

# Dialectica Interpretation with Marked Counterexamples

Trifon Trifonov\*

Faculty of Mathematics and Informatics, Sofia University

triffon@fmi.uni-sofia.bg

Gödel’s functional “Dialectica” interpretation can be used to extract functional programs from non-constructive proofs in arithmetic by employing two sorts of higher-order witnessing terms: positive realisers and negative counterexamples. In the original interpretation decidability of atoms is required to compute the correct counterexample from a set of candidates. When combined with recursion, this choice needs to be made for every step in the extracted program, however, in some special cases the decision on negative witnesses can be calculated only once. We present a variant of the interpretation in which the time complexity of extracted programs can be improved by marking the chosen witness and thus avoiding recomputation. The achieved effect is similar to using an abortive control operator to interpret computational content of non-constructive principles.

## 1 Introduction

Gödel’s Dialectica interpretation [5] is one of the first systematic methods for obtaining computational content from proofs in classical arithmetic. Different variants of the interpretation have been proposed to aid “proof mining” — looking for constructive information inside what appears to be a non-constructive argument (e.g. [9]). Particularly interesting is the case where terms of non-ground types (i.e., functional programs) are being automatically obtained from a weak existence proof. The extracted algorithms usually calculate the witness for the existential quantifier in an indirect manner, often obscure and surprising. There are also a number of competing techniques for program extraction like (refined)  $A$ -translation [4, 2], Krivine’s realisability, control operators [6], the  $\lambda\mu$ -calculus [11] and others. All of these methods systematically find correct programs, but the relations between them are still being investigated.

Another topic of ongoing research is whether such approaches are feasible for practical extraction of sufficiently efficient correct algorithms. Even though many automatic software systems for handling large proof objects are being actively developed, additional work is needed to identify and remove possible redundancies, so that the extracted programs are more readable, shorter and faster. Examples of such optimisations include uniform decorations [1, 7, 12], soundness-preserving program transformations [10], avoiding syntactic repetition [16]. In the present paper we suggest another such technique for the Dialectica interpretation, which marks computed counterexamples that are determined to be valid. We demonstrate how in certain cases this approach can reduce the average time complexity of the obtained program by terminating recursive search immediately after a correct counterexample is found. The suggested change is an extension of the interpretation variant given in [16]. The reason is that even though counterexample marking can be formally applied directly to the original Dialectica interpretation, its practical effects are only visible when (at least) syntactic repetition is avoided.

---

\*The author gratefully acknowledges financial support by the Bulgarian National Science Fund within project DO 02-102/23.04.2009 and by the European Social Fund within project BG 051PO001-3.3.04/28.08.2009.

## 2 Negative Arithmetic

We work in a restriction of Heyting Arithmetic with finite types (denoted  $\text{HA}^\omega$  in [17]) to the language of  $\rightarrow$  and  $\forall$ . We refer to the resulting system as Negative Arithmetic ( $\text{NA}^\omega$ ).

**Definition 1.** Types  $(\rho, \sigma)$ , (object) terms  $(s, t)$  and formulas  $(A, B)$  are defined as follows:

$$\begin{aligned}
\rho, \sigma & ::= \text{B} \mid \text{N} \mid \alpha \mid \rho \Rightarrow \sigma \mid \rho \times \sigma \\
s, t & ::= x^\rho \mid (\lambda x^\rho t^\sigma)^{\rho \Rightarrow \sigma} \mid (s^{\rho \Rightarrow \sigma} t^\rho)^\sigma \mid \langle s^\rho, t^\sigma \rangle^{\rho \times \sigma} \mid (t^{\rho \times \sigma} \_ )^\rho \mid (t^{\rho \times \sigma} \_ )^\sigma \mid \\
& \quad \text{tt}^\text{B} \mid \text{ff}^\text{B} \mid 0^\text{N} \mid \text{S}^{\text{N} \Rightarrow \text{N}} \mid \mathcal{C}^{\text{B} \Rightarrow \sigma \Rightarrow \sigma \Rightarrow \sigma} \mid \mathcal{R}^{\text{N} \Rightarrow \sigma \Rightarrow (\text{N} \Rightarrow \sigma \Rightarrow \sigma) \Rightarrow \sigma} \\
A, B & ::= \text{at}(t^\text{B}) \mid A \rightarrow B \mid \forall x^\rho A
\end{aligned}$$

The base types of booleans  $\text{B}$  and natural numbers  $\text{N}$  are equipped with the usual constructors and structural recursor constants. Here  $x$  denotes a typed object variable and  $\alpha$  denotes a type variable. Freely occurring type variables allow for a restricted form of polymorphism. The sets of free variables  $\text{FV}(t)$ ,  $\text{FV}(A)$  and bound variables  $\text{BV}(t)$ ,  $\text{BV}(A)$  are defined inductively as usual. Substitution of terms for object variables  $s[x := t]$ ,  $A[x := t]$  is by default assumed to be capture-free with respect to abstraction and quantification.

The operational semantics of object terms are given by the usual  $\beta$ -reduction rules and computation rules for the recursor constants:

$$\begin{array}{lll}
\langle s, t \rangle \_ \llcorner & \xrightarrow{r} & s \\
\langle s, t \rangle \_ \lrcorner & \xrightarrow{r} & t \\
(\lambda x s) t & \xrightarrow{r} & s[x := t] \\
\mathcal{C} \text{tt} t_1 t_2 & \xrightarrow{r} & t_1 \\
\mathcal{C} \text{ff} t_1 t_2 & \xrightarrow{r} & t_2 \\
\mathcal{R} 0 s t & \xrightarrow{r} & s \\
\mathcal{R} (\text{Sn}) s t & \xrightarrow{r} & t n (\mathcal{R} n s t)
\end{array}$$

We will make use of the following “let” notation for a  $\beta$ -redex:

$$\text{let } x := t \text{ in } s \quad := \quad (\lambda x s) t.$$

We express derivations in a natural deduction system with a similar syntax to that of object terms to stress the Curry-Howard correspondence. Proof terms are typed by their conclusion formulas and are built from *assumption variables*.

**Definition 2.** *Proof terms*  $(M, N)$  of  $\text{NA}^\omega$  are defined as follows:

$$\begin{aligned}
M, N & ::= u^A \mid (\lambda u^A M^B)^{A \rightarrow B} \mid (M^{A \rightarrow B} N^A)^B \mid \\
(*) & \quad (\lambda x^\rho M^{A(x)})^{\forall x^\rho A(x)} \mid (M^{\forall x^\rho A(x)} t)^{A(t)} \mid \\
\text{AxT} & : \text{at}(\text{tt}) \mid \text{Cases}^{A(b)} : \forall b^\text{B} (A(\text{tt}) \rightarrow A(\text{ff}) \rightarrow A(b)) \mid \\
\text{Ind}^{A(n)} & : \forall n^\text{N} (A(0) \rightarrow \forall n^\text{N} (A(n) \rightarrow A(\text{Sn})) \rightarrow A(n))
\end{aligned}$$

with the usual variable condition  $(*)$  that the object variable  $x$  does not occur freely in any of the open assumptions of  $M$ . The sets of free variables  $\text{FV}(M)$  and free (open) assumption variables  $\text{FA}(M)$  as well as capture-free substitutions  $M[x := t]$  and  $M[u := N]$  are defined inductively as usual.

The *truth axiom*  $\text{AxT}$  defines the logical meaning of  $\text{at}(\cdot)$  and allows us to consider any boolean valued function defined in our term system as a decidable predicate. When we write for example  $n = m$ , we actually mean  $\text{at}(\text{Eq} n m)$ , where  $\text{Eq}^{\text{N} \Rightarrow \text{N} \Rightarrow \text{B}}$  is a term defining the decidable equality for natural numbers.

In our negative language defining falsity as  $F := \text{at}(\text{ff})$  already gives us the full power of classical logic. In particular, for a formula  $A$  we can prove *Ex falso quodlibet* (efq):  $\vdash F \rightarrow A$  and *Stability*:  $\vdash ((A \rightarrow F) \rightarrow F) \rightarrow A$  by meta induction on  $A$ , using  $\text{AxT}$  and *Cases* for the base case. We will thus use the abbreviations  $\neg A := A \rightarrow F$  and  $\exists x^\rho A := \neg \forall x^\rho \neg A$ .

The term system in consideration is essentially Gödel's T and the reduction relation  $\mapsto^r$  is well-known to be strongly normalising and confluent. Thus, instead of insisting that object terms appearing in formulas of proof rules match exactly, we require them only to have the same  $\eta$ -long normal form; this equality will be denoted by  $\stackrel{r}{=}$ . Note that we make no such assumption for extracted programs or for proof terms themselves.

**Notation.** For technical convenience we will use  $\varepsilon$  for denoting a special *nulltype*, i.e., lack of computational content. By abuse of notation we also use  $\varepsilon$  to denote all terms of nulltype. We stipulate that the following simplifications are always carried out implicitly:

$$\begin{aligned} \rho \times \varepsilon \rightsquigarrow \rho, \quad t^{\rho \times \varepsilon} \lrcorner \rightsquigarrow t, \quad \langle t, \varepsilon \rangle \rightsquigarrow t, \quad \rho \Rightarrow \varepsilon \rightsquigarrow \varepsilon, \quad \lambda x \varepsilon \rightsquigarrow \varepsilon, \quad \varepsilon t \rightsquigarrow \varepsilon \\ \varepsilon \times \rho \rightsquigarrow \rho, \quad t^{\varepsilon \times \rho} \lrcorner \rightsquigarrow t, \quad \langle \varepsilon, t \rangle \rightsquigarrow t, \quad \varepsilon \Rightarrow \rho \rightsquigarrow \rho, \quad \lambda x^\varepsilon t \rightsquigarrow t, \quad t \varepsilon \rightsquigarrow t \\ \forall x^\varepsilon A \rightsquigarrow A, \quad M \varepsilon \rightsquigarrow M \end{aligned} \quad (\varepsilon)$$

Consequently, to simplify presentation all  $\varepsilon$  terms will be silently omitted, as they hold no computational content.

### 3 Quasi-linear Dialectica interpretation

We will shortly outline the variant of the Dialectica interpretation, which was presented in [16]. It allows extraction of more efficient programs by avoiding syntactic repetition of subterms. In particular, it turns out that the size of the extracted terms depends almost linearly on the proof size. The present paper will build upon this interpretation to improve efficiency of recursion even further.

The general idea behind avoidance of syntactic repetition is to factor out common subterms as much as possible in the positive and negative content during the extraction process. To achieve this we use *definition contexts* — a tool, which allows to gradually accumulate the common part of all witnesses of a given proof. In order to apply this technique, the definition of the Dialectica computational types needs to be slightly revised so that we use uncurried function types instead of curried ones, because both the partial and the full application of an uncurried function to a variable increase the term size with a constant<sup>1</sup>.

We start with some preliminary notations. We use  $[\cdot]$  to denote size of terms, formulas and proofs. For a proof  $M$  we define its *maximal sequent length*  $[\![M]\!]$  as  $\max_{N \leq M} |\text{FA}(N)|$ , where  $N \leq M$  is the subproof relation. The rest of the needed definitions are presented below.

**Definition 3.** Let us fix a type variable  $\diamond$  and an object variable  $\square$  of type  $\diamond$ , which will be referred to as “a hole”. A *definition context*  $E$  is a term built by the following rules:

$$E ::= \square^\diamond \mid (E^{\rho \Rightarrow \sigma} t^\rho)^\sigma \mid (\lambda x^\rho E^\sigma)^{\rho \Rightarrow \sigma},$$

where  $t$  does not contain the type  $\diamond$ . For a definition context  $E^\rho$  and term  $t^\sigma$ , we define the term  $E[t]$  ( $t$  in the context  $E$ ) as  $E[\diamond := \sigma][\square := t]$ , where, contrary to our usual convention, the free variables of  $t$  are allowed to be bound by abstractions in  $E$ .

<sup>1</sup>In contrast, full application of a curried function needs a variable for each parameter.

**Definition 4.** We define the partial application of the (uncurried) function  $f$  to the term  $t$  as

$$\begin{aligned} f^{\rho \Rightarrow \tau} \circ t^{\rho} &:= ft \\ f^{\rho \times \sigma \Rightarrow \tau} \circ t^{\rho} &:= \lambda x^{\sigma} f \langle t, x \rangle, \text{ where } x \text{ is a fresh variable.} \end{aligned}$$

**Definition 5.** We extend the projection operations  $\perp$  and  $\lrcorner$  to functions:

$$f^{\rho \Rightarrow \sigma \times \tau} \perp := \lambda x^{\rho} f x_{\perp}, \quad f^{\rho \Rightarrow \sigma \times \tau} \lrcorner := \lambda x^{\rho} f x_{\lrcorner}.$$

**Definition 6.** For a formula  $A$  we define the positive and negative computational types ( $\sigma^+(A)$  and  $\sigma^-(A)$ ). We will also denote  $\sigma^*(A) := \sigma^-(A) \Rightarrow \sigma^+(A)$ . We define:

$$\begin{aligned} \sigma^+(\text{at}(b)) &:= \varepsilon, & \sigma^-(\text{at}(b)) &:= \varepsilon, \\ \sigma^+(A \rightarrow B) &:= \sigma^+(B) \times \sigma^-(A), & \sigma^-(A \rightarrow B) &:= \sigma^*(A) \times \sigma^-(B), \\ \sigma^+(\forall x^{\rho} A) &:= \sigma^+(A), & \sigma^-(\forall x^{\rho} B) &:= \rho \times \sigma^-(B). \end{aligned}$$

**Definition 7.** For  $r : \sigma^*(A)$ ,  $s : \sigma^-(A)$  we define  $|A|_s^r$  as follows:

$$\begin{aligned} |\text{at}(b)| &:= \text{at}(b), & |\forall x A|_s^r &:= |A(s_{\perp})|_{s_{\lrcorner}}^{r \circ s_{\perp}}, \\ |A \rightarrow B|_s^r &:= |A|_{r s_{\lrcorner}}^{s_{\perp}} \rightarrow |B|_{s_{\lrcorner}}^{(r \circ s_{\perp})_{\perp}}. \end{aligned}$$

The soundness theorem for the new variant of the interpretation follows a similar pattern to the usual soundness proof. On every inductive step we define:

1. a definition context  $\llbracket M \rrbracket : \sigma^-(A) \Rightarrow \diamond$
2. a context-dependent positive witnessing term  $\llbracket M \rrbracket^+ : \sigma^+(A)$
3. context-dependent negative witnessing terms  $\llbracket M \rrbracket_i^- : \sigma^-(C_i)$

The final extracted term will be obtained by putting the context-dependent terms inside the context:

$$\{\llbracket M \rrbracket\} := \llbracket M \rrbracket[\langle \llbracket M \rrbracket^+, \dots, \llbracket M \rrbracket_i^-, \dots \rangle].$$

We will refer to the separate components put in the context as follows:

$$\{\llbracket M \rrbracket\}^+ := \llbracket M \rrbracket[\llbracket M \rrbracket^+], \quad \{\llbracket M \rrbracket\}_i^- := \llbracket M \rrbracket[\llbracket M \rrbracket_i^-].$$

**Theorem 1** (Soundness of quasi-linear extraction). *Let  $\mathcal{P} : A$  be a proof in  $\text{NA}^{\omega}$  from assumptions  $u_i : C_i$ . Let  $x_i : \sigma^*(C_i)$  and  $y_A : \sigma^-(A)$  be fresh variables. Then there is a term  $\{\llbracket \mathcal{P} \rrbracket\}$ , satisfying the following conditions:*

- i. we can prove  $|A|_{y_A}^{\{\llbracket \mathcal{P} \rrbracket\}^+}$  from  $|C_i|_{\{\llbracket \mathcal{P} \rrbracket\}_i^-}^{x_i}$ ,
- ii.  $\text{FV}(\{\llbracket \mathcal{P} \rrbracket\}) \subseteq \text{FV}(\mathcal{P}) \cup \{x_i\}$ ,
- iii.  $[\{\llbracket \mathcal{P} \rrbracket\}] \leq K([\mathcal{P}] + [\llbracket \mathcal{P} \rrbracket]^2)$  for a fixed constant  $K$ , not depending on  $\mathcal{P}$ .

## 4 A special case of recursion

A specific feature of the Dialectica interpretation which allows to embed classical logic into a quantifier-free constructive system is the extraction of counterexamples. In our negative language, in order to prove  $\exists x A$  we need to use the assumption  $\forall x \neg A$  to derive a contradiction. The non-trivial use of classical logic comes where we use this assumption more than once. In the extracted term this corresponds to deciding between counterexamples by checking the validity of the quantifier-free translation  $|A|_y^x$ . An extreme example of this phenomenon is the interpretation of induction<sup>2</sup>, which corresponds to using the induction hypothesis an unbounded number of times. This is reflected by a case distinction on every recursive step in the recursively defined programs for computing counterexamples for open assumptions. However, there is a special case of the induction scheme in which a case distinction on every step is redundant and, moreover, can lead to an unnecessary increase of complexity.

Let  $\mathcal{P} := \text{Ind}_{\mathbb{N}, A(n)} n M^{A(0)} (\lambda n, v^{A(n)} N^{A(n+1)})$  be a proof by induction from assumptions  $u_i : C_i$ . Consider the case where  $\sigma^-(A) = \varepsilon$ . For the sake of simplicity let us assume that we have only one open assumption  $u : C$  and let us omit all indices. By the usual soundness theorem (cf. [13]) we obtain the following extracted terms:

$$\begin{aligned} \llbracket \mathcal{P} \rrbracket^+ &:= \mathcal{R} n \llbracket M \rrbracket^+ (\lambda n, x_v \llbracket N \rrbracket^+) \\ \llbracket \mathcal{P} \rrbracket^- &:= \mathcal{R} n \llbracket M \rrbracket^- \left( \lambda n, p \left( \llbracket N \rrbracket^- \xi \right) \overset{u}{\bowtie} p \right), \quad \text{for } \xi := [x_v := \llbracket \mathcal{P} \rrbracket^+] , \end{aligned}$$

where  $\bowtie$  is a case distinction operator, defined as follows:

$$t_1 \overset{u}{\bowtie} t_2 := \begin{cases} t_1, & \text{if } u : C \notin \text{FA}(N), \\ t_2, & \text{if } u : C \notin \text{FA}(M), \\ \mathcal{C} |C|_{t_1}^{x_u} t_2 t_1, & \text{otherwise.} \end{cases}$$

Note that the case distinction operator depends not only on the assumption  $u$ , but on the proof branches  $M$  and  $N$  as well. To keep notation simpler we do not make this dependency explicit; the referred proof will be clear from the context.

We first notice that the computation of  $\llbracket \mathcal{P} \rrbracket^+$  is linear on  $n$ . However, in each recursive step of  $\llbracket \mathcal{P} \rrbracket^-$  in order to compute  $\llbracket N \rrbracket^-$ , we invoke a sub-computation of  $\llbracket \mathcal{P} \rrbracket^+$  for the current value of  $n$ . This makes the computation of  $\llbracket \mathcal{P} \rrbracket^-$  at least quadratic on  $n$ . In the general case of treatment of induction this cannot be avoided. However, in the case where  $\sigma^-(A) = \varepsilon$  it is easy to see that we can compute positive and negative content simultaneously:

$$\begin{aligned} \langle \llbracket \mathcal{P} \rrbracket^+, \llbracket \mathcal{P} \rrbracket^- \rangle &:= \mathcal{R} n \langle \llbracket M \rrbracket^+, \llbracket M \rrbracket^- \rangle \\ &\quad \left( \lambda n, x_v, \langle p_+, p_- \rangle \left\langle \llbracket N \rrbracket^+ p_+, (\llbracket N \rrbracket^- \xi') \overset{u}{\bowtie} p_- \right\rangle \right), \end{aligned}$$

where  $\xi' := [x_v := p_+]$ . Thus, as shown in [16], by avoiding recomputation we might improve worst time complexity of the program. However, in this special case we can optimise even further. For a fixed  $n$ ,  $\llbracket \mathcal{P} \rrbracket^-$  can be seen as performing a linear search for a counterexample for  $C$  among the  $n$  candidates in the list  $L^n := (\llbracket M \rrbracket^-, (\llbracket N \rrbracket^- \xi' [n := k])_{k < n-1})$ . Formally,

$$|C|_{\llbracket \mathcal{P} \rrbracket^-}^x \leftrightarrow \bigwedge_{k < n} |C|_{L_k^n}^x \quad \text{and} \quad \exists K < n \llbracket \mathcal{P} \rrbracket^- = L_K^n.$$

<sup>2</sup>Here we refer to the full induction rule, not to the commonly considered assumptionless induction rule.

The definition of  $\bowtie$  is asymmetric: it performs the counterexample check on one of its operands only (cf. [14]). In the considered case,  $\llbracket \mathcal{P} \rrbracket^-$  always returns the *last possible counterexample* in the list  $L^n$ , i.e.,  $\forall k > K |C|_{L_k}^x$ . This behaviour seems inefficient and a simple idea to change it is to reverse the operands of  $\bowtie$  in the definition of  $\llbracket \mathcal{P} \rrbracket^-$ . Indeed, in this case we would return the *first possible counterexample* from  $L^n$ , i.e.,  $\forall k < K |C|_{L_k}^x$ . Unfortunately, this will not improve the efficiency of the extracted program, because it will still perform  $n$  steps, computing all elements from the list  $L^n$  and performing  $n - 1$  case distinctions. It is clearly sufficient to terminate the recursion as soon as we find the first index  $K$  for which  $\neg |C|_{L_k}^x$ . Although this will not change the *worst time complexity* of the program, it might improve the *average time complexity*, in case the expected value of  $K$  is lower than  $O(n)$ .

Such an earlier terminating search could be implemented by adding a boolean flag  $b$ , which specifies whether a counterexample is already found. For example:

$$\begin{aligned} \langle \llbracket \mathcal{P} \rrbracket^+, \langle \llbracket \mathcal{P} \rrbracket^-, b \rangle \rangle &:= \mathcal{R}n \langle \llbracket M \rrbracket^+, \langle \llbracket M \rrbracket^-, \text{ff} \rangle \rangle \\ &\quad \left( \lambda n, x_v, \langle p_+, \langle p_-, b \rangle \rangle \left\langle \llbracket N \rrbracket^+ p_+, p_- \overset{b,u}{\bowtie} \llbracket N \rrbracket^- \xi' \right\rangle \right), \\ \text{where } t_1 \overset{b,u}{\bowtie} t_2 &:= \mathcal{C} b \langle t_1, \text{tt} \rangle (\mathcal{C} |C|_{t_1}^x \langle t_2, \text{ff} \rangle \langle t_1, \text{tt} \rangle). \end{aligned}$$

Note that the assumption  $\sigma^-(A) = \varepsilon$  is important, otherwise  $p_-$  would be a function, applied to a term depending on  $n$  on each recursive step. This would prevent us from using the information that a counterexample is found on an earlier step to terminate the recursion.

In [13] Ratiu and the author considered the Infinite Pigeonhole Principle as a case study for program extraction from non-constructive proofs. There we showed that the refined  $A$ -translation method [2] extracts a program, which has exponential worst time complexity, but polynomial average time complexity, while the program extracted by Dialectica is exponential in both the worst and the average case. However, the optimisation described in this section applies and we can obtain a Dialectica program, which has polynomial average time complexity, like the one extracted by refined  $A$ -translation.

The considered early termination of the recursive process is very reminiscent of an abortive control operator [3], where immediate transfer of the program flow control occurs. As discussed in [13], similar situations occur with programs, extracted by refined  $A$ -translation. For the case study discussed there, it seemed that this feature had an important contribution in achieving better average time complexity. This suggestion is reaffirmed by the fact that adding such an optimisation to the Dialectica interpretation has the same favourable effect on extracted programs.

## 5 Marked counterexamples

As was discussed in Section 4, the programs extracted with the original Dialectica interpretation do not take advantage of the information about the validity of counterexample where a case distinction is needed. The case distinction construction  $t_1 \overset{u}{\bowtie} t_2$  forces us to choose between two candidate counterexamples  $t_1$  and  $t_2$  for the assumption  $u : C$ . The choice is made by direct checking of the decidable Dialectica translation of the formula  $C$  for one of the counterexamples. What is not taken into account is that if the check confirms the existence of a counterexample, all further computation of witnesses and counterexamples is pointless. In a certain sense, this can be viewed as avoiding both

1. *recomputation* — the validity of the counterexample is rechecked if we have more than two occurrences of the assumption  $C$ ,

2. *redundant computation* — all further counterexamples and witnesses computed are not needed for a sound verification proof.

It is important to note that the common context approach from [16] seems inapplicable for avoiding such kind of recomputation. The reason is the underlying difference between repeated subterms and the recomputation considered here. We can detect duplicated terms during the extraction process and we use a shared context to avoid it. However, the counterexample decision happens during the evaluation of the program and, depending on the input parameters, recomputation might or might not occur. Attempting to use a shared context would imply precomputation of all possible case distinctions, which could be much worse than recomputing only one case distinction.

We will thus follow a different idea. As was already hinted in Section 4, an additional marker will be attached to each extracted counterexample, carrying information about its validity. We will use  $B^\perp$  with three constants: tt, ff and  $\perp$ , as a type for markers.

**Definition 8.** For a formula  $A$  we will re-define the positive and negative computational types denoting the new variants as  $\rho^+(A)$  and  $\rho^-(A)$ . We will also denote  $\rho^*(A) := \rho^-(A) \Rightarrow \rho^+(A)$  and  $\rho^\circ(A) := B^\perp \times \rho^-(A)$ . We define:

$$\begin{aligned} \rho^+(\text{at}(b)) &:= \varepsilon, & \rho^-(\text{at}(b)) &:= \varepsilon, \\ \rho^+(A \rightarrow B) &:= \rho^+(B) \times \rho^\circ(A), & \rho^-(A \rightarrow B) &:= \rho^*(A) \times \rho^-(B) \\ \rho^+(\forall x^\sigma A) &:= \rho^+(A), & \rho^-(\forall x^\sigma B) &:= \sigma \times \rho^-(B) \end{aligned}$$

For clarity  $t \blacktriangleright m$  will denote that  $t$  is marked by  $m$ . Consequently, when we write  $t \blacktriangleright m \stackrel{r}{=} s$ , we will mean that  $m \stackrel{r}{=} s_\perp$  and  $t \stackrel{r}{=} s_\perp$ . The marker constants have the following intended meaning:

- $t \blacktriangleright \perp$  — we have no information yet about the validity of  $|C_i|_t^{x_i}$ ,
- $t \blacktriangleright \text{ff}$  — we have checked that  $\neg |C_i|_t^{x_i}$ ,
- $t \blacktriangleright \text{tt}$  —  $t$  is an arbitrarily chosen term and we should prefer another candidate counterexample without the need to check  $|C_i|_t^{x_i}$ .

The change in the positive type in the implication case of the translation leads to a slight adjustment to the Dialectica translation (emphasized by a box below):

$$|A \rightarrow B|_z^f := |A|_{fz_\perp}^{\perp} \rightarrow |B|_{z_\perp}^{(f \circ z_\perp)_\perp}.$$

The essential use of the marker comes in the definition of case distinction terms.

**Lemma 1.** For every formula  $C$  and variable  $x$ :  $\rho^*(C)$  there is a term  $T_\boxtimes^C : \rho^\circ(C) \Rightarrow \rho^-(C) \Rightarrow \rho^-(C)$  with  $\text{FV}(T_\boxtimes^C) \subseteq \text{FV}(C) \cup \{x\}$ , such that for  $t_1, t_2 : \rho^\circ(C)$  from the assumptions  $u_i : (m_i = \text{ff} \rightarrow \neg |C|_{s_i}^x)$  we can prove

$$A_i : (m \neq \text{tt} \rightarrow |C|_s^x) \rightarrow (m_i \neq \text{tt} \rightarrow |C|_{s_i}^x),$$

$$B : m = \text{ff} \rightarrow \neg |C|_s^x,$$

where  $s_i \blacktriangleright m_i \stackrel{r}{=} t_i$  and  $s \blacktriangleright m \stackrel{r}{=} T_\boxtimes^C t_1 t_2$ .

*Proof.* Define

$$\begin{aligned} T_\boxtimes^C(s_1 \blacktriangleright m_1)(s_2 \blacktriangleright m_2) &:= \mathcal{C}(T_\vee(m_2 = \text{tt})(m_1 = \text{ff})) t_1 \\ &\quad \mathcal{C}(T_\vee(m_1 = \text{tt})(T_\vee(m_2 = \text{ff})(T_C x s_1))) t_2 (s_1 \blacktriangleright \text{ff}), \end{aligned}$$

		$m_1$		
		tt	ff	$\perp$
$m_2$	tt	$D_1$	$D_2$	$D_2$
	ff	$D_1$	$D_1$	$D_1$
	$\perp$	$D_1$	$D_2$	?

Table 1: Case analysis for  $T_{\bowtie}^C t_1 t_2$ 

where  $T_V := \lambda x, y \mathcal{C} x t t y$  and  $T_C$  is such that  $\text{at}(T_C x y) \leftrightarrow |C|_y^x$ . It is clear that

$$\text{at}(T_V x y) \leftrightarrow (\neg \text{at}(x) \rightarrow \text{at}(y)) \leftrightarrow (\neg \text{at}(y) \rightarrow \text{at}(x)).$$

Let us denote  $D_i := (T_{\bowtie}^C t_1 t_2 = t_i)$ . It is easy to see that

- $D_i$  immediately implies  $A_i$  and also  $B$  by  $u_i$ ,
- $m_i = \text{tt}$  immediately implies  $A_i$
- $D_i \wedge (m_i = \text{ff})$  implies  $\neg |C|_s^x$  by  $u_i$ , which contradicts with the premise  $m \neq \text{tt} \rightarrow |C|_s^x$ , implying both  $A_1$  and  $A_2$ .

Table 1 summarizes the validity of  $D_i$  depending on the values of  $m_1$  and  $m_2$ . It can be checked that the arguments above are sufficient to establish the validity of  $A_1, A_2$  and  $B$  in all cases except the one marked by '?. In order to complete the proof we assume that  $m_1 = m_2 = \perp$  and consider cases on the decidable formula  $|C|_{s_1}^x$ .

*Case  $|C|_{s_1}^x$ .*  $D_2$  holds, thus we have only  $A_1$  to prove. However, the conclusion of  $A_1$  is exactly what we assumed in this case.

*Case  $\neg |C|_{s_1}^x$ .* We have  $s = s_1$ , which implies that  $\neg |C|_s^x$ , proving  $B$ . On the other hand,  $m = \text{ff}$ , which contradicts with the premise  $m \neq \text{tt} \rightarrow |C|_s^x$ , implying both  $A_1$  and  $A_2$ .

Finally, assuming that  $T_C$  is a variable bound by an external definition context as in [16], we see that  $\lceil T_{\bowtie}^C t_1 t_2 \rceil \leq \lceil t_1 \rceil + \lceil t_2 \rceil + K$  for some constant  $K$ , not dependent on  $C$ ,  $t_1$  or  $t_2$ .  $\square$

We will prove soundness for the modified variant of the Dialectica interpretation using the constructions  $\llbracket M \rrbracket, \llbracket M \rrbracket^+, \llbracket M \rrbracket^-$  from Theorem 1. Their types will be as follows:

$$\llbracket M \rrbracket : \rho^-(A) \rightarrow \diamond, \quad \llbracket M \rrbracket^+ : \rho^+(A), \quad \llbracket M \rrbracket_i^- : \rho^{-\circ}(C_i).$$

**Theorem 2** (Soundness of counterexample marking). *Let  $\mathcal{P} : A$  be a proof from assumptions  $u_i : C_i$ . Let  $x_i : \rho^*(C_i)$  and  $y_A : \rho^-(A)$  be fresh variables. Then there is a term  $\{\{\mathcal{P}\}\}$ , satisfying the following conditions:*

- we can prove  $|A|_{y_A}^{\{\{\mathcal{P}\}\}^+}$  from  $(m_i \neq \text{tt} \rightarrow |C_i|_{s_i}^{x_i})$ ,*
- we can prove  $m_i = \text{ff} \rightarrow \neg |C_i|_{s_i}^{x_i}$ ,*
- $\text{FV}(\{\{\mathcal{P}\}\}) \subseteq \text{FV}(\mathcal{P}) \cup \{x_i\}$ ,*
- $\lceil \{\{\mathcal{P}\}\} \rceil = K(\lceil \mathcal{P} \rceil + \lceil \{\{\mathcal{P}\}\} \rceil^2)$  for a fixed constant  $K$ , not depending on  $\mathcal{P}$ ,*

where  $s_i \blacktriangleright m_i \stackrel{r}{=} \{\{\mathcal{P}\}\}_i^- y_A$ .



*Proof.* The proof is a modification of the argument needed for Theorem 1.

*Case  $u : A$ .* We set as before  $\llbracket \mathcal{P} \rrbracket := \lambda y_A \llbracket \cdot \rrbracket$ ,  $\llbracket \mathcal{P} \rrbracket^+ := x_u y_A$  and set  $\llbracket \mathcal{P} \rrbracket_u^- := y_A \blacktriangleright \perp$ . Then  $\{\llbracket \mathcal{P} \rrbracket\}^+ \stackrel{r}{=} \lambda y_A x_u y_A \stackrel{r}{=} x_u$  and  $\{\llbracket \mathcal{P} \rrbracket\}_i^- y_A \stackrel{r}{=} y_A \blacktriangleright \perp$ . The assumption premise  $m_u \neq \text{tt}$  holds, which is enough to conclude that  $|A|_{y_A}^{x_u}$ . On the other hand  $m_u \neq \text{ff}$ , which makes the second condition trivially true. The size bounds and the variable condition also hold as in Theorem 1.

*Case  $\lambda u^B M^C$ .* Let us denote  $s_i \blacktriangleright m_i \stackrel{r}{=} \{M\}_i^- y_C$  for  $i$  ranging over all assumption variables of  $M$ , including  $u$ . By induction hypothesis we have a proof of  $|C|_{y_C}^{\{M\}^+}$  from  $m_u \neq \text{tt} \rightarrow |B|_{s_u}^{x_u}$  and  $m_i \neq \text{tt} \rightarrow |C_i|_{s_i}^{x_i}$ , as well as proofs of  $m_u = \text{ff} \rightarrow \neg |B|_{s_u}^{x_u}$  and  $m_i = \text{ff} \rightarrow \neg |C_i|_{s_i}^{x_i}$ . The extracted terms which work for Theorem 1 are still applicable:

$$\begin{aligned} \llbracket \mathcal{P} \rrbracket &:= \lambda y_A \mathbf{let} \ x_u := y_A \blacktriangleleft \ \mathbf{in} \ \llbracket M \rrbracket (y_A \blacktriangleleft), \\ \llbracket \mathcal{P} \rrbracket^+ &:= \langle \llbracket M \rrbracket^+, \llbracket M \rrbracket_u^- \rangle, \\ \llbracket \mathcal{P} \rrbracket_i^- &:= \llbracket M \rrbracket_i^-. \end{aligned}$$

However, we consider an additional special case: if  $u \notin \text{FA}(M)$ , then we set  $\llbracket M \rrbracket_u^- := \sqcup \blacktriangleright \text{tt}$ , where  $\sqcup$  is an arbitrary term of type  $\rho^-(B)$ . Substituting  $y_A$  with  $\langle x_u, y_C \rangle$ , it suffices to prove

1.  $|B|_{s_u}^{x_u} \rightarrow |C|_{y_C}^{\{M\}^+}$  from  $m_i \neq \text{tt} \rightarrow |C_i|_{s_i}^{x_i}$ ,
2.  $m_i = \text{ff} \rightarrow \neg |C_i|_{s_i}^{x_i}$ .

2 follows directly from our induction hypothesis and for 1 we consider subcases on  $m_u$ . The case  $m_u = \perp$  is proved as in [16]. If  $m_u = \text{ff}$  then by induction hypothesis we have  $\neg |B|_{s_u}^{x_u}$  and we can conclude using eq. Finally, if  $u \notin \text{FA}(m)$  or  $m_u = \text{tt}$ , then by induction hypothesis we can actually prove  $|C|_{y_C}^{\{M\}^+}$  without using the assumption  $m_u \neq \text{tt} \rightarrow |B|_{s_u}^{x_u}$ . Therefore we can conclude by using the same proof with a void implication introduction.

*Case  $M^B \rightarrow^A N^B$ .* We define  $\llbracket \mathcal{P} \rrbracket$ ,  $\llbracket \mathcal{P} \rrbracket^+$  and  $\llbracket \mathcal{P} \rrbracket_i^-$  essentially as in Theorem 1, with the only difference that we redefine the case distinction operator  $\bowtie$  used to combine the negative extracted terms of type  $\rho^-(C_i)$  as follows:

$$t_1 \bowtie^{u_i} t_2 := \begin{cases} t_1, & \text{if } u_i \notin \text{FA}(N), \\ t_2, & \text{if } u_i \notin \text{FA}(M), \\ T_{\bowtie}^{C_i} t_1 t_2, & \text{otherwise.} \end{cases}$$

The results from Lemma 1 are sufficient to conclude the proof.

*Cases  $\lambda x^p M^B$  and  $M^{\forall x^p A} t^p$ .* The proof of the same case in Theorem 1 still applies, because in both cases we neither remove nor introduce assumptions.

*Case  $\text{Ind}_{B,A(b)} b M^{A(\text{tt})} N^{A(\text{ff})}$ .* We define the extracted terms as in Theorem 1 with the only change that  $\bowtie$  is again defined as above. Lemma 1 allows us to apply the usual soundness proof for this case.

*Case  $\text{Ind}_{N,A(n)} n M^{A(0)} (\lambda n, u^{A(n)} N^{A(n+1)})$ .* By induction hypothesis we have:

$$\begin{array}{lll} \text{a proof } M' \text{ of} & |A(0)|_{y_A}^{\{M\}^+} & \text{from } m_i \neq \text{tt} \rightarrow |C_i|_{s_i}^{x_i}, \\ \text{proofs } M_i'' \text{ of} & m_i = \text{ff} \rightarrow \neg |C_i|_{s_i}^{x_i}, & \\ \text{a proof } N' \text{ of} & |A(n+1)|_{y_A}^{\{N\}^+} & \text{from } n_i \neq \text{tt} \rightarrow |C_i|_{r_i}^{x_i} \text{ and} \\ & & n_u \neq \text{tt} \rightarrow |A(n)|_{r_u}^{x_u}, \end{array}$$

$$\begin{aligned} \text{proofs } N_i'' \text{ of } n_i = \text{ff} &\rightarrow \neg |C_i|_{r_i}^{x_i}, \\ \text{a proof } N_u'' \text{ of } n_u = \text{ff} &\rightarrow \neg |A(n)|_{r_u}^{x_u}, \end{aligned}$$

where  $s_i \blacktriangleright m_i \stackrel{r}{=} \{|M|\}_i^- y_A$  and  $r_j \blacktriangleright n_j \stackrel{r}{=} \{|N|\}_j^- y_A$  for  $i$  and  $j$  ranging over  $\text{FA}(M)$  and  $\text{FA}(N)$ , respectively.

We define the extracted terms  $\llbracket \mathcal{P} \rrbracket$ ,  $\llbracket \mathcal{P} \rrbracket^+$  and  $\llbracket \mathcal{P} \rrbracket_i^-$  as in Theorem 1, but using the marker-aware variant of  $\boxtimes$ . We have

$$\begin{aligned} \{\mathcal{P}\} [n := 0] &\stackrel{r}{=} \{|M|\} & (*) \\ \{\mathcal{P}\}^+ [n := n+1] &\stackrel{r}{=} \text{let } x_u := \{\mathcal{P}\} \text{ in } \{|N|\}^+, \\ \{\mathcal{P}\}_i^- [n := n+1] y_A &\stackrel{r}{=} \text{let } x_u := \{\mathcal{P}\} \text{ in } \{|N|\}_i^- y_A \boxtimes^u \{\mathcal{P}\}_i^- (\{|N|\}_u^- y_A). \end{aligned}$$

Let us denote  $t_i \blacktriangleright p_i \stackrel{r}{=} \{\mathcal{P}\}_i^- y_A$ . To prove soundness of the term  $\{\mathcal{P}\}$  we will use induction on  $n$  to prove the formulas

$$\begin{aligned} F(n) &:= \forall y_A \left( \left( \bigwedge_i (p_i \neq \text{tt} \rightarrow |C_i|_{t_i}^{x_i}) \right) \rightarrow |A|_{y_A}^{\{\mathcal{P}\}^+} \right), \\ G_i(n) &:= \forall y_A \left( p_i = \text{ff} \rightarrow \neg |C_i|_{t_i}^{x_i} \right). \end{aligned}$$

For  $n = 0$  we can directly use the proofs  $M'$  and  $M_i''$  from the induction hypothesis, with the necessary implication introductions. Now let us assume  $F(n)$  and  $G_i(n)$ , fix  $y_A$  and the premises  $p_i^+ \neq \text{tt} \rightarrow |C_i|_{t_i^+}^{x_i}$ , where  $t_i^+ \blacktriangleright p_i^+ \stackrel{r}{=} \{\mathcal{P}\}_i^- [n := n+1] y_A$ . By (\*) and the properties of  $\boxtimes$  from Lemma 1 we can conclude

$$\left( n_i \Xi \neq \text{tt} \right) \rightarrow |C_i|_{r_i \Xi}^{x_i}, \quad (1)$$

$$\left( p_i [y_A := r_u \Xi] \neq \text{tt} \right) \rightarrow |C_i|_{t_i [y_A := r_u \Xi]}^{x_i}, \quad (2)$$

where  $\Xi := [x_u := \{\mathcal{P}\}^+]$ .

$G_i(n+1)$  can be shown from the proofs  $N_i''$  and  $G_i(n)$  instantiated with  $y_A := r_u \Xi$  by using (\*) and Lemma 1. To prove  $F(n+1)$  we start by using the induction hypothesis  $F(n)$  for  $y_A := r_u \Xi$  and (2) to obtain  $|A|_{r_u \Xi}^{\{\mathcal{P}\}^+}$ . In order to continue further, we need to consider subcases on  $n_u \Xi$ .

In case  $n_u \Xi = \perp$ , by (1) we have all premises of  $N' \Xi$ , thus we can conclude  $|A|_{y_A}^{\{\mathcal{P}\}^+ [n := n+1]}$ , which was to be shown.

In case  $n_u \Xi = \text{ff}$ , we can use  $N_u'' \Xi$  to derive a contradiction and conclude by using eqf.

In case  $n_u \Xi = \text{tt}$ , the premise  $n_u \Xi \neq \text{tt} \rightarrow |A(n)|_{r_u \Xi}^{x_u}$  of  $N' \Xi$  is trivially true and by (1) we have all other premises, hence  $|A|_{y_A}^{\{\mathcal{P}\}^+ [n := n+1]}$ .  $\square$

## 6 Conclusion and future work

The presented variant of the Dialectica interpretation interleaves extracted programs with additional information, which is utilised during evaluation in order to omit redundant calculations. Other kinds of redundancies can be avoided by means of uniform annotations, as described in [8, 12]. It can be argued that extensions of such technical nature may obscure the obtained computational content. However, such

ideas seem to be practically applicable when implementing an automatic extraction method, with the goal to find a correct program, which is not less efficient than a non-verified hand-written program. A topic of further investigation would be to find with a suitable combination between the refined uniform annotations described in [15] and the current extension of the interpretation.

## References

- [1] Ulrich Berger (2005): *Uniform Heyting Arithmetic*. *Ann. Pure Appl. Logic* 133(1–3), pp. 125–148. Available at <http://dx.doi.org/10.1016/j.apal.2004.10.006>.
- [2] Ulrich Berger, Wilfried Buchholz & Helmut Schwichtenberg (2002): *Refined program extraction from classical proofs*. *Ann. Pure Appl. Logic* 114(1–3), pp. 3–25.
- [3] Matthias Felleisen & Daniel P. Friedman (1989): *A Syntactic Theory of Sequential State*. *Theor. Comput. Sci.* 69(3), pp. 243–287.
- [4] Harvey Friedman (1978): *Classically and intuitionistically provably recursive functions*. *Lecture Notes in Mathematics* 669, pp. 21–27.
- [5] Kurt Gödel (1958): *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*. *Dialectica* 12, pp. 280–287.
- [6] Timothy Griffin (1990): *A Formulae-as-Types Notion of Control*. In: *POPL*, pp. 47–58. Available at <http://doi.acm.org/10.1145/96709.96714>.
- [7] Mircea-Dan Hernest (2007): *Optimized programs from (non-constructive) proofs by the light (monotone) Dialectica interpretation*. Ph.D. thesis, Ecole Polytechnique. Available at <http://www.brics.dk/~danher/teza/thesfull.pdf>.
- [8] Mircea-Dan Hernest & Trifon Trifonov (2010): *Light Dialectica Revisited*. *Annals of Pure and Applied Logic* 161(11), pp. 1313–1430. Available at <http://dx.doi.org/10.1016/j.apal.2010.04.008>.
- [9] Ulrich Kohlenbach (2008): *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer. Available at <http://www.springer.com/mathematics/book/978-3-540-77532-4>.
- [10] Yevgeniy Makarov (2006): *Practical Program Extraction from Classical Proofs*. *Electr. Notes Theor. Comput. Sci.* 155, pp. 521–542. Available at <http://dx.doi.org/10.1016/j.entcs.2005.11.071>.
- [11] Michel Parigot (1992):  *$\lambda\mu$ -Calculus: An Algorithmic Interpretation of Classical Natural Deduction*. In: Andrei Voronkov, editor: *LPAR, Lecture Notes in Computer Science* 624, Springer, pp. 190–201. Available at <http://dx.doi.org/10.1007/BFb0013061>.
- [12] Diana Ratiu & Helmut Schwichtenberg (2009). *Decorating Proofs*. To appear in Mints Festschrift, draft at: <http://www.math.lmu.de/~schwicht/papers/mints09/deco20090728.pdf>.
- [13] Diana Ratiu & Trifon Trifonov (2009). *Exploring the Computational Content of the Infinite Pigeon-hole Principle*. Draft at <http://www.math.lmu.de/~trifonov/papers/iph.pdf>. Available at <http://dx.doi.org/10.1093/logcom/exq007>. To appear in Proceedings of CiE 2008, Journal of Logic and Computation.
- [14] Helmut Schwichtenberg (2008): *Dialectica interpretation of well-founded induction*. *Mathematical Logic Quarterly* 54(3), pp. 229–239. Available at <http://dx.doi.org/10.1002/malq.200710045>.
- [15] Trifon Trifonov (2009): *Dialectica Interpretation with Fine Computational Control*. In: Klaus Ambos-Spies, Benedikt Löwe & Wolfgang Merkle, editors: *Mathematical Theory and Computational Practice, LNCS 5635*, Springer Berlin/Heidelberg, pp. 467–477. Available at [http://dx.doi.org/10.1007/978-3-642-03073-4\\_48](http://dx.doi.org/10.1007/978-3-642-03073-4_48). Proceedings of 5th Conference on Computability in Europe, CiE 2009, Heidelberg, Germany, July 19-24, 2009.

- [16] Trifon Trifonov (2010): *Quasi-linear Dialectica Extraction*. In: Fernando Ferreira, Benedikt Löwe, Elvira Mayordomo & Luís Mendes Gomes, editors: *CiE, Lecture Notes in Computer Science* 6158, Springer, pp. 417–426. Available at [http://dx.doi.org/10.1007/978-3-642-13962-8\\_46](http://dx.doi.org/10.1007/978-3-642-13962-8_46).
- [17] A. S. Troelstra (1973): *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, *Lecture Notes in Mathematics* 344. Springer-Verlag.