Arch. Math. Logic (2015) 54:921–940
DOI 10.1007/s00153-015-0450-y

Mathematical Logic

Normalization proof for Peano Arithmetic

Annika Siders[1]

Received: 6 March 2014 / Accepted: 3 October 2015 / Published online: 13 October 2015

Abstract A proof of normalization for a classical system of Peano Arithmetic formulated in natural deduction is given. The classical rule of the system is the rule for indirect proof restricted to atomic formulas. This rule does not, due to the restriction, interfere with the standard detour conversions. The convertible detours, numerical inductions and instances of indirect proof concluding falsity are reduced in a way that decreases a vector assigned to the derivation. By interpreting the expressions of the vectors as ordinals each derivation is assigned an ordinal less than ε_0. The vector assignment, which proves termination of the procedure, originates in a normalization proof for Gödel's T by Howard (Intuitionism and proof theory. North-Holland, Amsterdam, 1970).

Keywords Normalization · Peano Arithmetic · Natural deduction

Mathematics Subject Classification 03F05 · 03F30

1 Introduction

The aim of this paper is to prove normalization of Peano Arithmetic formulated in a standard natural deduction with an ordinal assignment originating in Howard's normalization proof for Gödel's T [1]. The result is a generalization of an argument found in an earlier paper [2] containing a consistency proof for Heyting Arithmetic.

✉ Annika Siders
annika.siders@helsinki.fi

1 Department of Philosophy, University of Helsinki, Unioninkatu 40 A, P.O. Box 24, 00014 Helsinki, Finland

Normalization is a process of transforming a derivation into a form that contains no unnecessary detours. Non-normalities in a derivation can be of three kinds: a detour of introduction and elimination, a numerical induction or an instance of indirect proof that concludes falsity. The procedure of normalization consists of finding a suitable non-normality in a derivation and eliminating or simplifying it. The method closely resembles the procedure of reducing a derivation of a contradiction to show that inconsistency is impossible, but in a normalization proof the reductions have to be generalized to derivations with open assumptions and an arbitrary conclusion.

A well-known difficulty of normalization proofs is the reduction of non-normal instances of implication, so called implication detours. A detour of introduction and elimination means that the major premise of an elimination rule has been introduced by a corresponding introduction rule. The reduction of the implication detour composes the derivation of the minor premise of the elimination rule with the derivation of the premise of the introduction rule. The problem that arises is that the reduction in this case can produce a seemingly more complex derivation. Namely, the derivation of the minor premise of the implication elimination rule can be duplicated an arbitrary number of times if there are multiple uses of the discharged assumption of the implication introduction rule. The solution must be to find a measure, which shows that the reduction is in fact a simplification.

In a sequent calculus setting normalization corresponds to cut elimination. If an assumption has been used several times in a natural deduction derivation, then we have a number of contractions on this formula in the sequent calculus version of the derivation. The problem of duplication of subderivations in natural deduction therefore corresponds to the problem of dealing with contraction when eliminating cuts in a sequent calculus system. That contraction requires a special treatment in cut elimination proofs confirms that a multiple discharge requires a corresponding solution in natural deduction.

Prawitz has in [5] given a proof of normalization for the classical system of natural deduction using a Gentzen-measure originating in Gentzen's 1938 proof for the consistency of Peano Arithmetic. In Gentzen's consistency proof contraction on a reducible cut formula requires a heightline argument. This argument introduces cuts on additional formulas in the derivation instead of simply converting cuts on compound formulas into cuts on shorter formulas. Because the natural deduction system lacks a cut rule the reformulation of Gentzen's method in natural deduction requires an addition to or alteration of the system to capture the interaction of cut and contraction in Gentzen's proof.

Prawitz' proof adds an explicit composition rule which corresponds to a cut rule in sequent calculus. This allows him to treat deduction notations that correspond to standard derivations in natural deduction. The additional rule of the system allows him to specify at which point in a derivation a heightline should be added. A heightline in a derivation corresponds to a tower of exponentiation in the ordinal assigned to the derivation and therefore the explicit composition rule gives an additional method to regulate the growth of the ordinal.

The solution presented in the present proof allows us to directly use the standard natural deduction system. Basing the ordinal assignment on Howard's constructions instead of Gentzen's assignment circumvents the problem of cut and contraction.

Howard's vector assignment has previously been used to show that it is possible to give a Gentzen-style consistency proof without a heightline argument [6]. When Howard's constructions are adapted to a sequent calculus system one can obtain cut reductions that resemble proofs of direct cut elimination without multicut. In a natural deduction setting this means that the assignment allows for the kind of duplication of a subderivation, which occurs in a detour conversion for implications. Therefore, the generalization from a consistency to a normalization proof using this method gives us a proof that does not require heightlines or a composition rule. The details and an intuitive explanation of the construction of the ordinal assignment are found in Sect. 4.

Section 4.3 that formally defines two operations on vectors and Sect. 6 that states formal lemmas for the operations can be skipped by a reader who prefers the informal explanation of the main property of vectors in Sect. 4.2. The main property of vectors and a comparison with the vector assignment of Definition 6 is enough to give a rough understanding of the ordinal inequalities for the normalization reductions in Sect. 8.

2 The system of natural deduction

We consider a system for classical predicate logic with the logical constant $\perp$ and the logical connectives &, $\supset$ and $\forall$. Negation, disjunction and existential quantifier are taken as defined connectives. The system has the standard introduction and elimination rules, and the classical rule for indirect inference, *reductio ad absurdum* (RAA), where the conclusion of the rule is an atomic formula. Prawitz used this system in his thesis from 1965 [4], which was the first published normalization result for classical predicate logic.

Logical rules:

$$\frac{A \quad B}{A\&B}\,\&I \qquad \frac{A\&B}{A}\,\&E \qquad \frac{A\&B}{B}\,\&E$$

$$\frac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}\,\supset I \qquad \frac{A \supset B \quad A}{B}\,\supset E$$

$$\frac{A(y/x)}{\forall x A}\,\forall I \qquad \frac{\forall x A}{A(t/x)}\,\forall E$$

$$\frac{\begin{matrix}[\neg P]\\ \vdots \\ \perp\end{matrix}}{P}\,RAA$$

Definition 1 The *major premise* of an elimination rule is the premise that contains the connective or quantifier, which is eliminated. The *minor premise* of an implication elimination rule is the derivation of the antecedent of the implication.

The system used in this article restricts the classical rule of reductio ad absurdum to atomic formulas. This is a standard method to deal with normalization for a classical system that avoids mixing non-normalities with the classical rule. Thus, the major premise of an elimination rule cannot be derived by reductio ad absurdum. As is proved in [4] Theorem III.1, the rule of reductio ad absurdum for arbitrary compound formulas is an admissible rule in this system. In other words, if we have a derivation of falsity from an assumption, $\neg A$, where A is a compound formula, then A is derivable. If a general formulation of the rule was added to the system, then a major premise of an elimination rule could be derived by RAA. Therefore, the rule RAA is restricted to atomic formulas in order to not interfere with the main reductions, the convertible detours.

3 Peano Arithmetic

The system of natural deduction can be extended with arithmetical axioms converted into rules of inference and an induction rule. Von Plato introduced this system of rules in [3] in order to prove the disjunction and existential properties for Heyting Arithmetic. The system was also stated and used in the consistency proof for Heyting Arithmetic found in [2].

The system of Peano Arithmetic is a theory in the language $\{0, s, +, \cdot\}$, where 0 is a constant, the successor function, s, is a unary function symbol and addition, $+$, and multiplication, $\cdot$, are binary function symbols. Furthermore, the system contains the binary predicate symbol for equality, $=$.

Definition 2 *Terms* can be inductively defined. The constant 0 and variables are terms and if t and t' are terms then st, $t + t'$ and $t \cdot t'$ are also terms. Terms are *closed* if they do not contain any variable.

Formal expressions for the natural numbers, *numerals*, are inductively defined as successors of the constant 0. The constant 0 is a numeral and if $\overline{m}$ is a numeral, then $s\overline{m}$ is also a numeral.

Peano's axioms define the system. The system can be extended with rules for primitive recursive functions by adding recursion rules for each function. To obtain a system for Peano Arithmetic we add the arithmetical rules as stated in [2] and the following standard induction rule to the logical calculus.

Induction rule:

$$\dfrac{A(0/x) \quad \begin{matrix}[A(y/x)] \\ \vdots \\ A(sy/x)\end{matrix}}{A(t/x)}\, Ind$$

In the induction rule y is the eigenvariable of the rule and it should not occur free in the conclusion.

The following properties for derivations are provable.

Lemma 1 *(i) For an arbitrary closed term t there exists a numeral $\overline{n}$ such that $\rightarrow t = \overline{n}$ can be derived without Ind.*

(ii) *Let t and t' be closed terms for which $t = t'$ can be derived without Ind. Then for an arbitrary formula A the sequent $A(t/x) \supset A(t'/x)$ can be derived without Ind.*

Proof The proof is by induction on the complexity of the term and formula respectively. A complete proof is found in [2] Lemma 4.1(i) and (iv). □

4 Defining ordinal expressions with variables

4.1 An ordinal assignment with variables

To give a normalization proof one needs to define the reductions that are applied on derivations, a strategy for finding a suitable reduction and an assignment of ordinals to derivations. An assignment is needed as a measure to prove termination of the normalization procedure. If an ordinal is assigned to each derivation and the chosen reduction decreases this ordinal, then by the well-ordering of the ordinals the procedure must terminate in a normal derivation that cannot be reduced by the procedure.

As was mentioned in the introduction a well-known difficulty with constructing an ordinal assignment for normalization proofs is the reduction of non-normal instances of implication. An implication detour is by the standard reduction converted into a derivation composed of the premise of the introduction rule and multiple copies of the minor premise of the implication elimination rule. Where the position of the subderivations is parallel in the detour they are placed on top of each other in the composed derivation. Thus, the ordinal assignment should capture a decrease in the ordinal, despite a seemingly arbitrary duplication of a subderivation as well as a substitution of the discharged implication assumption with an arbitrarily complex derivation of the same formula. Both of these features are solved by a Howard-style assignment.

The ordinal assignment of [1] introduces variables in the assigned ordinals. The variables are assigned to assumptions, both open and discharged, thereby allowing a substitution of an arbitrarily large ordinal for the variable, thus solving the problem with reduction to a composed derivation. Moreover, the ordinal assignment introduces ordinal vectors and two operations, the delta- and box-operation, on these vectors. These elements of the assignment interact to produce an upper bound for the ordinal of a derivation with a seemingly arbitrary duplication of a subderivation due to discharge of multiple assumptions. The decrease of the ordinal of the reduced derivation is due to the provable main property of vectors (see Sect. 4.2). Were the derivation is reduced to a composed derivation, the ordinal (or more accurately the ordinal vector) is reduced from an ordinal containing the operations to an ordinal defined by substitution of ordinals for the ordinal variables.

4.2 The main property of vectors

The assignment of [1] is a vector assignment to derivations. Formal expressions are defined by an axiomatic theory and then taken as components of the assigned vectors. By a given interpretation (see Sect. 4.4) the first component of the vector is interpreted as an ordinal less than ε_0, thus well-ordering the derivations by complexity and proving termination of the procedure. The assignment can be adapted to a natural deduction setting for which the standard reductions decrease the ordinal of the derivation.

For the vectors of expressions of the defined axiomatic theory two operations, the box- and the delta-operation, can be defined. The two operations on vectors will be used in the assignment to create an interaction between the implication introduction and elimination rule that allows for a reduction of a convertible detour. The vector assignment can be designed so that the delta-operation is always applied when an assumption is discharged by implication introduction. When implication elimination is used, then the binary box-operation is applied, thus adding the vector of the minor premise to the vector of the major premise.

The delta-operation on a formula A produces a vector, $\delta^A \mathbf{h}$, in which the components of the corresponding variable vector, $\mathbf{x^A}$, have been eliminated. The purpose of the delta-operation is essentially to remove the dependence on the variable in the vector, because the discharge of an assumption, A, to which the variable vector, $\mathbf{x^A}$ is assigned removes dependence on the formula in the derivation. The operation eliminates the variable by substituting the constant 1 for each occurrence of the variable in question while reducing the level of exponentiation in the ordinal. The formal reduction of the exponentiation is seen in clause (4) of the definition of the delta-operation, because the pair-construction on two expressions, which corresponds to an exponentiation of the ordinals, is replaced with addition of the expressions.

The box-operation, on the other hand, is constructed to correspond to a form of addition and iteration of exponentiation in the ordinal. The length of an assigned vector is used as a parameter coding the complexity of the formula in the derivation to which the vector is assigned. The box-operation iterates the pair-construction on expressions a number of times depending on the length of the vector. This makes it possible to interpret the vector components as towers of exponentiation where the height of the exponent depends on the complexity of the formula.

To show that the reduction of an implication detour lowers the ordinal we will use the following main property of vectors (stated in Corollary 1 in Sect. 6).

$$((\delta^A \mathbf{h}) \Box \mathbf{e})_i \succ (\mathbf{h}[\mathbf{e}/\mathbf{x^A}])_i$$

for all $i \leqslant length(\mathbf{h})$. The property shows that the vector $\mathbf{h}[\mathbf{e}/\mathbf{x^A}]$, in which components of $\mathbf{e}$ have been substituted for the components of the variable vector $\mathbf{x^A}$ occurring in the vector $\mathbf{h}$, is less than the vector containing both of the defined operations.

If the vector of the premise of the implication introduction rule is $\mathbf{h}$, and the variable vector $\mathbf{x^A}$ is assigned to the discharged assumption A of the rule, then the delta-operation, $\delta^A \mathbf{h}$, essentially corresponds to the conclusion of the implication introduction. The box-operation adds the vector of the minor premise of the elimination rule, $\mathbf{e}$, to the vector of the implication introduction rule. Thus assigning the vector $(\delta^A \mathbf{h}) \Box \mathbf{e}$ to the conclusion of the reducible detour. The vector $\mathbf{h}[\mathbf{e}/\mathbf{x^A}]$ corresponds to the reduced derivation obtained by composition.

An explanation to the inequality can be given by comparing where the two vectors gain their main weight. The main weight of the reduced vector $\mathbf{h}[\mathbf{e}/\mathbf{x^A}]$ is gained when the variable of the discharged assumption, $\mathbf{x^A}$, which would be substituted with the constant 0 if the assumption was free, is substituted with an arbitrarily large vector **e** followed by the weight of the vector **h**. An upper bound for this reduced vector is found in the vector of the detour because the delta-operation, $\delta^A \mathbf{h}$, reduces the weight of the vector **h** but combines this with an increased weight to the application of the box-operation. The combination of the operations in the assignment therefore transfers weight from the vector when an assumption is discharged and reintroduces this weight when a corresponding elimination rule is applied creating a detour in the derivation.

Through the defined normalization strategy a suitable reduction is found so that no assumptions are discharged below the reduction and the confirmation of the decrease of the vector can be done without having to prove whether the delta-operation preserves an inequality (see Sect. 7).

4.3 Two operations on vectors of expressions

The axiomatic theory of expressions, the definition of vectors of expressions, the operations on vectors as well as the interpretation of the vectors as ordinals used in this proof are all taken as such from Howard's proof. These basic definitions were also used in the consistency proof [2], but the vector assignment of this paper has been adapted to prove normalization.

We define the level of a formula, which is a measure of the complexity of a formula. The level of a formula will be the length of the vector assigned to the formula.

Definition 3 The *level* of a formula A, denoted $l(A)$, is inductively defined.

1. The level of an atomic formula and falsity is 0.
2. The level of a conjunction $A\&B$ is $max\{l(A), l(B)\}$.
3. The level of an implication $A \supset B$ is $max\{l(A) + 1, l(B)\}$.
4. The level of a universally quantified formula $\forall x A$ is $l(A)$.

For a formal definition of the expressions, which are the components of the vectors, see [1] or section 5.1 in [2]. Expressions are constants (0, 1, ω), indexed variables for each formula (x_i^A), sums ($g + h$) and pairs ((g, h)). By an axiomatic theory for expressions an inequality relation, $\succ$, can be defined on the expressions.

We add the following axiom 14 to the axioms of the theory of expressions:

$$\textbf{Axiom14} : \text{If } g \prec \omega \text{ and } h \prec \omega, \text{ then } (g, h) \prec \omega. \tag{4.1}$$

Expressions are divided into classes, C_i, with the property that each expression in a class contains no variable that has a lower index than the class. This is a property that is crucial for the definition of the delta-operation to be able to eliminate variables from the vector. The class of vectors, $\mathbf{C}$, is then defined by taking the vectors, which have the i:th component from C_i, for each i. The vectors in $\mathbf{C}$ will be assigned to formulas in a derivation and the box- and the delta-operation are applied on vectors from this class.

Definition 4 The *box-operation* of two vectors $\mathbf{f}\Box\mathbf{g}$ is defined to be the vector $h = \langle h_0, \ldots, h_n\rangle$, where $n = max\{length(\mathbf{f}), length(\mathbf{g})\}$, such that

$$h_n = f_n + g_n \text{ and}$$
$$h_i = (h_{i+1}, f_i + g_i) \text{ for } 0 \leqslant i < n.$$

Definition 5 The *delta-operation* on a formula A of an expression h in $\cup C_i$, denoted $\delta^A h$, is a vector in $\mathbf{C}$ of length $l(A) + 1$ that does not contain any component of the vector $\mathbf{x^A}$. The vector is defined when the C_i, to which h belongs, is specified.

1. If h is in C_i and contains no component of $\mathbf{x^A}$, then $\delta^A h$ is the vector of length $l(A) + 1$ defined by $(\delta^A h)_i = h + 1$ and $(\delta^A h)_j = 1$, when $j \neq i$ and $0 \leqslant j \leqslant l(A) + 1$.
2. If h is x_i^A, then $(\delta^A h)_j = 1$ for $0 \leqslant j \leqslant l(A) + 1$.
3. If h contains a component of $\mathbf{x^A}$ and $h = f + g$, where f and g are in C_i, then $\delta^A h = \delta^A f + \delta^A g$.
4. If h contains a component of $\mathbf{x^A}$ and $h = (f, g)$, where f is in C_{i+1} and g is in C_i, then

$$(\delta^A h)_j = (\delta^A f)_j + (\delta^A g)_j \quad \text{if } 0 \leqslant j \leqslant l(A) \quad \text{and}$$
$$(\delta^A h)_j = 2(\delta^A f)_j + 2(\delta^A g)_j + 1 \quad \text{if } j = l(A) + 1.$$

The delta-operation is also defined for vectors $\mathbf{h} = \langle h_0, \ldots, h_n\rangle$ in $\mathbf{C}$.

$$(\delta^A \mathbf{h})_j = (\delta^A h_0)_j + \cdots + (\delta^A h_n)_j \quad \text{if } 0 \leqslant j \leqslant l(A) + 1$$

and if $n > l(A) + 1$, then we define

$$(\delta^A \mathbf{h})_j = h_j + 1 \quad \text{for } l(A) + 1 < j \leqslant n.$$

The vector $\delta^A \mathbf{h}$ has the length $max\{l(A) + 1, n\}$.

4.4 Interpretation of the vectors as ordinals

To obtain an ordinal less than ε_0 from a vector of expressions, the first component of the vector is seen as a function with its variables ranging over ordinals. If the function is applied to the constant 0, then by interpreting the resulting expression by the rules below an ordinal is obtained.

The inequality relation $a \succ b$ on expressions is interpreted as inequality of ordinals $a > b$ and $a + b$ as the natural sum $a\#b$. The definition of the pair-construction on expressions, (a, b), is separated into two cases depending on whether $b = 0$. The pair $(a, 0)$ is interpreted as 0. On the other hand, assume that $b \succ 0$ and represent b in Cantor normal form to the base 2, $b = 2^{b_1} + \cdots + 2^{b_n}$, where $b_1 > \cdots > b_n$. Then, (a, b) is $2^{c_1} + \cdots + 2^{c_n}$, where $c_i = a\#b_i$ and $1 \leqslant i \leqslant n$.

The described interpretation satisfies the axioms of the theory of expressions. Moreover, the additional Axiom 4.1 (Axiom 14) is satisfied, which is shown by the following argument. Assume that $g \prec \omega$ and $h \prec \omega$. If $h = 0$, then $(g, h) = 0 \prec \omega$. If $h \succ 0$ and $h = 2^{h_1} + \cdots + 2^{h_n}$, where $h_1 > \cdots > h_n$, then, $h_i \prec \omega$ and $(g, h) = 2^{c_1} + \cdots + 2^{c_n} \prec \omega$, because $c_i = g\#h_i \prec \omega$ for $1 \leqslant i \leqslant n$.

5 The vector assignment

We can now assign a vector, $\mathbf{f}$, to each formula, A, in a derivation by induction on the length of the derivation. The length of the vector will be the level of the formula to which the vector is assigned, $length(\mathbf{f}) = l(A)$. The main idea of the assignment is taken from [2], but it has been adapted for classical natural deduction and the task of normalization. The added constant vectors $\langle 1 \rangle$, that add 1 to the first component of the vectors, are introduced in order to ensure that application of each rule strictly increases the vector. This allows us to consider normalization procedures for subderivations in order to create a normalization strategy for the whole derivation because the shift to considering a subderivation lowers the ordinal.

The assignment of the implication introduction rule has been modified to let us substitute a constant 0 for all occurrences of a variable that corresponds to the discharged assumption of the rule. The addition of the constant vector makes it possible to consider the normalization of the subderivation of the premise of the implication introduction rule instead of the whole derivation if the introduction rule is not part of a non-normality regardless of the increase of open assumptions. Thus, we can consider subderivations even if the open assumptions of the derivation change.

The vector assignment for the induction rule has been adapted correspondingly as its reduction involves adding implication rules. The change also allows us to consider a normalization procedure for the subderivation of the second premise of the induction rule, which has discharged assumptions that change into open assumptions.

Another change form previous proofs is the additional vector that adds a factor for changing a numeral into an equal closed term. The vector $\mathbf{f}$ assigned to the derivation of $A(\overline{m}) \supset A(t)$ described in Lemma 1(ii) for some closed term t for which $t = \overline{m}$ is derivable was integrated into the vector assignment for inductions in the proof of [2]. Thus, this added vector depended on the form of the induction formula, but we will in the present proof instead add a vector with ω as components to make the assignment uniform. This change is possible, because the components of the dependent vector of [2] will always be smaller than ω and thus we can take an upper bound.

We define the restriction of a vector to increase readability: if $\mathbf{f} = \langle f_0, \ldots, f_n \rangle$, and $\mathbf{g} = \langle f_0, \ldots, f_m \rangle$ where $m \leqslant n$, then $\mathbf{g} = (\mathbf{f}) \restriction_m$ is the restricted vector.

Definition 6 The *vector assigned to a formula* in a derivation is defined as follows:

1. An assumption A is assigned the vector $\mathbf{x^A} = \langle x_0^A, \ldots, x_n^A \rangle$, where $n = l(A)$.
2. If the vector assigned to the premise of an instance of RAA with the discharged assumption $\neg P$ is $\mathbf{f} = \langle f_0 \rangle$, then the vector assigned to the conclusion of the rule is $(\delta^{\neg P} \mathbf{f} \Box \langle (1, 1), 1 \rangle) \restriction_0$.
3. The conclusion of an arithmetical rule without premises is assigned the vector $\langle 0 \rangle$.

4. If the premise of a one-premise arithmetical rule is $\langle f\rangle$, then the conclusion of the rule has the vector $\langle f+1\rangle$.
5. If the premises of an instance of transitivity are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule is assigned the vector $\mathbf{f}+\mathbf{g}+\langle 1\rangle$.
6. If the premises of an instance of $\&I$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule is assigned a vector $\mathbf{f}+\mathbf{g}+\langle 1\rangle$.
7. If the premise of an instance of $\&E$ is assigned the vector $\mathbf{f}$ and the conclusion is the formula A, then the conclusion of the rule is assigned a vector $\mathbf{f}\restriction_{l(A)} +\langle 1\rangle$. The vector is similar if the conclusion of the rule is B.
8. If the premise of $\supset I$ is assigned the vector $\mathbf{f}$ and the formula in the conclusion of the rule is $A\supset B$, then the vector assigned to the conclusion of the rule is $\delta^A\mathbf{f}\square\langle 0,\ldots,0\rangle$. The length of the constant vector is $l(A)$.
9. If the premises of $\supset E$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$ and the principal formula of the rule is $A\supset B$, then the conclusion of the rule is assigned the vector $(\mathbf{f}\square\mathbf{g})\restriction_{l(B)} +\langle 1\rangle$.
10. If the premise of $\forall I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule has the vector $\mathbf{f}+\langle 1\rangle$.
11. If the premise of $\forall E$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned the vector $\mathbf{f}+\langle 1\rangle$.
12. If the principal formula in the conclusion of an instance of Ind is $A(t)$ and $l(A)=n$, then the vector assigned to the conclusion of the rule depends on the term t.
    (a) If t is a closed term, then there is a derivation of $t=\overline{m}$ for some unique numeral $\overline{m}$ according to Lemma 1. If the vectors assigned to the first and second premise of the Ind-rule are $\mathbf{h}$ and $\mathbf{g}$ respectively, then the vector of the conclusion of the induction is

    $$((\langle\omega,\ldots,\omega,2(m+1)\rangle\square(\delta^{A(x)}\mathbf{g}\square\langle 0,\ldots,0\rangle))\restriction_{n+1}\square\mathbf{h})\restriction_n,$$

    where the length of the vector $\langle\omega,\ldots,\omega,2(m+1)\rangle$ is $n+2=l(A)+2$ and the length of the constant zero vector is n.
    (b) If on the other hand the term t contains a variable, then the vector of the conclusion of the induction is

    $$((\langle\omega,\ldots,\omega\rangle\square(\delta^{A(x)}\mathbf{g}\square\langle 0,\ldots,0\rangle))\restriction_{n+1}\square\mathbf{h})\restriction_n,$$

    where the length of the ω-vector, $\langle\omega,\ldots,\omega\rangle$, is $n+2=l(A)+2$ and the length of the constant zero vector is n.

The vector assigned to the conclusion of a derivation is the vector assigned to the whole derivation.

6 Properties of the operations on vectors

We state some lemmas that prove properties for the operations on vectors. The lemmas are based on or are restatements of lemmas found in [2].

The following lemma gives an upper limit for the components of vectors assigned to derivations without induction.

Lemma 2 *Let* $\mathbf{f} = \langle f_0, \ldots, f_{n+1} \rangle$*, where* $n = l(A)$*, be the vector assigned to the derivation of* $A(\overline{m}) \supset A(t)$ *described in Lemma 1(ii) for some closed term* t *for which* $t = \overline{m}$ *is derivable. Then* $f_i \prec \omega$ *for all* i.

Proof The lemma is proved by induction on the construction of the expression. The derivation contains no open assumptions and therefore the vector contains no variables. If f_i is the constant 0 or 1, then by axiom 6 of the theory of expressions the claim holds. If $f_i = g + h$ and $g \prec \omega$ and $h \prec \omega$, then by axiom 7 of the theory of expressions the claim holds. If $f_i = (g, h)$ and $g \prec \omega$ and $h \prec \omega$, then by the additional axiom 14 of the theory of expressions the claim holds. Because the derivation does not contain any induction inferences, the expression ω will never be introduced in the vector. □

The following monotonicity properties of the operations on vectors are stated in [2] as Lemmas 7.1–7.3.

Lemma 3 $(\mathbf{f}\Box\mathbf{g})_i \succcurlyeq f_i$ *for all* i.

The corresponding inequality holds for the vector $\mathbf{g}$ because of the commutativity of the box-operation.

Lemma 4 *Assume that* $length(\mathbf{f}) = length(\mathbf{g}) = n$*, and* $f_i \succcurlyeq g_i$ *for* $0 \leqslant i \leqslant n$*, then* $(\mathbf{f}\Box\mathbf{h})_i \succcurlyeq (\mathbf{g}\Box\mathbf{h})_i$ *for all* i.

Lemma 5 *Under the assumption of Lemma 4 and the additional assumption* $f_i \succ g_i$ *for* $0 \leqslant i \leqslant k$*, and some* $k \leqslant n$*, the inequality* $(\mathbf{f}\Box\mathbf{h})_i \succ (\mathbf{g}\Box\mathbf{h})_i$ *holds for* $0 \leqslant i \leqslant k$.

The following lemma, proved in [2] as lemma 7.8, shows that we can substitute a term for a free variable without increasing the components of the vector. The only change that can occur in the vector is the possible change of assignment to induction rules. This lemma will be used in the reductions of numerical inductions and detours with a universal maximum formula.

Lemma 6 *If there is a derivation to which the vector* $\mathbf{f}$ *is assigned and another derivation to which the vector* $\mathbf{g}$ *is assigned and the latter derivation is obtained from the former by substituting a term for a free variable, then* $f_i \succcurlyeq g_i$ *for* $0 \leqslant i \leqslant length(\mathbf{f})$.

The following lemma and corollary states the main property of vectors. The result is stated as lemma 7.5 and corollary 7.6 in [2].

Lemma 7 *Let* $\mathbf{e}$ *be a vector of length* $l(A)$ *and assume* h *is in* C_i*. Then* $((\delta^A h)\Box\mathbf{e})_i \succ h[\mathbf{e}/\mathbf{x^A}]$.

Proof The lemma is proved by induction on the number of times clauses 3 and 4 in the definition of the delta-operation are applied in $\delta^A h$. A complete proof is found in lemma 2.11 of [1]. □

Corollary 1 *If* $\mathbf{h}$ *is in* $\mathbf{C}$ *and* $\mathbf{e}$ *has length* $l(A)$*, then* $((\delta^A \mathbf{h})\Box\mathbf{e})_i \succ (\mathbf{h}[\mathbf{e}/\mathbf{x^A}])_i$*, for all* $i \leqslant length(\mathbf{h})$.

The following lemma shows that the assignment of a composed derivation corresponds to substitution in the vectors. The result is stated as lemma 7.7 in [2].

Lemma 8 *Assume that there is a derivation of A to which the vector* $\mathbf{f}$ *is assigned and a derivation of B to which the vector* $\mathbf{g}$ *is assigned and that A is an open assumption in the latter derivation. Assume furthermore that no open assumption in the derivation of A becomes discarded in the derivation of B, where all assumptions A have been replaced with the derivation of A. Then the vector assigned to this derivation is* $\mathbf{g}[\mathbf{f}/\mathbf{x}^{\mathbf{A}}]$.

We prove two lemmas that show how a constant vector can be moved relative to the box-operation.

Lemma 9 *For all vectors we have* $(\mathbf{a} + \langle 1\rangle)\Box\mathbf{b} = \mathbf{a}\Box(\mathbf{b} + \langle 1\rangle)$.

Proof We have that for all $i > 0$ $(\mathbf{a} + \langle 1\rangle)_i = a_i$ and similarly for $\mathbf{b}$, thus by the Definition 4 of the box-operation w$((\mathbf{a} + \langle 1\rangle)\Box\mathbf{b})_i = (\mathbf{a}\Box\mathbf{b})_i = (\mathbf{a}\Box(\mathbf{b} + \langle 1\rangle))_i$.

For $i = 0$ we get by the Definition 4 of the box-operation that $(\mathbf{a} + \langle 1\rangle)\Box\mathbf{b})_0 = ((\mathbf{a}\Box\mathbf{b})_1, a_0 + 1 + b_0) = (\mathbf{a}\Box(\mathbf{b} + \langle 1\rangle))_0$. □

Lemma 10 *For all vectors we have* $((\mathbf{a} + \langle 1\rangle)\Box\mathbf{b})_i \succcurlyeq ((\mathbf{a}\Box\mathbf{b}) + \langle 1\rangle)_i$ *for* $i \leqslant max\{length(\mathbf{a}), length(\mathbf{b})\}$.

Proof By the Definition 4 of the box-operation we have that for all $i > 0$ $(\mathbf{a}+\langle 1\rangle)_i = a_i$, thus $((\mathbf{a} + \langle 1\rangle)\Box\mathbf{b})_i = (\mathbf{a}\Box\mathbf{b})_i = ((\mathbf{a}\Box\mathbf{b}) + \langle 1\rangle)_i$.

For $i = 0$ by the Definition 4 of the box-operation, by Axiom 8 for the theory of expressions, by axiom 11, axiom 12 and by the definition of the box-operation respectively we get that $(\mathbf{a}+\langle 1\rangle)\Box\mathbf{b})_0 = ((\mathbf{a}\Box\mathbf{b})_1, a_0+1+b_0) = ((\mathbf{a}\Box\mathbf{b})_1, a_0+b_0)+((\mathbf{a}\Box\mathbf{b})_1, 1) \succ ((\mathbf{a}\Box\mathbf{b})_1, a_0+b_0)+(0, 1) = (\mathbf{a}\Box\mathbf{b})_1, a_0+b_0)+1 = ((\mathbf{a}\Box\mathbf{b})+\langle 1\rangle)_0$. □

The following lemma is a variant of the lemma 7.9 in [2] for vector calculations for reductions of numerical inductions, where the induction term equals a successor. We base the proof on the mentioned lemma 7.9 in [2].

Lemma 11 *Let* $\mathbf{g}' = \delta^{A(x)}\mathbf{g}\Box\langle 0, \ldots, 0\rangle$ *and*

$$\mathbf{e} = ((\langle\omega, \ldots, \omega, 2m + 4\rangle\Box\mathbf{g}') \upharpoonright_{n+1} \Box\mathbf{h}) \upharpoonright_n,$$

and let

$$\mathbf{e}' = (((((\langle\omega, \ldots, \omega, 2m + 2\rangle\Box\mathbf{g}') \upharpoonright_{n+1} \Box\mathbf{h}) \upharpoonright_n \Box\mathbf{g}') + \langle 1\rangle) \upharpoonright_n \Box\mathbf{f}) \upharpoonright_n +\langle 1\rangle,$$

where $length(\mathbf{f}) = n+1$ *and* $length(\mathbf{h}) = length(\mathbf{g}) = n$. *Furthermore,* $0 \prec f_i \prec \omega$ *for* $i \leqslant n + 1$.

Then $e'_i \prec e_i$ *for* $0 \leqslant i \leqslant n$.

Proof The differences in the vectors of the stated lemma compared to earlier results, such as lemma 7.9 of [2], is that we have replaced the vector $\delta^{A(x)}\mathbf{g}$ by $\delta^{A(x)}\mathbf{g}\Box$ $\langle 0, \ldots, 0\rangle$ because of the assignment to the implication introduction rule and we have replaced the vector components f_i with ω in the vector assigned to the induction.

To show that these changes are superficial we can define the following vector:

$$\mathbf{e}'' = ((((\langle\omega, \ldots, \omega, 2m+2\rangle \square \mathbf{g}') \restriction_{n+1} \square \mathbf{h}) \restriction_n \square \mathbf{g}') \restriction_n \square \langle\omega, \ldots, \omega\rangle) \restriction_n,$$

where the length of the constant ω-vector, $\langle\omega, \ldots, \omega\rangle$ is $n+1$. That is, the length of the ω-vector is the same as the length of the vector $\mathbf{f}$.

In the proof of lemma 7.9 of [2] we only use the fact that the vector $\mathbf{f}$ is $\delta^A \mathbf{f}'$ for some vector $\mathbf{f}'$ and the similar fact for the vector $\delta^{A(x)}\mathbf{g}$, which is used instead of $\delta^{A(x)}\mathbf{g}\square\langle 0, \ldots, 0\rangle$, to be able to assume that the vectors components are greater than 0. But this also holds for $\omega \succ 0$ and $(\delta^{A(x)}\mathbf{g}\square\langle 0, \ldots, 0\rangle)_i \succ 0$ for each $i \leqslant n+1$. Thus, we can replace the vector $\mathbf{f}$ by the ω-vector and the vector $\delta^{A(x)}\mathbf{g}$ by $\delta^{A(x)}\mathbf{g}\square\langle 0, \ldots, 0\rangle$ and by lemma 7.9 of [2] conclude that $e''_i \prec e_i$ for $0 \leqslant i \leqslant n$.

Therefore, if we can prove that $e'_i \prec e''_i$ for $0 \leqslant i \leqslant n$ then we are done.

To prove this claim we define the following vector

$$\mathbf{r} = (((\langle\omega, \ldots, \omega, 2m+2\rangle \square \mathbf{g}') \restriction_{n+1} \square \mathbf{h}) \restriction_n \square \mathbf{g}') \restriction_n .$$

We conclude that $\mathbf{e}' = ((\mathbf{r} + \langle 1\rangle)\square\mathbf{f}) \restriction_n + \langle 1\rangle$ and $\mathbf{e}'' = (\mathbf{r}\square\langle\omega, \ldots, \omega\rangle) \restriction_n$.

By Lemma 9 we have $e'_i = (((\mathbf{r}+\langle 1\rangle)\square\mathbf{f}) \restriction_n + \langle 1\rangle)_i = ((\mathbf{r}\square(\mathbf{f}+\langle 1\rangle)) \restriction_n + \langle 1\rangle)_i = ((\mathbf{r}\square(\mathbf{f}+\langle 1\rangle)) + \langle 1\rangle) \restriction_n)_i$.

By Lemma 10 we have $((\mathbf{r}\square(\mathbf{f}+\langle 1\rangle)) + \langle 1\rangle)_i \prec (\mathbf{r}\square(\mathbf{f}+\langle 2\rangle))_i$.

By Lemma 2 and Axiom 7 for the theory of expressions we have $f_i + 2 \prec \omega$ for all i and therefore by Lemma 5 we get $e'_i \prec (\mathbf{r}\square\langle\omega, \ldots, \omega\rangle)_i = e''_i$ for $i \leqslant n$.

This proves the claim. $\square$

7 Normalization of Peano Arithmetic

To prove normalization of Peano Arithmetic we will define a normalization strategy, which will show in which order we should choose the non-normalities that we reduce. The restricted reductions of [1] with a unique ordinal assignment correspond to a limitation in choice of the considered reducibility in the PA-derivation. The reducibility may not be a part of a subderivation that discharges assumptions, because then there would be an application of the delta-operation on the corresponding vector. The problem, which Howard states, is that general reductions, corresponding to a proof of strong normalization, do not necessarily reduce the vector because the delta-operation applied on vectors does not necessarily preserve inequalities. The reason for this is that the different vectors can fall under separate clauses of the definition of the delta-operation. If $f \prec g$, then the inequality $(\delta^A f)_i \prec (\delta^A g)_i$ does not hold in general when the expressions f and g differ in structure. However, even if general reductions cannot be treated, a suitable normalization strategy can be found. The normalization strategy we will use is based on the normalization strategy of [4].

We define what a normal derivation is and define a main elimination path that we will use to give the normalization strategy.

Definition 7 A *main elimination path* of a derivation is an inductively defined list of formulas, $(A_0, \ldots, A_n)$. The end formula of the derivation, A_0, belongs to the main elimination path. If a formula, A_i, which is included in the main elimination path is the conclusion of an elimination rule, then the major premise, A_{i+1}, of the rule is included in the main elimination path. The *length* of the main elimination path is the length, n, of the list. The formula A_n is the *end* of the main elimination path.

A *maximum formula* is a formula that has been concluded by an introduction rule and is the major premise of an elimination rule.

A *normal derivation* (i) contains no maximum formula and (ii) all inductions conclude a formula $A(t)$ where the term t contains a variable. Furthermore, (iii) no instance of RAA in a normal derivation concludes the formula $\perp$.

Thus, a derivation with a maximum formula, a numerical induction or an instance of indirect proof with falsity as the conclusion is called non-normal.

The method of finding a suitable non-normality to reduce is divided into four cases depending on the form of the derivation. The four cases are mutually exclusive and cover all the cases of rules. Therefore, one of the cases applies for each non-normal derivation.

Cases I–III apply if we have a non-normality at the end of the main elimination path. In Case IV normalization of the subderivations gives a normalization procedure for the whole derivation.

Definition 8 (*Normalization strategy)* We consider the formula at the end of the main elimination path. If this formula is derived by or part of a non-normality, then we normalize this non-normality. Otherwise we consider subderivaitons.

Case I If the formula we are considering is a maximum formula, then we reduce this non-normality by a detour conversion.

Case II If the formula we are considering is derived by a non-normal instance of an induction rule, then this case is reduced by an induction reduction.

Case III If the formula we are considering is derived by a non-normal instance of RAA, then this case is reduced by a RAA reduction.

Case IV If the formula we are considering is not part of a non-normality, then the formula is an assumption or is derived by an arithmetical rule, a normal instance of induction, a normal instance of RAA. The formula can also be derived by an introduction rule if the length of the main elimination path is 0. In these cases we normalize the derivation by normalizing the subderivations of the potential premises of the rule and the minor premises of potential implication elimination rules in the main elimination path.

We need to reduce the non-normalities found in the cases I-III (this will be done in Theorem 1), obtain a derivation with a lower ordinal and show that the rules of the main elimination path preserve the inequality (Lemma 12). We also need to prove that considering the subderivations as in Case IV decreases the ordinal. This will be done in Sect. 9.

8 The normalization reductions

We can show that reducing a non-normality if it occurs at the end of the main elimination path of the derivation decreases the ordinal of the derivation. These reductions are the Cases I–III of our normalization strategy in Definition 8.

Theorem 1 (Normalization reductions of PA) *If the last rule of a derivation, P to which the vector* $\mathbf{a}$ *is assigned, is a non-normality, then there is a derivation, P', of the same formula to which the vector* $\mathbf{b}$ *is assigned and $a_0 \succ b_0$ and $a_i \succcurlyeq b_i$ for* $0 < i \leqslant length(\mathbf{a}) = length(\mathbf{b})$.

Proof The reduction applied in the derivation depends on the kind of non-normality that is found in the derivation by our normalization strategy. All cases are listed below.

Reductio ad absurdum If there is a non-normal instance of RAA, which concludes $\perp$, then it is possible to eliminate the rule by replacing the derivation

$$\dfrac{\begin{array}{c}[\neg\perp]\\ \vdots\\ \perp\end{array}}{\perp}\,RAA$$

by the derivation

$$\begin{array}{c}\dfrac{[\perp]}{\perp\supset\perp}\supset I\\ \vdots\\ \perp\end{array}$$

We let the vector assigned to the assumption of the instance of RAA in the derivation P be $\mathbf{h}$. The vector assigned to the conclusion of the instance of RAA is $(\delta^{\neg\perp}\mathbf{h}\square\langle(1,1),1\rangle)\restriction_0$.

By Case 2 of the definition of the delta-operation we have that $\delta^{\perp}x_0^{\perp} = \langle 1, 1\rangle$ and therefore we calculate that $\delta^{\perp}x_0^{\perp}\square\langle 0\rangle = \langle(1,1),1\rangle$.

By Lemma 8 the vector assigned to the reduced derivation is $\mathbf{h}[(\delta^{\perp}x_0^{\perp}\square\langle 0\rangle)/\mathbf{x}^{\neg\perp}]$.

Thus, by Corollary 1 we get that $\mathbf{h}[(\delta^{\perp}x_0^{\perp}\square\langle 0\rangle)/\mathbf{x}^{\neg\perp}] = \mathbf{h}[\langle(1,1),1\rangle/\mathbf{x}^{\neg\perp}] \prec (\delta^{\neg\perp}\mathbf{h}\square\langle(1,1),1\rangle)\restriction_0$. This proves the claim for the RAA reduction.

Induction reduction. Assume that there is a non-normal instance of an induction rule. This rule concludes a formula $A(t)$ with a closed term t and there exists a derivation of $t = \overline{m}$ for some numeral $\overline{m}$ according to Lemma 1(i).

$$\dfrac{\begin{array}{c}\\ \vdots\\ A(0)\end{array}\quad\begin{array}{c}[A(x)]\\ \vdots\\ A(sx)\end{array}}{A(t)}\,Ind$$

The reduction now applied depends on the numeral $\overline{m}$.

Case 1 If $\overline{m} \equiv 0$, then according to Lemma 1(ii) there is a derivation of $A(0) \supset A(t)$ without inductions. The components of the vector assigned to this derivation $\mathbf{f}$ *will all be less than ω according to Lemma 2. The reduced derivation is composed by implication elimination with the first premise of the induction as minor premise.*

$$\dfrac{\begin{matrix}\vdots & \vdots\\ A(0)\supset A(t) & A(0)\end{matrix}}{A(t)}\supset E$$

Let $\mathbf{h}$ *and* $\mathbf{g}$ *be the vectors assigned to the premises of the induction in* P. *Then the vector assigned to the conclusion of the induction rule in* P *is* $((\langle\omega,\ldots,\omega,2\rangle\Box(\delta^{A(x)}\mathbf{g}\Box\langle 0,\ldots,0\rangle))\restriction_{n+1}\Box\mathbf{h})\restriction_n$. *The vector assigned to the reduced derivation is* $(\mathbf{f}\Box\mathbf{h})\restriction_n+\langle 1\rangle$.

By Lemma 10 we have that $((\mathbf{f}\Box\mathbf{h})\restriction_n+\langle 1\rangle)_i\preccurlyeq(((\mathbf{f}+\langle 1\rangle)\Box\mathbf{h})\restriction_n)_i$.

By Lemma 2 and axiom 7 for the theory of expressions we have $f_i+1\prec\omega$ *for all* i *and therefore by Lemmas 3 and 5 we get* $((((\langle\omega,\ldots,\omega,2\rangle\Box(\delta^{A(x)}\mathbf{g}\Box\langle 0,\ldots,0\rangle))\restriction_{n+1}\Box\mathbf{h})\restriction_n)_i\succcurlyeq(((\langle\omega,\ldots,\omega\rangle)\Box\mathbf{h})\restriction_n)_i\succ(((\mathbf{f}+\langle 1\rangle)\Box\mathbf{h})\restriction_n)_i\succcurlyeq((\mathbf{f}\Box\mathbf{h})\restriction_n+\langle 1\rangle)_i$.

Case 2 If $\overline{m}\equiv s(\overline{m'})$ *for some numeral* $\overline{m'}$, *then according to Lemma 1(ii) there is a derivation of* $A(s\overline{m'})\supset A(t)$ *without inductions with the vector* $\mathbf{f}$. *Thus, we have* $f_i\prec\omega$ *by Lemma 2. The following reduction on the derivation is applied:*

$$\dfrac{\begin{matrix}\vdots\\ A(s\overline{m'})\supset A(t)\end{matrix}\qquad \dfrac{\dfrac{\begin{matrix}[A(\overline{m'})]\\ \vdots\\ A(s\overline{m'})\end{matrix}}{A(\overline{m'})\supset A(s\overline{m'})}\supset I\qquad \dfrac{\begin{matrix}\vdots & [A(x)]\\ & \vdots\\ A(0) & A(sx)\end{matrix}}{A(\overline{m'})}\,Ind}{A(s\overline{m'})}\supset E}{A(t)}\supset E$$

The derivation of $A(s\overline{m'})$ *from* $A(\overline{m'})$ *is the second premise of the original induction with* $\overline{m'}$ *substituted for* x.

Let $\mathbf{h}$ *and* $\mathbf{g}$ *be as in Case 1 and let* $\mathbf{g}'=\delta^{A(x)}\mathbf{g}\Box\langle 0,\ldots,0\rangle$. *Then the vector assigned to the conclusion of the induction rule in* P *is*

$$\mathbf{e}=((\langle\omega,\ldots,\omega,2m+4\rangle\Box\mathbf{g}')\restriction_{n+1}\Box\mathbf{h})\restriction_n .$$

The vector assigned to the conclusion of the induction rule in the reduced derivation is $((\langle\omega,\ldots,\omega,2m+2\rangle\Box\mathbf{g}')\restriction_{n+1}\Box\mathbf{h})\restriction_n$. *The vector assigned to the conclusion of the implication introduction rule in the reduced derivation is* $\mathbf{g_1'}=\delta^{A(x)}\mathbf{g_1}\Box\langle 0,\ldots,0\rangle$, *where* $\mathbf{g_1}$ *is obtained from* $\mathbf{g}$ *by substitution of the numeral for the induction variable in the derivation. By Lemma 6 the vector* $\mathbf{g_1}$ *is not greater than* $\mathbf{g}$ *and furthermore the structure of the expressions of the vector is preserved during the substitution, so the inequality is preserved by the application of the delta-operation.*

Thus, the vector assigned to the conclusion of the reduced part of the derivation in P' *is less than or equal to*

$$\mathbf{e}'=(((((\langle\omega,\ldots,\omega,2m+2\rangle\Box\mathbf{g}')\restriction_{n+1}\Box\mathbf{h})\restriction_n\Box\mathbf{g}')+\langle 1\rangle)\restriction_n\Box\mathbf{f})\restriction_n+\langle 1\rangle.$$

The claim that $e_i\succ e_i'$ *for* $i\leqslant n$ *is proved by Lemma 11.*

Detour conversion *Assume that there is a convertible detour in the derivation. The reductions are standard conversions depending on the outermost logical connective of the maximum formula.*

Case 1 If the formula is an implication, then the derivation has the form:

$$\dfrac{\dfrac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}\supset I \quad \begin{matrix}\vdots\\ A\end{matrix}}{B}\supset E$$

and this is reduced into the following derivation:

$$\begin{matrix}\vdots\\ A\\ \vdots\\ B\end{matrix}$$

Let the vector assigned to the premise of the implication introduction rule in P be $\mathbf{f}$ *and the vector assigned to the minor premise of the implication elimination rule be* $\mathbf{g}$. *Then the vector assigned to the conclusion of the elimination is* $((\delta^A \mathbf{f} \Box \langle 0, \ldots, 0\rangle)\Box \mathbf{g}) \restriction_{l(B)}$ $+\langle 1\rangle$. *This vector is reduced to the vector* $\mathbf{f}[\mathbf{g}/\mathbf{x^A}]$ *by Lemma 8.*

By monotonicity of the box-operation, Lemma 3, we get that $(\delta^A \mathbf{f} \Box \langle 0, \ldots, 0\rangle)_i \succcurlyeq (\delta^A \mathbf{f})_i$.

Thus, $((\delta^A \mathbf{f} \Box \langle 0, \ldots, 0\rangle)\Box \mathbf{g})_i \succ (\delta^A \mathbf{f} \Box \mathbf{g})_i \succ (\mathbf{f}[\mathbf{g}/\mathbf{x^A}])_i$ *by Lemma 5 and Corollary 1.*

Case 2 We assume that the maximum formula is a conjunction. Let the vectors assigned to the premises of the conjunction introduction rule in P be $\mathbf{f}$ *and* $\mathbf{g}$. *Assume that A is the conclusion of the elimination rule. Then the vector assigned to the conclusion of the elimination is* $(\langle f_0 + g_0 + 2, \ldots, f_n + g_n\rangle) \restriction_{l(A)}$, *where* $n = max\{length(\mathbf{f}), length(\mathbf{g})\}$. *This vector is reduced to the vector* $\mathbf{f}$ *for which the claim holds.*

Case 3 We assume that the maximum formula is a universal formula. If the vector assigned to the premise of the universal introduction rule in P is $\mathbf{f}$, *then the vector assigned to the conclusion of the elimination rule is* $\langle f_0 + 2, \ldots, f_n\rangle$. *In the derivation of the premise of the introduction rule, $A(x)$, the term t can be substituted for x. If t contains a variable, then the vector,* $\mathbf{f}'$, *of the derivation that results from the reduction remains unchanged and equal to* $\mathbf{f}$ *and if t is closed, then by Lemma 6 the components of the vector are not increased. Thus,* $f_0 + 2 \succ f_0 \succcurlyeq f_0'$ *for the first component and the inequalities for the other components hold.* ◻

We show that the inequalities produced by the reductions of Theorem 1 are preserved if they occur at the end of a main elimination path.

Lemma 12 *If the length of the main elimination path is greater than 0, then the elimination rules preserve the ordinal inequality produced by a reduction of a non-normality at the end of the main elimination path.*

Proof This claim is proved by induction on the number of rules below the reduced part, that is, by the number of formulas in the main elimination path. Since the reduced part will only have major premises of elimination rules below it, the inductive step is proved by monotonicity of the box-operation and addition.

We assume that $\mathbf{f}_0 \succ \mathbf{g}_0$ and $\mathbf{f}_i \succcurlyeq \mathbf{g}_i$ for $0 < i \leqslant length(\mathbf{f})$, where $\mathbf{f}$ is the original vector and $\mathbf{g}$ is the reduced vector.

If the rule is an instance of $\&E$ and the conclusion is the formula A, then the conclusion of the rule in the original derivation has the vector $\mathbf{f} \restriction_{l(A)} +\langle 1\rangle$. Thus, for the first component of the vectors we get the desired inequality $\mathbf{f}_0 + 1 \succ \mathbf{g}_0 + 1$. The other components remain the same.

If the rule is an instance of $\supset E$ and the minor premise of the rule is assigned the vector $\mathbf{h}$ and the principal formula of the rule is $A \supset B$, then the conclusion of the rule in the original derivation has the vector $(\mathbf{f}\Box\mathbf{h}) \restriction_{l(B)} +\langle 1\rangle$. Thus, Lemma 5 proves the claim.

If the rule is an instance of $\forall E$, then the conclusion of the rule in the original derivation has the vector $\mathbf{f} + \langle 1\rangle$. Thus, the inequalities hold as for the conjunction rule. □

9 Considering subderivations

We show that considering subderivations in our normalization strategy of Definition 8 lets us deal with derivations with a lower ordinal. We will deal with Case IV and show that if there is no non-normality at the end of the main elimination path, then the subderivations of the premises will have lower ordinals and we can consider them instead. This might seem trivial, but we cannot for example assume that $(\delta^A\mathbf{f})_i \succ f_i$, because the delta-operation decreases the ordinal by removing the pair-construction on expressions. This is the reason for always combining the delta-operation with an application of the box-operation, which lets us use the main property of vectors (Corollary 1).

We have three lemmas. Lemma 13 shows that the minor premise of an implication elimination rule has a lower ordinal if the main elimination path is normal. Lemma 14 deals with the normal instances of induction and RAA. Lemma 15 deals with the normal instances of introduction and arithmetical rules.

Lemma 13 *The minor premise of an instance of an implication elimination rule, has a lower ordinal than the conclusion of the rule.*

Proof If the premises of $\supset E$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$ and the principal formula of the rule is $A \supset B$, then the conclusion of the rule has the vector $(\mathbf{f}\Box\mathbf{g}) \restriction_{l(B)} +\langle 1\rangle$. Thus, for the first component of the vectors we get the desired inequality for the minor premise $\mathbf{g}_0 \preccurlyeq (\mathbf{f}\Box\mathbf{g})_0 \prec (\mathbf{f}\Box\mathbf{g})_0 + 1$ by the monotonicity of addition and the box-operation (see Lemma 3). □

The second Lemma 14 shows that we can consider the subderivations if the main elimination path ends with a normal instance of induction or RAA.

Lemma 14 *The premises of a normal instance of induction or RAA have lower ordinals than the conclusion of the rule.*

Proof If the vector assigned to the premise of a normal instance of RAA with the discharged assumption $\neg P$ is $\mathbf{f}$, then the vector assigned to the conclusion of the rule is $(\delta^{\neg P}\mathbf{f}\Box\langle(1,1),1\rangle)\restriction_0$. The ordinal assigned to the subderivation obtained after the interpretation comes from the first component of the vector $\mathbf{f}$ or $\mathbf{f}_0[\langle 0,0\rangle/\mathbf{x}^{\neg\mathbf{P}}]$, after substituting the constant 0 for each component of the variable vector $x^{\neg P}$. By monotonicity of the box-operation we have the inequality $(\delta^{\neg P}\mathbf{f}\Box\langle(1,1),1\rangle)\restriction_0 \succ (\delta^{\neg P}\mathbf{f}\Box\langle 0,0\rangle)\restriction_0$. Thus, for the first component of the vectors we get the desired inequality $\mathbf{f}_0[\langle 0,0\rangle/\mathbf{x}^{\neg\mathbf{P}}] \prec (\delta^{\neg P}\mathbf{f}\Box\langle 0,0\rangle)_0 \prec (\delta^{\neg P}\mathbf{f}\Box\langle(1,1),1\rangle)_0$ by Corollary 1.

If the vectors assigned to the premises of a normal instance of induction are $\mathbf{h}$ and $\mathbf{g}$, then the vector of the conclusion of the induction is

$$((\langle\omega,\ldots,\omega,\omega\rangle\Box(\delta^{A(x)}\mathbf{g}\Box\langle 0,\ldots,0\rangle))\restriction_{n+1}\Box\mathbf{h})\restriction_n,$$

where the length of the ω-vector, $\langle\omega,\ldots,\omega,\omega\rangle$, is $n+2=l(A)+2$.

The claim for the first premise of the induction rule with the vector $\mathbf{h}$ follows from the monotonicity of the box-operation, Lemma 3.

Now, consider the second premise. By Corollary 1 the vector $\delta^{A(x)}\mathbf{g}\Box\langle 0,\ldots,0\rangle$ is greater than the vector $\mathbf{g}[\langle 0,\ldots,0\rangle/\mathbf{x}^{\mathbf{A}}]$. This vector is the vector $\mathbf{g}$ assigned to the derivation of the second premise of the induction after the substitution of the constant 0 for free variables in the interpretation of vectors as ordinals. The desired inequality then follows from monotonicity of the box-operation. □

The third Lemma 15 shows that we can consider the subderivations if the main elimination path ends with a normal instance of an introduction or an arithmetical rule.

Lemma 15 *The premises of a normal instance of an introduction or arithmetical rule will have a lower ordinal than the conclusion of the rule.*

Proof If the last rule of the derivation is an arithmetical one-premise or two-premise rule, then the claim follows from monotonicity of addition, because $\mathbf{f}_0 \preccurlyeq \mathbf{f}_0+1$, $\mathbf{f}_0 \preccurlyeq \mathbf{f}_0+\mathbf{g}_0+1$ and $\mathbf{g}_0 \preccurlyeq \mathbf{f}_0+\mathbf{g}_0+1$.

If the premises of an instance of $\&I$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule has the vector $\mathbf{f}+\mathbf{g}+\langle 1\rangle$. Thus, the claim is proved by the monotonicity of addition.

If the premise of $\supset I$ is assigned the vector $\mathbf{f}$ and the formula in the conclusion of the rule is $A\supset B$, then the vector assigned to the conclusion of the rule is $\delta^A\mathbf{f}\Box\langle 0,\ldots,0\rangle$, where the length of the constant vector is $l(A)$. By Corollary 1 this vector is greater than the vector $\mathbf{f}[\langle 0,\ldots,0\rangle/\mathbf{x}^{\mathbf{A}}]$. This vector is the vector f assigned to the derivation of the premise of the implication introduction after the substitution of the constant 0 for free variables in the interpretation of vectors as ordinals.

If the premise of $\forall I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule has the vector $\mathbf{f}+\langle 1\rangle$. Thus, the claim is proved by the monotonicity of addition. □

10 Termination of the normalization procedure

We can conclude that the normalization procedure of the normalization strategy found in Definition 8 terminates because by each step of the strategy we obtain a lower ordinal. Thus, by the well-ordering of the ordinals the procedure must terminate in a normal derivation.

Theorem 2 *The normalization procedure defined by the strategy of Definition 8 terminates in a normal derivation.*

Proof In the normalization strategy of Definition 8 we identify the main elimination path and then normalize the rule at the end of the path as in Case I–III, obtaining a derivation with a lower ordinal by Theorem 1. That the ordinal inequality is preserved by the elimination rules in the main elimination path is shown by Lemma 12.

If the main elimination path does not end in a non-normality, then we consider the subderivations of the premises of the formula at the end of the main elimination path as in Lemmas 14 and 15. We also consider the subderivations of the minor premises of the implication elimination rules in the main elimination path as in Lemma 13. These lemmas show that one normal rule increases the ordinal and by Lemma 12 the whole original derivation has a greater ordinal than the premises of the normal rule.

Therefore, considering these subderivations and their ordinals we will get a finite number of ordinals, which are all less than the ordinal of the previous step of normalization. Each of these ordinal branchings will terminate, showing that each subderivation normalizes. Thus, by the well-ordering of the ordinal ε_0 this reduction strategy for the derivation must terminate in a normal derivation. □

References

1. Howard, W.A.: Assignment of ordinals to terms for primitive recursive functionals of finite type. In: Kino, J.M.A., Vesley, R. (eds.) Intuitionism and Proof Theory. North-Holland, Amsterdam (1970)
2. Kanckos, A.: Consistency of Heyting arithmetic in natural deduction. Math. Log. Q. **56**(6), 611–624 (2010)
3. von Plato, J.: Normal form and existence property for derivations in Heyting arithmetic. Acta Philos. Fennica **78**, 159–163 (2006)
4. Prawitz, D.: Natural deduction: a proof-theoretical study, 2nd edn. Dover Publications, UK (2006)
5. Prawitz, D.: A note on Gentzen's 2nd consistency proof and normalization of natural deductions in 1st order arithmetic. In: Kahle, R., Rathjen, M. (eds.) Gentzen's Centenary: The Quest for Consistency. Springer (2015)
6. Siders, A.: Gentzen's consistency proof without heightlines. Arch. Math. Logic. **52**(3–4), 449–468 (2013)