True infinitesimal differential geometry

Mikhail G. Katz*

Vladimir Kanovei

Tahl Nowik

M. Katz, Department of Mathematics, Bar Ilan University, Ramat Gan 52900 Israel
Email address: katzmik@math.biu.ac.il

V. Kanovei, IPPI, Moscow, and MIIT, Moscow, Russia
Email address: kanovei@googlemail.com

T. Nowik, Department of Mathematics, Bar Ilan University, Ramat Gan 52900 Israel
Email address: tahl@math.biu.ac.il

*Supported by the Israel Science Foundation grant 1517/12.

Abstract. This text is based on two courses taught at Bar Ilan University. One is the undergraduate course 89132 taught to groups of about 120, 130, and 150 freshmen during the '14-'15, '15-'16, and '16-'17 academic years. This course exploits infinitesimals to introduce all the basic concepts of the calculus, mainly following Keisler's textbook. The other is the graduate course 88826 that has been taught to between 5 and 10 graduate students yearly for the past several years. Much of what we write here is deeply affected by this pedagogical experience.

The graduate course develops two viewpoints on infinitesimal generators of flows on manifolds. The classical notion of an infinitesimal generator of a flow on a manifold is a (classical) vector field whose relation to the actual flow is expressed via *integration*. A true infinitesimal framework allows one to re-develop the foundations of differential geometry in this area. The True Infinitesimal Differential Geometry (TIDG) framework enables a more transparent relation between the infinitesimal generator and the flow.

Namely, we work with a *combinatorial* object called a hyperreal *walk* with infinitesimal step, constructed by hyperfinite iteration. We then deduce the continuous flow as the real shadow of the said walk.

Namely, a vector field is defined via an infinitesimal displacement in the manifold to itself. Then the walk is obtained by *iteration* rather than *integration*. We introduce *synthetic* combinatorial conditions D^k for the regularity of a vector field, replacing the classical *analytic* conditions of C^k type. The D^k conditions guarantee the usual theorems such as uniqueness and existence of solution of ODE locally, Frobenius theorem, Lie bracket, analysis of small oscillations of the pendulum, and other concepts.

Here we cover vector fields, infinitesimal generator of flow, Frobenius theorem, hyperreals, infinitesimals, transfer principle, ultrafilters, ultrapower construction, hyperfinite partitions, microcontinuity, internal sets, halo, prevector, prevector field, regularity conditions for prevector fields, flow of a prevector field, relation to classical flow. The concluding mathematical chapters prove the transfer principle.

The historical part of the book takes a fresh look at several centuries of the development of infinitesimal procedures in analysis from the viewpoint enabled by Robinson's framework for infinitesimal mathematics. See also related publications at `http://u.math.biu.ac.il/~katzmik/infinitesimals.html`

Contents

Preface

Terence Tao has recently published a number of texts where he exploits ultraproducts in general, and Robinson's framework for infinitesimals in particular, as a fundamental tool; see e.g., [**Tao 2014**], [**Tao & Vu 2016**]. In the present text, we apply such an approach to the foundations of differential geometry.

Differential geometry has its roots in the study of curves and surfaces by methods of mathematical analysis in the 17th century. Accordingly it inherited the infinitesimal methods typical of these early studies in analysis. Weierstrassian effort at the end of the 19th century solidified mathematical foundations by breaking with the infinitesimal mathematics of Leibniz, Euler, and Cauchy, but in many ways the baby was thrown out with the water, as well. This is because mathematics, including differential geometry, was stripped of the intuitive appeal and clarity of the earlier studies.

Shortly afterwards, the construction of a non-Archimedean totally ordered field by Levi-Civita in 1890s paved the way for a mathematically correct re-introduction of infinitesimals. Further progress eventually lead to Abraham Robinson's *hyperreal* framework, a modern foundational theory which introduces and treats infinitesimals to great effect with full mathematical rigor. Since then, Robinson's framework has been used in a foundational role both in mathematical analysis [**Keisler 1986**] and other branches of mathematics, such as differential equations [**Benoit 1997**], measure, probability and stochastic analysis [**Loeb 1975**], [**Herzberg 2013**], [**Albeverio et al. 2009**], asymptotic series [**Van den Berg 1987**], as well as mathematical economics [**Sun 2000**]. These developments have led both to new mathematical results and to new insights and simplified proofs of old results.

The goal of the present monograph is to restore the original infinitesimal approach in differential geometry, on the basis of true infinitesimals in Robinson's framework. In the context of vector fields and flows, the idea is to work with a *combinatorial* object called a hyperreal walk constructed by hyperfinite iteration. We then deduce the continuous flow as the real shadow of the said walk.

We address the book to a reader willing to benefit from both the intuitive clarity and mathematical rigor of this modern approach. Readers with prior experience in differential geometry will find new (or rather old but somewhat neglected since Weierstrass; see `http://u.math.biu.ac.il/~katzmik/infinitesimals.html`) insights on known concepts and arguments, while readers with experience in applications of the hyperreal framework will find a new and fascinating application in a thriving area of modern mathematics.

Part 1

True Infinitesimal Differential Geometry

Introduction

> … Lübsen defined the differential quotient first by means of the limit notion; but along side of this he placed (after the second edition) what he considered to be the *true infinitesimal calculus* – a mystical scheme of operating with infinitely small quantities. [**Klein 1908**, p. 217]

The classical notion of an infinitesimal generator of a flow on a manifold is a (classical) vector field whose relation to the actual flow is expressed via *integration*. A true infinitesimal framework allows one to re-develop the foundations of differential geometry in this area in such a way that the relation between the infinitesimal generator and the flow becomes more transparent.

The idea is to work with a *combinatorial* object called a hyperreal walk constructed by hyperfinite iteration. We then deduce the continuous flow as the real shadow of the said walk.

Thus, a vector field is defined via an infinitesimal displacement defined by a self-map F of the manifold itself. The flow is then obtained by *iteration* $F^{\circ N}$ (rather than *integration*), as in Euler's method. As a result, the proof of the invariance of a vector field under a flow becomes a consequence of basic set-theoretic facts such as the associativity of composition, or more precisely the commutation relation $F \circ F^{\circ N} = F^{\circ N} \circ F$.

Furthermore, there are *synthetic*, or *combinatorial*, conditions D^k for the regularity of a (pre)vector field, replacing the usual *analytic* conditions of C^k type, that guarantee the usual theorems such as uniqueness and existence of solution of ODE locally, Frobenius theorem, Lie bracket, and other concepts.

A pioneering work in applying true infinitesimals to treat differential geometry of curves and surfaces is [**Stroyan 1977**]. Our approach is different because Stroyan's point of departure is the analytic class C^1 of vector fields, whereas we use purely combinatorial *synthetic* conditions D^1 (and D^2) in place of C^1 (and C^2). Similarly, the C^k conditions were taken as the point of departure in the following books: [**Stroyan & Luxemburg 1976**], [**Lutz & Goze 1981**], and [**Almeida, Neves & Stroyan 2014**].

To illustrate the advantages of this approach, we provide a proof of a case of the theorem of Frobenius on commuting flows. We also treat small oscillations of a pendulum (see Chapter 14), where the idea that the period of oscillations with infinitesimal amplitude should be independent of the amplitude finds precise mathematical expression.

One advantage of the hyperreal approach to solving a differential equation is that the hyperreal walk exists for all time, being defined combinatorially by iteration of a self-map of the manifold. The focus therefore shifts away from proving the *existence* of a solution, to establishing the *properties* of a solution.

Thus, our estimates show that given a uniform Lipschitz bound on the vector field, the hyperreal walk for all finite time stays in the finite part of ${}^*M$. If M is complete then the finite part is nearstandard. Our estimates then imply that the hyperreal walk for all finite time descends to a real flow on M.

0.1. Breakdown by chapters

In Chapters 3 and 4.7, we introduce the basic notions of an extended number system while avoiding excursions into mathematical logic that are not always accessible to students trained in today's undergraduate and graduate programs. We start with a syntactic account of a number system containing infinitesimals and infinite numbers, and progress to the relevant concepts such as filter, ultrapower construction, the extension principle (from sets and functions over the reals to their natural extensions over the hyperreals), and the transfer principle.

In Chapter 5.4, we present some typical notions and results from undergraduate calculus and analysis, such as continuity and uniform continuity, extreme value theorem, etc., from the point of view of the extended number system, introduce the notion of internal set generalizing that of a natural extention of a real set, and give a related construction of the reals out of hyperrationals.

In Chapter 14 we apply the above to treat infinitesimal oscillations of the pendulum. In Section 1.1, we present the traditional approach to differentiable manifolds, vector fields, flows, and 1-parameter families of transformations. In Chapter 9.2, we present the traditional "A-track" (see section 0.2) approach to the invariance of a vector field under its flow and to the theorem of Frobenius on the commutation of flows. In Chapters 10.2 and 11.2, we present an approach to vectors and vector fields as infinitesimal displacements, and obtain bounds necessary for proving the existence of the flow locally. In Chapters 12.5 and 13.7 we approach flows as iteration of (pre)vector fields. Chapters 16 through 18 contain more advanced material on superstructures, the transfer principle, definability and conservativity. Chapters 20 through 24 contain a historical update on infinitesimal mathematics.

0.2. Historical remarks

Many histories of analysis are based on a default Weierstrassian foundation taken as a *primary point of reference*.

In contrast, the article [**Bair et al. 2013**] developed an approach to the history of analysis as evolving along separate, and sometimes competing, tracks.

These are:

- the A-track, based upon an Archimedean continuum; and
- the B-track, based upon what we refer to as a Bernoullian (i.e., infinitesimal-enriched) continuum.[1]

Historians often view the work in analysis from the 17th to the middle of the 19th century as rooted in a background notion of continuum that is not punctiform. This necessarily creates a tension with modern, punctiform theories of the continuum, be it the A-type set-theoretic continuum as developed by Cantor, Dedekind, Weierstrass, and others, or B-type continua as developed by [**Hewitt 1948**], [**Łoś 1955**], [**Robinson 1966**], and others. How can one escape a trap of presentism in interpreting the past from the viewpoint of set-theoretic foundations commonly accepted today, whether of type A or B?

A possible answer to this query resides in a distinction between procedure and ontology. In analyzing the work of Fermat, Leibniz, Euler, Cauchy, and other great mathematicians of the past, one must be careful to distinguish between

(1) its practical aspects, i.e., actual mathematical practice involving procedures and inferential moves, and,
(2) semantic aspects related to the actual set-theoretic justification of the entities such as points of the continuum, i.e., issues

[1]Scholars attribute the first systematic use of infinitesimals as a foundational concept to Johann Bernoulli. While Leibniz exploited both infinitesimal methods and "exhaustion" methods usually interpreted in the context of an Archimedean continuum, Bernoulli never wavered from the infinitesimal methodology. To note the fact of such systematic use by Bernoulli is not to say that Bernoulli's foundation is adequate, or that it could distinguish between manipulations with infinitesimals that produce only true results and those manipulations that can yield false results.

of the ontology of mathematical entities such as numbers or points.

In Chapters 20 through 24 we provide historical comments so as to connect the notions of infinitesimal analysis and differential geometry with their historical counterparts, with the proviso that the connection is meant in the *procedural* sense as explained above.

0.3. Preview of flows and infinitesimal generators

As discussed in Section 9.1, the infinitesimal generator X of a flow $\theta_t(p) = \theta(t,p)\colon \mathbb{R} \times M \to M$ on a differentiable manifold M is a (classical) vector field defined by the relation

$$X_p(f) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \big(f(\theta_{\Delta t}(p)) - f(p) \big)$$

satisfied by all functions f on M. Even though the term "infinitesimal generator" uses the adjective "infinitesimal", this traditional notion does not actually exploit any infinitesimals. The formalism is somewhat involved, as can be sensed already in the proof of the invariance of the infinitesimal generator under the flow it generates, as well as the proof of the special case of the Frobenius theorem on commuting flows in Section 10.1.

We will develop an alternative formalism where the infinitesimal generator of a flow is an *actual infinitesimal* (pre)vector field. As an application, we will present more transparent proofs of both the invariance of a vector field under its flow and the Frobenius theorem in this case.

CHAPTER 1

Differentiable manifolds

1.1. Definition of differentiable manifold

A n-dimensional manifold is a set M possessing additional properties (a formal definition appears below as Definition 1.1.2). Namely, M is assumed to be covered by a collection of subsets (called coordinate charts or neighborhoods), typically denoted A or B, and having the following properties. For each coordinate neighborhood A we have an injective map $u\colon A \to \mathbb{R}^n$ whose image

$$u(A) \subseteq \mathbb{R}^n$$

is an open set in $\mathbb{R}^n$. Thus, the coordinate neighborhood is a pair

$$(A, u).$$

The maps are required to satisfy the following compatibility condition. Let

$$u\colon A \to \mathbb{R}^n, \quad u = (u^i)_{i=1,\ldots,n},$$

and similarly

$$v\colon B \to \mathbb{R}^n, \quad v = (v^\alpha)_{\alpha=1,\ldots,n}$$

be a pair of coordinate neighborhoods. Whenever the overlap $A \cap B$ is nonempty, it has a nonempty image $v(A \cap B)$ in Euclidean space. Both $u(A)$ and $u(A \cap B)$, etc., are assumed to be open subsets of $\mathbb{R}^n$.

DEFINITION 1.1.1. Let v^{-1} be the inverse map from the image (in $\mathbb{R}^n$) of the injective map v back to M.

Restricting to the subset $v(A \cap B)$, we obtain a one-to-one map

$$u \circ v^{-1}\colon v(A \cap B) \to \mathbb{R}^n \tag{1.1.1}$$

from an open set $v(A \cap B) \subseteq \mathbb{R}^n$ to $\mathbb{R}^n$.

Similarly, the map $v \circ u^{-1}$ from the open set $u(A \cap B) \subseteq \mathbb{R}^n$ to $\mathbb{R}^n$ is one-to-one.

DEFINITION 1.1.2. A smooth n-dimensional manifold M is a union

$$M = \cup_{\alpha \in I} A_\alpha,$$

where I is an index set, together with maps $u_\alpha\colon A_\alpha \to \mathbb{R}^n$, satisfying the following compatibility condition: the map (1.1.1) is differentiable for all choices of coordinate neighborhoods $A = A_\alpha$ and $B = A_\beta$ (where $\alpha, \beta \in I$) as above.

Definition 1.1.3. The maps $u \circ v^{-1}$ are called the *transition maps*.[1] The collection of coordinate charts as above is called an *atlas* for the manifold M.

Definition 1.1.4. A 2-dimensional manifold is called a surface.

Note that we have not said anything yet about a topology on M.

Definition 1.1.5. The coordinate charts induce a topology on M by imposing the usual conditions:

(1) If $S \subseteq \mathbb{R}^n$ is an open set then $v^{-1}(S) \subseteq M$ is defined to be open;
(2) arbitrary unions of open sets in M are open;
(3) finite intersections of open sets are open.

Remark 1.1.6. We will usually assume that M is connected. Given the manifold structure as above, connectedness of M is equivalent to path-connectedness[2] of M.

Remark 1.1.7 (Metrizability). There are some pathological non-Hausdorff examples like two copies of $\mathbb{R}$ glued along an open halfline of $\mathbb{R}$. These satisfy the compatibility condtion of Definition 1.1.2. To rule out such examples, the simplest condition is that of metrizability; see e.g., Example 1.4.1, Theorem 1.5.2.

Identifying $A \cap B$ with a subset of $\mathbb{R}^n$ by means of the coordinates (u^i), we can think of the map v as given by n real-valued functions

$$v^\alpha(u^1, \ldots, u^n), \quad \alpha = 1, \ldots, n. \tag{1.1.2}$$

1.2. Hierarchy of smoothness

The manifold condition stated in Definition 1.1.2 can be stated as the requirement that the n real-valued functions $v^\alpha(u^1, \ldots, u^n)$ of (1.1.2) are all smooth.

Definition 1.2.1. The usual hierarchy of smoothness (of functions), denoted C^k (or C^∞, or C^{an}), in Euclidean space generalizes to manifolds as follows.

[1] funktsiot maavar
[2] kshir-mesila

(1) For $k = 1$ a manifold M is C^1 if and only if all n^2 partial derivatives
$$\frac{\partial v^\alpha}{\partial u^i}, \quad \alpha = 1, \ldots, n,\ i = 1, \ldots, n$$
exist and are continuous.

(2) The manifold M is C^2 if all n^3 second partial derivatives
$$\frac{\partial^2 v^\alpha}{\partial u^i \partial u^j}$$
exist and are continuous.

(3) The manifold M is C^k if all n^k the partial derivatives
$$\frac{\partial^k v^\alpha}{\partial u^{i_1} \cdots \partial u^{i_k}}$$
exist and are continuous.

(4) The manifold M is C^∞ if for each $k \in \mathbb{N}$, all partial derivatives
$$\frac{\partial^k v^\alpha}{\partial u^{i_1} \cdots \partial u^{i_k}}$$
exist.

(5) The manifold M is C^{an} if for each $k \in \mathbb{N}$, all partial derivatives
$$\frac{\partial^k v^\alpha}{\partial u^{i_1} \cdots \partial u^{i_k}}$$
are real analytic functions.

The last condition is of course the strongest one.

1.3. Open submanifolds, Cartesian products

The notion of open and closed set in M is inherited from Euclidean space via the coordinate charts (see Definition 1.1.5).

DEFINITION 1.3.1. An open subset $C \subseteq M$ of a manifold M is itself a manifold, called an open submanifold, with differentiable structure obtained by the restriction of the coordinate map $u = (u^i)$ of (A, u). The restriction will be denoted $u\restriction_{A \cap C}$.

Let $\mathrm{Mat} = \mathrm{Mat}_{n,n}(\mathbb{R})$ be the set of square matrices with real coefficients. This is identified with Euclidean space of dimension n^2, and is therefore a manifold.

THEOREM 1.3.2. *Define a subset $GL(n, \mathbb{R}) \subseteq Mat_{n,n}(\mathbb{R})$ by setting*
$$GL(n, \mathbb{R}) = \{X \in Mat_{n,n}(\mathbb{R}) \colon \det(X) \neq 0\}.$$
Then $GL(n, \mathbb{R})$ is an open submanifold.

Proof. The determinant function is a polynomial in the entries x_{ij} of the matrix X. Therefore it is a continuous function of the entries, which are the coordinates in $\mathbb{R}^{n^2}$. Thus $GL(n,\mathbb{R})$ is the inverse image of the open set $\mathbb{R}\setminus\{0\}$ under a continuous map, and is therefore an open set, hence a manifold with respect to the restricted atlas. □

Remark 1.3.3. The complement D of $GL(n,\mathbb{R})$ in $\mathrm{Mat}_{n,n}$ is the closed set consisting of matrices of zero determinant. The set D for $n \geq 2$ is not a manifold.

Theorem 1.3.4. *Let M and N be two differentiable manifolds of dimensions m and n. Then the Cartesian product $M \times N$ is a differentiable manifold of dimension $m+n$. The differentiable structure is defined by coordinate neighborhoods of the form $(A \times B, u \times v)$, where (A,u) is a coordinate chart on M, while (B,v) is a coordinate chart on N. Here the function $u \times v$ on $A \times B$ is defined by*

$$(u \times v)(a,b) = (u(a), v(b))$$

for all $a \in A, b \in B$.

1.4. Circle, tori

Theorem 1.4.1. *The circle $S^1 = \{(x,y) \in \mathbb{R}^2 \colon x^2+y^2=1\}$ is a manifold.*

Proof. Let us give an explicit atlas for the circle. Let $A^+ \subseteq S^1$ be the open upper halfcircle

$$A^+ = \{a = (x,y) \in S^1 \colon y > 0\}.$$

Consider the coordinate chart (A^+, u), namely

$$u \colon A^+ \to \mathbb{R}, \tag{1.4.1}$$

defined by setting $u(x,y) = x$ (projection to the x-axis).

The open lower halfcircle $A^- = \{a = (x,y) \in S^1 \colon y < 0\}$ also gives a coordinate chart (A^-, u) where the coordinate u is defined by the same formula (1.4.1). We similarly define the right halfcircle

$$B^+ = \{a = (x,y) \in S^1 \colon x > 0\},$$

yielding a coordinate chart (B^+, v) where $v(x,y) = y$, and similarly for B^-.

The transition function between A^+ and B^+ is calculated as follows. Note that in the overlap $A^+ \cap B^+$ one has both $x > 0$ and $y > 0$. Let us calculate the transition function $u \circ v^{-1}$. The map v^{-1} sends $y \in \mathbb{R}^1$ to the point $(\sqrt{1-y^2}, y) \in S^1$, and then the coordinate map u

sends $(\sqrt{1-y^2}, y)$ to the first coordinate $\sqrt{1-y^2} \in \mathbb{R}^1$. Thus the composed map $u \circ v^{-1}$ given by

$$y \mapsto \sqrt{1-y^2} \tag{1.4.2}$$

is the transition function in this case. Since function (1.4.2) is smooth for all $y \in (0,1)$, the circle is a C^∞ manifold of dimension 1 (modulo spelling out the remaining transition functions).

Finally we discuss the metrizability condition (see Remark 1.1.7). We define a distance function by setting

$$d(p,q) = \arccos\langle p, q\rangle \tag{1.4.3}$$

This gives a metric on S^1 having all the required properties. It follows that S^1 is metrizable. □

EXAMPLE 1.4.2. The torus $T^2 = S^1 \times S^1$ is a 2-dimensional manifold by Theorem 1.3.4. Similarly, the n-torus $T^n = S^1 \times \cdots \times S^1$ (product of n copies of the circle) is an n-dimensional manifold.

EXAMPLE 1.4.3. The unit sphere $S^n \subseteq \mathbb{R}^{n+1}$ admits an atlas similar to the case of the circle. The distance function is defined by the same formula (1.4.3).

1.5. Projective spaces

Another basic example of a manifold is the projective space, defined as follows. Let $X = \mathbb{R}^{n+1} \setminus \{0\}$ be the collection of $(n+1)$-tuples $x = (x^0, \ldots, x^n)$ distinct from the origin. Define an equivalence relation $\sim$ between $x, y \in X$ by setting $x \sim y$ if and only if there is a real number $t \neq 0$ such that $y = tx$, i.e.,

$$y^i = tx^i, \quad i = 0, \ldots, n.$$

Denote by $[x]$ the equivalence class of $x \in X$.

DEFINITION 1.5.1. the real projective space, $\mathbb{RP}^n$, is the collection of equivalence classes $[x]$, i.e.,

$$\mathbb{RP}^n = \{[x] \colon x \in X\}.$$

THEOREM 1.5.2. *The space $\mathbb{RP}^n$ is a smooth n-dimensional manifold.*

PROOF. To show that $\mathbb{RP}^n$ is a manifold, we need to exhibit an atlas. We define coordinate neighborhoods A_i, where $i = 0, \ldots, n$ by setting

$$A_i = \{[x] \colon x^i \neq 0\}. \tag{1.5.1}$$

We will now define the coordinate pair (A_i, u_i), where $u_i\colon A_i \to \mathbb{R}^n$, namely the corresponding coordinate chart. We can set

$$u_i(x) = \left(\frac{x^0}{x^i}, \ldots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \ldots, \frac{x^n}{x^i}\right)$$

since division by x^i is allowed in the neighorhood A_i by condition (1.5.1). The coordinate u_i is well-defined because if $x \sim y$ then $u_i(x) = u_i(y)$ by canceling out the t in the numerator and denominator.

Let us calculate the transition maps. We let $u = u_i$ and $v = u_j$, where we assume for simplicity that $i < j$. We wish to calculate the map $u \circ v^{-1}$ associated with $A_i \cap A_j$. Take a point

$$z = (z^0, \ldots, z^{n-1}) \in \mathbb{R}^n.$$

Since we work with the condition $x^j \neq 0$, we can rescale the homogeneous coordinates so that $x^j = 1$. Thus we can represent $v^{-1}(z)$ by the $(n+1)$-tuple

$$v^{-1}(z) = (z^0, \ldots, z^{j-1}, 1, z^j, \ldots, z^{n-1}). \tag{1.5.2}$$

Now we apply $u = u_i$ to (1.5.2), obtaining

$$u \circ v^{-1}(z) = \left(\frac{z^0}{z^i}, \ldots, \frac{z^{i-1}}{z^i}, \frac{z^{i+1}}{z^i}, \ldots \frac{z^{j-1}}{z^i}, \frac{1}{z^i}, \frac{z^{j+1}}{z^i}, \ldots, \frac{z^{n-1}}{z^i}\right) \tag{1.5.3}$$

All transition functions appearing in (1.5.3) are rational functions and are therefore smooth. Thus $\mathbb{RP}^n$ is a differentiable manifold.

Let us check the metrizability condition. For unit vectors p, q we set

$$d(p, q) = \arccos |\langle p, q\rangle|$$

For arbitrary p, q we use the formula

$$d(p, q) = \arccos \frac{|\langle p, q\rangle|}{|p|\,|q|}. \tag{1.5.4}$$

Note that p and $-p$ represent the same point in projective space. Formula (1.5.4) provides a metric on $\mathbb{RP}^n$ with all the required properties, showing that $\mathbb{RP}^n$ is metrizable. □

1.6. Derivations

Let M be a differentiable manifold as defined in Section 1.1. The tangent space, denoted T_pM, at a point $p \in M$ is intuitively the collection of all tangent vectors at the point p.[3]

[3] A preliminary notion of a tangent space are developed in introductory courses based on a Euclidean embedding of the manifold; see e.g., `http://u.math.biu.ac.il/~katzmik/egreglong.pdf` (course notes for 88-201).

In modern differential geometry, a tangent vector can be defined via derivations.

Definition 1.6.1. Let $p \in M$. Let

$$\mathbb{D}_p = \{f \colon f \in C^\infty\}$$

be the ring of C^∞ real-valued functions f defined in an (arbitrarily small) open neighborhood of $p \in M$.

Definition 1.6.2. The ring operations in $\mathbb{D}_p$ are pointwise multiplication fg and pointwise addition $f + g$, where we choose the intersection of the domains of f and g as the domain of the new function (respectively sum or product). Thus, we set $(fg)(x) = f(x)\,g(x)$ for all x where both functions are defined.

Choose local coordinates $(u^1, \ldots, u^n)$ near $p \in M$. The following is proved in multivariate calculus.

Theorem 1.6.3. *A partial derivative* $\frac{\partial}{\partial u^i}$ *at* p *is a linear form, or* 1-*form, denoted* $\frac{\partial}{\partial u^i} \colon \mathbb{D}_p \to \mathbb{R}$ *on the space* $\mathbb{D}_p$, *satisfying the Leibniz rule*

$$\left.\frac{\partial(fg)}{\partial u^i}\right|_p = \left.\frac{\partial f}{\partial u^i}\right|_p g(p) + f(p) \left.\frac{\partial g}{\partial u^i}\right|_p \tag{1.6.1}$$

for all $f, g \in \mathbb{D}_p$.

Formula (1.6.1) can be written briefly as

$$\tfrac{\partial}{\partial u^i}(fg) = \tfrac{\partial}{\partial u^i}(f)g + f\tfrac{\partial}{\partial u^i}(g)$$

keeping in mind that both sides are evaluated only at the point p (*not* in a neighborhood of the point). Formula (1.6.1) motivates the following more general definition of a derivation at $p \in M$.

Definition 1.6.4. A *derivation* X at the point $p \in M$ is a linear form

$$X \colon \mathbb{D}_p \to \mathbb{R}$$

on the space $\mathbb{D}_p$ satisfying the Leibniz rule:

$$X(fg) = X(f)g(p) + f(p)X(g) \tag{1.6.2}$$

for all $f, g \in \mathbb{D}_p$.

CHAPTER 2

Derivations, tangent and cotangent bundles

2.1. The space of derivations

The notion of a manifold M was defined in Section 1.1. Recall that a derivation X at $p \in M$ is a linear form on the space of smooth functions $\mathbb{D}_p$ such that X satisfies the Leibniz rule at p; see Section 1.6. It turns out that the space of derivations is spanned by partial derivatives. Namely, we have the following theorem.

Theorem 2.1.1. *Let M be an n-dimensional manifold. Let $p \in M$. Then the collection of all derivations at p is a vector space of dimension n, denoted T_pM, and called the tangent space of M at p.*

Proof in case $n = 1$. We will prove the result in the case $n = 1$. For example, one could think of the 1-dimensional manifold $M = \mathbb{R}$ with the standard smooth structure. Thus we have a single coordinate u in a neighborhood of a point $p \in M$ which can taken to be 0, i.e., $p = 0$. Let $X \colon \mathbb{D}_p \to \mathbb{R}$ be a derivation at p. We would like to show that X coincides with the derivative $\frac{d}{dx}$ or its multiple. We argue in 4 steps as follows.

(1) Consider the constant function $1 \in \mathbb{D}_p$. Let us determine $X(1)$. We have $X(1) = X(1 \cdot 1) = 2X(1)$ by the Leibniz rule. Therefore $X(1) = 0$. Similarly for any constant a we have $X(a) = aX(1) = 0$ by linearity of X.
(2) Now consider the monic polynomial $u = u^1$ of degree 1, viewed as a linear function $u \in \mathbb{D}_{p=0}$. We evaluate the derivation X at the element $u \in \mathbb{D}_p$ and set $c = X(u)$. Thus $c \in \mathbb{R}$.
(3) By the Taylor remainder formula, every $f \in \mathbb{D}_{p=0}$ can be written as follows:

$$f(u) = a + bu + g(u)u, \quad a, b \in \mathbb{R},$$

where g is smooth and $g(0) = 0$. Since $f'(0) = b$, we have by linearity and Leibniz rule

$$\begin{aligned} X(f) &= X(a + bu + g(u)u) \\ &= bX(u) + X(g)u(0) + g(0) \cdot c \\ &= bc + 0 + 0 \\ &= c\,\frac{d}{du}(f). \end{aligned}$$

(4) It follows from $X(f) = c\frac{d}{du}(f)$ that derivation X coincides with the derivation $c\frac{d}{du}$ for all input functions $f \in \mathbb{D}_p$. Hence the tangent space is 1-dimensional and spanned by the element $\frac{d}{du}$, proving the theorem in this case.

The case of general n is treated similarly using a Taylor formula with partial derivatives. □

2.2. Tangent bundle, sections of a bundle

Let M be a differentiable manifold. In Section 2.1 we defined the tangent space T_pM at $p \in M$ as the space of derivations at p.

Definition 2.2.1. As a set, the *tangent bundle*, denoted TM, of an n-dimensional manifold M is the disjoint union of all tangent spaces T_pM as p ranges through M, or in formulas:

$$TM = \bigcup_{p \in M} T_pM.$$

Theorem 2.2.2. *The tangent bundle TM of an n-dimensional manifold M has a natural structure of a manifold of dimension $2n$.*

Proof. We coordinatize TM locally using $2n$ coordinate functions as follows. By Theorem 2.1.1, a tangent vector v at a point p decomposes as $v = v^i \frac{\partial}{\partial u^i}$ (with respect to the Einstein summation convention).

We combine the *coordinates* $(u^1, \ldots, u^n)$ of a point $p \in M$, together with the *components* of tangent vectors $v \in T_pM$, with respect to the basis $(\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}, \ldots, \frac{\partial}{\partial u^n})$, namely $v = v^i \frac{\partial}{\partial u^i}$. The resulting string of coordinates

$$(u^1, \ldots, u^n, v^1, \ldots, v^n)$$

of the pair (p, v) parametrizes a neighborhood of TM. It can be checked that the transition functions are smooth, showing that TM is a $(2n)$-dimensional manifold. □

THEOREM 2.2.3. *The tangent bundle TS^1 of the circle S^1 is a 2-dimensional manifold that can be identified with an (infinite) cylinder $S^1 \times \mathbb{R}$.*

PROOF. We represent the circle as the set of complex numbers of unit length:

$$S^1 = \{e^{i\theta}\} \subseteq \mathbb{C}.$$

Recall that $e^{i(\theta+2\pi n)} = e^{i\theta}$ for all $n \in \mathbb{Z}$. We use the coordinate θ on the circle to express a tangent vector at a point $e^{i\theta} \in S^1$ as $c\frac{d}{d\theta}$ where $c \in \mathbb{R}$. Then the pair $(e^{i\theta}, c)$ gives a parametrisation for the tangent bundle of S^1. □

DEFINITION 2.2.4 (Canonical projection). Given the tangent bundle TM of a manifold M, let

$$\pi_M : TM \to M, \quad (p, v) \mapsto p$$

be the canonical projection "forgetting" the tangent vector v and keeping only its initial point p.

DEFINITION 2.2.5. [Section] In the language of the theory of bundles, a vector field X on M is a *section* of the tangent bundle. Recall that the latter is given by $\pi_M : TM \to M$. Namely, a vector field is a map $X : M \to TM$ satisfying the condition

$$\pi_M \circ X = \mathrm{Id}_M.$$

We will express a vector field more concretely in terms of local coordinates in see Section 2.3.

2.3. Vector fields

Consider a coordinate chart (A, u) in M where $u = (u^i)_{i=1,\dots,n}$. We have a basis $(\frac{\partial}{\partial u^i})$ for T_pM by Theorem 2.1.1. Therefore an arbitrary vector $X \in T_pM$ is a linear combination

$$X^i \frac{\partial}{\partial u^i},$$

for appropriate coefficients $X^i \in \mathbb{R}$ depending on the point p. Here we use the Einstein summation convention.

Recall that the vectors $\left(\frac{\partial}{\partial u^i}\right)$ form a basis for the tangent space at every point of $A \subseteq M$.

DEFINITION 2.3.1. A choice of component functions $X^i(u^1, \dots, u^n)$ in the neighborhood will define a *vector field*

$$X^i(u^1, \dots, u^n) \frac{\partial}{\partial u^i}$$

in the neighborhood A.

Here the components X^i are required to be of an appropriate differentiability type. In more detail, we have the following definition.

DEFINITION 2.3.2. (see [**Boothby 1986**, p. 117]) Let M be a C^∞ manifold. A vector field X of class C^r on M is a map assigning to each point p of M, a vector $X_p \in T_pM$ whose components (X^i) in any local coordinate (A, u) are functions of class C^r.

EXAMPLE 2.3.3. Let M be the Euclidean plane $\mathbb{R}^2$. Via obvious identifications the Euclidean norm in the (x, y)-plane leads naturally to a Euclidean norm $|\ |$ on the tangent space (i.e., tangent plane) at every point with respect to which both tangent vectors $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ are orthogonal and have unit norm:

$$\left|\frac{\partial}{\partial x}\right| = \left|\frac{\partial}{\partial y}\right| = 1.$$

Note that each of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ defines a global vector field on $\mathbb{R}^2$ (defined at every point of the plane). Any combination

$$X = f(x, y)\frac{\partial}{\partial x} + g(x, y)\frac{\partial}{\partial y}$$

is also a vector field in the plane, with $|X| = \sqrt{f^2 + g^2}$ since the basis $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ is orthonormal.

REMARK 2.3.4 (Representation by path). The derivation i.e., vector $\frac{\partial}{\partial x}$ at a point $p = (a, b) \in \mathbb{R}^2$ is represented by the path $\alpha(s) = (a + s, b)$, in the sense that

$$\forall f \in \mathbb{D}_p, \quad \frac{\partial}{\partial x} f = \tfrac{d}{ds}\big|_{s=0} f(\alpha(s)),$$

and we have $|\frac{\partial}{\partial x}| = |\alpha'| = 1$.

EXAMPLE 2.3.5. Similarly, the vector $\frac{\partial}{\partial y}$ at a point $p = (a, b)$ is represented by the path $\alpha(s) = (a, b + s)$.

2.4. Vector fields defined by polar coordinates

REMARK 2.4.1. Eventually we will develop the notion of a differential k-form, generalizing the notion of a 1-form. The 1-forms, also known as covectors, are dual to vectors.

Interesting examples of vector fields are provided by polar coordinates. These may be undefined at the origin, i.e., apriori only defined in the open submanifold $\mathbb{R}^2 \setminus \{0\}$ of $\mathbb{R}^2$. Here the vector $\frac{\partial}{\partial \theta}$ at a point with polar coordinates (r, θ) is represented by the path $\alpha(\theta) = (r\cos\theta, r\sin\theta)$ with derivative

$$\alpha'(\theta) = (-r\sin\theta, r\cos\theta) = r(-\sin\theta, \cos\theta)$$

and therefore $|\alpha'| = r$. Hence at the point with polar coordinates (r, θ), we have

$$\left|\frac{\partial}{\partial\theta}\right| = r.^{1} \tag{2.4.2}$$

Corollary 2.4.2. *The rescaled vector* $\frac{1}{r}\frac{\partial}{\partial\theta}$ *is of norm* 1.

2.5. Source, sink, circulation

In this section we will describe some illustrative examples of vector fields.

Example 2.5.1 (Zero of type source/sink). The vector field $\frac{\partial}{\partial r}$ in the plane is undefined at the origin, but $r\frac{\partial}{\partial r}$ has a continuous extension (C^0) which is a vector field vanishing at the origin.[2]

Definition 2.5.2. The vector field in the plane defined by $r\frac{\partial}{\partial r}$ is called a *source* while the opposite vector field $X = -r\frac{\partial}{\partial r}$ is called a *sink*.[3]

Note that the integral curves of a source flow from the origin and away from it, whereas the integral curves of a sink flow into the origin, and converge to it for large time. At a point $p \in \mathbb{R}^2$ with polar coordinates (r, θ) the sink is given by

$$X(p) = \begin{cases} -r\frac{\partial}{\partial r} & \text{if } p \neq 0 \\ 0 & \text{if } p = 0. \end{cases}$$

[1]An alternative argument can be given in terms of differentials. Since $dr^2 = dx^2 + dy^2$ by Pythagoras, we have $|dr| = 1$ as well. Meanwhile $\theta = \arctan\frac{y}{x}$ and therefore $d\theta = \frac{1}{1+(y/x)^2}d(y/x) = \frac{x^2}{y^2+x^2}\frac{xdy-ydx}{x^2} = \frac{xdy-ydx}{r^2}$. Hence

$$|d\theta| = \frac{|xdy - ydx|}{r^2} = \frac{r}{r^2} = \frac{1}{r}. \tag{2.4.2}$$

Thus $rd\theta$ is a unit covector. We therefore have an orthonormal basis $(dr, rd\theta)$ for the cotangent space. Since $d\theta\left(\frac{\partial}{\partial\theta}\right) = 1$, equation (2.4.2) implies (2.4.1). Therefore $\frac{1}{r}\frac{\partial}{\partial\theta}$ is a unit vector.

[2]Moreover it is equal to $x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}$ and hence smooth.

[3]Would these be makor and kior (with kaf)? Actually this is known as bor, not kior.

EXAMPLE 2.5.3 (Zero of type "circulation" in the plane). The vector $\frac{\partial}{\partial\theta}$ in the plane, viewed as a tangent vector at a point at distance r from the origin, tends to zero as r tends to 0, as is evident from (2.4.1). Therefore the vector field defined by $\frac{\partial}{\partial\theta}$ on $\mathbb{R}^2 \setminus \{0\}$ extends by continuity to the point $p = 0$.[4] Thus we obtain a continuous vector field $p \mapsto X(p) = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$ on $\mathbb{R}^2$ which vanishes at the origin:

$$X(p) = \begin{cases} \frac{\partial}{\partial\theta} & \text{if } p \neq 0 \\ 0 & \text{if } p = 0. \end{cases}$$

Such a vector field is sometimes described as having *circulation*[5] around the point 0. The integral curves of a circulation are circles around the origin.

REMARK 2.5.4. The zero of the vector field associated with small oscillations of the pendulum are of circulation type. These were studied recently in e.g., [**Kanovei et al. 2016**].

EXAMPLE 2.5.5 (Zero of type *circulation* on a sphere). Spherical coordinates (ρ, θ, φ) in $\mathbb{R}^3$ restrict to the unit sphere $S^2 \subseteq \mathbb{R}^3$ to give coordinates (θ, φ) on S^2. The north pole is defined by $\varphi = 0$. At this point, the angle θ is undefined but the vector field $\frac{\partial}{\partial\theta}$ can be extended by continuity as in the plane (see Example 2.5.3), and we obtain a zero of type "circulation". Similarly the south pole is defined by $\varphi = \pi$. Here $\frac{\partial}{\partial\theta}$ has a zero of type "circulation" but going clockwise (with respect to the natural orientation on the 2-sphere).

Thus the vector field $\frac{\partial}{\partial\theta}$ on the sphere has two zeros of circulation type, namely north and south poles.

2.6. Duality in linear algebra

Let V be a real vector space. We will assume all vector spaces to be finite dimensional unless stated otherwise.

EXAMPLE 2.6.1. Euclidean space $\mathbb{R}^n$ is a real vector space of dimension n.

EXAMPLE 2.6.2. The tangent plane T_pM of a regular surface M (see Definition 1.1.4) at a point $p \in M$ is a real vector space of dimension 2.

DEFINITION 2.6.3. A *linear form*, also called 1-*form*, ϕ on V is a linear functional from V to $\mathbb{R}$.

[4] Moreover it equals $-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$ and hence smooth.

[5] machzor, tzirkulatsia.

DEFINITION 2.6.4. The *dual* space of V, denoted V^*, is the space of all linear forms ϕ on V. Namely,

$$V^* = \{\phi \colon \phi \text{ is a 1-form on } V\}.$$

Evaluating ϕ at an element $x \in V$ produces a scalar $\phi(x) \in \mathbb{R}$.

DEFINITION 2.6.5. The natural pairing between V and V^* is a linear map

$$\langle\ ,\ \rangle \colon V \times V^* \to \mathbb{R},$$

defined by setting $\langle x, y\rangle = y(x)$, for all $x \in V$ and $y \in V^*$.

THEOREM 2.6.6. *If V admits a basis of vectors (x_i), then V^* admits a unique basis, called the* dual basis (y_j), *satisfying*

$$\langle x_i, y_j\rangle = \delta_{ij}, \tag{2.6.1}$$

for all $i, j = 1, \ldots, n$, where δ_{ij} is the Kronecker delta function.

EXAMPLE 2.6.7. The vectors $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ form a basis for the tangent plane T_pE of the Euclidean plane E at each point $p \in E$. The dual space is denoted T_p^* and called the cotangent plane.

DEFINITION 2.6.8. The basis dual to $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ is denoted (dx, dy). Thus (dx, dy) is a basis for the cotangent plane T_p^* at every point $p \in E$.

Polar coordinates will be dealt with in detail in Subsection 2.7. They provide helpful examples of vectors and 1-forms, as follows.

EXAMPLE 2.6.9. In polar coordinates, we have a basis $(\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta})$ for the tangent plane T_pE of the Euclidean plane E at each point $p \in E \setminus \{0\}$. The dual space T_p^* has a basis dual to $\left(\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right)$ and denoted $(dr, d\theta)$.

EXAMPLE 2.6.10. In polar coordinates, the 1-form

$$r\,dr$$

occurs frequently in calculus. This 1-form vanishes at the origin (defined by the condition $r = 0$), and gets "bigger and bigger" as we get further away from the origin, as discussed in Section 2.7.

2.7. Polar, cylindrical, and spherical coordinates

Polar coordinates[6] (r, θ) satisfy $r^2 = x^2 + y^2$ and $x = r\cos\theta$, $y = r\sin\theta$. In $\mathbb{R}^2 \setminus \{0\}$, one way of defining the ranges for the variables is to require

$$r > 0 \quad \text{and} \quad \theta \in [0, 2\pi).$$

[6]koordinatot koteviot

It is shown in elementary calculus that the area of a region D in the plane in polar coordinates is calculated using the area element

$$dA = r\,dr\,d\theta.$$

Thus, the area is expressed by the following integral:

$$\text{area}(D) = \int_D dA = \iint r dr d\theta.$$

Cylindrical coordinates in Euclidean 3-space are studied in vector calculus.

Definition 2.7.1. Cylindrical coordinates (koordinatot gliliot)

$$(r, \theta, z)$$

are a natural extension of the polar coordinates (r, θ) in the plane.

The volume of an open region D is calculated with respect to cylindrical coordinates using the volume element

$$dV = r\,dr\,d\theta\,dz.$$

Thus the volume of D can be expressed as follows:

$$\text{vol}(D) = \int_D dV = \iiint r dr\,d\theta\,dz.$$

Example 2.7.2. Find the volume of a right circular cone with height h and base a circle of radius b.

Spherical coordinates[7]

$$(\rho, \theta, \varphi)$$

in Euclidean 3-space are studied in vector calculus.

Definition 2.7.3. Spherical coordinates (ρ, θ, φ) are defined as follows. The coordinate ρ is the distance from the point to the origin, satisfying

$$\rho^2 = x^2 + y^2 + z^2,$$

or $\rho^2 = r^2 + z^2$, where $r^2 = x^2 + y^2$. If we project the point orthogonally to the (x, y)-plane, the polar coordinates of its image, (r, θ), satisfy $x = r\cos\theta$ and $y = r\sin\theta$.

The last coordinate φ of a point in $\mathbb{R}^3$ is the angle between the position vector of the point and the third basis vector $e_3 = (0, 0, 1)^t$ in 3-space. Thus

$$z = \rho\cos\varphi \quad \text{while} \quad r = \rho\sin\varphi\,.$$

[7]koordinatot kaduriot

Here the ranges of the coordinates are often chosen as follows:

$$0 \leq \rho, \quad \text{while} \quad 0 \leq \theta \leq 2\pi, \quad \text{and} \quad 0 \leq \varphi \leq \pi$$

(note the different upper bounds for θ and φ). Recall that the volume of a region $D \subseteq \mathbb{R}^3$ is calculated using a volume element of the form

$$dV = \rho^2 \sin\varphi \; d\rho \, d\theta \, d\,\varphi,$$

so that the volume of a region D is

$$\mathrm{vol}(D) = \int_D dV = \iiint_D \rho^2 \sin\varphi \; d\rho \, d\theta \, d\,\varphi\,.$$

EXAMPLE 2.7.4. Calculate the volume of the spherical shell between spheres of radius $\alpha > 0$ and $\beta \geq \alpha$.

Now consider a sphere S_ρ of radius $\rho = \beta$. The area of a spherical region on S_ρ is calculated using the area element

$$dA_{S_\rho} = \beta^2 \sin\varphi \; d\theta \, d\,\varphi\,.$$

Thus the area of a spherical region $D \subseteq S_\beta$ is given by the integral

$$\mathrm{area}(D) = \int_D dA_{S_\rho} = \iint \beta^2 \sin\varphi \; d\theta \, d\,\varphi\,.$$

EXAMPLE 2.7.5. Calculate the area of the spherical region on a sphere of radius β included in the first octant, (so that all three Cartesian coordinates are positive).

2.8. Cotangent space and cotangent bundle

Derivations were already discussed in Section 1.6. Recall that the tangent space T_pM at $p \in M$ is the space of derivations at p.

DEFINITION 2.8.1. The vector space dual to the tangent space T_p is called the *cotangent space*, and denoted T_p^*.

Thus an element of a tangent space is a vector, while an element of a cotangent space is called a 1-form, or a *covector*.

DEFINITION 2.8.2. As a set, the *cotangent bundle*, denoted T^*M, of an n-dimensional manifold M is the disjoint union of all cotangent spaces T_p^*M as p ranges through M, or in formulas:

$$T^*M = \bigcup_{p \in M} T_p^*M.$$

DEFINITION 2.8.3. The basis dual to the basis $(\frac{\partial}{\partial u^i})$ is denoted

$$(du^i), \quad i = 1, \ldots, n.$$

Thus each du^i is by definition a 1-form on T_p, or a cotangent vector (covector for short). We are therefore working with dual bases $\left(\frac{\partial}{\partial u^i}\right)$ for vectors, and (du^i) for covectors. The pairing as in (2.6.1) gives

$$\left\langle \frac{\partial}{\partial u^i}, du^j \right\rangle = du^j \left(\frac{\partial}{\partial u^i} \right) = \delta_i^j, \tag{2.8.1}$$

where δ_i^j is the Kronecker delta: $\delta_i^j = 1$ if $i = j$ and $\delta_i^j = 0$ if $i \neq j$.

EXAMPLE 2.8.4. Examples of 1-forms in the plane are dx, dy, dr, rdr, $d\theta$.

CHAPTER 3

Number systems and infinitesimals

3.1. Successive extensions $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, ${}^*\mathbb{R}$

The page `http://u.cs.biu.ac.il/~katzmik/88-826.html` gives a link to these course notes as well as exams from previous years.

Our reference for true infinitesimal calculus is Keisler's textbook [**Keisler 1986**], downloadable at
`http://www.math.wisc.edu/~keisler/calc.html`

We start by motivating the familiar sequence of extensions of number systems

$$\mathbb{N} \hookrightarrow \mathbb{Z} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{R}$$

in terms of their applications in arithmetic, algebra, and geometry.

Remark 3.1.1. Each successive extension is introduced for the purpose of solving problems, rather than enlarging the number system for its own sake. Thus, the extension $\mathbb{Q} \hookrightarrow \mathbb{R}$ enables one to express the length of the diagonal of the unit square and the area of the unit disc in our number system.

The familiar continuum $\mathbb{R}$ is an Archimedean continuum, in the sense that it satisfies the following Archimedean property.

Definition 3.1.2. An ordered field extending $\mathbb{N}$ is said to satisfy the *Archimedean property* if

$$(\forall \epsilon > 0)(\exists n \in \mathbb{N})\,[n\epsilon > 1]\,.$$

We will provisionally denote the real continuum by $\mathbb{A}$, where "A" stands for *Archimedean.* Thus we obtain a chain of extensions

$$\mathbb{N} \hookrightarrow \mathbb{Z} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{A},$$

as above, where $\mathbb{A}$ is complete. In each case one needs an enhanced ordered number system to solve an ever broader range of problems from algebra or geometry.

Remark 3.1.3. Each real number can be represented by an unending decimal. The idea of representing each number by an unending decimal is due to Simon Stevin in the 16th century; some historical remarks on Stevin and his numbers can be found in Section 20.1.

The next stage is the extension

$$\mathbb{N} \hookrightarrow \mathbb{Z} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{A} \hookrightarrow \mathbb{B},$$

where $\mathbb{B}$ is a *Bernoullian continuum* containing infinitesimals, defined as follows.

DEFINITION 3.1.4. A Bernoullian extension of $\mathbb{R}$ is any *proper* extension which is an ordered field.

REMARK 3.1.5. The A-track approach to the calculus in the spirit of Weierstrass and his followers exploits the complete Archimedean continuum exclusively. The B-track approach exploits an infinitesimal-enriched continuum. Both of these approaches to the calculus and analysis have been present throughout the history of analysis starting from the 17th century onward; see Part 3.

Any Bernoullian extension allows us to define infinitesimals and do interesting things with those. But things become really interesting if we assume the Transfer Principle (Section 3.8), and work in a true hyperreal field, defined as in Definition 3.3.4 below. We will provide some motivating comments for the Transfer Principle in Section 3.3.

3.2. Motivating discussion for infinitesimals

Infinitesimals can be motivated from three different angles: geometric, algebraic, and arithmetic/analytic.

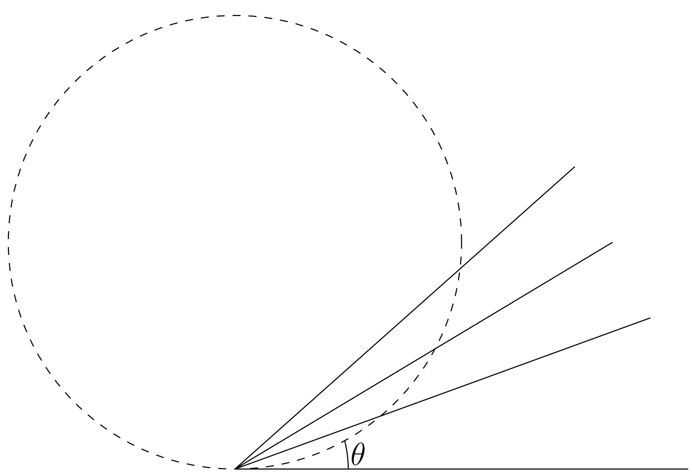

FIGURE 3.2.1. Hornangle θ is smaller than every rectilinear angle. Courtesy of Arkadius Kalka.

Remark 3.2.1 (Geometric approach (hornangles)). Some students have expressed the sentiment that they did not understand infinitesimals until they heard a geometric explanation of them in terms of what was classically known as hornangles. A hornangle is the crevice between a circle and its tangent line at the point of tangency. If one thinks of this crevice as a quantity, it is easy to convince oneself that it should be smaller than every rectilinear angle (see Figure 3.2.1). This is because a sufficiently small arc of the circle will be in the convex region cut out by the rectilinear angle no matter how small. One can define addition on these angles (e.g., the sum of two hornangles is the hornangle defined by the circle whose curvature is the sum of the curvatures of the original circles) with the result that a hornangle added to itself arbitrarily many times will still be smaller than any rectilinear angle. We cite this example merely as intuitive motivation (our actual construction of infinitesimals will be different).

Remark 3.2.2 (Algebraic approach (passage from ring to field)). The idea is to represent an infinitesimal by a sequence tending to zero. One can get something in this direction without reliance on any form of the axiom of choice. Namely, take the ring S of all sequences of real numbers, with arithmetic operations defined term-by-term. Now quotient the ring S by the equivalence relation that declares two sequences to be equivalent if they differ only on a finite set of indices. The resulting object S/K is a proper ring extension of $\mathbb{R}$, where $\mathbb{R}$ is embedded by means of the constant sequences. However, this object is not a field. For example, it has zero divisors. But quotienting it further in such a way as to get a field, by extending the ideal K to a *maximal* ideal K', produces a field S/K', namely a hyperreal field.

Remark 3.2.3 (Analytic/arithmetic approach). This approach is similar to 3.2.2 but with greater emphasis on analysis rather than algebra. One can mimick the construction of the reals out of the rationals as the set of equivalence classes of Cauchy sequences, and construct the hyperreals as equivalence classes of sequences of real numbers under an appropriate equivalence relation. This viewpoint is detailed in Section 5.1.

Some more technical comments on the definability of a hyperreal line can be found in Section 18.1.

3.3. Introduction to the transfer principle

The *transfer principle* is a type of theorem that, depending on the context, asserts that properties, rules, laws or procedures valid for a certain number system, still apply (i.e., are "transfered") to an extended number system.

EXAMPLE 3.3.1. The familiar extension $\mathbb{Q} \hookrightarrow \mathbb{R}$ preserves the property of being an ordered field.

EXAMPLE 3.3.2. To give a negative example, the frequently used extension $\mathbb{R} \hookrightarrow \mathbb{R} \cup \{\pm\infty\}$ of the real numbers to the so-called *extended reals* does not preserve the property of being an ordered field.

The hyperreal extension $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ (defined below) preserves *all* first-order properties (i.e., properties involving quantification over elements but not over sets; see Section 5.8 for a fuller discussion) of ordered fields.

EXAMPLE 3.3.3. The formula $\sin^2 x + \cos^2 x = 1$, true over $\mathbb{R}$ for all real x, remains valid over ${}^*\mathbb{R}$ for each hyperreal input x, including infinitesimal and infinite values of $x \in {}^*\mathbb{R}$.

Thus the transfer principle for the extension $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ is a theorem asserting that any statement true over $\mathbb{R}$ is similarly true over ${}^*\mathbb{R}$, and vice versa. Historically, the transfer principle has its roots in the procedures involving Leibniz's *Law of continuity*; see Section 21.3.

We will explain the transfer principle in several stages of increasing degree of abstraction. More details can be found in Sections 3.8, 5.5, 5.7, and chapter 16.

DEFINITION 3.3.4. An ordered field $\mathbb{B}$, properly including the field $\mathbb{A} = \mathbb{R}$ of real numbers (so that $\mathbb{A} \subsetneqq \mathbb{B}$) and satisfying the Transfer Principle, is called a hyperreal field.

Once such an extended field $\mathbb{B}$ is fixed, elements of $\mathbb{B}$ are called *hyperreal numbers*,[1] while the extended field itself is usually denoted ${}^*\mathbb{R}$.

THEOREM 3.3.5. *Hyperreal fields exist.*

For example, a hyperreal field can be constructed as the quotient of the ring $\mathbb{R}^{\mathbb{N}}$ of sequences of real numbers, by an appropriate maximal ideal; see Section 5.3.

3.4. Infinitesimals and infinite numbers

DEFINITION 3.4.1. A *positive infinitesimal* is a positive hyperreal number ϵ such that

$$(\forall n \in \mathbb{N})\,[n\epsilon < 1]\,.$$

[1]Similar terminology is used with regard to integers and hyperintegers; see Section 5.11.

Alternatively one could quantify over real n. More generally, we have the following.

DEFINITION 3.4.2. A hyperreal number ε is said to be *infinitely small* or *infinitesimal* if

$$-a < \varepsilon < a$$

for every positive real number a.

In particular, one has $\varepsilon < \frac{1}{2}$, $\varepsilon < \frac{1}{3}$, $\varepsilon < \frac{1}{4}$, $\varepsilon < \frac{1}{5}$, etc. If $\varepsilon > 0$ is infinitesimal then $H = \frac{1}{\varepsilon}$ is positive infinite, i.e., greater than every real number.

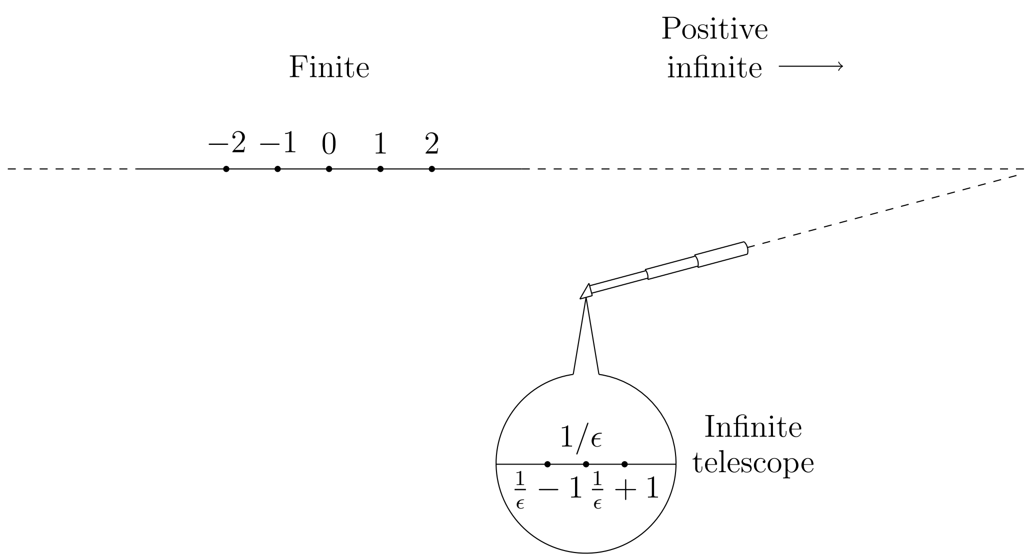

FIGURE 3.4.1. Keisler's telescope. Courtesy of Arkadius Kalka.

A hyperreal number that is not an infinite number is called *finite*; sometimes the term *limited* is used in place of *finite*.

3.5. Keisler's pictorial techniques

Keisler's textbook uses pictorial techniques.[2] Thus, it exploits the technique of representing the hyperreal line graphically by means of:

(1) dots indicating the separation between the finite realm and the infinite realm;
(2) One can view infinitesimals with microscopes as in Figure 4.1.1;
(3) One can also view infinite numbers with telescopes as in Figure 3.4.1.

[2]technikot tziuriot

3.6. Finite hyperreals

We have an important subset

$$\{\text{finite hyperreals}\} \subseteq {}^*\mathbb{R}$$

which is the domain of the function called the *standard part function* (also known as the *shadow*):

$$\mathbf{st}\colon \{\text{finite hyperreals}\} \to \mathbb{R}$$

which rounds off each finite hyperreal to the nearest real number. For details concerning the shadow see Section 4.1.

Example 3.6.1. Consider the slope calculation for $y = x^2$ at a point c. Here we exploit the standard part function (the *shadow*). If a curve is defined by $y = x^2$ we wish to find the slope at the point c. To this end we use an infinitesimal x-increment Δx and compute the corresponding y-increment $\Delta y = (c+\Delta x)^2 - c^2 = (c+\Delta x+c)(c+\Delta x-c) = (2c+\Delta x)\Delta x$. The corresponding "average" slope is therefore $\frac{\Delta y}{\Delta x} = 2c + \Delta x$ which is infinitely close to $2c$, and we are naturally led to the definition of the slope at c as the *shadow* of $\frac{\Delta y}{\Delta x}$, namely

$$\mathbf{st}\left(\frac{\Delta y}{\Delta x}\right) = 2c.$$

3.7. Extension principle

This section deals with the extension principle, which expresses the idea that all real objects have natural hyperreal counterparts. We will be mainly interested in *sets*, *functions* and *relations*. The *extension principle* asserts the following:

> The order relation on the hyperreal numbers extends the order relation on the real numbers. There exists a hyperreal number greater than zero but smaller than every positive real number. Every set $D \subseteq \mathbb{R}$ has a natural extension ${}^*D \subseteq {}^*\mathbb{R}$. Every real function f with domain D has a natural hyperreal extension ${}^*f$ with domain ${}^*D$.

Here "extension" means that

$${}^*f\lfloor_D = f.$$

We will also exploit extensions of arbitrary *relations*. Here the noun *principle* (in *extension principle*) means that we are going to assume that there is a function ${}^*f\colon \mathbb{B} \to \mathbb{B}$ which satisfies certain properties. It is a separate problem to define a continuum $\mathbb{B}$ which admits a coherent definition of ${}^*f$ for all functions $f\colon \mathbb{A} \to \mathbb{A}$. Such a problem will be solved in Section 5.3.

Here the *naturality* of the extension alludes to the fact that such an extension is unique, and the *coherence* refers to the fact that the domain of the natural extension of a function is the natural extension of its domain. We now present a more detailed version of Definition 3.4.2.

DEFINITION 3.7.1. The following three terms will be useful in the sequel:

(1) A *positive infinitesimal* is a positive hyperreal smaller than every positive real.
(2) A *negative infinitesimal* is a negative hyperreal greater than every negative real.
(3) An arbitrary *infinitesimal* is either a positive infinitesimal, a negative infinitesimal, or zero.

Ultimately it turns out counterproductive to employ asterisks for hyperreal functions; in fact we will drop them already in equation (3.8.1). See also Remark 5.8.4.

3.8. Transfer principle

DEFINITION 3.8.1. The Transfer Principle asserts that every first-order statement true over $\mathbb{R}$ is similarly true over ${}^*\mathbb{R}$, and vice versa.

REMARK 3.8.2. The adjective *first-order* alludes to the limitation on quantification to *numbers* as opposed to *sets of numbers*, as discussed in more detail in Section 5.8.

Listed below are a few examples of first-order statements.

EXAMPLE 3.8.3. The commutativity rule for addition $x+y=y+x$ is valid for all hyperreal x, y by the transfer principle.

EXAMPLE 3.8.4. The formula

$$\sin^2 x + \cos^2 x = 1 \tag{3.8.1}$$

is valid for all hyperreal x by the transfer principle.

EXAMPLE 3.8.5. The statement

$$0 < x < y \implies 0 < \frac{1}{y} < \frac{1}{x} \tag{3.8.2}$$

holds for all hyperreal x, y.

EXAMPLE 3.8.6. The indicator function $\chi_{\mathbb{Q}}$ of the rational numbers equals 1 on rational inputs and 0 on irrational inputs. By the transfer principle, its natural extension ${}^*\chi_{\mathbb{Q}} = \chi_{({}^*\mathbb{Q})}$ will be 1 on hyperrational numbers ${}^*\mathbb{Q}$ and 0 on hyperirrational numbers (namely, numbers in the complement ${}^*\mathbb{R} \setminus {}^*\mathbb{Q}$).

EXAMPLE 3.8.7. All *ordered field*-statements are subject to transfer. As we will see below, it is possible to extend transfer to a much broader category of statements, such as those containing the function symbols "exp" or "sin" or those that involve infinite sequences of reals.

We summarize the definitions that already appeared in Section 3.3.

DEFINITION 3.8.8. The following three terms will be useful in the sequel.

(1) A hyperreal number x is *finite* if there exists a real number r such that $|x| < r$.
(2) A hyperreal number is called *positive infinite* if it is greater than every real number;
(3) a hyperreal number *negative infinite* if it is smaller than every real number.

3.9. Three orders of magnitude

Hyperreal numbers come in three *orders of magnitude*: infinitesimal, appreciable, and infinite.

DEFINITION 3.9.1. A number is *appreciable* if it is finite but not infinitesimal.

Next, we will outline the rules for manipulating hyperreal numbers.

To give a typical proof, consider the rule that if ϵ is positive infinitesimal then $\frac{1}{\epsilon}$ is positive infinite. Indeed, for every positive real r we have $0 < \epsilon < r$. Now if r is real then $\frac{1}{r}$ is also real. It follows from (3.8.2) by transfer that $\frac{1}{\epsilon}$ is greater than every positive real, i.e., that $\frac{1}{\epsilon}$ is infinite.

3.10. Rules for manipulating hyperreals

Let ϵ, δ denote arbitrary infinitesimals. Let b, c denote arbitrary appreciable numbers. Let H, K denote arbitrary infinite numbers. We have the following theorem.

THEOREM 3.10.1. *We have the following rules for addition:*

- $\epsilon + \delta$ *is infinitesimal;*
- $b + \epsilon$ *is appreciable;*

- *$b + c$ is finite (possibly infinitesimal);*
- *$H + \epsilon$ and $H + b$ are infinite.*

We have the following rules for products.

- *$\epsilon\delta$ and $b\epsilon$ are infinitesimal;*
- *bc is appreciable;*
- *Hb and HK are infinite.*

We have the following rules for quotients.

- *$\frac{\epsilon}{b}$, $\frac{\epsilon}{H}$, $\frac{b}{H}$ are infinitesimal;*
- *$\frac{b}{c}$ is appreciable;*
- *$\frac{b}{\epsilon}$, $\frac{H}{\epsilon}$, $\frac{H}{b}$ are infinite provided $\epsilon \neq 0$.*

We have the following rules for roots, where n is a standard natural number.

- *if $\epsilon > 0$ then $\sqrt[n]{\epsilon}$ is infinitesimal;*
- *if $b > 0$ then $\sqrt[n]{b}$ is appreciable;*
- *if $H > 0$ then $\sqrt[n]{H}$ is infinite.*

REMARK 3.10.2. The traditional topic of the so-called "indeterminate forms" can be treated without introducing any ad-hoc terminology, by noting that we have *no rules* for the order of magnitude in certain cases, such as $\frac{\epsilon}{\delta}$, $\frac{H}{K}$, $H\epsilon$, and $H + K$. These cases correspond to what are known since François Napoléon Marie Moigno as *indeterminate forms.*

THEOREM 3.10.3. *Arithmetic operations on the hyperreal numbers are governed by the following rules.*

(1) *every hyperreal number between two infinitesimals is infinitesimal.*
(2) *Every hyperreal number which is between two finite hyperreal numbers, is finite.*
(3) *Every hyperreal number which is greater than some positive infinite number, is positive infinite.*
(4) *Every hyperreal number which is less than some negative infinite number, is negative infinite.*

EXAMPLE 3.10.4. The difference $\sqrt{H+1}-\sqrt{H-1}$ (where H is infinite) is infinitesimal. Namely,

$$\begin{aligned}\sqrt{H+1}-\sqrt{H-1} &= \frac{\left(\sqrt{H+1}-\sqrt{H-1}\right)\left(\sqrt{H+1}+\sqrt{H-1}\right)}{\left(\sqrt{H+1}+\sqrt{H-1}\right)}\\ &= \frac{H+1-(H-1)}{\left(\sqrt{H+1}+\sqrt{H-1}\right)}\\ &= \frac{2}{\sqrt{H+1}+\sqrt{H-1}}\end{aligned}$$

is infinitesimal. Once we introduce limits (see Section 4.3), this example can be reformulated as follows: $\lim_{n\to\infty}\left(\sqrt{n+1}-\sqrt{n-1}\right)=0$.

3.11. Two relations of proximity

DEFINITION 3.11.1. Two hyperreal numbers a, b are said to be infinitely close, written

$$a \approx b,$$

if their difference $a-b$ is infinitesimal.

It is convenient also to introduce the following terminology and notation. We will use Leibniz's notation $\ulcorner\!\urcorner$. Leibniz actually used a symbol that looks more like $\sqcap$ but the latter is commonly used to denote a product. Leibniz used the symbol to denote a generalized notion of equality "up to" a negligible term (though he did not distinguish it from the usual symbol "=" which he also used in the same sense). A prototype of such a relation (though not the notation) appeared already in Fermat under the name *adequality*. We will use it for a multiplicative relation among (pre)vectors.

DEFINITION 3.11.2. Two hyperreal numbers a, b are said to be *adequal*,[3] written

$$a \ulcorner\!\urcorner b,$$

if either $\frac{a}{b}\approx 1$ or $a=b=0$.

REMARK 3.11.3. The relation $\sin x \approx x$ for infinitesimal x is immediate from the continuity of sine at the origin (in fact both sides are infinitely close to 0), whereas the relation

$$\sin x \ulcorner\!\urcorner x$$

is a subtler relation equivalent to the computation of the first order Taylor approximation of sine.

[3]See Section 21.1 on Fermat.

CHAPTER 4

From infinitesimal calculus to ultrapower

4.1. Standard part principle

In addition to the extension principle, and the transfer principle, an important role in infinitesimal mathematics is played by the *standard part principle.*

THEOREM 4.1.1 (Standard Part Principle). *Every finite hyperreal number x is infinitely close to a unique real number.*

The proof will be given in Section 4.11.

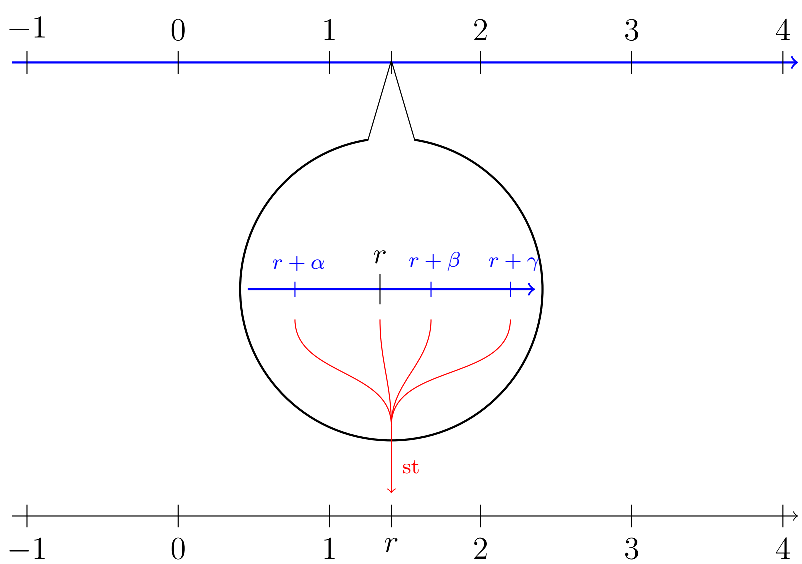

FIGURE 4.1.1. The standard part function, **st**, "rounds off" a finite hyperreal to the nearest real number. The function **st** is here represented by a vertical projection. Keisler's "infinitesimal microscope" is used to view an infinitesimal neighborhood of a standard real number r, where α, β, and γ represent typical infinitesimals. Courtesy of Wikipedia.

DEFINITION 4.1.2. The real number infinitely close to a finite hyperreal x is called the standard part, or *shadow*, denoted $\mathbf{st}(x)$, of x.

Thus we have $x \approx \mathbf{st}(x)$.

DEFINITION 4.1.3. The ring ${}^{\flat}\mathbb{R} \subseteq {}^{*}\mathbb{R}$ of finite hyperreals is the domain of the shadow $\mathbf{st}\colon {}^{\flat}\mathbb{R} \to \mathbb{R}$.

We will use the notation Δx, Δy for infinitesimals.

4.2. Differentiation

An infinitesimal increment Δx can be visualized graphically by means of a microscope as in the Figure 4.2.1.

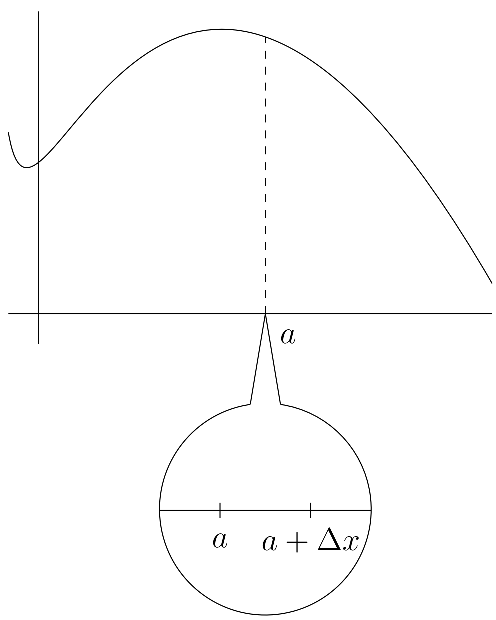

FIGURE 4.2.1. Infinitesimal increment Δx under the microscope. Courtesy of Arkadius Kalka.

DEFINITION 4.2.1. The *slope* s of a function f at a real point a is defined by setting

$$s = \mathbf{st}\left(\frac{f(a+\Delta x) - f(a)}{\Delta x}\right)$$

whenever the shadow exists (i.e., the ratio is finite) and is the same for each nonzero infinitesimal Δx. The construction is illustrated in Figure 4.2.2.

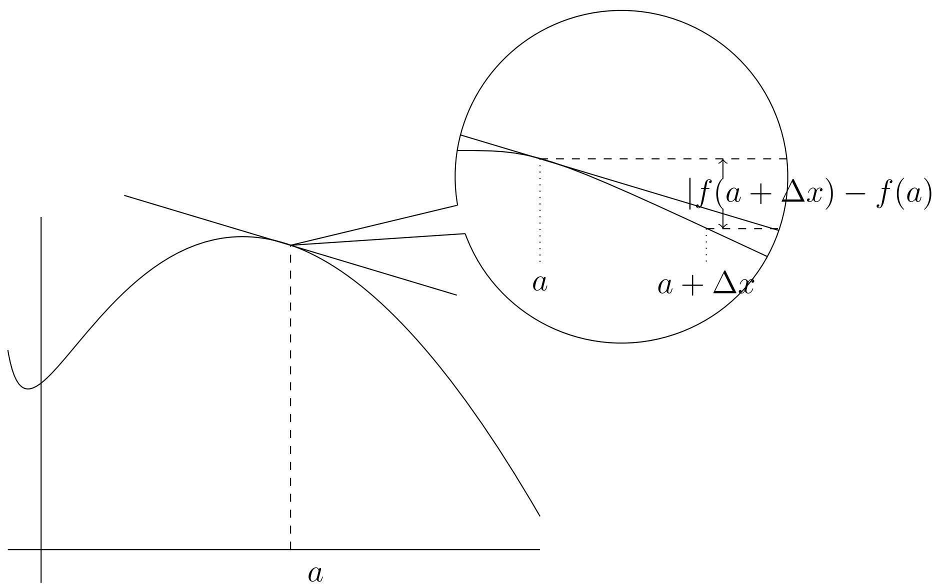

FIGURE 4.2.2. Defining slope of f at a. Courtesy of Arkadius Kalka.

DEFINITION 4.2.2. Let f be a real function of one real variable. The *derivative* of f is the new function f' whose value at a real point x is the slope of f at x. In symbols,

$$f'(x) = \mathbf{st}\left(\frac{f(x+\Delta x) - f(x)}{\Delta x}\right) \tag{4.2.1}$$

whenever the slope exists as specified in Definition 4.2.1.

Equivalently, we can characterize the real function f' using the relation $\approx$, by setting

$$f'(x) \approx \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

(wherever the slope exists, as before).

DEFINITION 4.2.3. When $y = f(x)$ we define a new dependent variable Δy by settting

$$\Delta y = f(x + \Delta x) - f(x)$$

called the y-increment.

Then we can write the derivative as $\mathbf{st}\left(\frac{\Delta y}{\Delta x}\right)$.

EXAMPLE 4.2.4. If $f(x) = x^2$ we obtain the derivative of $y = f(x)$ by the following direct calculation already performed in Example 3.6.1:

$$\begin{aligned} f'(x) \approx \frac{\Delta y}{\Delta x} &= \frac{(x+\Delta x)^2 - x^2}{\Delta x} \\ &= \frac{(x + \Delta x - x)(x + \Delta x + x)}{\Delta x} \\ &= \frac{\Delta x(2x + \Delta x)}{\Delta x} \\ &= 2x + \Delta x \\ &\approx 2x. \end{aligned}$$

4.3. Limit and standard part

DEFINITION 4.3.1. Let L and c be real numbers. We say that L is the *limit* of a function f at the point c, and write

$$\lim_{x \to c} f(x) = L,$$

if whenever x is infinitely close to c but different from c, the value $f(x)$ is infinitely close to L.

Equivalently, we have $\lim_{x\to c} f(x) = L$ if and only if each $x \approx c$, $x \neq c$, satisfies $\mathbf{st}(f(x)) = L$. Thus we can define the derivative by setting

$$f'(x) = \lim_{\Delta x \to 0} \left(\frac{f(x+\Delta x) - f(x)}{\Delta x} \right)$$

under the conditions specified in Definition 4.2.2.

One similarly defines infinite limits like $\lim_{x\to\infty} f(x)$, etc.

4.4. Differentials

Given a function $y = f(x)$ one defines the dependent variable $\Delta y = f(x + \Delta x) - f(x)$ as in Definition 4.2.3 above.

Definition 4.4.1. We define a new dependent variable dy by setting $dy = f'(x)\Delta x$ at a point where f is differentiable, and sets for symmetry $dx = \Delta x$.

Remark 4.4.2. We have the adequality

$$dy \sqcap \Delta y \;\text{ if }\; dy \neq 0$$

as in Definition 3.11.2.

Theorem 4.4.3. *Leibniz's notation* $\frac{dy}{dx}$ *is related to Lagrange's notation as follows:*

$$f'(x) = \frac{dy}{dx}$$

or equivalently $dy = f'(x)dx$.

In Leibniz's notation, rules such as the chain rule acquire an appealing form.

Remark 4.4.4. For a function $z = g(x, y)$ of two variables one can write $dz = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy$ with a similar meaning attached to the infinitesimal differentials dx, dy, and dz.

4.5. Second differences and second derivatives

Definition 4.5.1. Let $y = f(x)$. The *second difference* $\Delta^2 y = \Delta^2(x, y, \Delta x)$ is the dependent variable

$$\Delta^2 y = f(x) - 2f(x + \Delta x) + f(x + 2\Delta x).$$

Definition 4.5.2. The second derivative $f''(x)$ is defined by setting $f''(x) = g'(x)$ where $g(x) = f'(x)$.

If the derivative of the derivative of f is continuous then one can define the second derivative equivalently by setting $f''(x) = \mathbf{st}\big(\frac{\Delta^2 y}{\Delta x^2}\big)$; see Theorem 4.5.4.

As in the case of the first derivative, the second derivative is said to exist at x if the value is independent of the choice of Δx.

Definition 4.5.3. The dependent variable d^2y is $d^2y = f''(x)(dx)^2$.

Thus whenever $d^2y \neq 0$, one has $\Delta^2 y \sqcap d^2y$.

Theorem 4.5.4. *Assume that the derivative of the derivative of* f *is continuous. Then* $f''(x) = \mathbf{st}\big(\frac{\Delta^2 y}{\Delta x^2}\big)$.

PROOF OF EQUIVALENCE OF TWO DEFINITIONS. Since both sides of the claimed equality $(f'(x))' = \mathbf{st}\big(\frac{\Delta^2 y}{\Delta x^2}\big)$ are unaffected by addition of linear terms to the function f, we can assume without loss of generality that $f(x) = f'(x) = 0$ at a fixed point x.

Let $h = \Delta x$. We apply Taylor formula with remainder (generalizing the mean value theorem; see Theorem 7.4.1 below). Thus we have

$$f(x+h) = \frac{1}{2}h^2 f''(x+\vartheta h)$$

where $f''(x)$ is shorthand for $(f'(x))'$, for suitable $0 < \vartheta < 1$. Similarly, $f(x+2h) = \frac{1}{2}h^2 f''(x+\bar{\vartheta}\,2h)$. Then

$$\begin{aligned}\Delta^2 y &= f(x) - 2f(x+h) + f(x+2h)\\ &= f(x) - 2(\tfrac{1}{2})h^2 f''(x+\vartheta h) + \tfrac{1}{2}(2h)^2 f''(x+\bar{\vartheta}\,2h)\\ &= -h^2 f''(x+\vartheta h) + 2h^2 f''(x+\bar{\vartheta}\,2h)\end{aligned}$$

and therefore

$$\frac{\Delta^2 y}{\Delta x^2} = -f''(x+\vartheta h) + 2f''(x+\bar{\vartheta}\,2h) \approx f''(x)$$

by continuity of f''. □

4.6. Application: osculating circle of a curve, curvature

Second differences provide an intuitive approach to understanding the curvature of curves.

DEFINITION 4.6.1. Given an arclength parametrisation $\gamma(s)$ of a smooth curve in the plane, consider the shadow of the circle through three infinitely close points A, B, C on γ. This circle is called the osculating circle of γ at the point $P \in \gamma$ given by the common shadow of A, B, C.

DEFINITION 4.6.2. The center of curvature O at P of the curve γ is the center of the osculating circle to the curve at P.

DEFINITION 4.6.3. The curvature $k_\gamma(s)$ of the curve γ at a point $P = \gamma(s)$ is the reciprocal of the radius $|OP|$ of the osculating circle to the curve at P, i.e., $k_\gamma(s) = \frac{1}{|OP|}$.

THEOREM 4.6.4. *The curvature of the curve at a point P is the norm of the second derivative of $\gamma(s)$ at P.*

PROOF. Given a smooth regular plane curve in $\mathbb{R}^2$, consider a diamond formed by three consecutive infinitely close points A, B, C on the curve, such that $|AB| = |BC|$, together with a fourth point $B' \in \mathbb{R}^2$ symmetric to B with respect to the line AC. Thus $|AB| = |BC| =$

$|CB'| = |B'A|$. To construct such a diamond we intersect the curve with a circle of infinitesimal radius h centered at B, to produce points A and C with $|AB| = |BC| = h$. Then the center of curvature of the curve is the (shadow of the) center O of the circle passing through A, B, C. Let $R = \frac{1}{k}$ be the radius of this circle. By similarity of isosceles triangles $\triangle ABB' \sim \triangle OAB$, we obtain

$$\frac{|BB'|}{h} = \frac{h}{R}$$

where $h = |AB| = |BC|$ is the side of the diamond and $|BB'|$ is the length of the short diagonal. Therefore

$$\frac{|BB'|}{h^2} = \frac{1}{R} = k \tag{4.6.1}$$

is the curvature at the point. The expression (4.6.1) can be identified with the second derivative of the curve parametrized by arclength. Indeed, we have

$$BB' = (A - B) + (C - B) = A + C - 2B$$

and the second derivative, by Definition 4.5.2, is the shadow of

$$\frac{A + C - 2B}{h^2} = \frac{\Delta^2\gamma}{h^2}$$

where we used the fact that for an infinitesimal arc, the ratio of the arc AB (or BC) to the chord h is infinitely close to 1, i.e., the arc and the chord are *adequal* (see Section 3.11). The result now follows from Theorem 4.5.4 expressing the second derivative in terms of second differences. □

4.6.1. Epilogue: true infinitesimal calculus. This section is optional. The article [**Katz & Polev 2017**] describes a recent experience of teaching true infinitesimal calculus. The starting point of the article is the view that education and pedagogy are empirical sciences and therefore the effectiveness of this or that approach is most pertinently judged based on their classroom effect, rather than apriori partis pris.

Note the choice is not between infinitesimals and limits, since limits are present in both approaches (in the infinitesimal approach they are defined via the shadow; see Section 4.3. Rather, the choice is between infinitesimals and the *Epsilontik*. Moreover the goal is not to *replace* the ϵ, δ definitions by infinitesimal ones, but rather to use the latter to *prepare* for the former.

The infinitesimal approach was the main approach to analysis for several hundred years from Leibniz to Cauchy before the intuitive definitions were replaced by long-winded epsilontic paraphrases thereof.

Some mathematicians and following them also some historians tend to view the Weierstrassian approach as the benchmark from which other approaches are measured. That this is taken as a self-evident truth may be a reflection of a parti pris involving a belief in a *butterfly model* of development of mathematics. Philosopher Ian Hacking effectively challenges this belief in his recent book by contrasting such a model with a *Latin model*; see [**Hacking 2014**].

Hacking's distinction between the butterfly model and the Latin model involves the contrast between a model of a deterministic biological development of animals like butterflies, as opposed to a model of a contingent historical evolution of languages like Latin.

Hacking's dichotomy applies to the development of the field of mathematics as a whole. Some scholars view the development of mathematics as a type of organic process predetermined genetically from the start, even though the evolution of the field may undergo apparently sudden and dramatic changes, like the development of a butterfly which passes via a cocoon stage which is entirely unlike what it is pre-destined to produce in the end.

The Latin model acknowledges contingent factors in the development of an exact science (mathematics included), and envisions the possibility of other paths of development that may have been followed.

For example, had an axiomatic formalisation of infinitesimals been proposed earlier (e.g., by Du Bois-Reymond or other infinitesimalist of his generation, in conjunction with Frege and/or Peano), it might have been incorporated into the early formalisations of set theory, and spared us the verbal excesses of the Cantor–Russell opposition to infinitesimals, reflecting the state of affairs in mathematical foundations toward the end of the 19th century.

On such a view, there is no reason to view A-track Weierstrassian analysis as the benchmark by which infinitesimal analysis should be measured, and strengthens the Fermat–Leibniz–Euler–Cauchy–Robinson continuity in the development of infinitesimal analysis.

4.7. Introduction to the ultrapower

To motivate the material on filters contained in this chapter, we will first provide an outline of a construction of a hyperreal field ${}^*\mathbb{R}$ exploiting filters in this section. A more detailed technical presentation of the construction appears in Section 5.3.

Let $\mathbb{R}^{\mathbb{N}}$ denote the ring of sequences of real numbers, with arithmetic operations defined termwise. Then we will define ${}^*\mathbb{R}$ as

$$ {}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}} / \, \mathrm{MAX} \tag{4.7.1} $$

where $\mathrm{MAX} \subseteq \mathbb{R}^{\mathbb{N}}$ is an appropriate maximal ideal.

REMARK 4.7.1. What we wish to emphasize at this stage is the formal analogy between (4.7.1) and the construction of the real field as the quotient field of the ring of Cauchy sequences of rational numbers.

Note that in both cases, the subfield is embedded in the superfield by means of constant sequences. We will now describe a construction of such a maximal ideal.

REMARK 4.7.2. The idea is to define the ideal MAX as consisting of all "negligible" sequences $\langle u_n \colon n \in \mathbb{N} \rangle$, i.e., sequences which vanish for a set of indices of full measure 1; namely,

$$\xi\big(\{n \in \mathbb{N} \colon u_n = 0\}\big) = 1.$$

Let $\mathcal{P}(\mathbb{N})$ is the set of subsets of $\mathbb{N}$. Here a measure $\xi \colon \mathcal{P}(\mathbb{N}) \to \{0, 1\}$ has the following properties:

(1) ξ takes only two values, 0 and 1;
(2) ξ is a finitely additive measure;
(3) ξ takes the value 1 on each cofinite set.[1]

DEFINITION 4.7.3. The subset $\mathcal{F}_\xi \subseteq \mathcal{P}(\mathbb{N})$ consisting of sets of full measure 1 is called a free ultrafilter.

These originate with [**Tarski 1930**]. The construction of a Bernoullian continuum outlined above was therefore not available prior to that date. The construction outlined above is known as an ultrapower construction. The first construction of this type appeared in [**Hewitt 1948**], as did the term *hyper-real*.

REMARK 4.7.4. In the paragraph above, we motivated the construction in terms of finitely additive measures because the notion of a measure is more familiar to the broad mathematical public today than the notion of a filter. The two descriptions are in fact equivalent. The more detailed treatment below will rely on filters.

4.8. Introduction to filters

To present an infinitesimal-enriched continuum of hyperreals, we need some preliminaries on filters. Let I be a nonempty set (usually $\mathbb{N}$). The *power set* of I is the set

$$\mathcal{P}(I) = \{A \colon A \subseteq I\}$$

of all subsets of I.

[1]For each pair of complementary *infinite* subsets of $\mathbb{N}$, such a measure ξ *decides* in a coherent way which one is *negligible* (i.e., of measure 0) and which is *dominant* (measure 1).

Definition 4.8.1. A *filter* on I is a nonempty collection $\mathcal{F} \subseteq \mathcal{P}(I)$ of subsets of I satisfying the following axioms:

- Intersections: if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.
- Supersets: if $A \in \mathcal{F}$ and $A \subseteq B \subseteq I$, then $B \in \mathcal{F}$.

Thus to show that $B \in \mathcal{F}$, it suffices to show

$$A_1 \cap \cdots \cap A_n \subseteq B,$$

for some finite n and some $A_1, \ldots, A_n \in \mathcal{F}$.

Example 4.8.2. The full power set $\mathcal{P}(I)$ is itself a filter.

A filter $\mathcal{F}$ contains the empty set $\varnothing$ if and only if $\mathcal{F} = \mathcal{P}(I)$.

Definition 4.8.3. We say that a filter $\mathcal{F}$ is *proper* if $\varnothing \notin \mathcal{F}$.

Every filter contains I itself, and in fact the one-element set $\{I\}$ is the smallest filter on I.

Definition 4.8.4. An *ultrafilter* is a proper filter that satisfies the following additional property:

- for any $A \subseteq I$, either $A \in \mathcal{F}$ or $A^c \in \mathcal{F}$, where $A^c = I \setminus A$.

4.9. Examples of filters

We present several examples of filters.

Example 4.9.1 (Principal ultrafilter). Choose $i \in I$. Then the collection

$$\mathcal{F}^i = \{A \subseteq I : i \in A\}$$

is an ultrafilter, called the *principal ultrafilter* generated by i.

Remark 4.9.2. If I is finite, then every ultrafilter on I is of the form $\mathcal{F}^i$ for some $i \in I$, and so is principal.

Example 4.9.3 (Fréchet filter). The filter

$$\mathcal{F}^{Fre} = \{A \subseteq I : I \setminus A \text{ is finite}\}$$

is the cofinite, or Fréchet, filter on I, and is proper if and only if I is infinite. Note that $\mathcal{F}^{Fre}$ is not an ultrafilter.[2]

[2]More generally, one can consider a *filter generated by a collection.* If $\varnothing \neq \mathcal{H} \subseteq \mathcal{P}(I)$, then the filter generated by $\mathcal{H}$, i.e., the smallest filter on I including $\mathcal{H}$, is the collection $\mathcal{F}^{\mathcal{H}} = \{A \subseteq I : A \supset B_1 \cap \cdots \cap B_n$ for some n and some $B_i \in \mathcal{H}\}$ For $\mathcal{H} = \varnothing$ we set $\mathcal{F}^{\mathcal{H}} = \{I\}$. If $\mathcal{H}$ has a single member B, then $\mathcal{F}^{\mathcal{H}} = \{A \subseteq I : A \supset B\}$, which is called the principal filter generated by B. The ultrafilter $\mathcal{F}^i$ of Example 4.9.1 is the special case of this when $B = \{i\}$, namely a set containing a single element.

4.10. Properties of filters

The following results are immediate consequences of the definitions.

THEOREM 4.10.1. *An ultrafilter $\mathcal{F}$ has the following properties:*
(1) $A \cap B \in \mathcal{F}$ *if and only if* ($A \in \mathcal{F}$ and $B \in \mathcal{F}$);
(2) $A \cup B \in \mathcal{F}$ *if and only if* ($A \in \mathcal{F}$ or $B \in \mathcal{F}$);
(3) $A^c \in \mathcal{F}$ *if and only if* $A \notin \mathcal{F}$.

THEOREM 4.10.2. *Let $\mathcal{F}$ be an ultrafilter and $\{A_1, \ldots, A_n\}$ a finite collection of pairwise disjoint ($A_i \cap A_j = \varnothing$) sets such that*

$$A_1 \cup \ldots \cup A_n \in \mathcal{F}.$$

Then $A_i \in \mathcal{F}$ for exactly one i such that $1 \leq i \leq n$.

Recall that a free ultrafilter is an ultrafilter that is not principal.

THEOREM 4.10.3. *A nonprincipal ultrafilter must contain each cofinite set. Thus every free ultrafilter $\mathcal{F}$ includes the Fréchet filter $\mathcal{F}^{Fre}$, namely*

$$\mathcal{F}^{Fre} \subseteq \mathcal{F}.$$

This is a crucial property used in the construction of infinitesimals and infinitely large numbers; see e.g., Example 5.2.1.

REMARK 4.10.4. A filter $\mathcal{F}$ is an ultrafilter on I if and only if it is a *maximal* proper filter on I, i.e., a proper filter that cannot be extended to a larger proper filter on I.

DEFINITION 4.10.5. A collection $\mathcal{H} \subseteq \mathcal{P}(I)$ has the *finite intersection property* if the intersection of every nonempty finite subcollection of $\mathcal{H}$ is nonempty, i.e.,

$$B_1 \cap \cdots \cap B_n \neq \varnothing \text{ for any } n \text{ and any } B_1, \ldots, B_n \in \mathcal{H}.$$ [3]

4.11. Real continuum as quotient of sequences of rationals

To motivate the construction of the hyperreal numbers, we will first analyze the construction of the real numbers via Cauchy sequences. Let $\mathbb{Q}^{\mathbb{N}}$ denote the ring of sequences of rational numbers. Let

$$\mathbb{Q}^{\mathbb{N}}_C \subseteq \mathbb{Q}^{\mathbb{N}}$$

denote the subring consisting of Cauchy sequences.

[3] Note that the filter $\mathcal{F}^{\mathcal{H}}$ (see note 2) is proper if and only if the collection $\mathcal{H}$ has the finite intersection property.

Definition 4.11.1. The real field $\mathbb{R}$ is the quotient field

$$\mathbb{R} = \mathbb{Q}_C^{\mathbb{N}} \big/ \mathrm{MAX} \tag{4.11.1}$$

where $\mathrm{MAX} \subseteq \mathbb{Q}_C^{\mathbb{N}}$ is the maximal ideal consisting of null sequences (i.e., sequences tending to zero).

Note that $\mathbb{Q}_C^{\mathbb{N}}$ is only a ring, whereas the quotient (4.11.1) is a field.

The point we wish to emphasize is that a field extension is constructed starting with the base field and using *sequences* of elements in the base field.

An alternative construction of $\mathbb{R}$ from $\mathbb{Q}$ is via Dedekind cuts, as follows.

Definition 4.11.2 (Dedekind reals). A real number x is a pair $x = \{Q, Q'\}$ of two nonempty sets of rationals with the following three properties:

(1) the sets are complementary i.e., $Q \cup Q' = \mathbb{Q}$ and $Q \cap Q' = \varnothing$;
(2) $(\forall q \in Q)\,(\forall q' \in Q')\,[q < q']$;
(3) the "left" set Q does not have a maximal element.

Suppose $\mathbb{R}$ has already been constructed and let $x \in \mathbb{R}$. With respect to the natural order on $\mathbb{R}$, one can express the two sets $Q = Q_x$ and $Q' = Q'_x$ as follows: $Q_x = \{q \in \mathbb{Q} \colon q < x\}$ and $Q'_x = \{q \in \mathbb{Q} \colon q \geq x\}$. This will be exploited in Section 16.2.

The Dedekind reals provide a convenient framework for proving the standard part principle (Theorem 4.1.1), as follows.

Proof of standard part principle. The result holds generally for an arbitrary ordered field extension $\mathbb{R} \hookrightarrow E$. Indeed, let $x \in E$ be finite. If $x \approx r$ for $r \in \mathbb{Q}$ then we set $\mathbf{st}(x) = r$.

Now suppose x is not infinitely close to a rational number. Then x induces a Dedekind cut $\{Q_x, Q'_x\}$ on the subfield $\mathbb{Q} \subseteq \mathbb{R} \subseteq E$ via the total order of E, by setting $Q_x = \{q \in \mathbb{Q} \colon q < x\}$ and $Q'_x = \{q \in \mathbb{Q} \colon q \geq x\}$ as usual, where $Q'_x = \{q \in \mathbb{Q} \colon q > x\}$ since x is not infinitely close to a rational number.

The real number corresponding to the Dedekind cut is then infinitely close to x. □

CHAPTER 5

Hyperrationals, hyperreals, continuity

5.1. Extending $\mathbb{Q}$

Free ultrafilters $\mathcal{F}$ on $\mathbb{N}$ were defined in Section 4.10. We will think of a set $A \in \mathcal{F}$ in the ultrafilter as *dominant* and its complement $\mathbb{N} \setminus A$ as *negligible*.

Let $\mathbb{Q}^{\mathbb{N}}$ be the ring of sequences of rational numbers. We choose a free ultrafilter $\mathcal{F}$ on $\mathbb{N}$ and form an ideal $\text{MAX} = \text{MAX}_{\mathcal{F}} \subseteq \mathbb{Q}^{\mathbb{N}}$ as follows. Here MAX consisting of all sequences $\langle u_n : n \in \mathbb{N} \rangle \in \mathbb{Q}^{\mathbb{N}}$ such that

$$\{n \in \mathbb{N} : u_n = 0\} \in \mathcal{F}.$$

Definition 5.1.1 (Ultrapower construction). A field extension ${}^*\mathbb{Q}$ of $\mathbb{Q}$ is the quotient

$${}^*\mathbb{Q} = \mathbb{Q}^{\mathbb{N}} \big/ \text{MAX}.$$

Here we dropped the subscript $\mathcal{F}$ from MAX for simplicity.

Definition 5.1.2. The inclusion $\mathbb{Q} \hookrightarrow {}^*\mathbb{Q}$ is defined by identifying each $q \in \mathbb{Q}$ with the equivalence class of the constant sequence $\langle q, q, q, \ldots \rangle$.

Remark 5.1.3. This MAX is not the same maximal ideal as the one used in the construction of $\mathbb{R}$ from $\mathbb{Q}$ in Section 4.11. We employ similar notation to emphasize the similarity of the two constructions.

Definition 5.1.4. The equivalence class of a sequence $u = \langle u_n \rangle$ will be denoted $[u]$ or alternatively $[u_n] \in {}^*\mathbb{Q}$.

Theorem 5.1.5. *The field ${}^*\mathbb{Q}$ can be defined in an equivalent way as follows:*

$${}^*\mathbb{Q} = \left\{ [u] : u \in \mathbb{Q}^{\mathbb{N}} \right\},$$

where $[u]$ is the equivalence class of the sequence u relative to the ultrafilter $\mathcal{F}$.

Definition 5.1.6. The *order* on the field ${}^*\mathbb{Q}$ is defined by setting $[u] < [v]$ if and only if $\{n \in \mathbb{N} : u_n < v_n\} \in \mathcal{F}$.

5.2. Examples of infinitesimals

With respect to the construction presented in Section 5.1, we can now give some examples of infinitesimals.

PROPOSITION 5.2.1. *Let* $\alpha = \left[\frac{1}{n}\right]$ *i.e., the equivalence class of the sequence* $\langle \frac{1}{n} \colon n \in \mathbb{N}\rangle$*. Then* α *is smaller than* r *for every positive real number* $r > 0$*.*

PROOF. We have $\alpha = [u]$ where $u = \langle u_n \colon n \in \mathbb{N}\rangle$ is the null sequence (i.e., sequence tending to zero) $u_n = \frac{1}{n}$. Let $r \in \mathbb{R}$, $r > 0$. Consider the set

$$S = \{n \in \mathbb{N} \colon u_n < r\}.$$

Equivalently, we can write

$$S = \left\{n \in \mathbb{N} \colon \tfrac{1}{r} < n\right\}.$$

Since there are only finitely many integers that fail to satisfy the condition $\frac{1}{r} < n$, the set S is cofinite; more explicitly,

$$S = \left\{ \left\lceil \tfrac{1}{r} \right\rceil, \left\lceil \tfrac{1}{r} \right\rceil + 1, \left\lceil \tfrac{1}{r} \right\rceil + 2, \ldots \right\}.$$

Thus S is a member of the Fréchet filter $\mathcal{F}^{Fre}$ (see Section 4.9). But $\mathcal{F}^{Fre} \subseteq \mathcal{F}$ by Theorem 4.10.3 for each free ultrafilter $\mathcal{F}$, and in particular the one used in the ultrapower construction above. Hence $S \in \mathcal{F}$. Therefore $\alpha < r$ by definition of the order relation. □

EXAMPLE 5.2.2. The sequence

$$\left\langle \tfrac{(-1)^n}{n} \right\rangle \tag{5.2.1}$$

represents a nonzero infinitesimal in ${}^*\mathbb{Q}$, whose sign depends on whether or not the set $2\mathbb{N}$ of even natural numbers is a member of the ultrafilter. If $2\mathbb{N} \in \mathcal{F}$ then the sequence (5.2.1) is equivalent to $\left\langle \frac{1}{n} \right\rangle$ and therefore generates a positive infinitesimal.

5.3. Ultrapower construction of a hyperreal field

To obtain a full hyperreal field model of a B-continuum, we replace $\mathbb{Q}$ by $\mathbb{R}$ in the construction of Section 5.1, and form a similar quotient by the ideal $\mathrm{MAX} \subseteq \mathbb{R}^{\mathbb{N}}$ consisting of real sequences $\langle u_n \colon n \in \mathbb{N}\rangle \in \mathbb{R}^{\mathbb{N}}$ such that $\{n \in \mathbb{N} \colon u_n = 0\} \in \mathcal{F}$.

DEFINITION 5.3.1. We set

$${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}} / \mathrm{MAX} \tag{5.3.1}$$

where $\mathrm{MAX} = \mathrm{MAX}_{\mathcal{F}}$.

Equivalently,

$$ {}^*\mathbb{R} = \left\{ [u] \colon u \in \mathbb{R}^{\mathbb{N}} \right\}. $$

We wish to emphasize the analogy with formula (4.11.1) defining the A-continuum, and also a key difference: the basic ring is $\mathbb{R}^{\mathbb{N}}$ rather than $\mathbb{R}^{\mathbb{N}}_C$. In more detail, we proceed as follows.

(1) We define $\mathbb{R}^{\mathbb{N}}$ to be a ring with componentwise operations.
(2) We choose a nonprincipal ultrafilter $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$.
(3) We define $\mathrm{MAX}_{\mathcal{F}}$ to be the subset of $\mathbb{R}^{\mathbb{N}}$ consisting of real sequences $\langle u_n \rangle$ such that $\{n \in \mathbb{N} \colon u_n = 0\} \in \mathcal{F}$.
(4) We observe that $\mathrm{MAX}_{\mathcal{F}}$ is an ideal of the ring $\mathbb{R}^{\mathbb{N}}$.
(5) We observe that the quotient $\mathbb{R}^{\mathbb{N}} / \mathrm{MAX}_{\mathcal{F}}$ is a field and in fact an ordered field (see Definition 5.3.2).
(6) We denote the quotient by $\mathbb{R}^{\mathbb{N}} / \mathcal{F}$ to simplify notation.

Definition 5.3.2. A natural order relation ${}^*\!<$ on ${}^*\mathbb{R}$ is defined by setting

$$ [u_n] \;{}^*\!< [v_n] $$

if and only of the relation $<$ holds for a "dominant" set of indices, where "dominance" is determined by our fixed ultrafilter $\mathcal{F}$:

$$ [u_n] \;{}^*\!< [v_n] \quad \text{if and only if} \quad \{n \in \mathbb{N} \colon u_n < v_n\} \in \mathcal{F}. $$

Additional details on the ultrapower construction can be found e.g., in [**Davis 1977**].

Theorem 5.3.3. *The field $\mathbb{R}^{\mathbb{N}} / \mathcal{F}$ satisfies the transfer principle.*

We will present a detailed proof of the transfer principle in Chapters 16 and 17. See [**Chang & Keisler 1990**, Chapter 4], for a more general model theoretic study of ultrapowers and their applications.

5.4. Construction via equivalence relation

From now on, we will work with a variant of the construction (5.3.1). The variant is formulated in terms of an equivalence relation. This variant is more readily generalizable to other contexts (see Section 7.5).

Definition 5.4.1. We set ${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\!\sim$ where $\sim$ is an equivalence relation defined as follows:

$$ \langle u_n \rangle \sim \langle v_n \rangle \quad \text{if and only if} \quad \{n \in \mathbb{N} \colon u_n = v_n\} \in \mathcal{F}. $$

Since the relation $\sim$ depends on the choice of the free ultrafilter $\mathcal{F}$, so does the quotient $\mathbb{R}^{\mathbb{N}}/\!\sim$. For this reason the logicians, and they are definitely a special breed, decided to write directly

$$ \mathbb{R}^{\mathbb{N}} / \mathcal{F}, $$

as already noted in Section 5.3. But we will see that this abuse of notation offers some insights.

Definition 5.4.2. Subsets ${}^*\mathbb{Q}$ and ${}^*\mathbb{N}$ of the field ${}^*\mathbb{R}$ consist of $\mathcal{F}$-classes of sequences $\langle u_n \rangle$ with rational (respectively, natural) terms.

More generally, we have the following.

Definition 5.4.3. Given a subset $X \subseteq \mathbb{R}$, the subset ${}^*X$ of ${}^*\mathbb{R}$ consists of $\mathcal{F}$-classes of sequences $\langle u_n \rangle$ with terms $u_n \in X$ for each n.

Definition 5.4.4. If $X \subseteq \mathbb{R}$ and $f\colon X \to \mathbb{R}$, then the map

$${}^*f\colon {}^*X \to {}^*\mathbb{R}$$

sends each $\mathcal{F}$-class of a sequence $\langle u_n \rangle$ (where $u_n \in X$) to the $\mathcal{F}$-class of the sequence $\langle v_n \rangle$ where $v_n = f(u_n)$ for each n, or in formulas ${}^*f([\langle u_n \rangle]) = [\langle f(u_n) \rangle]$.

It is easy to check that this definition of ${}^*f$ is independent of the choices made in the construction.

As noted in Definition 5.3.2, the *order* ${}^*\!<$ on ${}^*\mathbb{R}$ is defined by setting $[u] \;{}^*\!<\; [v]$ if and only if $\{n \in \mathbb{N}\colon u_n < v_n\} \in \mathcal{F}$.

5.5. The ordered field language

Let the *ordered field language* be defined starting with a pair of operations $\cdot$ and $+$ and relation $<$, and enhanced as follows.

(1) We allow finite expressions like $(x + y) \cdot z + (x - 2y)$ (in free variables x, y, z, etc.) called *terms*;
(2) we can form further formulas like $T = T'$ and $T < T'$ where T, T' are terms;
(3) we allow more complex formulas by means of logical connectives[1] and quantifiers $\forall, \exists$;
(4) given a specific ordered field F, we allow the replacement of free variables in a formula by elements of F;
(5) this leads to the notion of a formula being *true* in F;
(6) this enables us to express and study various properties of the field F in a formal and well defined way.

5.6. Extending the language further

The ordered field language defined in Section 5.5 is not sufficient as a basis for either calculus or analysis. For instance there is no way to express phenomena related to non-algebraic functions like exp, sin, etc. Therefore we need to extend the ordered field language further.

[1] Kesher logi

Definition 5.6.1. We will use the term *extended real number language* to refer to the ordered field language extended further as follows:

(7) we can freely use expressions like $f(x)$, where $f\colon \mathbb{R} \to \mathbb{R}$ is any particular function;
(8) we can freely use formulas like $x \in D$, where D is any particular subset of $\mathbb{R}$;
(9) we can freely use any particular real numbers as parameters.

Remark 5.6.2. Returning to the fundamental idea of an extension $\mathbb{A} \hookrightarrow \mathbb{B}$ of an Archimedean continuum by a Bernoullian one, it may be helpful to point out that there is a triple of objects involved here:

(1) a symbol r in the language,
(2) the number $r_{\mathbb{A}}$ which is the interpretation of the symbol r in the Archimedean continuum, and
(3) the number $r_{\mathbb{B}}$ which is its interpretation in the Bernoullian continuum, where the star-transform of $r_{\mathbb{A}}$ is $r_{\mathbb{B}}$.

There is an important distinction here between syntax (i.e., the language itself) and semantics (referring to the model where the language is interpreted).

A more detailed treatment of formulas can be found for example in [**Loeb & Wolff 2015**]. We give a more precise version of the transfer principle as follows.

Theorem 5.6.3 (Transfer revisited). *Let A be a sentence of the* extended real number language*, and let ${}^*\!A$ be obtained from A by substituting ${}^*\!f$ and ${}^*\!D$ for each f or D which occur in A. Then A is true over $\mathbb{R}$ if and only if ${}^*\!A$ is true over ${}^*\mathbb{R}$.*

In Chapter 16 we will prove a generalisation of Theorem 5.6.3. In Remark 5.8.4 we will explain our approach to dropping the asterisks on functions.

Remark 5.6.4. In the statement of the theorem we wrote *true over* $\mathbb{R}$, or *over* ${}^*\mathbb{R}$, rather than the usual *true in* $\mathbb{R}$, to emphasize the fact that the sentences involved refer not only to real numbers themselves but also to sets of reals as well as real functions, namely, objects usually thought of as elements of a superstructure *over* $\mathbb{R}$; see Chapter 16 for more details.

Further analysis of the transfer principle appears in Section 5.7.

5.7. Upward and downward Transfer

Some motivating comments for the transfer principle already appeared in Section 3.3. Recall that by Theorem 5.3.3 we have an extension $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ of an Archimedean continuum by a Bernoullian one which furthermore satisfies transfer. To define ${}^*\mathbb{R}$, we fix a nonprincipal ultrafilter $\mathcal{F}$ over $\mathbb{N}$ and let ${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\mathcal{F}$. By construction, elements $[u]$ of ${}^*\mathbb{R}$ are represented by sequences $u = \langle u_n \colon n \in \mathbb{N}\rangle$ of real numbers, $u \in \mathbb{R}^{\mathbb{N}}$, with an appropriate equivalence relation defined in terms of $\mathcal{F}$ as in Section 5.3. Here an order is defined as follows. We have $[u] > 0$ if and only if $\{n \in \mathbb{N} \colon u_n > 0\} \in \mathcal{F}$.

Definition 5.7.1. The inclusion $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ is defined by sending a real number $r \in \mathbb{R}$ to the constant sequence $u_n = r$, i.e.,

$$\langle r, r, r, \ldots\rangle.$$

Since we view $\mathbb{R}$ as a subset of ${}^*\mathbb{R}$, we will denote the resulting (standard) hyperreal by the same symbol r.

The transfer principle was formulated in Theorem 5.6.3. To summarize, it asserts that truth over $\mathbb{R}$ is equivalent to truth over ${}^*\mathbb{R}$. Therefore there are two directions to transfer in.

Definition 5.7.2. Transfer of statements in the direction

$$\mathbb{R} \rightsquigarrow {}^*\mathbb{R}$$

is called *upward*, whereas transfer in the opposite direction

$$ {}^*\mathbb{R} \rightsquigarrow \mathbb{R}$$

is called *downward*.

5.8. Examples of first order statements

In this section we will illustrate the application of the transfer principle by several examples of transfer, applied to sentences familiar from calculus. The statements that transfer is applicable to are first-order quantified formulas, namely formulas involving quantification over field *elements* only, as in formulas of type $(\forall x \in \mathbb{R})$, etc. Quantification over all *subsets* of the field is disallowed.

Example 5.8.1. The completeness property of the reals is not transferable because its formulation involves quantification over all *subsets* of the field, as in $(\forall A \subseteq \mathbb{R})$, etc.

Example 5.8.2. Rational numbers $q > 0$ satisfy the following:

$$(\forall q \in \mathbb{Q}^+)(\exists n, m \in \mathbb{N}) \left[q = \frac{n}{m} \right],$$

where $\mathbb{Q}^+$ is the set of positive rationals. By upward transfer, we obtain the following statement, satisfied by all hyperrational numbers:

$$(\forall q \in {}^*\mathbb{Q}^+)(\exists n, m \in {}^*\mathbb{N}) \left[q = \frac{n}{m}\right]$$

where n, m may be infinite. Hypernatural numbers i.e., elements of ${}^*\mathbb{N}$, will be discussed in more detail in Section 5.11.

As discussed in Section 5.6, Transfer is applicable to functions, as well, in addition to the ordered field formulas.

DEFINITION 5.8.3. A function f is *continuous* at a real point $c \in \mathbb{R}$ if the following condition is satisfied:

$$(\forall \epsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall x \in \mathbb{R})\big[|x - c| < \delta \ \Rightarrow \ |f(x) - f(c)| < \epsilon\big].$$

By upward transfer, a continuous function f similarly satisfies the following formula over ${}^*\mathbb{R}$:

$$(\forall \epsilon \in {}^*\mathbb{R}^+)(\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*\mathbb{R})\big[|x - c| < \delta \ \Rightarrow \ |{}^*\!f(x) - {}^*\!f(c)| < \epsilon\big]. \tag{5.8.1}$$

Note that formula (5.8.1) is satisfied in particular for each positive infinitesimal $\epsilon \in {}^*\mathbb{R}$.

REMARK 5.8.4. In practice most calculations exploit the hyperreal extension of a function f, rather than f itself. We will therefore continue on occasion to denote by f the extended hyperreal function, as well.

DEFINITION 5.8.5. Let $D \subseteq \mathbb{R}$. A real function f is continuous in D if the following condition is satisfied:

$$\begin{aligned} &\underline{(\forall x \in D)(\forall \epsilon \in \mathbb{R}^+)}(\exists \delta \in \mathbb{R}^+)(\forall x' \in D) \\ &\qquad \big[|x' - x| < \delta \ \Rightarrow \ |f(x') - f(x)| < \epsilon\big], \end{aligned}$$

where we underlined the first pair of quantifiers for later purposes.

Switching the underlined pair of quantifiers, one obtains an equivalent formula

$$\begin{aligned} &(\forall \epsilon \in \mathbb{R}^+)(\forall x \in D)(\exists \delta \in \mathbb{R}^+)(\forall x' \in D) \\ &\qquad \big[|x' - x| < \delta \ \Rightarrow \ |f(x') - f(x)| < \epsilon\big]. \end{aligned} \tag{5.8.2}$$

REMARK 5.8.6. Switching the second and third quantifiers in (5.8.2) produces an *inequivalent* formula. Namely, we obtain the definition of *uniform* continuity as in (5.9.1) below.

By upward transfer, the hyperreal extension ${}^*\!f$ of such a function will similarly satisfy the following formula over ${}^*\mathbb{R}$:

$$\begin{aligned}&(\forall\epsilon\in{}^*\mathbb{R}^+)(\forall x\in{}^*\!D)(\exists\delta\in{}^*\mathbb{R}^+)(\forall x'\in{}^*\!D)\\&\qquad\big[|x'-x|<\delta\ \Rightarrow\ |{}^*\!f(x')-{}^*\!f(x)|<\epsilon\big],\end{aligned}\tag{5.8.3}$$

where ${}^*\!D\subseteq{}^*\mathbb{R}$ is the natural extension of $D\subseteq\mathbb{R}$ (see Section 3.1).

5.9. Uniform continuity

Definition 5.9.1. A real function f is *uniformly continuous* in a domain $D\subseteq\mathbb{R}$ if the following condition is satisfied:

$$\begin{aligned}&(\forall\epsilon\in\mathbb{R}^+)(\exists\delta\in\mathbb{R}^+)(\forall x\in D)(\forall x'\in D)\\&\qquad\big[|x'-x|<\delta\ \Rightarrow\ |f(x')-f(x)|<\epsilon\big].\end{aligned}\tag{5.9.1}$$

By upward transfer, the hyperreal extension ${}^*\!f$ of such a function will similarly satisfy the following formula over ${}^*\mathbb{R}$:

$$\begin{aligned}&(\forall\epsilon\in{}^*\mathbb{R}^+)(\exists\delta\in{}^*\mathbb{R}^+)(\forall x\in{}^*\!D)(\forall x'\in{}^*\!D)\\&\qquad\big[|x'-x|<\delta\ \Rightarrow\ |{}^*\!f(x')-{}^*\!f(x)|<\epsilon\big].\end{aligned}$$

An alternative characterisation of uniform continuity, of reduced quantifier complexity, is presented in Section 6.3.

5.10. Example of using downward transfer

To illustrate the use of *downward* transfer in a proof, consider the problem of showing that if f is differentiable at $c\in\mathbb{R}$ and $f'(c)>0$ then there is a point $x\in\mathbb{R}, x>c$ such that $f(x)>f(c)$. This proof was proposed by a student in the freshman course 89132 at Bar Ilan University.

Indeed, for infinitesimal $\alpha>0$ we have $\mathrm{st}\left(\frac{{}^*\!f(c+\alpha)-{}^*\!f(c)}{\alpha}\right)>0$ as in (4.2.1). Thus the quotient $\frac{{}^*\!f(c+\alpha)-{}^*\!f(c)}{\alpha}$ itself is appreciable and positive. Hence we have ${}^*\!f(c+\alpha)>{}^*\!f(c)$. Setting $x=c+\alpha$, we see that the following formula holds over the hyperreal field ${}^*\mathbb{R}$ as witnessed by this particular x:

$$(\exists x>c)[{}^*\!f(x)>{}^*\!f(c)].$$

We now apply downward transfer to obtain the formula

$$(\exists x>c)[f(x)>f(c)],\tag{5.10.1}$$

which holds over the real field $\mathbb{R}$. Namely, formula (5.10.1) asserts the existence of a real number $x>c$ such that $f(x)>f(c)$, as required.

Remark 5.10.1. Another example of downward transfer will be given in Section 6.2 following formula (6.2.3).

5.11. Dichotomy for hypernatural numbers

Let us consider the set of hypernatural numbers, ${}^*\mathbb{N}$ (positive hyperintegers) in more detail.

Definition 5.11.1. A hyperreal number x is called *finite* if $|x|$ is less than some real number. Equivalently, x is finite if $|x|$ is less than some natural number. A number is *infinite* if it is not finite.

Thus an infinite positive hypereal is a hyperreal number bigger than each real number.

Now consider the extension $\mathbb{N} \hookrightarrow {}^*\mathbb{N}$ constructed by an ultrapower as in Section 5.3, where ${}^*\mathbb{N} = \mathbb{N}^{\mathbb{N}}/\mathcal{F}$. Recall that we have the following definition.

Definition 5.11.2. A sequence $u \in \mathbb{N}^{\mathbb{N}}$ given by

$$u = \langle u_n \colon n \in \mathbb{N} \rangle$$

is said to be equivalent to a sequence $v \in \mathbb{N}^{\mathbb{N}}$ given by $v = \langle v_n \colon n \in \mathbb{N} \rangle$ if and only if the set of indices

$$\{n \in \mathbb{N} \colon u_n = v_n\}$$

is a member of a fixed ultrafilter $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$. We identify $\mathbb{N}$ with its image in ${}^*\mathbb{N}$ via the natural embedding generated by $n \mapsto \langle n, n, n, \ldots \rangle$.

Theorem 5.11.3. *Every finite hypernatural number is a natural number.*

Proof. Let $[u] \in {}^*\mathbb{N}$ where $u = \langle u_n \colon n \in \mathbb{N} \rangle$. Suppose $[u]$ is finite. Then there exists a real number

$$r > 0$$

such that $[u] < r$. This means that $\{n \in \mathbb{N} \colon u_n < r\} \in \mathcal{F}$. Modifying the sequence u on a negligible set of terms does not affect its equivalence class $[u]$. Therefore we can replace the remaining members of the sequence by 0 so that the condition $u_n < r$ is now satisfied for all $n \in \mathbb{N}$. Now consider the finite collection of integers

$$\big\{0, 1, 2, \ldots \lfloor r \rfloor \big\}.$$

For each index $n \in \mathbb{N}$, we have either $u_n = 0$, or $u_n = 1$, or $u_n = 2, \ldots$, or $u_n = \lfloor r \rfloor$. In other words, we represent $\mathbb{N}$ as a disjoint union of at most $\lfloor r \rfloor + 1$ sets, each of type $S_i = \{n \in \mathbb{N} \colon u_n = i\}$.

We now apply the defining property of an ultrafilter (see Theorem 4.10.2). It follows that exactly one of these sets, say

$$S_{i_0} = \{n \in \mathbb{N} \colon u_n = i_0\},$$

is dominant, i.e., it is a member of $\mathcal{F}$. Therefore $[u] = i_0$ and so $[u]$ is a natural number. □

Alternative proof. We provide an alternative proof exploiting the transfer principle instead of the special features of the ultrapower construction. Let $r \in \mathbb{R}^+$ be fixed. Note that natural numbers satisfy the formula

$$(\forall n \in \mathbb{N})\Big[n < r \quad \Rightarrow \quad n = 0 \ \vee n = 1 \ \vee \ n = 2 \ \vee \ \ldots \ \vee \ n = \lfloor r \rfloor\Big]$$

(of finite length dependent on r). Therefore by transfer, all hypernatural numbers satisfy the corresponding formula

$$(\forall n \in {}^*\mathbb{N})\big[n < r \quad \Rightarrow \quad n = 0 \ \vee n = 1 \ \vee \ n = 2 \ \vee \ \ldots \ \vee \ n = \lfloor r \rfloor\big].$$

Thus, if a hypernatural number $[u]$ is smaller than r then it must necessarily be one of the natural numbers $0, 1, 2, \ldots, \lfloor r \rfloor$. □

Corollary 5.11.4. *Every member of* ${}^*\mathbb{N}$ *is either a natural number or an infinite number.*

Proof. This is immediate from Theorem 5.11.3. □

CHAPTER 6

Galaxies, equivalence of definitions of continuity

6.1. Galaxies

Recall that, given a free ultrafilter $\mathcal{F}$ on $\mathbb{N}$, we set ${}^*\mathbb{N} = \mathbb{N}^{\mathbb{N}}/\text{MAX}$ where $u \in \text{MAX}$ if and only if $\{n \in \mathbb{N} \colon u_n = 0\} \in \mathcal{F}$. The hypernaturals ${}^*\mathbb{N}$ are partitioned into galaxies as follows.

Definition 6.1.1. An equivalence relation $\sim_g$ on ${}^*\mathbb{N}$ is defined as follows: for $n, m \in {}^*\mathbb{N}$, we set

$$n \sim_g m \;\text{ if and only if }\; |n - m| \text{ is finite.}$$

Definition 6.1.2. An equivalence class of the relation $\sim_g$ is called a *galaxy*. The galaxy of an element $x \in {}^*\mathbb{N}$ will be denoted $\text{gal}(x)$.[1]

Example 6.1.3. By Corollary 5.11.4, there is a unique galaxy containing finite numbers, namely the galaxy

$$\text{gal}(0) = \{0, 1, 2, \ldots\} = \mathbb{N},$$

where as before each element n of $\mathbb{N}$ is identified with the class in ${}^*\mathbb{N}$ of the constant sequence $\langle n, n, n, \ldots \rangle$.

Example 6.1.4. Each galaxy containing an infinite hyperinteger H is of the form

$$\text{gal}(H) = \{\ldots, H-3, H-2, H-1, H, H+1, H+2, H+3, \ldots\} \quad (6.1.1)$$

and is therefore order-isomorphic to $\mathbb{Z}$ (rather than $\mathbb{N}$).

Corollary 6.1.5. *The ordered set* $({}^*\mathbb{N}, <)$ *is not well-ordered.*

Indeed, it contains a copy of $\mathbb{Z}$ given by (6.1.1), which is not well-ordered.

[1]Sometimes galaxies are defined in the context of larger number systems, such as ${}^*\mathbb{R}$. Our main interest in galaxies lies in illustrating the concept of a non-internal set; see Section 7.7.

6.2. Equivalence of S-continuity and ϵ, δ continuity at c

Let $D = D_f \subseteq \mathbb{R}$ be the domain of a real function f.

Remark 6.2.1. It is a good exercise to prove that ${}^*(D_f) = D_{{}^*\!f}$. We will denote this set by ${}^*\!D_f$.

Definition 6.2.2. Let $x \in {}^*\!D_f$. We say that ${}^*\!f$ is *microcontinuous*[2] at x if whenever $x' \approx x$, one also has ${}^*\!f(x') \approx {}^*\!f(x)$ for x' in the domain ${}^*\!D_f$ of ${}^*\!f$. Here $x' \approx x$ means that $x' - x$ is infinitesimal.

Remark 6.2.3. This definition can be applied not only at a real point $x \in D_f$ but also at a hyperreal point $x \in {}^*\!D_f$; see Example 6.3.4 exploiting microcontinuity at an infinite point.

Theorem 6.2.4. *${}^*\!f$ is microcontinuous at a real point c if and only if* $\lim_{x\to c} f(x) = f(c)$.[3]

This was already discussed in Section 4.3.

Theorem 6.2.5. *Let $c \in \mathbb{R}$. A real function f is continuous in the ϵ, δ sense at c if and only if ${}^*\!f$ is microcontinuous at c.*

The result appeared in [**Robinson 1966**, Theorem 3.4.1] and is by now part of the logical toolkit of every mathematician working in Robinson's framework. We will follow [**Goldblatt 1998**, p. 76].

Proof of direction ($\Leftarrow$). Let $L = f(c)$. Recall that a real function f is continuous at c in the ϵ, δ sense if

$$(\forall \epsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall x \in D_f)\big[|x - c| < \delta \Rightarrow |f(x) - L| < \epsilon\big]. \quad (6.2.1)$$

Assume that ${}^*\!f$ is microcontinuous at c, so that

$$(\forall x \in {}^*\!D_f) \quad [x \approx c \;\Rightarrow\; {}^*\!f(x) \approx L]\,. \quad (6.2.2)$$

[2] In the literature the term *S-continuous* is sometimes used in place of *microcontinuous*.

[3] Microcontinuity at all points in a segment ("between two limits" i.e., endpoints) provides a useful modern proxy for Cauchy's definition of continuity in his 1821 text *Cours d'Analyse*: "the function $f(x)$ is continuous with respect to x between the given limits if, between these limits, an infinitely small increment in the variable always produces an infinitely small increment in the function itself." (translation from [**Bradley & Sandifer 2009**, p. 26]) In the current cultural climate it needs to be pointed out that Cauchy was obviously not familiar with our particular model of a B-continuum, based as it is on traditional set-theoretic *foundations*. On the other hand, his *procedures* and inferential moves find closer proxies in the context of modern infinitesimal-enriched continua than in the context of modern Archimedean continua; see Section 23.1. See Section 23.1 for further details on Cauchy's use of infinitesimals.

Let us prove that f is continuous in the sense of formula (6.2.1). Choose a real number $\epsilon > 0$ as in the leftmost quantifier in (6.2.1). Let $d > 0$ be infinitesimal. If $|x-c| < d$ then in particular $x \approx c$. By microcontinuity of ${}^*\!f$ we necessarily have

$$ {}^*\!f(x) \approx L, $$

and in particular $|{}^*\!f(x) - L| < \epsilon$ since ϵ is appreciable. Then the value $\delta = d$ is witness to the truth of the existence claim expressed by the formula

$$ (\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*\!D_f)\big[|x-c| < \delta \;\Rightarrow\; |{}^*\!f(x) - L| < \epsilon\big] \tag{6.2.3} $$

where our chosen ϵ is a fixed parameter in formula (6.2.3) (unlike formula (6.2.1) which quantifies over ϵ). We now apply *downward transfer* (see Section 5.7) to formula (6.2.3) to obtain

$$ (\exists \delta \in \mathbb{R}^+)(\forall x \in D_f)\big[|x-c| < \delta \;\Rightarrow\; |f(x) - L| < \epsilon\big]. $$

We conclude that there exists a real $\delta > 0$ as required, proving the direction ($\Leftarrow$).[4] □

Proof of direction ($\Rightarrow$). Conversely, assume that the ϵ, δ condition (6.2.1) holds. Let ϵ be a positive real number. Then by (6.2.1) there is an appropriate real number $\delta > 0$ such that the following sentence is true:

$$ (\forall x \in D_f)\ \big[|x-c| < \delta \Rightarrow |f(x) - L| < \epsilon\big]. \tag{6.2.4} $$

Applying upward transfer to formula (6.2.4), we obtain

$$ (\forall x \in {}^*\!D_f)\ \big[|x-c| < \delta \Rightarrow |{}^*\!f(x) - L| < \epsilon\big]. \tag{6.2.5} $$

But now whenever x satisfies $x \approx c$, the condition $|x - c| < \delta$ is automatically satisfied since $x-c$ is infinitesimal while δ is appreciable. It follows from (6.2.5) that if $x \approx c$ then the inequality $|{}^*\!f(x) - L| < \epsilon$ holds. This is true for all real numbers $\epsilon > 0$. Therefore ${}^*\!f(x) \approx L$, proving the relation (6.2.2) and the opposite implication ($\Rightarrow$). □

6.3. Uniform continuity in terms of microcontinuity

The ϵ, δ definition of uniform continuity appeared in Definition 5.9.1. We now introduce an equivalent definition in terms of microcontinuity. The equivalence is proved in Theorem 7.1.1.

---

[4] There is a shorter proof that exploits the overspill principle that will be elaborated later (see Section 7.8). Suppose f is S-continuous at c. Given a real $\epsilon > 0$, we define the set $A_\epsilon = \{\delta \colon (\forall x \in {}^*\!D_f)\,[\,|x-c| < \delta \implies |f(x) - f(c)| < \epsilon\,]\}$. The set is internal and contains all positive infinitesimals. By ovespill, A_ϵ must contain an appreciable value, as well, as required.

DEFINITION 6.3.1 (Definition via $\approx$). A real function f is *uniformly continuous* in its domain $D = D_f$ if and only if

$$(\forall x \in {}^*D_f)(\forall x' \in {}^*D_f)\,[x \approx x' \ \Rightarrow\ {}^*f(x) \approx {}^*f(x')]$$

where ${}^*f$ is its natural extension to the hyperreals.

Alternatively, one has the following definition via microcontinuity.

DEFINITION 6.3.2 (Alternative definition via $\approx$). A real function f is uniformly continuous in its domain D_f if

$$\text{for all } x \in {}^*D_f, \quad {}^*f \text{ is microcontinuous at } x.$$

REMARK 6.3.3. The above Definition 6.3.2 sounds startlingly similar to the definition of continuity itself. What is the difference between the two definitions? The point is that microcontinuity is now required at every point of the Bernoullian continuum rather than merely at the points of the Archimedean continuum, i.e., in the domain of ${}^*f$ which is the natural extension of the real domain of the real function f.

EXAMPLE 6.3.4. The function $f(x) = x^2$ fails to be uniformly continuous on its domain $D = \mathbb{R}$ because of the failure of microcontinuity of its natural extension ${}^*f$ at any single infinite hyperreal $H \in {}^*D_f = {}^*\mathbb{R}$ (cf. Definition 6.3.2). The failure of microcontinuity at H is checked as follows. Consider the infinitesimal $\epsilon = \frac{1}{H}$, and the point $H+\epsilon$ infinitely close to H. To show that ${}^*f$ is not microcontinuous at H, we calculate

$${}^*f(H+\epsilon) = (H+\epsilon)^2 = H^2 + 2H\epsilon + \epsilon^2 = H^2 + 2 + \epsilon^2 \approx H^2 + 2.$$

This value is not infinitely close to ${}^*f(H) = H^2$:

$${}^*f(H+\epsilon) - {}^*f(H) = 2 + \epsilon^2 \not\approx 0.$$

Therefore microcontinuity fails at the point $H \in {}^*\mathbb{R}$. Thus the squaring function $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$.

CHAPTER 7

Uniform continuity, EVT, internal sets

7.1. Equivalence of definitions of uniform continity

The definitions appeared in Section 6.3.

THEOREM 7.1.1. *Uniform continuity of f in its domain $D = D_f$ can be characterized in the following two equivalent ways:*

(1) *by means of the formula*

$$\begin{aligned}&(\forall \epsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)\\ &\quad \underline{(\forall x \in D)(\forall x' \in D)\ \big[|x'-x| < \delta \Rightarrow |f(x') - f(x)| < \epsilon\big]}\end{aligned} \tag{7.1.1}$$

(2) *and by means of the formula*

$$(\forall x \in {}^*\!D)(\forall x' \in {}^*\!D)\,[x \approx x' \ \Rightarrow\ {}^*\!f(x) \approx {}^*\!f(x')]\,. \tag{7.1.2}$$

PROOF OF ($\Rightarrow$). We first show that condition (7.1.1) implies (7.1.2). We treat ϵ and $\delta = \delta(\epsilon)$ as fixed real parameters. We then apply upward transfer to the underlined part of formula (7.1.1), we obtain

$$(\forall x \in {}^*\!D)(\forall x' \in {}^*\!D)\ \big[|x'-x| < \delta \Rightarrow |{}^*\!f(x') - {}^*\!f(x)| < \epsilon\big]. \tag{7.1.3}$$

This holds for each real $\epsilon > 0$, where real $\delta > 0$ was chosen as a function of ϵ. If $x \approx x'$ then the condition $|x - x'| < \delta$ is satisfied whatever the value of the real number $\delta > 0$. Formula (7.1.3) therefore implies

$$(\forall x \in {}^*\!D)(\forall x' \in {}^*\!D)\ \big[x \approx x' \Rightarrow |{}^*\!f(x') - {}^*\!f(x)| < \epsilon\big]. \tag{7.1.4}$$

Note that formula (7.1.4) is true for each real $\epsilon > 0$. We therefore conclude that ${}^*\!f(x') \approx {}^*\!f(x)$, proving (7.1.2). □

PROOF OF ($\Leftarrow$). We will show the contrapositive statement, namely that $\neg$(7.1.1) implies $\neg$(7.1.2). Assume the negation of (7.1.1). It follows that there exists a real number $\epsilon > 0$ such that

$$(\forall \delta \in \mathbb{R}^+)(\exists x \in D)(\exists x' \in D)\ \big[|x'-x| < \delta\ \wedge\ |f(x') - f(x)| > \epsilon\big]. \tag{7.1.5}$$

Applying upward transfer to formula (7.1.5) we obtain

$$(\forall \delta \in {}^*\mathbb{R}^+)(\exists x \in {}^*\!D)(\exists x' \in {}^*\!D)\ \big[|x'-x| < \delta\ \wedge\ |{}^*\!f(x') - {}^*\!f(x)| > \epsilon\big] \tag{7.1.6}$$

where ϵ is a fixed real parameter. The formula is true for all positive hyperreal δ. In particular it holds for an infinitesimal $\delta_0 > 0$. For this value, we obtain

$$(\exists x \in {}^*\!D)(\exists x' \in {}^*\!D)\ \big[|x' - x| < \delta_0 \ \wedge\ |{}^*\!f(x') - {}^*\!f(x)| > \epsilon\big]. \quad (7.1.7)$$

The condition $|x' - x| < \delta_0$ implies that $x \approx x'$. Therefore the formula (7.1.7) implies the following:

$$(\exists x \in {}^*\!D)(\exists x' \in {}^*\!D)\ \big[x' \approx x \ \wedge\ |{}^*\!f(x') - {}^*\!f(x)| > \epsilon\big].$$

Meanwhile the lower bound ϵ is real and therefore appreciable. It follows that ${}^*\!f(x') \not\approx {}^*\!f(x)$. Thus

$$(\exists x \in {}^*\!D)(\exists x' \in {}^*\!D)\big[x' \approx x \ \wedge\ {}^*\!f(x') \not\approx {}^*\!f(x)\big],$$

violating condition (7.1.2). This establishes the required contrapositive implication $\neg(7.1.1) \implies \neg(7.1.2)$. □

Remark 7.1.2. The term *microcontinuity* is exploited in the textbooks [**Davis 1977**] and [**Gordon et al. 2002**] in place of the term S-continuity. It reflects the existence of two definitions of continuity, one using infinitesimals, and one using epsilons. The former is what we refer to as microcontinuity. It is given a special name to distinguish it from the traditional definition of continuity. Note that microcontinuity at a nonstandard hyperreal does not correspond to any notion available in the epsilontic framework limited to an Archimedean continuum.

We give an additional example of the failure of uniform continuity seen from the viewpoint of Definition 6.3.2.

Example 7.1.3. Consider the function f given by $f(x) = \frac{1}{x}$ on the open interval $D = (0, 1) \subseteq \mathbb{R}$. Then ${}^*\!f$ fails to be microcontinuous at a positive infinitesimal. Indeed, choose an infinite hyperreal $H > 0$ and let $x = \frac{1}{H}$ and $x' = \frac{1}{H+1}$. Clearly $x \approx x'$. Both of these points are in the extended domain ${}^*\!D = {}^*(0, 1)$. Meanwhile,

$${}^*\!f(x') - {}^*\!f(x) = H + 1 - H = 1 \not\approx 0.$$

It follows from Theorem 7.1.1 that the real function f is not uniformly continuous on the open interval $(0, 1) \subseteq \mathbb{R}$.

7.2. Hyperreal extreme value theorem

First we clarify a notational point with regard to real and hyperreal intervals.

EXAMPLE 7.2.1. The unit interval $[0,1] \subseteq \mathbb{R}$ has a natural extension ${}^*[0,1] \subseteq {}^*\mathbb{R}$. Transfering the formula $[0,1] = \{x \in \mathbb{R}\colon 0 \leq x \leq 1\}$ we see that the segment ${}^*[0,1]$ contains all positive infinitesimals as well as all hyperreal numbers smaller than 1 and infinitely close to 1.

The extreme value theorem (EVT) is usually proved in two or more stages:

(1) one first shows that the function is bounded;
(2) then one proceeds to construct an extremum by one or another procedure involving choices of sequences.

The hyperreal approach is both more economical (there is no need to prove boundedness first) and less technical.

THEOREM 7.2.2 (EVT). *A continuous function f on $[0,1] \subseteq \mathbb{R}$ has a maximum.*

PROOF. Let $H \in {}^*\mathbb{N} \setminus \mathbb{N}$ be an infinite hypernatural number.[1] The real interval $[0,1]$ has a natural hyperreal extension

$$^*[0,1] = \{x \in {}^*\mathbb{R}\colon 0 \leq x \leq 1\}.$$

Consider its partition into H subintervals of equal infinitesimal length $\frac{1}{H}$, with partition points[2]

$$x_i = \tfrac{i}{H}, \quad i = 0, \ldots, H.$$

The function f has a natural extension ${}^*\!f$ defined on the hyperreals between 0 and 1. Note that in the real setting (when the number of partition points is finite), a point with the maximal value of f among the partition points x_i can always be chosen by induction.[3] By transfer, there is a hypernatural j such that $0 \leq j \leq H$ and

$$(\forall i \in {}^*\mathbb{N})\big[i \leq H \Rightarrow {}^*\!f(x_j) \geq {}^*\!f(x_i)\big]. \tag{7.2.2}$$

Consider the real point $c = \mathbf{st}(x_j)$ where $\mathbf{st}$ is the standard part function. Then $c \in [0,1]$ since non-strict inequalities are preserved under

[1] For instance, the one represented by the sequence $\langle 1,2,3,\ldots\rangle$ with respect to the ultrapower construction outlined in Section 5.3.

[2] The existence of such a partition follows by upward transfer (see Section 5.7) applied to the first order formula $(\forall n \in \mathbb{N})\ (\forall x \in [0,1])\ (\exists i < n)\ \left[\frac{i}{n} \leq x < \frac{i+1}{n}\right]$.

[3] We have the following first order property expressing the existence of a maximum of f over a finite collection:

$$(\forall n \in \mathbb{N})\ (\exists j \leq n)\ (\forall i \leq n)\ \left[f\left(\tfrac{j}{n}\right) \geq f\left(\tfrac{i}{n}\right)\right].$$

We now apply the transfer principle to obtain

$$(\forall n \in {}^*\mathbb{N})\ (\exists j \leq n)\ (\forall i \leq n)\ \left[{}^*\!f\left(\tfrac{j}{n}\right) \geq {}^*\!f\left(\tfrac{i}{n}\right)\right], \tag{7.2.1}$$

where ${}^*\mathbb{N}$ is the collection of hypernatural numbers. Formula (7.2.1) is true in particular for a specific infinite hypernatural value of n given by $H \in {}^*\mathbb{N} \setminus \mathbb{N}$.

passage to the standard part. Let $L = f(c)$. By microcontinuity of f at $c \in \mathbb{R}$, we have ${}^*\!f(x_j) \approx L$, i.e., $\mathbf{st}({}^*\!f(x_j)) = L$. Now consider an arbitrary real point $x \in [0,1]$. Then x lies in an appropriate sub-interval of the partition, namely $x \in [x_i, x_{i+1}]$, so that $\mathbf{st}(x_i) = x$, or $x_i \approx x$. Applying $\mathbf{st}$ to the inequality in the formula (7.2.2), we obtain

$$L = \mathbf{st}({}^*\!f(x_j)) \geq \mathbf{st}({}^*\!f(x_i)) = {}^*\!f(\mathbf{st}(x_i)) = {}^*\!f(x)$$

by microcontinuity at the point $x \in \mathbb{R}$. Hence $L \geq f(x)$, for all real x, proving c to be a maximum of f. □

Corollary 7.2.3. *The point c is a maximum of ${}^*\!f$ as well.*

This follows by upward transfer applied to the formula expressing the fact that c is a maximum of f.

Corollary 7.2.4 (Rolle's theorem). *A differentiable function on a compact interval with identical values at the endpoints has vanishing derivative at some interior point of the interval.*

Proof. By the extreme value theorem, f has a maximum c in the interval. We can assume that the maximum is in the interior by passing to $-f$ if necessary. Let $\epsilon > 0$ be infinitesimal. Since c is a maximum of ${}^*\!f$ by Corollary 7.2.3, we have $\frac{f(c+\epsilon)-f(c)}{\epsilon} \leq 0$ and $\frac{f(c-\epsilon)-f(c)}{-\epsilon} \geq 0$. Taking the shadow, we obtain both $f'(c) \leq 0$ and $f'(c) \geq 0$, as required. □

7.3. Intermediate value theorem

Theorem 7.3.1. *Let f be a continuous real function on $[0,1]$ and assume that $f(0)f(1) < 0$. Then there is a point $c \in [0,1]$ such that $f(c) = 0$.*

Proof. Let $H \in {}^*\mathbb{N} \setminus \mathbb{N}$ and consider the corresponding partition of ${}^*[0,1]$ with partition points $x_i = \frac{i}{H}$, $i = 0, \ldots, H$.

To fix ideas, assume that $f(0) < 0$ and $f(1) > 0$. Consider the set of all partition points x_i such that for all partition points before x_i, the function is nonpositive. By transfer this set has a *last* point x_j. By hypothesis

$$f(x_j) \leq 0 < f(x_{j+1}).$$

Applying standard part we obtain

$$\mathbf{st}(f(x_j)) \leq 0 \leq \mathbf{st}(f(x_{j+1})).$$

Let $c = \mathbf{st}(x_j)$. By microcontinuity at c, we have $\mathbf{st}(f(x_j)) = f(c) = \mathbf{st}(f(x_{j+1}))$. Therefore f vanishes at c. □

7.4. Mean Value Theorem

THEOREM 7.4.1 (Mean Value Theorem). *Let f be a differentiable function. Then*

$$(\forall x \in \mathbb{R})(\forall h \in \mathbb{R})(\exists \vartheta \in \mathbb{R})\big[f(x+h) - f(x) = h \cdot g(x + \vartheta h)\big]$$

where $0 < \vartheta < 1$ *and* $g(x) = f'(x)$.

Here differentiability of f is assumed on the appropriate interval $[x, x+h]$ where $h > 0$.

PROOF. The traditional proof passes via reduction (by subtracting a linear function of slope $\frac{f(x+h)-f(x)}{h}$) to Rolle's theorem (Theorem 7.2.4). Then the new function f satisfies the boundary condition $f(x) = f(x+h)$, and by Rolle's theorem, its derivative g satisfies $g(x + \vartheta h) = 0$ for a suitable ϑ, as in Section 7.2. □

7.5. Ultrapower construction applied to $\mathcal{P}(\mathbb{R})$

Consider the set of subsets of $\mathbb{R}$, denoted $\mathcal{P}(\mathbb{R})$. Thus saying that A is contained in $\mathcal{P}(\mathbb{R})$, i.e., $A \in \mathcal{P}(\mathbb{R})$ means that A is included in $\mathbb{R}$, i.e., $A \subseteq \mathbb{R}$. By the extension principle (see Section 3.3) we have the corresponding subset ${}^*\!A \subseteq {}^*\mathbb{R}$ called the natural extension of A. We would like to apply the ultrapower construction to the set $\mathcal{P}(\mathbb{R})$.

DEFINITION 7.5.1. It will be convenient to exploit the shorthand notation $\mathbb{P} = \mathcal{P}(\mathbb{R})$.

Then the natural extension ${}^*\mathbb{P}$ of $\mathbb{P}$ is constructed as before as a quotient of the set $\mathbb{P}^{\mathbb{N}}$ by an appropriate equivalence relation defined in terms of an ultrafilter using the dominant/negligible dichotomy, so that we have

$${}^*\mathbb{P} = \mathbb{P}^{\mathbb{N}}/\mathcal{F}$$

in the notation of Section 5.4.

The natural extensions of subsets of $\mathbb{R}$ constitute a useful family of subsets, but we will now construct an important *larger* class of subsets of ${}^*\mathbb{R}$ called *internal sets*. These are the members of ${}^*\mathbb{P}$.

DEFINITION 7.5.2. An element $\alpha \in {}^*\mathbb{P}$ is an $\mathcal{F}$-equivalence class $\alpha = [A]$ where A is a sequence $A = \langle A_n \in \mathbb{P} \colon n \in \mathbb{N}\rangle$ of elements of $\mathbb{P}$ (i.e., subsets of $\mathbb{R}$), where $A \sim B$ if and only if $\{n \in \mathbb{N} \colon A_n = B_n\} \in \mathcal{F}$.

We will provide further details in Section 7.6.

7.6. Internal sets

In more detail, we have the following definition. Let $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$ be the fixed free ultrafilter used in the construction of our hyperreal field.

DEFINITION 7.6.1 (Definition of internal set). Consider a sequence $A = \langle A_n \in \mathcal{P}(\mathbb{R})\colon n \in \mathbb{N}\rangle$ of subsets of $\mathbb{R}$. We use it to define a set $\alpha = [A] \subseteq {}^*\mathbb{R}$ as follows (by abuse of notation, the same symbols α and $[A]$ are used both for the equivalence class of sequences of sets, and for the subset of ${}^*\mathbb{R}$ that it represents). Namely, a hyperreal $[u] = [u_n]$ is an element of the set $\alpha = [A]$ if and only if one has

$$\{n \in \mathbb{N}\colon u_n \in A_n\} \in \mathcal{F}.$$

Such sets $\alpha \subseteq {}^*\mathbb{R}$ are called *internal*.

EXAMPLE 7.6.2. An example of a subset of ${}^*\mathbb{R}$ which is internal but is not a natural extension of any real set is the interval $[0, H]$ where H is an infinite hyperreal. To show that it is internal, represent H by a sequence $u = \langle u_n \colon n \in \mathbb{N}\rangle$, so that $H = [u]$. Then the set $\alpha = [0, H]$ is represented by the equivalence class $\alpha = [A]$ of the sequence $A = \langle A_n \colon n \in \mathbb{N}\rangle$ of real intervals $A_n = [0, u_n] \subseteq \mathbb{R}$.

7.7. The set of infinite hypernaturals is not internal

Does the construction of Section 7.5 give all possible subsets of ${}^*\mathbb{R}$? The answer turns out to be negative already for the hypernaturals, as we show in Theorem 7.7.4.

THEOREM 7.7.1. *Each internal subset of* ${}^*\mathbb{N}$ *has a least element.*

REMARK 7.7.2. In other words, ${}^*\mathbb{N}$ is *internally well-ordered* in the sense that the well-ordering property is satisfied if we only deal with internal sets.

PROOF. Assume that an internal α is given by $\alpha = [A]$ where $A = \langle A_n \rangle$. Here each set A_n can be assumed nonempty by Lemma 7.7.3 below. Since $\mathbb{N}$ is well-ordered we can choose a minimal element $u_n \in A_n$ for each n. Consider the sequence $u = \langle u_n \colon n \in \mathbb{N}\rangle$ formed by these minimal elements. Its equivalence class $[u]$ is the minimal element in the set $\alpha \in {}^*\mathbb{P}$. □

LEMMA 7.7.3. *A nonempty set* $\alpha = [A]$ *where* $A = \langle A_n \rangle$ *can always be represented by a sequence where each of the sets* A_n *in the sequence is nonempty.*

PROOF. The set of indices n for which $A_n = \varnothing$ is negligible (i.e., not a member of the ultrafilter) since otherwise α would be the empty set itself. For each n from this negligible set of indices n, we can replace the corresponding set $A_n = \varnothing$ by $A_n = \mathbb{N}$ without affecting the equivalence class $\alpha = [A]$. In the new sequence all the sets A_n are nonempty. □

Theorem 7.7.4. *The set of infinite hypernaturals,* ${}^*\mathbb{N} \setminus \mathbb{N}$*, is not an internal subset of* ${}^*\mathbb{N}$*.*

Proof. If the set ${}^*\mathbb{N} \setminus \mathbb{N}$ were internal it would have a least element by Theorem 7.7.1. But there is no minimal infinite hypernatural because if H is an infinite hypernatural then $H - 1$ is another infinite hypernatural. □

Corollary 7.7.5. *The subset* $\mathbb{N} \subseteq {}^*\mathbb{N}$ *is not internal.*

Proof. A subset of ${}^*\mathbb{N}$ is internal if and only if its complement is. Since ${}^*\mathbb{N} \setminus \mathbb{N}$ is not internal, the same is true of $\mathbb{N}$. □

7.8. Underspill

Definition 7.8.1 (Underspill). *Underspill* is the principle that every internal set including the set difference ${}^*\mathbb{N} \setminus \mathbb{N}$ necessarily contains also some elements that are finite natural numbers.

This principle will be exploited in Section 15.4.

7.9. Attempt to transfer the well-ordering of $\mathbb{N}$

Question 7.9.1. Is it possible to apply the transfer principle to the well-ordering property of the natural numbers, in such a way as to obtain a property of ${}^*\mathbb{N}$?

We showed in Section 7.7 that a nonempty *internal* subset of $\mathbb{N}$ has a least element. We will answer the more general question concerning subsets of $\mathbb{R}$ question in Section 7.10.

All the examples of transfer given in Section 5.7 deal with quantification over numbers, such as natural, rational, or real numbers.

As already pointed out, quantification over sets cannot be encompassed by the transfer principle. Thus, the least upper bound property for bounded sets in $\mathbb{R}$ fails when interpreted literally over ${}^*\mathbb{R}$, due to the following result.

Theorem 7.9.2. *The set of all infinitesimals in* ${}^*\mathbb{R}$ *does not admit a least upper bound.*

Proof. Suppose $C > 0$ were such a bound. Either C is infinitesimal or it is appreciable. If C were infinitesimal, then the infinitesimal $2C$ would be greater than C, contradicting the supposition that C

is a least upper bound. If C were appreciable, then $C/2$ would also be appreciable and therefore a *smaller* upper bound (for the set of infinitesimals) than C. The contradiction proves that there is no least upper bound. □

7.10. Quantification over internal sets; Henkin semantics

Quantification over sets in $\mathbb{N}$ can be transfered on condition of being interpreted as quantification over elements of $\mathbb{P} = \mathcal{P}(\mathbb{N})$, as follows.

EXAMPLE 7.10.1. The condition of well-ordering can be stated as follows. To simplify the formula we will use the symbol $\forall'$ to denote quantification over nonempty sets only:

$$(\forall' A \subseteq \mathbb{N})(\exists u \in A)(\forall x \in A)\big[u \leq x\big]. \tag{7.10.1}$$

To make this transferable, we replace the relation of inclusion $A \subseteq \mathbb{N}$ by the relation of containment $A \in \mathbb{P}$. We therefore reformulate condition (7.10.1) as follows:

$$(\forall A \in \mathbb{P} \setminus \{\varnothing\})(\exists u \in A)(\forall x \in A)\big[u \leq x\big],$$

where $\mathbb{P} = \mathcal{P}(\mathbb{N})$. At this point transfer can be applied, resulting in the following sentence:

$$\big(\forall A \in {}^*\mathbb{P} \setminus \{\varnothing\}\big)(\exists u \in A)(\forall x \in A)\big[u \leq x\big].$$

Here quantification is over elements of ${}^*\mathbb{P}$ (rather than over arbitrary subsets of $\mathbb{N}$).

In other words, quantification is over *internal* subsets of ${}^*\mathbb{N}$, resulting in a correct sentence. The crucial fact is the properness of the inclusion

$${}^*\mathcal{P}(\mathbb{N}) \subsetneqq \mathcal{P}({}^*\mathbb{N}).$$

REMARK 7.10.2. [**Robinson 1966**] refers to this approach to quantification as *Henkin semantics.*

All sentences involving quantification over subsets

$$A \subseteq \mathbb{R} \tag{7.10.2}$$

can similarly be transfered provided we interpret the quantification as ranging over

$$A \in \mathcal{P}(\mathbb{R}) \tag{7.10.3}$$

and transfering formula (7.10.3) instead of (7.10.2), to obtain

$$A \in {}^*\mathcal{P}(\mathbb{R}),$$

entailing quantification over internal subsets only.

EXAMPLE 7.10.3. The least upper bound property for bounded subsets of $\mathbb{R}$ holds over the hyperreals, provided we interpret it as applying to internal subsets only ("in the sense of Henkin").

7.11. From hyperrationals to reals

This section supplements the material of Section 5.3 on the ultrapower construction. We used the ultrapower construction to build the hyperreal field out of the real field. But the ultrapower construction can also be used to give an alternative construction of the real number system starting with the rationals. The star-transform of the field $\mathbb{Q}$ of rational numbers gives an extension $\mathbb{Q} \hookrightarrow {}^*\mathbb{Q}$ to the hyperrationals, where ${}^*\mathbb{Q} = \mathbb{Q}^{\mathbb{N}} / \mathrm{MAX}_{\mathcal{F}}$ (see Section 5.1 for details). Consider the subring $F \subseteq {}^*\mathbb{Q}$ consisting of all finite hyperrationals. Thus F is the galaxy of the number 0; see Definition 6.1.2. More precisely, we have the following definition.

DEFINITION 7.11.1. The ring F is defined by the following two equivalent conditions:
(1) F is the galaxy of $0 \in {}^*\mathbb{Q}$;
(2) $F = \left\{x \in {}^*\mathbb{Q} \colon (\exists r \in \mathbb{Q}) \left[|x| < r\right]\right\}$.

DEFINITION 7.11.2. The ideal $I \subseteq F$ of infinitesimal elements is defined by setting $I = \left\{x \in F \colon (\forall r \in \mathbb{Q}) \left[r > 0 \;\Rightarrow\; |x| < r\right]\right\}$.

Here the absolute value bars denote the natural extension of the usual absolute value function on the rationals (as usual for functions, the star superscript on the absolute value function is suppressed).

Note that F is only a ring (and not a field) because an infinitesimal is not invertible in F. In Theorem 8.1.1 we will now show that the ring F admits a natural quotient which is isomorphic to $\mathbb{R}$.

DEFINITION 7.11.3. The *halo* $\mathrm{hal}(x) \subseteq {}^*\mathbb{Q}$ of an element $x \in {}^*\mathbb{Q}$ is the set of elements of ${}^*\mathbb{Q}$ infinitely close to x.

EXAMPLE 7.11.4. We have $\mathrm{hal}(0) = I$.

CHAPTER 8

Halos, ihulls, manifolds

8.1. From hyperrationals to reals bis

Recall that

(1) the ring ${}^\flat\mathbb{R} \subseteq {}^*\mathbb{R}$ consists of finite hyperreals (see Definition 4.1.3);
(2) we have ${}^*\mathbb{Q} \subseteq {}^*\mathbb{R}$;
(3) $F \subseteq {}^*\mathbb{Q}$ is the ring of finite hyperrationals;
(4) $I \subseteq F$ of hyperrational infinitesimals;
(5) If $x \in {}^*\mathbb{Q}$ then its halo is $\mathrm{hal}(x) = x + I \subseteq {}^*\mathbb{Q}$.

THEOREM 8.1.1. *The ideal $I \subseteq F$ is maximal, and the quotient field $\hat{\mathbb{Q}} = F/I$ is naturally isomorphic to $\mathbb{R}$, so that we have a short exact sequence*[1]

$$0 \to I \to F \to \mathbb{R} \to 0.$$

We will provide the isomorphisms in each direction, denoted as follows: $\phi\colon \hat{\mathbb{Q}} \to \mathbb{R}$ and $\psi\colon \mathbb{R} \to \hat{\mathbb{Q}}$.

HOMOMORPHISM FROM $\hat{\mathbb{Q}}$ TO $\mathbb{R}$. A typical element of $\hat{\mathbb{Q}}$ is a halo, namely $\mathrm{hal}(x) \subseteq {}^*\mathbb{Q}$, where $x \in F$ can be viewed as an element of the ring ${}^\flat\mathbb{R} \subseteq {}^*\mathbb{R}$. Since x is finite, its standard part is well-defined, and we set

$$\phi(\mathrm{hal}(x)) = \mathbf{st}(x), \tag{8.1.1}$$

where $\mathbf{st}\colon {}^\flat\mathbb{R} \to \mathbb{R}$ is the standard part. Formula (8.1.1) provides the required homomorphism $\phi\colon \hat{\mathbb{Q}} \to \mathbb{R}$. □

COROLLARY 8.1.2. *Given a finite $x \in {}^*\mathbb{Q}$, consider the extended decimal expansion*

$$x = a.d_1d_2d_3\ldots d_nd_{n+1}\ldots;\ldots d_Hd_{H+1}\ldots. \tag{8.1.2}$$

Then the digits of x of finite rank (i.e., digits before the semicolon in (8.1.2)*) define a standard decimal expansion which uniquely determines the real number $\mathbf{st}(x)$, giving an alternative construction of the homomorphism ϕ.*

[1]Sidra meduyeket ktzara

Homomorphism from $\mathbb{R}$ to $\hat{\mathbb{Q}}$. Let us now construct a homomorphism in the opposite direction. The idea is to truncate the extended decimal expansion of a real number at a hypernatural rank. Choose a fixed $H \in {}^*\mathbb{N} \setminus \mathbb{N}$. Consider a positive real number $y \in \mathbb{R}$. The procedure will be similar for a negative number but we describe it for a positive number so as to fix ideas. Let $K = \lfloor 10^H y \rfloor$. Then we have $0 \leq 10^H y - K < 1$. Dividing by 10^H we obtain

$$0 \leq y - \frac{K}{10^H} < \frac{1}{10^H}.$$

It follows that $y \approx \frac{K}{10^H}$. In other words, we are considering the decimal approximation of y truncated at rank H, yielding a hyperrational number

$$\frac{K}{10^H} \in F \subseteq {}^*\mathbb{Q}.$$

Then the map $\psi : \mathbb{R} \to \hat{\mathbb{Q}}$ is defined by sending the real number y to the halo $\mathrm{hal}\left(\frac{K}{10^H}\right) \subseteq {}^*\mathbb{Q}$. In formulas, we have

$$\psi(y) = \frac{K}{10^H} + I.$$

Then we have $\phi(\psi(y)) = \mathbf{st}\left(\frac{K}{10^H}\right) = y$, proving the theorem. □

Second proof. This can be elaborated as follows in terms of the extended decimal expansion. Consider the decimal expansion of y as $y = a.a_1a_2a_3 \ldots a_n \ldots$ defined for each natural index $n \in \mathbb{N}$. By transfer, the decimal digits a_n are defined for each hypernatural rank, as well: $y = a.a_1a_2a_3 \ldots a_n \ldots ; \ldots a_{H-1}a_Ha_{H+1} \ldots$ Here ranks to the left of the semicolon are finite, while ranks to the right of the semicolon are infinite. Choose an infinite hypernatural $H \in {}^*\mathbb{N} \setminus \mathbb{N}$. The product $10^H y$ has the form $10^H y = a\, a_1a_2a_3 \ldots a_n \ldots a_{H-1}a_H \,.\, a_{H+1} \ldots$ with a decimal point after the digit a_H. Therefore its integer part (floor) is the hyperinteger $K = \lfloor 10^H y \rfloor = a\, a_1a_2a_3 \ldots a_n \ldots a_{H-1}a_H$, and we proceed as above. □

Third proof. Over $\mathbb{R}$ we have

$$(\forall \epsilon > 0)(\forall y \in \mathbb{R})(\exists q \in \mathbb{Q})[|y = q| < \epsilon].$$

By transfer,

$$(\forall \epsilon \in {}^*\mathbb{R}^+)(\forall y \in {}^*\mathbb{R})(\exists q \in {}^*\mathbb{Q})[|y = q| < \epsilon].$$

Now choose an infinitesimal $\epsilon > 0$. It follows that for each real y there is a hyperrational q with $y \approx q$. □

8.2. Ihull construction

In the previous section we described a construction of $\mathbb{R}$ starting from ${}^*\mathbb{Q}$. More generally, one has the following *ihull construction* ("i" for *infinitesimal*)[2] in the context of an arbitrary metric space M. Here the halo of $x \in {}^*M$ is defined to be the set of points at infinitesimal distance from x.

DEFINITION 8.2.1. Given a metric space (M, d), we build ${}^*M$ via the ultrapower. The distance function d extends to ${}^*M$ as usual. Let $\approx$ be the relation of infinite proximity in ${}^*M$, and denote by $F \subseteq {}^*M$ the set of finite elements of ${}^*M$ (i.e., the galaxy of any element in M). The quotient $F/\approx$ is called the *ihull* of M and denoted $\hat{M}$.

COROLLARY 8.2.2. *The ihull of* $\mathbb{Q}$ *is* $\mathbb{R}$.

This is a restatement of Theorem 8.1.1.

8.3. Repeating the construction

What happens if we apply the ihull construction to $M = \mathbb{R}$ in place of $\mathbb{Q}$? Namely, consider the star-transfer ${}^*\mathbb{R}$, and the ring of finite hyperreals ${}^*\mathbb{R}_F \subseteq {}^*\mathbb{R}$, as well as the ideal of infinitesimals

$$ {}^*\mathbb{R}_I \subseteq {}^*\mathbb{R}_F. $$

It turns out that we do not get anything new in this direction, as attested to by the following.

THEOREM 8.3.1. *The quotient* $\widehat{\mathbb{R}} = {}^*\mathbb{R}_F / {}^*\mathbb{R}_I$ *is naturally isomorphic to* $\mathbb{R}$ *itself; briefly,* $\mathbb{R}$ *is its own ihull, i.e.,* $\hat{\mathbb{R}} = \mathbb{R}$*, so that we have a short exact sequence*

$$ 0 \to {}^*\mathbb{R}_I \to {}^*\mathbb{R}_F \to \mathbb{R} \to 0. $$

The proof is essentially the same, with the main ingredient being the existence of the standard part function with range $\mathbb{R}$:

$$ \mathbf{st} : {}^*\mathbb{R}_F \to \mathbb{R} $$

with kernel precisely ${}^*\mathbb{R}_I$.

REMARK 8.3.2. Similar considerations apply to $M = \mathbb{R}^n$, showing that the quotient $\widehat{\mathbb{R}^n} = {}^*\mathbb{R}^n_F / {}^*\mathbb{R}^n_I$ is naturally isomorphic to $\mathbb{R}^n$ itself. We will return to this observation in Section 10.3.

[2]In the literature this construction is often referred to as the *nonstandard hull construction* (meatefet lo standartit).

8.4. Plane with a puncture

We will exploit the following useful notation for halos.

Definition 8.4.1. We set $\overset{\circ}{x} = \mathrm{hal}(x)$.

Remark 8.4.2. In general the ihull $\hat{X}$ of a metric space X consists of halos $\overset{\circ}{x}$ with distance d defined to be

$$d(\overset{\circ}{x}, \overset{\circ}{y}) = \mathbf{st}({}^*d(x,y)). \tag{8.4.1}$$

Here $\hat{X}$ may be larger than the metric completion of X. The metric completion of X is in general defined as the approachable part of $\hat{X}$, or the closure of X in $\hat{X}$. For a detailed discussion see [**Goldblatt 1998**, Chapter 18]. See next section for an example.

The usual flat metric $dx^2 + dy^2$ in the (x,y)-plane can be written in polar coordinates (r,θ) as $dx^2 + dy^2 = dr^2 + r^2 d\theta^2$ where θ is the usual polar angle in $\mathbb{R}/2\pi\,\mathbb{Z}$.

Let X be the universal cover of the plane minus the origin, coordinatized by (r,ζ) where $r > 0$ and ζ is an arbitrary real number. In formulas, X is a coordinate chart $r > 0$, $\zeta \in \mathbb{R}$ with metric

$$dr^2 + r^2 d\zeta^2, \tag{8.4.2}$$

giving the universal cover of the flat metric on $\mathbb{R}^2 \setminus \{0\}$, for which the covering map $X \to \mathbb{R}^2 \setminus \{0\}$, $(r,\zeta) \mapsto (r,\theta)$ induces an isometry of the Riemannian metric at each point.

Proposition 8.4.3. *Points of the ihull $\hat{X}$ of the form $\overset{\circ}{(r,\zeta)}$ for appreciable r and infinite ζ are not approachable from X.*

Proof. The distance function d of X extends to the ihull $(\hat{X}, d)$. Here points of $\hat{X}$ are halos in the finite part of ${}^*X$. Notice that in $\hat{X}$ the origin has been "restored" and can be represented by a point $(\epsilon, 0)$ in ${}^*X$ where $\epsilon > 0$ is infinitesimal.

Consider a point $(1,\zeta) \in {}^*X$ where ζ is infinite. Let us show that the point $(1,\zeta)$ is at a finite distance ${}^*d$ from the point $(\epsilon, 0)$, namely at distance infinitely close to 1. Indeed, the triangle inequality applied to the sequence of points $(1,\zeta)$, $(\frac{1}{\zeta^2},\zeta)$, $(\frac{1}{\zeta^2},0)$, $(\epsilon,0)$ yields the bound

$$d^*((1,\zeta),(\epsilon,0)) \le (1 - \tfrac{1}{\zeta^2}) + \tfrac{1}{\zeta^2}\zeta + |\tfrac{1}{\zeta^2} - \epsilon| \approx 1$$

and therefore

$$d(\overset{\circ}{(1,\zeta)}, \overset{\circ}{(\epsilon,0)}) = 1$$

by (8.4.1). Hence $\overset{\circ}{(1,\zeta)} \in \hat{X}$.

On the other hand, let us show that the point $\overparen{(1,\zeta)} \in \hat{X}$ is not approachable from X. Indeed, the metric d of $\hat{X}$ restricted to the rectangle

$$^*[\tfrac{1}{2}, 2] \times {}^*[\zeta - 1, \zeta + 1]$$

dominates the product metric $dr^2 + \frac{1}{4} d\zeta^2$ by (8.4.2), and therefore the rectangle includes the metric ball centered at $\overparen{(1,\zeta)}$ of radius $\frac{1}{2}$ without any standard points. ☐

Remark 8.4.4. In the example above, the closure $\bar{X} \subseteq \hat{X}$ is not locally compact. Thus, the boundary of the metric unit ball in $\bar{X}$ centered at the origin $\overparen{(\epsilon, 0)}$ is a line. Therefore the space $\bar{X}$ does not satisfy the hypotheses of Theorem 10.4.7.

CHAPTER 9

1-parameter groups of transformations, invariance

9.1. 1-parameter groups of transformations of a manifold

Let $\theta\colon \mathbb{R} \times M \to M$ be a smooth mapping (this θ is in general unrelated to polar coordinates). Denote the coordinates by $t \in \mathbb{R}$ and $p \in M$, so that we have a map $\theta = \theta(t,p)$. We will often write $\theta_t(p)$ for $\theta(t,p)$.

Definition 9.1.1. Assume θ satisifies the following two conditions:

(1) $\theta_0(p) = p$ for all $p \in M$;

(2) $\theta_t \circ \theta_s(p) = \theta_{t+s}(p)$ for all $p \in M$ and all $s, t \in \mathbb{R}$.

Then θ is called a 1-parameter group of transformations, a *flow*, or an *action* (of $\mathbb{R}$ on M) for short.

An action θ defines a vector field X on M called the *infinitesimal generator*[1] of θ, as follows.

Definition 9.1.2. The *infinitesimal generator* X of the flow θ_t at a point $p \in M$ is a derivation $X_p\colon \mathbb{D}_p \to \mathbb{R}$ defined by setting for each function f on M,

$$X_p(f) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(f(\theta_{\Delta t}(p)) - f(p)\right).$$

Remark 9.1.3. The term *infinitesimal generator* is a generally accepted term defined as above in a context where actual infinitesimals are not present, as already mentioned in footnote 1.

9.2. Invariance of infinitesimal generator under flow

In this section we follow [**Boothby 1986**]. Consider a flow θ_t on a manifold M as in Section 9.1. Thus we have a map $\theta\colon \mathbb{R} \times M \to M$, where we write $\theta_t(p)$ for $\theta(t,p)$, such that $\theta_0(p) = p$ for all $p \in M$, and $\theta_t \circ \theta_s(p) = \theta_{t+s}(p)$ for all $p \in M$ and all $s, t \in \mathbb{R}$.

Then one obtains an induced action θ_{t_*} on a vector field Y on M, producing a new vector field $\theta_{t_*}(Y)$, as follows. We view Y as a derivation acting on functions f on M. Given a function f on M, we need to

[1] This term is common in the literature and does not use infinitesimals in the sense of Robinson.

specify the manner in which the new vector field $\theta_{t_*}(Y)$ differentiates the function f.

Definition 9.2.1. The induced action θ_{t_*} on the vector field Y of the flow θ of the manifold M is defined by setting

$$\theta_{t_*}(Y_p)f = Y_p(f \circ \theta_t) \tag{9.2.1}$$

for all $f \in \mathbb{D}_q$ where $q = \theta_t(p)$.

Remark 9.2.2. The vector $\theta_{t_*}(Y_p)$ resides at the point $q = \theta_t(p) \in M$, so that $\theta_{t_*} : T_p \to T_{\theta_t(p)}$.

In more detail, let $q = \theta_t(p)$. If a function f is defined in a neighborhood of $q \in M$ then the composed function $f \circ \theta_t$ (for a fixed t) is defined in a neighorhood of p and therefore it makes sense apply the derivation Y_p to the function $f \circ \theta_t$, as in formula (9.2.1).

Recall that the infinitesimal generator X of a flow θ is the vector field satisfying $X_p(f) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\big(f(\theta_{\Delta t}(p)) - f(p)\big)$ (see Definition 9.1.2). What happens when we apply the induced action of the flow to the vector field which is the infinitesimal generator of the flow itself? An answer is provided by the following theorem.

Theorem 9.2.3 (Boothby p. 124). *The infinitesimal generator X of an action θ_t is invariant under the flow:*

$$\theta_{t_*}(X_p) = X_{\theta_t(p)}.$$

Proof. The proof is a direct computation, obtained by testing the field X on a function $f \in \mathbb{D}_q$ where $q = \theta_t(p)$, and $\mathbb{D}_q$ is the space of smooth functions defined in a neighborhood of q. Thus, we have

$$\begin{aligned} \theta_{t_*}(X_p)f &= X_p(f \circ \theta_t) \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\big(f \circ \theta_t\,(\theta_{\Delta t}(p)) - f \circ \theta_t(p)\big). \end{aligned}$$

By definition of a flow, we have $\theta_t \circ \theta_{\Delta t} = \theta_{t+\Delta t} = \theta_{\Delta t} \circ \theta_t$. Using this commutation relation, we can express the action as follows:

$$\begin{aligned} \theta_{t_*}(X_p)f &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\big(f \circ \theta_{\Delta t}(\theta_t(p)) - f(\theta_t(p))\big). \\ &= X_{\theta_t(p)}f. \end{aligned}$$

Thus $\theta_{t_*}(X_p) = X_{\theta_t(p)}$, proving the theorem. □

Remark 9.2.4. While Theorem 9.2.3 is intuitively "obvious", the received formalism involved in proving it is bulky. In particular we had to resort to a test function f which does not appear in the formulation of the invariance. We will develop a more direct approach in Section 12.4.

9.3. Lie derivative and Lie bracket

In this section as in the previous one we follow Boothby. Let θ_t be a flow on M with infinitesimal generator X. Let Y be another vector field on M.

DEFINITION 9.3.1 (Boothby p. 154). The vector field $L_X Y$, called the *Lie derivative* of Y with respect to X, is defined by setting

$$L_X Y = \lim_{t \to 0} \frac{1}{t} \left(\theta_{-t*}(Y_{\theta(t,p)}) - Y_p \right)$$

THEOREM 9.3.2. *If X and Y are vector fields on M then*

$$L_X Y = [X, Y],$$

where $[X,Y]f = X_p(Yf) - Y_p(Xf)$ is the Lie bracket.

PROOF. This can be checked in coordinates using a Taylor formula with remainder. □

EXAMPLE 9.3.3. The coordinate vector fields commute with respect to the Lie bracket, i.e.,

$$\left[\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \right] = 0$$

for all $i, j = 1, \ldots, n$. This is an equivalent way of stating the equality of mixed partials $\frac{\partial^2 f}{\partial x \, \partial y} = \frac{\partial^2 f}{\partial y \, \partial x}$ (sometimes called Schwarz's theorem or Clairaut's theorem). Thus $L_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} = 0$.

EXAMPLE 9.3.4. As an example of a nontrivial bracket, one has

$$\left[y \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right] = -\frac{\partial}{\partial x}.$$

In other words, vector fields $y\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ don't commute.

9.4. Diffeomorphisms

In this section as in Section 9.3 we follow Boothby. We first review some concepts from advanced calculus. If

$$\sigma \colon M \to N$$

is a smooth map such that $\sigma(p) = q \in N$, we obtain an induced map $\sigma_* \colon T_p M \to T_q N$. Here σ_* is the map induced on the tangent space. This can be defined as in Definition 9.4.2 below.

REMARK 9.4.1. A curve $\alpha(t)$ in M with $\alpha(0) = p$ defines a derivation X acting on functions $f \in \mathbb{D}_p$ by $Xf = \frac{d}{dt}\big|_{t=0} f(\alpha(t))$. We then write $X = \alpha'(0)$. We have the following extension of Definition 9.2.1.

DEFINITION 9.4.2. Choose a representing curve $\alpha(t)$ in M so that we have $X = \alpha'(0)$, and define $\sigma_*(X)$ to be the vector represented by the curve defined by the composition $\sigma \circ \alpha$ in the image manifold N.

In coordinate-free notation, chain rule takes the following form.

THEOREM 9.4.3 (Chain rule). *Chain rule for smooth maps σ, τ among manifolds asserts that the composition $\sigma \circ \tau$ satisfies*

$$(\sigma \circ \tau)_* = \sigma_* \circ \tau_*,$$

i.e., $(\sigma \circ \tau)_*(X) = \sigma_* \circ \tau_*(X)$ for all tangent vectors X.

DEFINITION 9.4.4. A *diffeomorphism* of M is a bijective one-to-one map $\phi\colon M \to M$ such that both ϕ and ϕ^{-1} are C^∞ maps for all coordinate charts.

REMARK 9.4.5. If ϕ is a diffeomorphism then $\phi_*\colon T_p \to T_{\phi(p)}$ is a vector space isomorphism.

9.5. Commutation and invariance under diffeomorphism

Recall that we have the notion of a flow

$$\theta_t\colon M \to M \tag{9.5.1}$$

depending on parameter $t \in \mathbb{R}$. Recall from Definition 9.1.2 that the vector field X (which is the infinitesimal generator of the flow) and the flow θ_t itself are related by

$$X_p(f) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(f(\theta_{\Delta t}(p)) - f(p) \right). \tag{9.5.2}$$

DEFINITION 9.5.1. A vector field X is said to *generate* a flow θ_t if X is the infinitesimal generator of θ_t.

DEFINITION 9.5.2. An *integral curve*, or *orbit*, of a vector field X through a fixed point $p \in M$ is a curve $\theta(t,p)$ satisfying $\theta(0,p) = p$ and $\frac{\partial \theta}{\partial t} = X$ (in the sense of Remark 9.4.1) where t ranges through an open interval containing 0.

Recall that a vector field X is said to be invariant under a smooth map σ if $\sigma_*(X) = X$ at every point. In more detail, if $q = \sigma(p) \in M$ then we require that $\sigma_*(X_p) = X_q$.

THEOREM 9.5.3 (Boothby p. 142). *Let X be a vector field generating a flow $\theta(t,p)$ on M and let $F\colon M \to M$ be a diffeomorphism. Then the following two conditions are equivalent:*

(1) *X is invariant under F;*
(2) *F commutes with the flow, i.e., $F(\theta(t,p)) = \theta(t, F(p))$.*

PROOF OF (1) $\Longrightarrow$ (2). Suppose X is invariant under F. Thus we have $F_*(X_p) = X_{F(p)}$ for all $p \in M$. Let $\sigma_p(t)$ be an integral curve through $p \in M$. Since we wish to use p as the index, we use the letter σ in place of θ to avoid clash of notation with θ_t introduced in (9.5.1). By the chain rule we have $(F \circ \sigma_p)_* = F_* \circ \sigma_{p*}$. Thus the curve $F \circ \sigma_p \colon \mathbb{R} \to M$ induces a map sending the vector $\frac{d}{dt}$ (a tangent vector to $\mathbb{R}$) to the vector $F_*(X_p)$. Here $\frac{d}{dt}$ is the natural basis for the tangent line $T_0\,\mathbb{R}$ at the origin, and $X_p = \dot{\sigma}_p(0) = \sigma_{p*}\left(\frac{d}{dt}\right)$ where $\dot{\sigma} = \frac{d\sigma}{dt}$. In other words, the curve $F \circ \sigma_p \colon \mathbb{R} \to M$ is an integral curve of the vector field $F_*(X)$. Hence the map F takes the integral curve σ_p of the vector field X to an integral curve of the vector field $F_*(X)$.

Since $F_*(X) = X$ by hypothesis, the uniqueness of integral curves implies that $F(\theta(t,p)) = \theta(t, F(p))$. $\square$

PROOF OF (2) $\Longrightarrow$ (1). We now assume the commutation relation $F(\theta(t,p)) = \theta(t, F(p))$. Let us prove (1). Note that the vector $F_*(X_p)$ is represented by the curve $F \circ \sigma_p(t)$ and by the commutation this is precisely $\sigma_{F(p)}(t)$, which indeed represents the vector field X at the point $F(p)$. Thus $F_*(X_p) = X_{F(p)}$ as required. $\square$

CHAPTER 10

Commuting vfields, flows, differential geom. via ID

10.1. Commutation of vector fields and flows

In this section as in the previous one we follow Boothby. Here we prove a special case of the Frobenius theorem on flows, in the case of commuting flows.

Recall from Definition 9.3.1 that if θ_t is a flow on M with infinitesimal generator X, and Y is a vector field on M, then the vector field $L_X Y$, called the *Lie derivative* of Y with respect to X, is defined by setting $L_X Y = \lim_{t\to 0} \frac{1}{t}\left(\theta_{-t*}(Y_{\theta(t,p)}) - Y_p\right)$ and furthermore $L_X Y = [X, Y]$.

THEOREM 10.1.1 (Boothby p. 156). *Let X and Y be C^∞ vector fields on a manifold M. Let $\theta = \theta_t$ be the flow generated by X, and $\eta = \eta_s$ be the flow generated by Y. Then the following two conditions are equivalent:*

(1) $[X, Y] = 0$;
(2) *for each $p \in M$ there exists a $\delta_p > 0$ such that $\eta_s \circ \theta_t(p) = \theta_t \circ \eta_s(p)$ whenever $|s| < \delta_p$ and $|t| < \delta_p$.*

PROOF. First we show the easy direction (2) $\Longrightarrow$ (1). We apply Theorem 9.5.3 with $F = \theta_t$ to conclude that the infinitesimal generator Y of the flow η_s is invariant under the map θ_t for a fixed small parameter t. It follows that $\theta_{t*}(Y_q) = Y_{\theta(t,q)}$, i.e., $\theta_{-t*}(Y_{\theta(t,q)}) = Y_q$ by Remark 9.4.5. Now consider the Lie derivative. We obtain[1]

$$\begin{aligned}
[X, Y]_q &= (L_X Y)_q \\
&= \lim_{t\to 0} \frac{1}{t}\left(\theta_{-t*}(Y_{\theta(t,q)}) - Y_q\right) \\
&= \lim_{t\to 0} \frac{1}{t}(Y_q - Y_q) \\
&= 0
\end{aligned}$$

as required. □

[1]The calculation in Boothby on page 157 contains some misprints involving misplaced parentheses.

PROOF OF THE OPPOSITE DIRECTION. Now let us show the opposite direction (1) $\Longrightarrow$ (2) in Theorem 10.1.1. Assume that $[X,Y]=0$ identically on M. Consider a point $q \in M$. Define a vector $Z_q(t)$ at the point q by setting

$$Z_q(t) = \theta_{-t*}\left(Y_{\theta(t,q)}\right) \in T_qM. \tag{10.1.1}$$

Let us show that the t-derivative $\dot{Z}(t)$ vanishes. Let $q' = \theta(t,q)$ so that $Z_q(t) = \theta_{-t*}(Y_{q'})$. Then

$$\dot{Z}_q(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\theta_{-(t+\Delta t)_*}\left(Y_{\theta(\Delta t,q')}\right) - \theta_{-t*}(Y_{q'})\right) \tag{10.1.2}$$

Applying chain rule, we decompose the action of $\theta_{-(t+\Delta t)_*}$ as composition

$$\theta_{-(t+\Delta t)_*} = \theta_{-t*} \circ \theta_{-\Delta t*}.$$

Thus we obtain from (10.1.2) and linearity and continuity of induced maps that

$$\dot{Z}_q(t) = \theta_{-t*}\left(\lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\theta_{-\Delta t*}\left(Y_{\theta(\Delta t,q')}\right) - Y_{q'}\right)\right) \tag{10.1.3}$$

Now we recognize the expression $\frac{1}{\Delta t}\left(\theta_{-\Delta t*}\left(Y_{\theta(\Delta t,q')}\right) - Y_{q'}\right)$ as the Lie derivative L_XY of Y at the point $q' = \theta(t,q)$. Therefore we obtain

$$\dot{Z}_q(t) = \theta_{-t*}\left((L_XY)_{q'}\right) = \theta_{-t*}(0) = 0$$

since by hypothesis (1), the Lie bracket vanishes everywhere, including at the point $q' = \theta(t,q)$.

Since the t-derivative vanishes, the vector $Z_q(t)$ of (10.1.1) is constant for $|t| < \delta$, and therefore equal to Y_q. This means that the vector field Y is invariant under the flow θ_t. Therefore by Theorem 9.5.3, we obtain the commutation of flows $\eta_s \circ \theta_t(q) = \theta_t \circ \eta_s(q)$. □

A more general version of the theorem of Frobenius is proved in Section 10.3.1.

The equivalence of commutation and vanishing Lie is done using prevector fields in Section 15.23.

10.2. Infinitesimals and infinitesimal generators

Here we follow [**Nowik & Katz 2015**]. In Chapter 9.2, we presented the traditional A-track formalism for dealing with the infinitesimal generator of a flow on a manifold. Note that the adjective *infinitesimal* in this context is a dead metaphor as it no longer refers to true infinitesimals. We will now develop a true infinitesimal formalism for dealing with the infinitesimal generator of a flow.

An infinitesimal (B-track) formalism enables a more intuitive approach not merely to the infinitesimal generator of a flow, but to many other topics in calculus and analysis. Thus, the concepts of derivative, continuity, integral, and limit can all be defined via infinitesimals. For example, in Section 7.2 we presented the infinitesimal approach to the extreme value theorem.

10.3. Relations of $\prec$ and $\prec\prec$

We will work in the hyperreal extension $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ defined in Chapter 4.7.

Definition 10.3.1. Let $r, s \in {}^*\mathbb{R}$. We define the relation $\prec$ by writing[2]

$$r \prec s$$

if $r = as$ for finite a. We also define the relation $\prec\prec$ by writing[3]

$$r \prec\prec s$$

if $r = as$ for infinitesimal a.

Example 10.3.2. The formula $r \prec 1$ means that r is finite; $r \prec\prec 1$ means that r is infinitesimal.

More generally, consider a finite dimensional vector space V over $\mathbb{R}$.

Definition 10.3.3. Given $v \in {}^*V$ and $s \in {}^*\mathbb{R}$, we write

$$v \prec s$$

if ${}^*\|v\| \prec s$ for some norm $\| \cdot \|$ on V.

We will generally omit the asterisk $*$ from function symbols and so will simply write $\|v\| \prec s$. This condition is independent of the choice of norm since all norms on V are equivalent.

Definition 10.3.4. We write

$$v \prec\prec s$$

if $\|v\| \prec\prec s$ (with similar remarks with regard to the norm).

Definition 10.3.5. Given $v, w \in {}^*V$, we write

$$v \approx w$$

when $v - w \prec\prec 1$.

We choose a basis for V thus identifying it with $\mathbb{R}^n$.

[2]Read: "Big-O"

[3]Read "little-o".

LEMMA 10.3.6. *An element $x = (x^1, \ldots, x^n) \in {}^*\mathbb{R}^n$ satisfies $x \prec s$ or $x \prec\prec s$ if and only if for each $i = 1, \ldots, n$, the component x^i satisfies the corresponding relation.*

This can be checked directly for the Euclidean norm on $\mathbb{R}^n$.

DEFINITION 10.3.7. Let $s \in {}^*\mathbb{R}$, $s > 0$. We set

$$\begin{cases} V_F^s = \{v \in {}^*V : v \prec s\} \\ V_I^s = \{v \in {}^*V : v \prec\prec s\} \end{cases}$$

Then $V_I^s \subseteq V_F^s \subseteq {}^*V$ are linear subspaces over $\mathbb{R}$.

THEOREM 10.3.8. *For each finite-dimensional vector space V over $\mathbb{R}$ we have a canonical isomorphism $V_F^s/V_I^s \cong V$, so that we obtain a short exact sequence*

$$0 \to V_I^s \to V_F^s \to V \to 0.$$

PROOF. This is a special case of the ihull construction. Indeed, by Section 8.3 we have $V_F^1/V_I^1 \cong V$, and multiplication by s maps V_F^1 onto V_F^s and V_I^1 onto V_I^s. Briefly, Euclidean space is its own ihull; see Remark 8.4.2. □

10.3.1. Proof of Frobenius.

This section is optional. We will denote by $[\cdot, \cdot]$ the classical Lie bracket.

THEOREM 10.3.9 (Frobenius). *Let $U \subseteq \mathbb{R}^n$ be open. Let $X_1, \ldots, X_k : U \to \mathbb{R}^n$ be k independent C^1 classical vector fields s.t. $[X_i, X_j] = \sum_{m=1}^k C_{ij}^m X_m$ for $C_{ij}^m : U \to \mathbb{R}$. Let g_t^i be the classical flow of X_i. Given $p \in U$, let $r > 0$ be such that $\varphi(t_1, \ldots, t_k) = g_{t_1}^1 \circ g_{t_2}^2 \circ \cdots \circ g_{t_k}^k(p)$ is defined on $(-r, r)^k$ and $\frac{\partial \varphi}{\partial t_1}, \ldots, \frac{\partial \varphi}{\partial t_k}$ are independent there. Then span $\left\{\frac{\partial \varphi}{\partial t_1}(a), \ldots, \frac{\partial \varphi}{\partial t_k}(a)\right\} =$ span$\{X_1(\sigma(a)), \ldots, X_k(\sigma(a))\}$ for every $a \in (-r, r)^k$.*

PROOF. By definition of $g_{t_i}^i$, we have $\frac{\partial}{\partial t_i}(g_{t_i}^i \circ \cdots \circ g_{t_k}^k)(p) = X_i$. So to show $\frac{\partial \varphi}{\partial t_i} \in span\{X_1, \ldots, X_k\}$ one needs to show that $(g_{t_1}^1 \circ \cdots \circ g_{t_{i-1}}^{i-1})_*(X_i) \in span\{X_1, \ldots, X_k\}$. Equivalently, one needs $(g_{t_j}^j)_*(X_i) \in span\{X_1, \ldots, X_k\}$ for every i, j, and for convenience of notation we let $j = 1$. Using the flow g_t^1 we may change coordinates so that $X_1 = \frac{\partial}{\partial x_1}$, and so the flow of X_1 is simply $g_t(x) = x + (t, 0, \ldots, 0)$, and we need to show that $span\{X_1, \ldots, X_k\}$ is invariant under the flow g_t, $t \in (-r, r)$.

For fixed q let $X_i(t) = X_i(q + (t, 0, \ldots, 0))$ then $\frac{d}{dt} X_i(t) = [X_1, X_i] = \sum_{m=1}^k C_{1i}^m(t) X_m(t)$. Since this simple flow maps a vector to the "same" (i.e. parallel) vector at the image point, $span\{X_1, \ldots, X_k\}$ being invariant under the flow means that $span\{X_1(t), \ldots, X_k(t)\}$ is the same subspace for all $t \in (-r, r)$.

Let us change basis in $\mathbb{R}^n$ such that $X_1(q), \dots, X_k(q)$ are the first k standard basis vectors.

Let us denote the coordinates of X_i by X_i^j, $j = 1, \dots, n$, so we must show that $X_i^j(t) = 0$ for all $1 \le i \le k$, $k+1 \le j \le n$ and $t \in (-r, r)$.

Let $W \subseteq (-r, r)$ be the set of all t satisfying $X_i^j(t) = 0$ for all $1 \le i \le k$, $k+1 \le j \le n$. Then W is clearly closed and $0 \in W$, so if we show that W is also open then $W = (-r, r)$ and we are done.

At this point we pass to the nonstandard extension. To show that W is open we must show that for every $w \in W$ and every $s \approx w$, $s \in {}^*W$. By transfer $s \in {}^*W$ means that $X_i^j(s) = 0$ for all $1 \le i \le k$, $k+1 \le j \le n$. Say $s > w$.

Let $A = \max |X_i^j(u)|$, $w \le u \le s$, $1 \le i \le k$, $k+1 \le j \le n$, and assume this maximum is attained at $u = u'$, with $i = a$, $j = b$.

Let $B = \max |\frac{d}{dt} X_i^j(u)|$, $w \le u \le s$, $1 \le i \le k$, $k+1 \le j \le n$, and assume this maximum is attained at $u = u''$, with $i = c$, $j = d$.

Then since $X_a^b(w) = 0$ we have

$$\begin{aligned} A &= |X_a^b(u')| \\ &= |X_a^b(u') - X_a^b(w)| \le |u' - w| B \\ &= |u' - w| \cdot |\frac{d}{dt} X_c^d(u'')| \\ &= |u' - w| \cdot |\sum_m C_{1c}^m(u'') X_m^d(u'')| \\ &\prec |u' - w| A \\ &\prec\prec A. \end{aligned}$$

We have $A \prec\prec A$ and so necessarily $A = 0$ and so $X_i^j(s) = 0$ for all $1 \le i \le k$, $k+1 \le j \le n$ (the notation $\prec$ was introduced in Section 10.3). □

10.4. Smooth manifolds, notion of nearstandardness

Our object of interest is a smooth manifold M together with an enlargement $M \hookrightarrow {}^*M$. Here as usual ${}^*M = M^{\mathbb{N}}/\mathcal{F}$ where $\mathcal{F}$ is a free ultrafilter on $\mathbb{N}$.

Definition 10.4.1. Let $p \in M$. The *halo* of p, which we denote by

$$\mathfrak{h}(p) = \mathfrak{h}_M(p), \tag{10.4.1}$$

is the set of all points $x \in {}^*M$, for which there is a coordinate neighborhood U of p such that $x \in {}^*U$ and $x \approx p$ in the given coordinates, or briefly $\mathfrak{h}_M(p) = \{x \in {}^*M : x \approx p\}$.

Remark 10.4.2. By including the index M in the notation for the halo as in (10.4.1), we emphasize its dependence on the ambient manifold. For example, the halo of the origin in ${}^*\mathbb{R}$ consists of hyperreal infinitesimals, while the halo of the origin in ${}^*\mathbb{R}^2$ consists of points both of whose coordinates are infinitesimal.

The definition of $\mathfrak{h}(p)$ does not require coordinates, but rather depends only on the *topology* of M. The points of M are called *standard*.

DEFINITION 10.4.3. A point of ${}^*M$ belonging to $\mathfrak{h}(p)$ for some $p \in M$ is called *nearstandard*.

DEFINITION 10.4.4. If a is nearstandard then $\mathbf{st}(a)$ is the unique $p \in M$ such that $a \in \mathfrak{h}(p)$.

DEFINITION 10.4.5. Let $A \subseteq M$ be a subset. The halo ${}^{\mathfrak{h}}A$ of A is a subset of ${}^*M$ defined by

$$ {}^{\mathfrak{h}}A = \bigcup_{a \in A} \mathfrak{h}_M(a). $$

EXAMPLE 10.4.6. For the ambient manifold M itself, then the halo ${}^{\mathfrak{h}}M$ is the set of all nearstandard points in ${}^*M$.

The following is proved in [**Davis 1977**, p. 90, Theorem 5.6]. Let X be a metric space, e.g., a finite-dimensional smooth manifold equipped with a Riemannian metric.

THEOREM 10.4.7 (Davis). *For a metric space X the following are equivalent:*

(1) *every bounded closed set in X is compact;*
(2) *every finite point*[4] *of ${}^*X$ is nearstandard.*

In particular, if M is a finite-dimensional complete Riemannian manifold, then a point of ${}^*M$ is *finite* if and only if it is *nearstandard*.

10.5. Results for general culture on topology

We have the following characterisations of open sets and compact sets. A good reference is [**Davis 1977**].

THEOREM 10.5.1. *A subset A is open in M if and only if ${}^{\mathfrak{h}}A \subseteq {}^*A$.*

EXAMPLE 10.5.2. The subset of $M = \mathbb{R}$ given by the closed interval $A = [0, 1]$ is not open since ${}^*A$ fails to contain negative infinitesimals which are in the halo of A, i.e., $\text{hal}_M(0) \not\subseteq {}^*A$.

THEOREM 10.5.3. [**Davis 1977**, p. 78, item 1.6] *A set $A \subseteq M$ is compact if and only if ${}^*A \subseteq {}^{\mathfrak{h}}A$.*

See Theorem 11.1.3 for a proof.

[4]Meaning that it is at finite distance from any $p \in M$.

EXAMPLE 10.5.4. The subset of $\mathbb{R}$ given by the open interval $A = (0,1)$ is not compact since a positive infinitesimal ϵ is in ${}^*\!A$ but ϵ is not in the halo of A, since it is not infinitely close to any point of A because every point of A is appreciable. Since every point of A is appreciable, so is every point of ${}^{\natural}\!A$, but infinitesimals are not appreciable.

In particular, if A is compact then ${}^*\!A \subseteq {}^{\natural}M$. Recall that the halo both of a point and of a subset are defined relative to the ambient manifold M (see Remark 10.4.2).

THEOREM 10.5.5 (Davis p. 78, item 1.7). *The ambient manifold M itself is compact if and only if* ${}^{\natural}M = {}^*M$.

REMARK 10.5.6. If M is noncompact then ${}^{\natural}M$ is an external set.[5]

EXAMPLE 10.5.7. The 1-dimensional manifold $M = \mathbb{R}$ is not compact. The halo ${}^{\natural}\mathbb{R}$ of $\mathbb{R}$ consists of all *finite* hyperreals. This is an external subset of ${}^*\mathbb{R}$. Recall that ${}^{\natural}\mathbb{R}$ is the domain of the standard part function $\mathbf{st}\colon {}^{\natural}\mathbb{R} \to \mathbb{R}$ defined in Section 4.1.

See Section 11.1 for some proofs.

10.6. Saturation

There is an important principle of *saturation*,[6] which we now introduce in a somewhat abridged or special form. We will need the following special case of the saturation property, formulated for nested[7] sequences.

THEOREM 10.6.1. *If $\langle A_n \colon n \in \mathbb{N}\rangle$ is a nested sequence of nonempty subsets of $\mathbb{R}$ then the sequence $\langle {}^*\!A_n \colon n \in \mathbb{N}\rangle$ has a common point.*

PROOF. Let $\mathbb{P} = \mathcal{P}(\mathbb{R})$ be the set of subsets of $\mathbb{R}$. Consider a sequence $\langle A_n \in \mathbb{P}\colon n \in \mathbb{N}\rangle$ viewed as a function $f\colon \mathbb{N} \to \mathbb{P}$, $n \mapsto A_n$. By the extension principle we have a function ${}^*\!f\colon {}^*\mathbb{N} \to {}^*\mathbb{P}$. Let $B_n = {}^*\!f(n)$. For each finite n we have $B_n = {}^*\!A_n \in {}^*\mathbb{P}$. For each infinite value of the index $n = H$ the entity $B_H \in {}^*\mathbb{P}$ is by definition internal but is not (necessarily) the natural extension of any subset of $\mathbb{R}$.

If $\langle A_n\rangle$ is a *nested* sequence in $P\setminus\{\varnothing\}$ then by transfer $\langle B_n : n \in {}^*\mathbb{N}\rangle$ is a nested sequence in ${}^*\mathbb{P} \setminus \{\varnothing\}$. Let H be a fixed infinite index. Then for each finite n the set ${}^*\!A_n \subseteq {}^*\mathbb{R}$ includes B_H. Choose any element $c \in B_H$. Then c is contained in ${}^*\!A_n$ for each finite n:

$$c \in \bigcap_{n\in\mathbb{N}} {}^*\!A_n$$

[5]A set is external if it is not internal. See Section 7.5.

[6]Revaya

[7]mekunenet

as required. □

Remark 10.6.2. The injective map $*\colon \mathbb{P} \to {}^*\mathbb{P}$ sends A_n to ${}^*\!A_n$. For each natural n we have a symbol a_n in the extended real number language (see Remark 5.6.4) whose standard interpretation is $A_n \in \mathbb{P}$. Meanwhile the nonstandard interpretation of a_n is the entity ${}^*\!A_n \in {}^*\mathbb{P}$. The sequence $\langle A_n \colon n \in \mathbb{N}\rangle$ in $\mathbb{P}$ is the standard interpretation of the symbol $a = \langle a_n \rangle$. The sequence $\langle B_n \colon n \in {}^*\mathbb{N}\rangle$ in ${}^*\mathbb{P}$ is the nonstandard interpretation of the symbol a, and in particular $B_n = {}^*\!A_n$ for finite n.

Corollary 10.6.3. *If the family of subsets* $\{A_n\}_{n\in\mathbb{N}}$ *has the finite intersection property (see Definition* 4.10.5*) then* $\exists c \in \bigcap_{n\in\mathbb{N}} {}^*\!A_n$.

This is an equivalent formulation of Theorem 10.6.1.

CHAPTER 11

Compactness, prevectors

11.1. Equivalent characterisations of compactness

Let X be a topological space. Let $p \in X$. A neighborhood of p is an open set that contains p. Recall the following.

(1) The halo of p, denoted $\mathfrak{h}(p) \subseteq {}^*X$, is the intersection of all ${}^*U$ where U runs over all neighborhoods of p in X.
(2) A point $y \in {}^*X$ is called nearstandard in X if there is $p \in X$ such that $y \in \mathfrak{h}(p)$ (see Section 10.4).
(3) The saturation property asserts that if sets $A_n \subseteq X$ form a decreasing (nested) sequence $\langle A_n \colon n \in \mathbb{N} \rangle$, then $\cap_{n\in\mathbb{N}} {}^*A_n \neq \varnothing$.

Example 11.1.1. Let $A_n = \{p \in \mathbb{N} \colon p \text{ prime and } p \geq n\}$. Then by saturation there exists an element $c \in \cap_{n\in\mathbb{N}} {}^*A_n$, called an *infinite prime.*

Lemma 11.1.2. *For a finite union, the star of union is the union of stars.*

Proof. We have $(\forall y \in X)\big[y \in A \cup B \iff (y \in A) \vee (y \in B)\big]$. This is a first-order formula. Applying upward transfer we obtain

$$(\forall y \in {}^*X)\big[y \in {}^*A \cup {}^*B \iff (y \in {}^*A) \vee (y \in {}^*B)\big].$$

The lemma now follows by induction. □

A more advanced application of saturation is the following theorem.

Theorem 11.1.3. *Suppose the topology of X admits a countable basis (e.g., X is a metric space). Then the following two conditions are equivalent:*

(1) *X is compact;*
(2) *every $y \in {}^*X$ is nearstandard in X.*

Proof of (1) ⇒ (2). Assume X is compact, and let $y \in {}^*X$. Let us show that y is nearstandard (this direction does not require saturation).

Suppose on the contrary that y is not nearstandard, i.e., y is not in the halo of any point $p \in X$. Thus, every $p \in X$ has a neighborhood U_p

such that

$$y \notin {}^*U_p. \tag{11.1.1}$$

Consider the collection $\{U_p\}_{p\in X}$. This collection is an open cover of X. Since X is compact, the collection has a finite subcover $U_{p_1}, \dots, U_{p_n}$, so that $X = U_{p_1} \cup \dots \cup U_{p_n}$. Applying Lemma 11.1.2 to this finite union, we obtain

$${}^*X = {}^*U_{p_1} \cup \dots \cup {}^*U_{p_n}.$$

Hence y is in one of the ${}^*U_{p_i}$, $i = 1, \dots, n$, contradicting (11.1.1). □

Proof of (2) ⇒ (1). This direction exploits saturation. Assume every $y \in {}^*X$ is nearstandard, and let $\{U_a\}$ be an open cover of X. We need to find a finite subcover.

Suppose on the contrary that the union of any finite collection of U_a is not all of X. Then the complements S_a of U_a form a collection of (closed) sets $\{S_a\}$ with the finite intersection property. It follows that the collection $\{{}^*S_a\}$ similarly has the finite intersection property.

At this point we use the condition that the family is countable (see Remark 11.1.4). By saturation (see Corollary 10.6.3), the intersection of all ${}^*S_a$ is non-empty. Let y be a point in this intersection. By hypothesis, there is a point $p \in X$ such that $y \in \mathfrak{h}(p)$. Now the $\{U_a\}$ form a cover of X so there is a U_b such that $p \in U_b$. But y is in ${}^*S_a$ for all a, in particular $y \in {}^*S_b$, so it is not in ${}^*U_b$, a contradiction to $y \in \mathfrak{h}(p)$. □

Remark 11.1.4. Given a basis for the topology of X, compactness is clearly equivalent to the property that every cover by basis sets has a finite subcover. If we have a countable basis, then all covers under consideration are countable, and so in the proof above we need only countable saturation.

Theorem 11.1.5 (Cantor's intersection theorem). *A nested decreasing sequence of nonempty compact sets has a common point.*

Proof. Given a nested sequence of compact sets K_n, we consider the corresponding decreasing nested sequence of internal sets, $\langle {}^*K_n \colon n \in \mathbb{N} \rangle$. This sequence has a common point x by saturation. But for a compact set K_n, every point of ${}^*K_n$ is nearstandard (i.e., infinitely close to a point of K_n). In particular, $\mathbf{st}(x) \in K_n$ for all n, as required. □

11.2. Properties of the natural extension

Given an open set $W \subseteq \mathbb{R}^n$ and a smooth function $f \colon W \to \mathbb{R}$ we note some properties of the extension ${}^*f \colon {}^*W \to {}^*\mathbb{R}$, obtained by transfer. When there is no risk of confusion, we will omit the asterisk $*$ from the function symbol ${}^*\!f$ and simply write f for both the original function and its extension. As for functions, we omit the asterisk $*$ from relation symbols, writing $\leq$ in place of ${}^*\!\leq$.

Recall that the external subset ${}^\natural W \subseteq {}^*W$ consists of all points infinitely close to some point of W (see Section 10.4). We can then consider the shadow map $\mathbf{st}\colon {}^\natural W \to W$.

LEMMA 11.2.1. *Let $W \subseteq \mathbb{R}^n$ be an open set. Assume $f\colon W \to \mathbb{R}$ is continuous. Then ${}^*\!f(a)$ is finite for each $a \in {}^\natural W$.*

PROOF. Given $a \in {}^\natural W$, let $U \subseteq W$ be a neighborhood of the point $\mathbf{st}(a) \in W$ such that $\overline{U} \subseteq W$ and $\overline{U}$ is compact.

By compactness, there is a constant $C \in \mathbb{R}$ such that $|f(x)| \leq C$ for all $x \in U$. By transfer $|f(x)| \leq C$ for all $x \in {}^*U$. In particular we have $|f(a)| \leq C$. Therefore $f(a)$ is finite. □

11.3. Remarks on gradient

Recall that ${}^\natural U$ is the set of nearstandard points of ${}^*U$ whose standard part is in U, where ${}^*U$ is the natural extension of U. Given an open set $U \subseteq \mathbb{R}^n$ and a smooth function $f\colon U \to \mathbb{R}$, we can compute the partial derivatives $\frac{\partial f}{\partial u^i}$ as usual.

REMARK 11.3.1. If $y = f(x)$ then we have

$$dy = \frac{\partial f}{\partial u^i}\, du^i$$

where the symbols dy and du^i can be interpreted either in the sense of Leibniz–Keisler differentials (see Section 4.4) or in the traditional sense of 1-forms (see Section 2.3).

We will consider the *row* vector

$$\left(\tfrac{\partial f}{\partial u^1} \; \tfrac{\partial f}{\partial u^2} \quad \ldots \quad \tfrac{\partial f}{\partial u^n} \right)$$

sometimes called the *gradient* of f in calculus.

DEFINITION 11.3.2. The partial derivatives of ${}^*\!f$ are by definition the functions

$${}^*\!\left(\frac{\partial f}{\partial u^i} \right),$$

i.e., the natural extensions of the partial derivatives of the real function f.

DEFINITION 11.3.3. Given a function $f \in C^1$ we consider the *row vector* D_a of partial derivatives at each point $a \in {}^*U$.

One similarly defines the higher partial derivatives of f at each point $a \in {}^*U$.

Definition 11.3.4. Given a function $f \in C^2$ we consider the Hessian matrix H_a of second partial derivatives of f at each point $a \in {}^*U$.

By Lemma 11.2.1, D_a and H_a are finite throughout the set ${}^\natural U$. For future applications, we would like to consider the infinitesimal segment between two infinitely close points in the halo of U.

Lemma 11.3.5. *Let $a, b \in {}^\natural U$ with $a \approx b$, then the segment between a and b is included in ${}^\natural U \subseteq {}^*U$.*

Proof. Each point in the segment between a and b is infinitely close to the point $\mathbf{st}(a) = \mathbf{st}(b) \in {}^\natural U$. □

Remark 11.3.6. We will not use the traditional bracket notation for this segment so as to avoid confusion.

We obtain the following version of the mean value theorem.

Theorem 11.3.7. *If $f\colon U \to \mathbb{R}$ is C^1 then ${}^*f(b) - {}^*f(a) = D_x\,(b-a)$ for a suitable x in the segment between a and b.*

Proof. This results by applying the transfer principle to the mean value theorem (see Section 7.4). Note that here $D_x\,(b-a)$ is a product of matrices, namely a row vector times a column vector. □

Equivalently, we can write

$$ {}^*f(b) - {}^*f(a) - D_x\,(b-a) = 0 \tag{11.3.1} $$

for the specific choice of the point x provided by the mean value theorem.

11.4. Version of MVT at an arbitrary point of the segment

At an arbitrary point y in place of x we have the following weaker statement. We will exploit the $\prec\!\prec$ notation (see Section 10.3).

Theorem 11.4.1. *Let f be C^1 and let $a, b \in {}^\natural U$ be infinitely close. Then for every y that is infinitely close to a, we have*

$$ f(b) - f(a) - D_y\,(b-a) \prec\!\prec \|b-a\|. $$

Proof. By hypothesis, the partial derivatives D_y are themselves continuous functions of y. Then the characterization of continuity via infinitesimals and microcontinuity in Section 6.2 implies that $D_x - D_y$ is infinitesimal for any $y \approx a$. Therefore using the special x from (11.3.1), we can write

$$\begin{aligned} f(b) - f(a) - D_y(b-a) &= \big(f(b) - f(a) - D_x(b-a)\big) + (D_x - D_y)(b-a) \\ &= (D_x - D_y)\,(b-a) \\ &\prec\prec \|b-a\| \end{aligned}$$

as required. □

Setting $y = a_0$, we obtain the following corollary.

COROLLARY 11.4.2. *Let $a_0 = \mathbf{st}(a)$. Then*

$$f(b) - f(a) - D_{a_0}(b-a) \prec\prec \|b-a\|.$$

REMARK 11.4.3. If the first partial derivatives are Lipschitz (e.g., if f is C^2), and we are given an infinitesimal constant β such that we have $\|x-y\| \prec \beta$ for all x in the segment between a and b, then the following stronger condition holds for such y:

$$f(b) - f(a) - D_y(b-a) \prec \beta\|b-a\|.$$

REMARK 11.4.4. If f is C^2 then by transfer of the Taylor approximation theorem we have

$$^*\!f(b) - {}^*\!f(a) = D_a(b-a) + \frac{1}{2}(b-a)^t H_x(b-a)$$

for a suitable x in the segment between a and b, and remarks similar to those we have made regarding $D_x - D_y$ apply to $H_x - H_y$.

REMARK 11.4.5. Let $\sigma = (\sigma^i)\colon U \to \mathbb{R}^n$ be a C^1 map. Then the n rows D_a^i corresponding to σ^i form the Jacobian matrix J_a of σ at a point a. By applying the above considerations to each σ^i we obtain that

$$\sigma(b) - \sigma(a) - J_y(b-a) \prec\prec \|b-a\|.$$

REMARK 11.4.6. If all first partial derivatives are Lipschitz (e.g., if σ is C^2) then we obtain the relation $\sigma(b) - \sigma(a) - J_y(b-a) \prec \beta\|b-a\|$ with β as in Remark 11.4.3.

11.5. Prevectors

We now choose a positive infinitesimal $\lambda \in {}^*\mathbb{R}$, and *fix it once and for all*. We will now define the concept of a prevector.

DEFINITION 11.5.1. Let $a \in {}^\natural M$. A *prevector* based at a is a pair (a, x), where $x \in {}^\natural M$, such that for each smooth function $f \in \mathbb{D}_{a_0}$, we have $^*\!f(x) - {}^*\!f(a) \prec \lambda$, where $a_0 = \mathbf{st}(a)$.

REMARK 11.5.2. The hypothesis ${}^*\!f(x) - {}^*\!f(a) \prec \lambda$ is stronger than merely requiring ${}^*\!f(x)$ and ${}^*\!f(a)$ to be infinitely close.

We can avoid quantification over functions occurring in Definition 11.5.1 by giving an equivalent condition of being a prevector in coordinates as follows. Consider coordinates in a neighborhood W of the standard point $a_0 = \mathbf{st}(a)$ in M, whose image in Euclidean space is $U \subseteq \mathbb{R}^n$. Let $\hat{a}, \hat{x} \in {}^*U$ be the coordinates for $a, x \in {}^*W$.

THEOREM 11.5.3. *The pair (a, x) is a prevector based at the point a if and only if*

$$\hat{x} - \hat{a} \prec \lambda,$$

where the difference $\hat{x} - \hat{a}$ is defined using the linear structure of the ambient vector space ${}^*\mathbb{R}^n \supseteq {}^*U$.

PROOF. Let us show that the two definitions are indeed equivalent. Assume the first definition, and let $(u^1, \ldots, u^n)$ be the chosen coordinate functions. Since each u^i in particular is a smooth function, we get $u^i(x) - u^i(a) \prec \lambda$ for each i, i.e., $\hat{x} - \hat{a} \prec \lambda$.

Conversely, assume that the pair (a, x) satisfies the second definition. Let $f \in \mathbb{D}_{a_0}$ be a smooth function. Then by the mean value theorem (Theorem 11.3.7), we have $f(x) - f(a) = D_c(\hat{x} - \hat{a})$ for a suitable c in the segment between $\hat{a}$ and $\hat{x}$. Now the components of D_c are finite by Lemma 11.2.1. Therefore $f(x) - f(a) \prec \|\hat{x} - \hat{a}\| \prec \lambda$. $\square$

11.6. Tangent space to manifold via prevectors, ivectors

DEFINITION 11.6.1. Let $a \in {}^\natural M$. We denote by $P_a = P_a(M)$ the set of prevectors based at a.

We will use P_a to define the tangent space of M at a. First we will define an equivalence relation $\equiv$ on P_a as follows.

DEFINITION 11.6.2. We write

$$(a, x) \equiv (a, y)$$

if one of the following two equivalent conditions is satisfied:

(1) ${}^*\!f(y) - {}^*\!f(x) \prec\prec \lambda$ for every smooth function $f\colon M \to \mathbb{R}$;
(2) in coordinates as above, $\hat{y} - \hat{x} \prec\prec \lambda$.

The equivalence of the two definitions follows by the argument of Section 11.5.

DEFINITION 11.6.3. We denote by

$$T_a = T_a(M)$$

the set of equivalence classes: $T_a = P_a/\equiv$.

REMARK 11.6.4. This is a generalisation of the ihull (nonstandard hull) construction; see Section 8.2.

DEFINITION 11.6.5. In the spirit of physics notation, we will denote the equivalence class of the pair $(a, x) \in P_a$ by

$$\overrightarrow{ax} \in T_a.$$

We will refer to it as an *ivector* (short for *infinitesimal vector*) so as to distinguish it from a classical vector (which has noninfinitesimal norm by definition).

THEOREM 11.6.6. *A choice of coordinates in a neighborhood of M gives an identification of T_a with $\mathbb{R}^n$ for every nearstandard a.*

PROOF. Given $a \in {}^{\natural}M$, let $W \subseteq M$ be a coordinate neighborhood of the point $\mathbf{st}(a) \in M$. Thus we have a map $W \to \mathbb{R}^n$ with image $U \subseteq \mathbb{R}^n$. Here by definition $x \mapsto \hat{x}$ for all $x \in W$. Recall that $(\mathbb{R}^n)^\lambda_F$ is the space of points in ${}^*\mathbb{R}^n$ that are finite compared to λ. We define a map

$$P_a \to (\mathbb{R}^n)^\lambda_F$$

by sending $(a, x) \mapsto \hat{x} - \hat{a}$. This induces an identification of the space $T_a = P_a/\!\equiv$ with the space $\mathbb{R}^n = (\mathbb{R}^n)^\lambda_F/(\mathbb{R}^n)^\lambda_I$ (see Theorem 10.3.8). Under this identification, the space T_a inherits the structure of a vector space over $\mathbb{R}$.[1] $\square$

REMARK 11.6.7. The entity denoted T_a is a mixture of standard and nonstandard notions. It is a vector space over $\mathbb{R}$, but defined at every nearstandard point $a \in {}^{\natural}M$.[2]

[1]If we choose a different coordinate patch in a neighborhood of $a_0 = \mathbf{st}(a)$ with image $U' \subseteq \mathbb{R}^n$, then if $\varphi\colon U \to U'$ is the change of coordinates, then by Definition 11.3.2, we have $\varphi(\hat{x}) - \varphi(\hat{a}) - J_{a_0}(\hat{x} - \hat{a}) \prec\!\prec \|\hat{x} - \hat{a}\| \prec \lambda$, where J is the Jacobian matrix. Hence we obtain $\varphi(\hat{x}) - \varphi(\hat{a}) - J_{a_0}(\hat{x} - \hat{a}) \prec\!\prec \lambda$. This means that the map $\mathbb{R}^n \to \mathbb{R}^n$ induced by the two identifications of T_a with $\mathbb{R}^n$ provided by the two coordinate maps, is given by multiplication by the matrix J_{a_0}, and therefore is linear. Thus it preserves the linear space structure of the quotient. Hence the vector space structure induced on T_a via coordinates is independent of the choice of coordinates, and so we have a well defined vector space structure on T_a over $\mathbb{R}$.

[2]If $a, b \in {}^{\natural}M$ and $a \approx b$, then given coordinates in a neighborhood of $\mathbf{st}(a)$, the identifications of T_a and T_b with $\mathbb{R}^n$ induced by these coordinates induces an identification between T_a and T_b. Given a different choice of coordinates, the matrix $J_{\mathbf{st}(a)}$ used in the previous paragraph is the same matrix for a and b, and so the identification of T_a with T_b is well defined, independent of a choice of coordinates. Thus when $a \approx b \in {}^{\natural}M$ we may unambiguously add an ivector $\overrightarrow{ax} \in T_a$ and an ivector $\overrightarrow{by} \in T_b$.

CHAPTER 12

Action, differentiation, prevector fields

12.1. Action of a prevector on smooth functions

Prevectors were defined in Section 11.5. They act on functions in a purely algebraic way that does not rely on either limits or standard part, as follows. Recall that $\lambda > 0$ is a fixed infinitesimal.

DEFINITION 12.1.1. A prevector $(a, x) \in P_a$ acts on a smooth function $f\colon M \to \mathbb{R}$ as follows:

$$(a, x)\, f = \frac{1}{\lambda}\big(f(x) - f(a)\big). \tag{12.1.1}$$

Note that the resulting function in the righthand side of (12.1.1) is not a real function in general.

REMARK 12.1.2. The righthand side is finite by definition of prevector.

This action is infinitely close to being a derivation in the following sense.

THEOREM 12.1.3. *The action $(a, x)f$ satisfies the Leibniz rule up to infinitesimals.*

PROOF. Indeed, given real functions f and g we compute the action on the product fg as follows:[1]

$$\begin{aligned}
(a, x)(fg) &= \frac{1}{\lambda}\Big(f(x)g(x) - f(a)g(a)\Big) \\
&= \frac{1}{\lambda}\Big(f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)\Big) \\
&= f(x)\frac{1}{\lambda}\Big(g(x) - g(a)\Big) + \frac{1}{\lambda}\Big(f(x) - f(a)\Big)g(a) \\
&\approx f(a)\frac{1}{\lambda}\Big(g(x) - g(a)\Big) + \frac{1}{\lambda}\Big(f(x) - f(a)\Big)g(a).
\end{aligned}$$

Here the final relation $\approx$ is justified by the continuity of f at $a_0 = \mathbf{st}(a)$ and finiteness of the action $\frac{1}{\lambda}\big(g(x) - g(a)\big)$. □

[1] We suppress the stars on the natural extensions of f and g to simplify the calculations.

12.2. Differentiation by ivectors

We use the action of a prevector to define the differentiation by an ivector as follows.

DEFINITION 12.2.1. The differentiation of a function $f\colon M \to \mathbb{R}$ by an ivector $\overrightarrow{ax} \in T_a$ is defined by setting

$$\overrightarrow{ax}\, f = \mathbf{st}\big((a,x)f\big) \tag{12.2.1}$$

where $(a,x)f$ is the action of Definition 12.1.1.

The action $\overrightarrow{ax}\, f$ is well defined by definition of the equivalence relation $\equiv$. Recall that $T_a = P_a/\equiv$.

REMARK 12.2.2. The ivector $\overrightarrow{ax}$ is a nonstandard object based at a possibly nonstandard point a, but it assigns a standard real number to the standard function f.

THEOREM 12.2.3. *Differentiation by an ivector $\overrightarrow{ax}$ satisfies the following version of the Leibniz rule:*

$$\overrightarrow{ax}(fg) = f(a_0)\cdot \overrightarrow{ax}\, g + \overrightarrow{ax} f \cdot g(a_0)$$

where $a_0 = \mathbf{st}(a)$.

PROOF. Applying Theorem 12.1.3 and the shadow, we obtain the following for the differentiation $\overrightarrow{ax}(fg)$:

$$\begin{aligned}\overrightarrow{ax}(fg) &= \mathbf{st}(f(a))\cdot \overrightarrow{ax}\, g + \overrightarrow{ax} f \cdot \mathbf{st}(g(a))\\ &= f(a_0)\cdot \overrightarrow{ax}\, g + \overrightarrow{ax} f \cdot g(a_0),\end{aligned}$$

where the second equality relies on the continuity of f and g at a_0. □

When $a = a_0 \in M$ we obtain the ordinary Leibniz rule for differentiation.

12.3. Relation to classical vectors

Ivectors can be used in place of classical vectors by the following corollary.

THEOREM 12.3.1. *Each classical tangent vector X at a standard point $a \in M$ defines a unique ivector $\overrightarrow{ax}$ where in coordinates $x = a + \lambda X$.*

PROOF. The classical derivative Xf of f in the direction of vector X at a_0 is by definition $Xf = \mathbf{st}\left(\frac{f(a_0+\lambda X)-f(a_0)}{\lambda}\right)$. Note that

$$\frac{f(a_0 + \lambda X) - f(a_0)}{\lambda} \approx \frac{f(a + \lambda X) - f(a)}{\lambda}$$

by Corollary 11.4.2. Thus Xf coincides with $\overrightarrow{ax}f$ as in formula (12.2.1). While the choice of the point $x = a + \lambda X$ is coordinate-dependent and therefore so is the prevector (a, x), the actions of vector X and ivector $\overrightarrow{ax}$ on $\mathbb{D}_{a_0}$ coincide when viewed as derivations. $\square$

12.4. Induced map on prevectors

Let M be a smooth manifold. We recall the following.

(1) $a \in {}^{\natural}M$ is a nearstandard point.
(2) λ is a fixed infinitesimal.
(3) P_a is the space of prevectors (a, x) satisfying $\hat{x} - \hat{a} \prec \lambda$ in coordinates.
(4) The equivalence relation $\equiv$ between prevectors (a, x) and (a, y) in P_a is defined by requiring in coordinates $\hat{y} - \hat{x} \prec\prec \lambda$ (see Section 11.6).
(5) $T_a = P_a/\!\equiv$ is the tangent space at a.
(6) The ivector $\overrightarrow{ax} \in T_a$ is the equivalence class of a prevector $(a, x) \in P_a$.

Let $f\colon M \to N$ be a smooth map between smooth manifolds with natural extension ${}^*\!f\colon {}^*\!M \to {}^*\!N$, or more generally an internal map $\Phi\colon {}^*\!M \to {}^*\!N$ (such as a prevector field defined in Section 12.5 below).

DEFINITION 12.4.1. Let $a \in {}^{\natural}M$. The *differential* df_a of f is the map

$$df_a\colon P_a \to P_{f(a)}, \quad (a, x) \mapsto (f(a), f(x)).$$

For a standard smooth map f one necessarily has $(f(a), f(x)) \in P_{f(a)}$. For an internal map, additional hypotheses may be required; see Section 13.3.

DEFINITION 12.4.2. Let $f\colon M \to N$ be a smooth map and ${}^*\!f$ its natural extension. Let $a \in {}^{\natural}M$. The *tangent map* Tf_a induced by the differential df_a is the map

$$Tf_a\colon T_a(M) \to T_{f(a)}(N)$$

defined as follows. We choose a prevector (a, x) representing an ivector in T_aM. We then set $Tf_a\left(\overrightarrow{ax}\right) = \overrightarrow{f(a)\ f(x)}$.

Here the ivector on the right-hand side, as the notation suggests, is the equivalence class in $T_{f(a)}$ of the prevector $(f(a), f(x)) \in P_{f(a)}$.

REMARK 12.4.3. The relation between f and df_a, as well as between f and Tf_a seems more transparent here than in the corresponding classical definition.

THEOREM 12.4.4. *For a standard point $a \in M$, the space T_a is naturally identified with the classical tangent space of M at a.*

PROOF. This was shown in Theorem 12.3.1. One can also check this in coordinates. It suffices to show this for an open set $U \subseteq \mathbb{R}^n$, where the classical tangent space at any point a is naturally identified with $\mathbb{R}^n$ itself. A classical vector $v \in \mathbb{R}^n$ is then identified with the ivector $\overrightarrow{ax}$ where $x = a + \lambda \cdot v$. Under this identification, our definition of the differentiation $\overrightarrow{ax}f$ in Section 12.1 coincides with the classical one. A similar remark applies to the tangent map $Tf_a(\overrightarrow{ax})$. $\square$

REMARK 12.4.5. When the manifolds M, N are open subsets of Euclidean space $\mathbb{R}^n$, and T_a is the classical tangent space, the tangent map $Tf_a\colon T_a \to T_{f(a)}$ is identified with the Jacobian matrix J_a (at the point a) whose *rows* are the gradients D_a of each of the n components of f.

12.5. Prevector fields, class D^0

Internal sets were defined in Section 7.5.

DEFINITION 12.5.1. A map $f\colon {}^*M \to {}^*N$ is said to be internal if its graph is an internal subset of ${}^*M \times {}^*N$.

Recall that λ is a fixed infinitesimal. A prevector field is, intuitively, the assignment of an *infinitesimal displacement* (of size comparable to λ) at every point, given by an internal map. More precisely, we have the following definition. Recall that P_a is the set of prevectors at a point $a \in {}^\natural M$, i.e., pairs (a, x) such that in coordinates $\hat{x} - \hat{a} \prec \lambda$.

DEFINITION 12.5.2. A *prevector field* Φ on a smooth manifold M is an *internal* map $\Phi\colon {}^*M \to {}^*M$ such that for every $a \in {}^\natural M$ we have

$$(a, \Phi(a)) \in P_a. \tag{12.5.1}$$

In coordinates where addition and subtraction are possible, this condition translates into $\Phi(a) - a \prec \lambda$ for every $a \in {}^\natural M$.

DEFINITION 12.5.3. The class of prevector fields will be denoted D^0. Thus $\Phi \in D^0$ if and only if $\Phi(a) - a \prec \lambda$ for every $a \in {}^\natural M$.

REMARK 12.5.4. Even though the *condition* (12.5.1) is imposed only at points of ${}^\natural M$, we need the map Φ to be *defined* on all of ${}^*M$. This is because we need Φ (and in particular both its domain and range) to be an *internal* entity, more specifically an element of ${}^*\mathrm{Map}(M, M)$. Being internal enables us to apply hyperfinite iteration which is the foundation of the hyperreal walk; see Section 12.9.

DEFINITION 12.5.5. Let $\Phi, G \in D^0$. We say Φ is equivalent to G and write $\Phi \equiv G$ if one of the following three equivalent conditions is satisfied for every $a \in {}^\natural M$:

(1) $(a, \Phi(a)) \equiv (a, G(a))$,
(2) in coordinates $\Phi(a) - G(a) \prec\prec \lambda$,
(3) $\overrightarrow{a\,\Phi(a)} = \overrightarrow{a\,G(a)}$.

12.6. Local prevector fields

DEFINITION 12.6.1. A *local* prevector field on an open $U \subseteq M$ is an internal map $\Phi\colon {}^*U \to {}^*V$ satisfying the conditions of Definition 12.5.5, where the set $V \supseteq U$ is open.

When the distinction is needed, we will call a prevector field defined on all of ${}^*M$ a *global* prevector field.

REMARK 12.6.2. Allowing the values of a local prevector field defined on ${}^*U$ to lie in a possibly larger range ${}^*V$ enables us to work with a restriction of a global prevector field Φ to a smaller domain which is not necessarily invariant under Φ.

EXAMPLE 12.6.3. Let $M = \mathbb{R}$. If we wish to restrict the prevector field Φ on M given by $\Phi(a) = a + \lambda$ to the domain ${}^*(0, 1) \subseteq {}^*\mathbb{R}$, then we need the flexibility of allowing a slightly larger range for Φ.

In the sequel we will usually not mention the slightly larger range V when describing a local prevector field, but will tacitly assume that we have such a V when needed. A second instance where it may be required for the range to be slightly larger than the domain is the following natural setting for defining a local prevector field.

EXAMPLE 12.6.4. Let $p \in M$. Let $V \subseteq M$ a coordinate neighborhood of p with image $V' \subseteq \mathbb{R}^n$ where $\hat{p} \in V'$ corresponds to $p \in V$. Let X be a classical vector field on V, given in coordinates by $X'\colon V' \to \mathbb{R}^n$. Thus if in coordinates one has $X = X^i \frac{\partial}{\partial u^i}$ then X' is given by the n-tuple $(X^1, \ldots, X^n)$. Then there is a neighborhood U' of $\hat{p} \in \mathbb{R}^n$, with $U' \subseteq V'$, such that we can define a local prevector field $\Phi'\colon {}^*U' \to {}^*V'$ by setting

$$\Phi'(a) = a + \lambda \cdot X'(a). \tag{12.6.1}$$

Indeed, it suffices to choose U' such that $\overline{U'}$ is compact and furthermore $\overline{U'} \subseteq V'$. For the corresponding open set $U \subseteq V$ in M, this induces a local prevector field $\Phi\colon {}^*U \to {}^*V$ which realizes the classical vector field X on U in the sense of Definition 12.7.1 below (cf. Theorem 12.4.4).

Remark 12.6.5. Realizing vector field (12.6.1) of Example 12.6.4 may require restricting to a smaller neighborhood U, as discussed in Example 12.6.3. For example, when $M = V = V' = (0,1)$ and $X = 1\frac{d}{dx}$ in classical notation, one needs to take $U = (0,r)$ for some $r \in \mathbb{R}$, $r < 1$ in order for $\Phi(a) = a + \lambda$ to lie always in ${}^*V$.

12.7. Realizing classical vector fields

Let U be an open neighborhood in the smooth manifold M. How does a prevectorfield realize[2] a classical vector field?

Definition 12.7.1. We say that a local prevector field Φ on U *realizes* a classical vector field X on U if for every smooth function $h\colon U \to \mathbb{R}$ we have the following relation for the new function Xh:

$$Xh\,(a) = \overrightarrow{a\ \Phi(a)}\,h$$

for all $a \in U$, where $\overrightarrow{a\ \Phi(a)}$ is the ivector represented by the prevector $(a, \Phi(a))$.

Remark 12.7.2. The condition of Definition 12.7.1 involves only *standard* points a.

Different coordinates for the same neighborhood U will induce equivalent realizations in ${}^*U$ as in Proposition 12.7.4 below. First we state the following classical result on coordinate change.

Lemma 12.7.3. *Let $U \subseteq \mathbb{R}^n$ and let $a \in U$. Let $X\colon U \to \mathbb{R}^n$ be a classical vector field. Let $\sigma\colon U \to W \subseteq \mathbb{R}^n$ be a (smooth) change of coordinates. Thus $T\sigma$ sends T_a to $T_{\sigma(a)}$. Let $Y\colon W \to \mathbb{R}^n$ be the classical vector field corresponding to X under the coordinate change. Then*

$$Y(\sigma(a)) = J_a X(a), \tag{12.7.1}$$

where J_a is the Jacobian matrix of σ at a.

In Example 12.6.4 we associated a prevector field to a classical vector field by setting $\Phi(a) = a + \lambda X(a)$. We now show that different coordinates for the same neighborhood U will induce realizations that are *equivalent* in the sense of the relation $\equiv$.

[2]Memamesh

PROPOSITION 12.7.4. *Consider a change of coordinates* σ *as in Lemma* 12.7.3*, so that we have* $Y(\sigma(a)) = J_a X(a)$*. Let* Φ*,* G *be the prevector fields given by*

(1) $\Phi(a) = a + \lambda X(a)$,
(2) $G(a) = a + \lambda Y(a)$

Then we have an equivalence $\Phi \equiv \sigma^{-1} \circ G \circ \sigma$*, or in coordinates,* $\sigma \circ \Phi(a) - G \circ \sigma(a) \prec\prec \lambda$ *for all* $a \in {}^{\natural}U$*.*

PROOF. Let σ^i, Y^i, G^i be the ith component of σ, Y, G respectively, and let D_a^i be the gradient (see Remark 11.3.3) of σ^i at a, so that D_a^i is the ith row of J_a. We apply the mean value theorem to σ^i together with (12.7.1) to obtain the following bound for a suitable point c in the segment between a and $a + \lambda X(a)$:

$$[l]\sigma^i \circ \Phi(a) - G^i \circ \sigma(a) = \sigma^i(a + \lambda X(a)) - G^i \circ \sigma(a)$$
$$= \sigma^i(a + \lambda X(a)) - \sigma^i(a) - \lambda Y^i(\sigma(a))$$
$$= D_c^i \lambda X(a) - \lambda D_a^i X(a)$$
$$= \lambda(D_c^i - D_a^i) X(a)$$
$$\prec\prec \lambda,$$

where D_c^i is the gradient of σ^i at c. Such a point c exists by Theorem 11.3.7. Since the bound $\sigma^i \circ \Phi(a) - G^i \circ \sigma(a) \prec\prec \lambda$ holds for each component i, we obtain the bound $\sigma \circ \Phi(a) - G \circ \sigma(a) \prec\prec \lambda$, proving the proposition. □

12.8. Internal induction

Internal induction is treated in most textbooks on Robinson's framework. We will follow [**Goldblatt 1998**, p. 129].

THEOREM 12.8.1. *If* $X \subseteq {}^*\mathbb{N}$ *is an internal subset containing the element* $1 \in {}^*\mathbb{N}$ *and closed under the successor function*[3] $n \mapsto n+1$ *then* $X = {}^*\mathbb{N}$*.*

PROOF. We argue by contradiction. Consider the set difference $Y = {}^*\mathbb{N} \setminus X$. Suppose Y is nonempty. since Y is internal, it has a least element $n \in Y$ (see Section 7.7). Hence $n - 1 \in X$. However, by closure under successor, the number n must also be in X, contradicting the hypothesis. This proves the theorem. □

EXAMPLE 12.8.2 (Counterexample). The set $\mathbb{N}$ contains 1, is closed under successor, but is different from ${}^*\mathbb{N}$. Thus the hypothesis that Y be internal is indispensable.

[3] "peulat ha'okev" according to hebrew wikipedia at peano axioms

Internal induction will be used in Section 12.9 and in the proof of Theorem 15.7.2.

12.9. Global prevector fields, walks, and flows

In this section we will define the hyperreal walk[4] of a prevector field and relate it to the classical flow. Recall that a prevector field Φ on M is an internal map $\Phi\colon {}^*\!M \to {}^*\!M$ from ${}^*\!M$ to itself.

REMARK 12.9.1. We wish to view a prevector field as the *first step* of the corresponding walk at time $t = \lambda$, and will define the hyperreal walk Φ_t and the real flow ϕ_t accordingly.

The hyperreal walk at arbitrary time t is defined by iterating Φ the appropriate number of times.

DEFINITION 12.9.2. Recall that $\Phi \in {}^*\mathrm{Map}(M)$. Consider the map $\mathrm{Map}(M)\times\mathbb{N} \to \mathrm{Map}(M)$ taking (ϕ, n) to $\phi^n = \phi\circ\phi\circ\cdots\circ\phi$ (composition altogether n times). Let ${}^*\mathrm{Map}(M) \times {}^*\mathbb{N} \to {}^*\mathrm{Map}(M)$ be its natural extension. This defines Φ^n for all hypernatural n.

We will now define the hyperreal walk Φ_t of a prevector field Φ. The flow is initially defined at times $t \in {}^*\mathbb{R}$ that are hyperinteger multiples of the base infinitesimal λ.

DEFINITION 12.9.3. Let $\Phi\colon {}^*\!M \to {}^*\!M$ be a global prevector field. For each hypernatural $n \in {}^*\mathbb{N}$, let $t = n\lambda$ and define the *hyperreal walk* Φ_t of Φ at time t by setting

$$\Phi_t(a) = \Phi^n(a). \tag{12.9.1}$$

DEFINITION 12.9.4. The *real flow* ϕ_t on M is $\phi_t(x) = \mathbf{st}\big(\Phi_{n\lambda}(x)\big)$ for all $x \in M$, where $n\lambda \le t < (n+1)\lambda$, i.e., $n = \left\lfloor \frac{t}{\lambda} \right\rfloor$.

REMARK 12.9.5. Under suitable hypotheses for Φ, the real flow ϕ_t will be defined at real t whenever t is sufficiently small so that the standard part is well-defined; see material around (15.9.1) for the details.

REMARK 12.9.6. The idea of a vector field being the *infinitesimal generator* of its flow receives a literal meaning in the TIDG setting.

[4]Hiluch

CHAPTER 13

Invariance, commutation

13.1. Proof of invariance

Let $\Phi\colon {}^*\!M \to {}^*\!M$ be a prevector field on M. Recall that the hyperreal walk is defined by setting $\Phi_t = \Phi^{\circ N}$ where $N = \lfloor \frac{t}{\lambda} \rfloor$. A prevector field Φ (not only *can be thought of as* but actually) *is* the infinitesimal generator of both the walk Φ_t and the flow ϕ_t as discussed in Section 12.9.

THEOREM 13.1.1. *Let Φ be a prevector field on M, and let Φ_t be its hyperreal walk as in* (12.9.1)*. Then the prevector field Φ is invariant under the differential of the map Φ_t, in the sense that $d\Phi_t\big(a, \Phi(a)\big) = (b, \Phi(b))$ where $b = \Phi(a)$.*

PROOF. Note that we have the relation

$$(\forall n \in {}^*\mathbb{N}) \quad \Phi^n \circ \Phi = \Phi \circ \Phi^n,$$

proved by internal induction (see Section 12.8). Now let $b = \Phi^n(a)$. The differential $d\Phi_t$ of the map $\Phi_t = \Phi^n\colon {}^*\!M \to {}^*\!M$ sends P_a to P_b. By Definition 12.4.1, it acts on each prevector $(a, \Phi(a))$ by

$$\begin{aligned} d\Phi_t(a, \Phi(a)) &= (\Phi^n(a), \Phi^n(\Phi(a))) \\ &= (\Phi^n(a), \Phi(\Phi^n(a))) \\ &= (b, \Phi(b)) \end{aligned}$$

by the associativity of composition, proving the invariance of the prevector field Φ. □

REMARK 13.1.2. This proof of invariance compares favorably to the proof of invariance under the flow in the A-track framework (see Theorem 9.2.3).

REMARK 13.1.3. For any *global* prevector field Φ, the hyperreal walk Φ_t is defined for all $t \geq 0$. This is of course not the case for the classical flow of a classical vector field. Similarly Φ_t is defined for all $t \leq 0$ if Φ is bijective.

A treatment of Lie brackets via infinitesimal displacements appears in Section 15.15. A treatment of commutation via infinitesimal displacements appears in Section 15.23. A summary appears in Section 13.5.

13.2. Hyperreal walk of a local prevector field

The hyperreal walk for a *local* prevector field is defined similarly. A bit of care is required to specify its domain.

Definition 13.2.1. Let $U \subseteq V$. Consider a local prevector field $\Phi\colon {}^*U \to {}^*V$. We extend Φ to an internal map $\Phi'\colon {}^*V \to {}^*V$ by defining $\Phi'(a) = a$ for all $a \in {}^*V - {}^*U$.

Remark 13.2.2. In Section 13.3 we will define classes D^k. If Φ is D^k on ${}^\natural U$, the extended field Φ' as in Definition 13.2.1 is internal but it may not be D^k on all of ${}^\natural V$. Therefore care needs to be taken in specifying the domain of the prevector field we wish to iterate.

Definition 13.2.3. We set

$$Y_n = \{a \in {}^*U\colon (\Phi')^n(a) \in {}^*U\}.$$

The domain of Φ_t is the set $Y_{n(t)} \subseteq {}^*U$, where $n(t) = \left\lfloor \frac{t}{\lambda} \right\rfloor$. The walk Φ_t is defined on $Y_{n(t)}$ by setting $\Phi_t(a) = \Phi'_t(a)$.

We wish also to consider the flow Φ_t for $t \leq 0$. This can be defined for a global prevector field Φ which is bijective on ${}^*M$, or for a local prevector field which is bijective in the sense of Remark 15.13.1, in particular a D^1 local prevector field. Namely, we define Φ_t for $t \leq 0$ to be $(\Phi^{-1})_{-t}$.

13.3. Regularity class D^1 for prevector fields

We present a *synthetic* approach to vector fields where instead of using the *analytic* classes C^k, we will use combinatorially defined classes denoted D^k.

Defining various operations on prevector fields, such as their flow, or Lie bracket, requires suitable regularity properties. To motivate the definition of the class D^1, we first recall the following classical definition.

Definition 13.3.1. A classical vector field $X\colon U \to \mathbb{R}^n$ is called *K-Lipschitz*, where $K \in \mathbb{R}$, if

$$\|X(a) - X(b)\| \leq K\|a - b\| \text{ for all } a, b \in U. \tag{13.3.1}$$

For the local prevector field Φ of Example 12.6.4, where $\Phi(a)-a=\lambda X(a)$, this translates into the inequality

$$\left\| \big(\Phi(a)-a\big)-\big(\Phi(b)-b\big) \right\| \le K\lambda\|a-b\| \tag{13.3.2}$$

for $a,b\in {}^*U$.

REMARK 13.3.2. Note the presence of the factor of λ on the right hand side of formula (13.3.2) which is not present in formula (13.3.1). This is due to the fact that (13.3.2) deals with infinitesimal prevectors of type $(a,\Phi(a))$ rather than classical vectors.

The bound (13.3.2) motivates the following definition.

DEFINITION 13.3.3. A prevector field Φ on a smooth manifold M is of class D^1 if whenever $a,b\in {}^\natural M$ and $(a,b)\in P_a$, in coordinates the following condition is satisfied:

$$\Phi(a)-a-\Phi(b)+b\prec\lambda\|a-b\|. \tag{13.3.3}$$

13.4. Second differences, class D^2

Working with Lie brackets (more precisely, ibrackets) of prevector fields requires a stronger condition than D^1. Second differences were already discussed in Section 4.5 in the context of the second derivative.

DEFINITION 13.4.1. Let v,w be infinitesimal vectors such that $v\prec\lambda$ and $w\prec\lambda$. The *second difference* of a prevector field Φ is defined by $\Delta^2_{v,w}\Phi(a)=\Phi(a)-\Phi(a+v)-\Phi(a+w)+\Phi(a+v+w)$.

DEFINITION 13.4.2. A prevector field Φ on a smooth manifold M is of class D^2 if for each $a\in {}^\natural M$, the following condition is satisfied in coordinates. For each pair $v,w\in {}^*\mathbb{R}^n$ with $v\prec\lambda$, $w\prec\lambda$, the *second difference* $\Delta^2_{v,w}\Phi(a)$ is sufficiently small:

$$\Delta^2_{v,w}\Phi(a)\prec\lambda\|v\|\|w\|.$$

In Proposition 15.2.1 we will show that the prevector field of Example 12.6.4, induced by a classical vector field of class C^k, is necessarily a D^k prevector field ($k=1,2$). This is in fact the central motivation for our definition of the class D^k.

REMARK 13.4.3. Our condition D^k is in fact a weaker condition than C^k. Thus, the prevectorfield Φ of Example 12.6.4 is D^1 if X is Lipschitz, which is weaker than C^1.

Our Definition 12.5.5 of the class D^0 is consistent with the above as it consists in requiring $\Phi(a)-a\prec\lambda$, i.e., Φ being a prevector field. Note that in our definitions above, being a prevector field is part of the definition of an entity in D^k.

REMARK 13.4.4. The class D^2 enables us to define an infinitesimal version of the Lie bracket as in Section 13.5.

13.5. Lie ibrackets, commutation

Lie ibrackets are dealt with in detail in Section 15.15 and Section 15.23. Let $N = \left\lfloor \frac{1}{\lambda} \right\rfloor$. Here we will provide a summary of the proof of commutation.

DEFINITION 13.5.1. The *Lie ibracket* $\dot{[}\Phi, G\dot{]}$ of prevector fields Φ and G is $\dot{[}\Phi, G\dot{]} = (G^{-1} \circ \Phi^{-1} \circ G \circ \Phi)^{\circ N}$.

(See Definition 15.15.1). The following theorem corresponds to the classical fact that the bracket of two vector fields vanishes if and only if their flows commute. The identity prevector field I is defined by $I(a) = a$ for all a.

THEOREM 13.5.2. *Let Φ, G be two D^2 prevector fields on M, and let ϕ_t, g_s be the associated real flows. Then we have $\dot{[}\Phi, G\dot{]} \equiv I$ if and only if $\phi_t \circ g_s = g_s \circ \phi_t$ for all $0 \le t, s \le T$ for some $0 < T \in \mathbb{R}$.*

(See Theorem 15.23.1). The idea of the proof is as follows. We partition the parameter rectangle $[0, t] \times [0, s]$ into N^2 infinitesimal rectangles. The composition $\Phi_t \circ G_s$ corresponds to following $2N$ infinitesimal displacements along the left and top sides of the parameter rectangle. Meanwhile, the composition $G_s \circ \Phi_t$ corresponds to following $2N$ infinitesimal displacements following the bottom and right sides of the rectangle. We proceed as follows.

(1) We deform left-top one path into the bottom-right path by a sequence of N^2 *elementary moves*.[1] Here an elementary move switches from left and top sides of an infinitesimal rectangle to bottom and right sides of the same rectangle.
(2) We show that if $\dot{[}\Phi, G\dot{]} \equiv I$ then each elementary move results in a change of size $\prec\prec \frac{1}{N^2}$ in the final outcome.
(3) Therefore the total change is $\prec\prec 1$.
(4) Therefore compositions of the hyperreal walks satisfy $\Phi_t \circ G_s \equiv G_s \circ \Phi_t$.
(5) Passing to the standard part we obtain that the compositions of the real flows satisfy $\phi_t \circ g_s = g_s \circ \phi_t$, as required.

[1]chiluf elementari.

13.6. Bounds for D^1 prevector fields

Recall that a D^1 prevector field Φ satisfies the condition

$$\Phi(a) - a - \Phi(b) + b \prec \lambda\|a - b\|$$

whenever $(a, b) \in P_a$ (see Section 13.3). We note that Φ satisfies the following adequality $\sqcap$ relation (see Section 3.11).

Proposition 13.6.1. *Let Φ be a D^1 prevector field on M. Assume $a, b \in {}^{\natural}M$ satisfy $a - b \prec \lambda$. Then $\|\Phi(a) - \Phi(b)\| \sqcap \|a - b\|$.*

Proof. By the triangle inequality we have

$$\begin{aligned} \big|\|\Phi(a) - \Phi(b)\| - \|a - b\|\big| &\le \|\Phi(a) - \Phi(b) - a + b\| \\ &= K\lambda\|a - b\| \end{aligned} \tag{13.6.1}$$

where the hyperreal constant K defined by the last equality in formula (13.6.1) is finite by the D^1 condition. Hence

$$-K\lambda\|a - b\| \le \|\Phi(a) - \Phi(b)\| - \|a - b\| \le K\lambda\|a - b\|.$$

Rearranging terms, we obtain

$$(1 - K\lambda)\|a - b\| \le \|\Phi(a) - \Phi(b)\| \le (1 + K\lambda)\|a - b\|,$$

Since λ is infinitesimal, we have $1 \pm K\lambda \sqcap 1$, proving the result. □

Corollary 13.6.2. *Under the hypotheses of Proposition* 13.6.1 *we have $\|a - b\| \prec \|\Phi(a) - \Phi(b)\| \prec \|a - b\|$.*

13.7. Operations on prevector fields

Addition of D^1 prevector fields in coordinates is equivalent, in the sense of the relation $\equiv$, to their composition. More precisely, we show the following. Ivectors $\overrightarrow{a\,x}$ were defined in Section 11.6.

Proposition 13.7.1. *Assume Φ be a D^1 prevector field on M and let $G \in D^0$. Then*

(1) *for every $a \in {}^{\natural}M$, we have $\overrightarrow{a\ \Phi(G(a))} = \overrightarrow{a\ \Phi(a)} + \overrightarrow{a\ G(a)}$;*
(2) *if G is also D^1 then $\Phi \circ G \equiv G \circ \Phi$.*

Proof. In a coordinate chart, let

$$x = a + (\Phi(a) - a) + (G(a) - a) = \Phi(a) + G(a) - a. \tag{13.7.1}$$

Passing to equivalence classes, we obtain the following relation among ivectors: $\overrightarrow{a\ \Phi(a)} + \overrightarrow{a\ G(a)} = \overrightarrow{a x}$. Now let $b = G(a)$. Since G is a prevector field we have $(a, b) \in P_a$. From formula (13.7.1) we obtain

$$\Phi(G(a)) - x = \Phi(b) - \Phi(a) - b + a \prec \lambda\|b - a\| \prec\prec \lambda$$

by the D^1 condition on Φ, proving part (1). Finally, part (2) results from part (1) since

$$\overrightarrow{a\Phi(a)} + \overrightarrow{aG(a)} = \overrightarrow{aG(a)} + \overrightarrow{a\Phi(a)} = \overrightarrow{aG(\Phi(a))},$$

as required. □

DEFINITION 13.7.2. We define the notation of the increment δ_Φ by setting $\delta_\Phi(a) = \Phi(a) - a$.

LEMMA 13.7.3. *The prevector field defined by the sum of the increments of two D^1 prevector fields is a D^1 prevector field.*

PROOF. Let Φ and G be D^1 prevector fields. The D^1 condition can be expressed as $\delta_\Phi(a) - \delta_\Phi(b) \prec \lambda\|a-b\|$. Therefore

$$\begin{aligned}\delta_{\Phi+G}(a) - \delta_{\Phi+G}(b) &= \delta_\Phi(a) + \delta_G(a) - \delta_\Phi(b) - \delta_G(b) \\ &= (\delta_\Phi(a) - \delta_\Phi(b)) + (\delta_G(a) - \delta_G(b)) \\ &\prec \lambda\|a-b| + \lambda\|a-b| \\ &\prec \lambda\|a-b\|\end{aligned}$$

proving that the prevector field $a \mapsto a + \delta_\Phi + \delta_G$ is D^1. □

REMARK 13.7.4. By the previous proposition, composition is equivalent to addition. From the previous lemma we deduce that the prevector field $a \mapsto a + \delta_\Phi + \delta_G$ is D^1 but the argument shows only that the composition of two D^1 prevector fields is equivalent to a D^1 prevector field, which does not necessarily imply that it is itself D^1; see the example on page 29 line -8 of [**Nowik & Katz 2015**]. Therefore a separate argument is needed to prove Lemma 13.7.3.

We will now show that the composition of D^1 prevector fields is again D^1.

PROPOSITION 13.7.5. *If prevector fields Φ, G are D^1 then the composition prevector field $\Phi \circ G$ is D^1.*

PROOF. Let $c = G(a)$ and $d = G(b)$. We have by the triangle inequality

$$\begin{aligned}&\|\Phi\circ G(a) - \Phi\circ G(b) - a + b\| \\ &\quad \le \|\Phi\circ G(a) - \Phi\circ G(b) - G(a) + G(b)\| + \|G(a) - G(b) - a + b\| \\ &\quad = \|\Phi(c) - \Phi(d) - c + d\| + \|G(a) - G(b) - a + b\| \\ &\quad \prec \lambda\|c-d\| + \lambda\|a-b\| \\ &\quad \prec \lambda\|a-b\|\end{aligned}$$

where Corollary 13.6.2 is applied in order to bound $c - d$ in terms of $a - b$. □

CHAPTER 14

Application: small oscillations of a pendulum

The material in this chapter first appeared in *Quantum Studies: Mathematics and Foundations* [**Kanovei et al. 2016**].

14.1. Small oscillations of a pendulum

In his 1908 book *Elementary Mathematics from an Advanced Standpoint*, Felix Klein advocated the introduction of calculus into the high-school curriculum. One of his arguments was based on the problem of small oscillations of the pendulum.[1] The problem had been treated until then using a somewhat mysterious *superposition principle*. The latter involves (a vertical plane projection of) a hypothetical circular motion of the pendulum. Klein advocated what he felt was a better approach, involving the differential equation of the pendulum; see [**Klein 1908**, p. 187].

The classical problem of the pendulum translates into a second order nonlinear differential equation

$$\ddot{x} = -\frac{g}{\ell}\sin x$$

for the variable angle x with the vertical direction, where g is the constant of gravity and ℓ is the length of the (massless) rod or string.

Remark 14.1.1. The problem of small oscillations deals with the case of small amplitude,[2] i.e., x is small, so that $\sin x$ is approximately x. Then the equation is boldly replaced by the linear one

$$\ddot{x} = -\frac{g}{\ell}x,$$

whose solution is harmonic motion with period $2\pi\sqrt{\ell/g}$ independent of the amplitude a (here x ranges through the interval $-a \leq x \leq a$).

This suggests that the period of small oscillations should be independent of their amplitude. The intuitive solution outlined above may be acceptable to a physicist, or at least to the mathematicians'

[1] Tnudat metutelet

[2] misraat

proverbial physicist. The solution Klein outlined in his book does not go beyond the physicist's solution (Klein goes on to derive an exact solution of the nonlinear equation, from which of course one can also derive the correct asymptotics for small oscillations).

REMARK 14.1.2. The Hartman–Grobman theorem [**Hartman 1960**], [**Grobman 1959**] provides a criterion for the flow of the nonlinear system to be conjugate to that of the linearized system, under the hypothesis that the linearized matrix has no eigenvalue with vanishing real part. However, the hypothesis is not satisfied for the pendulum problem.

To provide a rigorous mathematical treatment for the notion of *small oscillation*, it is tempting to exploit a hyperreal framework following [**Nowik & Katz 2015**]. Here the notion of small oscillation can be given a precise sense, namely infinitesimal amplitude. Note however the following.

REMARK 14.1.3. Even for infinitesimal x, one cannot boldly replace $\sin x$ by x.

Therefore additional arguments are required. The linearisation of the pendulum is treated in [**Stroyan 2015**] using Dieners' "Short Shadow" Theorem; see Theorem 5.3.3 and Example 5.3.4 there. This chapter can be viewed as a self-contained treatement of Stroyan's Example 5.3.4.

The traditional A-track setting in the context of the real continuum can only make sense of the claim that "the period of small oscillations is independent of the amplitude" by means of a paraphrase in terms of limits, rather than a specific oscillation. In the context of an infinitesimal-enriched continuum, such a claim can be formalized more literally; see Corollary 14.8.1.

REMARK 14.1.4. The breakdown of infinite divisibility at quantum scales makes irrelevant the mathematical definitions of derivatives and integrals in terms of limits as x tends to zero. Rather, quotients like $\frac{\Delta y}{\Delta x}$ need to be taken in a certain range, or level. In our article [**Nowik & Katz 2015**] we developed a general framework for differential geometry at level λ, where λ is an infinitesimal but the formalism is a better match for a situation where infinite divisibility fails and a scale for calculations needs to be fixed accordingly. Here we implement such an approach to give a rigorous account "at level λ" for small oscillations of the pendulum.

14.2. Vector fields and infinitesimal displacements

The framework developed in [**Robinson 1966**] involves a proper extension $\mathbb{R} \hookrightarrow {}^*\mathbb{R}$ preserving the properties of $\mathbb{R}$ to a large extent discussed in Remark 14.3.3. Elements of ${}^*\mathbb{R}$ are called hyperreal numbers. A positive hyperreal number is called *infinitesimal* if it is smaller than every positive real number.

We choose a fixed positive infinitesimal $\lambda \in {}^*\mathbb{R}$ (a restriction on the choice of λ appears in Section 14.8).

REMARK 14.2.1. We will work with flows in a two-dimensional context where the manifold can be conveniently identified with $\mathbb{C}$. We will also exploit the natural extension $\mathbb{C} \hookrightarrow {}^*\mathbb{C}$.

DEFINITION 14.2.2. Given a classical vector field $V = V(z)$ where $z \in \mathbb{C}$, one forms an *infinitesimal displacement* δ_Φ (of the intended prevectorfield Φ; see definition below) by setting $\delta_\Phi = \lambda V$.

Note that the infinitesimal displacement can also be a more general internal function.

Thus the aim is to construct the corresponding hyperreal walk, denoted Φ_t, in the plane. Note that a zero of δ_Φ corresponds to a fixed point of the flow.

DEFINITION 14.2.3. The *infinitesimal generator* of the flow (or walk) is the function $\Phi\colon {}^*\mathbb{C} \to {}^*\mathbb{C}$, also called a *prevector field*, defined by

$$\Phi(z) = z + \delta_\Phi(z), \tag{14.2.1}$$

where $\delta_\Phi(z) = \lambda V(z)$ in the case of a displacement generated by a classical vector field as above.

REMARK 14.2.4. An infinitesimal generator, or prevector field, could be a more general internal function. See Section 7.5 and Chapter 10.2 for details on internal sets, and Definition 13.3.3 for the D^1 condition for prevectorfields.

14.3. The hyperreal walk and the real flow

We propose a concept of solution of differential equation based on Euler's method with infinitesimal step, with well-posedness based on a property of adequality (see Section 14.4), as follows.

DEFINITION 14.3.1. The *hyperreal walk*, $\Phi_t(z)$ is a t-parametrized map ${}^*\mathbb{C} \to {}^*\mathbb{C}$ defined whenever t is a hypernatural multiple $t = N\lambda$ of λ, by setting

$$\Phi_t(z) = \Phi_{N\lambda}(z) = \Phi \circ \Phi \circ \ldots \circ \Phi(z) = \Phi^{\circ N}(z), \tag{14.3.1}$$

where $\Phi^{\circ N}$ is the N-fold composition.

Such hyperfinite compositions are discussed in more detail in Section 12.9. For arbitrary hyperreal t the walk can be defined by setting

$$\Phi_t = \Phi_{N\lambda} = \Phi^{\circ N}$$

where $N = \left\lfloor \frac{t}{\lambda} \right\rfloor$ and $\lfloor \ \rfloor$ is the floor function. Recall that the (natural extension of the) floor function $\lfloor x \rfloor$ rounds off the hyperreal number x to the nearest integer no greater than x.

REMARK 14.3.2. The fact that the infinitesimal generator Φ given by (14.2.1) is invariant under the flow Φ_t of (14.3.1) receives a transparent meaning in this framework, expressed by the commutation relation $\Phi \circ \Phi^{\circ N} = \Phi^{\circ N} \circ \Phi$ due to *transfer* of associativity of composition of maps (for an argument using internal induction see Section 12.8).

REMARK 14.3.3. The *transfer principle* is a type of theorem that, depending on the context, asserts that rules, laws or procedures valid for a certain number system, still apply (i.e., are *transfered*) to an extended number system, as discussed in Section 3.3; see there for some illustrative examples. For a more detailed discussion, see Chapter 16.

The standard part (or shadow) **st** rounds off each finite hyperreal to its nearest real number. The standard part function naturally extends from the finite part of $^*\mathbb{R}$ to that of $^*\mathbb{C}$ (e.g., componentwise).

DEFINITION 14.3.4. The *real flow* ϕ_t on $\mathbb{C}$ for $t \in \mathbb{R}$ when it exists is constructed as the shadow (i.e., standard part) of the hyperreal walk Φ_t by setting

$$\phi_t(z) = \mathbf{st}\left(\Phi_{N\lambda}(z)\right)$$

where $N = \left\lfloor \frac{t}{\lambda} \right\rfloor$.

For t sufficiently small, appropriate regularity conditions ensure that the point $\Phi_{N\lambda}(z)$ is finite so that the shadow is well-defined.

The usual relation of being infinitely close is denoted $\approx$. Thus for finite (i.e., non-infinite) z, w we have $z \approx w$ if and only if $\mathbf{st}(z) = \mathbf{st}(w)$. This relation is an additive one.

The appropriate relation for working with small prevector fields is a multiplicatively invariant one rather than an additively invariant one, as detailed in Section 14.4.

14.4. Adequality $\sqcap$

The following is a generalisation to the complex setting of the definition that appeared in Section 3.11.

Definition 14.4.1. Let $z, w \in {}^*\mathbb{C}$. We say that z and w are *adequal* and write $z \sqcap w$ if either one has $\frac{z}{w} \approx 1$ (i.e., $\frac{z}{w} - 1$ is infinitesimal) or $z = w = 0$.

This implies in particular that the angle between z, w is infinitesimal (when they are nonzero), but the relation $\sqcap$ entails a stronger condition. If one of the numbers is appreciable, then so is the other and the relation $z \sqcap w$ is equivalent to $z \approx w$. If one of z, w is infinitesimal then so is the other, and the difference $|z - w|$ is not merely infinitesimal, but so small that the quotients $|z - w|/z$ and $|z - w|/w$ are infinitesimal, as well.

We are interested in the behavior of orbits in a neighborhood of a fixed point 0, under the assumption that the infinitesimal displacement satisfies the Lipschitz condition. In such a situation, we have the following theorem. Recall that the D^1 condition was specified in Definition 13.3.3.

Theorem 14.4.2. *D^1 prevector fields defined by adequal infinitesimal displacements define hyperreal walks that are adequal at each finite time.*

Thus, if $\delta_\Phi \sqcap \delta_G$ then $\Phi_t \sqcap G_t$. It follows in particular that $\phi_t = g_t$ where ϕ_t and g_t are the corresponding real flows. This was shown in [**Nowik & Katz 2015**, Example 5.12].

14.5. Infinitesimal oscillations of the pendulum

Let x denote the variable angle between an oscillating pendulum and the downward vertical direction. By considering the projection of the force of gravity in the direction of motion, one obtains the equation of motion $m\ell\ddot{x} = -mg\sin x$ where m is the mass of the bob of the pendulum, ℓ is the length of its massless rod, and g is the constant of gravity. Thus we have a second-order nonlinear differential equation

$$\ddot{x} = -\tfrac{g}{\ell}\sin x. \tag{14.5.1}$$

The initial condition of releasing the pendulum at angle a (for *amplitude*) is described by

$$\begin{cases} x(0) = a \\ \dot{x}(0) = 0 \end{cases}$$

We replace the second-order equation (14.5.1) by the pair of first order equations

$$\begin{cases} \dot{x} = \sqrt{\tfrac{g}{\ell}}\, y, \\ \dot{y} = -\sqrt{\tfrac{g}{\ell}}\sin x, \end{cases}$$

and initial condition $(x, y) = (a, 0)$. We identify (x, y) with $x + iy$ and $(a, 0)$ with $a + i0$ as in Section 14.2. The classical vector field corresponding to this system is then

$$V(x,y) = \sqrt{\tfrac{g}{\ell}}\, y - \left(\sqrt{\tfrac{g}{\ell}}\sin x\right) i \tag{14.5.2}$$

(with a zero at the origin that turns out to be of circulation type; see Example 2.5.3). The corresponding prevector field (i.e., infinitesimal generator of the flow) Φ is defined by the infinitesimal displacement

$$\delta_\Phi = \left(\lambda\sqrt{\tfrac{g}{\ell}}\, y\right) - \left(\lambda\sqrt{\tfrac{g}{\ell}}\sin x\right) i$$

so that

$$\Phi(z) = z + \delta_\Phi(z).$$

We are interested in the hyperfinite walk of Φ, with initial condition $z_0 = a + 0i$. The flow is generated by hyperfinite iteration of Φ.

14.6. Linearized walk

Consider also a prevector field $E(z) = z + \delta_E(z)$ defined by the displacement

$$\delta_E(z) = \lambda\sqrt{\tfrac{g}{\ell}}\, y - i\lambda\sqrt{\tfrac{g}{\ell}}\, x$$

where as before $z = x + iy$. We are interested in small oscillations, i.e., the case of infinitesimal amplitude a.

Lemma 14.6.1. *We have* $\delta_E \sqcap \delta_\Phi$.

Proof. Since $\sin x \sqcap x$ for infinitesimal x we have $y - ix \sqcap y - i\sin x$. □

Remark 14.6.2. Due to the multiplicative invariance of the relation of adequality, the rescalings of E and Φ by change of variable $z = aZ$ remain adequal and therefore define adequal hyperreal walks by Theorem 14.4.2.

14.7. Adjusting linear prevector field

We will compare the linear field E to another linear prevector field H defined by

$$\begin{aligned} H(x+iy) &= e^{-i\lambda\sqrt{\frac{g}{\ell}}}(x+iy) \\ &= \left(x\cos\lambda\sqrt{\tfrac{g}{\ell}} + y\sin\lambda\sqrt{\tfrac{g}{\ell}}\right) + \left(-x\sin\lambda\sqrt{\tfrac{g}{\ell}} + y\cos\lambda\sqrt{\tfrac{g}{\ell}}\right) i \end{aligned}$$

given by clockwise rotation of the x, y plane by infinitesimal angle $\lambda\sqrt{\frac{g}{\ell}}$. The corresponding hyperreal walk, defined by hyperfinite iteration of prevector field H, satisfies the exact equality

$$H_t(a,0) = \left(a\cos\sqrt{\tfrac{g}{\ell}}\,t, -a\sin\sqrt{\tfrac{g}{\ell}}\,t\right) \tag{14.7.1}$$

whenever t is a hypernatural multiple of λ.

COROLLARY 14.7.1. *We have the periodicity property $H_{\frac{2\pi}{\sqrt{g/\ell}}}(z) = z$ and hence*

$$H_{t+\frac{2\pi}{\sqrt{g/\ell}}} = H_t \tag{14.7.2}$$

whenever both t and $\frac{2\pi}{\sqrt{\frac{g}{\ell}}}$ are hypernatural multiples of λ.

LEMMA 14.7.2. *We have an adequality $\delta_E \sqcap \delta_H$ whenever x and y are finite.*

PROOF. Let $\alpha = \lambda\sqrt{\frac{g}{\ell}} \approx 0$. Then $\delta_E(z) = \delta_E(x+iy) = \alpha(y - ix) = -i\alpha z$, while $H(z) = e^{-i\alpha}z$ and therefore $\delta_H(z) = (e^{-i\alpha}-1)\,z$. Therefore

$$\frac{\delta_H}{\delta_E} = \frac{e^{-i\alpha}-1}{-i\alpha} \approx 1$$

as required. □

By Theorem 14.4.2 the hyperfinite walks of Φ, E and H satisfy

$$\Phi_t(a,0) \sqcap E_t(a,0) \sqcap H_t(a,0)$$

for each finite initial amplitude a and for all finite time t which is a hypernatural multiple of λ.

14.8. Conclusion

The advantage of the prevector field H is that its hyperreal walk is given by an explicit formula (14.7.1) and is therefore periodic with period precisely $\frac{2\pi}{\sqrt{g/\ell}}$, provided we choose our base infinitesimal λ in such a way that $\frac{2\pi}{\lambda\sqrt{g/\ell}}$ is hypernatural. We obtain the following consequence of the periodicity property (14.7.2): modulo an appropriate choice of a representing prevector field (namely, H) in the adequality class, the hyperreal walk is periodic with period $2\pi\sqrt{\ell/g}$. This can be summarized as follows.

COROLLARY 14.8.1. *The period of infinitesimal oscillations of the pendulum represented by a hyperreal walk is independent of their amplitude.*

Remark 14.8.2. If one rescales such an infinitesimal oscillation to appreciable size by a change of variable $z = aZ$ where a is the amplitude, and takes standard part, one obtains a standard harmonic oscillation with period $2\pi\sqrt{\ell/g}$.

The formulation contained in Corollary 14.8.1 has the advantage of involving neither rescaling nor shadow-taking.

CHAPTER 15

Study of D^2

15.1. Second-order mean value theorem

We will need a well-known generalisation of the mean value theorem (Theorem 7.4.1) to two variables. We will follow [**Rudin 1976**, Theorem 9.40]. We will assume that f is C^2 for simplicity (Rudin presents weaker conditions).

THEOREM 15.1.1 (Rudin Theorem 9.40). *Suppose a two-variable function $f(u^1, u^2)$ is defined and of class C^2 in an open set in the plane including a closed rectangle Q with sides parallel to the coordinate axes, having points (a, b) and $(a+h, b+k)$ as opposite vertices, where $h > 0$ and $k > 0$. Set*

$$\Delta^2(f, Q) = f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b).$$

Then there is a point (x, y) in the interior of Q such that

$$\Delta^2(f, Q) = hk \left(\frac{\partial^2 f}{\partial u^1 \partial u^2} \right)(x, y).$$

REMARK 15.1.2. Note the analogy with the mean value theorem.

PROOF. We will exploit a 1-variable function

$$u(t) = f(t, b+k) - f(t, b) \text{ where } t \in [a, a+h].$$

Two applications of the mean value theorem will yield the result. Note that the mean value theorem is applicable to the first partial derivative of f since $f \in C^2$ by hypothesis. Namely there is an $x \in (a, a+h)$, and there is a $y \in (b, b+k)$, such that

$$\begin{aligned} \Delta^2(f, Q) &= u(a+h) - u(a) \\ &= hu'(x) \\ &= h\left(\left(\frac{\partial f}{\partial u^1} \right)(x, b+k) - \left(\frac{\partial f}{\partial u^1} \right)(x, b) \right) \\ &= hk \left(\frac{\partial^2 f}{\partial u^1 \partial u^2} \right)(x, y) \end{aligned}$$

proving the theorem. □

15.2. The condition C^2 implies D^2

Let M be a smooth manifold. Recall our "second difference" notation for a prevector field $\Phi\colon {}^*M \to {}^*M$ in a coordinate neighborhood of a nearstandard point $a \in {}^\natural M$:

$$\Delta^2_{v,w}\Phi(a) = \Phi(a) - \Phi(a+v) - \Phi(a+w) + \Phi(a+v+w).$$

The condition D^2 for prevector fields was defined in Section 13.4, namely $\Delta^2_{v,w}\Phi(a) \prec \lambda\|v\|\|w\|$ whenever $v, w \prec \lambda$. Given a classical vector field X, in coordinates one defines a prevector field $\Phi(a) = a + \lambda X(a)$. We will show that if X is of class C^2 then Φ is of class D^2.

Proposition 15.2.1. *Let $W \subseteq \mathbb{R}^n$ be open. Let $X\colon W \to \mathbb{R}^n$ be a classical C^2 vector field, and consider its natural extension to ${}^*W$. Then for each $a \in {}^\natural W$ and each pair of infinitesimal $v, w \in {}^*\mathbb{R}^n$, we have*

$$\Delta^2_{v,w}X(a) \prec \|v\|\|w\|.$$

The proof appears below. We obtain the following immediate corollary (see Example 12.6.4).

Corollary 15.2.2. *Let X be of class C^2. If Φ is the prevector field on ${}^*W$ defined by $\Phi(a) = a + \lambda X(a)$ then Φ is D^2.*

Proof of Proposition 15.2.1. Let $U \subseteq W$ be a smaller neighborhood of $\mathbf{st}(a)$ for which all second partial derivatives of X are bounded by a fixed real constant. Let $(X^1, \ldots, X^n)$ be the components of X. Given $p \in U$ and $v_1, v_2 \in \mathbb{R}^n$ such that we have $p+s_1v_1+s_2v_2 \in U$ for all $0 \le s_1, s_2 \le 1$, we define functions ψ^i by

$$\psi^i(s_1, s_2) = X^i(p + s_1v_1 + s_2v_2).$$

Then by the chain rule, the mixed second partial of ψ^i satisfies

$$\left|\frac{\partial^2\psi^i}{\partial s_1\partial s_2}\right| \le C_i\|v_1\|\|v_2\|$$

where C_i is determined by a bound for the second partial derivatives of X^i in U. To shorten the notation we will use summation over pairs $(e_1, e_2) \in \{0,1\}^2$. By Theorem 15.1.1 (second-order mean value theorem) there is a point $(t_1, t_2) \in [0,1] \times [0,1]$ such that

$$\sum_{(e_1,e_2)\in\{0,1\}^2} (-1)^{e_1+e_2}\psi^i(e_1, e_2) = \frac{\partial^2}{\partial s_1\partial s_2}\psi^i(t_1, t_2).$$

Therefore

$$\begin{aligned}\left|\Delta^2_{v_1,v_2}X^i(p)\right| &= \Big|\sum_{(e_1,e_2)\in\{0,1\}^2}(-1)^{e_1+e_2}X^i(p+e_1v_1+e_2v_2)\Big|\\ &= \Big|\sum_{(e_1,e_2)\in\{0,1\}^2}(-1)^{e_1+e_2}\psi^i(e_1,e_2)\Big|\\ &= \Big|\frac{\partial^2}{\partial s_1\partial s_2}\psi^i(t_1,t_2)\Big|\\ &\le C_i\|v_1\|\|v_2\|\end{aligned}$$

with C_i as above. Since this is valid for each X^i, $i = 1, \ldots, n$, there is a constant $K \in \mathbb{R}$ such that

$$\left|\Delta^2_{v_1,v_2}X^i(p)\right| \le K\|v_1\|\|v_2\| \qquad (15.2.1)$$

for every $p \in U$ and $v_1, v_2 \in \mathbb{R}^n$ such that $p + s_1v_1 + s_2v_2 \in U$ for all choices of $s_1, s_2 \in [0, 1] \subseteq \mathbb{R}$.

Applying the transfer principle to the bound (15.2.1), we conclude that the same formula holds, with the same K, for all $p \in {}^*U$ and $v_1, v_2 \in {}^*\mathbb{R}^n$ provided that $p + s_1v_1 + s_2v_2 \in {}^*U$ for all choices of $s_1, s_2 \in {}^*[0, 1] \subseteq {}^*\mathbb{R}$. In particular this is true for $p = a$ and all infinitesimal $v_1, v_2 \in {}^*\mathbb{R}^n$. □

15.3. Bounds for D^2 prevector fields

PROPOSITION 15.3.1. *If Φ and G are D^2 then $\Phi \circ G$ is D^2.*

PROOF. We will establish that $D^2 \implies D^1$ in Proposition 15.11.4. In a coordinate chart, let $p = G(a)$ and $x = G(a + v) - G(a)$ and $y = G(a+w) - G(a)$. By Propositions 15.11.4 and 13.6.1 we have $x \sqcap \|v\|$ and $y \sqcap \|w\|$. Also by Propositions 15.11.4 and 13.6.1 we obtain the following bound that will be exploited below:

$$\begin{aligned}&\|\Phi(p + x + y) - \Phi(G(a + v + w))\|\\ &\qquad \sqcap \|p + x + y - G(a + v + w)\|\\ &\qquad = \| - G(a) + G(a + v) + G(a + w) - G(a + v + w)\|\\ &\qquad = \|\Delta^2_{v,w}G(a)\|\\ &\qquad \prec \lambda\|v\|\|w\|\end{aligned}$$

since G is D^2. Thus,

$$\|\Phi(p + x + y) - \Phi(G(a + v + w))\| \prec \lambda\|v\|\|w\|. \qquad (15.3.1)$$

Next,

$$
\begin{aligned}
&\Delta^2_{v,w}\Phi\circ G(a)\\
&\qquad = \|\Phi\circ G(a) - \Phi\circ G(a+v) - \Phi\circ G(a+w) + \Phi\circ G(a+v+w)\|\\
&\qquad = \|\Phi(p) - \Phi(p+x) - \Phi(p+y) + \Phi\circ G(a+v+w)\|\\
&\qquad \le \|\Phi(p) - \Phi(p+x) - \Phi(p+y) + \Phi(p+x+y)\|\\
&\qquad\qquad + \|\Phi(p+x+y) - \Phi(G(a+v+w))\|\\
&\qquad = \left\|\Delta^2_{x,y}\Phi(p)\right\| + \|\Phi(p+x+y) - \Phi(G(a+v+w))\|\\
&\qquad \prec \lambda\|x\|\|y\| + \lambda\|v\|\|w\|\\
&\qquad \prec \lambda\|v\|\|w\|
\end{aligned}
$$

exploiting the bound (15.3.1). □

15.4. Sharpening the bounds via underspill

Here we are interested in sharpening the $\prec$ bounds for prevector fields on M in terms of a specific real constant, using underspill.[1]

Let Φ be a prevector field on W, where $W \subseteq M$ is an open coordinate neighborhood with image $U \subseteq \mathbb{R}^n$. Let $B \subseteq U$ be a closed ball. We will identify W with U to lighten the notation.

Proposition 15.4.1. *In the notation above we have the following.*

1. *There is $C \in \mathbb{R}$ such that $\|\Phi(a) - a\| \le C\lambda$ for all $a \in {}^*\!B$.*
2. *If G is another prevector field, then there is a finite $\beta \in {}^*\mathbb{R}$ such that $\|\Phi(a) - G(a)\| \le \beta\lambda$ for all $a \in {}^*\!B$.*
3. *If furthermore $\Phi \equiv G$ then the constant $\beta \in {}^*\mathbb{R}$ in (2) can be chosen to be infinitesimal.*

Proof. The first item is a special case of the second when $G(a) = a$ for all a. Now let us prove the second item. Define a set A of hypernaturals as follows:

$$A = \{n \in {}^*\mathbb{N} \colon \|\Phi(a) - G(a)\| \le n\lambda \ \text{ for every } \ a \in {}^*\!B\}.$$

Since B is closed, we have ${}^*\!B \subseteq {}^\natural U$. Hence every infinite hypernatural $n \in {}^*\mathbb{N}$ is in A by definition of a prevector field. Therefore ${}^*\mathbb{N} \setminus \mathbb{N} \subseteq A$. But A is an internal set, being defined in terms of internal entities Φ and G. Therefore by underspill there is a finite integer C in A.[2]

---

[1] glisha?

[2] Recall that *underspill* is the fact that if $A \subseteq {}^*\mathbb{N}$ is an internal set, and A contains each infinite n then it must also contain a finite n. This is based on the fact that infinite hypernaturals form an external set; see Section 7.7.

To prove the third statement (for $\Phi \equiv G$), consider the set

$$A = \left\{ n \in {}^*\mathbb{N} \colon \|\Phi(a) - G(a)\| \le \frac{\lambda}{n} \ \text{ for every } \ a \in {}^*B \right\}.$$

Every finite $n \in {}^*\mathbb{N}$ is in A by definition of the equivalence relation. Therefore by overspill there is an infinite $n \in {}^*\mathbb{N}$ in A. We can therefore choose the infinitesimal value $\beta = \frac{1}{n}$. □

15.5. Further bounds via underspill

PROPOSITION 15.5.1. *Let Φ be a D^1 prevector field on M. Let $W \subseteq M$ be an open coordinate neighborhood with image $U \subseteq \mathbb{R}^n$. Let $B \subseteq U$ be a closed ball. Then*

(1) *there is $K \in \mathbb{R}$ such that whenever $a, b \in {}^*B$, we have*

$$\|\Phi(a) - a - \Phi(b) + b\| \le K\lambda\|a - b\|; \tag{15.5.1}$$

(2) *furthermore, we have*

$$(1 - K\lambda)\|a - b\| \le \|\Phi(a) - \Phi(b)\| \le (1 + K\lambda)\|a - b\|. \tag{15.5.2}$$

PROOF. Let $N = \left\lfloor \frac{1}{\lambda} \right\rfloor$. Let $a, b \in {}^*B$. We construct a hyperfinite partition of the segment between a and b. For each $k = 0, \ldots, N$ let

$$a_k = a + \frac{k}{N}(b - a).$$

Then $a_k - a_{k+1} \prec \lambda$. Let $C_{ab} \in {}^*\mathbb{R}$ be the maximum of the ratios

$$\frac{\|\Phi(a_k) - a_k - \Phi(a_{k+1}) + a_{k+1}\|}{\lambda\|a_k - a_{k+1}\|}$$

over all $k = 0, \ldots, N-1$. Since Φ is D^1, it follows that the constant C_{ab} is finite. For every $0 \le k \le N - 1$ we have

$$\begin{aligned}\|\Phi(a_k) - a_k - \Phi(a_{k+1}) + a_{k+1}\| &\le C_{ab}\lambda\|a_k - a_{k+1}\| \\ &= C_{ab}\lambda\frac{\|a - b\|}{N} \\ &\sqcap\ C_{ab}\lambda^2\|a - b\|.\end{aligned}$$

Therefore by the triangle inequality we have

$$\begin{aligned}\|\Phi(a) - a - &\Phi(b) + b\| \\ &= \|\delta_\Phi(a) - \delta_\Phi(b)\| \\ &\le \sum_{k=0}^{N-1} \|\delta_{\Phi(a_k)} - \delta_{\Phi(a_{k+1})}\| \\ &\le C_{ab}\lambda\|a - b\|.\end{aligned} \tag{15.5.3}$$

To remove the dependence on a, b, let

$$A = \{n \in {}^*\mathbb{N} \colon \|\Phi(a) - a - \Phi(b) + b\| \le n\lambda\|a - b\| \quad \text{whenever} \quad a, b \in {}^*B\}.$$

Since each C_{ab} as in (15.5.3) is finite, every infinite hypernatural $n \in {}^*\mathbb{N}$ is in A. Hence by underspill, the internal set A also contains a finite integer K, proving (15.5.1) in item (1).

The estimate (15.5.2) follows from the first as in the proof of Proposition 13.6.1. □

15.6. Overspill sharpening for D^2

In Section 15.5 we used overspill to prove a sharpened bound for a D^1 prevector field Φ on M. A similar proposition can be proved for D^2. Recall that $\Delta^2_{v,w}\Phi(a) = \Phi(a) - \Phi(a+v) - \Phi(a+w) + \Phi(a+v+w)$, and that $\Phi \in D^2$ if and only if $\Delta^2_{v,w}\Phi(a) \prec \lambda\|v\|\|w\|$ whenever $a \in {}^\natural M$ and $v \prec \lambda$, $w \prec \lambda$.

Proposition 15.6.1. *Let Φ be a D^2 prevector field on a manifold M. Let $W \subseteq M$ be a coordinate neighborhood with image $U \subseteq \mathbb{R}^n$. Let $B \subseteq U$ be a closed ball. Then there is $K \in \mathbb{R}$ such that*

$$\|\Delta^2_{v,w}\Phi(a)\| \le K\lambda\|v\|\|w\|$$

for all $a \in {}^*B$ and $v, w \in {}^*\mathbb{R}^n$ such that $a + v, a + w, a + v + w \in {}^*B$.

Proof. The proof is similar to that of Proposition 15.5.1. We set $N = \left\lfloor \frac{1}{\lambda} \right\rfloor$. Given $a \in {}^*B$ and $v, w \in {}^*\mathbb{R}^n$ such that $a + v, a + w, a + v + w \in {}^*B$, let $a_{k,l} = a + \frac{k}{N}v + \frac{l}{N}w$, $0 \le k, l \le N$. Let C_{avw} be the maximum of the values

$$\frac{\|\Phi(a_{k,l}) - \Phi(a_{k+1,l}) - \Phi(a_{k,l+1}) + \Phi(a_{k+1,l+1})\|}{\lambda\|v/N\|\|w/N\|}$$

over all $0 \le k, l \le N - 1$. Then C_{avw} is finite. It follows that whenever we have $0 \le k, l \le N - 1$, we have a bound

$$\|\Phi(a_{k,l}) - \Phi(a_{k+1,l}) - \Phi(a_{k,l+1}) + \Phi(a_{k+1,l+1})\| \le C_{avw}\lambda \left\|\frac{v}{N}\right\| \left\|\frac{w}{N}\right\|.$$

Summing the resulting double telescoping sum over all $0 \le k, l \le N - 1$ we get

$$\|\Phi(a) - \Phi(a + v) - \Phi(a + w) + \Phi(a + v + w)\| \le C_{avw}\lambda\|v\|\|w\|.$$

By underspill as in the proof of Proposition 15.5.1, there is a single finite K which works for all a, v, w. Namely, to remove the dependence on a, v, w, let

$$A = \{n \in {}^*\mathbb{N} \colon \|\Delta^2_{v,w}\Phi(a)\| \le n\lambda\|v\|\|w\| \text{ whenever all the points are in } {}^*B\}.$$

Since each C_{avw} is finite, every infinite hypernatural $n \in {}^*\mathbb{N}$ is in A. Hence by underspill, the internal set A also contains a finite integer K, proving the proposition. □

15.7. Dependence on initial conditions

In this section we deal with initial conditions.[3]

REMARK 15.7.1. We showed in Proposition 15.5.1 that each D^1 prevector field Φ defined in an open neighborhood of a closed Euclidean ball B satisfies the bound $\|\Phi(a) - a - \Phi(b) + b\| \leq K\lambda\|a - b\|$ for all $a, b \in {}^*B$ for an appropriate finite K. Such a K will be exploited in Theorem 15.7.2.

Hyperreal walk was defined in Section 12.9, and its local version in Section 13.2. We show that the hyperreal walk is Lipschitz, in the following precise sense.

THEOREM 15.7.2. *Let Φ be a local D^1 prevector field on ${}^*U$. Given a point $p \in U$ and a coordinate neighborhood of p with image $W \subseteq \mathbb{R}^n$, let $B' \subseteq B \subseteq W$ be closed balls of radii respectively $r/2$ and r around the image of p. Suppose $\|\Phi(a) - a - \Phi(b) + b\| \leq K\lambda\|a - b\|$ for all $a, b \in {}^*B$, with K a finite constant,[4] then*

(1) *there is a positive $T \in \mathbb{R}$ such that $\Phi_t(a) \in {}^*B$ whenever $a \in {}^*B'$ and $-T \leq t \leq T$;*
(2) *whenever $a, b \in {}^*B'$ and $0 \leq t \leq T$, we have $\|\Phi_t(a) - \Phi_t(b)\| \leq e^{Kt}\|a - b\|$.*
(3) *If we take a slightly larger constant $K' = K/(1 - K\lambda)^2$, then whenever $a, b \in {}^*B'$ and $-T \leq t \leq T$, we have*

$$e^{-K'|t|}\|a - b\| \leq \|\Phi_t(a) - \Phi_t(b)\| \leq e^{K'|t|}\|a - b\|.$$

PROOF. We will give a proof for positive t.[5] Let $C \in \mathbb{R}$ be as in Proposition 15.4.1 (on sharpening bounds via underspill and overspill), so that $\|\Phi(a) - a\| \leq C\lambda$ for all $a \in {}^*B$. We define $T \in \mathbb{R}$ by setting

$$T = \frac{r}{2C} > 0.$$

Let $a \in {}^*B'$ and $0 \leq t \leq T$, and set

$$n = n(t) = \left\lfloor \frac{t}{\lambda} \right\rfloor. \tag{15.7.1}$$

[3] tnai hatchalati

[4] Such finite K exists by Proposition 15.5.1.

[5] The case of negative t then follows from Remark 15.13.3.

We then obtain a bound for $\Phi^n(a) - a$ by means of a hyperfinite sum as follows:

$$\|\Phi^n(a) - a\| \le \sum_{m=1}^{n} \|\Phi^m(a) - \Phi^{m-1}(a)\| \le nC\lambda \le \frac{r}{2}.$$

By the triangle inequality we obtain

$$\|\Phi^n(a)\| \le \|\Phi^n(a) - a\| + \|a\| \le \frac{r}{2} + \frac{r}{2} = r,$$

and therefore $\Phi^n(a) \in {}^*\!B$, proving item (1).

Next, by estimate (15.5.2) we have

$$(1 - K\lambda)\|a - b\| \le \|\Phi(a) - \Phi(b)\| \le (1 + K\lambda)\|a - b\|$$

for all $a, b \in {}^*\!B$, and so by internal induction (see Theorem 12.8.1)[6] we obtain

$$(1 - K\lambda)^n\|a - b\| \le \|\Phi^n(a) - \Phi^n(b)\| \le (1 + K\lambda)^n\|a - b\|,$$

proving (2) in view of the fact that

$$(1 + K\lambda)^n = \left(1 + \frac{Kt}{n}\right)^n \le e^{Kt} \tag{15.7.2}$$

by formula (15.7.1) defining n. The choice of K' results by elementary algebra. □

15.8. Dependence on prevector field

In this section we will provide bounds on how fast the flows of a pair of prevector fields can diverge from each other.

In particular, this will show that if prevector fields are equivalent then the corresponding classical flows coincide.

We will deal with a more general situation (than that of balls $B' \subseteq B$ dealt with in Section 15.7) where A, A' are internal rather than standard sets. Here A is the set where by assumption the flow is contained in up to time T. The set A will typically be larger than A'.

[6]This step requires internal induction and cannot be obtained by iterating the formula n times, writing down the resulting formula with integer parameter n, and applying the transfer principle. The problem is that the entities occurring in the formula are internal rather than standard, so there is no real statement to apply transfer to. However, one can use the same trick as in the definition of the hyperfinite iterate in the first place, by introducing an additional parameter as in Section 12.9. Namely, the formula holds for all triples $(\Phi_1, \Phi_2, n) \in \text{Map}(M) \times \text{Map}(M) \times \mathbb{N}$, and we apply the star transform.

REMARK 15.8.1. Let Φ be a D^1 local prevector field on ${}^*U$ and let G be any local prevector field on ${}^*U$. Given a coordinate neighborhood included in U with image $W \subseteq \mathbb{R}^n$, let $A' \subseteq A \subseteq {}^*W$ be internal sets. Assume that β is such that

$$\|\Phi(a) - G(a)\| \leq \beta\lambda \tag{15.8.1}$$

for all $a \in A$ (in particular if β is infinitesimal then the prevector fields are equivalent). Assume also that $K > 0$ is a finite constant such that $\|\Phi(a) - a - \Phi(b) + b\| \leq K\lambda\|a - b\|$ for all $a, b \in A$. Assume that $0 < T \in \mathbb{R}$ is such that $\Phi_t(a)$ and $G_t(a)$ are in A for all $a \in A'$ and $0 \leq t \leq T$.

THEOREM 15.8.2. *In the notation of Remark* 15.8.1*, we have the following.*

(1) *for all $a \in A'$ and $0 \leq t \leq T$, we have the following bound:*

$$\|\Phi_t(a) - G_t(a)\| \leq \frac{\beta}{K}(e^{Kt} - 1) \leq \beta t e^{Kt}.$$

(2) *If G^{-1} exists, e.g., if G is also D^1, and if $\Phi_t(a)$ and $G_t(a)$ are in A for all $a \in A'$ and $-T \leq t \leq T$, then*

$$\|\Phi_t(a) - G_t(a)\| \leq \frac{\beta}{K}(e^{K|t|} - 1) \leq \beta|t|e^{K|t|}$$

for all $-T \leq t \leq T$.

PROOF. It suffices to provide a proof for positive t. We will prove by internal induction that for each n,

$$\|\Phi^n(a) - G^n(a)\| \leq \frac{\beta}{K}\Big((1 + K\lambda)^n - 1\Big). \tag{15.8.2}$$

This implies the statement when applied to $n = n(t)$; cf. (15.7.2). The basis of the induction is provided by the estimate (15.8.1). By triangle inequality and Proposition 15.5.1(2) we have

$$\begin{aligned}
&\|\Phi^{n+1}(a) - G^{n+1}(a)\| \\
&\qquad \leq \|\Phi(\Phi^n(a)) - \Phi(G^n(a))\| + \|\Phi(G^n(a)) - G(G^n(a))\| \\
&\qquad \leq (1 + K\lambda)\|\Phi^n(a) - G^n(a)\| + \beta\lambda.
\end{aligned}$$

Exploiting the inductive hypothesis (15.8.2), we obtain

$$\begin{aligned}\|\Phi^{n+1}(a) - G^{n+1}(a)\| & \\ &\le (1+K\lambda)\frac{\beta}{K}\big((1+K\lambda)^n - 1\big) + \beta\lambda \\ &= \frac{\beta}{K}(1+K\lambda)^{n+1} - \frac{\beta}{K} - \frac{\beta K\lambda}{K} + \beta\lambda \\ &= \frac{\beta}{K}\left((1+K\lambda)^{n+1} - 1\right)\end{aligned}$$

proving the inductive step from n to $n+1$. □

Corollary 15.8.3. *Assume Φ and G are as in Theorem 15.8.2. If we have $\Phi \equiv G$, then there is an infinitesimal β for the statement of Theorem 15.8.2, and therefore $\Phi_t(a) \approx G_t(a)$ for all $0 \le t \le T$.*

Proof. By Proposition 15.4.1. □

15.9. Inducing a real flow

The hyperreal walk Φ_t of a prevector field Φ induces a classical flow on M as follows.

Definition 15.9.1. Let Φ be a D^1 local prevector field. Let $B' \subseteq \mathbb{R}^n$ and $[-T,T]$ be as in Theorem 15.7.2. The *real flow* $\phi_t\colon B' \to M$ induced by Φ_t is defined by setting

$$\phi_t(x) = \mathbf{st}(\Phi_t(x)) \tag{15.9.1}$$

(see Definition 12.9.4). Recall that Φ_t is defined by hyperfinite iteration (of Φ) precisely $\lfloor t/\lambda \rfloor$ times.

The following are immediate consequences of Theorems 15.7.2 and Corollary 15.8.3.

Theorem 15.9.2. *Given a D^1 prevector field Φ the following hold:*

1. *the flow ϕ_t is Lipschitz continuous with constant $e^{K|t|}$.*
2. *the real flow ϕ_t is injective.*
3. *If G is another D^1 prevector field with real flow g_t, and $\Phi \equiv G$, then $\phi_t = g_t$.*

Remark 15.9.3. Suppose Φ is obtained from a classical vector field X by the procedure of Example 12.6.4. Then [**Keisler 1976**, Theorem 14.1] shows that ϕ_t is in fact the flow of the classical vector field X in the classical sense. By Theorem 15.9.2(3), each prevector field Φ that realizes X will yield a hyperreal walk whose standard part will produce the same classical flow.

The results of this subsection have the following application to the classical setting.

Corollary 15.9.4. *Consider an open set* $U \subseteq \mathbb{R}^n$. *Let* $X, Y : U \to \mathbb{R}^n$ *be classical vector fields, where* X *is Lipschitz with constant* K, *and there is a constant* $b \in \mathbb{R}$ *such that* $\|X(x) - Y(x)\| \leq b$ *for all* $x \in U$. *If* $x(t), \tilde{x}(t)$ *are integral curves (i.e., the* ϕ_t *with different initial conditions) of* X *then*

$$\|x(t) - \tilde{x}(t)\| \leq e^{Kt}\|x(0) - \tilde{x}(0)\|. \tag{15.9.2}$$

If $y(t)$ *is an integral curve of* Y *with* $x(0) = y(0)$ *then*

$$\|x(t) - y(t)\| \leq \frac{b}{K}(e^{Kt} - 1) \leq bte^{Kt}. \tag{15.9.3}$$

Proof. Define prevector fields on ${}^*U$ as usual by setting $\Phi(a) - a = \lambda X(a)$ and $G(a) - a = \lambda Y(a)$ as in Example 12.6.4. We apply Theorem 15.7.2 to prove (15.9.2). We use Theorem 15.8.2, and Remark 15.9.3 with $b = \beta$ to prove (15.9.3). □

15.10. Geodesic flow

One advantage of the hyperreal approach to solving a differential equation is that the hyperreal walk exists for all time, being defined combinatorially by iteration of a self-map of the manifold. The focus therefore shifts away from proving the *existence* of a solution, to establishing the *properties* of a solution.

Thus, our estimates show that given a uniform Lipschitz bound on the vector field, the hyperreal walk for all finite time stays in the finite part of ${}^*M$.

Theorem 15.10.1. *Suppose* M *is complete and* Φ *is a* D^1 *prevector field on* M *satisfying a* uniform *Lipschitz bound. Then the hyperreal walk* Φ_t *for all finite time* t *descends to a real flow on* M.

Proof. On a complete manifold M, the walk Φ_t is nearstandard for all finite t. The estimates of Corollary 15.9.4 imply that the real flow ϕ_t exists and is unique for all $t \in \mathbb{R}$. □

In particular, consider the geodesic equations on an n-dimensional manifold M.

Consider a Riemannian metric $\langle\, ,\, \rangle$ on M with metric coefficients written in coordinates $(u^1, \ldots, u^n)$ as (g_{ij}) meaning that $\left\langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \right\rangle = g_{ij}$. The symbols Γ^k_{ij} can be expressed as $\Gamma^k_{ij} = \frac{g^{k\ell}}{2}\left(g_{i\ell;j} - g_{ij;\ell} + g_{j\ell;i}\right)$ where (g^{ij}) is the inverse matrix of the metric (g_{ij}).

Definition 15.10.2. Given a regular curve $\alpha(s)$ in M expressed in coordinates by $(\alpha^1(s), \ldots, \alpha_n(s))$, the geodesic equation is

$$(\forall k) \quad \alpha^{k\prime\prime} + \Gamma^k_{ij}\alpha^{i\prime}\alpha^{j\prime} = 0 \tag{15.10.1}$$

with the Einstein summation convention.

The equations (15.10.1) can be interpreted as a first-order ODE on the $(2n-1)$-dimensional manifold SM (union of unit spheres in all tangent spaces). The latter is known to satisfy a uniform Lipschitz bound. Therefore the geodesic flow on a complete Riemannian manifold M exists for all time $t \in \mathbb{R}$.

15.11. D^2 implies D^1

We study further properties of D^1 and D^2 fields.

Remark 15.11.1. An ivector is obtained from a vector by multiplying by λ. For the sake of providing an intuitive explanation for the implication $D^2 \implies D^1$, note that being D^1 roughly corresponds to the first derivative being finite, while being D^2 roughly corresponds to the second derivative being finite. We want to show that if the second derivative is finite then the first derivative is finite. It turns out that there is a bound for the first derivative in terms of bounds on the function itself and the second derivative.

Lemma 15.11.2. *If a C^2 function $f\colon [0,1] \to \mathbb{R}$ satisfies $|f| \le A$ and $|f''| \le B$ then $|f'| \le 4A + B$.*

Proof. By Taylor formula with remainder $f(x) = f(c) + (x-c)f'(c) + \frac{1}{2}(x-c)^2 f''(d)$ where d is between x and c. Therefore $f'(c) = \frac{f(x)-f(c)-\frac{1}{2}(x-c)^2 f''(d)}{x-c}$. Now let c be a maximum of $|f'|$. We can assume without loss of generality that $c \ge \frac{1}{2}$ (otherwise apply the same argument to the function $f(\frac{1}{2}-x)$ to get the same bound). Then

$$\begin{aligned} |f'(c)| &\le 2\left|f(0) - f(c) - \frac{1}{2}(0-c)^2 f''(d)\right| \\ &\le 2|f(0)| + 2|f(c)| + c^2|f''(d)| \\ &\le 4A + B \end{aligned}$$

as required. □

First we prove a lower bound (15.11.1) on the change in a displacement function δ over a suitably long interval. Later the bound will be applied to $\delta = \delta_\Phi$.

LEMMA 15.11.3. *Let $B \subseteq \mathbb{R}^n$ be an open ball around the origin* 0, *let $a \in \mathfrak{h}(0)$, and let $0 \neq v \in {}^*\mathbb{R}^n$ be infinitesimal. If $\delta \colon {}^*B \to {}^*\mathbb{R}^n$ is an internal function satisfying*

$$\delta(x) - 2\delta(x+v) + \delta(x+2v) \prec \|v\| \|\delta(a) - \delta(a+v)\|$$

for all $x \in {}^\natural B$, then there is $m \in {}^*\mathbb{N}$ such that $a + mv \in {}^\natural B$ and

$$\delta(a) - \delta(a+v) \prec \|v\| \|\delta(a) - \delta(a+mv)\|. \tag{15.11.1}$$

PROOF. We choose real $r > 0$ slightly smaller than the radius of B and let $N = \left\lfloor \frac{r}{\|v\|} \right\rfloor$ where $\lfloor\ \rfloor$ is the integer part function. Then for each $m \leq N$ we have $a + mv \in {}^\natural B$. Let $x_j = a + jv$ for $0 \leq j \leq N$. Let $A = \delta(a) - \delta(a+v)$. By our assumption on δ we have

$$\delta(x_j) - 2\delta(x_{j+1}) + \delta(x_{j+2}) = C_j \|v\| \|A\|$$

with C_j finite. Let C be the maximum of $C_0, \ldots, C_N$. Then C is finite and

$$\delta(x_j) - 2\delta(x_{j+1}) + \delta(x_{j+2}) \leq C\|v\| \|A\| \text{ for each } 0 \leq j \leq N. \tag{15.11.2}$$

Given $k \leq N$ we have

$$\begin{aligned}
&\Big\| A - \Big(\delta(x_k) - \delta(x_{k+1}) \Big) \Big\| \\
&\quad = \Big\| \Big(\delta(x_0) - \delta(x_1) \Big) - \Big(\delta(x_k) - \delta(x_{k+1}) \Big) \Big\| \\
&\quad \leq \sum_{j=0}^{k-1} \Big\| \Big(\delta(x_j) - \delta(x_{j+1}) \Big) - \Big(\delta(x_{j+1}) - \delta(x_{j+2}) \Big) \Big\| \\
&\quad \leq Ck\|v\| \|A\|
\end{aligned}$$

by estimate (15.11.2). Thus for each $m \leq N$,

$$\begin{aligned}
\Big\| mA - \Big(\delta(x_0) - \delta(x_m) \Big) \Big\| &= \Big\| \sum_{k=0}^{m-1} \Big(A - \Big(\delta(x_k) - \delta(x_{k+1}) \Big) \Big) \Big\| \\
&\leq Cm^2 \|v\| \|A\|,
\end{aligned}$$

and therefore

$$\Big\| mA - \Big(\delta(x_0) - \delta(x_m) \Big) \Big\| = Km^2 \|v\| \|A\|$$

with K finite. It follows that

$$m\|A\| - Km^2 \|v\| \|A\| \leq \|\delta(x_0) - \delta(x_m)\|,$$

and so, multiplying by $\|v\|$, we obtain

$$m\|v\| \|A\| (1 - Km\|v\|) \leq \|v\| \|\delta(x_0) - \delta(x_m)\|.$$

Now let $m = \min\{\ N\ ,\ \lfloor \frac{1}{2K\|v\|} \rfloor\ \}$, then

$$m\|v\|\|A\|/2 \le \|v\|\|\delta(x_0) - \delta(x_m)\|.$$

By definition of N and since K is finite, we have that $m\|v\|$ is appreciable, i.e., not infinitesimal, and so finally $A \prec \|v\|\|\delta(x_0) - \delta(x_m)\|$, that is, $\delta(a) - \delta(a+v) \prec \|v\|\|\delta(a) - \delta(a+mv)\|$. □

PROPOSITION 15.11.4. *If prevector field Φ is D^2 for some choice of coordinates in $W \subseteq M$, then Φ is D^1.*

PROOF. Given $a \in {}^\natural W$, in the given coordinates take some ball B around $\mathbf{st}(a)$. Let $\delta_\Phi(x) = \Phi(x) - x$. We must show that whenever $v \prec \lambda$ it follows that $\delta_\Phi(a) - \delta_\Phi(a+v) \prec \lambda\|v\|$ where $v = b - a$ in Definition 13.3.3. If $\delta_\Phi(a) - \delta_\Phi(a+v) \prec\prec \lambda\|v\|$ then we are certainly done. Otherwise we have

$$\lambda v \prec \|\delta_\Phi(a) - \delta_\Phi(a+v)\|. \tag{15.11.3}$$

On the other hand, we have

$$\begin{aligned}\delta_\Phi(x) - 2\delta_\Phi(x+v) + \delta_\Phi(x+2v) &= \Phi(x) - 2\Phi(x+v) + \Phi(x+2v)\\ &= \Delta^2_{v,v}\Phi(x)\\ &\prec \lambda\|v\|^2,\end{aligned} \tag{15.11.4}$$

by taking $v = w$ in Definition 13.4.2. Combining the two inequalities (15.11.3) and (15.11.4) we obtain

$$\delta_\Phi(x) - 2\delta_\Phi(x+v) + \delta_\Phi(x+2v) \prec \|v\|\|\delta_\Phi(a) - \delta_\Phi(a+v)\|.$$

Now Lemma 15.11.3 yields an $m \in {}^*\mathbb{N}$ such that $a + mv \in {}^\natural B$ and

$$\begin{aligned}\delta_\Phi(a) - \delta_\Phi(a+v) &\prec \|v\|\|\delta_\Phi(a) - \delta_\Phi(a+mv)\|\\ &\le \|v\|\Big(\|\delta_\Phi(a)\| + \|\delta_\Phi(a+mv)\|\Big)\\ &\prec \|v\|\lambda\end{aligned}$$

since Φ is a prevector field so that $\delta_\Phi(x) = \Phi(x) - x \prec \lambda$ for all x. This shows that $\delta_\Phi(a) - \delta_\Phi(a+v) \prec \lambda\|v\|$ and hence Φ is D^1. □

15.12. Injectivity

When speaking about local D^1 or D^2 prevector fields, whenever needed we will assume, perhaps by passing to a smaller domain, that a constant K as in Propositions 15.5.1, 15.6.1 exists.

LEMMA 15.12.1. *If Φ is a D^1 prevector field then Φ is injective on ${}^\natural M$.*

PROOF. Let $a \neq b \in {}^\natural M$. If $\mathbf{st}(a) \neq \mathbf{st}(b)$ then clearly $\Phi(a) \neq \Phi(b)$. Otherwise there exists a B such that $a, b \in {}^\natural B$ as in Proposition 15.5.1. Therefore part (15.5.2) of the proposition gives the lower bound $\|\Phi(a) - \Phi(b)\| \geq (1 - K\lambda)\|a - b\|$ and hence $\Phi(a) \neq \Phi(b)$. □

We will now show that a D^1 prevector field is in fact bijective on ${}^\natural M$. We first prove local surjectivity, as follows.

LEMMA 15.12.2. *Let $B_1 \subseteq B_2 \subseteq B_3 \subseteq \mathbb{R}^n$ be closed balls centered at the origin, of radii $r_1 < r_2 < r_3$. Fix $0 < s \in \mathbb{R}$ smaller than $r_2 - r_1$ and $r_3 - r_2$. Consider a function $f\colon B_2 \to B_3$ such that*

(1) *$\|f(x) - f(y)\| \leq 2\|x - y\|$ for all $x, y \in B_2$;*
(2) *$\|f(x) - x\| < s$ for all $x \in B_2$.*

Then $f(B_2) \supseteq B_1$.

PROOF. Our assumptions on f imply that it is continuous. For every $x \in \partial B_2$ the straight line segment between x and $f(x)$ is included in $B_3 - B_1$, so the restriction $f|_{\partial B_2}$ is homotopic in $B_3 - B_1$ to the inclusion of ∂B_2. Now if some $p \in B_1$ is not in the image $f(B_2)$ then the restriction $f|_{\partial B_2}$ is null-homotopic in $B_3 - \{p\}$, and so the same is true for the inclusion of ∂B_2, a contradiction. □

PROPOSITION 15.12.3. *Let $B_1 \subseteq B_2 \subseteq B_3 \subseteq \mathbb{R}^n$ be closed balls centered at the origin, of radii $r_1 < r_2 < r_3$. If $\Phi\colon {}^*B_2 \to {}^*B_3$ is a local D^1 prevector field then $\Phi({}^*B_2) \supseteq {}^*B_1$.*

PROOF. Fix $0 < s \in \mathbb{R}$ smaller than $r_2 - r_1$ and $r_3 - r_2$ as before. We apply transfer to the statement of Lemma 15.12.2 to obtain that for every internal function $f\colon {}^*B_2 \to {}^*B_3$, if $\|f(x) - f(y)\| \leq 2\|x - y\|$ for all $x, y \in {}^*B_2$ and $\|f(x) - x\| < s$ for all $x \in {}^*B_2$ then $f({}^*B_2) \supseteq {}^*B_1$. In particular this is true for a D^1 prevector field $\Phi\colon {}^*B_2 \to {}^*B_3$, by Proposition 15.5.1(2). □

The following is immediate from Corollary 15.12.1 and Propostion 15.12.3.

THEOREM 15.12.4. *If $\Phi\colon {}^*M \to {}^*M$ is a D^1 prevector field then the restriction $\Phi\restriction_{{}^\natural M}\colon {}^\natural M \to {}^\natural M$ is bijective.*

PROOF. Let $p \in {}^\natural M$ and let $p_0 = \mathbf{st}(p)$. Arguing with balls B_i centered at p_0 we obtain that $p \in {}^*B_1 \subseteq \Phi({}^*B_2) \subseteq \Phi({}^\natural M)$. □

REMARK 15.12.5. On all of ${}^*M$, a D^1 prevector field may be non-injective and nonsurjective, e.g., take $M = (0, 1)$ and $\Phi\colon {}^*M \to {}^*M$ given by $\Phi(x) = \lambda$ for $x \leq \lambda$ and $\Phi(x) = x$ otherwise. Recall that the definition of D^1 prevector field imposes no restrictions in the complement ${}^*M - {}^\natural M$ other than being internal on ${}^*M$.

15.13. Time reversal

REMARK 15.13.1. Given a prevector field Φ of class D^1, the restricted map $\Phi\lfloor_{{}^\natural M}\colon {}^\natural M \to {}^\natural M$ and its inverse $(\Phi\lfloor_{{}^\natural M})^{-1}$ are not internal if M is noncompact, since their domain is not internal. On the other hand, for any $A \subseteq M$, the restriction $\Phi\lfloor_{{}^*A}$ is internal. Furthermore, on ${}^*B_1$ of Proposition 15.12.3, Φ has an inverse $\Phi^{-1}\colon {}^*B_1 \to {}^*B_2$ in the sense that $\Phi \circ \Phi^{-1}(a) = a$ for all $a \in {}^*B_1$, and Φ^{-1} is internal. So, for a local D^1 prevector field $\Phi\colon {}^*U \to {}^*V$ we may always assume (perhaps for slightly smaller domain) that $\Phi^{-1}\colon {}^*U \to {}^*V$ also exists, in the above sense. As mentioned, we will usually not mention the range ${}^*V$ but rather speak of a local prevector field on ${}^*U$.

PROPOSITION 15.13.2. *If Φ is D^1 then Φ^{-1} is D^1, where Φ^{-1} is as in Remark* 15.13.1.

PROOF. Let $x = \Phi^{-1}(a)$, $y = \Phi^{-1}(b)$, then

$$\begin{aligned}\|\Phi^{-1}(a) - \Phi^{-1}(b) - a + b\| &= \|x - y - \Phi(x) + \Phi(y)\| \\ &\prec \lambda\|x - y\| \\ &\prec \lambda\|\Phi(x) - \Phi(y)\| \\ &= \lambda\|a - b\|\end{aligned}$$

by Proposition 13.6.1. □

REMARK 15.13.3. If one follows the proofs of Proposition 15.13.2 and Proposition 13.6.1 one sees that if $\|\Phi(a)-\Phi(b)-a+b\| \le K\lambda\|a-b\|$ in some domain ${}^*U \subseteq {}^*\mathbb{R}^n$, then

$$\|\Phi^{-1}(a) - \Phi^{-1}(b) - a + b\| \le K'\lambda\|a - b\|$$

in a corresponding domain for Φ^{-1}, with K' only slightly larger than K, namely $K' = K/(1 - K\lambda)$.

15.14. Equivalence

LEMMA 15.14.1. *Let Φ, G be D^1 prevector fields. If $\Phi(a) \equiv G(a)$ for all* standard a, *then $\Phi(b) \equiv G(b)$ for all b and therefore $\Phi \equiv G$.*

Proof. Let K be as in Proposition 15.5.1 for both Φ and G. Let $a = \mathbf{st}(b)$, then

$$\begin{aligned}
\|\Phi(b) - G(b)\| &= \|\delta_\Phi(b) - \delta_G(b)\| \\
&= \|\delta_\Phi(b) - \delta_\Phi(a) + \delta_\Phi(a) - \delta_G(a) + \delta_G(a) - \delta_G(b)\| \\
&= \|\delta_\Phi(b) - \delta_\Phi(a) + \Phi(a) - G(a) + \delta_G(a) - \delta_G(b)\| \\
&\le \|\delta_\Phi(b) - \delta_\Phi(a)\| + \|\Phi(a) - G(a)\| + \|\delta_G(a) - \delta_G(b)\| \\
&\le K\lambda\|a - b\| + \|\Phi(a) - G(a)\| + K\lambda\|a - b\| \\
&\prec\prec \lambda
\end{aligned}$$

since $a \approx b$ and $\Phi(a) \equiv G(a)$ by hypothesis. □

Recall that Definition 12.7.1, which defines when a prevector field Φ realizes a classical vector field X, involves only *standard* points. It follows from Lemma 15.14.1 that if Φ is D^1 then this determines Φ up to equivalence. Namely, we have the following.

Corollary 15.14.2. *Let $U \subseteq \mathbb{R}^n$ be open, and let $X\colon U \to \mathbb{R}^n$ be a classical vector field. If Φ, G are two D^1 prevector fields that realize X then $\Phi \equiv G$. In particular, if X is Lipschitz and G is a D^1 prevector field that realizes X, then $\Phi \equiv G$, where Φ is the prevector field obtained from X as in Example* 12.6.4 *namely in coordinates $\Phi(a) = a + \lambda \cdot X(a)$.*

15.15. Lie ibracket

Lie ibrackets were discussed briefly in Section 13.5. Relation to flows will be studied in Section 15.23.

Given a pair of local or global prevector fields Φ, G whose inverses Φ^{-1}, G^{-1} exist, e.g., if Φ, G are D^1, we define their Lie ibracket, denoted $[\Phi, G]^{\cdot}$, as follows. Its relation to the classical Lie bracket will be clarified below. Let $N = \left\lfloor \frac{1}{\lambda} \right\rfloor$.

Definition 15.15.1. The *Lie ibracket* $[\Phi, G]^{\cdot}$ of prevector fields Φ and G is $[\Phi, G]^{\cdot} = (G^{-1} \circ \Phi^{-1} \circ G \circ \Phi)^N$.

Remark 15.15.2. Note that this definition of the ibracket is global in character in that it does not rely on a choice of a coordinate patch where addition can be used. We will compare this definition to one using addition in Section 15.17.

Lie ibracket is related to Lie bracket (see Section 15.17). A pair of noncommuting vector fields appears in Example 9.3.4.

15.16. The canonical pvf of a flow

Given a prevector field Φ, we defined the hyperreal walk Φ_t and the real flow ϕ_t. The canonical representative prevector field is obtained once we have the real flow $\phi_t = \mathbf{st}(\Phi_t)$. Namely we can extend it to the nonstandard domain as usual, and use it to define a new prevector field $\widetilde{\Phi}$ as follows.

DEFINITION 15.16.1. The prevector field $\widetilde{\Phi}$ is the map ${}^*\phi_t$ at time $t = \lambda$, or $\widetilde{\Phi} = {}^*\phi_\lambda$.

15.17. Relation to classical Lie bracket

The following theorem justifies our definition of the ibracket $[\Phi, G]^{\cdot}$, by relating it to the classical notion of Lie bracket.

THEOREM 15.17.1. *Let X, Y be two classical C^2 vector fields and let $[X, Y]$ denote their classical Lie bracket. Let Φ, G be D^2 prevector fields that realize X, Y respectively. Then $[\Phi, G]^{\cdot}$ realizes $[X, Y]$.*

PROOF. By Remark 15.9.3, the flows ϕ_t, g_t obtained as shadows of Φ_t, G_t coincide with the classical flows of X, Y. It is well known that the classical bracket $[X, Y]$ is related in coordinates to the classical flow as follows:

$$[X, Y](p) = \lim_{t \to 0} \frac{1}{t^2} \Big(g_t^{-1} \circ \phi_t^{-1} \circ g_t \circ \phi_t(p) - p \Big). \tag{15.17.1}$$

The characterisation of limits via infinitesimals (Definition 4.3.1) yields

$$[X, Y](p) \approx \frac{1}{\lambda^2} \Big(\widetilde{G}^{-1} \circ \widetilde{\Phi}^{-1} \circ \widetilde{G} \circ \widetilde{\Phi}(p) - p \Big).$$

Now, if $v \approx w$ then $\lambda v \equiv \lambda w$, so by Example 12.6.4, the bracket $[X, Y]$ can be realized by the prevector field

$$A(a) = a + \frac{1}{\lambda} \Big(\widetilde{G}^{-1} \circ \widetilde{\Phi}^{-1} \circ \widetilde{G} \circ \widetilde{\Phi}(a) - a \Big).$$

Thus it remains to show that $[\Phi, G]^{\cdot} \equiv A$. The proof involves several steps that will be developed below and can be summarized as follows.

By Proposition 15.18.1 the fields $\widetilde{\Phi}, \widetilde{G}$ are D^2, and so by Proposition 15.21.2 we have $[\widetilde{\Phi}, \widetilde{G}]^{\cdot} \equiv A$. By Theorem 15.22.1 we have $\Phi \equiv \widetilde{\Phi}$ and $G \equiv \widetilde{G}$, and so by Theorem 15.22.2 we conclude that $[\Phi, G]^{\cdot} \equiv A$. $\square$

15.18. The canonical prevector field of a D^2 flow is D^2

PROPOSITION 15.18.1. *If* $\Phi \in D^2$ *then* $\widetilde{\Phi} \in D^2$.

PROOF. We will obtain estimates for the real flow ϕ_t and then transfer them to obtain the desired estimates for $\widetilde{\Phi}$.

Assume $\|\Delta^2_{v,w}\Phi(a)\| \le K\lambda \|v\|\|w\|$ for all a, v, w in some ${}^*B$ as in Proposition 15.6.1. Recall that

$$\Delta^2_{v,w}\Phi^n(a) = \Phi^n(a) - \Phi^n(a+v) - \Phi^n(a+w) + \Phi^n(a+v+w).$$

We will prove by internal induction that

$$\|\Delta^2_{v,w}\Phi^n(a)\| \le K\lambda \sum_{i=n-1}^{2n-2} (1+K\lambda)^i \|v\|\|w\|. \tag{15.18.1}$$

Let $p = \Phi^n(a)$, $x = \Phi^n(a+v) - \Phi^n(a)$ and $y = \Phi^n(a+w) - \Phi^n(a)$. Since Φ is D^2 and hence also D^1 by Proposition 15.11.4, we have $\|x\| \le (1+K\lambda)^n\|v\|$ and $\|y\| \le (1+K\lambda)^n\|w\|$. We exploit the inductive hypothesis (15.18.1) to obtain the following estimate that will be useful later:

$$\begin{aligned}
&\|\Phi(p+x+y) - \Phi^{n+1}(a+v+w))\| \\
&\quad = \|\Phi(p+x+y) - \Phi \circ \Phi^n(a+v+w))\| \\
&\quad \le (1+K\lambda)\|p+x+y - \Phi^n(a+v+w)\| \\
&\quad = (1+K\lambda)\| - \Phi^n(a) + \Phi^n(a+v) + \Phi^n(a+w) - \Phi^n(a+v+w)\| \\
&\quad = (1+K\lambda)\|\Delta^2_{v,w}(a)\Phi^n\|.
\end{aligned}$$

Therefore by the inductive hypothesis (15.18.1),

$$\begin{aligned}
\|\Phi(p+x+y) - \Phi^{n+1}(a+v+w))\| &\le (1+K\lambda)K\lambda \sum_{i=n-1}^{2n-2} (1+K\lambda)^i \|v\|\|w\| \\
&= K\lambda \sum_{i=n}^{2n-1} (1+K\lambda)^i \|v\|\|w\|.
\end{aligned}$$

Thus we have

$$\|\Phi(p+x+y) - \Phi^{n+1}(a+v+w))\| \le K\lambda \sum_{i=n}^{2n-1} (1+K\lambda)^i \|v\|\|w\|. \tag{15.18.2}$$

Now

$$\begin{aligned}
&\Delta^2_{v,w}\Phi^{n+1}(a)\\
&\quad= \|\Phi^{n+1}(a) - \Phi^{n+1}(a+v) - \Phi^{n+1}(a+w) + \Phi^{n+1}(a+v+w)\|\\
&\quad\le \|\Phi(p) - \Phi(p+x) - \Phi(p+y) + \Phi(p+x+y)\|\\
&\qquad\quad + \|\Phi(p+x+y) - \Phi^{n+1}(a+v+w)\|\\
&\quad\le K\lambda\|x\|\|y\| + K\lambda\sum_{i=n}^{2n-1}(1+K\lambda)^i\|v\|\|w\|
\end{aligned}$$

by (15.18.2). Therefore

$$\begin{aligned}
\Delta^2_{v,w}\Phi^{n+1}(a) &\le K\lambda(1+K\lambda)^{2n}\|v\|\|w\| + K\lambda\sum_{i=n}^{2n-1}(1+K\lambda)^i\|v\|\|w\|\\
&= K\lambda\sum_{i=n}^{2n}(1+K\lambda)^i\|v\|\|w\|
\end{aligned}$$

which completes the inductive step and proves (15.18.1). Now let $n = \lfloor t/\lambda \rfloor$. We obtain

$$\begin{aligned}
\|\Delta^2_{v,w}\Phi^n(a)\| &\le K\lambda\sum_{i=n-1}^{2n-2}(1+K\lambda)^i\|v\|\|w\|\\
&= K\lambda\sum_{i=n-1}^{2n-2}\left(1+\frac{Kt}{n}\right)^i\|v\|\|w\|\\
&\le K\lambda\sum_{i=n-1}^{2n-2}\left(1+\frac{Kt}{n}\right)^{2n-2}\|v\|\|w\|\\
&\le K\lambda n e^{2Kt}\|v\|\|w\|\\
&\le Kte^{2Kt}\|v\|\|w\|.
\end{aligned}$$

a bound valid for all a, v, w in some appreciable $^*B$ as in Proposition 15.6.1. We now take standard parts to obtain a similar bound for the real flow ϕ_t. Thus for standard a, v, w we have

$$\|\Delta^2_{v,w}\phi_t(a)\| \le Kte^{2Kt}\|v\|\|w\|. \tag{15.18.3}$$

Transfering estimate (15.18.3) to the nonstandard domain and evaluating at $t = \lambda$ we obtain $\|\Delta^2_{v,w}\widetilde{\Phi}(a)\| \le K\lambda e^{2K\lambda}\|v\|\|w\|$ as required. □

15.19. Ibracket of D^1 pvfs is a pvf

THEOREM 15.19.1. *If Φ, G are local D^1 prevector fields then the ibracket $\dot{[}\Phi, G\dot{]}$ is a prevector field, i.e., $\dot{[}\Phi, G\dot{]}(a) - a \prec \lambda$ for all a.*

PROOF. Substituting $x = a$ and $y = \Phi^{-1} \circ G \circ \Phi(a)$ in the relation $\delta_\Phi(x) - \delta_\Phi(y) \prec \lambda\|x - y\|$ gives

$$\begin{aligned}\delta_\Phi(a) - \delta_\Phi(\Phi^{-1} \circ G \circ \Phi(a)) &\prec \lambda\|a - \Phi^{-1} \circ G \circ \Phi(a)\| \\ &\prec \lambda^2\end{aligned}$$

since composition of D^1 pvfs is a D^1 pvf. Thus

$$\Phi(a) - a - G \circ \Phi(a) + \Phi^{-1} \circ G \circ \Phi(a) \prec \lambda^2. \tag{15.19.1}$$

Now substituting $x = \Phi(a)$ and $y = G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a)$ in the relation $G(x) - x - G(y) + y \prec \lambda\|x - y\|$ gives

$$\begin{aligned}G \circ \Phi(a) - \Phi(a) - \Phi^{-1} \circ G \circ \Phi(a) + G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a)& \\ \prec \lambda\|\Phi(a) - G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a)\|& \\ \prec \lambda^2.&\end{aligned} \tag{15.19.2}$$

Adding estimates (15.19.1) and (15.19.2) gives the estimate

$$G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - a \prec \lambda^2.$$

By underspill in an appropriate ${}^*U$ there exists a finite $C > 0$ such that

$$\|G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - a\| \le C\lambda^2$$

for all $a \in {}^*U$. Exploiting a telescoping sum, we obtain

$$\begin{aligned}\|\delta_{\dot{[}\Phi,G\dot{]}}(a)\| &= \|(G^{-1} \circ \Phi^{-1} \circ G \circ \Phi)^{\frac{1}{\lambda}}(a) - a\| \\ &\le \sum_{k=1}^{\frac{1}{\lambda}} \|(G^{-1} \circ \Phi^{-1} \circ G \circ \Phi)^k(a) - (G^{-1} \circ \Phi^{-1} \circ G \circ \Phi)^{k-1}(a)\| \\ &\le C\lambda\end{aligned}$$

as required. Therefore $\dot{[}\Phi, G\dot{]} \in D^0$. □

15.20. An estimate

LEMMA 15.20.1. *Let Φ be D^2 and G be D^1. Then for all a, b with $a - b \prec \lambda$, we have*

$$\delta_\Phi(a) - \delta_\Phi(b) - \delta_\Phi(G(a)) + \delta_\Phi(G(b)) \prec \lambda^2\|a - b\| \tag{15.20.1}$$

or equivalently

$$\Phi(a) - \Phi(b) - \Phi(G(a)) + \Phi(G(b)) - a + b + G(a) - G(b) \prec \lambda^2\|a - b\|.$$

PROOF. Let $v = b - a$ and $w = G(a) - a$. The bound (15.20.1) is similar to the D^2 condition $\Delta^2_{v,w}\Phi(a) \prec \lambda\|v - w\|$ and in fact this will be exploited in the proof. Since Φ is D^1 (by Proposition 15.11.4) we have

$$\Phi(a + v + w) - \Phi(G(b)) - (a + v + w) + G(b) \prec \lambda\|a + v + w - G(b)\|. \tag{15.20.2}$$

But $a + v + w = b + G(a) - a$ and so (15.20.2) yields

$$\begin{aligned} \Phi(a + v + w) - \Phi(G(b)) - b - G(a) + a + G(b) & \\ \prec \lambda\|b + G(a) - a - G(b)\| & \\ \prec \lambda^2\|a - b\| & \end{aligned}$$

since G is D^1. Thus

$$\begin{aligned} &\|\Phi(a) - \Phi(b) - \Phi(G(a)) + \Phi(G(b)) - a + b + G(a) - G(b)\| \\ &\quad = \|\Phi(a) - \Phi(a + v) - \Phi(a + w) + \Phi(a + v + w) \\ &\qquad\quad - \Phi(a + v + w) + \Phi(G(b)) - a + b + G(a) - G(b)\| \\ &\quad \leq \|\Delta^2_{v,w}\Phi(a)\| + \| - \Phi(a + v + w) + \Phi(G(b)) - a + b + G(a) - G(b)\| \\ &\quad \prec \lambda\|v\|\|w\| + \lambda^2\|a - b\| \\ &\quad = \lambda\|b - a\|\|G(a) - a\| + \lambda^2\|a - b\| \\ &\quad \prec \lambda^2\|a - b\| \end{aligned}$$

since $\|G(a) - a\| \prec \lambda$ by virtue of G being a pvf, as required. □

LEMMA 15.20.2. *Let Φ be a local D^1 prevector field defined on ${}^*U$. Assume*

$$\|\Phi(a) - \Phi(b) - a + b\| \leq K\lambda\|a - b\|$$

for all $a, b \in {}^*U$. Then the walk of Φ satisfies:

$$\|\Phi^n(a) - \Phi^n(b) - a + b\| \leq K\lambda n e^{K\lambda n}\|a - b\|.$$

PROOF. We have

$$\begin{aligned} \|\Phi^n(a) - \Phi^n(b) - a + b\| &\leq \sum_{i=1}^{n} \|\Phi^i(a) - \Phi^i(b) - \Phi^{i-1}(a) + \Phi^{i-1}(b)\| \\ &\leq \sum_{i=1}^{n} K\lambda\|\Phi^{i-1}(a) - \Phi^{i-1}(b)\| \\ &\leq \sum_{i=1}^{n} K\lambda(1 + K\lambda)^{i-1}\|a - b\| \\ &\leq K\lambda n e^{K\lambda n}\|a - b\|. \end{aligned}$$

Here the third inequality is by internal induction as in the proof of Theorem 15.7.2. □

THEOREM 15.20.3. *If* Φ, G *are* D^2 *then* $[\dot{\Phi}, \dot{G}]$ *is* D^1.

PROOF. By Proposition 15.11.4 (D^2 implies D^1), Proposition 15.13.2 (inverse is also D^1), and Proposition 13.7.5 (composition of D^1s is again D^1), we obtain that $\Phi^{-1} \circ G \circ \Phi$ is D^1. Now in Lemma 15.20.1 take G to be $\Phi^{-1} \circ G \circ \Phi$ then we get for $a - b \prec \lambda$:

$$\Phi(a) - \Phi(b) - G \circ \Phi(a) + G \circ \Phi(b) - a + b + \Phi^{-1} \circ G \circ \Phi(a) - \Phi^{-1} \circ G \circ \Phi(b) \prec \lambda^2 \|a - b\|. \tag{15.20.3}$$

As above $G^{-1} \circ \Phi^{-1} \circ G$ is D^1 and now take in Lemma 15.20.1 the entities a, b, Φ, G to be respectively $\Phi(a), \Phi(b), G, G^{-1} \circ \Phi^{-1} \circ G$ then we get

$$\begin{aligned} G \circ \Phi(a) &- G \circ \Phi(b) - \Phi^{-1} \circ G \circ \Phi(a) + \Phi^{-1} \circ G \circ \Phi(b) \\ &- \Phi(a) + \Phi(b) + G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(b) \\ &\prec \lambda^2 \|\Phi(a) - \Phi(b)\| \\ &\prec \lambda^2 \|a - b\| \end{aligned} \tag{15.20.4}$$

by Corollary 13.6.2 (namely $\|a - b\| \prec \|\Phi(a) - \Phi(b)\| \prec \|a - b\|$). Adding inequalities (15.20.3) and (15.20.4), we obtain

$$G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(b) - a + b \prec \lambda^2 \|a - b\|.$$

Let $H = G^{-1} \circ \Phi^{-1} \circ G \circ \Phi$. Then $[\dot{\Phi}, \dot{G}] = H^{\frac{1}{\lambda}}$. We need to show that $H^{\frac{1}{\lambda}}(a) - H^{\frac{1}{\lambda}}(b) - a + b \prec \lambda \|a - b\|$ and we know

$$H(a) - H(b) - a + b \prec \lambda^2 \|a - b\|.$$

By underspill in an appropriate ${}^*U$ there exists a finite $C > 0$ such that $\|H(a) - H(b) - a + b\| \leq C\lambda^2 \|a - b\|$ for all $a, b \in {}^*U$. So by Lemma 15.20.2 with $K = C\lambda$ and $n = \frac{1}{\lambda}$, we obtain

$$\|H^{\frac{1}{\lambda}}(a) - H^{\frac{1}{\lambda}}(b) - a + b\| \leq C\lambda e^{C\lambda} \|a - b\|$$

as required. □

15.21. Equivalence

LEMMA 15.21.1. *Let* Φ *be a local prevector field defined on* ${}^*U$*. Assume* $\|\Phi(a) - a\| \leq C\lambda$ *and* $\|\Phi(a) - a - \Phi(b) + b\| \leq K\lambda \|a - b\|$ *for all* $a, b \in {}^*U$*. Then the walk* Φ^n *satisfies the bound*

$$\|\Phi^n(a) - a - n(\Phi(a) - a)\| \leq KCn^2\lambda^2.$$

PROOF. We have

$$\begin{aligned}\|\Phi^n(a) - a - n(\Phi(a) - a)\| &= \left\|\sum_{i=1}^{n}\Big(\Phi^i(a) - \Phi^{i-1}(a) - (\Phi(a) - a)\Big)\right\| \\ &\le \sum_{i=1}^{n}\left\|\Phi(\Phi^{i-1}(a)) - \Phi^{i-1}(a) - \Phi(a) + a\right\| \\ &\le \sum_{i=1}^{n} K\lambda\|\Phi^{i-1}(a) - a\| \\ &\le \sum_{1\le j<i\le n} K\lambda\|\Phi^j(a) - \Phi^{j-1}(a)\| \\ &\le n^2 K\lambda C\lambda\end{aligned}$$

as required. □

PROPOSITION 15.21.2. *Let Φ, G be D^2 prevector fields. Then we have an equivalence $\dot{[}\Phi, G\dot{]}(a) \equiv a + \frac{1}{\lambda}\Big(G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - a\Big)$ for all a.*

PROOF. Let $H = G^{-1} \circ \Phi^{-1} \circ G \circ \Phi$. The proof of Theorem 15.19.1 provides $C' \prec 1$ such that $\|H(a) - a\| \le C'\lambda^2$ for all a. The proof of Theorem 15.20.3 provides $C'' \prec 1$ such that

$$\|H(a) - H(b) - a + b\| \le C''\lambda^2\|a - b\|$$

for all a, b. Taking $C = C'\lambda$, $K = C''\lambda$ and $n = \frac{1}{\lambda}$ in Lemma 15.21.1 we get $\|H^{\frac{1}{\lambda}}(a) - a - \frac{1}{\lambda}(H(a) - a)\| \le C'C''\lambda^2 \prec\prec \lambda$. □

15.22. Two theorems

THEOREM 15.22.1.

THEOREM 15.22.2.

15.23. Ibracket and flows

The following theorem corresponds to the classical fact that the Lie bracket of two vector fields vanishes if and only if their flows commute. We recall that prevector field I is defined by $I(a) = a$ for all a.

THEOREM 15.23.1. *Let Φ, G be two D^2 prevector fields and ϕ_t, g_t the associated flows. Then we have $\dot{[}\Phi, G\dot{]} \equiv I$ if and only if $\phi_t \circ g_s = g_s \circ \phi_t$ for all $0 \le t, s \le T$ for some $0 < T \in \mathbb{R}$.*

Proof. Assume first that $\dot{[}\Phi, G\dot{]} \equiv I$, i.e., $\dot{[}\Phi, G\dot{]}(a) - a \prec\prec \lambda$ for all a. So by Proposition 15.21.2 we obtain

$$\frac{1}{\lambda}\Big(G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - a\Big) \prec\prec \lambda,$$

so $G^{-1} \circ \Phi^{-1} \circ G \circ \Phi(a) - a \prec\prec \lambda^2$, which implies by Proposition 13.6.1 that $G \circ \Phi(a) - \Phi \circ G(a) \prec\prec \lambda^2$ for all a. Now let $n = \lfloor t/\lambda \rfloor$ and $m = \lfloor s/\lambda \rfloor$. We need to show $\Phi^n \circ G^m(a) \approx G^m \circ \Phi^n(a)$ for all a. This involves nm interchanges of Φ and G, where a typical move is from

$$\Phi^k \circ G^r \circ \Phi \circ G^{m-r} \circ \Phi^{n-k-1}$$

to $\Phi^k \circ G^{r+1} \circ \Phi \circ G^{m-r-1} \circ \Phi^{n-k-1}$. Applying the bound $\Phi \circ G(p) - G \circ \Phi(p) \prec\prec \lambda^2$ at the point $p = G^{m-r-1} \circ \Phi^{n-k-1}(a)$ we obtain

$$\Phi \circ G^{m-r} \circ \Phi^{n-k-1}(a) - G \circ \Phi \circ G^{m-r-1} \circ \Phi^{n-k-1}(a) \prec\prec \lambda^2.$$

By Propositions 15.11.4, 15.5.1 there is a constant $K \in \mathbb{R}$ such that

$$\|\Phi(a) - a - \Phi(b) + b\| \leq K\lambda\|a - b\|$$

and $\|G(a) - a - G(b) + b\| \leq K\lambda\|a - b\|$ for all a, b in an appropriate domain. Then by Theorem 15.7.2 applied to G^r and then to Φ^k,

$$\begin{aligned} &\|\Phi^k \circ G^r \circ \Phi \circ G^{m-r} \circ \Phi^{n-k-1}(a) - \Phi^k \circ G^{r+1} \circ \Phi \circ G^{m-r-1} \circ \Phi^{n-k-1}(a)\| \\ &\leq e^{K(t+s)}\|\Phi \circ G^{m-r} \circ \Phi^{n-k-1}(a) - G \circ \Phi \circ G^{m-r-1} \circ \Phi^{n-k-1}(a)\| \prec\prec \lambda^2. \end{aligned}$$

Adding the nm contributions when passing from $\Phi^n \circ G^m(a)$ to $G^m \circ \Phi^n(a)$ we obtain

$$\Phi^n \circ G^m(a) - G^m \circ \Phi^n(a) \prec\prec 1;$$

this is because among the nm differences that we add, there is a maximal one, which is say $\beta\lambda^2$ with β infinitesimal, and so the sum of all nm contributions is at most $nm\beta\lambda^2 \leq ts\beta \prec\prec 1$.

Conversely, assume $\phi_t \circ g_t = g_t \circ \phi_t$. Then by transfer $\widetilde{\Phi} \circ \widetilde{G} = \widetilde{G} \circ \widetilde{\Phi}$, so $\widetilde{G}^{-1} \circ \widetilde{\Phi}^{-1} \circ \widetilde{G} \circ \widetilde{\Phi} = I$, and so $[\widetilde{\Phi}, \widetilde{G}] = I$. By Proposition 15.18.1 and Theorems 15.22.1, 15.22.2 we get $\dot{[}\Phi, G\dot{]} \equiv I$. □

We have the following application to the traditional setting.

Corollary 15.23.2. *Let X, Y be classical C^2 vector fields. Then the flows of X and Y commute if and only if their Lie bracket vanishes.*

It follows that if $X_1, \ldots, X_k$ are k independent vector fields with $[X_i, X_j] = 0$ (classical Lie bracket) for $1 \leq i, j \leq k$, then there are coordinates in a neighborhood of any given point such that $X_1, \ldots, X_k$ are the first k coordinate vector fields.

Proof. Define prevector fields by $\Phi(a) = a + \lambda X(a)$ and $G(a) = a + \lambda Y(a)$ as in Example 12.6.4, and apply Proposition 15.2.1, Remark 15.9.3, and Theorems 15.17.1, 15.23.1. The final statement is a straightforward conclusion in the classical setting. □

Part 2

Foundations of true infinitesimal analysis

CHAPTER 16

Universes and extensions

This chapter develops a detailed rigorous set-theoretic setting for the TIDG framework.

16.1. Universes

Robinson's idea was to enlarge the mathematical world we are studying, in a way that does not change any of its properties (in a sense to be made precise). The enlarged view leads to new insights on the original world. Thus, our starting point is some set X, which in the case of differential geometry may be (the underlying set of) a smooth n-dimensional manifold M, or the disjoint union of M with finitely many other sets one might want to refer to, say $\mathbb{R}^n$, $\mathbb{R}$, and $\mathbb{N}$.

Let us review some set-theoretic background related to universes. Recall that given a set A, the symbol $\mathcal{P}(A)$ denotes the set of all subsets of A. We define the universe $V(X)$ as follows. Let

$$V_0(X) = X.$$

At the next step we set $V_1(X) = V_0(X) \cup \mathcal{P}(V_0(X))$. More generally, we define recursively for $n \in \mathbb{N}$,

$$V_{n+1}(X) = V_n(X) \cup \mathcal{P}(V_n(X)).$$

Finally the universe $V(X)$ is defined as follows.

Definition 16.1.1. The *universe*, or *superstructure*, over a set X is the set

$$V(X) = \bigcup_{n\in\mathbb{N}} V_n(X).$$

Corollary 16.1.2. *If $a \in V(X) \setminus X$ then $a \subseteq V(X)$.* □

Exercise 16.1.3. By definition, we have $V_m(X) \subseteq V_{m+1}(X)$. Prove that in fact $V_m(X) \subsetneqq V_{m+1}(X)$, and even, by Cantor's theorem, that $V_{m+1}(X)$ has a strictly greater cardinality than $V_m(X)$.

For special needs of Robinson's framework, one usually requires X to have the following key technical property; see e.g., [**Keisler 1976**, 15B] or [**Chang & Keisler 1990**, 4.4].

Definition 16.1.4 (base sets). A set X is a *base set* if $X \neq \varnothing$ and if $x \in X$ then x is an *atom* in $V(X)$, meaning that $x \neq \varnothing$ while $x \cap V(X) = \varnothing$.

The following are important consequences.

Corollary 16.1.5 ([**Chang & Keisler 1990**], Corollary 4.4.1). *Assume that X is a base set. If $a \in V(X)$ and $a \in b \in V_m(X)$ then $m \geq 1$ and $a \in V_{m-1}(X)$.* □

Corollary 16.1.6. *Assume that X is a base set. If $a, b \in V(X)$ and $a \cap V(X) = b \cap V(X)$ then either $a = b$ or $a, b \in X \cup \{\varnothing\}$.*

Proof. Suppose that $a \notin X$. Then $a \in \bigcup_{m \geq 1} V_m(X)$ by construction, and hence $a \subseteq V(X)$ and $a \cap V(X) = a$. If also $b \notin X$ then similarly $b \cap V(X) = b$, thus $a \cap V(X) = b \cap V(X)$ implies $a = b$. If $b \in X$ then $b \cap V(X) = \varnothing$ (as X is a base set), and we conclude that $a = a \cap V(X) = \varnothing$, as required. □

For examples of base sets see Section 16.2.

16.2. The real numbers as a base set

We recall Dedekind's construction of the real field, already mentioned in Section 4.11. There are several formally different definitions of the real field leading to essentially the same (modulo isomorphism) structure by results widely known since Dedekind. Of those, we are going to use the Dedekind definition via cuts, for technical reasons.

Definition 16.2.1 (Dedekind reals). A real number x is a pair $x = \{Q, Q'\}$ of two complementary sets of rationals, where $(\forall q \in Q)\,(\forall q' \in Q')\,[q < q']$ and the set $Q (= Q_x)$ does not have a maximal element.

With respect to the natural order on $\mathbb{R}$, one can express the two sets $Q = Q_x$ and $Q' = Q'_x$ as follows: $Q_x = \{q \in \mathbb{Q} \colon q < x\}$ and $Q'_x = \{q \in \mathbb{Q} \colon q \geq x\}$.

The key property of $\mathbb{R}$ which implies that $\mathbb{R}$ is a base set, is the fact that each real x consists of two elements Q_x and Q'_x, which are sets of von Neumann rank (see below) equal exactly to ω.

For those not versed in set theory we provide a definition.

Definition 16.2.2. The *von Neumann rank* of a set x is an ordinal $\texttt{rank}(x)$ defined so that $\texttt{rank}(\varnothing) = 0$ and if $x \neq \varnothing$ then $\texttt{rank}(x)$ is the least ordinal strictly greater than all ordinals $\texttt{rank}(y)$, $y \in x$.

The existence and uniqueness of $\texttt{rank}(x)$ follow from the axioms of modern Zermelo–Fraenkel set theory ZFC, with the key role of the *regularity*, or *foundation* axiom in the argument. Informally, $\mathtt{rank}(x)$ can be viewed as the finite or transfinite length of the cumulative construction of the set x from the empty set.

DEFINITION 16.2.3. Let γ be an ordinal. A set X is a γ-*base set* if $X \neq \varnothing$, X consists of non-empty elements, and we have $\mathtt{rank}(a) = \gamma$ whenever $a \in x \in X$.

LEMMA 16.2.4 ([**Chang & Keisler 1990**], Exercise 4.4.1). *If γ is an infinite ordinal, and X is a γ-base set, then X is a base set.*

PROOF. We show that if $y \in V_m(X)$ then either $\gamma+1 \leq \mathtt{rank}(y) \leq \gamma+m+1$ or $\mathtt{rank}(y) < m$, hence one never has $\mathtt{rank}(y) = \gamma$. This rules out the possibility of $y \in x$ for any $x \in X = V_0(X)$.

The claim is proved by induction on m.

If $y \in X = V_0(X)$ then $\mathtt{rank}(y) = \gamma + 1$ by the choice of X.

If $y \in V_1(X)$ then either $y \in X$ with $\mathtt{rank}(y) = \gamma+1$, or $y = \varnothing$ (the empty set) with $\mathtt{rank}(\varnothing) = 0$, or $\varnothing \neq y \subseteq V_0(X)$ with $\mathtt{rank}(y) = \gamma+2$.

Similarly if $y \in V_2(X)$ then either $y \in V_1(X)$ with $\mathtt{rank}(x) = 0$ or $\gamma+1 \leq \mathtt{rank}(y) \leq \gamma+2$ by the above, or $y \subseteq \{\varnothing\}$ with $\mathtt{rank}(y) \leq 1$, or $\varnothing \neq y \subseteq V_1(X) \setminus (X \cup \{\varnothing\})$ with $\mathtt{rank}(y) = \gamma + 3$.

Extending this argument, one easily proves the claim.[1] □

If X is a base set then there is another rank function $\mathtt{rank}_X(b)$ which reflects the inner structure of the superstructure $V(X)$ and is defined as follows.

DEFINITION 16.2.5. Let X be a base set. We let $\mathtt{rank}_X(b) = 0$ for each $b \in X$, $\mathtt{rank}_X(\varnothing) = 1$, and if $b \in V(X) \setminus X$, $b \neq \varnothing$, then $\mathtt{rank}_X(b)$ is the least natural number strictly bigger than all natural numbers $\mathtt{rank}_X(y)$, $y \in b$.

Recall that if $y \in b \in V(X)\setminus X$ then $y \in V(X)$ by Corollary 16.1.2. Thus, with some resemblance to the definition of the von Neumann rank above, $\mathtt{rank}_X(b)$ is the (finite, in this case) length of the cumulative construction of the set b from atoms in X. The empty set satisfies $rank_X(\varnothing) = 1$ (as any other subset of X) since it does not belong to X.

[1] An alternative argument runs as follows. If $x \in X = V_0(X)$ then $\mathtt{rank}(x) = \gamma + 1$ by the choice of X. If $x \in V_1(X)$ then either $x \in X$ with $\mathtt{rank}(x) = \gamma + 1$, or $x = \varnothing$ (the empty set) with $\mathtt{rank}(\varnothing) = 0$, or $x \in V_1(X)\setminus X$ with $\mathtt{rank}(x) = \gamma+2$. Similarly if $x \in V_2(X)$ then either $x \in X$ with $\mathtt{rank}(x) = \gamma + 1$, or $\mathtt{rank}(x) \leq 1$ (for $x = \varnothing$ and $x = \{\varnothing\}$), or $x \in V_2(X) \setminus X$ with $\gamma + 2 \leq \mathtt{rank}(x) \leq \gamma + 3$. Extending this argument, one easily shows by induction that if $m \geq 1$ then every set $y \in V_m(X)\setminus X$ satisfies either $\gamma+2 \leq \mathtt{rank}(y) \leq \gamma+m+1$ or $\mathtt{rank}(Y) < m$, hence never $\mathtt{rank}(y) = \gamma$. This rules out the possibility of $y \in x$ for any $x \in X = V_0(X)$.

EXERCISE 16.2.6. Prove that if X is a base set and $m \geq 1$ then every set $b \in V_m(X) \setminus V_{m-1}(X)$ satisfies $\texttt{rank}_X(b) = m$.

Next, we consider the most important case $X = \mathbb{R}$ (the real numbers). Recall that ω is the least infinite ordinal.

LEMMA 16.2.7. *The set $X = \mathbb{R}$ is an ω-base set, and hence a base set.*

PROOF. Set theory defines natural numbers so that $0 = \varnothing$ and $n = \{0, 1, 2, \ldots, n-1\}$ for all $n > 0$. It follows that $\texttt{rank}(n) = n$ for any natural number n. Next, according to the set theoretic definition of an ordered pair $\langle x, y\rangle$ as

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\},$$

all pairs and triples of natural numbers have finite ranks. Therefore each rational number q, identified with a triple $\langle s, m, n\rangle$, where $s \in \{0, 1, 2\}$, $m = 1, 2, 3, \ldots$, $n = 0, 1, 2, 3, \ldots$, and $q = (s-1) \cdot \frac{n}{m}$, also has a finite rank $\texttt{rank}(q)$.

We conclude that every *infinite* set S of rational numbers satisfies $\texttt{rank}(S) = \omega$ exactly. Therefore each Dedekind real x consists of two sets (infinite sets of rationals Q_x and Q'_x) of rank $\texttt{rank}(Q_x) = \texttt{rank}(Q'_x) = \omega$. It remains to apply Lemma 16.2.4 with $\gamma = \omega$. □

16.3. More sets in universes

From now on we concentrate on the case $X = \mathbb{R}$. By definition the set $V_0(\mathbb{R}) = \mathbb{R}$ has cardinality $\operatorname{card}(\mathbb{R}) = \mathfrak{c}$ (the continuum), and then by induction $\operatorname{card}(V_n(\mathbb{R})) = \exp^n(\mathfrak{c})$. This array of cardinalities easily exceeds all needs of conventional mathematics. For instance, the underlying set $|M|$ of a smooth connected manifold M of dimension $n \geq 1$ is a set of cardinality $\mathfrak{c}$. Therefore we can identify $|M|$ with a set in the complement $V_1(\mathbb{R}) \setminus V_0(\mathbb{R})$; this choice excludes the possibility of nonempty intersection of $|M|$ and $\mathbb{R}$. Then every object of interest for studying the smooth manifold M will be an element of $V(\mathbb{R})$, e.g.,

(1) The topology of M is an element of $V_3(\mathbb{R})$;
(2) With the usual encoding of an ordered pair $\langle a, b\rangle$ in terms of the set $\{\{a\}, \{a, b\}\}$, the set $M \times M$ is an element of $V_4(\mathbb{R})$;
(3) The set $\operatorname{Map}(M)$ of all maps from M to itself is an element of $V_5(\mathbb{R})$.

Similarly, the field operations of $\mathbb{R}$, the vector space operations of $\mathbb{R}^n$, and the natural embedding of $\mathbb{N}$ into $\mathbb{R}$, are elements of $V(\mathbb{R})$, as is each function from $\mathbb{R}^n$ to $\mathbb{R}$. See [**Chang & Keisler 1990**, Lemma 4.4.3].

16.4. Membership relation

Statements about any fixed universe of the form $V(X)$ are made using a first order language with a binary relation $\in$, the *membership relation*, as the only primary relation. This language will be called $\in$-language. Its formulas are called $\in$-formulas. Free variables in $\in$-formulas can be routinely replaced by elements of $V(X)$ called *constants* in such case.

DEFINITION 16.4.1. An $\in$-formula ϕ is called *bounded* if and only if quantifiers $\forall x$ and $\exists x$ always appear in ϕ in the form

$$\forall x\,(x \in a \Rightarrow \text{such-and-such})$$

and

$$\exists x\,(x \in a \wedge \text{such-and-such})$$

where a is either a variable or a constant, where ϕ can contain other free variables and/or constants in addition to a.

These forms are abbreviated respectively as $\forall x \in a\,(\text{such-and-such})$ and $\exists x \in a\,(\text{such-and-such})$.

16.5. Ultrapower construction of nonstandard universes

In this section we will present a generalisation of the ultrapower construction of Section 5.3.

For the purpose of such a generalisation, recall that $f, g, \ldots$ denote functions in a standard universe $V(X)$, while their extensions in a nonstandard extension ${}^*V(X)$ are denoted ${}^*f, {}^*g, \ldots$

DEFINITION 16.5.1. A *nonstandard extension* of $V(X)$ consists of a base set ${}^*X$ with $X \subsetneqq {}^*X$, a subuniverse ${}^*V(X) \subseteq V({}^*X)$, and a map

$$*\colon V(X) \to {}^*V(X) \tag{16.5.1}$$

satisfying the following two conditions:

(I) one has ${}^*X \subseteq {}^*V(X)$, and the map $*$ is an injective map from $V(X)$ into ${}^*V(X)$ such that if $x \in X$ then ${}^*x \in {}^*X$;

(II) for every bounded formula $\phi(x_1, \ldots, x_k)$, and every k-tuple of constants $a_1, \ldots, a_k \in V(X)$, the formula $\phi(a_1, \ldots, a_k)$ is true in $V(X)$ if and only if the formula $\phi({}^*a_1, \ldots, {}^*a_k)$ is true in ${}^*V(X)$.

REMARK 16.5.2. If $a \neq b$ then ${}^*a \neq {}^*b$ by (II), so that $*$ is a bijection of $V(X)$ onto a subset of ${}^*V(X)$. Yet we introduce the bijectivity condition separately by (I) for the sake of convenience.

Example 16.5.3. An example of a bounded formula is the formula expressing the continuity of a function; see Section 5.8. Therefore by (II) the same formula holds in the extension for ${}^*\!f$ where the quantification is now over the extended domain.

Again by (II), if $a \in A \in V(X)$ then ${}^*a \in {}^*\!A$, so if one identifies a with its image ${}^*a$, then one may think of A as a subset of ${}^*\!A$, that is, one can think of ${}^*\!A$ as an *extension* of A. Since functions are also sets, for every function $f \in V(X)$, $f\colon A \to B$, we can consider the set ${}^*\!f$, which is also a function, ${}^*\!f\colon {}^*\!A \to {}^*\!B$. If A and B are considered to be subsets of ${}^*\!A$ and ${}^*\!B$ respectively, then the function ${}^*\!f$ is an extension of the function f.

Theorem 16.5.4. *Nonstandard extensions exist.*

This is a well-known result presented in many sources, including the fundamental monograph on model theory [**Chang & Keisler 1990**]. We will provide a proof here so as to make the exposition self-contained. The proof occupies the remainder of this Chapter 16 as well as the next Chapter 17.

Proof. The construction we employ is referred to as the *reduced (or bounded) ultrapower*. It consists of two parts:

(1) the ultrapower itself, along with the Łoś theorem, namely Lemma 17.1.2;
(2) the transformation of the *bounded* ultrapower into a universe.

We start with an arbitrary ultrafilter $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$ containing all cofinite subsets of $\mathbb{N}$, namely a *nonprincipal* ultrafilter. The starting point for the construction of the full ultrapower is the product set $V(X)^{\mathbb{N}}$ which consists of all infinite sequences $\langle x_n \rangle = \langle x_n\colon n \in \mathbb{N}\rangle$ of elements $x_n \in V(X)$, or equivalently, all functions $f\colon \mathbb{N} \to V(X)$, via the identification of each f with the sequence $\langle f(n)\colon n \in \mathbb{N}\rangle$. The subsets $V_m(X)^{\mathbb{N}} \subseteq V^{\mathbb{N}}$ are defined accordingly.

16.6. Bounded ultrapower

Let X be a base set (see Definition 16.1.4).

Definition 16.6.1. The set union

$$V(X)^{\mathbb{N}}_{\mathtt{bd}} = \bigcup_m \left(V_m(X)^{\mathbb{N}}\right) \subsetneqq V(X)^{\mathbb{N}}$$

is the starting point for the construction of the *bounded* ultrapower.

Exercise 16.6.2. Find an element in $V(X)^{\mathbb{N}} \setminus V(X)^{\mathbb{N}}_{\mathtt{bd}}$.

Let m be a natural number. We have $b \in V_m(X) \setminus V_{m-1}(X)$ if and only if $\mathtt{rank}_X(b) = m$; see Exercise 16.2.6.

DEFINITION 16.6.3. A sequence $\langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ is *of type m* if the set $D_m = \{n \in \mathbb{N} \colon \mathtt{rank}_X(x_n) = m\}$ belongs to $\mathcal{F}$.

DEFINITION 16.6.4. In the notation as above, a sequence is *totally of type m* if $D_m = \mathbb{N}$.

In case $m = 0$, we naturally assume that $V_{-1}(X) = \varnothing$, so that $D_0 = \{n \colon x_n \in V_0(X)\}$, and $\langle x_n \rangle$ is totally of type 0 if and only if $x_n \in X = V_0(X)$ for all n.

LEMMA 16.6.5. *Assume that $\langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$. Then there is a unique number $m \in \mathbb{N}$ such that $\langle x_n \rangle$ is of type m.*

PROOF. By definition we have $\langle x_n \rangle \in V_M(X)^{\mathbb{N}}$ for some M and therefore we have a pairwise disjoint finite union $\mathbb{N} = \bigcup_{m \le M} D_m$ (partitioned by $\mathtt{rank}_X$). The lemma now follows from the fact that $\mathcal{F}$ is an ultrafilter. □

Following Section 5.4, we define an equivalence relation $\sim$ (or more precisely $\sim_{\mathcal{F}}$ as the relation depends on $\mathcal{F}$) on $V(X)^{\mathbb{N}}_{\mathtt{bd}}$ as follows:

$$\langle x_n \rangle \sim \langle y_n \rangle \quad \text{if and only if} \quad \{n \in \mathbb{N} \colon x_n = y_n\} \in \mathcal{F}.$$[2]

LEMMA 16.6.6. *Assume that $\langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ is of type m. There is a sequence $\langle y_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ totally of type m, such that $\langle x_n \rangle \sim_{\mathcal{F}} \langle y_n \rangle$.*

PROOF. If $n \in D_m$ then simply let $y_n = x_n$. If $n \notin D_m$ then put $y_n = a$, where a is a fixed element in $V_m(X) \setminus V_{m-1}(X)$. (Or $a \in X = V_0(X)$ in the case $m = 0$.) □

We continue with the proof of Theorem 16.5.4. The $\mathcal{F}$-quotient

$$V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F} = \left\{ \langle x_n \rangle_{\mathcal{F}} \colon \langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}} \right\}$$

consists of all $\mathcal{F}$-classes, or $\sim_{\mathcal{F}}$-classes

$$\langle x_n \rangle_{\mathcal{F}} = \left\{ \langle y_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}} \colon \langle x_n \rangle \sim \langle y_n \rangle \right\}$$

of sequences $\langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$. We define

$$V_m(X)^{\mathbb{N}}/\mathcal{F} = \left\{ \langle x_n \rangle_{\mathcal{F}} \colon \langle x_n \rangle \in V_m(X)^{\mathbb{N}} \right\}$$

for all m, so that $V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F} = \bigcup_m V_m(X)^{\mathbb{N}}/\mathcal{F}$.

COROLLARY 16.6.7. *Each element of $V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$ is of the form $\langle x_n \rangle_{\mathcal{F}}$, where $\langle x_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ is a sequence totally of type m for some (unique) m, so that $x_n \in V_m(X) \setminus V_{m-1}(X)$ for all n.*

[2]In such case there is a set $A \in \mathcal{F}$ such that we have $x_n = y_n$, for all $n \in A$.

Proof. Use lemmas 16.6.5, 16.6.6. □

If $\langle x_n\rangle_{\mathcal{F}}, \langle y_n\rangle_{\mathcal{F}}$ belong to $V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$ then we define a binary relation

$$\langle x_n\rangle_{\mathcal{F}} \;{}^*\!\!\in \langle y_n\rangle_{\mathcal{F}} \quad \text{if and only if} \quad \{n \in \mathbb{N}\colon x_n \in y_n\} \in \mathcal{F}.$$

Lemma 16.6.8. *If $\langle x_n\rangle_{\mathcal{F}}, \langle y_n\rangle_{\mathcal{F}}$ belong to $V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$ and we have $\langle x_n\rangle_{\mathcal{F}} \;{}^*\!\!\in \langle y_n\rangle_{\mathcal{F}}$ then there is a set $K \in \mathcal{F}$ and an index $m \geq 1$ such that $x_n \in V_{m-1}(X)$, $y_n \in V_m(X)$, and $x_n \in y_n$, for all $n \in K$.*

Proof. Let $L = \{n \in \mathbb{N}\colon x_n \in y_n\}$. By definition, the set L is a member of $\mathcal{F}$, and, as $\langle x_n\rangle_{\mathcal{F}}, \langle y_n\rangle_{\mathcal{F}} \in V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$, there is an index M such that $x_n, y_n \in V_M(X)$ for all n. Consider the sets

$$K_m = \{n \in L\colon x_n \in V_{m-1}(X) \text{ and } y_n \in V_m(X)\}, \quad 1 \leq m \leq M.$$

It follows by Corollary 16.1.5 that $L = \bigcup_{1 \leq m \leq M} K_m$, therefore at least one of K_m belongs to $\mathcal{F}$, and we can set $K = K_m$. □

Let $x \in V(X)$. Let $\langle x\rangle$ be the infinite constant sequence $\langle x, x, x, \dots\rangle = \langle x_n\rangle$, where $x_n = x$ forall n. We let $\langle x\rangle_{\mathcal{F}}$ denote its $\mathcal{F}$-class as above. The proof of Theorem 16.5.4 continues in Chapter 17.

CHAPTER 17

Łoś theorem and transfer principle

Here we continue the proof of Theorem 16.5.4.

Our main goal in Chapter 16 and here is a proof of the Transfer Principle, an assertion to the effect that, roughly speaking, any sentence true or false in the standard universe remains true, resp., false in the nonstandard universe, Corollary 17.2.1 and ultimately Corollary 17.3.1. This is not a really easy task, in part because such a typical tool of logic as *proof by induction* on the logical complexity of the sentence considered, does not seem to work. However there is another, and stronger claim which happens to admit such an inductive proof. This stronger claim is contained in the proof of Lemma 17.1.2. The lemma is in fact a cornerstone of the theory of ultraproducts.

17.1. Łoś theorem

The theorem connects the truth of sentences related to $V(X)^{\mathbb{N}}_{\mathtt{bd}}$ with their "truth sets" in the ultrafilter $\mathcal{F}$. A more detailed treatment of formulas can be found in [**Loeb & Wolff 2015**].

Definition 17.1.1. The truth set $T \subseteq \mathbb{N}$ of a formula ϕ is defined by setting

$$T_{\phi\left(\langle a_n^1\rangle,\ldots,\langle a_n^k\rangle\right)} = \left\{n \in \mathbb{N} \colon \phi(a_n^1, \ldots, a_n^k) \text{ is true in } V(X)\right\}.$$

Lemma 17.1.2 (the Łoś theorem). *If $\phi(x^1, \ldots, x^k)$ is an arbitrary bounded $\in$-formula and sequences $\langle a_n^1\rangle, \ldots, \langle a_n^k\rangle$ belong to $V(X)^{\mathbb{N}}_{\mathtt{bd}}$, then the sentence $\phi(\langle a_n^1\rangle_{\mathcal{F}}, \ldots, \langle a_n^k\rangle_{\mathcal{F}})$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$ if and only if the truth set $T_{\phi\left(\langle a_n^1\rangle,\ldots,\langle a_n^k\rangle\right)}$ is a member of $\mathcal{F}$.*

Here we use an upper index in the enumeration of free variables to distinguish it from the enumeration of terms of infinite sequences in $V(X)^{\mathbb{N}}_{\mathtt{bd}}$.

Example 17.1.3. Let ϕ be the formula $x \in y$. Assume that sequences $\langle a_n\rangle$, $\langle b_n\rangle$ belong to $V(X)^{\mathbb{N}}_{\mathtt{bd}}$. Then the formula $\langle a_n\rangle_{\mathcal{F}} \in \langle b_n\rangle_{\mathcal{F}}$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$ if and only if $\langle a_n\rangle_{\mathcal{F}} \mathrel{{}^*{\in}} \langle b_n\rangle_{\mathcal{F}}$ if and only if (by definition) the set $T_{\langle a_n\rangle\in\langle b_n\rangle} = \{n \in \mathbb{N} \colon a_n \in b_n\}$ belongs to $\mathcal{F}$.

PROOF. The proof of Lemma 17.1.2 proceeds by induction on the construction of formula ϕ out of elementary formulas $x \in y$ by means of logical connectives and quantifiers. Among them, it suffices to consider only $\neg$ (the negation), $\wedge$ (conjunction), and $\exists$ (the existence quantifier) since the remaining ones are known to be expressible via combinations of those three. For instance, $\Phi \vee \Psi$ is equivalent to $\neg(\neg\Phi \wedge \neg\Psi)$, and the quantifier $\forall$ is equivalent to $\neg\exists\neg$.

The base of induction is the case when ϕ is the formula $x \in y$, already checked in Example 17.1.3. As we go through the induction steps, we will reduce the arbitrarily long list of free variables $x^1, \dots, x^k$ to a minimally nontrivial list for each step, for the sake of brevity.

The step $\neg$. Let $\phi(x)$ be a bounded $\in$-formula with a single free variable x, and $\langle a_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$. Suppose that the lemma is established for $\phi(\langle a_n \rangle)$, and let us prove it for $\neg\, \phi(\langle a_n \rangle)$.

Indeed, formula $\neg\, \phi(\langle a_n \rangle_{\mathcal{F}})$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*\!\!\in \rangle$ if and only if $\phi(\langle a_n \rangle_{\mathcal{F}})$ is false in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*\!\!\in \rangle$. The latter holds if and only if (by the inductive hypothesis) $T_{\phi(\langle a_n \rangle)} \notin \mathcal{F}$ or equivalently $T_{\neg\,\phi(\langle a_n \rangle)} \in \mathcal{F}$ since obviously $T_\phi = \mathbb{N} \setminus T_{\neg\,\phi}$, and one and only one set in a pair of complementary sets belongs to $\mathcal{F}$ (as $\mathcal{F}$ is an ultrafilter).

The step $\wedge$. Here one proves the result for $\phi(a) \wedge \psi(a)$ assuming that the lemma is established separately for $\phi(a)$ and $\psi(a)$. We leave it as an exercise.

The step $\exists$. Assume that the lemma is established for a bounded formula $\phi(x, y, z)$, and prove it for $(\exists z \in x)\, \phi(x, y, z)$. Let $\langle a_n \rangle, \langle b_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$. We have to prove that the statement

(A) $(\exists z \in \langle a_n \rangle_{\mathcal{F}})\ \phi(\langle a_n \rangle_{\mathcal{F}}, \langle b_n \rangle_{\mathcal{F}}, z)$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*\!\!\in \rangle$

is equivalent to

(B) the set $L = T_{\exists z \in \langle a_n \rangle\, \phi(\langle a_n \rangle, \langle b_n \rangle, z)}$ belongs to $\mathcal{F}$.

Note that $L = \{n : \exists z \in a_n\, (\phi(a_n, b_n, z) \text{ is true in } V(X))\}$.

By definition, (A) is equivalent to

(C) there is a sequence $\langle c_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ such that both $\langle c_n \rangle_{\mathcal{F}} \ {}^*\!\!\in \langle a_n \rangle_{\mathcal{F}}$ and $\phi(\langle a_n \rangle_{\mathcal{F}}, \langle b_n \rangle_{\mathcal{F}}, \langle c_n \rangle_{\mathcal{F}})$ are true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*\!\!\in \rangle$.

By the inductive hypothesis for ϕ and the base of induction result, the requirements of (C) hold if and only if each of the sets

$$\begin{aligned} U'_{\langle c_n \rangle} = T_{\langle c_n \rangle \in \langle a_n \rangle} &= \{n \colon c_n \in a_n\}, \text{ and} \\ U''_{\langle c_n \rangle} = T_{\phi(\langle a_n \rangle, \langle b_n \rangle, \langle c_n \rangle)} &= \{n \colon \phi(a_n, b_n, c_n) \text{ is true in } V(X)\} \end{aligned}$$

belongs to $\mathcal{F}$, which is equivalent (as $\mathcal{F}$ is an ultrafilter) to $U_{\langle c_n \rangle} \in \mathcal{F}$, where

$$U_{\langle c_n \rangle} = U'_{\langle c_n \rangle} \cap U''_{\langle c_n \rangle} = T_{\langle c_n \rangle \in \langle a_n \rangle \wedge \phi(\langle a_n \rangle, \langle b_n \rangle, \langle c_n \rangle)}.$$

On the other hand, the set L of (B) satisfies $U_{\langle c_n\rangle} \subseteq L$ because if $n \in U_{\langle c_n\rangle}$ then taking $z = c_n$ satisfies $\phi(a_n, b_n, z)$ in $V(X)$. Therefore if $T \in \mathcal{F}$ then $L \in \mathcal{F}$ as $\mathcal{F}$ is an ultrafilter. This completes the proof that (A) implies (B).

To prove the converse suppose that $L \in \mathcal{F}$. Recall that $V(X)^{\mathbb{N}}_{\mathtt{bd}} = \bigcup_m V_m(X)^{\mathbb{N}}$, hence the sequence $\langle a_n\rangle$ belongs to some $V_m(X)$. If $n \in L$ then by definition there is some $z = c_n \in a_n$ such that $\phi(a_n, b_n, z)$ is true in $V(X)$.[1] It follows that $m > 0$ and $c_n \in V_{m-1}(X)$, by Corollary 16.1.5. If $n \notin L$ then let $c_n \in V_{m-1}(X)$ be arbitrary. The construction of the sequence $\langle c_n \colon n \in \mathbb{N}\rangle \in V_{m-1}(X)^{\mathbb{N}}$ uses the axiom of choice. Thus $\langle c_n\rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ and if $n \in L$ then n belongs to the set $U_{\langle c_n\rangle} = U'_{\langle c_n\rangle} \cap U''_{\langle c_n\rangle}$ as above by construction. It follows that the sets $U'_{\langle c_n\rangle}, U''_{\langle c_n\rangle}$ belong to $\mathcal{F}$, and hence we have (C) and (A).

This completes the proof of the Łoś theorem. □

17.2. Elementary embedding

Corollary 17.2.1. *The map $x \longmapsto \langle x\rangle_{\mathcal{F}}$ is an* elementary embedding *of the structure $\langle V(X); \in\rangle$ to $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$.*

In other words, if $\phi(x^1, \ldots, x^k)$ is a bounded $\in$-formula and we have $a^1, \ldots, a^k \in V(X)$, then the sentence $\phi(a^1, \ldots, a^k)$ is true in $V(X)$ if and only if the sentence $\phi(\langle a^1\rangle_{\mathcal{F}}, \ldots, \langle a^k\rangle_{\mathcal{F}})$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$.

Proof. Recall that $\langle a^j\rangle$ is the infinite sequence $\langle a^j_n : n \in \mathbb{N}\rangle$ such that $a^j_n = a^j$ for all n. Therefore the truth set $T_{\phi(\langle a^1\rangle, \ldots, \langle a^k\rangle)}$ is equal to $\mathbb{N}$ if $\phi(a^1, \ldots, a^k)$ is true in $V(X)$ and is equal to $\varnothing$ otherwise. Now an application of Lemma 17.1.2 concludes the proof. □

This result ends the first part of the proof of Theorem 16.5.4. We defined an embedding $x \longmapsto \langle x\rangle_{\mathcal{F}}$ of the universe $\langle V(X); \in\rangle$ to

$$\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$$

which satisfies Transfer by Corollary 17.2.1. The final step of the proof of the theorem is embedding $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$ into the universe $V({}^*X)$, where ${}^*X = \{\langle a_n\rangle_{\mathcal{F}} \colon \langle a_n\rangle \in X^{\mathbb{N}}\}$, and, we recall, $X = V_0(X)$.

Lemma 17.2.2. *Assume that γ is an infinite ordinal and X is a γ-base set. Then ${}^*X$ is a $(\gamma + 4)$-base set, hence a base set by Lemma* 16.2.4.

Proof. By definition, each $\langle a_n\rangle_{\mathcal{F}} \in {}^*X$ is a non-empty set of pairwise $\sim_{\mathcal{F}}$-equivalent sequences $\langle a_n\rangle = \langle a_n \colon n \in \mathbb{N}\rangle$ of elements $a_n \in X$.

[1] Making such choices simultaneously for all $n \in L$ requires Countable Choice.

It remains to prove that $\mathtt{rank}(\langle a_n\rangle) = \gamma + 4$. Note that a sequence of this form is formally equal to the set of all ordered pairs

$$\langle n, a_n\rangle = \big\{\{\{n\}, \{n, a_n\}\} : n \in \mathbb{N}\big\},$$

where

$\mathtt{rank}(n) = n,$
$\mathtt{rank}(a_n) = \gamma + 1$ (as X is a γ-base set),
$\mathtt{rank}(\{n\}) = n + 1,$
$\mathtt{rank}(\{n, a_n\}) = \gamma + 2,$
$\mathtt{rank}(\{\{n\}, \{n, a_n\}\}) = \gamma + 3,$
$\mathtt{rank}(\langle a_n : n \in \mathbb{N}\rangle) = \gamma + 4,$

as required. □

The lemma shows that $V({}^*X)$ is a legitimate superstructure which satisfies all the results of Sections 16.1 and 16.2.

Definition 17.2.3. Define a map $h\colon V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F} \to V({}^*X)$ as follows. Suppose that $\langle a_n\rangle_{\mathcal{F}} \in V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$. It can be assumed, by Corollary 16.6.7, that $\langle a_n\rangle$ is totally of type m for some m, so that $a_n \in V_m(X) \setminus V_{m-1}(X)$ (or just $a_n \in X = V_0(X)$ provided $m = 0$) for all n. The definition of $h(\langle a_n\rangle_{\mathcal{F}})$ goes on by induction on m.

(1) If $m = 0$ then we put $h(\langle a_n\rangle_{\mathcal{F}}) = \langle a_n\rangle_{\mathcal{F}}$.
In other words, if $\langle a_n\rangle \in X^{\mathbb{N}}$ then $h(\langle a_n\rangle_{\mathcal{F}}) = \langle a_n\rangle_{\mathcal{F}} \in {}^*X$.

(2) If $m \geq 1$ then we put

$$h(\langle a_n\rangle_{\mathcal{F}}) = \big\{h(\langle b_n\rangle_{\mathcal{F}}) : \langle b_n\rangle \in V_{m-1}(X)^{\mathbb{N}} \text{ and } \langle b_n\rangle_{\mathcal{F}} \mathrel{{}^*{\in}} \langle a_n\rangle_{\mathcal{F}}\big\}.$$

This type of definition is called *the Mostowski collapse*; see Theorem 4.4.9 in [**Chang & Keisler 1990**].

Exercise 17.2.4. Prove by induction on m that if $\langle a_n\rangle \in V_m(X)^{\mathbb{N}}$ then $h(\langle a_n\rangle_{\mathcal{F}}) \in V_m({}^*X)$. □

Lemma 17.2.5. *The map h is a bijection, and h is an isomorphism in the sense that if $\langle a_n\rangle, \langle b_n\rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ then*

$$\langle a_n\rangle_{\mathcal{F}} \mathrel{{}^*{\in}} \langle b_n\rangle_{\mathcal{F}} \quad \textit{if and only if} \quad h(\langle a_n\rangle_{\mathcal{F}}) \in h(\langle b_n\rangle_{\mathcal{F}}).$$

Proof. To prove that h is a bijection, suppose that $\langle a_n\rangle, \langle b_n\rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}$ and $\langle a_n\rangle_{\mathcal{F}} \neq \langle b_n\rangle_{\mathcal{F}}$. Then the truth set $T = \{n \in \mathbb{N} : a_n \neq b_n\}$ belongs to $\mathcal{F}$ by Lemma 17.1.2. It can be assumed, by Corollary 16.6.7, that $\langle a_n\rangle$, $\langle b_n\rangle$ are totally of type m, resp., m', for some m, m'.

Case 1: $m = m' = 0$, so that $a_n, b_n \in X = V_0(X)$ for all n. Then $h(\langle a_n\rangle_{\mathcal{F}}) = \langle a_n\rangle_{\mathcal{F}}$ and $h(\langle b_n\rangle_{\mathcal{F}}) = \langle b_n\rangle_{\mathcal{F}}$ by construction, and on the other hand, we have $\langle a_n\rangle_{\mathcal{F}} \neq \langle b_n\rangle_{\mathcal{F}}$ since $T \in \mathcal{F}$.

Case 2: $m' = 0$ and $m \geq 1$, so that $b_n \in X = V_0(X)$ and $a_n \in V_m(X) \setminus V_{m-1}(X)$ for all n. Then $h(\langle b_n \rangle_{\mathcal{F}}) = \langle b_n \rangle_{\mathcal{F}} \in {}^*X$ while $h(\langle a_n \rangle_{\mathcal{F}}) \subseteq V({}^*X)$ by construction, so that $h(\langle b_n \rangle_{\mathcal{F}}) \neq h(\langle a_n \rangle_{\mathcal{F}})$ by Lemma 17.2.2. (Recall Definition 16.1.4.)

Case 3: $m = 0$ and $m' \geq 1$, similar.

Case 4: $m = m' \geq 1$. Then, for any n, a_n and b_n belong to $V_m(X) \setminus V_{m-1}(X)$, therefore $a_n \subseteq V_{m-1}(X)$ and $b_n \subseteq V_{m-1}(X)$. In addition, if $n \in T$ then $a_n \neq b_n$, hence there exists some $c_n \in V_{m-1}(X)$ satisfying $c_n \in a_n \setminus b_n$ or $c_n \in b_n \setminus a_n$. It follows that one of the sets

$$T' = \{n \in T \colon c_n \in a_n \setminus b_n\}, \quad T'' = \{n \in T \colon c_n \in b_n \setminus a_n\}$$

belongs to $\mathcal{F}$; suppose that $T' \in \mathcal{F}$. If $n \notin T$ then let $c_n \in V_{m-1}(X)$ be arbitrary. Then $\langle c_n \rangle \in V_{m-1}(X)$. Now we have $\langle c_n \rangle \ {}^*\!\!\in \langle a_n \rangle$ but $\neg \langle c_n \rangle \ {}^*\!\!\in \langle b_n \rangle$, hence, by Definition 17.2.3, $h(\langle c_n \rangle_{\mathcal{F}}) \in h(\langle a_n \rangle_{\mathcal{F}})$ but $h(\langle c_n \rangle_{\mathcal{F}}) \notin h(\langle b_n \rangle_{\mathcal{F}})$. Therefore $h(\langle a_n \rangle_{\mathcal{F}}) \neq h(\langle b_n \rangle_{\mathcal{F}})$.

Case 5: $m, m' \geq 1$ and $m \neq m'$, say $m' < m$. For any n we have $a_n \in V_m(X) \setminus V_{m-1}(X)$, hence $a_n \subseteq V_{m-1}(X)$. Note that $a_n \subseteq V_k(X)$ is impossible for $k < m - 1$ as otherwise

$$a_n \in V_{k+1}(X) \subseteq V_{m-1}(X),$$

a contradiction. Thus there is an element $c_n \in a_n$, $c_n \in V_{m-1}(X) \setminus V_{m-2}(X)$, in particular, $c_n \in V_{m-1}(X) \setminus V_{m'-1}(X)$. Then $\langle c_n \rangle \in V_{m-1}(X)$ and $\langle c_n \rangle \ {}^*\!\!\in \langle a_n \rangle$, hence $h(\langle c_n \rangle_{\mathcal{F}}) \in h(\langle a_n \rangle_{\mathcal{F}})$. On the other hand, we have $c_n \notin b_n$ by Corollary 16.1.5, hence $\neg \langle c_n \rangle \ {}^*\!\!\in \langle b_n \rangle$ and $h(\langle c_n \rangle_{\mathcal{F}}) \notin h(\langle b_n \rangle_{\mathcal{F}})$. Thus we still have $h(\langle a_n \rangle_{\mathcal{F}}) \neq h(\langle b_n \rangle_{\mathcal{F}})$.

The claim that h is a bijection is established.

Now we prove the isomorphism claim. Let $\langle a_n \rangle, \langle b_n \rangle$ be sequences in $V(X)^{\mathbb{N}}_{\mathtt{bd}}$. As above, it can be assumed that $\langle a_n \rangle$, $\langle b_n \rangle$ are totally of type m, resp., m', for some $m, m' \in \mathbb{N}$. Suppose that $\langle a_n \rangle_{\mathcal{F}} \ {}^*\!\!\in \langle b_n \rangle_{\mathcal{F}}$. By definition the set $N = \{n \in \mathbb{N} \colon a_n \in b_n\}$ belongs to $\mathcal{F}$. Then $a_n \in V_{m'-1}(X)$ for all $n \in N$ by Corollary 16.1.5, and hence we have $h(\langle a_n \rangle_{\mathcal{F}}) \in h(\langle b_n \rangle_{\mathcal{F}})$ by Definition 17.2.3.

If conversely $h(\langle a_n \rangle_{\mathcal{F}}) \in h(\langle b_n \rangle_{\mathcal{F}})$ then immediately $\langle a_n \rangle_{\mathcal{F}} \ {}^*\!\!\in \langle b_n \rangle_{\mathcal{F}}$ still by Definition 17.2.3. □

Definition 17.2.6. Let ${}^*V(X) = \{h(\langle a_n \rangle_{\mathcal{F}}) \colon \langle a_n \rangle \in V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}\}$, the range of h. If $m \in \mathbb{N}$ then let

$${}^*V_m(X) = \{h(\langle a_n \rangle_{\mathcal{F}}) \colon \langle a_n \rangle \in V_m(X)^{\mathbb{N}}/\mathcal{F}\}. \qquad \square$$

Exercise 17.2.7. Prove using Lemma 17.2.5 that the sets ${}^*V_m(X)$ satisfy ${}^*V_m(X) = {}^*V(X) \cap V_m({}^*X)$, ${}^*V_{m+1}(X) \subseteq \mathcal{P}({}^*V_m(X))$, and finally ${}^*V(X) = \bigcup_m {}^*V_m(X) \subseteq V({}^*X)$. □

Thus ${}^*V(X)$ is a subuniverse of the full universe $V({}^*X)$ over ${}^*X = {}^*V_0(X)$. Let us use letters ξ, η to denote arbitrary elements of $V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$, so that if $\xi \in V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$ then $h(\xi) \in {}^*V(X)$. As h is an isomorphism by Lemma 17.2.5, we immediately obtain the following.

Corollary 17.2.8. *If $\phi(x^1, \dots, x^k)$ is a bounded $\in$-formula and $\xi^1, \dots, \xi^k \in V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}$, then $\phi(\xi^1, \dots, \xi^k)$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$ if and only if the sentence $\phi(h(\xi^1), \dots, h(\xi^k))$ is true in $\langle {}^*V(X); \in\rangle$.* $\square$

17.3. Completing the proof

Now we are ready to complete the proof of Theorem 16.5.4. We claim that ${}^*X$, ${}^*V(X)$, and the map $x \longmapsto {}^*x$ constitute a nonstandard extension of $V(X)$. The proof is obtained by combining Corollaries 17.2.1 and 17.2.8. Namely, if $x \in V(X)$ then let ${}^*x = h(\langle x\rangle_{\mathcal{F}})$, so that $x \longmapsto {}^*x$ maps $V(X)$ onto ${}^*V(X) \subseteq V({}^*X)$. Furthermore, Lemma 17.2.5 ensures that condition (I) of Section 16.5 holds. Now let us check condition (II) of Section 16.5. Consider a bounded formula $\phi(x_1, \dots, x_k)$, and arbitrary elements $a^1, \dots, a^k \in V(X)$. Then the formula $\phi(a^1, \dots, a^k)$ is true in $V(X)$ if and only if $\phi(\langle a^1\rangle_{\mathcal{F}}, \dots, \langle a^k\rangle_{\mathcal{F}})$ is true in $\langle V(X)^{\mathbb{N}}_{\mathtt{bd}}/\mathcal{F}; {}^*{\in}\rangle$ (by Corollary 17.2.1), if and only if the formula $\phi({}^*a^1, \dots, {}^*a^k)$ is true in ${}^*V(X)$ (by Corollary 17.2.8).

Note that both $V(X)$ and ${}^*V(X)$ are viewed as $\in$-structures with the true membership relation $\in$.

It remains to prove that $X \subsetneqq {}^*X$ properly, or, to be more precise, there is an element $y \in {}^*X$ not equal to any ${}^*x$, $x \in X$.

Let $a_0, a_1, a_2, \dots$ be an infinite sequence of pairwise distinct elements $a_n \in X$. Then $\langle a_n\rangle = \langle a_n : n \in \mathbb{N}\rangle$ belongs to $X^{\mathbb{N}} = V_0(X)^{\mathbb{N}}$. It follows that $\xi = h(\langle a_n\rangle_{\mathcal{F}}) \in {}^*X$. We now prove that if $x \in X$ then $\xi \neq {}^*x$. By definition one has to show that $\langle a_n\rangle \not\sim \langle x\rangle$, that is, the set $N = \{n : a_n \neq x\}$ belongs to $\mathcal{F}$. However N contains all natural numbers with the possible exception of a single index n satisfying $a_n = x$. Hence $N \in \mathcal{F}$. This completes the proof of Theorem 16.5.4. $\square$

Rephrasing condition (II) of Section 16.5, we obtain:

Corollary 17.3.1 (Transfer). *The map $x \longmapsto {}^*x$ is an elementary $\in$-embedding of the structure $\langle V(X); \in\rangle$ to $\langle {}^*V(X); \in\rangle$. In other words, if $\phi(x^1, \dots, x^k)$ is a bounded $\in$-formula and $a^1, \dots, a^k \in V(X)$, then $\phi(a^1, \dots, a^k)$ is true in $V(X)$ if and only if $\phi({}^*a^1, \dots, {}^*a^k)$ is true in ${}^*V(X)$.* $\square$

CHAPTER 18

Further issues in foundations

18.1. Definability

In Chapter 4.7 we presented a construction of the hyperreal line via filters and ultrapowers. For a broader picture, it may be useful to comment on the issue of definability.

At the end of the 20th century, many mathematicians felt that there could be no way of defining a suitable hyperreal line, since its definition involves nonconstructive foundational material. Therefore it came as a surprise to many people when Kanovei and Shelah proved in 2004 that, in fact, a hyperreal line is indeed definable [**Kanovei & Shelah 2004**].

What this means is that there exists a specific set-theoretic formula that defines a hyperreal line in the set-theoretic universe, in terms of nothing else beyond the commonly accepted Zermelo–Fraenkel axioms. This might seem surprising. To clarify the situation, one would compare it with that for Lebesgue measure. It is widely known and accepted as fact that the Lebesgue measure is σ-additive. Two items need to be pointed out here:

(1) to give an initial characterisation of the Lebesgue measure m, one does not need the axiom of choice (AC), but AC plays a crucial role in establishing suitable properties of m;
(2) in particular, to prove that the Lebesgue measure m is σ-additive, one requires AC.

Here item (2) follows from the existence of a model of ZF (Zermelo–Fraenkel axioms without AC) where the real line is a countable union of countable sets. This model is called the Feferman–Levy model [**Feferman & Levy 1963**].[1] See also [**Jech 1973**, chapter 10].

We see that, while some rather elementary properties of m (finite additivity or measurability of every open set of real numbers, to name a few) can be established in ZF, a proof of the σ-additivity of m necessarily requires (some version of) AC.

[1] One can show furthermore that in this model there exists a positive function with zero Lebesgue integral.

With regard to its foundational status, the situation is similar for a hyperreal line. One can give an initial characterisation of a hyperreal line without AC (cf. Kanovei–Shelah),[2] but actually to prove something about it one would need some version of AC, which can be just the full AC of ZFC (as in Kanovei–Shelah); recently it has been shown that the combination of the countable choice for sets of reals and the existence of an ultrafilter over $\mathbb{N}$ with a wellorderable base also suffices [**Herzberg et al. 2017**].

18.2. Conservativity

The various frameworks developed by Robinson and since, including those of Nelson and Hrbacek, are conservative extensions of ZFC, and in this sense are part of classical mathematics.

As far as conservativity is concerned, the following needs to be kept in mind. Robinson showed that, given a proof exploiting infinitesimals, there always exists an infinitesimal-free paraphrase.

However, this is merely a theoretical result. In practice, such a paraphrase may involve an uncontrolled increase in the complexity of the proof, and correspondingly a decrease in its comprehensibility, at least in principle.

From the point of computer scientist one should certainly be able to appreciate the difference.

Similarly, one can note that any result using the real numbers can be restated in terms of the rationals using sequences, and even in terms of the integers, etc. The ancient Greeks used formulations that would be very difficult for us to follow if we did not use modern translations in terms of extended number systems that have since become widely accepted and are considered *intuitive*. The related controversy over Unguru's proposal is well known.

The following two additional points should be kept in mind.

First, the conservativity property of the pair ZFC, IST does not hold for pairs like A_n, A_n^*, where A_n is the n-th order arithmetic and A_n^* its nonstandard extension, in fact A_n^* is rather comparable to the standard A_{n+1}. This profound result of Henson–Keisler around 1994 is well known to specialists working in Robinson's framework.[3]

Second, IST yields some principally new mathematics with respect to ZFC, in the sense that there are IST sentences not equivalent to any ZFC sentence. This was established in [**Kanovei & Reeken 2004**].

[2] We might want to be more precise here.

[3] provide references and exact formulations.

This can be compared to the obvious fact that any sentence about $\mathbb{C}$ admits an equivalent sentence about $\mathbb{R}$.

The following comments by Henson and Keisler usefully illustrate the point:

> It is often asserted in the literature that any theorem which can be proved using nonstandard analysis can also be proved without it. The purpose of this paper is to show that this assertion is wrong, and in fact there are theorems which can be proved with nonstandard analysis but cannot be proved without it. There is currently a great deal of confusion among mathematicians because the above assertion can be interpreted in two different ways. First, there is the following correct statement: any theorem which can be proved using nonstandard analysis can be proved in Zermelo-Fraenkel set theory with choice, ZFC, and thus is acceptable by contemporary standards as a theorem in mathematics. Second, there is the erroneous conclusion drawn by skeptics: any theorem which can be proved using nonstandard analysis can be proved without it, and thus there is no need for nonstandard analysis. [**Henson & Keisler 1986**, p. 377]

Henson and Keisler go on to analyze the reasons for such a confusion:

> The reason for this confusion is that the set of principles which are accepted by current mathematics, namely ZFC, is much stronger than the set of principles which are actually used in mathematical practice. It has been observed (see [F] and [S]) that almost all results in classical mathematics use methods available in second order arithmetic with appropriate comprehension and choice axiom schemes. This suggests that mathematical practice usually takes place in a conservative extension of some system of second order arithmetic, and that it is difficult to use the higher levels of sets. In this paper we shall consider systems of nonstandard analysis consisting of second order nonstandard arithmetic with saturation principles (which are frequently used in practice in nonstandard arguments). We shall prove that nonstandard analysis (i.e. second order nonstandard arithmetic) with the ω_1-saturation axiom scheme has the same

> strength as third order arithmetic. This shows that in principle there are theorems which can be proved with nonstandard analysis but cannot be proved by the usual standard methods. (ibid.)

In short, A_n^* is comparable to the standard A_{n+1} rather than to A_n itself.

CHAPTER 19

Burgers equation

We are interested in the article [**Benci–Luperi Baglini 2017**]. For preliminaries we mostly follow [**Pinchover–Rubinstein 2005**].

19.1. Normal direction of a surface

A crucial role in the method of solving quasilinear PDEs called the method of characteristics is played by the specific form of the normal direction to a surface in $M \subseteq \mathbb{R}^3$ given by the graph of a function $u = u(x, y)$.

Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^3$ be a regular parametrized surface, and let $\rho_x = \frac{\partial\rho}{\partial x}$ and $\rho_y = \frac{\partial\rho}{\partial y}$.

Definition 19.1.1. The unit normal vector $n(x, y)$ to a regular surface $M \subseteq \mathbb{R}^3$ at the point $\rho(x, y) \in \mathbb{R}^3$ is defined in terms of the vector product, as follows:

$$n = \frac{\rho_1 \times \rho_2}{|\rho_1 \times \rho_2|},$$

so that $\langle n, \rho_x\rangle = \langle n, \rho_y\rangle = 0$.

Either the vector $\frac{\rho_1\times\rho_2}{|\rho_1\times\rho_2|}$ or the vector $\frac{\rho_2\times\rho_1}{|\rho_1\times\rho_2|} = -\frac{\rho_1\times\rho_2}{|\rho_1\times\rho_2|}$ can be taken to be a normal vector to the surface.

Theorem 19.1.2. *Consider the surface $M \subseteq \mathbb{R}^3$ given by the graph of a function $u(x, y)$. Then the vector $\begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix}$ is perpendicular to the tangent plane T_p to M at the point $p = (x, y, u(x, y)) \in M$.*

Proof. A standard parametrisation of the graph of u is $\rho(x, y) = (x, y, u(x, y))$. Then $\rho_x = (1, 0, u_x)^t$ and $\rho_y = (0, 1, u_y)^t$. Taking the cross product, we obtain

$$\det\begin{pmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ 1 & 0 & u_x \\ 0 & 1 & u_y \end{pmatrix} = -u_x\overrightarrow{i} - u_y\overrightarrow{j} + \overrightarrow{k}$$

which is the opposite of the vector $\begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix}$. □

19.2. Quasilinear PDEs

The general form of a quasilinear PDE is

$$a(x,y,u)u_x + b(x,y,u)u_y = c(x,y,u). \qquad (19.2.1)$$

The graph of $u(x,y)$ is a surface in $\mathbb{R}^3$, called a solution surface M.

EXAMPLE 19.2.1. Consider the PDE

$$u_x = c_0 u + c_1(x,y). \qquad (19.2.2)$$

To obtain a unique solution we must provide an initial condition. A natural initial condition is given by a curve expected to lie on the solution surface M. Such a problem is called a Cauchy problem. For example, we can choose the initial condition

$$u(0,y) = y.$$

For a fixed value of y we obtain an ODE. The corresponding homogeneous equation $u_x = c_0 u$ is solved as usual by separation of variables to yield $\frac{du}{u} = c_0 dx$, so that $\ln u = c_0 x$ and therefore $u = e^{c_0 x}$, etc. The inhomogeneous equation (19.2.2) is solved by the function

$$u(x,y) = e^{c_0 x}\left[\int_0^x e^{-c_0 \xi} c_1(\xi,y)d\xi + y\right]$$

as is easily checked by differentiating.

REMARK 19.2.2. The solution surface M can be thought of as stitched together all the curves from a 1-parameter family of curves resulting from solving the ODE separately for each value of y.

19.3. The method of characteristics

This method of solution of first-order PDEs was developed by Hamilton. Consider the case of a *linear* PDE

$$a(x,y)u_x + b(x,y)u_y = c_0(x,y)u + c_1(x,y). \qquad (19.3.1)$$

We present the initial curve on the solution surface M parametrically as the curve Γ given by

$$\Gamma(s) = (x_0(s), y_0(s), u_0(s)), \quad s \in I = (\alpha, \beta). \qquad (19.3.2)$$

The PDE (19.3.1) with $c = c_0 u + c_1(x, y)$ can then be rewritten in the form

$$\left\langle \begin{pmatrix} a \\ b \\ c_0 u + c_1 \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0. \tag{19.3.3}$$

REMARK 19.3.1. The second vector gives the normal direction of the surface M as in Theorem 19.1.2.

DEFINITION 19.3.2. The system of ordinary differential equations

$$\begin{cases} \frac{dx}{dt}(t) = a(x(t), y(t)) \\ \frac{dy}{dt}(t) = b(x(t), y(t)) \\ \frac{du}{dt}(t) = c_0(x(t), y(t))u(t) + c_1(x(t), y(t)) \end{cases}$$

is referred to as the characteristic equations.

Since the vector formed by the functions on the right-hand side of the characteristic equations is always tangent to the surface by formula (19.3.3), the solution curve is constrained to stay on the surface M provided suitable smoothness and nondegeneracy conditions are ensured.

DEFINITION 19.3.3. The solution curves are called the characteristic curves. Their projections to the x, y plane are called the characteristics.

Since each characteristic curve emanates from a different point of the initial curve $\Gamma(s)$, we will incorporate the parameter s into the notation by writing the family of curves as

$$(x(t, s), y(t, s), u(t, s)).$$

The initial conditions can then be written as

$$x(0, s) = x_0(s), y(0, s) = y_0(s), u(0, s) = u_0(s),$$

where x_0, y_0, u_0 are the components of the initial curve Γ of (19.3.2). The characteristic equations then take the form

$$\begin{cases} x_t(t, s) = a(x(t), y(t)) \\ y_t(t, s) = b(x(t), y(t)) \\ u_t(t, s) = c_0(x(t), y(t))u(t) + c_1(x(t), y(t)) \end{cases}$$

with initial conditions $x(0, s) = x_0(s), y(0, s) = y_0(s), u(0, s) = u_0(s)$.

19.4. Passing from parametric representation to graph rep.

EXAMPLE 19.4.1. Solve the equation $u_x + u_y = 2$ subject to the initial condition $u(x,0) = x^2$.

To illustrate the application of the method of characteristics, we write the equation as a scalar product

$$\left\langle \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0. \tag{19.4.1}$$

The characteristic equations are then

$$\begin{cases} x_t(t,s) = 1 \\ y_t(t,s) = 1 \\ u_t(t,s) = 2 \end{cases}$$

with initial conditions $x(0,s) = s, y(0,s) = 0, u(0,s) = s^2$. Integrating, we obtain the family of characteristic curves

$$x(s,t) = t + f_1(s), \quad y(t,s) = t + f_2(s), \quad u(s,t) = 2t + f_3(s).$$

The initial conditions given in Example 19.4.1 force the choice of the functions f_i, so that we obtain

$$x(t,s) = t + s, \quad y(t,s) = t, \quad u(t,s) = 2t + s^2. \tag{19.4.2}$$

This provides a parametric representation of the solution surface.

The next step is the passage from the parametric representation in terms of t, s to a representation as graph of a function of x, y. In this case, the relevant equations (19.4.2) are easily inverted to yield

$$t = y, \quad s = x - y.$$

Hence the graph representation of the solution surface is immediate from the last equation in (19.4.2) and is given the graph of the function

$$u(x,y) = 2y + (x-y)^2.$$

19.5. Study of transversality condition

We study the PDE $au_x + bu_y = c$. Here $\Gamma(s) = (x_0(s), y_0(s), u_0(s))$, $s \in I = (\alpha, \beta)$ is the initial curve. In the linear case, the characteristic equations are

$$\begin{cases} \frac{dx}{dt}(t) = a(x(t), y(t)) \\ \frac{dy}{dt}(t) = b(x(t), y(t)) \\ \frac{du}{dt}(t) = c_0(x(t), y(t))u(t) + c_1(x(t), y(t)) \end{cases}$$

We applied the method of characteristics to a simple equation in Section 19.4. Applying the method to more complicated equations may involve several problems:

(1) The inversion is impossible even locally for small time t, if $\det J = \frac{\partial(x,y)}{\partial(t,s)} = 0$ along the initial curve ($t = 0$). Here $\det J = \frac{\partial x}{\partial t}\frac{\partial y}{\partial s} - \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} = (y_0)_s\, a - (x_0)_s\, b$. The condition $\det J \neq 0$ is the transversality condition.
(2) A possible global problem is that distinct characteristic curves may collide, or even if the characteristics (in the x, y plane) collide.

EXAMPLE 19.5.1. Solve the linear equation $-yu_x + xu_y = u$ subject to the initial condition $u(x, 0) = \psi(x)$.

Note that here y plays the role of time and the solution develops from the initial curve at time $y = 0$, similar to Burgers equation (see Section 19.7).

For illustrative purposes we write the equation in the form of a scalar product:

$$\left\langle \begin{pmatrix} -y \\ x \\ u \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0. \tag{19.5.1}$$

Therefore the characteristic equations take the form

$$\begin{cases} x_t(t,s) = -y \\ y_t(t,s) = x \\ u_t(t,s) = u \end{cases}$$

with initial conditions specified by the curve $\Gamma(s)$ given by

$$x(0, s) = s, \quad y(0, s) = 0, \quad u(0, s) = \psi(s).$$

Let us examine the transversality condition. Note that $\frac{\partial y}{\partial t} = x$ and along the initial curve we have $x = s$. Thus along the initial curve we have $\frac{\partial y}{\partial t} = s$. Therefore we compute

$$\det J = \frac{\partial x}{\partial t}\frac{\partial y}{\partial s} - \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} = 0 \cdot 0 - 1 \cdot s = -s.$$

Thus we expect a unique solution locally near the initial curve, except perhaps at the point $x = 0$.

The first two equations describe the linear oscillator. The solution of the characteristic equations is given by

$$\begin{cases} x(t,s) = f_1(s)\cos t + f_2(s)\sin t \\ y(t,s) = f_1(s)\sin t - f_2(s)\cos t \\ u(t,s) = e^t f_3(s). \end{cases} \tag{19.5.2}$$

Substituting the initial condition into the solution (19.5.2) leads to the parametric form of the solution surface:

$$(x,y,u) = (s\cos t, s\sin t, e^t\psi(s)).$$

The characteristics obtained as the projection to the x, y plane are the family of concentric circles. Isolating s and t one obtains

$$u(x,y) = \psi(\sqrt{x^2+y^2})e^{\arctan(y/x)}$$

Note that each of the characteristics meets the initial curve (the x-axis) twice.

19.6. A quasilinear equation

New phenomena appear for quasilinear equations.

EXAMPLE 19.6.1. Solve the quasilinear equation $(y+u)u_x + yu_y = x - y$ subject to initial condition $u(x,1) = 1 + x$.

The dot product form of this equation is

$$\left\langle \begin{pmatrix} y+u \\ y \\ x-y \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0. \tag{19.6.1}$$

Note that the coefficient $a = y + u$ of u_x is dependent on u making the equation quasilinear (rather than linear). A consequence is that the characteristic equations also involve u:

$$\begin{cases} x_t(t,s) = y + u \\ y_t(t,s) = y \\ u_t(t,s) = x - y. \end{cases} \tag{19.6.2}$$

The initial condition can be described by the parametrized curve Γ given by

(1) $x(0,s) = s$,
(2) $y(0,s) = 1$,
(3) $u(0,s) = 1 + s$.

Let us check the transversality condition along the initial curve Γ. We have $a = y + u = 2 + s$ while $b = 1$. Along the initial curve we have $\frac{\partial x}{\partial t} = y + u = 2 + s$, $\frac{\partial y}{\partial t} = y = 1$, $\frac{\partial x}{\partial s} = 1$, $\frac{\partial y}{\partial s} = 0$. We compute

$$\det J = \frac{\partial x}{\partial t}\frac{\partial y}{\partial s} - \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} = (2+s)\cdot 0 - 1 \cdot 1 = -1 \neq 0$$

and therefore inversion is possible at least locally.

Solving the characteristic equation for y we find $y(t,s) = e^t$.

Remark 19.6.2. To make further progress we need to deal with the fact that the equation for x_t depends on u so it is not a traditional ODE.

We use the following trick. Consider the sum $x_t + u_t = y + u - x - y = x + u$ as is immediate from (19.6.2). Setting $v = x + u$ we obtain the equation $v_t = v$. Solving this ODE we obtain $v = f(s)e^t$. The initial conditions (1) and (3) give $v = (1+2s)e^t$, i.e., $x + u = (1+2s)e^t$, so that

$$u = (1+2s)e^t - x. \tag{19.6.3}$$

At this point [**Pinchover–Rubinstein 2005**] turn pretty terse.

Substituting (19.6.3) into the first equation of (19.6.2) we obtain

$$x_t = y + (1+2s)e^t - x.$$

Meanwhile $y = f_2(s)e^t$ and the initial condition (2) yields $y = e^t$. Therefore $x_t = e^t + (1+2s)e^t - x$ or

$$x_t = -x + (2+2s)e^t.$$

This is an ordinary nonlinear ODE and one can apply integrating factors to solve the problem.

In [**Pinchover–Rubinstein 2005**] one finds a general result on local existence of solutions. The example treated in this section is meant to illustrate the change of coordinates involved in the proof.

19.7. Burgers equation

There is a local existence theorem for solutions of PDE. However there are cases of physical interest where we need solution beyond the point where the smooth solution breaks down. A case in point is the equation

$$u_y + uu_x = 0. \tag{19.7.1}$$

The equation is known as the inviscid Burger's equation (or Euler equation of hydrodynamics). We make several remarks.

(1) The equation models flow of mass with concentration u, where speed of flow depends on concentration.

(2) Here y is the time parameter.
(3) Solutions develop a singularity called a shock wave.
(4) One appication of the equation is to the study of traffic flow.

We start with simple case of a *linear* equation similar to (19.7.1):

$$u_y + cu_x = 0$$

with initial condition

$$u(x,0) = h(x).$$

Thus the initial curve can be parametrized by

$$(x,y,u) = (s,0,h(s)). \tag{19.7.2}$$

Writing the equation in dot product form we obtain

$$\left\langle \begin{pmatrix} c \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0. \tag{19.7.3}$$

The characteristic equations are

$$\begin{cases} x_t(t,s) = c \\ y_t(t,s) = 1 \\ u_t(t,s) = 0. \end{cases} \tag{19.7.4}$$

Solving this along the initial curve (19.7.2) we obtain

$$x = s + ct, \quad y = t, \quad u = h(s).$$

Eliminating s and t gives the explicit solution

$$u(x,y) = h(x - cy).$$

The solution implies that the initial profile does not change. It merely moves with speed c along the x-axis.

Remark 19.7.1. Here a wave of fixed shape is moving to the right with speed c.

19.8. Blow up of solutions of Burgers equation

Let us turn to the *quasilinear* Burgers (a.k.a. Euler) equation (19.7.1):

$$u_y + uu_x = 0$$

with initial condition $u(x,0) = h(x)$. This equation is solved similarly to the linear example treated in Section 19.7. We view the equation as a dot product

$$\left\langle \begin{pmatrix} u \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} u_x \\ u_y \\ -1 \end{pmatrix} \right\rangle = 0.$$

The characteristic equations are therefore

$$\begin{cases} x_t(t,s) = u \\ y_t(t,s) = 1 \\ u_t(t,s) = 0. \end{cases} \tag{19.8.1}$$

The initial condition for u is

$$u(x,0) = h(x)$$

and therefore the last equation from (19.8.1) implies $u(t,s) = h(s)$. Substituting this expression for u into the first equation, we obtain

$$x_t = h(s)$$

and the initial condition for x yields $x = s + h(s)t$. Also $y = t$ (as in the linear case). Thus $s = x - h(s)y$ and eliminating the variables s and t from u we obtain

$$u = h(x - uy). \tag{19.8.2}$$

We make 2 remarks in connection with formula (19.8.2):

(1) one can still think of u as the speed of the wave (as in the linear case) but now the speed u is variable;
(2) The u is given only implicitly (more precisely it is a functional equation for u);

Eliminating t we obtain the following equation for the characteristic (in the xy plane):

$$x = s + h(s)y.$$

The third characteristic equation $u_t = 0$ implies that for each fixed s, i.e., along a characteristic curve, u preserves its initial value $u = h(s)$. Equations

$$(x, y, u) = (s + h(s)t, t, h(s))$$

indicate that the characteristic curves are straight lines, and in particular that the characteristics are straight lines. The initial data $h(s)$ determine the speed of the characteristic emanating from a given point on $\Gamma(s)$.

Remark 19.8.1. If $s_1 < s_2$ and the characteristic leaving s_1 has higher speed than the characteristic leaving s_2, then they will intersect, causing the solution to blow up.

This can be seen algebraically as follows. We differentiate (19.8.2) with respect to x by chain rule, to obtain

$$u_x = h' \cdot (1 - yu_x).$$

Recall that $y > 0$ is the time parameter. Thus we have $u_x = h' - h'yu_x$ and therefore $u_x(1 + h'y) = h'$. We conclude that the solution blows up at critical time $y = y_c$ with

$$y_c = -\frac{1}{h'(s)}. \tag{19.8.3}$$

A necessary condition for the creation of a singularity is that one has $h'(s) < 0$ at least at one point, for otherwise an earlier characteristic will never overcome a later one.

19.9. Extending solution past blow-up

We will write the Burgers equation in integral form. We first rewrite it as follows:

$$\frac{\partial}{\partial y} u + \frac{1}{2} \frac{\partial}{\partial x} \left(u^2 \right) = 0. \tag{19.9.1}$$

Next we integrate with respect to x for a fixed y over an arbitrary but fixed interval $[a, b]$, obtaining

$$\frac{\partial}{\partial y} \int_a^b u(\xi, y) d\xi \ + \ \frac{1}{2} \left[u^2(b, y) - u^2(a, y) \right] = 0. \tag{19.9.2}$$

This type of equation is referred to in the literature as an "integral balance". Every solution of the differential equation (in particular, necessarily a C^1 solution) is also a solution of the integral balance (19.9.2), but the integral balance is well defined for functions not in C^1, called "weak solutions".

Consider a weak solution u that is C^1 except for discontinuities along a specific curve $x = \gamma(y)$. We split the interval of integration into two:

$$[a, b] = [a, \gamma(y)] \cup [\gamma(y), b]$$

and write the integral balance in the form

$$\frac{\partial}{\partial y} \left(\int_a^{\gamma(y)} u(\xi, y) d\xi + \int_{\gamma(y)}^b u(\xi, y) d\xi \right) \ + \ \frac{1}{2} \left[u^2(b, y) - u^2(a, y) \right] = 0.$$

Next, consider the expression

$$\frac{\partial}{\partial y} \left(\int_a^{\gamma(y)} u(\xi, y) d\xi \right).$$

We consider the primitive

$$F(x, y) = \int_a^x u(\xi, y) d\xi.$$

Then

$$\int_a^{\gamma(y)} u(\xi, y) d\xi = F(\gamma(y), y).$$

Therefore

$$\frac{\partial}{\partial y}\left(\int_a^{\gamma(y)} u(\xi, y) d\xi\right) = \frac{\partial}{\partial y} F(\gamma(y), y) = \frac{d\gamma}{dy} F_x(\gamma(y), y) + F_y(\gamma(y), y).$$

Differentiating under the integral sign, we obtain

$$\frac{\partial}{\partial y}\left(\int_a^{\gamma(y)} u(\xi, y) d\xi\right) = \frac{d\gamma}{dy}\int_a^{\gamma(y)} (u(\xi, y))_\xi \, d\xi + \int_a^{\gamma(y)} (u(\xi, y))_y \, d\xi$$

By the fundamental theorem of calculus applied to the first integral on the right-hand side, we have

$$\frac{\partial}{\partial y}\left(\int_a^{\gamma(y)} u(\xi, y) \, d\xi\right) = \frac{d\gamma}{dy}\,(u_-) + \int_a^{\gamma(y)} (u(\xi, y))_y \, d\xi$$

where u_- denotes the value of u when we approach the curve γ from the left. Next, we exploit the partial differential equation (19.9.1) to obtain

$$\frac{\partial}{\partial y}\left(\int_a^{\gamma(y)} u(\xi, y) \, d\xi\right) = \frac{d\gamma}{dy}\,(u_-) - \int_a^{\gamma(y)} \frac{1}{2}\,\left(u^2(\xi, y)\right)_\xi \, d\xi$$

Thus we obtain

$$\gamma_y(y) u_- - \frac{1}{2}\left(\int_a^{\gamma(y)} \left(u^2(\xi, y)\right)_\xi \, d\xi\right).$$

Applying a similar maneuver to $\int_{\gamma(b)}^b$, we obtain

$$\gamma_y(y)u_- - \gamma_y(y)u_+ - \frac{1}{2}\left(\int_a^{\gamma(y)} (u^2(\xi, y))_\xi + \int_{\gamma(y)}^b (u^2(\xi, y))_\xi\right) + \frac{1}{2}\left(u^2(b, y) - u^2(a, y)\right) = 0.$$

Performing the integration again and canceling the boundary terms of u^2 at a and b gives

$$\gamma_y(y)(u_- - u_+) = \frac{1}{2}\left(u_-^2 - u_+^2\right) = \frac{1}{2}\left(u_- + u_+\right)\left(u_- - u_+\right)$$

and therefore

$$\gamma_y(y) = \frac{1}{2}(u_- + u_+). \tag{19.9.3}$$

Thus the curve γ moves at a speed that is the average of the speeds on its left and the right sides.

19.10. Example

By formula (19.9.3), we have

$$\frac{d\gamma}{dy} = \frac{1}{2}(u_- + u_+). \tag{19.10.1}$$

Thus the discontinuity moves with speed $\frac{1}{2}(u_- + u_+)$.

REMARK 19.10.1. By assumption, the discontinuity of u at time y in the variable x is at location $x = \gamma(y)$. In other words, the speed of propagation of the discontinuity is precisely $\frac{d\gamma}{dy}$.

Consider Burgers' equation with initial condition

$$u(x,0) = h(x) = \begin{cases} 1 & \text{if } x \le 0 \\ 1 - \frac{x}{\alpha} & \text{if } 0 < x < \alpha \\ 0 & \text{if } x \ge \alpha. \end{cases} \tag{19.10.2}$$

Note that h is not monotone increasing. By formula (19.8.3) we have $y_c = -\frac{1}{h'(s)}$. Hence the solution develops a singularity at time $y_c = \alpha$. When $y < \alpha$ (before the shock) the solution is given by

$$u(x,y) = \begin{cases} 1 & \text{if } x \le y \\ \frac{x-\alpha}{y-\alpha} & \text{if } y < x < \alpha \\ 0 & \text{if } x \ge \alpha. \end{cases} \tag{19.10.3}$$

After the critical time y_c we can only talk about weak solutions. We seek a weak solution with a single discontinuity. Equation (19.10.1) gives the speed of propagation of the discontinuity.

Therefore the weak solution compatible with the integral balance even for time $y > \alpha$ is the solution

$$u(x,y) = \begin{cases} 1 & \text{if } x < \alpha + \frac{1}{2}(y-\alpha) \\ 0 & \text{if } x > \alpha + \frac{1}{2}(y-\alpha) \end{cases}$$

The solution has the structure of a moving jump discontinuity, called a *shock wave*.

The problem is the non-uniqueness of the weak solutions.

19.11. More general PDE

Consider the PDE

$$u_y + \frac{\partial}{\partial x}F(u) = 0. \tag{19.11.1}$$

Equations of this type are called *conservation laws*. To understand the name, relate to the quantity F as the *flux*. Equation (19.11.1) is supplemented with initial condition

$$u(x,0) = \begin{cases} u_- & \text{if } x < 0 \\ u_+ & \text{if } x > 0. \end{cases}$$

We assume that the solution takes the form of a shock wave

$$u(x,y) = \begin{cases} u_- & \text{if } x < \gamma(y) \\ u_+ & \text{if } x > \gamma(y). \end{cases} \tag{19.11.2}$$

Then the problem is to determine the *shock orbit* $x = \gamma(y)$. We find γ by integrating (19.11.1) with respect to x between the bounds x_1 and x_2 where $x_1 < \gamma$ and $x_2 > \gamma$. Using (19.11.2) we obtain

$$\frac{\partial}{\partial y}\big((x_2 - \gamma(y))u_+ + (\gamma(y) - x_1)u_-\big) = F(u_+) - F(u_-). \tag{19.11.3}$$

It follows from (19.11.3) that

$$\frac{d\gamma}{dy} = \frac{F(u_+) - F(u_-)}{u_+ - u_-}. \tag{19.11.4}$$

We denote the jump across the shock by square brackets, e.g., $[u] = u_+ - u_-$. Then (19.11.4) becomes

$$\frac{d\gamma}{dy} = \frac{[F]}{[u]}. \tag{19.11.5}$$

We make several related remarks.

(1) In many areas of continuum mechanics, the jump equation (19.11.5) is called the Rankine–Hugoniot condition.

(2) From our analysis of Example we expect that shock would only occur if characteristics collide. In the case of general conservation laws, this condition is expressed as **The entropy condition:** Characteristics must enter the shock curve, and are not allowed to emanate from it.

(3) The motivation for the entropy condition is rooted in gas dynamics and the second law of hydrodynamics: amount of information present in a closed system only decreases with time.

Part 3

History of infinitesimal math: Stevin to Skolem

In this part of the book we provide a historical perspective. The history of analysis, differential geometry, and related fields is often viewed as a process inevitably leading to rigorous Weierstrassian foundations stripped of infinitesimals. From such a perspective a return to infinitesimals becomes difficult to motivate. The perspective provided in this part is more favorable toward infinitesimals. Such a perspective can provide additional motivation for approaching differential geometry using true infinitesimals as we have done in this book.

CHAPTER 20

Sixteenth century

20.1. Simon Stevin

This section contains a discussion of Simon Stevin, unending decimals, and the real numbers. The material in this section is based in part on the article [**Błaszczyk, Katz & Sherry 2013**] published in *Foundations of Science*.

Simon Stevin developed an adequate system for representing ordinary numbers, including all the ones that were used in his time, whether rational or not. Moreover his scheme for representing numbers by unending decimals works well for all of them, as is well known.

Stevin developed specific notation for decimals (more complicated than the one we use today) and did actual technical work with them rather than merely envisioning their possibility, unlike some of his predecessors like E. Bonfils in 1350. Bonfils wrote that "the unit is divided into ten parts which are called Primes, and each Prime is divided into ten parts which are called Seconds, and so on into infinity" [**Gandz 1936**, p. 39] but his ideas remained in the realm of potentiality, as he did not develop any notation to ground them.

Even earlier, the Greeks developed techniques for solving problems that today we may solve using more advanced number systems. But to Euclid and Eudoxus, only $2, 3, 4, \ldots$ were numbers: everything else was proportion. The idea of attributing algebraic techniques in disguise to the Greeks is known as *Geometric Algebra* and is considered a controversial thesis. Our paper in no way depends on this thesis.

Stevin dealt with unending decimals in his book *l'Arithmetique* rather than the more practically-oriented *De Thiende* meant to teach students to work with decimals (of course, finite ones).

20.1.1. Stevin and the intermediate value theorem. As far as using the term *real* to describe the numbers Stevin was concerned with, the first author to describe the common numbers as *real* may have been Descartes. Representing common numbers (including both rational and not rational) by unending decimals was to Stevin not merely a matter of speculation, but the background of, for example, his work on proving the intermediate value theorem for polynomials (he worked with the example of a specific cubic) using subdivision into ten subintervals of equal length.

Stevin's accomplishment seems all the more remarkable if one recalls that it dates from before Vieta, meaning that Stevin had no notation beyond the tool inherited from the Greeks namely that of proportions $a : b :: c : d$. He indeed proceeds to write down a cubic equation as a proportion, which can be puzzling to an unprepared modern reader. The idea of an *equation* that we take for granted was in the process of emerging at the time. Stevin presented a divide-and-conquer algorithm for finding the root, which is essentially the one reproduced by Cauchy 250 years later in *Cours d'Analyse*.

In this sense, Stevin deserves the credit for developing a representation for the real numbers to a considerable extent, as indeed one way of introducing the real number field $\mathbb{R}$ is via unending decimals. He was obviously unaware of the existence of what we call today the transcendental numbers but then again Cantor and Dedekind were obviously unaware of modern developments in real analysis.

Cantor, as well as Méray and Heine, sought to characterize the real numbers axiomatically by means of Cauchy Completeness (CC). This property however is insufficient to characterize the real numbers; one needs to require the Archimedean property in addition to CC. Can we then claim that they (i.e., Cantor, Heine, and Méray) really knew what the real numbers are? Apparently, not any more than Stevin, if a sufficient axiom system is a prerequisite for *knowing the real numbers*.

Dedekind was convinced he had a proof of the existence of an infinite set;[1] see [**Ferreirós 2007**, p. 111 and section 5.2, p. 244]. Thus, Joyce comments on Dedekind's concept of things being objects of our thought and concludes:

> That's an innocent concept, but in paragraph 66 it's used to justify the astounding theorem that infinite sets exist. [**Joyce 2005**]

Do such aspects of the work of Cantor and Dedekind invalidate their constructions of the real number system? Surely not. Similarly, Stevin's

[1]The proof exploits the assumption that there exists a set S of all things, and that a mathematical thing is an object of our thought. Then if s is such a thing, then the thought, denoted s', that "s can be an object of my thought" is a mathematical object is a thing distinct from s. Denoting the passage from s to s' by ϕ, Dedekind gets a self-map ϕ of S which is some kind of blend of the successor function and the brace-forming operation. From this Dedekind derives that S is infinite, QED.

proposed construction should not be judged by the yardstick of awareness of future mathematical developments.

20.1.2. Modern decimals. In the approach to the real numbers via decimals, one needs to identify each terminating decimal with the corresponding string with an infinite tail of 9s, as in $1.0 = 0.999\ldots$ The more common approaches to $\mathbb{R}$ are (1) via Dedekind cuts, or (2) via equivalence classes of Cauchy sequences, an approach sometimes attributed to Georg Cantor, even though the concept of an equivalence relation did not exist yet at the time. The publication of [**Cantor 1872**] was preceded by [**Heine 1872**] by a few months but Heine explicitly attributes the idea of *Fundamentalrheine* to Cantor.

Even earlier, Charles Méray published his "Remarques sur la nature des quantités définies par la condition de servir de limites à des variables données" [**Méray 1869**]; see [**Dugac 1970**] for a detailed analysis. However, Méray's paper seems to have been unknown among German mathematicians.

While Stevin had no idea of the set-theoretic underpinnings of the received *ontology* of modern mathematics, *procedurally* speaking his approach to arithmetic was close to the modern one, meaning that he envisioned a certain homogeneity among all numbers with no preferential status for the rationals; see [**Malet 2006**], [**Katz & Katz 2012b**], [**Błaszczyk, Katz & Sherry 2013**] for further details.

Stevin's decimals cannot be placed on equal footing with the 1872 constructions, when both representations and algebraic operations were developed as well as the continuity axioms, while Stevin only gave the representation.

In 1923, A. Hoborski, a mathematician involved, like Stevin, in applied rather than pure mathematics, developed an arithmetic of real numbers based on unending decimal representations [**Hoborski 1923**].

CHAPTER 21

Seventeenth century

Among the many 17th century pioneers of infinitesimal geometry, we will focus on Fermat, Gregory, and Leibniz.

21.1. Pierre de Fermat

Pierre de Fermat (1601/1607–1665) developed a procedure known as *adequality* for finding maxima and minima of algebraic expressions, tangents to curves, etc. The name of the procedure derives from the $\pi\alpha\rho\iota\sigma\acute{o}\tau\eta\varsigma$ of Diophantus. Some of its applications amount to variational techniques exploiting a small variation E. Fermat's treatment of geometric and physical applications suggests that an aspect of approximation is inherent in adequality, as well as an aspect of smallness on the part of E. Fermat relied on Bachet's reading of Diophantus, who coined the term $\pi\alpha\rho\iota\sigma\acute{o}\tau\eta\varsigma$ for mathematical purposes and used it to refer to the way in which 1321/711 is approximately equal to 11/6. In translating Diophantus, Bachet performed a semantic calque, passing from *parisoō* to *adaequo*, which is the source for Fermat's term rendered in English as *adequality*.

21.1.1. Summary of the algorithm. To give a summary of Fermat's algorithm for finding the maximum or minimum value of an algebraic expression in a variable A, we will write such an expression in modern functional notation as $f(A)$. One version of the algorithm can be broken up into five steps in the following way:

(1) Introduce an auxiliary symbol E, and form $f(A+E)$;
(2) Set *adequal* the two expressions $f(A+E) \sqcap f(A)$ (the notation " $\sqcap$ " for adequality is ours, not Fermat's);
(3) Cancel the common terms on the two sides of the adequality. The remaining terms all contain a factor of E;
(4) Divide by E (in a parenthetical comment, Fermat adds: "or by the highest common factor of E");
(5) Among the remaining terms, suppress all terms which still contain a factor of E. Solving the resulting equation for A yields the desired extremum of f.

In simplified modern form, the algorithm entails expanding the difference quotient $\frac{f(A+E)-f(A)}{E}$ in powers of E and taking the constant term.

There are two crucial points in trying to understand Fermat's reasoning: first, the meaning of "adequality" in step (2); and second, the justification for suppressing the terms involving positive powers of E in step (5).

21.1.2. The parabola. As an example consider Fermat's determination of the tangent line to the parabola. To simplify Fermat's notation, we will work with the parabola $y = x^2$ thought of as the level curve

$$\frac{x^2}{y} = 1$$

of the two-variable function $\frac{x^2}{y}$. Given a point (x, y) on the parabola, Fermat seeks the tangent line through the point, exploiting the geometric fact that by convexity, a point (p, q) on the tangent line lies *outside* the parabola. He therefore obtains an inequality equivalent in our notation to $\frac{p^2}{q} > 1$, or $p^2 > q$. Here $q = y - E$, and E is Fermat's magic symbol we wish to understand. Thus, we obtain

$$\frac{p^2}{y-E} > 1. \tag{21.1.1}$$

At this point Fermat proceeds as follows:

(i) he writes down the inequality $\frac{p^2}{y-E} > 1$, or $p^2 > y - E$;
(ii) he invites the reader to *adégaler* (to "adequate");
(iii) he writes down the adequality $\frac{x^2}{p^2} \sqcap \frac{y}{y-E}$;
(iv) he uses an identity involving similar triangles to substitute $\frac{x}{p} = \frac{y+r}{y+r-E}$ where r is the distance from the vertex of the parabola to the point of intersection of the tangent to the parabola at y with the axis of symmetry,
(v) he cross multiplies and cancels identical terms on right and left, then divides out by E, *discards* the remaining terms containing E, and obtains $y = r$ as the solution.

What interests us are steps (i) and (ii). How does Fermat pass from an inequality to an adequality? Giusti observes: "Comme d'habitude, Fermat est autant détaillé dans les exemples qu'il est réticent dans les explications. On ne trouvera donc presque jamais des justifications de sa règle des tangentes." [**Giusti 2009**, p. 80] In fact, Fermat provides no explicit explanation for this step. However, what he does is to apply the defining relation for a curve to points on the tangent line to the curve. Note that here the quantity E, as in $q = y - E$, is positive: Fermat did not have the facility we do of assigning negative values to variables.

Fermat says nothing about considering points $y + E$ "on the other side", i.e., further away from the vertex of the parabola, as he does in the context of applying a related but different method, for instance in his two letters to Mersenne (see [**Strømholm 1968**, p. 51]), and in his letter to Brûlart [**Fermat 1643**]. Now for positive values of E, Fermat's inequality (21.1.1) would be satisfied by a *transverse ray* (i.e., secant ray) starting at (x, y) and lying outside the parabola, just as much as it is satisfied by a tangent ray starting at (x, y). Fermat's method therefore presupposes an additional piece of information, privileging the tangent ray over transverse rays. The additional piece of information is geometric in origin: he applies the defining relation (of the curve itself) to a point on the tangent ray to the curve. Such a procedure is only meaningful when the increment E is small.

In modern terms, we would speak of the tangent line being a "best approximation" to the curve for a small variation E; however, Fermat does not explicitly discuss the size of E.

The procedure of "discarding the remaining terms" in step (v) admits of a proxy in the hyperreal context in terms of the standard part principle (every finite hyperreal number is infinitely close to a real number). Fermat does not elaborate on the justification of this step, but he is always careful to speak of the *suppressing* or *deleting* the remaining term in E, rather than setting it equal to zero. Perhaps his rationale for suppressing terms in E consists in ignoring terms that don't correspond to a possible measurement, prefiguring Leibniz's *inassignable quantities*. Fermat's inferential moves in the context of his adequality are akin to Leibniz's in the context of his calculus.

While Fermat never spoke of his E as being *infinitely small*, the technique based on what eventually came to be known as infinitesimals was known to Fermat's contemporaries like Galileo (see [**Bascelli 2014a**], [**Bascelli 2014b**]) and Wallis (see [**Katz & Katz 2012a**, Section 13]). The technique was familiar to Fermat, as his correspondence with Wallis makes clear; see [**Katz, Schaps & Shnider 2013**, Section 2.1].

Fermat was very interested in Galileo's treatise *De motu locali*, as we know from his letters to Marin Mersenne dated apr/may 1637, 10 august, and 22 october 1638. Galileo's treatment of infinitesimals in *De motu locali* is discussed in [**Settle 1966**] and [**Wisan 1974**, p. 292].

The clerics in Rome forbade the doctrine of *indivisibles* on 10 august 1632 (a month before Galileo was summonded to stand trial over heliocentrism); this may help explain why the catholic Fermat may have been reluctant to speak of them explicitly.

The problem of the parabola could of course be solved purely in the context of polynomials using the idea of a double root, but for transcendental curves like the cycloid Fermat does *not* study the order of multiplicity of the zero of an auxiliary polynomial. Rather, Fermat explicitly stated that he applied the defining property of the curve to points on the tangent line: "Il faut donc adégaler (à cause de la propriété spécifique de la courbe qui est à considérer sur la tangente)" (see [**Katz, Schaps & Shnider 2013**] for more details).

Fermat's approach involves applying the defining relation of the curve, to a point on a *tangent* line to the curve where the relation is *not* satisfied exactly. Fermat's approach is therefore consistent with the idea of approximation. His method involves a negligible distance (whether infinitesimal or not) between the tangent and the original curve when one is near the point of tangency. This line of reasoning is related to the ideas of the differential calculus. Fermat correctly solves the cycloid problem by obtaining the defining equation of the tangent line.

21.1.3. The cycloid curve. Fermat treated numerous problems concerning maxima and minima as well as tangents to curves using a procedure called *adequality*. For a transcendental curve like the cycloid (see e.g., [**Cifoletti 1990**, p. 70]), he solved the problem of finding the tangent line at an arbitrary point of the curve as follows. Fermat starts with the defining equation of the cycloid, considers the tangent line at a point P of the curve, chooses a nearby point Q *on the tangent* at distance E from P, and substitutes Q into the defining equation of the cycloid (Fermat refers to such a defining equation as *la propriété spécifique de la courbe*), *as if* it satisfied the latter (hence not *equality* but *adequality* or approximate equality).

Since the cycloid is a transcendental curve, it would be difficult to interpret Fermat's solution as a purely algebraic procedure involving a formal variable E.

In more detail, note that the cycloid is generated by marking a point on a circle and tracing the path of the point as the circle rolls along a horizontal straight line. If the marked point is the initial point of contact of the circle with the line, and the circle rolls to the right, then the ordinate[1] of the marked point is given by the difference of the

[1] In this case the *ordinate* refers to the horizontal coordinate.

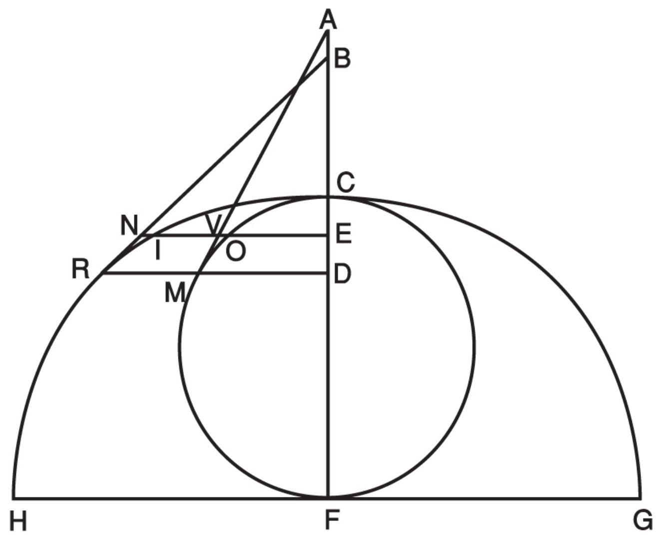

FIGURE 21.1.1. Fermat's cycloid

length of arc traversed (the distance the center of the circle has moved) and the distance of the point from the vertical line through the center of the circle.[2]

21.1.4. Fermat's description. Fermat's description of the cycloid is based on a diagram [**Tannery & Henry 1891**, Figure 103, p. 163] reproduced in Figure 21.1.1. Let R be a point on the cycloid and D the point of intersection of the horizontal line ℓ through R with the axis of symmetry of the cycloid generated by one full revolution of the circle. If M is the point of intersection of ℓ with the generating circle when centered on the axis of symmetry, and C is the apex of that circle then in the words of Fermat:

> La propriété spécifique de la courbe est que la droite RD est égale à la somme de l'arc de cercle CM et de l'ordonnée DM.[3] ([**Fermat circ. 1637**, p. 144])

Let r be the tangent line to the cycloid at R, and m the tangent line to the circle at M. To determine the defining relation of line r, Fermat considers the horizontal line $NIVOE$ passing through a point $N \in r$.

[2]Assuming the circle to have radius 1, the equation of the cycloid as described above is $x = \theta - \sin\theta, \quad y = 1 - \cos\theta$.

[3]To compare Fermat's description with the parametric description given in the previous footnote, we note that length of the segment RD is $\pi - x = \pi - \theta + \sin\theta$, while $\pi - \theta$ is the length of the arc CM, and the length DM equals $\sin\theta$.

Here I is the first point of intersection with the cycloid, while V is the point of intersection with m, and O is the point of intersection with the generating circle, and E is the point of intersection with the axis of symmetry.

The defining relation for r is derived from the defining relation for the cycloid using adequality. The defining relation for the point I on the cycloid is

$$IE = OE + \operatorname{arc} CO.$$

By *adequality*, Fermat first replaces I by the point $N \in r$:

$$NE \sqcap OE + \operatorname{arc} CO = OE + \operatorname{arc} CM - \operatorname{arc} MO. \qquad (21.1.2)$$

Then Fermat replaces O by the point $V \in m$, and the arc MO, by the length of the segment $MV \subset m$. This produces the linear relation

$$NE \sqcap VE + \operatorname{arc} CM - MV,$$

yielding the equation of the tangent line r to the cycloid at R as a graph over the axis of symmetry. The distance NE is expressed in terms of the distance VE, where $V \in m$, and the distance MV along that tangent line. Thinking in terms of slope relative to the variable distance DE (which corresponds to the parameter e in the example of the parabola), Fermat's equation says that the slope of r relative to DE is the slope of m minus the proportionality factor of MV relative to DE.[4] To summarize, Fermat exploited two adequations in his calculation:

(1) the length of a segment along m adequals the length of a segment of a circular arc, and
(2) the distance from the axis of symmetry to a point on r (or m) adequals the distance from the axis to a corresponding point on the cycloid (or circle).

As Fermat explains,

> Il faut donc *adégaler* (à cause de la proprété spécifique de la courbe qui est à considérer sur la tangente) cette droite $\frac{za-ze}{a}$ [i.e., NE] à la somme $OE+\operatorname{arc} CO \ldots$ [et] d'après la remarque précédente, substituer, à OE, l'ordonnée EV de la tangente, et à l'arc MO, la portion de la tangente MV qui lui est adjacente (Fermat [**Fermat circ. 1637**, p. 144]; [**Tannery & Henry 1891**, p. 228]).

[4] The slope of the tangent line relative to the axis of symmetry, or equivalently, relative to the y axis, given by elementary calculus is $\frac{d(\pi-x)}{d\theta}/\frac{dy}{d\theta} = \frac{-1}{\sin\theta} + \frac{\cos\theta}{\sin\theta}$. The length MV equals $e/(\sin\theta)$ and the slope of the tangent line to the circle relative to the y axis is $\frac{\cos\theta}{\sin\theta}$, in agreement with Fermat's equation.

The procedure can be summarized in modern terms by the following principle: *The tangent line to the curve is defined by using adequality to linearize the defining relation of the curve, or* "adégaler *(à cause de la propréte spécifique de la courbe qui est à considérer sur la tangente)."*

Fermat uses the same argument in his calculation of the tangents to other transcendental curves whose defining property is similar to the cycloids and involves arc length along a generating curve. For a discussion of some of these examples, see Giusti [**Giusti 2009**] and Itard [**Itard 1949**].

21.2. James Gregory

In his attempt to prove the irrationality of π, James Gregory (1638–1675) broadened the scope of mathematical *procedures* available at the time by introducing what he called a sixth operation (on top of the existing four arithmetic operations as well as extraction of roots). He referred to the new procedure as the *termination* of a (convergent) sequence: "And so by imagining this [sequence] to be continued to infinity, we can imagine the ultimate convergent terms *to be equal*; and we call those equal ultimate terms the termination of the [sequence]." [**Gregory 1667**, p. 18–19] Referring to sequences of inscribed and circumscribed polygons, he emphasized that

> if the abovementioned series of polygons can be terminated, that is, if that ultimate inscribed polygon is found to be equal (so to speak) to that ultimate circumscribed polygon, it would undoubtedly provide the quadrature of a circle as well as a hyperbola. But since it is difficult, and in geometry perhaps unheard-of, for such a series to come to an end [lit.: be terminated], we have to start by showing some Propositions by means of which it is possible to find the terminations of a certain number of series of this type, and finally (if it can be done) a general method of finding terminations of all convergent series.

Note that in a modern infinitesimal framework like [**Robinson 1966**], sequences possess terms with infinite indices. Gregory's relation can be formalized in terms of the standard part principle in Robinson's framework. This principle asserts that every finite hyperreal number is infinitely close to a unique real number.

If each term with an infinite index n is indistinguishable (in the sense of being infinitely close) from some real number, then we "terminate the series" (to exploit Gregory's terminology) with this number, meaning that this number is the limit of the sequence. Gregory considered the lengths of inscribed (I_n) and circumscribed (C_n) polygons, and obtained recursive relations $I_{n+1}^2 = C_n I_n$ and $C_{n+1} = \frac{2C_n I_{n+1}}{C_n + I_{n+1}}$; see [**Lützen 2014**, p. 225].

Gregory's definition of coincidence of lengths of the I_n and the C_n corresponds to a relation of infinite proximity in a hyperreal framework. Namely we have $I_n \approx C_n$ where $\approx$ is the relation of being infinitely close (i.e., the difference is infinitesimal), and the common standard part of these values is what is known today as the *limit* of the sequence.

Our proposed formalisation does not mean that Gregory is a pre-Robinsonian, but rather indicates that Robinson's framework is more helpful in understanding Gregory's procedures than a Weierstrassian framework. For additional details see [**Bascelli et al. 2017**].

21.3. Gottfried Wilhelm von Leibniz

Gottfried Wilhelm Leibniz (1646–1716) was a co-founder of infinitesimal calculus. When we trace the diverse paths through mathematical history that have led from the infinitesimal calculus of the 17th century to its version implemented in Abraham Robinson's framework in the twentieth, we notice patterns often neglected in received historiography focusing on the success of Weierstrassian foundations.

We have argued that the final version of Leibniz's infinitesimal calculus was free of logical fallacies, owing to its *procedural* implementation in ZFC via Robinson's framework.

21.3.1. Berkeley on shakier ground. Both Berkeley as a philosopher of mathematics, and the strength of his criticisms of Leibniz's infinitesimals have been overestimated by many historians of mathematics. Such criticisms stand on shakier ground than the underestimated mathematical and philosophical resources available to Leibniz for defending his theory. Leibniz's theoretical strategy for dealing with infinitesimals includes the following aspects:

(1) Leibniz clearly realized that infinitesimals violate the so-called Archimedean property[5] which Leibniz refers to as Euclid V.5;[6] in a letter to L'Hospital he considers infinitesimals as non-Archimedean quantities, in reference to Euclid's theory of proportions [**De Risi 2016**, p. 64, note 15].

[5]In modern notation this can be expressed as $(\forall x, y > 0)(\exists n \in \mathbb{N})[nx > y]$.

[6]In modern editions of *The Elements* this appears as Definition V.4.

$$\left\{\begin{matrix} \textit{assignable} \\ \textit{quantities} \end{matrix}\right\} \overset{\text{LC}}{\rightsquigarrow} \left\{\begin{matrix} \textit{assignable and inassignable} \\ \textit{quantities} \end{matrix}\right\} \overset{\text{TLH}}{\rightsquigarrow} \left\{\begin{matrix} \textit{assignable} \\ \textit{quantities} \end{matrix}\right\}$$

FIGURE 21.3.1. Leibniz's law of continuity (LC) takes one from assignable to inassignable quantities, while his transcendental law of homogeneity (TLH) returns one to assignable quantities.

(2) Leibniz introduced a distinction between assignable and inassignable numbers. Ordinary numbers are assignable while infinitesimals are inassignable. This distinction enabled Leibniz to ground the procedures of the calculus relying on differentials on the *transcendental law of homogeneity* (TLH), asserting roughly that higher order terms can be discarded in a calculation since they are negligible (in the sense that an infinitesimal is negligible compared to an ordinary quantity like 1).
(3) Leibniz exploited a generalized relation of *equality up to*. This was more general than the relation of strict equality and enabled a formalisation of the TLH (see previous item).
(4) Leibniz described infinitesimals as *useful fictions* akin to imaginary numbers. Leibniz's position was at variance with many of his contemporaries and allies who tended to take a more realist stance. We interpret Leibnizian infinitesimals as *pure fictions* at variance with a post-Russellian *logical fiction* reading involving a concealed quantifier ranging over ordinary values; see [**Bascelli et al. 2016**].
(5) Leibniz formulated a law of continuity (LC) governing the transition from the realm of assignable quantities to a broader one encompassing infinite and infinitesimal quantities: "il se trouve que les règles du fini réussissent dans l'infini ... et que vice versa les règles de l'infini réussissent dans le fini." [**Leibniz 1702**]
(6) Meanhile, the TLH returns to the realm of assignable quantities.

The relation between the two realms can be represented by the diagram of Figure 21.3.1.

Leibniz is explicit about the fact that his *incomparables* violate Euclid V.5 (when compared to other quantities) in his letter to l'Hospital from the same year: "J'appelle *grandeurs incomparables* dont l'une multipliée par quelque nombre fini que ce soit, ne sçauroit exceder l'autre, de la même facon qu'Euclide la pris dans sa cinquieme definition du cinquieme livre."[7] [**Leibniz 1695a**, p. 288]

Leading Leibniz scholar Jesseph in [**Jesseph 2015**] largely endorses Bos' interpretation of Leibnizian infinitesimals as fictions.

Modern proxies for Leibniz's procedures expressed by LC and TLH are, respectively, the *transfer principle* and the *standard part principle* in Robinson's framework. Leibniz's theoretical strategy for dealing with infinitesimals and infinite numbers was explored in the articles [**Katz & Sherry 2012**], [**Katz & Sherry 2013**], [**Sherry & Katz 2014**], and [**Bascelli et al. 2016**].

21.3.2. Ellipse with infinite focus. Leibniz gives several examples of the application of his Law of Continuity, including the following three examples.

(1) In the context of a discussion of parallel lines, he writes:

> when the straight line BP ultimately becomes parallel to the straight line VA, even then it converges toward it or makes an angle with it, only that the angle is then infinitely small [**Child 1920**, p. 148].

(2) Invoking the idea that the term equality may refer to equality up to an infinitesimal error, Leibniz writes:

> when one straight line[8] is equal to another, it is said to be unequal to it, but that the difference is infinitely small [**Child 1920**, p. 148].

(3) A conception of a parabola expressed by means of an ellipse with an infinitely removed focal point is evoked in the following terms:

> a parabola is the ultimate form of an ellipse, in which the second focus is at an infinite distance from the given focus nearest to the given vertex [**Child 1920**, p. 148].

Example (2) can be interpreted as follows. Leibniz denotes a finite positive quantity by

$$(d)x$$

([**Bos 1974**, p. 57] replaced this by $\underline{d}x$). The assignable quantity $(d)x$ as it varies passes via infinitesimal dx on its way to absolute 0. Then the infinitesimal dx is the Leibnizian *status transitus*. Zero is merely

[7] This can be translated as follows: "I use the term *incomparable magnitudes* to refer to [magnitudes] of which one multiplied by any finite number whatsoever, will be unable to exceed the other, in the same way [adopted by] Euclid in the fifth definition of the fifth book [of the *Elements*]."

[8] Here Leibniz is using the term *line* in its generic meaning of a *segment*.

the assignable *shadow* of the infinitesimal. Then a *line* (i.e., segment) of length $2x+dx$ will be equal to one of length $2x$, up to an infinitesimal. This particular *status transitus* is the foundation rock of the Leibnizian definition of the differential quotient.

Example (1) of parallel lines can be elaborated as follows. Let us follow Leibniz in building the line through $(0,1)$ parallel to the x-axis in the plane. Line L_H with y-intercept 1 and x-intercept H is given by $y = 1 - \frac{x}{H}$. For infinite H, the line L_H has negative infinitesimal slope, meets the x-axis at an infinite point, and forms an infinitesimal angle with the x-axis at the point where they meet. We will denote by $\mathrm{st}(x)$ the assignable (i.e., real) shadow of a finite x. Then every finite point $(x,y) \in L_H$ satisfies

$$\begin{aligned}\mathrm{st}(x,y) &= (\mathrm{st}(x), \mathrm{st}(y))\\ &= \left(\mathrm{st}(x), \mathrm{st}\left(1 - \tfrac{x}{H}\right)\right)\\ &= (\mathrm{st}(x), 1).\end{aligned}$$

Hence the finite portion of L_H is infinitely close to the line $y = 1$ parallel to the x-axis, which is its *shadow*. Thus, the parallel line is constructed by varying the oblique line depending on a parameter. Such variation passes via the *status transitus* defined by an infinite value of H.

To implement example (3), let us follow Leibniz in deforming an ellipse, via a *status transitus*, into a parabola. The ellipse with vertex (apex) at $(0,-1)$ and with foci at the origin and at $(0;H)$ is given by

$$\sqrt{x^2+y^2} + \sqrt{x^2+(y-H)^2} = H+2 \tag{21.3.1}$$

We square (21.3.1) to obtain

$$x^2+y^2+x^2+(H-y)^2+2\sqrt{(x^2+y^2)(x^2+(H-y)^2)} = H^2+4H+4 \tag{21.3.2}$$

We move the radical to one side

$$2\sqrt{(x^2+y^2)(x^2+(H-y)^2)} = H^2+4H+4-\left(x^2+y^2+x^2+(H-y)^2\right) \tag{21.3.3}$$

and square again. After cancellation we see that (21.3.1) is equivalent to

$$\left(y+2+\tfrac{2}{H}\right)^2 - (x^2+y^2)\left(1+\tfrac{4}{H}+\tfrac{4}{H^2}\right) = 0. \tag{21.3.4}$$

The calculation (21.3.1) through (21.3.4) depends on habits of *general reasoning* such as:

- squaring undoes a radical;
- the binomial formula;
- terms in an equation can be transfered to the other side; etc.

General reasoning of this type is familiar in the realm of ordinary assignable (finite) numbers, but why does it remain valid when applied to the, fictional, "realm" of inassignable (infinite or infinitesimal) numbers? The validity of transfering such *general reasoning* originally *instituted* in the finite realm, to the "realm" of the infinite is precisely the content of Leibniz's law of continuity.[9]

We therefore apply Leibniz's law of continuity to equation (21.3.4) for an infinite H. The resulting entity is still an ellipse of sorts, to the extent that it satisfies all of the equations (21.3.1) through (21.3.4). However, this entity is no longer finite. It represents a Leibnizian *status transitus* between ellipse and parabola. This *status transitus* has foci at the origin and at an infinitely distant point $(0, H)$. Assuming x and y are finite, we set $x_0 = \mathrm{st}(x)$ and $y_0 = \mathrm{st}(y)$, to obtain an equation for a real shadow of this entity:

$$\begin{aligned}
\mathrm{st}&\left(\left(y+2+\tfrac{2}{H}\right)^2 - (x^2+y^2)\left(1+\tfrac{4}{H}+\tfrac{4}{H^2}\right)\right) = \\
&= \left(y_0+2+\mathrm{st}\left(\tfrac{2}{H}\right)\right)^2 - \left(x_0^2+y_0^2\right)\left(1+\mathrm{st}\left(\tfrac{4}{H}+\tfrac{4}{H^2}\right)\right) \\
&= (y_0+2)^2 - \left(x_0^2+y_0^2\right) \\
&= 0.
\end{aligned}$$

Simplifying, we obtain

$$y_0 = \frac{x_0^2}{4} - 1. \tag{21.3.5}$$

Thus, the finite portion of the *status transitus* (21.3.4) is infinitely close to its *shadow* (21.3.5), namely the real parabola $y = \frac{x^2}{4} - 1$. This is the kind of payoff Leibniz is seeking with his law of continuity.

[9]When the *general reasoning* being transfered to the infinite "realm" is generalized to encompass arbitrary elementary properties (i.e., the first order properties), one obtains the Łoś–Robinson transfer principle, proved in Chapter 16.

CHAPTER 22

Eighteenth century

22.1. Leonhard Euler

Leonhard Euler (1707–1783) routinely relied on procedures exploiting infinite numbers in his work, as in applying the binomial formula to an expression raised to an infinite power so as to obtain the development of the exponential function into power series.

Euler's comments on infinity indicate an affinity with Leibnizian fictionalist views: "Even if someone denies that infinite numbers really exist in this world, still in mathematical speculations there arise questions to which answers cannot be given unless we admit an infinite number." [**Euler 2000**, § 82].

Euler's dual notion of *arithmetic* and *geometric* equality which indicate that, like Leibniz, he was working with generalized notions of equality. Thus, Euler wrote:

> Since the infinitely small is actually nothing, it is clear that a finite quantity can neither be increased nor decreased by adding or subtracting an infinitely small quantity. Let a be a finite quantity and let dx be infinitely small. Then $a + dx$ and $a - dx$, or, more generally, $a \pm ndx$, are equal to a. Whether we consider the relation between $a \pm ndx$ and a as arithmetic or as geometric, in both cases the ratio turns out to be that between equals. The arithmetic ratio of equals is clear: Since $ndx = 0$, we have $a \pm ndx - a = 0$. On the other hand, the geometric ratio is clearly of equals, since $\frac{a \pm ndx}{a} = 1$. From this we obtain the well-known rule that the infinitely small vanishes in comparison with the finite and hence can be neglected. For this reason the objection brought up against the analysis of the infinite, that it lacks geometric rigor, falls to the ground under its own weight, since nothing is neglected except that which is actually nothing. Hence with perfect justice we can affirm that in this sublime science we keep the same perfect geometric rigor that

> is found in the books of the ancients. [**Euler 2000**, § 87]

Like Leibniz, Euler did not distinguish notationwise between different modes of comparison, but we could perhaps introduce two separate symbols for the two relations, such as $\approx$ for the arithmetic comparison and the Leibnizian symbol $\sqcap$ for the geometric comparison. See [**Bair et al. 2016**] for further details.

CHAPTER 23

Nineteenth century

23.1. Augustin-Louis Cauchy

A. L. Cauchy (1789–1857) was a transitional figure. His significance stems from the fact that he championed greater rigor in mathematics. Historians enamored of set-theoretic foundations tend to translate the term *rigor* as *epsilon-delta*, and sometimes even attribute an epsilon-delta definition of continuity to Cauchy.

In reality, to Cauchy rigor stood for the traditional ideal of *geometric* rigor, meaning the rigor of Euclid's geometry as it was admired throughout the centuries. What lies in the background is Cauchy's opposition to certain summation techniques of infinite series as practiced by Euler and Lagrange without necessarily paying attention to convergence. To Cauchy rigor entailed a rejection of these techniques that he referred to as the *generality of algebra*.

In his textbooks, Cauchy insists on reconciling rigor with infinitesimals. By this he means not the elimination of infinitesimals but rather the reliance thereon, as in his definition of continuity. As late as 1853, Cauchy still defined continuity as follows in a research article:

> ... une fonction u de la variable réelle x sera *continue*, entre deux limites données de x, si, cette fonction admettant pour chaque valeur intermédiaire de x une valeur unique et finie, un accroissement infiniment petit attribué à la variable produit toujours, entre les limites dont il s'agit, un accroissement infiniment petit de la fonction elle-même. [**Cauchy 1853**] [emphasis in the original]

Already in 1821, Cauchy denoted his infinitesimal α and required $f(x+\alpha) - f(x)$ to be infinitesimal as the definition of the continuity of f. In differential geometry, Cauchy routinely defined the center of curvature of a plane curve by intersecting a pair of *infinitely close* normals to the curve. These issues are explored further in [**Cutland et al. 1988**], [**Katz & Katz 2011**], [**Borovik & Katz 2012**], [**Katz & Tall 2013**], [**Bascelli et al. 2014**], and [**Błaszczyk et al. 2016b**].

23.2. Bernhard Riemann

Bernhard Riemann (1826–1866) was a pioneer of modern differential geometry. Riemann's famous 1854 *Habilitation* lecture "On the hypotheses lying at the foundations of geometry" was followed by a more explicit but lesser known sequel dating from 1861 and entitled "Commentatio mathematica". Unlike the 1854 lecture, Riemann's 1861 sequel contains enough formulas to enable researchers to attempt to reconstruct Riemann's train of thought and his remarkable anticipation of later work of Levi-Civita and other differential geometers. The 1861 text is briefly discussed in [**Jost 2016**, p. 62, note 19]. A reader interested in a more detailed analysis can consult [**Darrigol 2015**].

Riemann's relative lack of familiarity with the field in 1854 turned out to be an advantage in that he was able to develop a perspective unaffected by Kantian notions of a priori space that influenced many geometers at the time. Riemann was influenced by the philosopher Johann Friedrich Herbart (see below).

As Jost notes, Riemann's 1854 lecture "penetrates as deeply as never before into a field that had occupied and challenged the greatest thinkers of mankind since classical antiquity, and it even hints at the greatest discovery of the physics of the following century" namely Einstein's relativity theory.

As Jost notes, a number of famous scientists, including psychologist Wilhelm Wundt and the philosopher Bertrand Russell, entered the stage with errors of judgment on both the topic and content of the 1854 lecture after it had been posthumously published by Riemann's friend Dedekind. As Jost notes, "subsequent generations of mathematicians worked out the ideas outlined in the brief lecture and confirmed their full validity and soundness and extraordinary range and potential."

Herbart's influence on Riemann is mentioned in [**Jost 2016**, note 52, p. 27]. Jost notes that this was analyzed in [**Erdmann 1877**, pp. 29–33], [**Boi 1995**, pp. 129–136], [**Scholz 1982**].

The latter source downplays Herbart's influence on Riemann's lecture, while historian David E. Rowe notes that in some of his writing, Riemann identifies himself as a Herbartian; see [**Rowe 2017**].

One could also mention the article [**Nowak 1989**]. Nowak argues that "Riemann was aware of the philosophical implications of his mathematics and structured the Habilitationsvortrag as a philosophical argument which used mathematics to demonstrate the untenability of Kant's position that Euclidean geometry constituted a set of synthetic a priori truths about physical space."

Nowak gives considerable attention to Herbart's influence on Riemann, and notes that "Herbart's discussion of space inspired Riemann to create a more fruitful combination of higher-dimensional geometry and Gauss's differential geometry than he might otherwise have been able to."

Nowak goes on to makes the following three points.

1. Herbart's constructive approach to space, already cited, mirrored the content of Riemann's reference to Gauss in that both discussed construction of spaces rather than construction in space.

2. Riemann followed Herbart in rejecting Kant's view of space as an a priori category of thought, instead seeing space as a concept which possessed properties and was capable of change and variation. Riemann copied some passages from Herbart on this subject, and the Fragmente philosophischen Inhalts included in his published works contain a passage in which Riemann cites Herbart as demonstrating the falsity of Kant's view.

3. Riemann took from Herbart the view that the construction of spatial objects were possible in intuition and independent of our perceptions in physical space. Riemann extended this idea to allow for the possibility that these spaces would not obey the axioms of Euclidean geometry. We know from Riemann's notes on Herbart that he read Herbart's Psychologie als Wissenschaft...

As mentioned above, Russell failed to appreciate the significance of Riemann's pioneering writings in differential geometry. Russell's *Essay on the Foundations of Geometry* (1897) is cited by Jost on page 127 but not analyzed.

Here Russell wrote the following in his section 65:

"[Riemann's] philosophy is chiefly vitiated, to my mind, by this fallacy, and by the uncritical assumption that a metrical coordinate systern can be set up independently of any axioms as to space-measurement. Riemann has failed to observe, what I have endeavoured to prove in the next chapter, that, unless space had a strictly constant measure of curvature, Geometry would become impossible; also that the absence of constant measure of curvature involves absolute position, which is an absurdity."

What Russell is claiming here, in a nutshell, is the impossibility of doing what has come to be called Riemannian geometry, such impossibility apparently derived from first philosophical principles. For similar non-sequiturs from Russell's pen in the matter of infinitesimal analysis see [**Katz & Sherry 2013**].

The only substantial formula in Riemann's lecture (addressed mainly to non-mathematicians) is that of the length element in a space of constant curvature. This is Riemann's formula $\frac{1}{1+\frac{\alpha}{4}\sum x^2}\sqrt{\sum dx^2}$ appearing in [**Jost 2016**, p. 37]. Today of course we would incorporate a summation index i as part of the notation, as in $\frac{1}{1+\frac{\alpha}{4}\sum_i (x^i)^2}\sqrt{\sum_i (dx^i)^2}$.

After a technical introduction to (modern) Riemannian geometry and tensors, occupying pages 63-113, Jost finally gets to Riemann's formula on page 114. However, what Jost doesn't mention explicitly enough is that the symbol dx had a different meaning to Riemann than the one given to it by Jost. Riemann viewed the length element as a combination of infinitesimal displacements dx^i of the coordinates x^i weighed by suitable functions. Meanwhile, in the modern formalism, dx^i is a covector dual to the vector $\frac{\partial}{\partial x^i}$. No entity like $\frac{\partial}{\partial x^i}$ appeared in Riemann's writing. Riemann needed no tensors to perform integration over curves in the manifold.

23.3. Otto Stolz

Otto Stolz (1842–1905) introduced the term Archimedean property.

23.4. Hermann Cohen

Hermann Cohen (1842–1918) sought a working logic of infinitesimals and operated with a distinction between extensive and intensive quantity, parallel to Leibnizian distinction of assignable vs inassignable.

CHAPTER 24

Twentieth century

24.1. Skolem

Nonstandard integers were first constructed by T. Skolem in the 1930s (see [**Skolem 1933**], [**Skolem 1934**]; an English version may be found in [**Skolem 1955**]). Skolem's accomplishment is generally regarded as a major milestone in the development of 20th century logic.

D. Scott (see [**Scott 1959**, p. 245]) compares Skolem's predicative approach with the ultrapower approach (Skolem's nonstandard integers are also discussed in [**Bell & Slomson 1969**] and [**Stillwell 1977**, pp. 148–150]). Scott notes that Skolem used the ring *DF* of algebraically (first-order) definable functions from integers to integers. The quotient DF/P of *DF* by a minimal prime ideal P produces Skolem's non-standard integers. The ideal P corresponds to a prime ideal in the Boolean algebra of idempotents. Note that the idempotents of *DF* are the characteristic functions of (first-order) definable sets of integers. Such sets give rise to a denumerable Boolean algebra $\mathfrak{P}$ and therefore can be given an *ordered basis*. Such a basis for $\mathfrak{P}$ is a nested sequence[1]

$$X_n \supset X_{n+1} \supset \dots$$

such that $Y \in \mathfrak{P}$ if and only if $Y \supset X_n$ for a suitable n. Choose a sequence (s_n) such that

$$s_n \in X_n \setminus X_{n+1}.$$

Then functions $f, g \in DF$ are in the same equivalence class if and only if

$$(\exists N)(\forall n \geq N)\, f(s_n) = g(s_n).$$

The sequence (s_n) is the *comparing function* used by Skolem to partition the definable functions into congruence classes. Note that, even though Skolem places himself in a context limited to definable functions, a key role in the theory is played by the comparing function which is *not* definable.

[1]We reversed the inclusions as given in [**Scott 1959**, p. 245] so as to insist on the analogy with a filter.

Including these sequences in $\mathbb{R}^{\mathbb{N}}$ yields an embedding of the Skolem nonstandard integers in the hyperintegers (via the ultrapower construction).

Bibliography

[Albeverio et al. 2009] Albeverio, S.; Fenstad, J.; Høgh-Krohn, R.; Lindstrøm, T. *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press 1986, Dover 2009.

[Almeida, Neves & Stroyan 2014] Almeida, R.; Neves, V.; Stroyan, K. "Infinitesimal differential geometry: cusps and envelopes." *Differ. Geom. Dyn. Syst.* **16** (2014), 1–13.

[Bair et al. 2013] Bair, J.; Błaszczyk, P.; Ely, R.; Henry, V.; Kanovei, V.; Katz, K.; Katz, M.; Kutateladze, S.; McGaffey, T.; Schaps, D.; Sherry, D.; Shnider, S. "Is mathematical history written by the victors?" *Notices of the American Mathematical Society* **60** (2013) no. 7, 886-904.
See `http://www.ams.org/notices/201307/rnoti-p886.pdf`
and `http://arxiv.org/abs/1306.5973`

[Bair et al. 2016] Bair, J.; Błaszczyk, P.; Ely, R.; Henry, V.; Kanovei, V.; Katz, K.; Katz, M.; Kutateladze, S.; McGaffey, T.; Reeder, P.; Schaps, D.; Sherry, D.; Shnider, S. "Interpreting the infinitesimal mathematics of Leibniz and Euler." *Journal for General Philosophy of Science*, online first.
See `http://dx.doi.org/10.1007/s10838-016-9334-z`
and `http://arxiv.org/abs/1605.00455`

[Bangert et al. 2007] Bangert, V; Croke, C.; Ivanov, S.; Katz, M. Boundary case of equality in optimal Loewner-type inequalities. *Trans. Amer. Math. Soc.* **359** (2007), no. 1, 1–17. See `https://arxiv.org/abs/math/0406008`

[Bascelli 2014a] Bascelli, T. "Galileo's quanti: understanding infinitesimal magnitudes." *Archive for History of Exact Sciences* **68**, no. 2, 121–136.

[Bascelli 2014b] Bascelli, T. "Infinitesimal issues in Galileo's theory of motion." *Revue Roumaine de Philosophie* **58**, no. 1, 23-41.

[Bascelli et al. 2014] Bascelli, T., Bottazzi, E., Herzberg, F., Kanovei, V., Katz, K., Katz, M., Nowik, T., Sherry, D., Shnider, S. "Fermat, Leibniz, Euler, and the gang: The true history of the concepts of limit and shadow." *Notices of the American Mathematical Society* **61**, no. 8, 848-864.

[Bascelli et al. 2016] Bascelli, T.; Błaszczyk, P.; Kanovei, V.; Katz, K.; Katz, M.; Schaps, D.; Sherry, D. "Leibniz vs Ishiguro: Closing a quarter-century of syncategoremania." *HOPOS: Journal of the Internatonal Society for the History of Philosophy of Science* **6**, no. 1, 117–147.
See `http://dx.doi.org/10.1086/685645`
and `http://arxiv.org/abs/1603.07209`

[Bascelli et al. 2017] Bascelli, T.; Błaszczyk, P.; Kanovei, V.; Katz, K.; Katz, M.; Kutateladze, S.; Nowik, T.; Schaps, D.; Sherry. D. "Gregory's sixth operation." *Foundations of Science*, to appear.
See `http://dx.doi.org/10.1007/s10699-016-9512-9`

[Bell & Slomson 1969] Bell, J.; Slomson, A. *Models and ultraproducts: An introduction.* North-Holland Publishing, Amsterdam-London.

[Benci–Luperi Baglini 2017] Benci, V.; Luperi Baglini, L. Generalized solutions in PDEs and the Burgers' equation. *J. Differential Equations* **263** (2017), no. 10, 6916–6952.

[Benoit 1997] Benoit, E. "Applications of Nonstandard Analysis in Ordinary Differential Equations." *Nonstandard Analysis, Theory and Applications.* Editors: Leif O. Arkeryd, Nigel J. Cutland, C. Ward Henson, NATO ASI Series Volume 493, pp. 153–182.

[Błaszczyk, Katz & Sherry 2013] Błaszczyk, P.; Katz, M.; Sherry, D. "Ten misconceptions from the history of analysis and their debunking." *Foundations of Science*, **18** (2013), no. 1, 43–74.
See http://dx.doi.org/10.1007/s10699-012-9285-8
and http://arxiv.org/abs/1202.4153

[Błaszczyk et al. 2016b] Błaszczyk, P.; Kanovei, V.; Katz, M.; Sherry, D. "Controversies in the foundations of analysis: Comments on Schubring's Conflicts." *Foundations of Science*, online first.
See `http://dx.doi.org/10.1007/s10699-015-9473-4`
and `http://arxiv.org/abs/1601.00059`

[Boi 1995] Boi, Luciano. *Le problème mathématique de l'espace. Une quête de l'intelligible.* With a preface by René Thom. Springer-Verlag, Berlin, 1995.

[Boothby 1986] Boothby, W. *An introduction to differentiable manifolds and Riemannian geometry.* Second edition. Pure and Applied Mathematics, 120. Academic Press, Orlando, FL, 1986.

[Borovik & Katz 2012] Borovik, A., Katz, M. "Who gave you the Cauchy–Weierstrass tale? The dual history of rigorous calculus." *Foundations of Science* **17**, no. 3, 245–276.
See `http://dx.doi.org/10.1007/s10699-011-9235-x`

[Bos 1974] Bos, H. J. M. "Differentials, higher-order differentials and the derivative in the Leibnizian calculus." *Arch. History Exact Sci.* **14**, 1–90.

[Boyer 1949] Boyer, C. *The concepts of the calculus.* Hafner Publishing Company.

[Bradley & Sandifer 2009] Bradley, R., Sandifer, C. *Cauchy's Cours d'analyse. An annotated translation.* Sources and Studies in the History of Mathematics and Physical Sciences. Springer, New York.

[Cajori 1928] Cajori. *A history of mathematical notations, Volumes 1-2.* The Open Court Publishing Company, La Salle, Illinois, 1928/1929. Reprinted by Dover, 1993.

[Cantor 1872] Cantor, G. "Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen." *Mathematische Annalen* **5**, 123–132.

[Cauchy 1821] Cauchy, A. L. *Cours d'Analyse de L'Ecole Royale Polytechnique. Première Partie. Analyse algébrique* (Paris: Imprimérie Royale, 1821).

[Cauchy 1853] Cauchy, A. L. "Note sur les séries convergentes dont les divers termes sont des fonctions continues d'une variable réelle ou imaginaire, entre des limites données." In *Oeuvres complètes*, Series 1, Vol. 12, pp. 30–36. Paris: Gauthier–Villars, 1900.

[Chang & Keisler 1990] Chang, C.; Keisler, H.J. *Model Theory.* Studies in Logic and the Foundations of Mathematics (3rd ed.), Elsevier (1990) [Originally published in 1973].

[Child 1920] Child, J. M. (ed.) *The early mathematical manuscripts of Leibniz.* Translated from the Latin texts published by Carl Immanuel Gerhardt with critical and historical notes by J. M. Child. Chicago-London: The Open Court Publishing.

[Cifoletti 1990] Cifoletti, G. *La méthode de Fermat: son statut et sa diffusion. Algèbre et comparaison de figures dans l'histoire de la méthode de Fermat.* Cahiers d'Histoire et de Philosophie des Sciences. Nouvelle Série **33**. Société Française d'Histoire des Sciences et des Techniques, Paris.

[Crossley 1987] Crossley, J. *The emergence of number.* Second edition. World Scientific Publishing, Singapore.

[Cutland et al. 1988] Cutland, N.; Kessler, C.; Kopp, E.; Ross, D. "On Cauchy's notion of infinitesimal." *British J. Philos. Sci.* **39**, no. 3, 375–378.

[Darrigol 2015] Darrigol, O. "The mystery of Riemann's curvature." *Historia Math.* **42** (2015), no. 1, 47–83.

[Davis 1977] Davis, M. *Applied nonstandard analysis.* Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1977. Reprinted by Dover, NY, 2005.
see `http://store.doverpublications.com/0486442292.html`

[De Risi 2016] De Risi, V. *Leibniz on the Parallel Postulate and the Foundations of Geometry. The Unpublished Manuscripts.* Springer.

[Dugac 1970] Dugac, P. "Charles Méray (1935–1911) et la notion de limite." *Revue d'Histoire des Sciences et de leurs Applications* **23** (1970), no. 4, 333–350.

[Ehrlich 2005] Ehrlich, P. "Continuity." In Encyclopedia of philosophy, 2nd edition (Donald M. Borchert, editor), Macmillan Reference USA, Farmington Hills, MI, 2005 (The online version contains some slight improvements), pp. 489–517.

[Erdmann 1877] Erdmann, Benno. *Die Axiome der Geometrie. Eine philosophische Untersuchung der Riemann-Helmholtzschen Raumtheorie.* Leipzig, Leopold Voss, 1877.

[Euclid 2007] Euclid. Euclid's Elements of Geometry, edited, and provided with a modern English translation, by Richard Fitzpatrick, 2007.
See `http://farside.ph.utexas.edu/euclid.html`

[Euler 1755] Euler, L. *Institutiones Calculi Differentialis.* Saint Petersburg.

[Euler 2000] Euler, L. *Foundations of Differential Calculus.* English translation of Chapters 1–9 of [**Euler 1755**] by J. Blanton, Springer, N.Y.

[Euler 1770] Euler, L. *Elements of Algebra (3rd English edition).* John Hewlett and Francis Horner, English translators. Orme Longman, 1822.

[Fearney-Sander 1979] Fearnley-Sander, D. "Hermann Grassmann and the creation of linear algebra." *Amer. Math. Monthly* **86**, no. 10, 809–817.

[Feferman & Levy 1963] Feferman, S.; Levy, A. "Independence results in set theory by Cohen's method II." *Notices Amer. Math. Soc.* **10** (1963).

[Fermat circ. 1637] Fermat, P. "Méthode pour la recherche du maximum et du minimum." In Tannery's edition [**Tannery & Henry 1896**], p. 121–156.

[Fermat 1643] Fermat, P. Letter to Brûlart. 31 march 1643. Oeuvres, Vol. 5, pp. 120–125.

[Ferreirós 2007] Ferreirós, J. *Labyrinth of thought. A history of set theory and its role in modern mathematics.* Second edition. Birkhäuser Verlag, Basel.

[Fowler 1992] Fowler, D. "Dedekind's theorem: $\sqrt{2} \times \sqrt{3} = \sqrt{6}$." *American Mathematical Monthly* **99**, no. 8, 725–733.

[Gandz 1936] Gandz, S. "The invention of the decimal fractions and the application of the exponential calculus by Immanuel Bonfils of Tarascon (c. 1350)." *Isis* **25**, 16–45.

[Gerhardt 1850-63] Gerhardt, C. I. (ed.) *Leibnizens mathematische Schriften.* Berlin and Halle: Eidmann.

[Giusti 2009] Giusti, E. "Les méthodes des maxima et minima de Fermat." *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6* **18**, Fascicule Special, 59–85.

[Goldblatt 1998] Goldblatt, R. *Lectures on the hyperreals. An introduction to nonstandard analysis.* Graduate Texts in Mathematics, 188. Springer-Verlag, New York.

[Gordon et al. 2002] Gordon, E.; Kusraev, A.; Kutateladze, S. *Infinitesimal analysis.* Updated and revised translation of the 2001 Russian original. Translated by Kutateladze. *Mathematics and its Applications*, **544**. Kluwer Academic Publishers, Dordrecht, 2002.

[Gregory 1667] Gregory, J. *Vera Circuli et Hyperbolae Quadratura.* Padua edition, 1667. Patavia edition, 1668.

[Grobman 1959] Grobman, D. "Homeomorphisms of systems of differential equations." *Doklady Akademii Nauk SSSR* **128**, 880–881.

[Hacking 2014] Hacking, I. *Why is there philosophy of mathematics at all?* Cambridge University Press.

[Hartman 1960] Hartman, P. "A lemma in the theory of structural stability of differential equations." *Proc. A.M.S.* **11** (4), 610–620.

[Heine 1872] Heine, E. "Elemente der Functionenlehre." *Journal fur die reine und angewandte Mathematik* **74**, 172–188.

[Henson & Keisler 1986] Henson, C. W.; Keisler, H. J. "On the strength of nonstandard analysis." *J. Symbolic Logic* **51**, no. 2, 377–386.

[Herzberg 2013] Herzberg, F. *Stochastic Calculus with Infinitesimals.* Lecture Notes in Mathematics Volume 2067, Springer, 2013

[Herzberg et al. 2017] Herzberg, F.; Kanovei, V.; Katz, M.; Lyubetsky, V. "Minimal axiomatic frameworks for definable hyperreals with transfer." In preparation.

[Hewitt 1948] Hewitt, E. "Rings of real-valued continuous functions. I." *Trans. Amer. Math. Soc.* **64** (1948), 45–99.

[Hoborski 1923] Hoborski, A. "Nowa teoria liczb niewymiernych." [New theory of irrational numbers] Kraków 1923. See also his "Aus der theoretischen Arithmetik." *Opuscucla Mathematica* **2**, 11-12, Kraków, 1938.
See `https://zbmath.org/?q=an:64.1000.03`

[Itard 1949] Itard, J. "Fermat, précurseur du calcul différentiel." *Archives Internationales d'Histoire des Sciences* **27**, no. 4, 589–610.

[Jech 1973] Jech, T. *The axiom of choice.* Studies in Logic and the Foundations of Mathematics, Vol. 75. North-Holland Publishing, Amsterdam-London; Amercan Elsevier Publishing, New York.

[Jesseph 2015] Jesseph, D. "Leibniz on the Elimination of infinitesimals." In *G.W. Leibniz, Interrelations between Mathematics and Philosophy*, pp. 189–205.

Norma B. Goethe, Philip Beeley, and David Rabouin, eds. Archimedes Series **41** Springer Verlag.

[Jost 2016] Jost, J. (Ed.) *Riemann, Bernhard. On the hypotheses which lie at the bases of geometry.* Edited and with commentary by Jürgen Jost. Expanded English translation of the German original. Classic Texts in the Sciences. Birkhäuser/Springer, Cham, 2016.

[Joyce 2005] Joyce, D. "Notes on Richard Dedekind's *Was sind und was sollen die Zahlen?*" Preprint.
See `http://aleph0.clarku.edu/~djoyce/numbers/dedekind.pdf`

[Kanovei et al. 2016] Kanovei, V.; Katz, K.; Katz, M.; Nowik, T. "Small oscillations of the pendulum, Euler's method, and adequality." Quantum Studies: Mathematics and Foundations 3 (2016), no. 3, 231-236. See `http://dx.doi.org/10.1007/s40509-016-0074-x` and `https://arxiv.org/abs/1604.06663`

[Kanovei & Reeken 2004] Kanovei, V.; Reeken, M. *Nonstandard analysis, axiomatically.* Springer Monographs in Mathematics, Berlin: Springer.

[Kanovei & Shelah 2004] Kanovei, V.; Shelah, S. A definable nonstandard model of the reals." *J. Symbolic Logic* **69** (2004), no. 1, 159–164.

[Katz & Katz 2011] Katz, K.; Katz, M. "Cauchy's continuum." *Perspectives on Science* **19**, no. 4, 426–452.
See `http://dx.doi.org/10.1162/POSC_a_00047`
and `http://arxiv.org/abs/1108.4201`

[Katz & Katz 2012a] Katz, K.; Katz, M. "A Burgessian critique of nominalistic tendencies in contemporary mathematics and its historiography." *Foundations of Science* **17**, no. 1, 51–89.
See `http://dx.doi.org/10.1007/s10699-011-9223-1`
and `http://arxiv.org/abs/1104.0375`

[Katz & Katz 2012b] Katz, K.; Katz, M. "Stevin numbers and reality." *Foundations of Science* **17**, no. 2, 109–123.
See `http://dx.doi.org/10.1007/s10699-011-9228-9`
and `http://arxiv.org/abs/1107.3688`

[Katz–Leichtnam 2013] Katz, M.; Leichtnam, E. "Commuting and noncommuting infinitesimals." *American Mathematical Monthly* **120** (2013), no. 7, 631–641. See `http://dx.doi.org/10.4169/amer.math.monthly.120.07.631` and `https://arxiv.org/abs/1304.0583`

[Katz & Polev 2017] Katz, M.; Polev, L. "From Pythagoreans and Weierstrassians to true infinitesimal calculus." *Journal of Humanistic Mathematics* **7** (2017), no. 1, 87–104. See `https://arxiv.org/abs/1701.05187`

[Katz, Schaps & Shnider 2013] Katz, M.; Schaps, D.; Shnider, S. "Almost Equal: The Method of Adequality from Diophantus to Fermat and Beyond." *Perspectives on Science* **21**, no. 3, 283–324.
See `http://www.mitpressjournals.org/doi/abs/10.1162/POSC_a_00101`
and `http://arxiv.org/abs/1210.7750`

[Katz & Sherry 2012] Katz, M.; Sherry, D. "Leibniz's laws of continuity and homogeneity." *Notices of the American Mathematical Society* **59** (2012), no. 11, 1550–1558.
See `http://www.ams.org/notices/201211/rtx121101550p.pdf`
and `http://arxiv.org/abs/1211.7188`

[Katz & Sherry 2013] Katz, M.; Sherry, D. "Leibniz's infinitesimals: Their fictionality, their modern implementations, and their foes from Berkeley to Russell and beyond." *Erkenntnis* **78**, no. 3, 571–625.
See `http://dx.doi.org/10.1007/s10670-012-9370-y`
and `http://arxiv.org/abs/1205.0174`

[Katz & Tall 2013] Katz, M.; Tall, D. "A Cauchy–Dirac delta function." *Foundations of Science* **18**, no. 1, 107–123.
See `http://dx.doi.org/10.1007/s10699-012-9289-4`
and `http://arxiv.org/abs/1206.0119`

[Katz, V. 1993] Katz, V. J. "Using the history of calculus to teach calculus." *Science & Education. Contributions from History, Philosophy and Sociology of Science and Education* **2**, no. 3, 243–249.

[Keisler 1986] Keisler, H. J. *Elementary Calculus: An Infinitesimal Approach. Second Edition.* Prindle, Weber & Schimidt, Boston, 1986. See online version at `http://www.math.wisc.edu/~keisler/calc.html`

[Keisler 1976] H. J. Keisler. *Foundations of Infinitesimal Calculus.* On-line Edition, 2007, at the author's web page:
`http://www.math.wisc.edu/~keisler/foundations.html`
(Originally published by Prindle, Weber & Schmidt 1976.)

[Klein 1908] Klein, F. *Elementary Mathematics from an Advanced Standpoint. Vol. I. Arithmetic, Algebra, Analysis.* Translation by E. R. Hedrick and C. A. Noble [Macmillan, New York, 1932] from the third German edition [Springer, Berlin, 1924]. Originally published as *Elementarmathematik vom höheren Standpunkte aus* (Leipzig, 1908).

[Leibniz 1695a] Leibniz, G. To l'Hospital, 14/24 june 1695, in [**Gerhardt 1850-63**], vol. I, pp. 287–289.

[Leibniz 1702] Leibniz, G. To Varignon, 2 febr., 1702, in [**Gerhardt 1850-63**] IV, pp. 91–95.

[Leibniz 1989] Leibniz, G. W. *La naissance du calcul différentiel.* 26 articles des Acta Eruditorum. Translated from the Latin and with an introduction and notes by Marc Parmentier. With a preface by Michel Serres. Mathesis. Librairie Philosophique J. Vrin, Paris, 1989.

[Loeb 1975] Loeb, Peter A. "Conversion from nonstandard to standard measure spaces and applications in probability theory." *Trans. Amer. Math. Soc.* **211**, 113–122.

[Loeb & Wolff 2015] Loeb, P.; Wolff, M. (Eds.) *Nonstandard Analysis for the Working Mathematician.* Second Edition. Springer, Dordrecht.

[Łoś 1955] Łoś, J. "Quelques remarques, théorèmes et problèmes sur les classes définissables d'algèbres." In *Mathematical interpretation of formal systems*, 98–113, North-Holland Publishin, Amsterdam, 1955.

[Lützen 2014] Lützen, J. "17th century arguments for the impossibility of the indefinite and the definite circle quadrature." *Revue d'histoire des mathématiques* **20**, no. 2, 211-251.

[Lutz & Goze 1981] Lutz, R.; Goze, M. *Nonstandard Analysis, a Practical Guide with Applications.* Lecture Notes in Mathematics 881, Springer-Verlag, 1981.

[Malet 2006] Malet, A. "Renaissance notions of number and magnitude." *Historia Mathematica* **33**, no. 1, 63–81.

[Marsh 1742] Marsh, J. *Decimal arithmetic made perfect.* London.

[Méray 1869] Méray, C. "Remarques sur la nature des quantités définies par la condition de servir de limites à des variables données." *Revue des Sociétés savantes, Sciences mathém. phys. et naturelles* 2, IV, 281–289.

[Newton 1671] Newton, I. "A Treatise on the Methods of Series and Fluxions," 1671. In [**Whiteside 1969**] (vol. III), pp. 33–35.

[Nowak 1989] Nowak, Gregory. "Riemann's Habilitationsvortrag and the synthetic a priori status of geometry." The history of modern mathematics, Vol. I (Poughkeepsie, NY, 1989), 17–46, Academic Press, Boston, MA, 1989.

[Nowik & Katz 2015] Nowik, T., Katz, M. "Differential geometry via infinitesimal displacements." *Journal of Logic and Analysis* **7**:5, 1–44. See `http://www.logicandanalysis.com/index.php/jla/article/view/237` and `http://arxiv.org/abs/1405.0984`

[Pinchover–Rubinstein 2005] Pinchover, Y.; Rubinstein, J. *An introduction to partial differential equations.* Cambridge University Press, Cambridge, 2005.

[Robinson 1966] Robinson, A. *Non-standard analysis.* North-Holland Publishing, Amsterdam 1966.

[Rowe 2017] Rowe, D. Review of the 2013 German version of the volume on Riemann edited by Jost. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 2017.

[Rudin 1976] Rudin, W. Principles of mathematical analysis. Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book, New York-Auckland-Düsseldorf, 1976.

[Scholz 1982] Scholz, Erhard. "Herbart's influence on Bernhard Riemann." *Historia Math.* **9** (1982), no. 4, 413–440.

[Scott 1959] Scott, D. "On constructing models for arithmetic." 1961 Infinitistic Methods (Proc. Sympos. Foundations of Math., Warsaw, 1959) pp. 235–255, Pergamon, Oxford; Państwowe Wydawnictwo Naukowe, Warsaw.

[Settle 1966] Settle, T. *Galilean science: essays in the mechanics and dynamics of the* Discorsi. Ph.D. dissertation, Cornell University, 288 pages.

[Sherry & Katz 2014] Sherry, D., Katz, M. "Infinitesimals, imaginaries, ideals, and fictions." *Studia Leibnitiana* **44** (2012), no. 2, 166–192 (the article appeared in 2014 even though the year given by the journal is 2012). See `http://arxiv.org/abs/1304.2137`

[Skolem 1933] Skolem, T. "Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems." *Norsk Mat. Forenings Skr., II. Ser.* No. 1/12, 73–82.

[Skolem 1934] Skolem, T. "Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen." *Fundamenta Mathematicae* **23**, 150–161.

[Skolem 1955] Skolem, T. "Peano's axioms and models of arithmetic." In *Mathematical interpretation of formal systems*, pp. 1–14. North-Holland Publishing, Amsterdam.

[Smith 1920] Smith, D. *Source Book in Mathematics.* Mc Grow-Hill, New York.

[Spivak 1979] Spivak, M. *A comprehensive introduction to differential geometry.* Vol. 4. Second edition. Publish or Perish, Wilmington, Del., 1979.

[Stevin 1585] Stevin, S. (1585) L'Arithmetique. In Albert Girard (Ed.), Les Oeuvres Mathematiques de Simon Stevin (Leyde, 1634), part I, p. 1–101.

[Stevin 1625] Stevin, S. L'Arithmetique. In Albert Girard (Ed.), 1625, part II. Online at
http://www.archive.org/stream/larithmetiqvedes00stev#page/353/mode/1up

[Stevin 1958] Stevin, S. The principal works of Simon Stevin. Vols. IIA, IIB: Mathematics. Edited by D. J. Struik. C. V. Swets & Zeitlinger, Amsterdam 1958. Vol. IIA: v+pp. 1–455 (1 plate). Vol. IIB: 1958 iv+pp. 459–976.

[Stolz 1883] Stolz, O. "Zur Geometrie der Alten, insbesondere über ein Axiom des Archimedes." *Mathematische Annalen* **22** (4), 504–519.

[Strømholm 1968] Strømholm, P. "Fermat's methods of maxima and minima and of tangents. A reconstruction." *Archive for History of Exact Sciences* **5**, no. 1, 47–69.

[Stillwell 1977] Stillwell, J. "Concise survey of mathematical logic." *Australian Mathematical Society. Journal. Series A. Pure Mathematics and Statistics* **24**, no. 2, 139–161.

[Stroyan 1977] Stroyan, K. "Infinitesimal analysis of curves and surfaces." In *Handbook of Mathematical Logic*, J. Barwise, ed. North-Holland Publishing Company, 1977.

[Stroyan 2015] Stroyan, K. *Advanced Calculus using Mathematica: NoteBook Edition.*

[Stroyan & Luxemburg 1976] Stroyan, K.; Luxemburg, W. *Introduction to the Theory of Infinitesimals.* Academic Press, 1976.

[Sun 2000] Sun, Y. "Nonstandard Analysis in Mathematical Economics." In: Nonstandard Analysis for the Working Mathematician, Peter A. Loeb and Manfred Wolff, eds, Mathematics and Its Applications, Volume 510, 2000, pp 261-305.

[Tannery & Henry 1891] Tannery, P., Henry, C. *Oeuvres de Fermat, Vol. 1* Gauthier-Villars.

[Tannery & Henry 1896] Tannery, P., Henry, C. *Oeuvres de Fermat, Vol. 3* Gauthier-Villars.

[Tao 2014] Tao, T. *Hilbert's fifth problem and related topics.* Graduate Studies in Mathematics, 153. American Mathematical Society, Providence, RI, 2014.

[Tao & Vu 2016] Tao, T.; Van Vu, V. "Sum-avoiding sets in groups." *Discrete Analysis* 2016:15, 31 pp, see http://dx.doi.org/10.19086/da.887
and http://arxiv.org/abs/1603.03068

[Tarski 1930] Tarski, A. "Une contribution à la théorie de la mesure." *Fundamenta Mathematicae* **15**, 42–50.

[Van den Berg 1987] Van den Berg, I. *Nonstandard asymptotic analysis.* Lecture Notes in Mathematics 1249, Springer 1987.

[Van der Waerden 1985] Van der Waerden, B. L. *A history of algebra. From al-Khwarizmi to Emmy Noether.* Springer-Verlag, Berlin.

[We55] Weatherburn, C. E.: Differential geometry of three dimensions. Vol. 1. Cambridge University Press, 1955.

[Whiteside 1969] Whiteside, D. T. (Ed.): The mathematical papers of Isaac Newton. Vol. III: 1670–1673. Edited by D. T. Whiteside, with the assistance in publication of M. A. Hoskin and A. Prag, Cambridge University Press, London-New York, 1969.

[Wisan 1974] Wisan, W. "The new science of motion: a study of Galileo's *De motu locali.*" *Archive for History of Exact Sciences* **13**, 103–306.