More infinity for a better finitism

Sam Sanders

MORE INFINITY FOR A BETTER FINITISM

SAM SANDERS

Abstract. Elementary Recursive Nonstandard Analysis, in short ERNA, is a constructive system of nonstandard analysis with a PRA consistency proof, proposed around 1995 by Patrick Suppes and Richard Sommer. It is based on an earlier system developed by Rolando Chuaqui and Patrick Suppes. Here, we discuss the inherent problems and limitations of the classical nonstandard framework and propose a much-needed refinement of ERNA, called $\text{ERNA}^{\mathbb{A}}$, in the spirit of Karel Hrbacek's stratified set theory. We study the metamathematics of $\text{ERNA}^{\mathbb{A}}$ and its extensions. In particular, we consider several transfer principles, both classical and 'stratified', which turn out to be related. Finally, we show that the resulting theory allows for a truly general, elegant and elementary treatment of basic analysis.

1. Introduction

By now, it is well-known that large parts of 'ordinary' mathematics can be developed in systems much weaker than ZFC ([20], [21]). However, most theories under consideration are at least as strong as WKL_0, which is conservative over $I\Sigma_1$. It is usually mentioned (see e.g. [1], [2] and [20]) that it should be possible to develop a large part of mathematics in much weaker systems, in particular in $I\Delta_0 + \exp$ and related systems. Most notably, there is Friedman's Grand Conjecture (see [2] and [6]):

> *Every theorem published in the Annals of Mathematics whose statement involves only finitary mathematical objects* (*i.e. what logicians call an arithmetical statement*) *can be proved in* EFA.

In 1929, Jacques Herbrand already made a similar claim, but without specifying the underlying logical system (see [9, p152]).

In this way, there have been attempts at developing analysis in nonstandard versions of $I\Delta_0 + \exp$ (see [1], [4], [12], [23], [24] and [25]). In particular, the theory ERNA and its predecessor NQA^+ (see [12] and [17]) are such systems. According to Chuaqui, Sommer and Suppes, the latter theories 'provide a foundation that is close to mathematical practice characteristic of theoretical physics'. In order to achieve this goal, the systems satisfy the following three conditions, listed in [4]:

(i) The formulation of the axioms is essentially a free-variable one with no use of quantifiers.
(ii) We use infinitesimals in an elementary way drawn from nonstandard analysis, but the account here is axiomatically self-contained and deliberately elementary in spirit.
(iii) Theorems are left only in approximate form; that is, strict equalities and inequalities are replaced by approximate equalities and inequalities. In

University of Ghent, Department of Pure Mathematics and Computer Algebra, Galglaan 2, B-9000 Gent (Belgium)
Telephone/Fax: +3292648573 / +3292644993
E-mail address: sasander@cage.ugent.be.

2000 *Mathematics Subject Classification.* 03H15, 26E35.
Key words and phrases. Relative, stratified, nonstandard, quantifier-free, analysis, ERNA.

particular, we use neither the notion of standard function nor the standard part function.

It is also mentioned in [4], that another standard practice of physics, namely the use of physically intuitive but mathematically unsound reasoning, is not reflected in the system.

By limiting the strength of the systems according to (i)-(iii), the consistency of ERNA can be proved in PRA, using Herbrand's theorem in the following form (see [4] and [23]).

1. **Theorem** (Herbrand)**.** *A quantier-free theory T is consistent if and only if every finite set of instantiated axioms of T is consistent.*

In this respect, the item (i) is not merely a technicality to suit Herbrand's theorem: the quantifier-free axioms reflect the absence of existential quantifiers in physics. As all ε-δ definitions of basic analysis are equivalent to universal nonstandard formulas, it indeed seems plausible that one can develop calculus inside ERNA and NQA$^+$ in a quantifier-free way, particularly, without the use of ε-δ-statements. However, we discuss two compelling arguments why such a development is impossible.

First, as exemplified by item (iii), NQA$^+$ has no 'standard part' function 'st', which maps every finite number x to the unique standard number y such that $x \approx y$. Thus, nonstandard objects like integrals and derivatives are only defined 'up to infinitesimals'. This leads to problems when trying to prove e.g. the fundamental theorems of calculus, which express that differentiation and integration cancel each other out. Indeed, in [4, Theorem 8.3], Chuaqui and Suppes prove the first fundamental theorem of calculus, using the previously proved corollary 7.4. The latter states that differentiation and integration cancel each other out *on the condition* that the mesh du of the hyperfinite Riemann sum of the integral and the infinitesimal y used in the derivative satisfy $du/y \approx 0$. Thus, for every y, there is a du such that for all meshes $dv \leq du$ the corresponding integral and derivative cancel each other out. The definition of the Riemann integral ([4, Axiom 18]) absorbs this problem, but the former is quite complicated as a consequence. Also, it does not change the fact that ε-δ-statements occur, be it swept under the proverbial nonstandard carpet. Similary, ERNA only proves a version of Peano's existence theorem with a condition similar to $du/y \approx 0$, contrary to Sommer and Suppes' claim in [24] (see [18]). Thus, ERNA and NQA$^+$ cannot develop basic analysis without invoking ε-δ statements.

Second, we consider to what extent classical nonstandard analysis is actually free of ε-δ-statements. For all functions in the standard language, the well-known classical ε-δ definitions of continuity or Riemann integrability, which are Π_3, can be replaced by universal nonstandard formulas (see e.g. [22, p70]). Given that even most mathematicians find it difficult to work with a formula with more than two quantifier alternations, this is a great virtue. Indeed, using the nonstandard method greatly reduces the sometimes tedious 'epsilon management' when working with several ε-δ statements, see [27]. Yet, nonstandard analysis is not completely free of ε-δ statements. For instance, consider the function $\delta(x) = \frac{1}{\pi}\frac{\varepsilon}{\varepsilon^2+x^2}$, with $\varepsilon \approx 0$ and let $f(x)$ be a standard C^∞ function with compact support. Calculating the (nonstandard) Riemann integral of $\delta(x) \times f(x)$ yields $f(0)$. Hence $\delta(x)$ is a nonstandard version of the Dirac Delta. However, not every Riemann sum with infinitesimal mesh is infinitely close to the Riemann integral: the mesh has to be *small enough* (compared to ε). Moreover, $\delta(x) \approx \delta(y)$ is not true for all $x \approx y$, only for x and y *close enough*. In general, most functions which are not in the standard language do *not* have an elegant universal definition of continuity or integrability and we have to resort to ε-δ statements. Thus, nonstandard analysis only partially removes the ε-δ formalism.

These two arguments show that the 'regular' nonstandard framework does not allow us to develop basic analysis in a quantifier-free way in weak theories of arithmetic. Moreover, for treating more advanced analysis, like the Dirac Delta, prevalent in physics, we would have to resort to ε-δ-statements anyway. Inspired by Hrbacek's 'stratified analysis' (see [10] and [11]), we introduce a weak theory of arithmetic, called ERNA$^{\mathbb{A}}$, which will allow us to develop analysis in a quantifier-free way. To this end, the theory ERNA$^{\mathbb{A}}$ has a multitude of sets of infinite numbers instead of the usual dichotomy of one set of finite numbers O, complemented with one set of infinite numbers Ω. Indeed, in ERNA$^{\mathbb{A}}$ there is a linear ordering $(\mathbb{A}, \preceq)$ with least number $\mathbf{0}$, such that for all nonzero $\alpha, \beta \in \mathbb{A}$, the infinite number ω_α is finite *compared to* ω_β for $\beta \succ \alpha$. Hence there are many 'degrees' or 'levels' of infinity and the least number $\mathbf{0}$ in the ordering $(\mathbb{A}, \preceq)$ corresponds to the standard level. It should be noted that the first nonstandard set theory involving different levels of infinity was introduced by Péraire in [16]. Another approach was developed by Gordon in [7].

In the second section, we describe ERNA$^{\mathbb{A}}$ and its fundamental features and in the third section, we prove the consistency of ERNA$^{\mathbb{A}}$ inside PRA. Though important in its own right, in particular for 'strict' finitism (see [26]), we not only wish to do quantifier-free analysis in ERNA$^{\mathbb{A}}$, but also study its metamathematics. Thus, in the fourth section, we introduce the 'Stratified Transfer Principle', which expresses that a true formula should hold at all levels (see [10]). As ERNA$^{\mathbb{A}}$ is a weak theory of arithmetic, we limit ourselves to transfer for *universal* formulas. This will turn out to be sufficient for developing analysis. Stratified Transfer equally applies to external formulas and is thus very different from transfer principles in regular nonstandard arithmetic. In the fifth section, we introduce various transfer principles for ERNA$^{\mathbb{A}}$, which are based on transfer principles for ERNA (see [12] and [13]). It turns out that these 'regular' transfer principles imply the Stratified Transfer Principle, which is remarkable, given the fundamental difference in scope between both. In the sixth section, we prove several important theorems of analysis in ERNA$^{\mathbb{A}}$ and extensions. In the last section, we argue that Stratified Transfer yields a good formal framework for theoretical physics.

2. ERNA$^{\mathbb{A}}$, THE SYSTEM

In this section, we describe ERNA$^{\mathbb{A}}$ and some of its fundamental features.

2.1. **The language.** Let $(\mathbb{A}, \preceq)$ be a fixed linear order with least element $\mathbf{0}$, e.g. $(\mathbb{N}, \leq)$ or $(\mathbb{Q}^+, \leq)$. For brevity, we write '$\alpha \prec \beta$' instead of '$\alpha \preceq \beta \wedge \alpha \neq \beta$'.

2. **Definition.** The language L of ERNA$^{\mathbb{A}}$ includes ERNA's, minus the symbols 'ω', 'ε' and '$\approx$'. Additionally, it contains, for every nonzero $\alpha \in \mathbb{A}$, two constants 'ω_α' and 'ε_α' and, for every $\alpha \in \mathbb{A}$, a binary predicate '$\approx_\alpha$'.

The set $\mathbb{A}$ and the predicate $\preceq$ are not part of the language of ERNA$^{\mathbb{A}}$. However, we shall sometimes informally refer to them in theorems and definitions. Note that there are no constants $\omega_{\mathbf{0}}$ and $\varepsilon_{\mathbf{0}}$ in L.

3. **Definition.** For all $\alpha \in \mathbb{A}$, the formula '$x \approx_\alpha 0$' is read 'x is α-infinitesimal', 'x is α-infinite' stands for '$x \neq 0 \wedge 1/x \approx_\alpha 0$'; '$x$ is α-finite' stands for 'x is not α-infinite'; 'x is α-natural' stands for 'x is hypernatural and α-finite'.

4. **Definition.** If L is the language of ERNA$^{\mathbb{A}}$, then $L^{\alpha\text{-}st}$, the α-standard language of ERNA$^{\mathbb{A}}$, is L without $\approx_\beta$ for all $\beta \in \mathbb{A}$ and without ω_β and ε_β for $\beta \succ \alpha$.

For $\alpha = \mathbf{0}$, we usually drop the addition '$\mathbf{0}$'. For instance, we write 'natural' instead of '$\mathbf{0}$-natural' and '$\approx$' instead of '$\approx_{\mathbf{0}}$'. Note that in this way, $L^{\mathbf{0}\text{-}st}$ is L^{st}, the *standard* language of ERNA$^{\mathbb{A}}$.

5. **Definition.** A term or formula is called *internal* if it does not involve $\approx_\alpha$ for any $\alpha \in \mathbb{A}$; if it does, it is called *external*.

2.2. **The axioms.** The axioms of ERNA$^{\mathbb{A}}$ include ERNA's, minus axiom 7.(4) (Hypernaturals), axiom set 11 (Infinitesimals) and axiom set 37 (External minimum). Additionally, ERNA$^{\mathbb{A}}$ contains the following axiom set.

6. **Axiom set** (Infinitesimals)**.**
   (1) *If x and y are α-infinitesimal, so are $x + y$ and $x \times y$.*
   (2) *If x is α-infinitesimal and y is α-finite, xy is α-infinitesimal.*
   (3) *An α-infinitesimal is α-finite.*
   (4) *If x is α-infinitesimal and $|y| \leq x$, then y is α-infinitesimal.*
   (5) *If x and y are α-finite, then so is $x + y$.*
   (6) *The number ε_α is β-infinitesimal for all $\beta \prec \alpha$.*
   (7) *The number $\omega_\alpha = 1/\varepsilon_\alpha$ is hypernatural and α-finite.*

7. **Theorem.** *The number ω_α is β-infinite for all $\beta \prec \alpha$.*

Proof. Immediate from items (6) and (7) of the previous axiom set. □

8. **Theorem.** *x is α-finite iff there is an α-natural n such that $|x| \leq n$.*

Proof. The statement is trivial for $x = 0$. If $x \neq 0$ is α-finite, so is $|x|$ because, assuming the opposite, $1/|x|$ would be α-infinitesimal and so would $1/x$ be by axiom 6.(4). By axiom 6.(5), the hypernatural $n = \lceil |x| \rceil < |x| + 1$ is then also α-finite. Conversely, let n be α-natural and $|x| \leq n$. If $1/|x|$ were α-infinitesimal, so would $1/n$ be by axiom 6.(4), and this contradicts the assumption that n is α-finite. □

Thus, we see that $L^{\alpha\text{-}st}$ is just L^{st} with all α-finite constants added.

9. **Corollary.** *$x \approx_\alpha 0$ iff $|x| < 1/n$ for all α-natural $n \geq 1$.*

For completeness, we list ERNA's 'weight' axioms and the related theorems, as we will repeatedly use them.

10. **Axiom set** (Weight)**.**
   (1) *if $\|x\|$ is defined, then $\|x\|$ is a nonzero hypernatural.*
   (2) *if $|x| = m/n \leq 1$ (m and $n \neq 0$ hypernaturals), then $\|x\|$ is defined, $\|x\|.|x|$ is hypernatural and $\|x\| \leq n$*
   (3) *if $|x| = m/n \geq 1$ (m and $n \neq 0$ hypernaturals), then $\|x\|$ is defined, $\|x\|/|x|$ is hypernatural and $\|x\| \leq m$.*

11. **Theorem.**
   (1) *If x is not a hyperrational, then $\|x\|$ is undefined.*
   (2) *If $x = \pm p/q$ with p and $q \neq 0$ relatively prime hypernaturals, then*

$$\| \pm p/q\| = \max\{|p|, |q|\}.$$

12. **Theorem.**
   (1) $\|0\| = 1$
   (2) *if $n \geq 1$ is hypernatural, $\|n\| = n$*
   (3) *if $\|x\|$ is defined, then $\|1/x\| = \|x\|$ and $\| \lceil x \rceil \| \leq \|x\|$*

(4) *if* $\|x\|$ *and* $\|y\|$ *are defined,* $\|x+y\|$*,* $\|x-y\|$*,* $\|xy\|$ *and* $\|x/y\|$ *are at most equal to* $(1+\|x\|)(1+\|y\|)$*, and* $\|x\hat{\ }y\|$ *is at most* $(1+\|x\|)\hat{\ }(1+\|y\|)$*.*

13. **Notation.** For any $0 < n \in \mathbb{N}$ we write $\|(x_1,\dots,x_n)\| = \max\{\|x_1\|,\dots,\|x_n\|\}$.

3. The consistency of $\mathrm{ERNA}^{\mathbb{A}}$

In this section, we prove the consistency of $\mathrm{ERNA}^{\mathbb{A}}$ inside PRA. We need the details of this proof for the proof of theorem 21.

As $\mathrm{ERNA}^{\mathbb{A}}$ is a quantifier-free theory, we can use Herbrand's theorem in the same way as in [12], [13] and [23], for more details, see [3] or [8]. To obtain ERNA's original consistency proof from the following, omit $\approx_\alpha$ for $\alpha \neq \mathbf{0}$ from the language.

14. **Theorem.** *The theory* $\mathrm{ERNA}^{\mathbb{A}}$ *is consistent and this consistency can be proved in* PRA.

Proof. In view of Herbrand's theorem, it suffices to show the consistency of every finite set of instantiated axioms of $\mathrm{ERNA}^{\mathbb{A}}$. Let T be such a set. We will define a mapping val_α on T, similar to the mapping val in ERNA's consistency proof. Thus, val_α maps the terms of T to rationals and the relations of T to relations on rationals, in such a way that all axioms of T are true under val_α. Hence T is consistent and the theorem follows.

First of all, as there are only finitely many elements of $\mathbb{A}$ in T, we interpret $(\mathbb{A},\preceq)$ as a suitable initial segment of $(\mathbb{N},\leq)$.

Second, like in the consistency proof of ERNA, all standard terms of T, except for min, are interpreted as their homomorphic image in the rationals: for all terms occurring in T, except min, ε_α, ω_α, we define

$$\mathrm{val}_\alpha(f(x_1,\dots,x_k)) := f(\mathrm{val}_\alpha(x_1),\dots,\mathrm{val}_\alpha(x_k)) \tag{1}$$

and for all relations R occurring in T, except $\approx_\alpha$, we define

$$\mathrm{val}_\alpha(R(x_1,\dots,x_k)) \text{ is true} \leftrightarrow R(\mathrm{val}_\alpha(x_1),\dots,\mathrm{val}_\alpha(x_k)). \tag{2}$$

Third, we need to gather some technical machinery. Let D be the maximum depth of the terms in T and let $\alpha_0,\alpha_1,\alpha_2,\dots,\alpha_{N-1}$ be all numbers of $\mathbb{A}$ that occur in T, with $\alpha_0 = \mathbf{0}$. As $\mathrm{ERNA}^{\mathbb{A}}$ has the same axiom schema for recursion as ERNA, no standard term of $\mathrm{ERNA}^{\mathbb{A}}$ grows faster than 2_k^x, for $k \in \mathbb{N}$. Hence, by [12, Theorem 30], there is a $0 < B \in \mathbb{N}$ such that for every term $f(\vec{x})$ occurring in T, not involving min, we have

$$||f(\vec{x})|| \leq 2_B^{||\vec{x}||}. \tag{3}$$

Further assume that t_D is the number of terms of depth D one can create using only function symbols occurring in T, and define $t := 3t_D + 3$.
With t and D, define the following functions:

$$f_0(x) = 2_B^x \text{ and } f_{n+1}(x) = f_n^t(x) = \underbrace{f_n(f_n(\dots(f_n(x))))}_{t\ f_n\text{'s}}. \tag{4}$$

Furthermore, define $a_0 := 1$ and

$$b_0^1 := f_{D+1}(a_0), c_0^1 := b_0^1, b_0^2 := f_{D+1}(c_0^1), c_0^2 := b_0^2, \dots, b_0^N := f_{D+1}\big(c_0^{N-1}\big), \tag{5}$$

and finally $c_0^N := b_0^N$ and $d_0 := f_{D+1}(c_0^N)$.

The numbers b_0^l allow us to interpret ε_α and ω_α:

$$\mathrm{val}_\alpha(\omega_{\alpha_1}) := b_0^1, \mathrm{val}_\alpha(\omega_{\alpha_2}) := b_0^2, \dots, \mathrm{val}_\alpha(\omega_{\alpha_{N-1}}) := b_0^{N-1} \tag{6}$$

and

$$\mathrm{val}_\alpha(\varepsilon_{\alpha_1}) := 1/b_0^1, \mathrm{val}_\alpha(\varepsilon_{\alpha_2}) := 1/b_0^2, \dots, \mathrm{val}_\alpha(\varepsilon_{\alpha_{N-1}}) := 1/b_0^{N-1}. \tag{7}$$

Hence we have an interpretation of all terms τ of depth zero such that $|\mathrm{val}_\alpha(\tau)| \in [0, a_0] \cup [b_0^1, c_0^1] \cup \dots \cup [b_0^N, c_0^N]$. For $i = 0$ and $1 \leq l \leq N-1$, we have

$$b_i^1 := f_{D-i+1}(a_i), b_i^{l+1} := f_{D-i+1}(c_i^l) \text{ and } d_i = f_{D-i+1}(c_i^N). \tag{8}$$

Then suppose that for $i \geq 0$ the numbers a_i, b_i^l, c_i^l and d_i have already been calculated and satisfy (8) and suppose val_α interprets all terms τ of depth i in such a way that $|\mathrm{val}_\alpha(\tau)| \in [0, a_i] \cup [b_i^1, c_i^1] \cup \dots \cup [b_i^N, c_i^N]$. We will now define a_{i+1}, b_{i+1}^l, c_{i+1}^l and d_{i+1}, which will satisfy (8) for $i+1$ and interpret all terms τ of depth $i+1$ in such a way that $|\mathrm{val}_\alpha(\tau)| \in [0, a_{i+1}] \cup [b_{i+1}^1, c_{i+1}^1] \cup \dots \cup [b_{i+1}^N, c_{i+1}^N]$.

In order to obtain a suitable interpretation for min, we define,

$$n_\varphi(\vec{x}) := (\mu n \leq d_i)\varphi(n, \mathrm{val}_\alpha(\vec{x})). \tag{9}$$

Let S_{i+1} be the set of all numbers $n_\varphi(\mathrm{val}_\alpha(\vec{\tau}))$ such that $\min_\varphi(\vec{\tau})$ has depth $i+1$ and is in T.

Now observe that, due to (8), the intervals $[a_i, b_i^1]$, $[c_i^l, b_i^{l+1}]$ and $[c_i^N, d_i]$ can be respectively partitioned in t intervals of the form

$$[f_{D-i}^j(a_i), f_{D-i}^{j+1}(a_i)], [f_{D-i}^j(c_i^l), f_{D-i}^{j+1}(c_i^l)] \text{ and } [f_{D-i}^j(c_i^N), f_{D-i}^{j+1}(c_i^N)] \tag{10}$$

for $j = 0, \dots, t-1 = 3t_D + 2$. Let V_{i+1} be the set of all numbers $n_\varphi(\vec{\tau})$ in S_{i+1} and all other terms $f(\vec{x})$ of T of depth at most $i+1$. Close V_{i+1} under taking the inverse and the weight, keeping in mind that $\|x\| = \|1/x\|$. Then V_{i+1} has at most $3t_D$ elements and recall that each partition in (10) has $3t_D + 3$ elements. Using the pigeon-hole principle, we can pick an interval, say the j_0-th one, which has empty intersection with V_{i+1}. Note that we can assume $1 \leq j_0 \leq 3t_D + 1$, because we have a surplus of three intervals. Finally we can define

$$a_{i+1} := f_{D-i}^{j_0}(a_i) \text{ and } b_{i+1}^1 := f_{D-i}^{j_0+1}(a_i) \tag{11}$$

The numbers b_{i+1}^l, c_{i+1}^l and d_{i+1} are defined in the same way. Hence (8) holds for $i+1$. Finally, we define

$$\mathrm{val}_\alpha(\min_\varphi(\vec{x})) := (\mu n \leq c_{i+1}^N)\varphi(n, \mathrm{val}_\alpha(\vec{x})) \tag{12}$$

for all $\min_\varphi(\vec{\tau})$ with depth $i+1$ in T. This definition, together with (3), yields that val_α interprets all terms τ of depth $i+1$ in such a way that $|\mathrm{val}_\alpha(\tau)| \in [0, a_{i+1}] \cup [b_{i+1}^1, c_{i+1}^1] \cup \dots \cup [b_{i+1}^N, c_{i+1}^N]$. Note that the latter property holds for all terms in V_{i+1}, in particular for $1/|\mathrm{val}_\alpha(\tau)|$.

After repeating this process D times, we obtain numbers a_D, b_D^l, c_D^l and d_D which allow us to interpret all terms of T. Finally, we give an interpretation to the relations $\approx_{\alpha_l}$:

$$\mathrm{val}_\alpha(\tau \approx_{\alpha_l} 0) \text{ is true} \leftrightarrow |\tau| \leq 1/b_D^{l+1}, \tag{13}$$

for $0 \leq l \leq N-1$. What is left is to show that under this interpretation val_α, all the axioms of T recieve the predicate true, which is done next.

Because most axioms of $\mathrm{ERNA}^{\mathbb{A}}$ hold for the rational numbers, the formulas (1) and (2) guarantee that all axioms of T have received a valid intepretation under val_α, except for axiom set 6 (Infinitesimals) above and ERNA's axiom set 31 (internal minimum).

First we treat the first axiom of 'Infinitesimals'. When either is zero, there is nothing to prove. Assume $\mathrm{val}_\alpha(\sigma \approx_{\alpha_l} 0)$ and $\mathrm{val}_\alpha(\tau \approx_{\alpha_l} 0)$ are true and that $\sigma + \tau$ appears in T. By (13), this implies $|\mathrm{val}_\alpha(\sigma)|, |\mathrm{val}_\alpha(\tau)| \leq 1/b_D^{l+1}$ or $1/|\mathrm{val}_\alpha(\tau)|, 1/|\mathrm{val}_\alpha(\sigma)| \geq b_D^{l+1}$. But since σ and τ have depth at most $D-1$, we have $1/|\mathrm{val}_\alpha(\tau)|, 1/|\mathrm{val}_\alpha(\sigma)| \in [0, a_{D-1}] \cup [b^1_{D-1}, c^1_{D-1}] \cup \cdots \cup [b^N_{D-1}, c^N_{D-1}]$ and since $a_{D-1} \leq a_D \leq b^{l+1}_D \leq b^{l+1}_{D-1}$, they must be in $\cup_{l+1 \leq k \leq N}[b^k_{D-1}, c^k_{D-1}]$. Hence we have $1/|\mathrm{val}_\alpha(\tau)|, 1/|\mathrm{val}_\alpha(\sigma)| \geq b^{l+1}_{D-1}$ or $|\mathrm{val}_\alpha(\tau)|, |\mathrm{val}_\alpha(\sigma)| \leq 1/b^{l+1}_{D-1}$, from which $|\mathrm{val}_\alpha(\sigma+\tau)| \leq 2/b^{l+1}_{D-1} < 1/b^{l+1}_D$. This last inequality is true, since $b^{l+1}_D > 2$ and $(b^{l+1}_D)^2 < b^{l+1}_{D-1}$. We have proved that $|\mathrm{val}_\alpha(\sigma+\tau)| \leq 1/b^{l+1}_D$, which is equivalent to $\mathrm{val}_\alpha(\sigma + \tau \approx_{\alpha_l} 0)$ being true. Hence the first axiom of the set 'Infinitesimals' receives the predicate true under val_α.

The second axiom of 'Infinitesimals' is treated in the same way as the first one.

The third axiom of 'Infinitesimals' holds trivially under val, since we cannot have that $|\mathrm{val}_\alpha(\tau)| \leq 1/b^{l+1}_D$ and $1/|\mathrm{val}_\alpha(\tau)| \leq 1/b^{l+1}_D$ hold at the same time. The fact that zero is α_l-finite, is immediate by the definition of the predicate 'x is α_l-finite'.

The fourth axiom of 'Infinitesimals' holds trivially, thanks to (13).

The fifth axiom of 'Infinitesimals' is treated like the first and second axiom of the same set.

The sixth and seventh item of 'Infinitesimals' both follow from (6), (7) and (13).

Now we will treat the axioms of the schema 'internal minimum'. First, note that the interval $[c^N_{i+1}, d^N_{i+1}]$, defined as in (11), has empty intersection with V_{i+1}. In particular, no term $n_\varphi(\vec{\tau})$ of T ends up in this interval. Thus, for terms $\min_\varphi$ of depth $i+1$, we have

$$\mathrm{val}_\alpha(\min_\varphi(\vec{\tau})) = (\mu n \leq c^N_{i+1})\varphi(n, \mathrm{val}_\alpha(\vec{\tau})) = (\mu n \leq c^N_D)\varphi(n, \mathrm{val}_\alpha(\vec{\vec{\tau}})) \qquad (14)$$

as c^N_D is in the interval $[c^N_{i+1}, d^N_{i+1}]$. We are ready to consider items (1)-(3) of the internal minimum schema. It is clear that item (1) always holds. For item (2), assume that the antecedent holds, i.e. $\mathrm{val}_\alpha(\min_\varphi(\vec{\tau}) > 0)$ is true. By the definition of $\mathrm{val}_\alpha(\min_\varphi)$ in (12), the consequent $\varphi(\mathrm{val}_\alpha(\min_\varphi(\vec{\tau})), \mathrm{val}_\alpha(\vec{\tau}))$ holds too. Hence item (2) holds. For item (3), assume that the antecedent holds, i.e. $\varphi(\mathrm{val}_\alpha(\sigma), \mathrm{val}_\alpha(\vec{\tau}))$ holds for some σ in T. This implies $\mathrm{val}_\alpha(\sigma) \leq c^N_D$ and thus there is a number $n \leq c^N_D$ such that $\varphi(n, \mathrm{val}_\alpha(\vec{\tau}))$. By (14), $\mathrm{val}_\alpha(\min_\varphi(\vec{\tau}))$ is the least of these and hence the formulas '$\min_\varphi(\vec{\tau}) \leq \sigma$' and '$\varphi(\min_\varphi(\vec{\tau}), \vec{\tau})$' receive a true interpretation under val_α. Thus, item (3) is also interpreted as true and we are done with this schema.

All axioms of T have received a true interpretation under val_α, hence T is consistent and, by Herbrand's theorem, $\mathrm{ERNA}^{\mathbb{A}}$ is. Now, Herbrand's theorem is provable in $I\Sigma_1$ and this theory is Π_2-conservative over PRA (see [3, 8]). As consistency can be formalized by a Π_1-formula, it follows immediately that PRA proves the consistency of $\mathrm{ERNA}^{\mathbb{A}}$. □

Note that if we define, in (5), a_0 as a number larger than 1 and any c^l_0 as a number larger than b^l_0, we still obtain a valid interpretation val_α for T and the consistency proof goes through.

The choice of $(\mathbb{A}, \preceq)$ is arbitrary, hence it is consistent with $\mathrm{ERNA}^{\mathbb{A}}$ that $\mathbb{A}$ is dense. It is possible to make this explicit by adding the following axiom to $\mathrm{ERNA}^{\mathbb{A}}$, for all nonzero $\alpha, \beta \in \mathbb{A}$.

$$\omega_\alpha < \omega_\beta \to \omega_\alpha < \omega_{\frac{\alpha+\beta}{2}} < \omega_\beta. \qquad (15)$$

The notation '$\frac{\alpha+\beta}{2}$' is of course purely symbolic. This axiom receives a valid interpretation by interpreting $(\mathbb{A}, \preceq)$ as $(\mathbb{Q}, \leq)$.

In the following, we repeatedly need overflow and underflow. Thus, we prove it explicitly in $\mathrm{ERNA}^{\mathbb{A}}$.

15. **Theorem.** *Let $\varphi(n)$ be an internal quantifier-free formula, not involving* min.

(1) *If $\varphi(n)$ holds for every α-natural n, it holds for all hypernatural n up to some α-infinite hypernatural $\overline{n}$* ***(overflow)***.

(2) *If $\varphi(n)$ holds for every α-infinite hypernatural n, it holds for all hypernatural n from some α-natural $\underline{n}$ on* ***(underflow)***.

Both numbers $\overline{n}$ and $\underline{n}$ are given by explicit $\text{ERNA}^{\mathbb{A}}$*-formulas not involving* min.

Proof. Let ω be some α-infinite number. For the first item, define

$$\overline{n} := (\mu n \leq \omega)\neg\varphi(n+1), \tag{16}$$

if $(\exists n \leq \omega)\neg\varphi(n+1)$ and zero otherwise. By theorem [12, Theorem 58], this term is available in ERNA and hence in $\text{ERNA}^{\mathbb{A}}$. Likewise for underflow. $\square$

The previous theorem shows that overflow holds for all $\alpha \in \mathbb{A}$, i.e. at all levels of infinity. As no one level is given exceptional status, this seems only natural. Furthermore, one intuitively expects formulas that do not explicitly depend on a certain level to be true at all levels if they are true at one. In the following section, we investigate a general principle that transfers universal formulas to all levels of infinity.

4. $\text{ERNA}^{\mathbb{A}}$ and Stratified Transfer

In nonstandard mathematics, Transfer expresses Leibniz's principle that the 'same' laws hold for standard and nonstandard objects alike. Typically, Transfer only applies to formulas involving standard objects, excluding e.g. ERNA's cosine $\sum_{i=0}^{\omega}(-1)^i\frac{x^{2i}}{(2i)!}$. In set theoretical approaches to nonstandard analysis, the standard part function 'st' applied to such an object, results in a standard object, thus solving this problem. The latter function is not available in ERNA, but 'generalized' transfer principles for objects like ERNA's cosine can be obtained (see [13, Theorem 19] and [18]), at the cost of introducing '$\approx$'. Unfortunately, formulas with occurrences of the predicate '$\approx$' are always excluded from Transfer, even in the classical set-theoretical approach.

For $\text{ERNA}^{\mathbb{A}}$, we wish to obtain a transfer principle that applies to all universal formulas, possibly involving $\approx$. As an example, consider the following formula, expressing the continuity of the standard function f on $[0,1]$:

$$(\forall x, y \in [0,1])(x \approx y \to f(x) \approx f(y)). \tag{17}$$

Assuming (17), it seems only natural that if $x \approx_\alpha y$ for $\alpha \succ \mathbf{0}$, then $f(x) \approx_\alpha f(y)$. In other words, there should hold, for all $\alpha \in \mathbb{A}$,

$$(\forall x, y \in [0,1])(x \approx_\alpha y \to f(x) \approx_\alpha f(y)), \tag{18}$$

which is (17), with $\approx$ replaced with $\approx_\alpha$. Incidentally, when f is a polynomial, an easy computation shows that (18) indeed holds, even for polynomials in $L^{\alpha\text{-}st}$. Below, we turn this into a general principle.

16. **Notation.** Let Φ^α be a formula of $L^{\alpha\text{-}st} \cup \{\approx_\alpha\}$. Then Φ^β is Φ^α with all occurrences of $\approx_\alpha$ replaced with $\approx_\beta$.

17. **Principle** (Stratified Transfer). *Assume $\alpha \succeq \mathbf{0}$ and let Φ^α be a quantifier-free formula of $L^{\alpha\text{-}st} \cup \{\approx_\alpha\}$, not involving* min. *For every $\beta \succ \alpha$,*

$$(\forall \vec{x})\Phi^\alpha(\vec{x}) \leftrightarrow (\forall \vec{x})\Phi^\beta(\vec{x}). \tag{19}$$

Note that Φ may involve α-standard parameters. We always tacitly allow (α-standard) parameters in all transfer principles in this paper, unless explicitly stated otherwise.

18. **Principle** (Weak Stratified Transfer)**.** *Assume $\alpha \succeq \mathbf{0}$ and let $f(\vec{x}, k)$ be a function of $L^{\alpha\text{-}st}$, not involving* min *and weakly increasing in k. For all $\beta \succ \alpha$, the following statements are equivalent*

$$\text{‘}f(\vec{x}, k) \text{ is } \alpha\text{-infinite for all } \vec{x} \text{ and all } \alpha\text{-infinite } k\text{’}$$

and

$$\text{‘}f(\vec{x}, k) \text{ is } \beta\text{-infinite for all } \vec{x} \text{ and all } \beta\text{-infinite number } k\text{’}.$$

The second transfer principle is a special case of the first. However, by the following theorem, the seemingly weaker second principle is actually equivalent to the first. We sometimes abbreviate ‘for all α-infinite ω’ by ‘$(\forall^{\alpha}\omega)$’.

19. **Theorem.** *In* ERNA$^{\mathbb{A}}$*, Weak Stratified Transfer is equivalent to Stratified Transfer.*

Proof. First, assume the Weak Stratified Transfer Principle and let $\Phi^{\alpha}(\vec{x})$ be as in the Stratified Transfer Principle. Replace in $\Phi^{\alpha}(\vec{x})$ all positive occurrences of $\tau_i(\vec{x}) \approx_{\alpha} 0$ with $(\forall^{\alpha\text{-}st} n_i)(|\tau_i(\vec{x})| < 1/n_i)$, where n_i is a new variable not yet appearing in $\Phi^{\alpha}(\vec{x})$. Do the same for the negative occurrences, using new variables m_i. Bringing all quantifiers in $(\forall\vec{x})\Phi^{\alpha}(\vec{x})$ to the front, we obtain

$$(\forall\vec{x})(\forall^{\alpha\text{-}st} n_1, \ldots, n_l)(\exists^{\alpha\text{-}st} m_1, \ldots, m_k)\Psi(\vec{x}, n_1, \ldots, n_l, m_1, \ldots, m_k),$$

where Ψ is quantifier-free and in $L^{\alpha\text{-}st}$. Using pairing functions, we can reduce all n_i to one variable n and reduce all m_i to one variable m. Hence the previous formula becomes

$$(\forall\vec{x})(\forall^{\alpha\text{-}st} n)(\exists^{\alpha\text{-}st} m)\Xi(\vec{x}, n, m),$$

where Ξ is quantifier-free and in $L^{\alpha\text{-}st}$. Fix some α-infinite number ω_1, we obtain

$$(\forall\vec{x})(\forall^{\alpha\text{-}st} n)(\exists m \le \omega_1)\Xi(\vec{x}, n, m),$$

Applying overflow, with $\omega = \omega_1$ in (16), yields

$$(\forall\vec{x})(\forall n \le \overline{n}(\vec{x}, \omega_1))(\exists m \le \omega_1)\Xi(\vec{x}, n, m).$$

Hence the function $\overline{n}(\vec{x}, k)$ is α-infinite for all $\vec{x}$ and α-infinite k and weakly increasing in k. By the Weak Stratified Transfer Principle, $\overline{n}(\vec{x}, k)$ is β-infinite for all $\vec{x}$ and all β-infinite k, for $\beta \succ \alpha$. Hence, for all $\vec{x}$, β-finite n and β-infinite k, we have

$$(\exists m \le k)\Xi(\vec{x}, n, m).$$

Fix $\vec{x}_0$ and β-finite n_0. Since $(\exists m \le k)\Xi(\vec{x}_0, n_0, m)$ holds for all β-infinite k, underflow yields $(\exists^{\beta\text{-}st} m)\Xi(\vec{x}_0, n_0, m)$. This implies

$$(\forall\vec{x})(\forall^{\beta\text{-}st} n)(\exists^{\beta\text{-}st} m)\Xi(\vec{x}, n, m).$$

Unpairing the variables n and m and bringing the quantifiers back in the formula, we obtain $(\forall\vec{x})\Phi^{\beta}(\vec{x})$. Thus, we have proved the forward implication in (19).

In the same way, it is proved that $(\forall\vec{x})\Phi^{\beta}(\vec{x})$ implies $(\forall\vec{x})\Phi^{\alpha}(\vec{x})$, i.e., the reverse implication in (19), assuming the Weak Stratified Transfer Principle.

Hence we proved that the Weak Stratified Transfer Principle implies the Stratified Transfer Principle. As the reverse implication is trivial, we are done. □

By the previous theorem, it suffices to prove the consistency of ERNA$^{\mathbb{A}}$ with the Weak Stratified Transfer Principle. Instead of proving this consistency directly, we show, in the next section, that Weak Stratified Transfer follows from Π_3^{α}-TRANS. The latter is ERNA$^{\mathbb{A}}$'s version of the classical transfer principle limited to Π_3-formulas. The schema Π_3^{α}-TRANS is analogous to Π_1-TRANS and Σ_2-TRANS, introduced in [12] and [13]. We suspect that PRA cannot prove the consistency of ERNA$^{\mathbb{A}}$ + Π_3^{α}-TRANS.

To conclude this section, we point to [10], where the importance of Stratified Transfer is discussed. Moreover, analysis developed in $\text{ERNA}^{\mathbb{A}}$ in section 6 is more elegant when Stratified Transfer is available. Also, Stratified Transfer (in some form or other) seems to be compatible with the spirit of 'strict' finitism (see [26]), as it merely lifts true universal formulas to higher levels. It would be interesting to know the exact logical strength of Stratified Transfer and how it can be weakened by imposing certain 'constructive' limitations on $\mathbb{A}$.

5. $\text{ERNA}^{\mathbb{A}}$ and regular Transfer

In this section, we will introduce the 'new' transfer principles Π_1^α-TRANS and Σ_2^α-TRANS, which are $\text{ERNA}^{\mathbb{A}}$-versions of the 'old' schemas Π_1-TRANS and Σ_2-TRANS. The adaptations made to the latter schemas to obtain the former are both natural and in line with the Stratified Transfer Principle above. We give a consistency proof for the extended theory, which requires significant changes to the consistency proof in [12]. We only sketch a consistency proof for $\text{ERNA}^{\mathbb{A}} + \Sigma_2^\alpha$-TRANS. Finally, using the new transfer principles, we prove that transfer for Π_3-formulas is sufficient for the Stratified Transfer Principle.

5.1. **Transfer for Π_1 and Σ_1-formulas.** Here, we introduce a 'stratified' version of transfer for Π_1 and Σ_1-formulas for $\text{ERNA}^{\mathbb{A}}$ and show that the extended theory is consistent. The following axiom schema is $\text{ERNA}^{\mathbb{A}}$'s version of Π_1-TRANS.

20. **Axiom schema** (Stratified Π_1-transfer)**.** *For every quantifier-free formula $\varphi(n)$ of $L^{\alpha\text{-}st}$, not involving min, we have*

$$(\forall^{\alpha\text{-}st} n)\varphi(n) \to (\forall n)\varphi(n). \tag{20}$$

The previous axiom schema is denoted by Π_1^α-TRANS and its parameter-free counterpart is denoted by Π_1^α-TRANS$^-$. Similarly, let Π_1-TRANS$^-$ be the parameter-free version of Π_1-TRANS (see also remark 52). After the consistency proof, the reasons for the restrictions on φ will become apparent. Resolving the implication in (20), we see that this formula is equivalent to

$$(0 < \min_{\neg\varphi} \text{ is } \alpha\text{-finite}) \vee (\forall n)\varphi(n). \tag{21}$$

Thus, $\text{ERNA}^{\mathbb{A}} + \Pi_1^\alpha$-TRANS$^-$ is equivalent to a quantifier-free theory and we may use Herbrand's theorem to prove its consistency. To obtain the consistency proof in [12] from the following proof, omit $\approx_\alpha$ for $\alpha \neq \mathbf{0}$ from the language.

21. **Theorem.** *The theory* $\text{ERNA}^{\mathbb{A}} + \Pi_1^\alpha$-TRANS$^-$ *is consistent and this consistency can be proved by a finite iteration of* $\text{ERNA}^{\mathbb{A}}$*'s consistency proof.*

Proof. Despite the obvious similarities between the theories $\text{ERNA} + \Pi_1$-TRANS$^-$ and $\text{ERNA}^{\mathbb{A}} + \Pi_1^\alpha$-TRANS$^-$, the consistency proof of the former (see [12, Theorem 44]) breaks down for the latter. The reason is that one of the explicit conditions for the consistency proof of $\text{ERNA} + \Pi_1$-TRANS$^-$ to work, is that φ must be in L^{st}. But in Π_1^α-TRANS$^-$, φ is in $L^{\alpha\text{-}st}$ and as such, the formula φ in (21) may contain the nonstandard number ω_β for $\beta \preceq \alpha$.

However, it is possible to salvage the original proof. We use Herbrand's theorem in the same way as in the consistency proof of $\text{ERNA}^{\mathbb{A}}$. Thus, let T be any finite set of instantiated axioms of $\text{ERNA}^{\mathbb{A}} + \Pi_1^\alpha$-TRANS$^-$. Leaving out the transfer axioms from T, we are left with a finite set T' of instantiated $\text{ERNA}^{\mathbb{A}}$ axioms. Let val_α be its interpretation into the rationals as in $\text{ERNA}^{\mathbb{A}}$'s consistency proof. However, nothing guarantees that the instances of Π_1^α-TRANS$^-$ in T are also interpreted as 'true' under val_α. We will adapt val_α by successively increasing the starting values defined in (5), if necessary. The resulting map will interpret all axioms in T as true, not just those in T'.

Let T and T' be as in the previous paragraph. Let D be the maximum depth of the terms in T. Let $\alpha_0, \dots, \alpha_{N-1}$ be all elements of $\mathbb{A}$ in T, with $\alpha_0 = \mathbf{0}$. For notational convenience, for φ as in Π_1^α-TRANS$^-$, we shall write $\varphi(n, \vec{\tau})$ instead of $\varphi(n)$, where $\vec{\tau}$ contains all numbers occurring in φ that are not in L^{st}. Finally, let the list $\varphi_1(n, \vec{\tau}_1), \dots, \varphi_M(n, \vec{\tau}_M)$ consist of the quantifier-free formulas whose Π_1^α-transfer axiom (21) occurs in T. If necessary, we arrange this list of formulas in such a way that $i < j$ implies that all ω_α in the range of $\vec{\tau}_i$ satisfy $\omega_\alpha \preceq \omega_\beta$ for some ω_β in the range of $\vec{\tau}_j$.

By (13), $\Omega_l := \bigcup_{l+1 \le i \le N} [b_D^i, c_D^i]$ is the set where val_α maps the α_l-infinite numbers. Also, $O_l := [0, a_D] \cup [b_D^1, c_D^1] \cup \dots \cup [b_D^l, c_D^l]$ is the set where val_α maps the α_l-finite numbers. If we have, for all $i \in \{1, ..., M\}$ and all $l \in \{0, ..., N-1\}$ such that $\gamma \preceq \alpha_l$ for all ω_γ in the range of $\vec{\tau}_i$, that

$$(\exists m \in O_l) \neg \varphi_i(m, \mathrm{val}_\alpha(\vec{\tau}_i)) \vee (\forall n \in [0, a_D] \cup \Omega_0) \varphi_i(n, \mathrm{val}_\alpha(\vec{\tau}_i)), \tag{22}$$

we see that val_α provides a true interpretation of the whole of T, not just T', as every instance of (21) receives a valid interpretation, in this case. However, nothing guarantees that (22) holds for all such numbers i and l. Thus, assume there is an exceptional $\varphi'(n, \vec{\tau}') := \varphi_i(n, \vec{\tau}_i)$ and l, for which

$$(\forall m \in O_l) \varphi'(m, \mathrm{val}_\alpha(\vec{\tau}')) \wedge \big(\exists n \in \big[b_D^{l+1}, c_D^{l+1}\big]\big) \neg \varphi'(n, \mathrm{val}_\alpha(\vec{\tau}')). \tag{23}$$

Now fix $\vec{\tau}'$ and let l_0 be the least l satisfying the previous formula. Then (23) implies $(\exists n \in \Omega_{l_0}) \neg \varphi'(n, \mathrm{val}(\vec{\tau}'))$, i.e. there is an '$\alpha_{l_0}$-infinite' n such that $\neg \varphi'(n, \mathrm{val}(\vec{\tau}'))$. Now choose a number $n_0 > c_D^N$ (for notational clarity, we write $a_0 = c_0^0$, for the case $l_0 = 0$) and construct a new interpretation val'_α with the same starting values as in (5), except for $(c_0^{l_0})' := n_0$. This val'_α continues to make the axioms in T' true and does the same with the instances in T of the axiom

$$(0 < \min_{\neg\varphi'}(\vec{\tau}') \text{ is } \alpha_{l_0}\text{-finite}) \vee (\forall n) \varphi'(n, \vec{\tau}') \tag{24}$$

Indeed, if a number $n \in \Omega_{l_0}$ is such that $\neg \varphi'(n, \mathrm{val}_\alpha(\vec{\tau}'))$, the number n is interpreted by val'_α as an α_{l_0}-finite number because $n \le c_D^N \le (c_0^{l_0})' \le (c_D^{l_0})'$ by our choice of $(c_0^{l_0})'$. Thus, the sentence $(\exists n \in O'_{l_0}) \neg \varphi'(n, \mathrm{val}_\alpha(\vec{\tau}'))$ is true. By definition, $\vec{\tau}'$ only contains numbers ω_{α_i} for $i \le l_0$ and (6) implies $\mathrm{val}_\alpha(\omega_{\alpha_i}) = b_0^i$, for $1 \le i \le N$. But increasing $c_0^{l_0}$ to $(c_0^{l_0})'$, as we did before, does not change the numbers $b_0^1, \dots, b_0^{l_0}$. Hence $\mathrm{val}_\alpha(\vec{\tau}') = \mathrm{val}'_\alpha(\vec{\tau}')$ and so $(\exists n \in O'_{l_0}) \neg \varphi'(n, \mathrm{val}_\alpha(\vec{\tau}'))$ implies $(\exists n \in O'_{l_0}) \neg \varphi'(n, \mathrm{val}'_\alpha(\vec{\tau}'))$. Thus, $(0 < \min_{\neg\varphi'}(\vec{\tau}') \text{ is } \alpha_{l_0}\text{-finite})$ is true under val'_α and so is the whole of (24).

Define T'' as T' plus all instances of (24) occurring in T. If there is another exceptional φ_i and l_0 such that (23) holds, repeat this process. Note that if we increase another c_0^j for $j \ge l_0$ and construct val''_α, the latter still makes the axioms of T' true, but the axioms of T'' as well, since increasing c_0^j does not change the interpretations of the numbers ω_{α_i} for $i \le l_0$ either. Hence (24) is true under val'' for the same reason as for val'. Recall that the list $\varphi_1(n, \vec{\tau}_1), \dots, \varphi_M(n, \vec{\tau}_M)$ is arranged in such a way that $i < j$ implies that all ω_α in the range of $\vec{\tau}_i$ satisfy $\omega_\alpha \preceq \omega_\beta$ for some ω_β in the range of $\vec{\tau}_j$. This arrangement of the list guarantees that the changes we make to val_α to satisfy a certain transfer axiom, do not invalidate a transfer axiom treated earlier.

This process, repeated, will certainly halt: either the two lists $\{1, \dots, M\}$ and $\{1, \dots, N-1\}$ become exhausted or, at some earlier stage, a valid interpretation is found for T. Note that this consistency proof is a finite iteration of ERNA$^{\mathbb{A}}$'s. $\square$

The restrictions on the formulas φ admitted in (20) are imposed by our consistency proof. Indeed, for every α_i occurring in T, the interpretation of ω_{α_j} for $j > i$ depends on the choice of c_0^i. By our changing $c_0^{l_0}$ into $(c_0^{l_0})' > c_0^{l_0}$, formulas like (24) could loose their 'true' interpretation from one step to the next, if they contain such ω_j. Likewise, the changing of c_0^l can change the interpretation of $\approx_\beta$, for any $\beta \in \mathbb{A}$, and hence this predicate cannot occur in φ. The exclusion of min has, of course, a different reason: $\min_\varphi$ is only allowed in ERNA when φ does not rely on min.

For convenience, we will usually use Π_1^α-TRANS instead of Π_1^α-TRANS$^-$. By contraposition, the schema Π_1^α-TRANS implies the following schema, which we denote Σ_1^α-TRANS.

22. **Axiom schema** (Stratified Σ_1-transfer)**.** *For every quantifier-free formula $\varphi(n)$ of $L^{\alpha\text{-}st}$, not involving min, we have*

$$(\exists n)\varphi(n) \to (\exists^{\alpha\text{-}st} n)\varphi(n). \tag{25}$$

Note that both in (20) and (25), the reverse implication is trivial. For $\varphi \in L^{\alpha\text{-}st}$, the levels $\beta \succeq \alpha$ are sometimes called the 'context' levels of φ and α is the called the 'minimial' context level, i.e. the lowest level on which all constants occurring in φ exist. In this respect, Σ_1^α-transfer expresses that true existential formulas can be pushed down to their minimal context level, which corresponds to their level of standardness.

5.2. **Transfer for Σ_2 and Π_2-formulas.** In order to obtain transfer for Σ_2 and Π_2-formulas in ERNA, we added a certain axiom schema to ERNA + Π_1-TRANS and showed that the resulting theory has transfer for Σ_2 and Π_2-formulas, see [13] for details. We also discussed why this approach is preferable to a more 'direct' approach. Here, we shall employ the same method to obtain 'Stratified Transfer' for Σ_2 and Π_2-formulas. As the method is similar to that used in [13], we only sketch the proofs. Our goal is to obtain the following transfer principle.

23. **Axiom schema** (Stratified Σ_2-transfer)**.** *For every quantifier-free formula φ from $L^{\alpha\text{-}st}$, not involving* min*, we have*

$$(\exists n)(\forall m)\varphi(n,m) \leftrightarrow (\exists^{\alpha\text{-}st} n)(\forall^{\alpha\text{-}st} m)\varphi(n,m). \tag{26}$$

We denote this schema by Σ_2^α-TRANS. By contraposition, it is equivalent to the Π_2^α-transfer principle

$$(\forall n)(\exists m)\varphi(n,m) \leftrightarrow (\forall^{\alpha\text{-}st} n)(\exists^{\alpha\text{-}st} m)\varphi(n,m). \tag{27}$$

In view of the equivalence between (26) and (27), we will only mention Π_2^α-transfer in the sequel if it is explicitly required. We will add certain axioms to ERNA$^{\mathbb{A}}$ + Π_1^α-TRANS and prove the consistency of the resulting theory. Then we show that the extended theory proves the above Σ_2^α-transfer principle.

First consider the following theorem of ERNA$^{\mathbb{A}}$ + Π_1^α-TRANS.

24. **Theorem.** *In* ERNA$^{\mathbb{A}}$ + Π_1^α-TRANS *we have, for every quantifier-free formula $\varphi(n,m)$ of $L^{\alpha\text{-}st}$ not involving* min*, the implication*

$$(\exists n)(\forall m)\varphi(n,m) \to (\forall^{\alpha\text{-}st} k)(\exists^{\alpha\text{-}st} n)(\forall m \le k)\varphi(n,m). \tag{28}$$

Proof. If the antecedent holds, we have $(\exists n)(\forall m \le k)\varphi(n,m)$ for every α-finite k. By Σ_1^α-transfer, $(\exists^{\alpha\text{-}st} n)(\forall m \le k)\varphi(n,m)$, hence the consequent of (28). □

By the previous theorem, (29) implies the forward implication in (26).

25. **Axiom schema** (TRANS^+_α)**.** *For every quantifier-free formula* $\varphi(n,m)$ *of* $L^{\alpha\text{-}st}$ *not involving* min*, we have*

$$(\exists n)(\forall m)\varphi(n,m) \to \left(\begin{array}{c} (\forall^{\alpha\text{-}st}k)(\exists^{\alpha\text{-}st}n)(\forall m \leq k)\varphi(n,m) \\ \downarrow \\ (\exists^{\alpha\text{-}st}n)(\forall^{\alpha\text{-}st}m)\varphi(n,m) \end{array}\right) \tag{29}$$

Theorem 27 will show that Σ_2^α-transfer as stated in (26) is provable in $\text{ERNA}^{\mathbb{A}}+\Pi_1^\alpha\text{-TRANS}+\text{TRANS}^+_\alpha$. Therefore, the latter theory will be abbreviated to $\text{ERNA}^{\mathbb{A}}+\Sigma_2^\alpha$-TRANS. The schema TRANS^+_α can be skolemized in exactly the same way as the schema TRANS^+, see [13, Theorem 3] for details. We have the following theorem.

26. **Theorem.** *The theory* $\text{ERNA}^{\mathbb{A}}+\Sigma_2^\alpha$-TRANS *is consistent.*

Proof. The proof of the consistency of ERNA + Σ_2-TRANS in [13] can easily be converted into a proof for the theorem at hand. The adaptations are minimal, as the skolemization of (29) is also a tautology in the finite setting of the model for an arbitrary finite subset of instantiated $\text{ERNA}^{\mathbb{A}}+\Pi_1^\alpha$-TRANS-axioms. □

Now we prove the main result of this section, viz. that $\text{ERNA}^{\mathbb{A}}+\Sigma_2^\alpha$-TRANS has Σ_2^α-transfer.

27. **Theorem.** *In* $\text{ERNA}^{\mathbb{A}}+\Sigma_2^\alpha$-TRANS*, the* Σ_2^α*-transfer principle, stated in* (26)*, holds.*

Proof. By theorem 26 we know that we can consistently add the axiom schema 25 to $\text{ERNA}^{\mathbb{A}}+\Pi_1^\alpha$-TRANS. In the extended theory, theorem 24 yields that (29) implies the forward implication in (26). For the inverse implication, assume that $(\exists^{\alpha\text{-}st}n)(\forall^{\alpha\text{-}st}m)\varphi(n,m)$ and fix α-finite n_0 such that $(\forall^{\alpha\text{-}st}m)\varphi(n_0,m)$. By Π_1^α-transfer, this implies $(\forall m)\varphi(n_0,m)$ and hence $(\exists n)(\forall m)\varphi(n,m)$. □

Using pairing functions, we immediately obtain Stratified Σ_2^α and Π_2^α-transfer for formulas involving blocks of quantifiers. As for Σ_1^α-transfer, Σ_2^α-transfer as in (26) expresses that a true Σ_2-formula can be pushed down to its minimal context level

5.3. **Transfer for Σ_3 and Π_3-formulas.** Here, we show that a certain transfer principle for Π_3-formulas, called Π_3^α-TRANS, is sufficient to obtain Weak Stratified Transfer. We first introduce the former. Note that it is the natural extension of Σ_2^α and Π_1^α-transfer.

28. **Axiom schema** (Stratified Π_3-transfer)**.** *For every quantifier-free formula* φ *of* $L^{\alpha\text{-}st}$*, not involving min, we have,*

$$(\forall^{\alpha\text{-}st}n)(\exists^{\alpha\text{-}st}m)(\forall^{\alpha\text{-}st}k)\varphi(n,m,k) \leftrightarrow (\forall n)(\exists m)(\forall k)\varphi(n,m,k). \tag{30}$$

We denote this schema by Π_3^α-TRANS. We now prove the main theorem of this section, namely that Π_3^α-transfer is sufficient to obtain Stratified Transfer.

29. **Theorem.** *The theory* $\text{ERNA}^{\mathbb{A}}+\Pi_3^\alpha$-TRANS *proves the Weak Stratified Transfer Principle.*

Proof. Assume $\mathbf{0} \preceq \alpha \prec \beta$ and let f be as in the Weak Stratified Transfer Principle and assume that $f(n,\vec{x})$ is α-infinite for all $\vec{x}$ and all α-infinite n. This implies that

$$(\forall\vec{x})(\forall^{\alpha\text{-}st}n)(\forall^\alpha\omega)(f(\omega,\vec{x})>n),$$

where the notation '$(\forall^\alpha\omega)$' denotes 'for all α-infinite numbers ω'. Fixing $\vec{x}_0$ and α-finite n_0 and applying underflow to the formula $(\forall^\alpha\omega)(f(\omega,\vec{x}_0)>n_0)$, yields the existence of an α-finite number k_0 such that $(f(k_0,\vec{x}_0)>n_0)$. Hence,

$$(\forall\vec{x})(\forall^{\alpha\text{-}st}n)(\exists^{\alpha\text{-}st}m)(f(m,\vec{x})>n), \tag{31}$$

and, by [12, Theorem 58], there is a function $g(n,\vec{x})$ which calculates the least m such that $f(m,\vec{x})>n$, for any $\vec{x}$ and α-finite n. Thus, (31) implies

$$(\forall^{\alpha\text{-}st}n)(\forall\vec{x})(f(g(n,\vec{x}),\vec{x})>n), \tag{32}$$

where $g(n,\vec{x})$ is α-finite for α-finite n and any $\vec{x}$. Now fix an α-infinite hypernatural ω_1 and define $h(n)$ as $\max_{\|\vec{x}\|\leq\omega_1} g(n,\vec{x})$. By definition, the function $h(n)$ is α-finite for α-finite n. As f is weakly increasing in its first argument, (32) implies

$$(\forall^{\alpha\text{-}st}n)(\forall\vec{x})(\|\vec{x}\|\leq\omega_1\to f(h(n),\vec{x})>n),$$

and also

$$(\forall^{\alpha\text{-}st}n)(\exists m\leq h(n))(\forall\vec{x})(\|\vec{x}\|\leq\omega_1\to f(m,\vec{x})>n).$$

We previously showed that $h(n)$ is α-finite for α-finite n. Thus,

$$(\forall^{\alpha\text{-}st}n)(\exists^{\alpha\text{-}st}m)(\forall\vec{x})(\|\vec{x}\|\leq\omega_1\to f(m,\vec{x})>n),$$

and also

$$(\forall^{\alpha\text{-}st}n)(\exists^{\alpha\text{-}st}m)(\forall^{\alpha\text{-}st}\vec{x})(f(m,\vec{x})>n). \tag{33}$$

By Π_3^α-transfer, this implies that

$$(\forall^{\beta\text{-}st}n)(\exists^{\beta\text{-}st}m)(\forall^{\beta\text{-}st}\vec{x})(f(m,\vec{x})>n). \tag{34}$$

Fixing appropriate β-finite n_0 and m_0, and applying Π_1^α-transfer, yields

$$(\forall^{\beta\text{-}st}n)(\exists^{\beta\text{-}st}m)(\forall\vec{x})(f(m,\vec{x})>n).$$

This formula implies that $f(k,\vec{x})$ is β-infinite for all $\vec{x}$ and all β-infinite k. The other implication in the Weak Stratified Transfer Principle is proved in the same way. □

It is clear form the proof why the theorem fails for β such that $\mathbf{0}\preceq\beta\prec\alpha$. Indeed, as f may contain ω_α, we cannot apply Π_3^α-transfer to (33) for such β.

Note that (Weak) Stratified Transfer is fundamentally different from the other transfer principles, as $\approx_\alpha$ can occur in the former, but not in the latter. In this respect, it is surprising that a 'regular' transfer principle such as Π_3^α-TRANS implies (Weak) Stratified Transfer.

However, if we consider things from the point of view of set theory, we can explain this remarkable correspondence between 'regular' and 'stratified' transfer. Internal set theory is an axiomatic approach to nonstandard mathematics (see [14] for details). Examples include Nelson's **IST** ([15]), Kanovei's **BST** ([14]), Péraire's **RIST** ([16]) and Hrbacek's **FRIST*** and **GRIST** ([10] and [11]), which inspired parts of ERNA$^{\mathbb{A}}$. These set theories are extensions of **ZFC** and most have a so called 'Reduction Algorithm'. This effective procedure applies to certain general classes of formulas and removes any predicate not in the original $\in$-language of **ZFC**. The resulting formula agrees with the original formula on standard objects. Thus, in **GRIST**, it is possible to remove the relative standardness predicate '$\sqsubseteq$' and hence transfer for formulas in the $\in$-$\sqsubseteq$-language follows from transfer for formulas in the $\in$-language. Similarly, in theorem 19, we show that transfer for formulas involving the relative standardness predicate $\approx_\alpha$ can be reduced to a very specific instance, involving fewer predicates $\approx_\alpha$. Later, in theorem 29, we prove that the remaining standardness predicates can be removed from the formula too, producing (33) and (34). Thus, we have reduced 'stratified' transfer to 'regular' transfer. In turn, it is surprising that a set-theoretical metatheorem such as the Reduction Algorithm appears in theories with strength far below **ZFC**.

6. Analysis in $\text{ERNA}^{\mathbb{A}}$

In this section, we obtain some basic theorems of analysis. We shall work in $\text{ERNA}^{\mathbb{A}} + \Pi_3^\alpha$-TRANS, i.e. we may use the Stratified Transfer Principle. Most theorems can be proved in $\text{ERNA}^{\mathbb{A}}$, at the cost of adding extra technical conditions. This is usually mentioned in a corollary.

For the rest of this section, we assume that $\mathbf{0} \prec \alpha \prec \beta$, that a and b are α-finite and that the functions f and g do not involve the minimum operator $\min_\varphi$.

6.1. **Continuity.** Here, we define the notion of continuity in $\text{ERNA}^{\mathbb{A}}$ and prove some fundamental theorems.

30. **Definition.** A function f is α-continuous at a point x_0, if $x \approx_\alpha x_0$ implies $f(x) \approx_\alpha f(x_0)$. A function is α-continuous over $[a, b]$ if

$$(\forall x, y \in [a, b])(x \approx_\alpha y \to f(x) \approx_\alpha f(y)).$$

As usual, we write 'continuous' instead of '$\mathbf{0}$-continuous'. If f is α and β-continuous for $\alpha \neq \beta$, we say that f is 'α, β-continuous'

31. **Theorem.** *If f is α-continuous over $[a, b]$ and α-finite in one point of $[a, b]$, it is α-finite for all x in $[a, b]$.*

Proof. Let f be as in the theorem, fix α-finite k_0 and consider

$$(\forall x, y \in [a, b])(|x - y| \le 1/N \wedge \|x, y\| \le \omega_\beta \to |f(x) - f(y)| < 1/k_0). \tag{35}$$

As f is α-continuous, this formula holds for all α-infinite N. By [12, Corollary 53], (35) is quantifier-free and applying underflow yields that it holds for all $N \ge N_0$, where N_0 is α-finite. Then let $x_0 \in [a, b]$ be such that $f(x_0)$ is α-finite. We may assume it satisfies $\|x_0\| \le \omega_\beta$. Using (35) for $N = N_0$, it easily follows that $f(x)$ deviates at most $(N_0\lceil b - a\rceil)/k_0$ from $f(x_0)$ for $\|x\| \le \omega_\beta$. As the points $x_n := a + \frac{n(b-a)}{\omega_\beta}$ partition the interval $[a, b]$ in α-infinitesimal subintervals, the theorem follows. □

32. **Corollary.** *If $f \in L^{\alpha\text{-}st}$ is α-continuous over $[a, b]$, it is α-finite for all $x \in [a, b]$.*

Proof. Let $f(x, \vec{x})$ be the function $f(x)$ from the corollary with all nonstandard numbers replaced with free variables. By [12, Theorem 30], there is a $k \in \mathbb{N}$ such that $\|f(x, \vec{x})\| \le 2_k^{\|x, \vec{x}\|}$. Thus, $f(x)$ is α-finite for α-finite x. Applying the theorem finishes the proof. □

By Stratified Transfer, an α-continuous function of $L^{\alpha\text{-}st}$ $\big($e.g. $\text{ERNA}^{\mathbb{A}}$'s cosine $\sum_{n=0}^{\omega_\alpha}(-1)^n \frac{x^{2n}}{(2n)!}\big)$ is also β-continuous for all $\beta \succeq \alpha$. Similar statements hold for integrability and differentiability. For the sake of brevity, we mostly do not explicitly mention these properties.

6.2. **Differentiation.** Here, we define the notion of differentiability in $\text{ERNA}^{\mathbb{A}}$ and prove some fundamental theorems. To this end, we need some notation.

33. **Notation.**

(1) A nonzero number x is '$\overline{\alpha}$-infinitesimal' or 'strict α-infinitesimal' (with respect to β) if $x \approx_\alpha 0 \wedge x \not\approx_\beta 0$. We denote this by $x \approx_{\overline{\alpha}} 0$.
(2) We write '$a \ll_\alpha b$' instead of '$a < b \wedge a \not\approx_\alpha b$' and '$a \lessapprox_\beta b$' instead of '$a < b \vee a \approx_\beta b$'.
(3) We write $\Delta_h(f)(x)$ instead of $\frac{f(x+h)-f(x)}{h}$.

We use the following notion of differentiability.

34. **Definition.**

(1) A function f is 'α-differentiable at x_0' if $\Delta_\varepsilon f(x_0) \approx_\alpha \Delta_{\varepsilon'} f(x_0)$ for all nonzero $\varepsilon, \varepsilon' \approx_\alpha 0$ and both quotients are α-finite.
(2) If f is α-differentiable at x_0 and $\varepsilon \approx_\alpha 0$, then $\Delta_\varepsilon f(x_0)$ is called 'the derivative of f at x_0' and is denoted $D_\alpha f(x_0)$.
(3) A function f is called 'α-differentiable over (a, b)' if it is α-differentiable at every point $a \ll_\alpha x \ll_\alpha b$.
(4) The concepts '$\overline{\alpha}$-differentiable' and '$\overline{\alpha}$-derivative' are defined by replacing, in the previous items, '$\varepsilon, \varepsilon' \approx_\alpha 0$' by '$\varepsilon, \varepsilon' \approx_{\overline{\alpha}} 0$'. We use the same notation for the $\overline{\alpha}$-derivative as for the α-derivative.

The choice of ε is arbitrary and hence the derivative is only defined 'up to infinitesimals'. There seems to be no good way of defining it more 'precisely', i.e. not up to infinitesimals, without the presence of a 'standard part' function 'st_α' which maps α-finite numbers to their α-standard part.

35. **Theorem.** *If a function f is α-differentiable over (a, b), it is α-continuous at all $a \ll_\alpha x \ll_\alpha b$.*

Proof. Immediate from the definition of differentiability. □

36. **Theorem.** *Let $f(x)$ and $g(x)$ be α-standard and α-differentiable over (a, b). Then $f(x)g(x)$ is α-differentiable over (a, b) and*

$$D_\alpha(fg)(x) \approx_\alpha D_\alpha f(x)\, g(x) + f(x) D_\alpha g(x) \tag{36}$$

for all $a \ll_\alpha x \ll_\alpha b$.

Proof. Assume f and g are α-differentiable over (a, b). Let ε be an α-infinitesimal and x such that $a \ll_\alpha x \ll_\alpha b$. Then,

$$\begin{aligned}
D_\alpha(fg)(x) &\approx_\alpha \tfrac{1}{\varepsilon}(f(x+\varepsilon)g(x+\varepsilon) - f(x)g(x)) \\
&= \tfrac{1}{\varepsilon}(f(x+\varepsilon)g(x+\varepsilon) - f(x)g(x+\varepsilon) + f(x)g(x+\varepsilon) - f(x)g(x)) \\
&= \tfrac{1}{\varepsilon}((f(x+\varepsilon) - f(x))g(x+\varepsilon) + f(x)(g(x+\varepsilon) - g(x))) \\
&= \tfrac{f(x+\varepsilon)-f(x)}{\varepsilon} g(x+\varepsilon) + f(x)\tfrac{g(x+\varepsilon)-g(x)}{\varepsilon} \\
&\approx_\alpha D_\alpha f(x) g(x+\varepsilon) + f(x) D_\alpha g(x) \approx_\alpha D_\alpha f(x) g(x) + f(x) D_\alpha g(x).
\end{aligned}$$

The final two steps follow from theorem 35 and corollary 32. Hence $f(x)g(x)$ is α-differentiable over (a, b) and (36) indeed holds. □

By theorem 31, the requirement '$f, g \in L^{\alpha\text{-}st}$' in the previous theorem, can be dropped if we additionally require fg to be α-finite in one point of (a, b). In the following theorem, there is no such requirement.

37. **Theorem** (Chain rule)**.** *Let g be α-differentiable at a and let f be α-differentiable at $g(a)$. Then $f \circ g$ is α-differentiable at a and*

$$D_\alpha(f \circ g)(a) \approx_\alpha D_\alpha f(g(a))\, D_\alpha g(a). \tag{37}$$

Proof. Let f and g be as in the theorem and assume $0 \neq \varepsilon \approx_\alpha 0$. First of all, since g is α-differentiable at a, we have, that $D_\alpha g(a) \approx_\alpha \frac{g(a+\varepsilon)-g(a)}{\varepsilon}$, which implies

$$g(a+\varepsilon) = \varepsilon D_\alpha g(a) + g(a) + \varepsilon\varepsilon'$$

for some $\varepsilon' \approx_\alpha 0$. Then $\varepsilon'' = \varepsilon D_\alpha g(a) + \varepsilon\varepsilon'$ is also α-infinitesimal. If $\varepsilon'' \neq 0$, then, as f is α-differentiable at $g(a)$, we have $D_\alpha f(g(a)) \approx_\alpha \frac{f(g(a)+\varepsilon'')-f(g(a))}{\varepsilon''}$. This implies

$$f(g(a) + \varepsilon'') = \varepsilon'' D_\alpha f(g(a)) + f(g(a)) + \varepsilon''\varepsilon'''$$

for some $\varepsilon''' \approx_\alpha 0$. If $\varepsilon'' = 0$, then the previous formula holds trivially for the same ε'''. Note that $\frac{\varepsilon''\varepsilon'''}{\varepsilon} \approx_\alpha 0$. Hence we have

$$\begin{aligned}\Delta_\varepsilon(f \circ g)(a) = \tfrac{f(g(a+\varepsilon))-f(g(a))}{\varepsilon} &= \tfrac{f(g(a)+\varepsilon'')-f(g(a))}{\varepsilon} \\ &= \tfrac{\varepsilon'' D_\alpha f(g(a))+\varepsilon''\varepsilon'''+f(g(a))-f(g(a))}{\varepsilon} \quad \approx_\alpha \tfrac{\varepsilon''}{\varepsilon} D_\alpha f(g(a)).\end{aligned}$$

By definition, $\frac{\varepsilon''}{\varepsilon} \approx_\alpha D_\alpha g(a)$ and hence $f \circ g$ is α-differentiable at a and (37) holds. □

It is easily verified that the theorems of this section still hold if we replace 'α-differentiable' with '$\overline{\alpha}$-differentiable'.

6.3. **Integration.** Here, we define the notion of Riemann integral in ERNA$^{\mathbb{A}}$ and prove some fundamental theorems.

In classical analysis, the Riemann-integral is defined as the limit of Riemann sums over ever finer partitions. In ERNA$^{\mathbb{A}}$, we adopt the following definition for the concept 'partition'.

38. **Definition.** A partition π of $[a, b]$ is a vector $(x_1, \ldots, x_n, t_1, \ldots t_{n-1})$ such that $x_i \leq t_i \leq x_{i+1}$ for all $1 \leq i \leq n-1$ and $a = x_1$ and $b = x_n$. The number $\delta = \max_{2 \leq i \leq n}(x_i - x_{i-1})$ is called the 'mesh' of the partition π.

For the definition of integrability, we need to quantify over all partitions of an interval. In [12], it is proved that ERNA contains pairing functions, which can uniquely code vectors of numbers into numbers (and decode them back). As partitions are merely vectors, it is intuitively clear that quantifying over all partitions of an interval is possible in ERNA, and thus in ERNA$^{\mathbb{A}}$. Also, in [18], the previous claim is proved explicitly. Incidentally, Riemann integration inside NQA$^+$, the predecessor of ERNA, uses equidistant partitions.

Assume that ω is α-infinite and that $a \ll_\alpha b$. Let n_0 be the least n such that $\frac{n}{\omega} > a$ and let n_1 be the least n such that $\frac{n}{\omega} > b$. Define $a_\omega := \frac{n_0}{\omega}$ and $b_\omega := \frac{n_1 - 1}{\omega}$. Like the derivative, the Riemann integral can only be defined 'up to infinitesimals'. Hence, for α-Riemann integrable functions, it does not matter whether we use the interval $[a, b]$ or the interval $[a_\omega, b_\omega]$ in its definition. From now on, we tacitly assume that $a \ll_\alpha b$.

39. **Definition** (Riemann Integration)**.** Let f be a function defined on $[a, b]$.
   (1) Given a partition $(x_1, \ldots, x_n, t_1, \ldots, t_{n-1})$ of $[a, b]$, the Riemann sum corresponding to f is defined as $\sum_{i=2}^n f(t_{i-1})(x_i - x_{i-1})$.
   (2) The function f is α-Riemann integrable on $[a, b]$, if for all partitions of $[a, b]$ with mesh $\approx_\alpha 0$, the Riemann sums are α-finite and α-infinitely close.
   (3) If f is α-Riemann integrable on $[a, b]$, then the integral of f over $[a, b]$, denoted as $\int_a^b f(x)\, d(x, \alpha)$, is the Riemann sum corresponding to f of the equidistant partition of $[a_\omega, b_\omega]$ with mesh $\varepsilon = \frac{1}{\omega} \approx_\alpha 0$ and points $t_i = \frac{x_{i+1}+x_i}{2}$.

40. **Theorem.** *A function f which is α-continuous and α-finite over $[a, b]$, is α-Riemann integrable over $[a, b]$.*

Proof. The proof for $\alpha = \mathbf{0}$ is given in [18] and can easily be adapted to $\alpha \succ \mathbf{0}$. □

41. **Theorem.** *Let f be α-continuous and α-finite over $[a, b]$ and assume $a \ll_\alpha c \ll_\alpha b$. Hence,*

$$\int_a^b f(x)\, d(x, \alpha) \approx_\alpha \int_a^c f(x)\, d(x, \alpha) + \int_c^b f(x)\, d(x, \alpha).$$

Proof. Immediate from the previous theorem and the definition of the Riemann integral. □

42. **Theorem.** *Let c be an α-finite positive constant such that $c \not\approx_\alpha 0$ and let f be α-continuous and α-finite over $[a, b+c]$. We have*

$$\int_a^b f(x+c)\,d(x,\alpha) \approx_\alpha \int_{a+c}^{b+c} f(x)\,d(x,\alpha).$$

Proof. Immediate from theorem 40 and the definition of the Riemann integral. □

43. **Theorem** (Second fundamental theorem)**.** *Let $f \in L^{\alpha\text{-}st}$ be α-continuous on $[a,b]$ and let $F(x)$ be $\int_a^x f(t)d(t,\beta)$. Then $F(x)$ is $\overline{\alpha}$-differentiable over (a,b) and the equation $D_\alpha F(x) \approx_\alpha f(x)$ holds for all $a \ll_\alpha x \ll_\alpha b$.*

Proof. Fix $\varepsilon \approx_{\overline{\alpha}} 0$ and x such that $a \ll_\alpha x \ll_\alpha b$. We have

$$\tfrac{F(x+\varepsilon)-F(x)}{\varepsilon} = \tfrac{1}{\varepsilon}\left(\int_a^{x+\varepsilon} f(t)d(t,\beta) - \int_a^x f(t)d(t,\beta)\right) \approx_\beta \tfrac{1}{\varepsilon}\int_x^{x+\varepsilon} f(t)d(t,\beta), \quad (38)$$

as ε is not β-infinitesimal. Let ω_1 be β-infinite and define $x_i = x + \frac{i\varepsilon}{\omega_1}$. Let $f(y_1)$ and $f(y_2)$ be the least and the largest $f(x_i)$ for $i \leq \omega_1$. As f is α,β-continuous, $m := f(y_1)$ and $M := f(y_2)$ are such that $m \lessapprox_\beta f(y) \lessapprox_\beta M$ for $y \in [x, x+\varepsilon]$ and $m \approx_\alpha M \approx_\alpha f(x)$. This implies

$$\varepsilon m \lessapprox_\beta \int_x^{x+\varepsilon} f(t)d(t,\beta) \lessapprox_\beta \varepsilon M,$$

and hence

$$m \lessapprox_\beta \tfrac{1}{\varepsilon}\int_x^{x+\varepsilon} f(t)d(t,\beta) \lessapprox_\beta M,$$

as ε is not β-infinitesimal. Thus,

$$m \approx_\alpha \tfrac{1}{\varepsilon}\int_x^{x+\varepsilon} f(t)d(t,\beta) \approx_\alpha M \approx_\alpha f(x).$$

By (38), F is $\overline{\alpha}$-differentiable and the theorem follows. □

44. **Corollary.** *The condition '$f \in L^{\alpha\text{-}st}$' in the theorem can be dropped if we require f to be α,β-continuous over $[a,b]$ and α-finite in one point of $[a,b]$.*

Proof. It is an easy verification that the proof of the theorem still goes through with these conditions. □

45. **Example.** Define $\varepsilon = \varepsilon_\alpha^4$. The function $d(x) = \frac{\varepsilon}{\varepsilon^2+x^2}$ is α,β-continuous for α-finite x and at most $1/\varepsilon_\alpha^4$. The function $\arctan x := \int_0^x \frac{d(t,\beta)}{1+t^2}$ is $\overline{\alpha}$-differentiable in all α-finite x and we have $D_\alpha\big(\arctan(x/\varepsilon)\big) \approx_\alpha \frac{\varepsilon}{\varepsilon^2+x^2}$ for all α-finite x.

46. **Theorem** (First fundamental theorem)**.** *Let $f \in L^{\alpha\text{-}st}$ be $\overline{\alpha}$-differentiable over (a,b) and such that $D_\alpha f$ is β-continuous over $[a,b]$. For $a \ll_\alpha c \ll_\alpha d \ll_\alpha b$, we have $\int_c^d D_\alpha f(x)\,d(x,\beta) \approx_\alpha f(d) - f(c)$.*

Proof. Let c, d be as stated and let ε be strict α-infinitesimal. Note that $d - c$ is α-finite. We have

$$\begin{aligned}
\int_c^d D_\alpha f(x)\,d(x,\beta) &\approx_\alpha \int_c^d \tfrac{f(x+\varepsilon)-f(x)}{\varepsilon}\,d(x,\beta) \\
&\approx_\beta \tfrac{1}{\varepsilon}\left(\int_c^d f(x+\varepsilon)\,d(x,\beta) - \int_c^d f(x)\,d(x,\beta)\right) \\
&\approx_\beta \tfrac{1}{\varepsilon}\left(\int_{c+\varepsilon}^{d+\varepsilon} f(x)\,d(x,\beta) - \int_c^d f(x)\,d(x,\beta)\right) \\
&\approx_\beta \tfrac{1}{\varepsilon}\left(\int_d^{d+\varepsilon} f(x)\,d(x,\beta) - \int_c^{c+\varepsilon} f(x)\,d(x,\beta)\right).
\end{aligned}$$

As in the proof of the second fundamental theorem, we have $\int_c^{c+\varepsilon} f(x)\,d(x,\beta) \approx_\alpha f(c)$ and $\int_d^{d+\varepsilon} f(x)\,d(x,\beta) \approx_\alpha f(d)$ and we are done. □

47. **Corollary** (Partial Integration)**.** *Let $f, g \in L^{\alpha\text{-}st}$ be $\overline{\alpha}$-differentiable over (a,b) and let $D_\alpha f$ and $D_\alpha g$ be β-continuous over $[a,b]$. For $a \ll_\alpha c \ll_\alpha d \ll_\alpha b$,*

$$\int_c^d f(x) D_\alpha g(x)\, d(x,\beta) \approx_\alpha \big[f(x)g(x)\big]_c^d - \int_c^d D_\alpha f(x) g(x)\, d(x,\beta).$$

Proof. Immediate from the second fundamental theorem and theorem 36. □

By theorem 35, we can drop the requirement '$f, g \in L^{\alpha\text{-}st}$' if we additionally require fg to be α-finite in one point of (a,b).

For simulating the Dirac Delta distribution, we need to introduce an extra level γ such that $\mathbf{0} \prec \gamma \prec \alpha$. We also need the function arctan.

48. **Theorem.** *Define the (finite) constant π as $4\arctan(1)$.*

(1) *For all α-finite x,* $\arctan(\pm|x|) + \arctan\big(\pm\frac{1}{|x|}\big) \approx_\alpha \pm\pi/2$.
(2) *We have* $\arctan(\pm\omega_\alpha^3) \approx_\gamma \pm\pi/2$.

Proof. The first item follows by calculating the $\overline{\alpha}$-derivative of $\arctan x + \arctan 1/x$ using the chain rule and noting that the result is α-infinitesimally close to zero. Thus, there is a constant C such that $\arctan x + \arctan 1/x \approx_\alpha C$, for all α-finite positive x. Substituting $x = 1$ yields $C = \pi/2$. The case $x < 0$ is treated in the same way. The second item follows from the previous item and the fact that $\arctan x$ is γ-continuous at zero. □

49. **Definition.** A function $f \in L^{\gamma\text{-}st}$ is said to have a 'compact support' if it is zero outside some interval $[a,b]$ with a, b γ-finite.

50. **Theorem.** *Let $f \in L^{\gamma\text{-}st}$ be an γ-differentiable function with compact support such that $D_\alpha f(x)$ is β-continuous for $x \approx_\gamma 0$. We have*

$$\frac{1}{\pi}\int_{-\omega_\alpha}^{\omega_\alpha} d(x) f(x) d(x,\beta) \approx_\gamma f(0).$$

Proof. Assume that $f(x)$ is zero outside $[a,b]$, with a, b γ-finite. First, we prove that $\int_{\varepsilon_\alpha}^b f(x)d(x)\, d(x,\beta) \approx_\gamma 0$. As $|x| \geq \varepsilon_\alpha$ implies $x^2 \geq \varepsilon_\alpha^2$ we have $d(x) = \frac{\varepsilon}{\varepsilon^2 + x^2} \leq \frac{\varepsilon}{x^2} \leq \frac{\varepsilon}{{\varepsilon_\alpha}^2} = \varepsilon_\alpha^2 < \varepsilon_\alpha$. Hence the integral $\int_{\varepsilon_\alpha}^b |d(x)|\,|f(x)|\, d(x,\beta)$ is at most $\varepsilon_\alpha \int_{\varepsilon_\alpha}^b |f(x)|\, d(x,\beta)$. As f is γ-finite and γ-continuous on $[a,b]$, we have $\int_{\varepsilon_\alpha}^b f(x)d(x)\, d(x,\beta) \approx_\gamma 0$. In the same way, we have $\int_a^{\varepsilon_\alpha} f(x)d(x)\, d(x,\beta) \approx_\gamma 0$ and $\int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} \arctan(x/\varepsilon)) D_\alpha f(x)\, d(x,\beta) \approx_\gamma 0$. Hence we have

$$\int_{-\omega_\alpha}^{\omega_\alpha} d(x) f(x) d(x,\beta) \approx_\beta \int_a^b d(x) f(x)\, d(x,\beta) \approx_\gamma \int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} d(x) f(x)\, d(x,\beta).$$

If $0 \notin [a,b]$, then $f(0) = 0$ and the theorem follows. Otherwise, by example 45, the function $d(x)$ is α-infinitesimally close to $\frac{1}{\pi} D_\alpha \arctan(x/\varepsilon)$, yielding

$$\int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} d(x) f(x)\, d(x,\beta) \approx_\alpha \frac{1}{\pi}\int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} D_\alpha(\arctan(x/\varepsilon))\, f(x)\, d(x,\beta).$$

The product $\arctan(x/\varepsilon)f(x)$ satisfies all conditions for partial integration, implying

$$\begin{aligned}
&\int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} D_\alpha(\arctan(x/\varepsilon))\, f(x)\, d(x,\beta) \\
&\approx_\alpha \big[\arctan(x/\varepsilon)\ \ f(x)\big]_{-\varepsilon_\alpha}^{\varepsilon_\alpha} - \textstyle\int_{-\varepsilon_\alpha}^{\varepsilon_\alpha} \arctan(x/\varepsilon_\alpha) D_\alpha f(x) d(x,\beta) \\
&\approx_\gamma [\arctan(x/\varepsilon)\ \ f(x)]_{-\varepsilon_\alpha}^{\varepsilon_\alpha} \\
&= \big(\arctan(\varepsilon_\alpha/\varepsilon)\, f(\varepsilon_\alpha) - \arctan(-\varepsilon_\alpha/\varepsilon)\, f(-\varepsilon_\alpha)\big) \\
&\approx_\gamma \big(\arctan(\omega_\alpha^3)\ f(0) - \arctan(-\omega_\alpha^3)\ f(0)\big) \approx_\gamma \pi f(0).
\end{aligned}$$

□

The function $d(x)$ has the typical 'Dirac Delta' shape: 'infinite at zero and zero everywhere else' and many functions like $d(x)$ exist. Also, if we define $H(x) = \frac{1}{\pi}\arctan(x/\varepsilon) + \frac{1}{2}$, we have $D_\alpha H(x) \approx_\alpha d(x)$ and $H(x)$ only differs from the 'usual' Heaviside function by an infinitesimal. In the same way as in the previous theorem, it is possible to prove statements like

$$\int_{-\omega_\alpha}^{\omega_\alpha} D_\xi d(x) f(x) d(x,\beta) \approx_\gamma -\int_{-\omega_\alpha}^{\omega_\alpha} d(x) D_\xi f(x) d(x,\beta) \approx_\gamma -\pi D_\xi f(0).$$

in $\mathrm{ERNA}^{\mathbb{A}}$, for $\alpha \prec \xi \prec \beta$. We have introduced the function $\arctan x$, because we needed its properties in theorem 50. The rest of the basic functions of analysis are easily defined and their well-known properties are almost immediate, thanks to Stratified Transfer.

In this section, we have shown that analysis can be developed inside $\mathrm{ERNA}^{\mathbb{A}}$ and its extensions in a concise and elegant way. We did not attempt to give an exhaustive treatment and have deliberately omitted large parts of analysis like e.g. higher order derivatives. It is interesting, however, to briefly consider the latter. In [10], Hrbacek argues that stratified analysis yields a more elegant way of defining higher order derivatives than regular nonstandard analysis. In this way, a function $D_\alpha f(x)$ is differentiable, if it is β-differentiable for $\beta \succ \alpha$ and $f''(x)$ is defined as $D_\beta D_\alpha f(x)$. Thus, to manipulate an object such as $D_\alpha f(x)$, which is not part of $L^{\alpha\text{-}st}$, we need to go to a higher level β, where $D_\alpha f(x)$ is standard. The same principle is at the heart of most theorems in this section, in particular the first fundamental theorem (theorem 46). This principle is the essence of stratified analysis, and occurs in all of mathematics: to study a set of objects, we extend it and gain new insights (e.g. real versus complex analysis). Thanks to Stratified Transfer, all levels have the same standard properties and thus, the extension to a higher level is always uniform.

7. Towards a formal framework for physics

We have introduced $\mathrm{ERNA}^{\mathbb{A}}$ and proved its consistency inside PRA. We subsequently obtained several results of analysis using the elegant framework of stratified analysis. Thus, $\mathrm{ERNA}^{\mathbb{A}}$ is a good formal framework for doing finitistic analysis in a quantifier-free way, akin to the way mathematics is done in physics. As it turns out, Stratified Transfer gives us an even better framework. We sketch an example to illustrate this claim.

It seems only fair to say that physicists employ a lower standard of mathematical rigor than mathematicians (see [5] for details). In this way, limits are usually pushed inside or outside integrals without a second thought. Moreover, a widely held 'rule of thumb' is that if, after performing a mathematically dubious manipulation, the result still makes physical and (to a lesser extent) mathematical sense, the manipulation was probably sound. As it turns out, stratified nonstandard analysis is a suitable formal framework for this sort of 'justification a posteriori'. We illustrate this with an example.

51. **Example.** Let f_i, a and b be standard objects. According to the previously mentioned 'rule of thumb', the following manipulation

$$\int_a^b \sum_{i=0}^{\infty} f_i(x,y)\,dx = \sum_{i=0}^{\infty} \int_a^b f_i(x,y)\,dx =: \sum_{i=0}^{\infty} g_i(y) =: g(y)$$

is considered valid in physics as long as the function $g(y)$ is physically and/or mathematically meaningful. In stratified analysis, assuming $\mathbf{0} \prec \alpha \prec \beta$, the previous becomes

$$\int_a^b \sum_{i=0}^{\omega_\alpha} f_i(x,y)\,d(x,\beta) \approx \sum_{i=0}^{\omega_\alpha} \int_a^b f_i(x,y)\,d(x,\beta) =: \sum_{i=0}^{\omega_\alpha} h_i(y) =: h(y).$$

The first step follows from Stratified Transfer. Indeed, as a finite summation can be pushed through a Riemann integral, a β-finite summation can be pushed through a β-Riemann integral. Thus, we can always obtain $h(y)$ and if it is finite (the very least for it to be physically meaningful), we have $h(y) \approx g(y)$, thus justifying our 'rule of thumb'.

52. **Remark.** In [12], the authors introduce the transfer principle Π_1-TRANS without stating whether standard parameters are allowed or not. Define Π_1-TRANS (Π_1-TRANS$^-$) as schema 43 of [12] with (without) standard parameters in φ. The proof of theorem 44 in [12] is obviously only correct for ERNA + Π_1-TRANS$^-$, as ERNA + Π_1-TRANS interprets $I\Sigma_1$, by theorem 45 in the same paper. In the rest of [12], in particular §4 and §6, the schema Π_1-TRANS is used. The authors hereby apologize for this oversight. Although the schemas Π_1^α-TRANS$^-$ and Π_1-TRANS$^-$ originate from technical considerations, they turn out to play an important role in the context of Reverse Mathematics. We will explore this avenue of research in [19].

53. **Acknowledgement.** The author thanks Professor Karel Hrbacek (City University of New York, City College) for his valuable advice.

References

[1] Jeremy Avigad, *Weak theories of nonstandard arithmetic and analysis*, Reverse mathematics 2001, Lect. Notes Log., vol. 21, Assoc. Symbol. Logic, La Jolla, CA, 2005, pp. 19–46.
[2] ———, *Number theory and elementary arithmetic*, Philos. Math. (3) **11** (2003), 257–284.
[3] Samuel R. Buss, *An introduction to proof theory*, Handbook of proof theory, Stud. Logic Found. Math., vol. 137, North-Holland, Amsterdam, 1998, pp. 1–78.
[4] Rolando Chuaqui and Patrick Suppes, *Free-variable Axiomatic Foundations of Infinitesimal Analysis: A Fragment with Finitary Consistency Proof*, Journal of Symbolic Logic **60** (1995), 122-159.
[5] Kevin Davey, *Is mathematical rigor necessary in physics?*, British J. Philos. Sci. **54** (2003), no. 3, 439–463.
[6] Harvey Friedman, *Grand Conjectures*, FOM mailing list (16 April 1999).
[7] E. I. Gordon, *Nonstandard methods in commutative harmonic analysis*, Vol. 164, American Mathematical Society, Providence, RI, 1997.
[8] Petr Hájek and Pavel Pudlák, *Metamathematics of First-Order Aritmetic*, Springer, 1998.
[9] Jacques Herbrand, *Écrits logiques*, Presses Universitaires de France, Paris, 1968 (French).
[10] Karel Hrbacek, *Stratified analysis?*, The strength of nonstandard analysis, Springer Wien, New York, Vienna, 2007, pp. 47–63.
[11] ———, *Relative set theory: internal view*, J. Log. Anal. **1** (2009), Paper 8, 108.
[12] Chris Impens and Sam Sanders, *Transfer and a supremum principle for ERNA*, Journal of Symbolic Logic **73** (2008), 689-710.
[13] ———, *Saturation and Σ_2-transfer for ERNA*, J. Symbolic Logic **74** (2009), no. 3, 901–913.

[14] Vladimir Kanovei and Michael Reeken, *Nonstandard analysis, axiomatically*, Springer, 2004.

[15] Edward Nelson, *Internal set theory: a new approach to nonstandard analysis*, Bull. Amer. Math. Soc. **83** (1977), no. 6, 1165–1198.

[16] Yves Péraire, *Théorie relative des ensembles internes*, Osaka J. Math. **29** (1992), no. 2, 267–297 (French).

[17] Michal Rössler and Emil Jeřábek, *Fragment of nonstandard analysis with a finitary consistency proof*, Bulletin of Symbolic Logic **13** (2007), 54-70.

[18] Sam Sanders, ERNA *and Friedman's Reverse Mathematics*, To appear in Journal of Symbolic Logic.

[19] ______, *Levels of Reverse Mathematics*, In preparation.

[20] Stephen G. Simpson, *Subsystems of second order arithmetic*, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1999.

[21] ______ (ed.), *Reverse mathematics 2001*, Lecture Notes in Logic, vol. 21, Association for Symbolic Logic, La Jolla, CA, 2005.

[22] Keith D. Stroyan and Willem A.J. Luxemburg, *Introduction to the theory of infinitesimals*, Academic Press, 1976.

[23] Richard Sommer and Patrick Suppes, *Finite Models of Elementary Recursive Nonstandard Analysis*, Notas de la Sociedad Matematica de Chile **15** (1996), 73-95.

[24] ______, *Dispensing with the Continuum*, Journal of Math. Psy. **41** (1997), 3-10.

[25] Patrick Suppes and Rolando Chuaqui, *A finitarily consistent free-variable positive fragment of Infinitesimal Analysis*, Proceedings of the IXth Latin American Symposium on Mathematical Logic Notas de Logica Mathematica **38** (1993), 1-59.

[26] Willem W. Tait, *Finitism*, The Journal of Philosophy **78** (1981), 524-564.

[27] Terence Tao, *Ultrafilters, nonstandard analysis, and epsilon management*, 2007. Wordpress blog, `http://terrytao.wordpress.com/`.