A Constructive Version of the Hilbert Basis Theorem

Aaron Hertz

May 7, 2004

1 Introduction

The Hilbert Basis Theorem was the first major example of a non-constructive proof recognized in mathematics. Gordan said, on the subject of the theorem, “das ist keine Mathematik, das ist Theologie!” — “this is not Mathematics, this is Theology!” [8] Although there are several equivalent statements of the theorem, in this paper we will consider the version which states, in essence, that all rings of polynomials over countable fields are finitely generated. (More generally, the theorem holds for polynomial ideals over any Notherian ring. All the proofs in this paper can easily be adapted to this more general situation.)

In this paper, we will consider two different constructive proofs. Each is accomplished by applying Gödel’s Dialectica Interpretation to a classical proof of the theorem. Both yield algorithms that are instances of primitive recursive functionals of finite types, essentially a simple programming language in which one can only express total functions. The first, from a standard proof, yields a constructive version requiring higher-type primitive recursion. The second, obtained by applying the interpretation to a proof by Simpson [9], yields a more efficient algorithm in a sense which will be explained later.

An overview of this thesis is as follows: In sections 2 and 3 we present logical and algebraic preliminaries, respectively. In section 4 we present our first proof of Dickson’s Lemma. In section 5 we show how to derive the Hilbert Basis Theorem from Dickson’s Lemma. Finally, in section 6 we give our second, more elegant constructive proof of Dickson’s Lemma.

2 Logical Preliminaries

The main result of this paper is the application of a formal translation from a classical proof of the Hilbert Basis Theorem to a constructive version of the theorem, and the extraction of an algorithm from that proof. In order to perform that translation, we must understand the logical fundamentals of the systems we are working in, and the details of how the translation works.

2.1 Theories of arithmetic

The set of *primitive recursive functions* is the smallest set of functions from the natural numbers to the natural numbers (of various arities) containing 0, the successor function $S(x) = x + 1$, projections $p_i^n(x_1, \ldots, x_n) = x_i$, and closed under composition and *primitive recursion.* Here, primitive recursion means that given two functions g and h, one can define a new function f by

$$f(0, \vec{z}) = g(\vec{z}), \quad f(x+1, \vec{z}) = h(f(x, \vec{z}), x, \vec{z})$$

With the primitive recursive functions one can define functions that deal with ordered pairs and sequences, and then code integers, rational numbers, lists, graphs, trees, finite sets, and so on. The theory *primitive recursive arithmetic* (*PRA*) has symbols for all the primitive recursive functions. The axioms consist of the defining equations and a principle of induction,

$$\varphi(0) \land \forall x\ (\varphi(x) \to \varphi(x+1)) \to \forall x\ \varphi(x),$$

for quantifier-free formulas φ. Note that we never explicitly reason in these systems, so we do not present the axioms for these systems.

First-order arithmetic, or *Peano arithmetic* (*PA*), is a formal theory in the language with symbols $0, S, +, \times, <$, which are intended to denote the usual operations on the natural numbers. The axioms consist of

- defining equations for the basic symbols
- induction for all formulas in the language.

Heyting arithmetic, or *HA*, is the same theory based on intuitionistic logic. Since in these theories one can define the primitive recursive functions and prove their defining axioms, it is convenient to act as though *PA* and *HA* are extensions of *PRA*.

These theories suffice to handle most reasoning about finitary objects. For dealing with infinitary objects like real numbers, functions from the natural numbers to the natural numbers, infinite sequences of natural numbers, and so on, we need a more expressive framework. The *finite types* are defined as follows:

- $\mathbb{N}$ is a finite type
- If σ and τ are finite types, so are $\sigma \times \tau$ and $\sigma \to \tau$

For example, the real numbers can be represented as objects of type $\mathbb{N} \to \mathbb{N}$, and sequences of real numbers can be represented as objects of type $\mathbb{N} \to (\mathbb{N} \to \mathbb{N})$. The set of *primitive recursive functionals of finite type* allows the following principles of definition:

- λ abstraction, application, pairing, projection
- Higher-type primitive recursion:

$$F(0) = G, \quad F(n+1) = H(F(n), n)$$

The theory PRA^ω axiomatizes these, just as PRA axiomatizes the primitive recursive functions. Gödel called PRA^ω by the name T instead. HA^ω and PA^ω are extensions of PRA^ω which allow full induction, differing in the same way as PA and HA differ. Using higher type primitive recursion, we can define functions, such as Ackermann's function, which are not primitive recursive [1]. Restricted primitive recursion yields only functions of type $\mathbb{N} \to \mathbb{N}$ which are primitive recursive. Our first proof of Dickson's Lemma uses the more general form of primitive recursion, while the second uses the "sharper" restricted primitive recursion.

2.2 The Dialectica interpretation

In a 1958 paper in the Swiss journal *Dialectica* [4], Gödel presented a translation which translates formulas φ in HA^ω to formulas φ^D in PRA^ω of the form $\exists \underline{x}\ \forall \underline{y}\ \varphi_D(\underline{x}, \underline{y})$, where φ_D is a quantifier-free formula, and $\underline{x}$ and $\underline{y}$ are sequences of variables. (For convenience, we will drop the underlining.) Furthermore, the structure of the translation makes it easy to extract an algorithm which explicitly witnesses φ^D.

The translation translates φ to φ^D, which is a formula of the form $\exists x\ \forall y\ \varphi_D$, where φ_D is a quantifier-free formula in the language of PRA^ω. Here the free variables of φ_D are those of φ, together with the (possibly empty) tuples of variables x and y. If one or more of the free variables z of φ are exhibited, as $\varphi(z)$, then we write $\varphi_D(x, y, z)$ for φ_D. Similarly, the free variables of ψ_D are the free variables of ψ, together with the (possibly empty) sequences u and v. The associations $(\quad)^D$ and $(\quad)_D$ are defined inductively as follows, where

$$\varphi^D = \exists x\, \forall y\ \varphi_D \quad \text{and} \quad \psi^D = \exists u\, \forall v\, \psi_D.$$

1. For φ an atomic formula, x and y are both empty and $\varphi^D = \varphi_D = \varphi$

2. $$\begin{aligned} (\varphi \wedge \psi)^D &= \exists x, u\, \forall y, v\ (\varphi \wedge \psi)_D \\ &= \exists x, u\, \forall y, v\ (\varphi_D(x, y) \wedge \psi_D(u, v)) \end{aligned}$$

3. $$\begin{aligned} (\varphi \vee \psi)^D &= \exists z, x, u\, \forall y, v\ (\varphi \vee \psi)_D \\ &= \exists z, x, u\, \forall y, v\ ((z = 0 \wedge \varphi_D(x, y)) \vee (z = 1 \wedge \psi_D(u, v))) \end{aligned}$$

4. $(\forall z\ \varphi(z))^D = \exists X\, \forall z, y\ (\forall z\ \varphi(z))_D = \exists X\, \forall z, y\ \varphi_D(X(z), y, z)$

5. $(\exists z\ \varphi(z))^D = \exists z, x\, \forall y\ (\exists z\ \varphi(z))_D = \exists z, x\, \forall y\ \varphi_D(x, y, z)$.

6. $$\begin{aligned} (\varphi \to \psi)^D &= \exists U, Y\, \forall x, v\ (\varphi \to \psi)_D \\ &= \exists U, Y\, \forall x, v\ (\varphi_D(x, Y(x, v)) \to \psi_D(U(x), v)) \end{aligned}$$

Since we define $\neg\varphi$ to be $\varphi \to \bot$, from 6 we obtain

7. $(\neg\varphi)^D = \exists Y\, \forall x\ \ (\neg\varphi)_D = \exists Y\, \forall x\ \neg\varphi_D(x, Y(x))$.

The types of x and y in the final translation depend only on the structure of φ. We recursively define $\mathit{TEX}(\varphi)$ to be the type of x in the final translation, and $\mathit{TFA}(\varphi)$ to be the type of y in the final translation. Here, we let $\emptyset$ be the "empty type", with the property that for any type ρ, $\rho \times \emptyset = \rho$, $\rho \to \emptyset = \emptyset$, and $\emptyset \to \rho = \rho$.

1. $\mathit{TEX}(\varphi) = \mathit{TFA}(\varphi) = \emptyset$ if φ is an atomic formula.

2. $\mathit{TEX}(\varphi \wedge \psi) = \mathit{TEX}(\varphi) \times \mathit{TEX}(\psi)$
   $\mathit{TFA}(\varphi \wedge \psi) = \mathit{TFA}(\varphi) \times \mathit{TFA}(\psi)$

3. $\mathit{TEX}(\varphi \vee \psi) = \mathbb{N} \times \mathit{TEX}(\varphi) \times \mathit{TEX}(\psi)$
   $\mathit{TFA}(\varphi \vee \psi) = \mathit{TFA}(\varphi) \times \mathit{TFA}(\psi)$

4. $\mathit{TEX}(\forall z\ \varphi(z)) = \rho \to \mathit{TEX}(\varphi)$ where ρ is the type of z ρ
   $\mathit{TFA}(\forall z\ \varphi(z)) = \rho \times \mathit{TFA}(\varphi)$ where ρ is the type of z

5. $\mathit{TEX}(\exists z\ \varphi(z)) = \rho \times \mathit{TEX}(\varphi)$ where ρ is the type of z
   $\mathit{TFA}(\exists z\ \varphi(z)) = \mathit{TFA}(\varphi)$

6. $TEX(\varphi \to \psi) = (TEX(\varphi) \to TEX(\psi)) \times ((TEX(\varphi) \times TFA(\psi)) \to TFA(\varphi))$
   $TFA(\varphi \to \psi) = TEX(\varphi) \times TFA(\psi)$
7. $TEX(\neg\varphi) = TEX(\varphi) \to TFA(\varphi)$
   $TFA(\neg\varphi) = TEX(\varphi)$

Theorem 2.1 *If HA^ω proves a formula φ, then PRA^ω proves φ^D.*

Proof. See [6][7.4] and [1][2.4.1].

2.3 The double-negation interpretation

The Dialectica Interpretation takes HA^ω to PRA^ω. To extend our translation to PA^ω, we use the Gödel-Gentzen double-negation translation. The "double negation" translation provides the first (and simplest) translation of classical logic into intuitionistic logic. It was discovered independently by Gödel [3] and Gentzen [2].

This translation works as follows: A is translated to A^N, where is defined inductively as:

1. $\varphi^N = \neg\neg\varphi$, if φ is an atomic formula
2. $\bot^N{=}\bot$
3. $(\varphi \wedge \psi)^N = \varphi^N \wedge \psi^N$
4. $(\varphi \vee \psi)^N = \neg(\neg\varphi^N \wedge \neg\psi^N)$
5. $(\varphi \to \psi)^N = \varphi^N \to \psi^N$
6. $(\forall x\ \varphi)^N = \forall x\ (A^N)$
7. $(\exists x\ \varphi)^N = \neg\forall x\ (\neg A^N)$

Theorem 2.2 *If some set A of axioms prove φ, the set A^N of translations of A prove φ^N.*

Proof. See [10][2.3.4] □

Corollary 2.3 *If PA^ω proves φ, HA^ω proves φ^N.*

Corollary 2.4 *If PA^ω proves φ, PRA^ω proves the Dialectica Interpretation of φ^N.*

Proof. By Corollary 2.3 and Theorem 2.1. □

2.4 The no-counterexample interpretation

Suppose the formula we are translating is in prenex normal form, that is, of the form

$$A = \exists x_1 \, \forall y_1 \; \ldots \exists x_n \, \forall y_n \; A_0(x_1, y_1, \ldots, x_n, y_n)$$

where A_0 is quantifier free. Then the Double Negation-Dialectica interpretation ("ND interpretation") is equivalent to a simpler interpretation, the 'no-counterexample interpretation,' due to Kreisel [7]:

$$\exists \Phi_1, \ldots, \Phi_n \, \forall \underline{f} \; A_0(\Phi_1 \underline{f}, f_1(\Phi_1 \underline{f}), \ldots, \Phi_n \underline{f}, f_n(\Phi_1 \underline{f}, \ldots, \Phi_n \underline{f}))$$

where $\underline{f}$ is an n-tuple of functions, such that f_1 is a function of 1 variable, f_2 is a function of two variables, and so on.

Suppose that A is of the form above. Then, it is clear that $\neg A$ is classically equivalent to:

$$\forall x_1 \, \exists y_1 \; \ldots \forall x_n \, \exists y_n \; \neg A_0(x_1, y_1, \ldots, x_n, y_n)$$

So a counterexample to A is given by functions $f_1, \ldots, f_n$ such that

$$\forall \_x \neg A_0(x_1, f_1(x_1), \ldots, x_n, f_n(x_1, \ldots, x_n)) \tag{1}$$

Hence, functionals $\underline{\Phi}$ which satisfy the no-counterexample interpretation provide a counterexample to (1), and thus show that no counterexample can exist to the (intuitionistic version of the) original formula.

The key statements which occur in the proofs of the Hilbert Basis Theorem we consider are in prenex normal form, so the no-counterexample interpretation applies. This provides a guide in our search for functionals that satisfy the conditions of the translation.

2.5 Applying the Dialectica interpretation

The logical ideas just discussed provide a method of translating many nonconstructive theorems of mathematics into constructive versions: apply the ND-interpretation to the statement of the theorem, and then every line of the proof.

In practice, one usually does not need a fully formalized representation of the entire proof. It suffices to translate the main lemmas and intermediary statements, and then, informally, try to find the appropriate Dialectica witnesses. This is the methodology we will follow below.

All the associated algorithms will be described precisely but informally. They could be described more formally by terms of the language of PRA^{ω}, but since the goal is ultimately to obtain an ordinary *mathematical* proof, we do not need to do this here.

3 Preliminaries from Algebra

In this section we discuss various ideas from abstract algebra necessary to understand the Hilbert Basis Theorem, and our proof of it. We assume that the reader is familiar with the basic definitions of groups, rings, and fields.

3.1 Polynomials

Informally, we define the ring of polynomials over a ring R by analogy to the familiar polynomials over the natural numbers.

Definition 3.1 (Ring of Polynomials $R[x]$) *Let R be a ring and let $R[x]$ denote the set of all sequences of elements of R $(a_0, a_1, \ldots)$ such that $a_i = 0$ for all but a finite number of indices i. We call $R[x]$ the ring of polynomials in x over R. We interpret an element $p \in R[x]$ as $p_0 + p_1 x + \cdots + p_n x^n + \cdots$. We define addition and multiplication in the usual manner.*

Theorem 3.2 *If R is a commutative [resp. a ring with identity or a ring with no zero divisors], then so is $R[x]$. In particular, if R is a field, $R[x]$ is a ring with no zero divisors and an identity (i.e. an* integral domain*).*

Proof. See [5][III.5.1] □

The ring of polynomials in two variables is viewed simply as the ring of polynomials in the second variable over the ring of polynomials in the first variable over the base ring. That is, $R[x_1, x_2] = (R[x_1])[x_2]$, and analogously for any number of variables.

For ease of notation, we define $\underline{x}^{\underline{\alpha}}$, where $\underline{x}$ and $\underline{\alpha}$ are vectors of length n to mean $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$.

A polynomial of the form $\underline{x}^{\underline{\alpha}}$ is called a *monomial.* If, for two monomials $\underline{x}^{\underline{\alpha}}$ and $\underline{x}^{\underline{\beta}}$ we have that for each $i < n$, $\alpha_i \leq \beta_i$ then we say $\underline{x}^{\underline{\alpha}}$ *divides* $\underline{x}^{\underline{\beta}}$, written $\underline{x}^{\underline{\alpha}} | \underline{x}^{\underline{\beta}}$.

Definition 3.3 (Degree of a Monomial) *Let $\underline{x}^{\underline{\alpha}}$ be a monomial in n variables. The* degree of $\underline{x}^{\underline{\alpha}}$ *$(\deg(a\underline{x}^{\underline{\alpha}}))$ is defined to be $\sum_{i=1}^{n} \alpha_i$. The* degree in x_k of $\underline{x}^{\underline{\alpha}}$ *$(\deg_k(\underline{x}^{\underline{\alpha}}))$ is defined to be α_k.*

3.2 Ideals

Since the Hilbert Basis Theorem makes a statement about ideals and their generators, it is necessary to understand what an ideal is.

Definition 3.4 (Ideal) *Let R be a ring, and S a non-empty subset of R that is closed under the operations of multiplication and addition in R. If S is itself a ring under these operations then S is called a* subring *of R. A subring I of a ring R is an* ideal *if for all $r \in R$ and for all $x \in I$, $rx \in I$ and $xr \in I$.*

The most intuitive example of an ideal is the even numbers ($2\mathbb{N}$) in the natural numbers ($\mathbb{N}$). The even numbers are a subring of the natural numbers, and the product of any even number by any natural number yields another even number. Furthermore, for every ring R, two trivial ideals exist: R itself and $\{0_R\}$.

Definition 3.5 (Ideal Generated by a Set) *Let X be a subset of a ring R. Let $\{A_i : i \in I\}$ be the family of all ideals in R which contain X. Then $\bigcap_{i \in I} A_i$ is called the* ideal generated by X. *This ideal is denoted $\langle X \rangle$.*

If $X = \{x_1, x_2, \ldots, x_n\}$ then $\langle X \rangle$ is said to be *finitely generated*. In a commutative ring, $a \in \langle x_1, x_2, \ldots, x_n \rangle$ if and only if there are $a_1, a_2, \ldots, a_n \in R$ such that $a = a_1x_1 + a_2x_2 + \cdots + a_nx_n$. In a field, every ideal is finitely generated [5][III.2.21].

3.3 The Division Algorithm

Let k be a countable field. The division algorithm for polynomials in n variables lets us express a polynomial $f \in k[x_1, x_2, \ldots, x_n]$ in terms of given polynomials $f_1, f_2, \ldots, f_s \in k[x_1, x_2, \ldots, x_n]$ to get a result of the form

$$f = a_1f_1 + a_2f_2 + \cdots + a_sf_s + r$$

where $a_1, a_2, \ldots, a_s$ and $r \in k[x_1, x_2, \ldots, x_n]$.

First, we must make some definitions:

Definition 3.6 (Lexicographic Ordering on Monomials) *Let $\underline{x}^{\alpha}$ and $\underline{x}^{\beta}$ be two monomials in n variables. We say $\underline{x}^{\alpha} <_{lex} \underline{x}^{\beta}$ if for some $i \leq n$ $\alpha_i < \beta_i$ and for all j such that $0 \leq j < i$, $\alpha_j = \beta_j$.*

Definition 3.7 (Leading Term of a Polynomial) *Let $P \in k[x_1, \ldots, x_n]$ be a polynomial. Let M be the set of monomials in P (that is, $P = \sum_{m \in M} a_m m$, where $a_m \in k$ and $a_m \neq 0$). The* leading term of P ($LT(P)$) *is the monomial of highest degree which is lexicographically first.*

Theorem 3.8 *Let $F = (f_1, \ldots, f_s)$ be an ordered s-tuple of polynomials in $k[x_1, \ldots, x_n]$. Then every $f \in k[x_1, \ldots, x_n]$ can be written as*

$$f = a_1 f_1 + \cdots a_s f_s + r$$

where $a_i, r \in k[x_1, \ldots, x_n]$ and either $r = 0$ or r is a linear combination, with coefficients in k, of monomials in k, none of which is divisible by any of $LT(f_1), \ldots, LT(f_s)$. We call r a remainder *of f on division by F.*

Proof. from Cox, Little & O'Shea 2.3.3. We will present only the algorithm.
Input: $f_1, \ldots, f_s, f$
Output: $a_1, \ldots, a_s, r$

```
$a_1 := 0; \ldots; a_s := 0; r := 0$
$p := f$
WHILE $p \neq 0$ DO
    $i := 1$
    divisionoccurred := false
    WHILE $i \leq s$ AND divisionoccurred = false DO
        IF $LT(f_i)|LT(p)$ THEN
            $a_i := a_i + LT(p)/LT(f_i)$
            $p := p - (LT(p)/LT(f_i))f_i$
            divisionoccurred := true
        ELSE
            $i := i + 1$
    IF divisionoccurred = false THEN
        $r := r + LT(p)$
        $p := p - LT(p)$
```
□

We note that this algorithm can be expressed primitive recursively.

4 A Constructive Proof of Dickson's Lemma

Both proofs of the Hilbert Basis Theorem which we will consider follow the same pattern. First we prove the ND-interpretation of Dickson's Lemma, which states that all monomial ideals are finitely generated. Then, we prove that given the ND-interpretation of Dickson's Lemma we can prove the ND-Interpretation of Hilbert Basis Theorem.

The constructive version of Dickson's Lemma we will be proving is that for a given sequence of monomials I, for any function M of type $\mathbb{N} \to \mathbb{N}$

there is an $n \in \mathbb{N}$ and an $i < n$ such that $I(i)|I(M(n))$. Furthermore, n can be calculated as $N(M)$, where N is a functional of type $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$.

4.1 Some Notes on Notation

In this proof, we will use a small amount of non-standard notation in order to make some of the concepts of this proof more clear. In this section, we will explain those notations

Formally, a sequence of elements in a set S is a mapping of the type $\mathbb{N} \to S$. Traditionally, a sequence is written using subscript indices (so that a sequence might have terms $p_0, p_1, p_2, \ldots$). Instead, to make the functional nature of a sequence more clear, we will write sequences using directly functional notation (so the terms of such a sequence would be $P(0), P(1), P(2), \ldots$).

Additionally, normal sequences are not allowed to have gaps. It is useful for our purposes to allow a sequence to simply not take a value at an index. If a sequence P does not take a value at an index i, we write $P(i) = *$, where $*$ is an element not in S.

Given that we will be using sequences of monomials, which may have gaps in them, we must slightly redefine our definition of divisibility of monomials to allow divisibility to make sense in this context. If s and t are are either monomials or "$*$", we define $s|t$ to mean "if $t \neq *$ then $s \neq *$ and $s|t$" where the second "$|$" is the standard divisibility relation for monomials.

4.2 Classical Statement and Proof

The first step in our translation is to prove the classical version of Dickson's Lemma, which states that every monomial ideal is finitely generated. The classical proof we will be translating follows:

Theorem 4.1 *Given a sequence of monomials I, there is an n such that for every m, there is an $i < n$ such that $I(i)|I(m)$.*

Proof. By induction on $k \geq 1$, where k is the number of variables. When $k = 1$, I is a sequence of monomials in one variable. Let r be the least degree of x which appears in the sequence. Let j be the first index such that the degree of x in $I(j)$ is r. Set $n = j + 1$. Since for each i, $I(j)|I(i)$, n satisfies the conditions of the theorem.

Inductively, suppose that the theorem holds for k variables. We wish to show that the theorem holds for $k + 1$ variables.

So, suppose I is a sequence of monomials in $k+1$ variables. For clarity, we will call the $k+1$st variable y.

Define I^∞ to be the projection of I into k variables. That is, if $I(i) = \underline{x}^{\underline{\alpha}} y^r$, then $I^\infty(i) = \underline{x}^{\underline{\alpha}}$. Define

$$I^l(i) = \begin{cases} \underline{x}^{\underline{\alpha}} & \text{if } I(i) = \underline{x}^{\underline{\alpha}} y^l \\ * & \text{otherwise} \end{cases}$$

By the induction hypothesis, we may choose an n such that for all m there is an $i < n$ such that

$$I^\infty(i) | I^\infty(m) \tag{2}$$

Let some index m be given, and suppose

$$\deg_{k+1} I(m) \geq \max\{\deg_{k+1} I(j) : j < n\}$$

By (2) we can choose an $i < n$ such that $I^\infty(i)|I^\infty(m)$. Then, for each j in $1 \ldots k+1$, $\deg_j I(i) < deg_j I(m)$. Therefore, $I(i)|I(m)$.

Therefore, there is an n such that for all m there is an $i < n$ such that

$$deg_{k+1} I(m) \geq \max\{deg_{k+1} I(j) : j \leq n\} \rightarrow \quad I(i)|I(m)) \tag{3}$$

For any l, we can apply the induction hypothesis to I^l , and choose an $n^{(l)}$ such that for all m there is an $i < n^{(l)}$ such that $(I^l(i)|I^l(m))$. For any l, we define $n_{(l)} = \max\{n^{(j)} : j \leq l\}$. We claim that for any l $n_{(l)}$ satisfies the condition that for all m there is an $i < n$ such that

$$\deg_{k+1} I(m) < l \rightarrow I(i)|I(m) \tag{4}$$

Let some l and m be given such that $\deg_{k+1} I(m) < l$. Let $j = \deg_{k+1} I(m)$. By the definition of $n^{(j)}$, there is an $i < n^{(j)}$ (and thus less than $n_{(l)}$) such that $I(i)|I(m)$. Thus the condition is satisfied.

Pick n_0 which satisfies (3), and let $l_0 = max\{deg_{k+1} I(j) : j < n_0\}$. Also, pick n_1 which satisfies (4) for l_0, that is $n_1 = n^{(l_0)}$. Finally, set $n = max\{n_0, n_1\}$.

Let m be given. If $deg_{k+1} I(m) \geq l_0$, then n suffices by (3). Otherwise, n suffices by (4). Therefore, this n works for any m, as required. □

We note that this proof is inherently non-constructive. The basis of the induction argument requires us to find the minimal degree of x which appears in the sequence. This requires an unbounded search of the sequence, and there is no way to do this computably, since any algorithm can consider only finitely many terms. For the same reason, the entire theorem is non-constructive — any proof would suffer from a similar problem.

4.3 Applying the Dialectica Interpretation

In this section, we will formalize the important steps of the classical proof, and show what the Dialectica interpretation of these statements is. It would be cumbersome to formally translate every step of the classical proof. Therefore, we will only translate the most important steps of the proof. In the following section, we will show how to use these ideas to find a constructive proof of Dickson's Lemma.

The formal statement of the classical version of Dickson's Lemma is:

$$\exists n \, \forall m \, \exists i < n \; (I(i)|I(m)) \tag{5}$$

The ND-interpretation of this statement is:

$$\exists N \, \forall M \, \exists i < N(M) \; (I(i)|I(M(N(M)))) \tag{5a}$$

where N is a functional of type $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ and M is a function of type $\mathbb{N} \to \mathbb{N}$.

We show directly that (5) holds if I is a sequence in 1 variable, and suppose as our induction hypothesis that (5) holds if I is a sequence in k variables. As we will see, the constructive version will be almost the same.

The next step is to apply our induction hypothesis to I^∞, and find

$$\exists n \, \forall m \, \exists i < n \; (I^\infty(i)|I^\infty(m)) \tag{6}$$

which translates to

$$\exists N \, \forall M \, \exists i < N(M) \; (I^\infty(i)|I^\infty(M(N(M)))) \tag{6a}$$

From this we conclude

$$\exists n \, \forall m \, \exists i < n \; (deg_{k+1}(I(m) \geq \\ \max\{deg_{k+1} I(j) : j \leq n\} \to I(i)|I(m))) \tag{7}$$

which has the translation:

$$\exists N \, \forall M \, \exists i < N(M) \; (\deg_{k+1} I(M(N(M))) \geq \\ max_{j<N(M)} \deg_{k+1} I(j) \to I(i)|I(M(N(M)))) \tag{7a}$$

Similarly, for any given l we apply the induction hypothesis and find $\exists n \, \forall m \, \exists i < n \; I^l(i)|I^l(m)$. We apply induction, and find

$$\forall l \, \exists n \, \forall m \, \exists i < n \; deg_{k+1} < l \to I(i)|I(m) \tag{8}$$

which translates to

$$\forall l \, \exists N \, \forall M \, \exists i < N(M) \; deg_{k+1} < l \rightarrow I(i)|I(M(N(M))) \tag{8a}$$

As above, we choose n_0, l_0, n_1 and n. By either (7) or (8) (as appropriate), n satisfies (5) for the given sequence I. Similar reasoning applies in the constructive case.

4.4 Our Constructive Proof of Dickson's Lemma

Using the previous section as a guide, we present our proof of the constructive version of Dickson's Lemma.

Theorem 4.2 *Let I be a sequence of monomials in k variables. Then, for any function M of type $\mathbb{N} \rightarrow \mathbb{N}$ there is an $n \in \mathbb{N}$ and an $i < n$ such that*

$$I(i)|I(M(n)) \tag{9}$$

and n can be calculated by applying a functional N of type $(\mathbb{N} \rightarrow \mathbb{N}) \rightarrow (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N}$ to I and M, so that $n = N(I, M)$. For notational clarity, we suppress the parameter I.

Proof. By induction on k.

First, we show that we can find a functional N^1 which works for sequences in one variable. We define:

$$N'(a, M) = \begin{cases} a & \text{if} \quad \exists i \leq a \; I(i)|I(M(a)) \\ N'(M(a), M) & \text{otherwise} \end{cases}$$

and define $N^1(M) = N'(0, M) + 1$. Note that at each step, if the algorithm does not terminate the degree of $I(M(a))$ must be less than the degree of $I(a)$. Otherwise, $I(a)|I(M(a))$. Therefore, the algorithm will stop in at most $deg_1 I(0)$ steps.

Now, by the induction hypothesis, we choose N_0 which satisfies the conditions of the theorem for I^∞. Let any M be given. By our hypothesis, we can find an $i < N_0(M)$ such that $I^\infty(i)|I^\infty(M(N_0(M)))$. Suppose that

$$deg_{k+1} I(M(N_0(M))) \geq \max\{deg_{k+1} I(j) : j < N_0(M)\}$$

Then, $I(i)|I(M(N_0(M)))$.

For any given j, we apply the induction hypothesis, and find a functional $N^{(j)}$ which satisfies the theorem for the sequence I^j. For a given l we define

$N_1(l, M) = \max\{N^j(M) : j < l\}$. Suppose that $j = deg_{k+1} I(M(N_1(M))) < l$. Then, we can find an $i < N^j(M)$ (and hence less than $N_1(M)$) such that $I(i)|I(M(N_1(M)))$.

For a given M, we define the following:

- $n_0 = N_0(M)$
- $l_0 = \max\{\deg_{k+1} I(j) : j < n_0\}$
- $n_1 = N_1(l_0, M)$
- $\hat{n} = \max\{n_0, n_1\}$

We define $N(M) = \hat{n}$.

To show that this satisfies the conditions of the theorem for the given sequence I, we need to consider two cases:

Case 1: $deg_{k+1} I(M(\hat{n})) \geq l_0$. By the definition of $\hat{n}$, $\hat{n} \geq n_0$, so we can choose an $i < \hat{n}$ such that $I^\infty(i)|I^\infty(M(\hat{n}))$. But, since $deg_{k+1} I(i) \leq deg_{k+1} I(M(\hat{n}))$, $I(i)|I(M(\hat{n}))$, as required.

Case 2: $deg_{k+1} I(M(\hat{n})) < l_0$. Let $l = deg_{k+1} I(M(\hat{n}))$. We note that $\hat{n} \geq n_1$. By the definition of n_1, we may choose an $i < \hat{n}$ such that $I^l(i)|I^l(M(\hat{n}))$. But, therefore $I(i)|I(M(\hat{n}))$, as required. □

Note that the passage from a function that works for sequences in k variables to a function which works for sequences in $k+1$ variables is uniform. Therefore, using higher type primitive recursion, one can define a single function $\hat{N}(k, I, M)$ that works for every k.

5 From Dickson's Lemma to the Hilbert Basis Theorem

Finally, we come to the main result of this paper. We will use the same format to present this proof as we did in the proof of Dickson's Lemma.

5.1 Classical Statement and Proof

The following method of going from Dickson's Lemma to the Hilbert Basis Theorem comes from Simpson [9].

Theorem 5.1 *Let F be a sequence of polynomials in $K[x_1, \ldots, x_l]$, where K is a countable field. Then there is an n such that for all m*

$$F(m) \in \langle F(1), \ldots, F(n) \rangle \tag{10}$$

Proof. We define H to be another sequence of polynomials, such that each element of H is of the form $H(j) = g_1F(1) + g_2F(2) + \cdots + g_lF(l)$ for some $l \in \mathbb{N}$. (We can let $H(j) = g_1F(1) + g_2F(2) + \cdots + g_lF(l)$ if j is the Gödel coding of a tuple $(g_1, g_2, \ldots, g_l)$, and 0 if the j codes no such tuple).

We define $LT(i)$ to be the sequence of leading terms of $H(i)$ (that is, the lexicographically first term of highest total degree). LT is a sequence of monomials in l variables, so by Dickson's Lemma, there is an n such that for all m there is an $i < n$ such that $LT(i)|LT(m)$.

We claim that this n also suffices for H. Suppose not. Then, there is some j such that $H(j) \notin \langle H(1), \ldots, H(n)\rangle$. Using the division algorithm, we can write $H(j) = g_1H(1) + g_2H(2) + \cdots + g_nH(n) + r$, where for every $i < n$, $LT(i) \nmid LT(r)$. But, by the construction of H, $r = H(a)$ for some a. Therefore, by the definition of n, there is an $i < n$ such that $LT(i)|LT(r)$. This contradicts the definition of r. Therefore, n must satisfy the conditions of the theorem.

Let $\hat{n}$ be the maximum j such that $F(j)$ appears in the definition of $H(1), \ldots, H(n)$. Clearly, this fulfills the requirement of the theorem for F. □

5.2 Dialectica Interpretation of the Hilbert Basis Theorem

Again, we will present a sketch of a formal version of the proof, with key equations and their translations highlighted.

The formal version of the theorem we wish to prove is: For a given sequence $F \in k[x_1, \ldots, x_l]$, where k is a countable field,

$$\exists n\, \forall m\; F(m) \in \langle F(1) \ldots F(n)\rangle \tag{11}$$

which translates to:

$$\exists N\, \forall M\; F(M(N(M))) \in \langle F(1) \ldots F(N(M))\rangle \tag{11a}$$

Where N is a functional of type $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ and M is a function of type $\mathbb{N} \to \mathbb{N}$.

We define H and LT in the same fashion as before. By Dickson's Lemma,

$$\exists n\, \forall m\; \exists i < n\; LT(i)|LT(m) \tag{12}$$

which translates to

$$\exists N_{LT}\, \forall M\; \exists i < N_{LT}(M)\; (LT(i)|LT(N_{LT}(M(N_{LT})))) \tag{12a}$$

We choose such an n, and let an m be given. By the division algorithm, we can write $H(m) = g_1H(1) + \cdots + g_nH(n) + r$, such that for each $i \leq n$, $LT(i) \nmid LT(r)$. But, $r = H(m) - g_1H(1) - \cdots - g_nH(n)$, so for some j $r = H(j)$. But, by (12), $\exists i\ < n(H(i)|H(j))$. Therefore, it must be the case that $r = 0$, so $H(m) \in \langle H(1), \ldots, H(n)\rangle$.

We define $\hat{n}$ the same way as in the informal proof. It is clear that $\hat{n}$ satisfies (11). □

5.3 A Constructive Proof of the Hilbert Basis Theorem

In this section, we present our constructive proof of the Hilbert Basis theorem from the constructive version of Dickson's Lemma.

Theorem 5.2 *For a given sequence F in $k[x_1, \ldots, x_l]$, and for any function M of type $\mathbb{N} \to \mathbb{N}$ there is an n such that*

$$F(M(n)) \in \langle F(1) \ldots F(n)\rangle$$

Furthermore, n can be computed from F and M by a functional N of type $(\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$, i.e. $n = N(F, M)$.

Proof. We define H to be another sequence of polynomials, such that each element of H is of the form $H(j) = g_1F(1) + g_2F(2) + \cdots + g_lF(l)$ for some $l \in \mathbb{N}$. (We can let $H(j) = g_1F(1) + g_2F(2) + \cdots + g_lF(l)$ if j is the Gödel coding of a tuple $(g_1, g_2, \ldots, g_l)$, and 0 if the j codes for no such tuple).

We define $LT(i)$ to be the leading term of $H(i)$ (that is, the lexicographically first term of highest total degree). LT is a sequence of monomials in l variables. Therefore, by Dickson's Lemma we may choose a functional N_{LT} such that for any function M of type $\mathbb{N} \to \mathbb{N}$ there is an $i < N_{LT}(M)$ such that:

$$LT(i)|LT(N_{LT}(M(N_{LT})))$$

Given an M which finds counterexamples on H, we wish to find an $\hat{M}$ that finds counterexamples on LT. That is, we wish it to have to property that for any n, if $H(M(n)) \notin \langle H(1), \ldots, H(n)\rangle$ then there is no $i < n$ such that $LT(i)|LT(M(n))$.

We define it as follows: $\hat{M}(n) = 0$ if $M(n) \in \langle H(1), \ldots, H(n)\rangle$. Otherwise, using the division algorithm, we can write $H(M(n)) = g_1H(1) + \cdots + g_nH(n) + r$, where $r = H(j)$ for some j, and for each $i < n$, $LT(i) \nmid LT(j)$. Define $\hat{M}(n) = j$.

Define $N(M) = N_{LT}(\hat{M})$. Suppose that

$$H(M(N(M))) \notin \langle H(1), \ldots, H(N(M)) \rangle$$

Then, by the above, there is no $i < N_{LT}(\hat{M})$ such that $LT(i) | LT(\hat{M}(N_{LT}(\hat{M})))$. But, this contradicts our constructive version of Dickson's Lemma. Therefore,

$$H(M(N(M))) \in \langle H(1), \ldots, H(N(M)) \rangle$$

□

6 A More Refined Proof via Ordinal Numbers

In his 1988 paper [9], Simpson presents a proof that the Hilbert Basis Theorem is equivalent to the well-orderedness of ordinals less than ω^ω. The latter can be expressed by saying that there is no infinitely decreasing sequence of ordinals below ω^ω. He does this by defining an explicit mapping from "bad" sequences of polynomials to decreasing sequences of ordinals. We will present his proof, and show how we can use it to find a proof of our constructive version of the Hilbert Basis Theorem. The resulting proof uses only the restricted form of primitive recursion defined in Section 2.1, in contrast to the previous proof.

6.1 Definitions

In order to present this proof, we must define several concepts. First, we define an ordering on k-tuples in $\mathbb{N}^k$.

Definition 6.1 *We say $\underline{a} \preceq \underline{b}$ if for every i in $1 \ldots k$, $a_i \leq b_i$.*

It is clear that $\preceq$ is a partial order on $\mathbb{N}^k$. A major step in the proof will be to show that this is a *well* partial order - that there is no infinite descending sequence of tuples. Note that if $\underline{\alpha}, \underline{\beta} \in \mathbb{N}^k$ such that $\alpha \preceq \beta$, then $\underline{x}^{\underline{\alpha}} | \underline{x}^{\underline{\beta}}$.

Standard definitions of addition and multiplication on ordinals are not commutative. In order to define arithmetic which better reflects our intuition of arithmetic on tuples of natural numbers, we define the *natural sum* and *natural product* as follows:

Definition 6.2 *We define the* natural sum *of ordinals as follows:*

$$(\omega^{\alpha_1} + \cdots + \omega^{\alpha_m}) + (\omega^{\beta_1} + \cdots + \omega^{\beta_n}) = \omega^{\gamma_1} + \cdots + \omega^{\gamma_{m+n}}$$

where $\langle \gamma_1, \ldots, \gamma_{m+n} \rangle$ *is a permutation of* $\langle \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n \rangle$ *such that* $\gamma_1 \geq \ldots \geq \gamma_{m+n}$. *The* natural product *is defined by*

$$(\omega^{\alpha_1} + \cdots + \omega^{\alpha_m}) \times (\omega^{\beta_1} + \cdots + \omega^{\beta_n}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \omega^{\alpha_i + \beta_j}$$

Essentially, for the natural product and natural sum we treat an ordinal as merely a polynomial in ω. Note that when referring to ordinals, we will always mean natural sum and natural product when we use + and ×.

We define $\mathrm{Bad}(\mathbb{N}^m)$ to be the set of all (finite or infinite) sequences S of elements in $(N)^m$ such that for all i and j, if $i < j$ then $S(i) \not\preceq S(j)$.

For sequences s and t, we define $s\hat{\ }t$ to be t appended to s. For any finite sequence s of elements in $\mathbb{N}^m$, we define $(\mathbb{N}^m)_s$ to be $\{a \in \mathbb{N}^m : s\hat{\ }\langle a \rangle \in \mathrm{Bad}(\mathbb{N}^m)\}$.

Additionally, we define two additional pieces of notation. For $u \leq v \leq \omega$ we write $[u, v) = \{a : u \leq a < v\}$. Finally, given an m-fold Cartesian product $\prod_{i=1}^{m} [u_i, v_i)$ with $u_i \leq v_i \leq \omega$ for each i, we define

$$\left| \prod_{i=1}^{m} [u_i, v_i) \right| = \prod_{i=1}^{m} (v_i - u_i)$$

where on the right-hand side $\prod$ denotes the natural product as defined above. Note that by the definition of ordinal subtraction, $\omega - k = \omega$ for any natural number k.

6.2 Mapping bad sequences to ordinals

In this section, we wish to show that $\preceq$ is a well partial order. This means that for any sequence of tuples A, there are an i and j such that $i < j$ and $A(i) \preceq A(j)$ (so there cannot be a sequence such that each element is either less than or incomparable with each previous element). To do so, we create what Simpson calls a *reification* from "bad" finite sequences of tuples to (that is, finite sequences which do not satisfy the above property) and ordinals. A reification is defined as a function f with the property that if $\langle a_1, \ldots, a_n \rangle$ and $\langle a_1, \ldots, a_n, a_{n+1} \rangle$ are both bad sequences, then $f(\langle a_1, \ldots, a_n \rangle) > f(\langle a_1, \ldots, a_n, a_{n+1} \rangle)$. That is, as we add elements to a bad sequence, the associated ordinal decreases.

Lemma 6.3 *Suppose that* $\langle a_1, \ldots, a_m\rangle \in \prod_{i=1}^m [u_i, v_1)$ *where* $u_i \leq v_i \leq \omega$ *for each* i*. Let* $\sum$ *denote the natural sum,* $\sigma = \langle \sigma_i : 1 \leq i \leq m\rangle$ *range over all* m*-tuples of zeros and ones that do not consist entirely of ones, and*

$$[u_i(\sigma), v_i(\sigma)) = \begin{cases} [u_i, a_i) & \textit{if} \quad \sigma_i = 0 \\ [a_i, v_i) & \textit{if} \quad \sigma_i = 1 \end{cases}$$

Then,

$$\sum_{\sigma} \left| \prod_{i=1}^m [u_i(\sigma), v_i(\sigma)) \right| < \left| \prod_{i=1}^m [u_i, v_i) \right| \tag{13}$$

Proof. Let k be the number of i's such that $v_i = \omega$. Suppose that $k = 0$. In this case, the lemma follows by observing that the disjoint union

$$\bigcup_{\sigma} \prod_{i=1}^m [u_i(\sigma), v_i(\sigma))$$

is a proper subset of the set $\prod_{i=1}^m [u_i, v_i)$, which has finite "measure".

Suppose that $k > 0$. The right hand side of (13) is of the form $\omega^k \times n$ where $n < \omega$. Let us say that σ is *wild* if $\sigma_i = 0$ for some i such that $v_i = \omega$, otherwise σ is *tame.* If σ is wild, the contribution of σ to the left-hand side of (13) is of the form $\omega^{k'} \times n'$, where $k' < k$ and $n' < \omega$. Hence, since any sum of ordinals less than ω^k will be strictly less then ω^k, the total contribution of all wild σ's is $< \omega^k$. On the other hand, the total contribution of all the tame σ's is of the form $\omega^k \times n''$ where $n'' < n$. (The inequality $n'' < n$ follows from the special case $k = 0$, which was proved above.) Thus the total left hand side is less than $\omega^k + (\omega^k \times n'') \leq \omega^k \times n$. □

Lemma 6.4 *For each* $m \in \mathbb{N}$ *there is a reification of* $\mathbb{N}^m$ *by* ω^m*.*

Proof. Fix $m \in \mathbb{N}$. We will define a reification $f : \mathrm{Bad}(\mathbb{N}^M) \to \omega^m + 1$. For $s \in \mathrm{Bad}(\mathbb{N}^m)$ we will define $f(s) \leq \omega^m$ by primitive recursion on the length of s. The value of $f(s)$ will be defined in terms of a decomposition of $(\mathbb{N}^m)_s$ into a disjoint union,

$$(\mathbb{N}^m)_s \subseteq \bigcup_{j \in J} \prod_{i=1}^m [u_{ij}, v_{ij})$$

where J is a finite index set, and $u_{ij} \leq v_{ij} \leq \omega$ for all $j \in J$ and $i = 1, \ldots, m$. We then define

$$f(s) = \sum_{j \in J} \left| \prod_{i=1}^m [u_{ij}, v_{ij}) \right| \tag{14}$$

We begin with the trivial decomposition $(\mathbb{N}^m)_{\langle\quad\rangle} = \mathbb{N}^m = \prod_{i=1}^m [0,\omega)$, and accordingly we define

$$f(\langle\quad\rangle) = \left|\prod_{i=1}^m [0,\omega)\right| = \omega^m$$

Now, fix $s \in \text{Bad}(\mathbb{N}^m)$, and assume inductively that we have already defined $f(s)$ according to a decomposition (14) of $(\mathbb{N}^m)_s$. Given $s' = s\hat{}\langle a\rangle \in \text{Bad}(\mathbb{N}^m)$, we want to define $f(s')$. Since by definition $a \in (\mathbb{N}^m)_s$, there is a unique $j' \in J$ such that $a = (a_1, \ldots, a_m) \in \prod_{i=1}^m [u_{ij'}, v_{ij'})$. As our decomposition of $(\mathbb{N}^m)_{s'}$ we take (14) with $\prod_{i=1}^m [u_{ij'}, v_{ij'})$ replaced by

$$\bigcup_\sigma \prod_{i=1}^m [u_{ij'}(\sigma), v_{ij'}(\sigma))$$

as in Lemma 6.3. It is easy to check that this provides a decomposition of $(\mathbb{N}^m)_{s'}$. The fact that $f(s') < f(s)$ follows from Lemma 6.3. □

Theorem 6.5 *$\preceq$ is a well partial order on $\mathbb{N}^m$ for any $m \in \mathbb{N}$.*

Proof. By Lemma 6.4, any bad sequence in $\mathbb{N}^m$ can be mapped to a decreasing sequence of ordinals less than $\omega^m + 1$. Therefore, an infinite bad sequence in $\mathbb{N}^m$ could be mapped to an infinite descending sequence in ordinals less $\omega^m + 1$. But, we know the ordinals are well-ordered, so this is a contradiction. Therefore, there can be no infinite bad sequence in $\mathbb{N}^m$, and therefore $\preceq$ is a well-partial order. □

6.3 Dickson's Lemma, via Ordinals

With the use of Lemma 6.5, we can prove a version of Dickson's Lemma.

Theorem 6.6 *For any countable partial ordering $\preceq$ on a set S, the following assertions are equivalent:*

1. *$\preceq$ is well partially ordered*

2. *For all infinite sequences A of elements in S, there exists an n such that for all m there exists an $i \leq n$ such that $A(i) \preceq A(m)$.*

Proof. The implication from 2 to 1 is trivial. To prove the implication from 1 to 2, let S be given such that $\preceq$ is a well partial order on S. Suppose the conclusion of 2 fails. Then, for every n there is an m such that for every $i < n$, $A(i) \not\preceq A(m)$. We define a subsequence $A(i_0), A(i_1), \ldots$ as follows: Set $i_0 = 0$, and i_{n+1} to be an m such that for every $i < i_n$, $A(i) \not\preceq A(m)$. But, then $A(i_n)$ must have the property that for each n and each $j < n$, $A(i_j) \not\preceq A(i_n)$. This contradicts the assumption that $\preceq$ is a well partial order. □

If we apply this theorem to $\preceq$ on $\mathbb{N}^m$, we get Dickson's Lemma, from which we can find the Hilbert Basis Theorem, as in the previous section.

6.4 A Constructive Version of the Ordinal Principles

Once again, we apply the Dialectica Interpretation (specifically the no-counterexample interpretation) to find specific algorithms witnessing the interpreted version of Dickson's Lemma.

The first step of the proof is to demonstrate that the ordinals are well ordered, and specifically a functional which witnesses the fact.

The standard statement of the well orderedness of the ordinals is that, given a sequence of ordinals A,

$$\forall A\ \exists i, j\ (i < j \wedge A(i) \leq A(j))$$

Note that the Dialectica interpretation of this statement is simply

$$\exists I, J\ \forall A\ (I(A) < J(A) \wedge A(I(A)) \leq A(J(A))$$

which has no "counterexample function." Therefore, we should be able to find a primitive recursive functional which witnesses that statement outright.

The theorem we wish to prove is:

Theorem 6.7 *For any sequence of ordinals A, there is an i and j such that $i < j$ and $A(i) \leq A(j)$. Furthermore, we can explicitly calculate i and j for each A.*

Proof. We wish to define a functional $O(A, m, o, d)$ which maps from sequences of ordinals to $\mathbb{N}$. If A is a sequence of ordinals less than ω^m, then there is an i and $j \leq O(A, m, 0, 0)$ such that $i < j$ and $A(i) \leq A(j)$. That is, O provides a bound on how long the sequence can decrease. We can then find an exact i and j by a finite search. For compactness of notation, we will omit the first parameter in our definition.

We define the arguments to the functional as follows:

- m is an upper bound on ordinals in the sequence — that is, all ordinals in the sequence have value less than ω^m (after applying the other arguments).
- o is an offset into the sequence. That is, instead of considering the sequence starting at 0, we consider the sequence starting at o.
- d, an ordinal, is a displacement of the sequence — essentially, we subtract d from the value of each term before performing our calculations.

First, we define

$$O(1,o,d) = \begin{cases} A(o)-d & \text{if} \quad A(o) \geq d \\ 0 & \text{otherwise} \end{cases}$$

A sequence of ordinals less than ω is a sequence of natural numbers. In the worst case, the sequence will decrease by 1 at each step, and will reach 0 in $A(o)-d$ steps (given our interpretations of o and d).

Suppose inductively that we have defined $O(m-1,o,d)$, and we wish to define $O(m,o,d)$.

$$O(m,o,d) = \begin{cases} O(m-1,o,d) & \text{if} \quad A(o)-d < \omega^{m-1} \\ O(m-1,o,d+\omega^m \times n) & \text{if} \quad O(m,o+O(m-1,o, \\ & \quad d+\omega^m \times n),d) \geq A(o) \\ O(m-1,o,d+\omega^m \times n)+ & \\ \quad O(m,o+O(m-1,o, & \\ \quad d+\omega^m \times n),d) & \text{otherwise} \end{cases}$$

This functional works in a straightforward fashion. It divides the sequence into an initial part (with leading term $\omega^m \times n$), and the rest of the sequence, with leading term at most $\omega^m \times (n-1)$. Since at each step either the coefficient of ω^m or the value of m is decreasing, this functional will eventually terminate. Clearly, we can find an i and j less than or equal to $O(m,0,0)$ such that $i < j$ and $A(i) \leq A(j)$ by finite search. □

The next step in the classical proof is to show that for sequences of m-tuples,

$$\forall N \; \exists i,j \; (i < j \wedge N(i) \preceq N(j))$$

which translates to

$$\exists I,J \; \forall N \; (I(N) < J(N) \wedge N(I(N)) \preceq N(J(N)))$$

Again, there is no counterexample term, so we should be able to find a functional witnessing the translation outright. Therefore, we need to prove:

Theorem 6.8 *For any sequence N of elements of $\mathbb{N}^m$, there is an i and j such that $i < j$ and $N(i) \preceq N(j)$. Furthermore, we can explicitly calculate such an i and j for each N.*

Proof. Our strategy will be the same as in the previous section. We will define a functional $T(N, m)$ which finds an upper bound on an i and j, which can then be found by finite search.

Given a sequence N of m-tuples in $\mathbb{N}^m$, we define a sequence of ordinals A such that $O(A, m, 0, 0)$ gives us a bound for N. For a given n, we define

$$A(n) = \begin{cases} f(\langle N(0), \ldots, N(n)\rangle) & \text{if} \quad \langle N(0), \ldots, N(n)\rangle \in \mathrm{Bad}(\mathbb{N}^m) \\ 0 & \text{otherwise} \end{cases}$$

where f is the reification function defined in Lemma 6.4. Note that f is primitive recursive.

Finally, we define $T(N, m) = O(A, m+1, 0, 0)$. By the previous theorem, we know that any sequence longer than $T(N, m)$ will not be "bad." Therefore, we can find an i and j less than or equal to $T(N, m)$ satisfying the theorem by bounded search. □

6.5 Another Constructive Proof of Theorem 5.2

The next step of our constructive proof is the translation of Theorem 6.6. The formalized version of the classical proof goes from the statement that (for a given A)

$$\exists i, j\ (i < j \wedge A(i) \preceq A(j))$$

to the statement

$$\exists n\ \forall m\ \exists i < n\ (A(i) \preceq A(m))$$

The Dialectica Interpretation of these statements is (again for a given A):

$$\exists i, j\ (i < j \wedge A(i) \preceq A(j)) \tag{15}$$

and

$$\exists N\ \forall M\ \exists i < N(M)\ (A(i) \preceq A(M(N(M)))) \tag{16}$$

So, we wish to prove:

Theorem 6.9 *Let A be a sequence in $\mathbb{N}^m$. For any function M of type $\mathbb{N} \to \mathbb{N}$ there is an n such that there is an $i < n$ such that $A(i) \preceq A(M(n))$. Furthermore, we can find a primitive recursive functional N of type* $(\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ *which calculates such an n for each A and M.*

Proof. Our goal is to find a functional N, depending on A, of type $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ such that given a function M of type $\mathbb{N} \to \mathbb{N}$ trying to find counterexamples to (16), N foils it. Let such an M be given.

First, we define $N(M) = 0$ if $M(0) = 0$ (that is, if M cannot find an effective counterexample, we have no need to continue). Otherwise, we define a sequence Q in $\mathbb{N}$ as follows:

$$Q(0) = 0$$

$$Q(n) = \begin{cases} M(Q(n-1)) & \text{if} \quad \neg\exists i < M(Q(n-1))\ (A(i) \preceq A(M(Q(n-1))) \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, we define a sequence S in $\mathbb{N}^m$ as $S(n) = A(Q(n))$.

By the previous section, we can constructively find an i and a j such that $i < j$ and $S(i) \preceq S(j)$. But, this means that there is an $k < Q(j)$ such that $A(k) \preceq A(M(Q(j)))$. Set $N(M) = Q(j)$. Then, N satisfies the conditions of the theorem. □

7 Conclusions

In this paper we have presented two constructive proofs of the Hilbert Basis Theorem, corresponding to two different classical proofs of Dickson's Lemma. The standard proof of Dickson's Lemma leads us to a constructive version requiring "higher type" primitive recursion, while a more complicated (and direct) proof of Dickson's Lemma leads to a constructive version requiring only standard primitive recursion.

Of course, this is only a simple example of the proof mining possibilities which arise from the Dialectica Interpretation (especially given that we only use the no-counterexample interpretation, and not the full power of the Dialectica Interpretation). Nevertheless, it is interesting to see that this interpretation allows us to find a straightforward and expressive constructive form of one of the most famous examples of a nonconstructive proof.

References

[1] Avigad, J. and Feferman, S. "Gödel's Functional ('Dialectica') Interpretation," *Handbook of Proof Theory*, S. R. Buss, Editor. 338–400. Elsevier, 1998.

[2] Gentzen, G. "Über das Verhältnia zwische intuitionistischer und klassicher Logik," *Archiv für Mathematische Logik und Grundlagenforschung*, 16:199-132, 1974.

[3] Gödel, K. "Zur intuitionistischen Arithmetik und Zahlentheorie," *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38, 1933.

[4] Gödel, K. "Über eine bisher noch nicht benüzte Erweiterung des finiten Standpunktes," *Dialectica*, 12:280–287, 1958.

[5] Hungerford, T. *Algebra.* Springer Press, 1974.

[6] Kohlenbach, U. *Proof Interpretations and the Computational Content of Proofs.* Unpublished Draft, 2002.

[7] Kreisel, G. "On the interpretations of non-finitist proofs, part I," *Journal of Symbolic Logic*, 16:241–267, 1951.

[8] Noether, M. "Paul Gordan," *Mathematische Annalen*, 75:1–41, 1914.

[9] Simpson, S. "Ordinal numbers and the Hilbert basis theorem," *Journal of Symbolic Logic*, 53:961–974, 1988.

[10] Troelstra, A.S. and Schwichtenberg, H. *Basic Proof Theory.* Cambridge University Press, 1996.