B \rightarrow \text{Set}$ lifts predicates Q_A on data of type A and Q_B on data of type B to a predicate on pairs of type $A \times B$, so that $\text{Pair}^\wedge A B Q_A Q_B (a, b) = Q_A a \times Q_B b$, and $\text{PTree}^\wedge : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{PTree } A \rightarrow \text{Set}$ lifts a predicate Q on data of type A to a predicate on data of type $\text{PTree } A$ asserting that Q holds for every element of type A contained in its argument perfect tree.

Moreover, for types admitting both deep induction and structural induction, the former generalizes the latter. Indeed, structural induction rules can be derived from deep induction rules by always choosing the constantly true predicate as the custom predicate traversing each data element contained in the data type. That way, deep induction only inspects the structure of the data type and not its content, just like structural induction does. A concrete example of such derivation is demonstrated in Section 4.1.

3 Introducing GADTs

As noted in Subsection 2.1, the return types of the constructors of a nested type have to be the same type instance of the type being defined. As a further generalization of ADTs and nested types, *generalized algebraic data types* (GADTs) [3,14,12] relax the restriction on the type instances appearing in a data type definition by allowing their constructors both to take as arguments *and return as results* data whose types involve type instances of the GADT other than the one being defined.

GADTs are used in precisely those situations in which different behaviors at different instances of a data type 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. Applications of 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. GADTs have also been used, e.g., to implement tagless interpreters [8,11,10], to improve memory performance [7], and to design APIs [9].

Example 3.1 As a first and notable example of GADT, we consider the `Equal` type. This GADT is parametrized by two type indices, but it is only possible to construct a data element if the two indices are instantiated at the same type. In Agda, we code it as

$$\begin{aligned} \text{data Equal} &: \text{Set} \rightarrow \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{refl} &: \text{Equal } A \ A \end{aligned} \tag{4}$$

`Equal` has thus a single data element when its two type arguments are the same and no data elements otherwise. The importance of the `Equal` type lies in the fact that we can understand other GADTs in terms of equality, as shown in the examples below.

Example 3.2 A more complex example of GADT is the type of sequences,

$$\begin{aligned} \text{data Seq} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{const} &: A \rightarrow \text{Seq } A \\ \text{sPair} &: \text{Seq } A \rightarrow \text{Seq } B \rightarrow \text{Seq } (A \times B) \end{aligned} \tag{5}$$

which comprises sequences of any type `a` and sequences obtained by pairing the data in two already existing sequences. Such GADTs can be understood in terms of the `Equal` data type [3,12]. For example, we can rewrite the `Seq` type as

$$\begin{aligned} \text{data Seq} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{const} &: A \rightarrow \text{Seq } A \\ \text{sPair} &: \forall (B \ C : \text{Set}) \rightarrow \text{Equal } A \ (B \times C) \rightarrow \text{Seq } B \rightarrow \text{Seq } C \rightarrow \text{Seq } A \end{aligned} \tag{6}$$

where the requirement that the `sPair` constructor produces an instance of `Seq` at a product type has been replaced with the requirement that the instance of `Seq` returned by `sPair` is *equal* to some product type. This encoding is particularly convenient when representing GADTs as Church encodings [13,1].

Example 3.3 GADTs are expressive enough to represent, for example, the types and terms of the simply typed lambda calculus. We will use a variant of the simply typed lambda calculus including a base type of booleans, and type constructors for arrow types and list types. We define a GADT `LType` to encode the set of permissible types for this calculus:

$$\begin{aligned} \text{data LType} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{tBool} &: \forall (B : \text{Set}) \rightarrow \text{Equal } A \ \text{Bool} \rightarrow \text{LType } A \\ \text{tArr} &: \forall (B \ C : \text{Set}) \rightarrow \text{Equal } A \ (B \rightarrow C) \rightarrow \text{LType } B \rightarrow \text{LType } C \rightarrow \text{LType } A \\ \text{tList} &: \forall (B : \text{Set}) \rightarrow \text{Equal } A \ (\text{List } B) \rightarrow \text{LType } B \rightarrow \text{LType } A \end{aligned} \tag{7}$$

The type system includes a `tBool` constructor representing the type of booleans; an arrow type constructor, `tArr`, which takes two types `B` and `C` and produces the arrow type `B → C`; and a `tList` constructor which takes a type `B` and constructs the type `List B`.

The term calculus for this type system includes variables, abstraction, application, and also includes an introduction rule that takes a list of lambda terms of type `B` and produces a lambda term of type `List B`.

$$\begin{aligned} \text{data LTerm} &: \text{Set} \rightarrow \text{Set} \text{ where} \\ \text{var} &: \text{String} \rightarrow \text{LType } A \rightarrow \text{LTerm } A \\ \text{abs} &: \forall (B \ C : \text{Set}) \rightarrow \text{Equal } A \ (B \rightarrow C) \rightarrow \text{String} \rightarrow \text{LType } B \rightarrow \text{LTerm } C \rightarrow \text{LTerm } A \\ \text{app} &: \forall (B : \text{Set}) \rightarrow \text{LTerm } (B \rightarrow A) \rightarrow \text{LTerm } B \rightarrow \text{LTerm } A \\ \text{listC} &: \forall (B : \text{Set}) \rightarrow \text{Equal } A \ (\text{List } B) \rightarrow \text{List } (\text{LTerm } B) \rightarrow \text{LTerm } A \end{aligned} \tag{8}$$

The type parameter for `LTerm` tracks the types of simply typed lambda calculus terms. So the type `LTerm A` contains lambda terms of type `A`. Variables are represented as strings and are tagged with their type. The `var`

constructor takes a variable name and a type for the variable and produces a term of the given type. The **abs** constructor takes a variable name, a type B for the variable, and a lambda term of type C for the body of the abstraction, and it produces a term of type $B \rightarrow C$. The **app** constructor applies a lambda term of type $B \rightarrow A$ to a lambda term of type B , producing a lambda term of type A . The **listC** constructor takes a list of lambda terms of type B and produces a lambda term of type $\text{List } B$.

Variables are tagged with their types in the **var** (variable introduction) and **abs** (abstraction) constructors. The role of **LType** in the **var** and **abs** constructors is to enforce that variables can only be tagged with a legal type, e.g., **Bool**, $\text{Bool} \rightarrow \text{Bool}$, $\text{List } (\text{Bool} \rightarrow \text{Bool})$, etc. Therefore the presence of **LType** in these constructors ensures that all lambda terms produced by **var**, **abs**, **app**, and **listC** are well-typed.

4 (Deep) induction for GADTs

Reasoning about GADTs can generally be quite difficult, even more so than for ADTs and nested types. To help with that, we want to extend the deep induction principle seen in Section 2.2 to GADTs.

Induction rules, and specifically deep induction rules for nested types, are traditionally derived using the functorial semantics of data types in the setting of a parametric model [5]. In particular, relational parametricity is used to validate the induction principle because induction is, itself, a form of unary parametricity, where binary relations have been replaced with predicates, which are essentially unary relations. [Is there a reference for this?](#)

Unfortunately, this approach cannot possibly be employed to prove a deep induction rule for GADTs, as these types do not allow for a functorial interpretation, at least in a parametric model [?].

Nevertheless, this paper shows how to extend deep induction to some GADTs. We will first demonstrate how to derive the deep induction rule for some example cases, and then provide a general principle that works for generic GADTs, as long as these do not feature nesting in their definition.

4.1 (Deep) induction for Equal

As a first example, we derive the induction rule for the **Equal** type from example 3.1. This will provide a simple case study that will inform the investigation of more complex GADTs. Moreover, since we define GADTs using the **Equal** type, as for example in Equation 5, this example will be instrumental in stating and deriving the induction rule of other GADTs.

To define an induction rule for a type G we first need a predicate-lifting operation which takes predicates on a type A and lifts them to predicates on GA . the predicate-lifting function for **Equal** is the function

$$\text{Equal}^\wedge : \forall (A B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}$$

defined as

$$\text{Equal}^\wedge A A Q Q' \text{ Refl} = \forall (a : A) \rightarrow \text{Equal } (Q a) (Q' a)$$

i.e., the function that takes two predicates on the same type and tests them for extensional equality.

Next, we need to associate each constructor of the GADT under consideration to the expression that a given predicate is preserved by such constructor. Let **dIndRefl** be the following function associated to the **refl** constructor:

$$\begin{aligned} \lambda (P : \forall (A B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}) \\ \rightarrow \forall (C : \text{Set}) (Q Q' : C \rightarrow \text{Set}) \rightarrow \text{Equal}^\wedge C C Q Q' \text{ refl} \rightarrow P C C Q Q' \text{ refl} \end{aligned}$$

The induction rule states that, if a predicate is preserved by all of the constructors of the GADT under consideration, then the predicate is satisfied by any element of the GADT. The induction rule for **Equal** is thus the type

$$\begin{aligned} \forall (P : \forall (A B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}) \\ \rightarrow \text{dIndRefl } P \rightarrow \forall (A B : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (e : \text{Equal } A B) \rightarrow \text{Equal}^\wedge A B Q_A Q_B e \rightarrow P A B Q_A Q_B e \end{aligned}$$

To validate the induction rule we need to provide it with a witness, i.e., we need to show that the associated type is inhabited. We thus need to define a term **dIndEqual** of the above type, i.e., we need to define

$$\text{dIndEqual } P \text{ crefl } A B Q_A Q_B e \text{ liftE} : P A B Q_A Q_B e$$

for $\text{crefl} : \text{dIndRefI } P, Q_A : A \rightarrow \text{Set}, Q_B : B \rightarrow \text{Set}, e : \text{Equal } A B$ and $\text{liftE} : \text{Equal}^\wedge A B Q_A Q_B e$. By pattern-matching, we only need to consider the case where $A = B$ and $e = \text{refl}$, and define

$$\text{dIndEqual } P \text{crefl } A A Q_A Q'_A \text{refl liftE} = \text{crefl } A Q_A Q'_A \text{liftE}$$

Having provided a well-defined term for it, we have shown that the induction rule for **Equal** is sound.

The type **Equal** also has a standard structural induction rule indEqual ,

$$\forall (Q : \forall (A B : \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}) \rightarrow (\forall (C : \text{Set}) \rightarrow P C C \text{refl}) \rightarrow \forall (A B : \text{Set}) (e : \text{Equal } A B) \rightarrow P A B e$$

As is the case for ADTs and nested types, the structural induction rule for **Equal** is a consequence of the deep induction rule. Indeed, we can define indEqual as

$$\text{indEqual } Q \text{srefl } A B e = \text{dIndEqual } P \text{srefl } A B K_\top^A K_\top^B e \text{liftE}$$

where $Q : \forall (A B : \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}$, $\text{srefl} : \forall (C : \text{Set}) \rightarrow P C C \text{refl}$ and $e : \text{Equal } A B$, and

- $P : \forall (A B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow \text{Equal } A B \rightarrow \text{Set}$ is defined as $P A B Q_A Q_B e = Q A B e$;
- K_\top^A and K_\top^B are the constantly \top -valued predicates on, respectively, A and B ;
- $\text{liftE} : \text{Equal}^\wedge A B K_\top^A K_\top^B e$ is defined by pattern matching, i.e., in case $A = B$ and $e = \text{refl}$, it is defined as $\text{liftE } a = \text{refl} : \text{Equal } A A$ for every $a : A$.

That the structural induction rule is a consequence of the deep induction one is also true for all the examples below, even though we will not remark it every time.

4.2 (Deep) induction for Seq

Next, we shall provide an induction rule for the **Seq** type defined in Example 3.2. Again, the first step in deriving the induction rule for **Seq** consists in defining the predicate-lifting function over it,

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

which is given by pattern-matching as

$$\text{Seq}^\wedge A Q_A (\text{const } a) = Q_A a$$

where $Q_A : A \rightarrow \text{Set}$ and $a : A$, and

$$\begin{aligned} & \text{Seq}^\wedge A Q_A (\text{sPair } B C e s_B s_C) \\ &= \sum [Q_B : B \rightarrow \text{Set}] [Q_C : C \rightarrow \text{Set}] \text{Equal}^\wedge A (B \times C) Q_A (Q_B \times Q_C) e \times \text{Seq}^\wedge B Q_B s_B \times \text{Seq}^\wedge C Q_C s_C \end{aligned}$$

where $e : \text{Equal } A (B \times C)$, $s_B : \text{Seq } B$ and $s_C : \text{Seq } C$.

Finally, let dIndConst be the function

$$\lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq } A \rightarrow \text{Set}) \rightarrow \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (a : A) \rightarrow Q_A a \rightarrow P A Q_A (\text{Const } a)$$

associated to the **const** constructor, and let dIndSPair be the function

$$\begin{aligned} & \lambda (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq } A \rightarrow \text{Set}) \\ & \rightarrow \forall (A B C : \text{Set}) (Q_A : A \rightarrow \text{Set}) (Q_B : B \rightarrow \text{Set}) (Q_C : C \rightarrow \text{Set}) (s_B : \text{Seq } B) (s_C : \text{Seq } C) (e : \text{Equal } A (B \times C)) \\ & \rightarrow \text{Equal}^\wedge A (B \times C) Q_A (\text{Pair}^\wedge B C Q_B Q_C) e \rightarrow P B Q_B s_B \rightarrow P C Q_C s_C \rightarrow P A Q_A (\text{sPair } B C e s_B s_C) \end{aligned}$$

associated to the **sPair** constructor.

With these tools we can formulate an induction rule for **Seq**,

$$\begin{aligned} & \forall (P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq } A \rightarrow \text{Set}) \rightarrow C \text{Const } P \rightarrow C \text{sPair } P \\ & \rightarrow \forall (A : \text{Set}) (Q_A : A \rightarrow \text{Set}) (s_A : \text{Seq } A) \rightarrow \text{Seq}^\wedge A Q_A s_A \rightarrow P A Q_A s_A \end{aligned}$$

To validate the induction rule, we define a term dIndSeq for the above type. We have to define

$$\text{dIndSeq } P \text{ cconst cspair } A Q_A s_A \text{ liftA} : P A Q_A s_A$$

where $P : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Seq } A \rightarrow \text{Set}$, $\text{cconst} : \text{cConst } P$, $\text{cspair} : \text{cSPair } P$, $Q_A : A \rightarrow \text{Set}$, $s_A : \text{Seq } A$ and $\text{liftA} : \text{Seq}^A A Q_A s_A$, and we proceed by pattern-matching on s_A . Let $s_A = \text{const } a$ for $a : A$, and define

$$\text{dIndSeq } P \text{ cconst cspair } A Q_A (\text{const } a) \text{ liftA} = \text{cconst } A Q_A a \text{ liftA}$$

Notice that $\text{Seq}^A A Q_A (\text{const } a) = Q_A a$, and thus $\text{liftA} : Q_A a$, making the right-hand-side in the above expression type-check. Now, let $s_A = \text{sPair } B C e s_B s_C$ for $e : \text{Equal } A (B \times C)$, $s_B : \text{Seq } B$ and $s_C : \text{Seq } C$, and define

$$\text{dIndSeq } P \text{ cconst cspair } A Q_A (\text{sPair } B C e s_B s_C) (Q_B, Q_C, \text{liftE}, \text{liftB}, \text{liftC}) = \text{cspair } A B C Q_A Q_B Q_C s_B s_C e \text{ liftE } p_B p_C$$

where $(Q_B, Q_C, \text{liftE}, \text{liftB}, \text{liftC}) : \text{Seq}^A A Q (\text{sPair } B C e s_B s_C)$, i.e.,

- $Q_B : B \rightarrow \text{Set}$ and $Q_C : C \rightarrow \text{Set}$;
- $\text{liftE} : \text{Equal}^A A (B \times C) Q_A (Q_B \times Q_C) e$;
- $\text{liftB} : \text{Seq}^A B Q_B s_B$ and $\text{liftC} : \text{Seq}^A C Q_C s_C$;

and p_B and p_C are defined as follows:

$$\begin{aligned} p_B &= \text{DlSeq } P \text{ cconst cspair } B Q_B s_B \text{ liftB} : P B Q_B s_B \\ p_C &= \text{DlSeq } P \text{ cconst cspair } C Q_C s_C \text{ liftC} : P C Q_C s_C \end{aligned}$$

4.3 (Deep) induction for LTerm

We now present an induction rule for the LTerm type defined in Example 3.3. For that, we need liftings for both LType and LTerm . We will also need a lifting of predicates for arrow types, since arrow types appear in LType and LTerm . The lifting for arrow types

$$\text{Arr}^A : \forall (A B : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (B \rightarrow \text{Set}) \rightarrow (A \rightarrow B) \rightarrow \text{Set}$$

is defined as

$$\text{Arr}^A A B Q_A Q_B f = \forall (a : A) \rightarrow Q_A a \rightarrow Q_B (f a)$$

Now we can define the lifting for LType , which has type

$$\text{LType}^A : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LType } A \rightarrow \text{Set}$$

and is defined by pattern-matching as

$$\text{LType}^A A Q_A (\text{tBool } B e) = \sum [Q_B : B \rightarrow \text{Set}] \text{Equal}^A A B Q_A K_T^{\text{Bool}} e$$

where $Q_A : A \rightarrow \text{Set}$, K_T^{Bool} is the constantly T -valued predicate on Bool and $e : \text{Equal } A \text{ Bool}$;

$$\text{LType}^A A Q_A (\text{tArr } B C e T_B T_C)$$

$$= \sum [Q_B : B \rightarrow \text{Set}] [Q_C : C \rightarrow \text{Set}] \text{Equal}^A A (B \rightarrow C) Q_A (\text{Arr}^A B C Q_B Q_C) e \times \text{LType}^A B Q_B T_B \times \text{LType}^A C Q_C T_C$$

where $Q_A : A \rightarrow \text{Set}$, $e : \text{Equal } A (B \rightarrow C)$, $T_B : \text{LType } B$ and $T_C : \text{LType } C$; and

$$\text{LType}^A A Q_A (\text{tList } B e T_B) = \sum [Q_B : B \rightarrow \text{Set}] \text{Equal}^A A (\text{List } B) Q_A (\text{List}^A B Q_B) e \times \text{LType}^A B Q_B T_B$$

where $Q_A : A \rightarrow \text{Set}$, $e : \text{Equal } A (\text{List } B)$, and $T_B : \text{LType } B$. We can define the lifting for LTerm . Again, the lifting

$$\text{LTerm}^A : \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}$$

is defined by pattern-matching as

$$\text{LTerm}^\wedge A Q_A (\text{var } s T_A) = \text{LType}^\wedge A Q_A T_A$$

where $s : \text{String}$ and $T_A : \text{LType } A$;

$$\begin{aligned} & \text{LTerm}^\wedge A Q_A (\text{abs } B C \text{ es } T_B t_C) \\ &= \sum [Q_B : B \rightarrow \text{Set}] [Q_C : C \rightarrow \text{Set}] \text{Equal}^\wedge A (B \rightarrow C) Q_A (\text{Arr}^\wedge B C Q_B Q_C) e \times \text{LType}^\wedge B Q_B T_B \times \text{LTerm}^\wedge C Q_C t_C \end{aligned}$$

where $e : \text{Equal } A (B \rightarrow C)$, $s : \text{String}$, $T_B : \text{LType } B$, and $t_C : \text{LTerm } C$;

$$\text{LTerm}^\wedge A Q_A (\text{app } B t_{BA} t_B) = \sum [Q_B : B \rightarrow \text{Set}] \text{LTerm}^\wedge (B \rightarrow A) (\text{Arr}^\wedge B A Q_B Q_A) t_{BA} \times \text{LTerm}^\wedge B Q_B t_B$$

where $t_{BA} : \text{LTerm } (B \rightarrow A)$ and $t_B : \text{LTerm } B$; and

$$\text{LTerm}^\wedge A Q_A (\text{listC } B \text{ e ts}) = \sum [Q_B : B \rightarrow \text{Set}] \text{Equal}^\wedge A (\text{List } B) Q_A (\text{List}^\wedge B Q_B) e \times \text{List}^\wedge (\text{LTerm } B) (\text{LTerm}^\wedge B Q_B) ts$$

where $e : \text{Equal } A (\text{List } B)$ and $ts : \text{List } (\text{LTerm } B)$. Notice that we use the lifting for List in the case for the listC constructor.

With these liftings, we can define the deep induction principle for LTerm . Let dIndVar be the function

$$\begin{aligned} & \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}) \\ & \rightarrow \forall(A : \text{Set})(Q_A : A \rightarrow \text{Set})(s : \text{String})(T_A : \text{LType } A) \rightarrow \text{LType}^\wedge A Q_A T_A \rightarrow P A Q_A (\text{var } s T_A) \end{aligned}$$

associated to the var constructor. Let dIndAbs be the function

$$\begin{aligned} & \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}) \\ & \rightarrow \forall(A B C : \text{Set})(Q_A : A \rightarrow \text{Set})(Q_B : B \rightarrow \text{Set})(Q_C : C \rightarrow \text{Set})(e : \text{Equal } A (B \rightarrow C))(s : \text{String}) \\ & \rightarrow (T_B : \text{LType } B) \rightarrow (t_C : \text{LTerm } C) \rightarrow \text{Equal}^\wedge A (B \rightarrow C) Q_A (\text{Arr}^\wedge B C Q_B Q_C) e \\ & \rightarrow \text{LType}^\wedge B Q_B T_B \rightarrow P C Q_C t_C \rightarrow P A Q_A (\text{abs } B C \text{ es } T_B t_C) \end{aligned}$$

associated to the abs constructor. Let dIndApp be the function

$$\begin{aligned} & \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}) \\ & \rightarrow \forall(A B : \text{Set})(Q_A : A \rightarrow \text{Set})(Q_B : B \rightarrow \text{Set})(t_{BA} : \text{LTerm } (B \rightarrow A))(t_B : \text{LTerm } B) \\ & \rightarrow P (B \rightarrow A) (\text{Arr}^\wedge B A Q_B Q_A) t_{BA} \rightarrow P B Q_B t_B \rightarrow P A Q_A (\text{app } B t_{BA} t_B) \end{aligned}$$

associated to the app constructor. Let dIndListC be the function

$$\begin{aligned} & \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}) \\ & \rightarrow \forall(A B : \text{Set})(Q_A : A \rightarrow \text{Set})(Q_B : B \rightarrow \text{Set})(e : \text{Equal } A (\text{List } B))(ts : \text{List } (\text{LTerm } B)) \\ & \rightarrow \text{Equal}^\wedge A (\text{List } B) Q_A (\text{List}^\wedge B Q_B) e \rightarrow \text{List}^\wedge (\text{LTerm } B) (P B Q_B) ts \rightarrow P A Q_A (\text{listC } B \text{ e ts}) \end{aligned}$$

associated to the listC constructor.

The type of the deep induction principle for LTerm is

$$\begin{aligned} & \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{LTerm } A \rightarrow \text{Set}) \rightarrow \text{dIndVar } P \rightarrow \text{dIndAbs } P \rightarrow \text{dIndApp } P \rightarrow \text{dIndListC } P \\ & \rightarrow \forall(A : \text{Set})(Q_A : A \rightarrow \text{Set})(t_A : \text{LTerm } A) \rightarrow \text{LTerm}^\wedge A Q_A t_A \rightarrow P A Q_A t_A \end{aligned}$$

To prove the induction principle, we define a term dIndLTerm for it. We have to define

$$\text{dIndLTerm } P \text{ cvar cabs capp clistc } A Q_A t_A \text{ liftA} : P A Q_A t_A$$

where $\text{cvar} : \text{dIndVar}$, $\text{cabs} : \text{dIndAbs}$, $\text{capp} : \text{dIndApp}$, $\text{clstc} : \text{dIndListC}$, $Q_A : A \rightarrow \text{Set}$, $t_A : \text{LTerm } A$, and $\text{lift}_A : \text{LTerm}^\wedge A Q_A t_A$. As before, we prove the induction principle by pattern matching on t_A . For the var constructor case, let $t_A = \text{var } s T_A$ and define

$$\text{dIndLTerm } P \text{ cvar cabs capp clstc } A Q_A (\text{var } s T_A) \text{ lift}_A = \text{cvar } A Q_A s T_A \text{ lift}_A$$

for $s : \text{String}$ and $T_A : \text{LType } A$. Notice that $\text{LTerm}^\wedge A Q_A (\text{var } s T_A) = \text{LType}^\wedge A Q_A T_A$, so $\text{lift}_A : \text{LType}^\wedge A Q_A T_A$. For the abs case, let $t_A = \text{abs } B C e s T_B t_C$ and define

$$\begin{aligned} \text{dIndLTerm } P \text{ cvar cabs capp clstc } A Q_A (\text{abs } B C e s T_B t_C) (Q_B, Q_C, \text{lift}_E, \text{lift}_{T_B}, \text{lift}_{t_C}) \\ = \text{cabs } A B C Q_A Q_B Q_C e s T_B t_C \text{ lift}_E \text{ lift}_{T_B} p_C \end{aligned}$$

where $(Q_B, Q_C, \text{lift}_E, \text{lift}_{T_B}, \text{lift}_{t_C}) : \text{LTerm}^\wedge A Q_A (\text{abs } B C e s T_B t_C)$, i.e.,

- $Q_B : B \rightarrow \text{Set}$;
- $Q_C : C \rightarrow \text{Set}$;
- $\text{lift}_E : \text{Equal}^\wedge A (B \rightarrow C) Q_A (\text{Arr}^\wedge B C Q_B Q_C) e$;
- $\text{lift}_{T_B} : \text{LType}^\wedge B Q_B T_B$;
- $\text{lift}_{t_C} : \text{LTerm}^\wedge C Q_C T_C$;

and p_C is defined as:

$$p_C = \text{dIndLTerm } P \text{ cvar cabs capp clstc } C Q_C t_C \text{ lift}_{t_C} : P C Q_C t_C$$

For the app case, let $t_A = \text{app } B t_{BA} t_B$ and define

$$\text{dIndLTerm } P \text{ cvar cabs capp clstc } A Q_A (\text{app } B t_{BA} t_B) (Q_B, \text{list}_{t_{BA}}, \text{list}_{t_B}) = \text{capp } A B Q_A Q_B t_{BA} t_B p_{BA} p_B$$

where $(Q_B, \text{list}_{t_{BA}}, \text{list}_{t_B}) : \text{LTerm}^\wedge A Q_A (\text{app } B t_{BA} t_B)$, i.e.,

- $Q_B : B \rightarrow \text{Set}$;
- $\text{list}_{t_{BA}} : \text{LTerm}^\wedge (B \rightarrow A) (\text{Arr}^\wedge B A Q_B Q_A) t_{BA}$;
- $\text{list}_{t_B} : \text{LTerm}^\wedge B Q_B t_B$;

and p_{BA} and p_B are defined as:

$$\begin{aligned} p_{BA} &= \text{dIndLTerm } P \text{ cvar cabs capp clstc } (B \rightarrow A) (\text{Arr}^\wedge B A Q_B Q_A) t_{BA} \text{ lift}_{t_{BA}} : P (B \rightarrow A) (\text{Arr}^\wedge B A Q_B Q_A) t_{BA} \\ p_B &= \text{dIndLTerm } P \text{ cvar cabs capp clstc } B Q_B t_B \text{ lift}_{t_B} : P B Q_B t_B \end{aligned}$$

Finally, for the listC case, let $t_A = \text{listC } B e ts$ and define

$$\text{dIndLTerm } P \text{ cvar cabs capp clstc } A Q_A (\text{listC } B e ts) (Q_B, \text{lift}_E, \text{lift}_{\text{List}}) = \text{clstc } A B Q_A Q_B e ts \text{ lift}_E p_{\text{List}}$$

where $(Q_B, \text{lift}_E, \text{lift}_{\text{List}}) : \text{LTerm}^\wedge A Q_A (\text{listC } B e ts)$, i.e.,

- $Q_B : B \rightarrow \text{Set}$;
- $\text{list}_E : \text{Equal}^\wedge A (\text{List } B) Q_A (\text{List}^\wedge B Q_B) e$;
- $\text{lift}_{\text{List}} : \text{List}^\wedge (\text{LTerm } B) (\text{LTerm}^\wedge B Q_B) ts$;

and p_{List} is defined as:

$$p_{\text{List}} = \text{liftListMap } (\text{LTerm } B) (\text{LTerm}^\wedge B Q_B) (P B Q_B) p_{ts} \text{ lift}_{\text{List}} : \text{List}^\wedge (\text{LTerm } B) (P B Q_B) ts$$

and p_{ts} is defined as:

$$p_{ts} = \text{dIndLTerm } P \text{ cvar cabs capp clstc } B Q_B : \text{PredMap } (\text{LTerm } B) (\text{LTerm}^\wedge B Q_B) (P B Q_B)$$

where

$$\text{liftListMap} : \forall (A : \text{Set}) \rightarrow (Q_A Q'_A : A \rightarrow \text{Set}) \rightarrow \text{PredMap } A Q_A Q'_A \rightarrow \text{PredMap } (\text{List } A) (\text{List}^\wedge A Q_A) (\text{List}^\wedge A Q'_A)$$

takes a morphism of predicates and produces a morphism of lifted predicates. **PredMap** is the type constructor that produces the type of morphisms between predicates, and it is defined as follows:

$$\begin{aligned} \text{PredMap} &: \forall (A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow \text{Set} \\ \text{PredMap } A \, Q_A \, Q'_A &= \forall (a : A) \rightarrow Q_A \, a \rightarrow Q'_A \, a \end{aligned}$$

4.4 General case

Finally, we generalize the approach taken in the previous examples and provide a general framework to derive induction rules for arbitrary GADTs. For that, we need to give a grammar for the types we will be considering. A generic GADT

$$\begin{aligned} \text{data } G : \text{Set}^\alpha \rightarrow \text{Set} \text{ where} \\ c_i : F_i \, G \, \overline{B} \rightarrow G(\overline{K_i} \, \overline{B}) \end{aligned}$$

is defined by a finite number of constructors c_i . In the definition above, F_i is a type constructor with signature $(\text{Set}^\alpha \rightarrow \text{Set}) \rightarrow \text{Set}^\beta \rightarrow \text{Set}$ and each K_i is a type constructor with signature $\text{Set}^\beta \rightarrow \text{Set}$ (i.e. a type constructor of arity β). The overline notation denotes a finite list: each of the α -many K_i is a type constructor of arity β so that it can be applied to the list of types \overline{B} of length β . Moreover, notice that the arity of G matches the number of type constructors K_i . We allow each F_i to be inductively built in the following ways (and with the following restrictions):

- $F_i = F'_i \times F''_i$ where F'_i and F''_i have the same signature as F_i and are built recursively from the same induction rules.
- $F_i = F'_i + F''_i$ where F'_i and F''_i have the same signature as F_i and are built recursively from the same induction rules.
- $F_i = F'_i \rightarrow F''_i$ where F'_i does not contain the recursive variable, i.e., $F'_i : \text{Set}^\beta \rightarrow \text{Set}$ is a type constructor of arity β , and F''_i has the same signature as F_i and is built recursively from the same induction rules.
- $F_i \, G \, \overline{B} = G(\overline{F_j} \, \overline{B})$ where none of the α -many F_j contains the recursive variable, i.e., $F_j : \text{Set}^\beta \rightarrow \text{Set}$ is a type constructor of arity β for each $j = 1, \dots, \alpha$. Such restriction is necessary to prevent nesting, as that would break the induction rule as discussed in Section 5.
- $F_i \, G \, \overline{B} = H \, \overline{B}$ where H is a type constructor of arity β not containing the recursive variable, i.e., $H : \text{Set}^\beta \rightarrow \text{Set}$. Notice that this covers the case in which F_i is a closed type, so, in particular, the unit and empty types, 1 and 0, and the case in which F_i is a variable.
- $F_i \, G \, \overline{B} = H(\overline{F_k} \, \overline{G} \, \overline{B})$ where H is a γ -ary type constructor not containing the recursive variable, i.e., $H : \text{Set}^\gamma \rightarrow \text{Set}$, and F_k has the same signature as F_i and is built recursively from the same induction rules, for every $k = 1 \dots \gamma$. Moreover, we require that H is not a GADT itself (but we allow it to be an ADT or even a nested type). This way we know that H admits functorial semantics [6], and thus there is a map function for H^\wedge ,

$$\text{HLM}ap : \forall (\overline{C} : \text{Set}) (\overline{Q_C} \, Q'_C : C \rightarrow \text{Set}) \rightarrow \overline{\text{PredMap } C \, Q_C \, Q'_C} \rightarrow \text{PredMap } (H \, \overline{C}) \, (H^\wedge \, \overline{C} \, Q_C) \, (H^\wedge \, \overline{C} \, Q'_C)$$

A concrete way to define **HLM**ap is to proceed by induction on the structure of the type H , and give an inductive definition when H is an ADT or a nested type. Such details are not essential to the present discussion, and thus we omit them.

We can summarize the above inductive definition with the following grammar (but beware that the above restrictions and requirements still apply):

$$F_i \, G \, \overline{B} := F'_i \, G \, \overline{B} \times F''_i \, G \, \overline{B} \mid F'_i \, G \, \overline{B} + F''_i \, G \, \overline{B} \mid F'_i \, \overline{B} \rightarrow F''_i \, G \, \overline{B} \mid G(\overline{F_j} \, \overline{B}) \mid H \, \overline{B} \mid H(\overline{F_k} \, \overline{G} \, \overline{B})$$

A further requirement that applies to all of the types appearing above, including the types K_i , is that every type needs to have a predicate-lifting function. This is not an overly restrictive condition, though: all types made by sums, products, arrow types and type application do, and so do GADTs as defined above. A concrete way to define the predicate-lifting function for a type is to proceed by induction on the structure of the type, and we have seen in the previous sections examples of how to do so for products and type application. We do not give here the general definition of lifting, as that would require to first present a full type calculus, and that is beyond the scope of the paper.

Consider a generic GADT as defined above,

$$\begin{aligned} \text{data } G : \text{Set} \rightarrow \text{Set} \text{ where} \\ c : F G \bar{B} \rightarrow G(K \bar{B}) \end{aligned} \tag{9}$$

which, for ease of notation, we assume to be a unary type constructor (i.e. it depends on a single type parameter A) and to have only one constructor c . Extending the argument to GADTs of arbitrary arity and with multiple constructors presents no difficulty other than heavier notation. In the definition above, F has signature $(\text{Set} \rightarrow \text{Set}) \rightarrow \text{Set}^\beta \rightarrow \text{Set}$ and each K has signature $\text{Set}^\beta \rightarrow \text{Set}$. The constructor c can be rewritten using the `Equal` type as

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

which is the form we shall use from now on.

In order to state the induction rule for G , we first need to define G 's associated predicate-lifting function

$$G^\wedge : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}$$

as

$$G^\wedge A Q_A (c \bar{B} \text{ex}) = \sum [\bar{Q}_B : \bar{B} \rightarrow \text{Set}] \text{Equal}^\wedge A (K \bar{B}) Q_A (K^\wedge \bar{B} \bar{Q}_B) e \times F^\wedge G \bar{B} G^\wedge \bar{Q}_B x$$

where $Q_A : A \rightarrow \text{Set}$ and $c \bar{B} \text{ex} : G A$, i.e., $e : \text{Equal } A (K \bar{B})$ and $x : F G \bar{B}$. As already mentioned before, we also assume to have liftings for F ,

$$F^\wedge : \forall(G : \text{Set}^\alpha \rightarrow \text{Set})(\bar{B} : \text{Set}) \rightarrow (\forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \rightarrow (\bar{B} \rightarrow \text{Set}) \rightarrow F G \bar{B} \rightarrow \text{Set}$$

and for K ,

$$K^\wedge : \forall(\bar{B} : \text{Set}) \rightarrow (\bar{B} \rightarrow \text{Set}) \rightarrow K \bar{B} \rightarrow \text{Set}$$

Finally, associate the function

$$\begin{aligned} \text{dIndC} &= \lambda(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \\ &\rightarrow \forall(A : \text{Set})(\bar{B} : \text{Set})(Q_A : A \rightarrow \text{Set})(\bar{Q}_B : \bar{B} \rightarrow \text{Set})(e : \text{Equal } A (K \bar{B})) (x : F G \bar{B}) \\ &\rightarrow \text{Equal}^\wedge A (K \bar{B}) Q_A (K^\wedge \bar{B} \bar{Q}_B) e \rightarrow F^\wedge G \bar{B} P \bar{Q}_B x \rightarrow P A Q_A (c \bar{B} \text{ex}) \end{aligned}$$

to the constructor c .

The induction rule for G is

$$\forall(P : \forall(A : \text{Set}) \rightarrow (A \rightarrow \text{Set}) \rightarrow G A \rightarrow \text{Set}) \rightarrow \text{dIndC } P \rightarrow \forall(A : \text{Set})(Q_A : A \rightarrow \text{Set})(y : G A) \rightarrow G^\wedge A Q_A y \rightarrow P A Q_A y$$

As we already did in the previous examples, we validate the induction rule by providing a term dIndG for the type above. Define

$$\text{dIndG } P \text{cc } A Q_A (c \bar{B} \text{ex}) (\bar{Q}_B, \text{liftE}, \text{liftF}) = \text{cc } A \bar{B} Q_A \bar{Q}_B \text{ex liftE } (p \times \text{liftF})$$

where $\text{cc} : \text{dIndC } P$ and

- $c \bar{B} \text{ex} : G A$, i.e., $e : \text{Equal } A (K \bar{B})$, and $x : F G \bar{B}$;
- $(\bar{Q}_B, \text{liftE}, \text{liftF}) : G^\wedge A Q_A (c \bar{B} \text{ex})$, i.e., $\bar{Q}_B : \bar{B} \rightarrow \text{Set}$ for each \bar{B} , $\text{liftE} : \text{Equal}^\wedge A (K \bar{B}) Q_A (K^\wedge \bar{B} \bar{Q}_B) e$, and $\text{liftF} : F^\wedge G \bar{B} G^\wedge \bar{Q}_B x$.

Finally, the morphism of predicates

$$p : \text{PredMap } (F G \bar{B}) (F^\wedge G \bar{B} G^\wedge \bar{Q}_B) (F^\wedge G \bar{B} P \bar{Q}_B)$$

is defined by structural induction on F as follows:

- Case $F = F_1 \times F_2$ where F_1 and F_2 have the same signature as F . We have that

$$F^\wedge G \bar{B} P \bar{Q}_B = \text{Pair}^\wedge(F_1 G \bar{B})(F_2 G \bar{B})(F_1^\wedge G \bar{B} P \bar{Q}_B)(F_2^\wedge G \bar{B} P \bar{Q}_B)$$

By inductive hypothesis, there exist morphisms of predicates

$$\begin{aligned} p_1 &: \text{PredMap}(F_1 G \bar{B})((F_1^\wedge G \bar{B} G^\wedge \bar{Q}_B)((F_1^\wedge G \bar{B} P \bar{Q}_B)) \\ p_2 &: \text{PredMap}(F_2 G \bar{B})((F_2^\wedge G \bar{B} G^\wedge \bar{Q}_B)((F_2^\wedge G \bar{B} P \bar{Q}_B)) \end{aligned}$$

Thus, we define $p(x_1, x_2)(\text{lift}F_1, \text{lift}F_2) = (p_1 x_1 \text{lift}F_1, p_2 x_2 \text{lift}F_2)$ for $x_1 : F_1 G \bar{B}$, $\text{lift}F_1 : F_1^\wedge G \bar{B} G^\wedge \bar{Q}_B x_1$, $x_2 : F_2 G \bar{B}$ and $\text{lift}F_2 : F_2^\wedge G \bar{B} G^\wedge \bar{Q}_B x_2$.

- Case $F = F_1 + F_2$ where F_1 and F_2 have the same signature as F . Analogous to case $F = F_1 \times F_2$.
- Case $F = F_1 \rightarrow F_2$ where F_1 does not contain the recursive variable, i.e., $F_1 : \text{Set}^\beta \rightarrow \text{Set}$, and F_2 has the same signature as F . We have that

$$F^\wedge G \bar{B} P \bar{Q}_B x = \forall(z : F_1 \bar{B}) \rightarrow F_1^\wedge \bar{B} \bar{Q}_B z \rightarrow F_2^\wedge G \bar{B} P \bar{Q}_B (x z)$$

where $x : F G \bar{B} = F_1 \bar{B} \rightarrow F_2 G \bar{B}$. By inductive hypothesis, there exist a morphism of predicates

$$p_2 : \text{PredMap}(F_2 G \bar{B})(F_2^\wedge G \bar{B} G^\wedge \bar{Q}_B)(F_2^\wedge G \bar{B} P \bar{Q}_B)$$

Thus, we define $p x \text{lift}F : F^\wedge G \bar{B} P \bar{Q}_B x$ for $\text{lift}F : F^\wedge G \bar{B} G^\wedge \bar{Q}_B x$ as $p x \text{lift}F z \text{lift}F_1 = p_2(x z)(\text{lift}F z \text{lift}F_1)$ for $z : F_1 \bar{B}$ and $\text{lift}F_1 : F_1^\wedge \bar{B} \bar{Q}_B z$. Notice that F_1 not containing the recursive variable is a necessary restriction, as the proof relies on $F^\wedge G \bar{B} G^\wedge \bar{Q}_B x$ and $F^\wedge G \bar{B} P \bar{Q}_B x$ having the same domain $F_1^\wedge \bar{B} \bar{Q}_B z$.

- Case $F G \bar{B} = G(F' \bar{B})$ where F' does not contain the recursive variable, i.e., $F' : \text{Set}^\beta \rightarrow \text{Set}$. Thus, $F^\wedge G \bar{B} P \bar{Q}_B = P(F' \bar{B})(F'^\wedge \bar{B} \bar{Q}_B)$. So, p is defined as

$$p = \text{dInd} G P \text{cc}(F' \bar{B})(F'^\wedge \bar{B} \bar{Q}_B)$$

- Case $F G \bar{B} = H \bar{B}$ where H is a β -ary type constructor not containing the recursive variable, i.e., $H : \text{Set}^\beta \rightarrow \text{Set}$. In such case, $p : \text{PredMap}(H \bar{B})(H^\wedge \bar{B} \bar{Q}_B)(H^\wedge \bar{B} \bar{Q}_B)$ is just the identity morphism of predicates.
- Case $F G \bar{B} = H(F_k G \bar{B})$ where H is a γ -ary type constructor not containing the recursive variable, i.e., $H : \text{Set}^\gamma \rightarrow \text{Set}$, and F_k has the same signature as F , for every $k = 1 \dots \gamma$. Moreover, we assume that H has an associated predicate-lifting function,

$$H^\wedge : \forall(\bar{C} : \text{Set}) \rightarrow (\bar{C} \rightarrow \text{Set}) \rightarrow H \bar{C} \rightarrow \text{Set}$$

and that this predicate-lifting function has a map function HLMMap of type

$$\forall(\bar{C} : \text{Set})(\bar{Q}_C \bar{Q}'_C : \bar{C} \rightarrow \text{Set}) \rightarrow \text{PredMap} \bar{C} \bar{Q}_C \bar{Q}'_C \rightarrow \text{PredMap}(H \bar{C})(H^\wedge \bar{C} \bar{Q}_C)(H^\wedge \bar{C} \bar{Q}'_C)$$

That means that H cannot be a GADT, as GADTs have no functorial semantics [?] and incur in the issue exposed in Section 5, but it can be an ADT or even a nested type as those types have functorial semantics [5,4]. Thus,

$$F^\wedge G \bar{B} P \bar{Q}_B = H^\wedge(\bar{F}_k G \bar{B})(\bar{F}_k^\wedge G \bar{B} P \bar{Q}_B)$$

By induction hypothesis, there is a morphism of predicates

$$p_k : \text{PredMap}(F_k G \bar{B})(F_k^\wedge G \bar{B} G^\wedge \bar{Q}_B)(F_k^\wedge G \bar{B} P \bar{Q}_B)$$

for every $k = 1 \dots \gamma$. So, p is defined as

$$p = \text{HLMMap}(\bar{F}_k G \bar{B})(\bar{F}_k^\wedge G \bar{B} G^\wedge \bar{Q}_B)(\bar{F}_k^\wedge G \bar{B} P \bar{Q}_B) \bar{p}_k$$

5 Induction for GADTs with nesting

In the previous sections, we derive induction rules for examples of GADTs that do not feature nesting, in the sense that their constructors contain no nested calls of the recursive variable, as truly nested types (such as Bush, Equation 3) do. Since both nested types and GADTs without nesting admit induction rules, as seen in the previous sections, it is just natural to expect that GADTs with nesting would as well. Surprisingly, that is not the case: indeed, the induction principle generally relies on (unary) parametricity of the semantic interpretation, and in the case of nested types it also relies on functorial semantics [6], but GADTs cannot admit both functorial and parametric semantics at the same time [?]. In this section we show how induction for GADTs featuring nesting goes wrong by analyzing the following concrete example of such a type.

$$\begin{aligned} \text{data } G(a : \text{Set}) : \text{Set where} \\ C : G(Ga) \rightarrow G(a \times a) \end{aligned} \tag{10}$$

The constructor C can be rewritten as

$$C : \exists(b : \text{Set}) \rightarrow \text{Equal } a(b \times b) \rightarrow G(Gb) \rightarrow Ga$$

which is the form we shall use from now on. The predicate-lifting function of G ,

$$G^\wedge : \forall(a : \text{Set}) \rightarrow (a \rightarrow \text{Set}) \rightarrow G\bar{a} \rightarrow \text{Set}$$

is defined as

$$G^\wedge a Q_a (C b e x) = \exists(Q_b : b \rightarrow \text{Set}) \rightarrow \text{Equal}^\wedge a(b \times b) Q_a (\text{Pair}^\wedge b b Q_b Q_b) e \times G^\wedge (Gb) (G^\wedge b Q_b) x$$

where $Q_a : a \rightarrow \text{Set}$, $e : \text{Equal } a(b \times b)$ and $x : G(Gb)$. Finally, let CC be the function

$$\begin{aligned} \lambda(P : \forall(a : \text{Set}) \rightarrow (a \rightarrow \text{Set}) \rightarrow Ga \rightarrow \text{Set}) \\ \rightarrow \forall(a b : \text{Set})(Q_a : a \rightarrow \text{Set})(Q_b : b \rightarrow \text{Set})(e : \text{Equal } a(b \times b))(x : G(Gb)) \\ \rightarrow \text{Equal}^\wedge a(b \times b) Q_a (\text{Pair}^\wedge b b Q_b Q_b) e \rightarrow P(Gb) (P b Q_b) x \rightarrow Pa Q_a (C b e x) \end{aligned}$$

associated to the C constructor.

The induction rule for G is

$$\begin{aligned} \forall(P : \forall(a : \text{Set}) \rightarrow (a \rightarrow \text{Set}) \rightarrow Ga \rightarrow \text{Set}) \rightarrow CCP \\ \rightarrow \forall(a : \text{Set})(Q_a : a \rightarrow \text{Set})(y : Ga) \rightarrow G^\wedge a Q_a y \rightarrow Pa Q_a y \end{aligned}$$

Consistently with the previous examples, to validate the induction rule we try to define a term of the above type, DIG , as

$$DIG P cc a Q_a (C b e x) (Q_b, L_E, L_G) = cc a b Q_a Q_b e \times L_E p$$

where $cc : CCP$ and

- $C b e x : Ga$, i.e., $e : \text{Equal } a(b \times b)$ and $x : G(Gb)$;
- $(Q_b, L_E, L_G) : G^\wedge a Q_a (C b e x)$, i.e., $Q_b : b \rightarrow \text{Set}$, $L_E : \text{Equal}^\wedge a(b \times b) Q_a (\text{Pair}^\wedge b b Q_b Q_b) e$, and $L_G : G^\wedge (Gb) (G^\wedge b Q_b) x$.

We still need to define $p : P(Gb) (P b Q_b) x$. We do so by using the induction rule and letting

$$p = DIG P cc (Gb) (P b Q_b) x q$$

where we still need to provide $q : G^\wedge (Gb) (P b Q_b) x$. If we had the map function of G^\wedge ,

$$GLMap : \forall(a : \text{Set})(Q_a Q'_a : a \rightarrow \text{Set}) \rightarrow \text{PredMap } a Q_a Q'_a \rightarrow \text{PredMap } (Ga) (G^\wedge a Q_a) (G^\wedge a Q'_a)$$

then we would be able to define

$$q = GLMap (Gb) (G^\wedge b Q_b) (P b Q_b) (DIG P cc b Q_b) x L_G$$

Unfortunately, we cannot define such a `GLMap`. Indeed, its definition would have to be

$$\text{GLMap } a \ Q_a \ Q'_a \ M \ (C \text{ b ex}) \ (Q_b, L_E, L_G) = (Q'_b, L'_E, L'_G)$$

where $Q_a : a \rightarrow \text{Set}$, $Q'_a : a \rightarrow \text{Set}$, $M : \text{PredMap } a \ Q_a \ Q'_a$, $C \text{ b ex} : G a$, i.e.,

- $e : \text{Equal } a \ (b \times b)$;
- $x : G \ (G b)$;
- $(Q_b, L_E, L_G) : G^\wedge a \ Q_a \ (C \text{ b ex})$, i.e.,
- $Q_b : b \rightarrow \text{Set}$;
- $L_E : \text{Equal}^\wedge a \ (b \times b) \ Q_a \ (\text{Pair}^\wedge b \ b \ Q_b \ Q_b) \ e$;
- $L_G : G^\wedge \ (G b) \ (G^\wedge b \ Q_b) \ x$;
- and $(Q'_b, L'_E, L'_G) : G^\wedge a \ Q'_a \ (C \text{ b ex})$, i.e.,
- $Q'_b : b \rightarrow \text{Set}$;
- $L'_E : \text{Equal}^\wedge a \ (b \times b) \ Q'_a \ (\text{Pair}^\wedge b \ b \ Q'_b \ Q'_b) \ e$;
- $L'_G : G^\wedge \ (G b) \ (G^\wedge b \ Q'_b) \ x$;

In other words, we have a proof L_E of the (extensional) equality of the predicates Q_a and $\text{Pair}^\wedge b \ b \ Q_b \ Q_b$ and a morphism of predicates M from Q_a to Q'_a , and we need to use those to deduce a proof of the (extensional) equality of the predicates Q'_a and $\text{Pair}^\wedge b \ b \ Q'_b \ Q'_b$, for some for some predicate Q'_b on b . But that is not generally possible: the facts that Q_a is equal to $\text{Pair}^\wedge b \ b \ Q_b \ Q_b$ and that there is a morphism of predicates M from Q_a to Q'_a do not guarantee that Q'_a is equal to $\text{Pair}^\wedge b \ b \ Q'_b \ Q'_b$ for some Q'_b .

At a deeper level, the fundamental issue is that the `Equal` type does not have functorial semantics, so that having morphisms $A \rightarrow A'$ and $B \rightarrow B'$ and a proof that A is equal to A' does not provide a proof that B is equal to B' . This is because GADTs can either have a syntax-only semantics or a functorial-completion semantics. Since we are interested in induction rules, we considered the syntax-only semantics, which is parametric but not functorial. Had we considered the functorial-completion semantics, which is functorial, we would have forfeited parametricity instead. In both cases, thus, we cannot derive an induction rule for GADTs featuring nesting. Unlike nestes types, indeed, GADTs do not admit a semantic interpretation that is both parametric and functorial [?].

6 Applications

Can get rid of `Maybe` using non-empty lists and postulates

In this section we use deep induction for the `LTerm` GADT to extract the type from a lambda term. We have a predicate

$$\begin{aligned} \text{getType} &: \forall (a : \text{Set}) \rightarrow (t : \text{LTerm } a) \rightarrow \text{Set} \\ \text{getType } a \ t &= \text{Maybe } (\text{LType } a) \end{aligned}$$

that takes a lambda term and produces its type (using `Maybe` to represent potential failure). We want to show this predicate is satisfied for every element of `LTerm a`. Because of the `ListC` constructor, this cannot be achieved without deep induction. In particular, deep induction is required to apply the induction to the individual terms in a list of terms.

So, using deep induction, we want to prove:

$$\text{getTypeProof} : \forall (a : \text{Set}) \rightarrow (t : \text{LTerm } a) \rightarrow \text{getType } a \ t$$

which we prove by

$$\text{getTypeProof } a \ t = \text{DILTerm } (\lambda b \ Q_b \ t \rightarrow \text{getType } b \ t) \ \text{gtVar} \ \text{gtABs} \ \text{gtApp} \ \text{gtListC} \ a \ K1 \ t \ (\text{LTerm}^\wedge K1 \ a \ t)$$

where $K1 : a \rightarrow \text{Set}$ is the constantly true predicate:

$$K1 \ x = \top$$

and $\text{LTerm}^\wedge \text{K1 a t} : \text{LTerm}^\wedge \text{a K1 t}$. Notice that there is no space in $\text{LTerm}^\wedge \text{K1}$, because

$$\text{LTerm}^\wedge \text{K1} : \forall (a : \text{Set}) (t : \text{LTermA}) \rightarrow \text{LTerm}^\wedge \text{a K1 t}$$

is a function that we will define. In addition to defining $\text{LTerm}^\wedge \text{K1}$, we also have to give a proof for each constructor `Var`, `Abs`, `App`, `ListC`:

$$\text{gtVar} : \forall (a : \text{Set}) (Q_a : a \rightarrow \text{Set}) (s : \text{String}) (T_a : \text{LType a}) \rightarrow \text{LType}^\wedge \text{a Q}_a T_a \rightarrow \text{Maybe (LType a)}$$

$$\begin{aligned} \text{gtAbs} : \forall (a \ b \ c : \text{Set}) (Q_a : a \rightarrow \text{Set}) (Q_b : b \rightarrow \text{Set}) (Q_c : c \rightarrow \text{Set}) (e : \text{Equal a (b} \rightarrow \text{c)}) (s : \text{String}) \\ (T_b : \text{LType b}) (t_c : \text{LTerm c}) \rightarrow \text{Equal}^\wedge \text{a (b} \rightarrow \text{c)} Q_a (\text{Arr}^\wedge \text{b c Q}_b Q_c) e \rightarrow \text{LType}^\wedge \text{b Q}_b T_b \\ \rightarrow \text{Maybe (LType c)} \rightarrow \text{Maybe (LType a)} \end{aligned}$$

$$\begin{aligned} \text{gtApp} : \forall (a \ b : \text{Set}) (Q_a : a \rightarrow \text{Set}) (Q_b : b \rightarrow \text{Set}) (t_{ba} : \text{LTerm (b} \rightarrow \text{a)}) (t_b : \text{LTerm b}) \\ \rightarrow \text{Maybe (LType (b} \rightarrow \text{a)})} \rightarrow \text{Maybe (LType b)} \rightarrow \text{Maybe (LType a)} \end{aligned}$$

$$\begin{aligned} \text{gtListC} : \forall (a \ b : \text{Set}) (Q_a : a \rightarrow \text{Set}) (Q_b : b \rightarrow \text{Set}) (e : \text{Equal a (List b)}) (ts : \text{List (LTerm b)}) \\ \rightarrow \text{Equal}^\wedge \text{a (List b)} Q_a (\text{List}^\wedge \text{b Q}_b) e \rightarrow \text{List}^\wedge (\text{LTerm b}) (\text{getType b}) ts \rightarrow \text{Maybe (LType a)} \end{aligned}$$

For variables we simply return the type T_a , and the cases for abstraction and application are similar. The interesting case is `gtListC`, in which we have to use the results of $(\text{List}^\wedge (\text{LTerm b}) (\text{getType b}) ts)$ in order to extract the type of one of the terms in the list. To define `gtListC` we pattern-match on the list of terms `ts`.

If `ts` is the empty list (denoted by `[]`), we cannot extract a type, so we return `nothing`.
Maybe handle this case differently, by using non-empty lists, for example

$$\text{gtListC a b Q}_a Q_b e [] L_e L_{ts} = \text{nothing}$$

If `ts` is a non-empty list, we pattern match on L_{ts} and use the result to construct the type we need:

$$\begin{aligned} \text{gtListC a b Q}_a Q_b e (t :: ts) L_e (\text{nothing}, L_{ts}) &= \text{nothing} \\ \text{gtListC a b Q}_a Q_b e (t :: ts) L_e (\text{just } T_b, L_{ts}) &= \text{just } (T \text{List b e } T_b) \end{aligned}$$

where $e : \text{Equal a (List b)}$ and $T_b : \text{LType b}$.

6.1 Defining $\text{LTerm}^\wedge \text{K1}$

Maybe this section can be deleted by assuming $\text{LTerm}^\wedge \text{a K1 t} = \text{K1}$. Maybe we can say this derives from parametricity. Currently we say that $\text{K1} \times \text{K1} = \text{K1}$ based on the fact that products are a built-in type and so this seems to be obviously true.

The last piece of infrastructure we need to define `getTypeProof` is a function

$$\text{LTerm}^\wedge \text{K1} : \forall (a : \text{Set}) \rightarrow (t : \text{LTermA}) \rightarrow \text{LTerm}^\wedge \text{a K1 t}$$

that provides a proof of $\text{LTerm}^\wedge \text{a K1 t}$ for any term $t : \text{LTermA}$. Because LTerm^\wedge is defined in terms of LType^\wedge , Arr^\wedge , and List^\wedge , we will need analogous functions for these liftings as well. We only give the definition of $\text{LTerm}^\wedge \text{K1}$, but the definitions for LType^\wedge , Arr^\wedge , and List^\wedge are analogous.

$\text{LTerm}^\wedge \text{K1}$ is defined by pattern matching on the lambda term t . For the `Var` case, let $t = (\text{Var s } T_a)$ and define

$$\text{LTerm}^\wedge \text{K1 a (Var s } T_a) = \text{LType}^\wedge \text{K1 a } T_a$$

For the **Abs** case, let $t = (\text{Abs } b \text{ c } e \text{ } T_b \text{ } t_c)$ and recall the definition of LTerm^\wedge for the **Abs** constructor, instantiating the predicate Q_a to $K1$:

$$\begin{aligned} \text{LTerm}^\wedge a \text{ } K1 (\text{Abs } b \text{ c } e \text{ } T_b \text{ } t_c) \\ = \exists (Q_b : b \rightarrow \text{Set}) (Q_c : c \rightarrow \text{Set}) \rightarrow \text{Equal}^\wedge a (b \rightarrow c) K1 (\text{Arr}^\wedge b \text{ c } Q_b \text{ } Q_c) e \\ \times \text{LType}^\wedge b \text{ } Q_b \text{ } T_b \times \text{LTerm}^\wedge c \text{ } Q_c \text{ } t_c \end{aligned}$$

so to define the **Abs** case of $\text{LTerm}^\wedge K1$, we need a proof of

$$\text{Equal}^\wedge a (b \rightarrow c) K1 (\text{Arr}^\wedge b \text{ c } Q_b \text{ } Q_c) e$$

i.e., that $K1$ is (extensionally) equal to the lifting $(\text{Arr}^\wedge b \text{ c } Q_b \text{ } Q_c)$ for some predicates Q_b, Q_c . The only reasonable choice for Q_b and Q_c is to let both be $K1$, which means we need a proof of:

$$\text{Equal}^\wedge a (b \rightarrow c) K1 (\text{Arr}^\wedge b \text{ c } K1 \text{ } K1) e$$

Since we are working with proof-relevant predicates (i.e., functions into **Set** rather than functions into **Bool**), the lifting $(\text{Arr}^\wedge b \text{ c } K1 \text{ } K1)$ of $K1$ to arrow types is not identical to $K1$ on arrow types, but the predicates are (extensionally) isomorphic. We discuss this issue in more detail at the end of the section. For now, we assume a proof

$$\text{Equal}^\wedge \text{Arr} K1 : \text{Equal}^\wedge a (b \rightarrow c) K1 (\text{Arr}^\wedge b \text{ c } K1 \text{ } K1) e$$

and define the **Abs** case of $\text{LTerm}^\wedge K1$ as

$$\text{LTerm}^\wedge K1 a (\text{Abs } b \text{ c } e \text{ } T_b \text{ } t_c) = (K1, K1, \text{Equal}^\wedge \text{Arr} K1, \text{LType}^\wedge K1 \text{ } b \text{ } T_b, \text{LTerm}^\wedge K1 \text{ } c \text{ } t_c)$$

For the **App** case, let $t = (\text{App } b \text{ } t_{ba} \text{ } t_b)$ and just as we did for the **Abs** case, recall the definition of $\text{LTerm}^\wedge a (\text{App } b \text{ } t_{ba} \text{ } t_b)$ with all of the predicates instantiated with $K1$:

$$\text{LTerm}^\wedge (b \rightarrow a) (\text{Arr}^\wedge b \text{ a } K1 \text{ } K1) t_{ba} \times \text{LTerm}^\wedge b \text{ } K1 \text{ } t_b$$

The second component can be given using $\text{LTerm}^\wedge K1$, and we can define the first component using a proof of

$$\text{LTerm}^\wedge (b \rightarrow a) K1 \text{ } t_{ba}$$

and a map-like function

$$\begin{aligned} \text{LTerm}^\wedge \text{EqualMap} : \forall \{a : \text{Set}\} \rightarrow (Q_a \text{ } Q'_a : a \rightarrow \text{Set}) \rightarrow (\text{Equal}^\wedge a \text{ } a \text{ } Q_a \text{ } Q'_a \text{ } \text{rfl}) \\ \rightarrow \text{PredMap} (\text{LTerm}^\wedge a \text{ } Q_a) (\text{LTerm}^\wedge a \text{ } Q'_a) \end{aligned}$$

that takes two (extensionally) equal predicates with the same carrier and produces a morphism of predicates between their liftings. The definition is straightforward enough, so we omit the details. But it is worth noting that while a true **HLM**ap function, which takes a *morphism* of predicates instead of a proof of equality, cannot be defined for GADTs in general, an analogue of $\text{LTerm}^\wedge \text{EqualMap}$ should be definable for every GADT. These analogues of $\text{LTerm}^\wedge \text{EqualMap}$ will be required to define $G^\wedge K1$ whenever G has a constructor of the form $(c : G (F b) \rightarrow G (K b))$.

Using $\text{LTerm}^\wedge \text{EqualMap}$, we can define the **App** case of $\text{LTerm}^\wedge K1$ as

$$\text{LTerm}^\wedge K1 a (\text{App } b \text{ } t_{ba} \text{ } t_b) = (K1, L_{\text{Arr}^\wedge K1}, \text{LTerm}^\wedge K1 \text{ } b \text{ } t_b)$$

where $L_{\text{Arr}^\wedge K1} : \text{LTerm}^\wedge (b \rightarrow a) (\text{Arr}^\wedge b \text{ a } K1 \text{ } K1) t_{ba}$ is defined as

$$L_{\text{Arr}^\wedge K1} = \text{LTerm}^\wedge \text{EqualMap} K1 (\text{Arr}^\wedge b \text{ a } K1 \text{ } K1) \text{Equal}^\wedge \text{Arr} K1 \text{ } t_{ba} \text{ } L_{K1}$$

where $L_{K1} = \text{LTerm}^\wedge K1 (b \rightarrow a) t_{ba} : \text{LTerm}^\wedge (b \rightarrow a) K1 \text{ } t_{ba}$.

Finally, we define the **ListC** case for $\text{LTerm}^\wedge K1$. Let $t = (\text{ListC } b \text{ e } t_s)$ and recall the definition of $\text{LTerm}^\wedge a (\text{ListC } b \text{ e } t_s)$ with all of the predicates instantiated to $K1$:

$$\text{Equal}^\wedge a (\text{List } b) K1 (\text{List}^\wedge b \text{ } K1) e \times \text{List}^\wedge (\text{LTerm}^\wedge b) (\text{LTerm}^\wedge b \text{ } K1) t_s$$

We can give the first component by assuming a proof $\text{Equal}^\wedge \text{ListK1} : \text{Equal}^\wedge a (\text{List } b) K1 (\text{List}^\wedge b K1) e$, but for the second component we again have multiple liftings nested together. In this case, we can get a proof of

$$\text{List}^\wedge (\text{LTerm } b) (\text{LTerm}^\wedge b K1) ts$$

using

$$\text{List}^\wedge \text{map} : \forall (a : \text{Set}) \rightarrow (Q_a Q'_a : a \rightarrow \text{Set}) \rightarrow \text{PredMap } Q_a Q'_a \rightarrow \text{PredMap } (\text{List}^\wedge a Q_a) (\text{List}^\wedge a Q'_a)$$

to map a morphism of predicates

$$\text{PredMap } (K1) (\text{LTerm}^\wedge b K1)$$

to a morphism of lifted predicates

$$\text{PredMap } (\text{List}^\wedge (\text{LTerm } b) K1) (\text{List}^\wedge (\text{LTerm } b) (\text{LTerm}^\wedge b K1))$$

We define the ListC case of $\text{LTerm}^\wedge K1$ as

$$\text{LTerm}^\wedge K1 a (\text{ListC } b e ts) = (K1, \text{Equal}^\wedge \text{ListK1}, L_{\text{List}^\wedge \text{LTerm}^\wedge K1})$$

where $L_{\text{List}^\wedge \text{LTerm}^\wedge K1} : \text{List}^\wedge (\text{LTerm } b) (\text{LTerm}^\wedge b K1) ts$

$$L_{\text{List}^\wedge \text{LTerm}^\wedge K1} = \text{List}^\wedge \text{map } (\text{LTerm } b) K1 (\text{LTerm}^\wedge b K1) m_{K1} ts (\text{List}^\wedge K1 (\text{LTerm } b) ts)$$

and $m_{K1} : \text{PredMap } (K1) (\text{LTerm}^\wedge b K1)$

$$m_{K1} t * = \text{LTerm}^\wedge K1 b t$$

where $*$ is the single element of $(K1 t)$. The use of $\text{List}^\wedge \text{map}$ is required in the ListC case because ListC takes an argument of type $\text{List } (\text{LTerm } b)$. The same technique can be used to define $G^\wedge K1$ whenever G has a constructor of the form $(c : F (G a) \rightarrow G (K b))$. We only allow constructors of this form when F is a nested type or ADT, so we are guaranteed to have a $F^\wedge \text{map}$ function.

6.2 Liftings of $K1$

To provide a proof of $G^\wedge a K1 t$ for every term $t : G a$, we need to know that the lifting of $K1$ by a type H is extensionally equal to $K1$ on H . For example, we might need a proof that $\text{Pair}^\wedge a b K1 K1$ is equal to the predicate $K1$ on pairs. Given a pair $(x, y) : a \times b$, we have

$$\begin{aligned} & \text{Pair}^\wedge a b K1 K1 (x, y) \\ &= K1 x \times K1 y \\ &= \top \times \top \end{aligned}$$

while

$$K1 (x, y) = \top$$

and while these types are not equal they are clearly isomorphic. So for simplicity of presentation, we assume $(F^\wedge a K1)$ is equal to $K1$ for every nested type and ADT F .

7 Conclusion/Related work

Mention Patricia/Neil2008 paper

No induction with primitive representation (reference Haskell Symposium paper and [5] and paper Patricia Neil Clement 2010)

8 TODO

- find correct entcsmacro file (current one is for 2018). Maybe ask Ana Sokolova (anas@cs.uni-salzburg.at).
- reference (correctly) Haskell Symposium paper
- reference inspiration for STLC GADT : <https://www.seas.upenn.edu/~cis194/spring15/lectures/11-stlc.html>

References

- [1] Atkey, R., *Relational parametricity for higher kinds*, in: *Computer Science Logic*, 2012, pp. 46–61.
- [2] Bird, R. and L. Meertens, *Nested datatypes*, in: *Mathematics of Program Construction*, 1998.
- [3] Cheney, J. and R. Hinze, *First-class phantom types*, Technical report, Cornell University (2003).
- [4] Johann, P., E. Ghiorzi and D. Jeffries, *Parametricity for primitive nested types*, in: *Foundations of Software Science and Computation Structures*, 2021, pp. 324–343.
- [5] Johann, P. and A. Polonsky, *Higher-kinded data types: Syntax and semantics*, in: *Logic in Computer Science*, 2019, pp. 1–13.
- [6] Johann, P. and A. Polonsky, *Deep induction: Induction rules for (truly) nested types*, in: *Foundations of Software Science and Computation Structures*, 2020, pp. 339–358.
- [7] Minsky, Y., *Why GADTs matter for performance*, <https://blog.janestreet.com/why-gadts-matter-for-performance/> (2015).
- [8] Pasalic, E. and N. Linger, *Meta-programming with typed object-language representations*, in: *Generic Programming and Component Engineering*, 2004, pp. 136–167.
- [9] Penner, C., *Simpler and safer API design using GADTs*, <https://chrispenner.ca/posts/gadt-design> (2020).
- [10] Peyton Jones, S. L., D. Vytiniotis, G. Washburn and S. Weirich, *Simple unification-based type inference for gadts*, in: *International Conference on Functional Programming*, 2006, pp. 50–61.
- [11] Pottier, F. and Y. Régis-Gianas, *Stratified type inference for generalized algebraic data types*, in: *Principles of Programming Languages*, 2006, pp. 232–244.
- [12] Sheard, T. and E. Pasalic, *Meta-programming with built-in type equality*, in: *Fourth International Workshop on Logical Frameworks and Meta-Languages*, 2004.
- [13] Vytiniotis, D. and S. Weirich, *Parametricity, type equality, and higher-order polymorphism* (2010).
- [14] Xi, H., C. Chen and G. Chen, *Guarded recursive datatype constructors*, in: *Proceedings of the 30th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, 2003, pp. 224–235.