Nonstandard Analysis

Theory and Applications

NATO ASI Series

Advanced Science Institutes Series

A Series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The Series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A Life Sciences **B Physics**	Plenum Publishing Corporation London and New York
C Mathematical and Physical Sciences **D Behavioural and Social Sciences** **E Applied Sciences**	Kluwer Academic Publishers Dordrecht, Boston and London
F Computer and Systems Sciences **G Ecological Sciences** **H Cell Biology** **I Global Environmental Change**	Springer-Verlag Berlin, Heidelberg, New York, London, Paris and Tokyo

PARTNERSHIP SUB-SERIES

1. Disarmament Technologies	Kluwer Academic Publishers
2. Environment	Springer-Verlag / Kluwer Academic Publishers
3. High Technology	Kluwer Academic Publishers
4. Science and Technology Policy	Kluwer Academic Publishers
5. Computer Networking	Kluwer Academic Publishers

The Partnership Sub-Series incorporates activities undertaken in collaboration with NATO's Cooperation Partners, the countries of the CIS and Central and Eastern Europe, in Priority Areas of concern to those countries.

NATO-PCO-DATA BASE

The electronic index to the NATO ASI Series provides full bibliographical references (with keywords and/or abstracts) to more than 50000 contributions from international scientists published in all sections of the NATO ASI Series.
Access to the NATO-PCO-DATA BASE is possible in two ways:

– via online FILE 128 (NATO-PCO-DATA BASE) hosted by ESRIN, Via Galileo Galilei, I-00044 Frascati, Italy.

– via CD-ROM "NATO-PCO-DATA BASE" with user-friendly retrieval software in English, French and German (© WTV GmbH and DATAWARE Technologies Inc. 1989).

The CD-ROM can be ordered through any member of the Board of Publishers or through NATO-PCO, Overijse, Belgium.

Series C: Mathematical and Physical Sciences – Vol. 493

Nonstandard Analysis

Theory and Applications

edited by

Leif O. Arkeryd
University of Gothenburg, Sweden

Nigel J. Cutland
University of Hull, England

and

C. Ward Henson
University of Illinois,
Urbana-Champaign, U.S.A.

Springer-Science+Business Media, B.V.

Proceedings of the NATO Advanced Study Institute on
Nonstandard Analysis and its Applications
International Centre for Mathematical Sciences, Edinburgh, Scotland
30 June – 13 July 1996

A C.I.P. Catalogue record for this book is available from the Library of Congress

ISBN 978-94-010-6335-7 ISBN 978-94-011-5544-1 (eBook)
DOI 10.1007/978-94-011-5544-1

Printed on acid-free paper

CONTENTS

PREFACE

More than thirty years after its discovery by Abraham Robinson[1], the ideas and techniques of Nonstandard Analysis (NSA) are being applied across the whole mathematical spectrum, as well as constituting an important field of research in their own right. The current methods of NSA now greatly extend Robinson's original work with infinitesimals. However, while the range of applications is broad, certain fundamental themes recur. The nonstandard framework allows many informal ideas (that could loosely be described as idealisation) to be made precise and tractable. For example, the real line can (in this framework) be treated simultaneously as both a continuum and a discrete set of points; and a similar dual approach can be used to link the notions infinite and finite, rough and smooth. This has provided some powerful tools for the research mathematician – for example Loeb measure spaces in stochastic analysis and its applications, and nonstandard hulls in Banach spaces. The achievements of NSA can be summarised under the headings (i) *explanation* - giving fresh insight or new approaches to established theories; (ii) *discovery* - leading to new results in many fields; (iii) *invention* - providing new, rich structures that are useful in modelling and representation, as well as being of interest in their own right.

The aim of the present volume is to make the power and range of applicability of NSA more widely known and available to research mathematicians. The twelve articles originated as lecture notes provided to students at the NATO Advanced Study Institute (ASI) *Nonstandard Analysis and its Applications*, held at the International Centre for Mathematical Sciences (ICMS), Edinburgh in July 1996. At this institute, leaders in the subject offered an in-depth and comprehensive series of lecture courses at an advanced level, laying the foundations of NSA as well as giving an in-

[1]Robinson, A., (1961) Non-standard analysis, *Proc. Roy. Acad. Amsterdam Ser. A*, **64**, pp. 432–40.

Robinson, A., (1966) *Nonstandard Analysis*, North-Holland, Amsterdam. 2nd revised edition, 1974; reissued in paperback by Princeton University Press, 1996.

troduction to the fundamental techniques of the methodology as applied in many areas of pure and applied mathematics.

The ASI formed the first phase of a seven week research programme *NSA96*, hosted by the ICMS during the summer of 1996. Phase two was a four week International Workshop during which leading experts (with expertise in both 'standard' and 'nonstandard' techniques) met for an intensive period of interaction and research on problems of common interest, and *NSA96* concluded with a one week International Research Symposium.

For the present purpose the original lecture notes have been carefully reworked in order that the benefits of the ASI may be extended to the wider community of mathematicians. The aim has been to provide a multi-authored text book that offers a careful introduction to NSA and a view of the current 'state of the art' – both the foundations of the subject *and* its rôle in current mathematical research in many areas.

The first four articles cover the fundamentals of NSA, and are designed to equip the reader for the study of more advanced theory and applications. The '*Gentle Introduction*' by Henson is designed for mathematicians who do not necessarily have any background in formal logic. It provides a foundation for NSA that avoids the features that often inhibit the discovery of the essence of this fascinating and powerful subject. The novelty of Henson's *Foundations* is a natural geometric definition of a nonstandard extension, from which the small amount of logic needed develops easily. By this means he dispenses with the need for heavy logical formalism, excessive set theoretic apparatus, and over dependence on one particular explicit construction of a nonstandard universe – the features that are often obstacles to mathematicians wishing to get to the heart of the subject.

Throughout Henson's '*Gentle Introduction*' the ultraproduct construction is used to give explicit examples of nonstandard extensions. These are often helpful to beginners who benefit from having a definite meaning for such concepts as '*internal set.*' For those who are familiar with the ultraproduct construction or prefer its explicit nature, NSA can be viewed as a systematic way to introduce the power of ultraproducts[2] into *all* areas of mathematics, combined with a language with which to understand and exploit the full extent of their properties. However, Henson's point of view here is that it is better to aim at working within an axiomatic framework, and this is what is placed at the forefront in his article.

Following Henson's *Foundations* are articles introducing the use of NSA in real analysis (Cutland), topology (Loeb) and measure theory (Ross) – which provide the foundation for more advanced applications. The starting point is the extension of the real number field $\mathbb{R}$ to a field $^*\mathbb{R}$ of *hyper-*

[2]Ultraproducts are often used to good effect on an *ad hoc* basis in several parts of mathematics, notably in functional analysis.

real (or *nonstandard*) *numbers*, which contains infinitesimal and infinite 'real numbers'. Our intuitions about the properties of such numbers can be formulated and proved, and this allows classical analysis to be developed rigorously in a natural way with the aid of these new numbers and concepts. Cutland's article gives an exposition of these ideas, all due to Robinson, covering limits, continuity, differentiation and integration.

The natural extension of these ideas to topology, again going back to Robinson, is expounded in Loeb's article. For each topological space X there is a nonstandard extension ${}^*X$ containing new 'ideal' elements, that correspond to infinitesimals in ${}^*\mathbb{R}$. The fundamental idea here is that the notion *infinitely close* can thus be made precise for *any* topological setting – with obvious pleasant consequences for the definition and use of notions such as open and closed sets, limits, and continuity.

Loeb measures have occupied a central position in NSA and its applications for over twenty years, both in measure theory and probability theory and those parts of mathematics where these play a foundational rôle, such as functional analysis, mathematical physics and mathematical finance theory. This is amply illustrated by the later articles in this volume. The article by Ross gives a new exposition of the Loeb measure construction; using it, *any* 'nonstandard' measure space gives rise to a *standard* measure space (a *Loeb space*). Thus we have a rich source of new measure and probability spaces, which turn out to be very powerful and at the same time tractable. Ross provides a sample of the myriad applications of Loeb measures, as well as paving the way for later articles that use them.

A second area in which NSA has had a significant impact is that of functional analysis, with the *nonstandard hull* construction playing a central rôle akin to that of the Loeb measure construction (indeed, these are related). Every nonstandard normed space gives rise (via this construction) to a *standard* Banach space – called its nonstandard hull. In his article here, Wolff gives an introduction to some of those fields of functional analysis where methods of NSA (and especially nonstandard hulls) have been found to be helpful and illuminating – including the structure of Banach spaces, elementary spectral theory, and the theory of operators; he includes some results as yet unpublished.

The field of differential equations (DEs) – including partial DEs and stochastic DEs (both ordinary and partial) – has long been one that has seen significant applications of NSA. A variety of powerful techniques has been developed, such as:

(a) the representation of DEs by nonstandard difference equations (with infinitesimal time steps); difference equations are easily solved, and the standard part mapping then yields a solution of the original DE.

(b) the use of infinitesimals to study perturbations of DEs;

(c) the use of hyperfinite dimensional ODEs to represent PDEs, and thus solve them using the 'transfer' of the elementary theory of ODEs.

Four of the advanced articles[3] in this volume give, among them, a sample of the work that is being done in this area. The first of these is Benoit's, which illustrates the uses of NSA in ODE theory, as developed extensively by the French school of nonstandard analysts. He touches on the solution of ODEs, and the use of infinitesimal techniques for studying questions of stability and perturbations of solutions. As is customary among the members of the French school (which extends beyond national boundaries), Benoit uses the dialect of NSA known as Internal Set Theory. This was first proposed as an alternative, axiomatic foundation for NSA by E. Nelson [4] in 1977. The essence of the subject is very much the same whichever approach is used – that of Nelson or the approach of Robinson, as cast in user friendly form by Henson in this volume. With a few important exceptions, arguments can be readily translated from one dialect to the other.

Following Benoit's article is that by Jin, which touches on advanced foundational issues of NSA. A nonstandard extension (or universe) that is adequate for many purposes is one that is $\aleph_1$-*saturated* (besides having the all important *transfer principle*). $\aleph_1$-saturation is the property of a nonstandard universe that ensures that Loeb measures are countably additive and that nonstandard hulls are complete with respect to their norms. It is possible, however, and sometimes necessary, to work with nonstandard universes with stronger 'saturation-like' properties, which Jin carefully describes. He indicates the relationship between these 'better' nonstandard universes, and also gives some powerful consequences of their existence, in areas such as descriptive set theory, database theory and Loeb measure theory.

A series of articles on stochastic analysis follows, beginning with Lindstrøm's, which builds on Ross' paper to lay the foundation of the hyperfinite (infinitesimal) approach to stochastic analysis. He commences with one of the best known contributions to the subject, Anderson's construction of Brownian motion and the Itô integral. This is still arguably the best construction available since it makes completely rigorous the intuition that $db^2 = dt$, from which much of stochastic calculus follows. Lindstrøm goes on to develop the more general theory of stochastic integration with respect to martingales, and outlines applications to stochastic DEs, extending the difference equation approach discussed by Benoit for ODEs. Further applications that demonstrate the power of this approach are to Brownian local time and the infinite dimensional Ornstein–Uhlenbeck process.

[3]Those by Benoit, Keisler, Arkeryd and Capiński.

[4]Nelson, E., Internal set theory, *Bull. Amer. Math. Soc.* **83**, pp.1165-93.

Keisler continues the stochastic DE theme in his article, showing how the hyperfinite difference approach that he first developed gives existence of solutions with extra properties (for example, solutions that are Markov processes, or are optimal in some sense). He then goes on to relate this work to very recent developments of his, involving notions of *neocompact sets* and *neocontinuous functions.* These play a central rôle in an ongoing programme to describe in purely standard terms the properties of Loeb spaces that make them so powerful in applications – particularly in proving existence results in analysis and probability.

The article by Kopp describes the application of the hyperfinite approach to stochastic integration to modern mathematical finance theory – which is currently a very fashionable field of applied stochastic analysis. Standardly there are two approaches – using either *discrete* or *continuous* mathematical models for the processes involved. The use of hyperfinite (nonstandardly discrete) models, whose standard parts are continuous, clarifies the relationship between these two approaches, and in fact has been used to establish new results concerning convergence of the discrete theory to its continuous counterpart.

From the beginning, NSA was seen as having great potential for mathematical physics, both in providing powerful analytic tools and also for modelling (using the existence of infinitesimal and infinite numbers, and other 'ideal' mathematical objects that NSA offers). Arkeryd's article gives three illustrations of the way in which this potential has been realised in his own work: the first uses infinitesimal techniques to answer a longstanding question from kinetic theory; the second discusses time asymptotics for equations of Boltzmann type, a field which has previously seen a number of successful applications of nonstandard machinery. Arkeryd's final example considers properties of the Schrödinger equation when Planck's constant $\hbar$ is taken to be infinitesimal, with applications to semiclassical limits.

The volume concludes with Capiński's brief survey of a nonstandard approach to hydromechanics – in particular the Navier–Stokes equations – which has been and is continuing to be fruitful. The essential idea is to model the Hilbert space setting of the equations by hyperfinite dimensional (nonstandard) equations which can be solved easily; then standard parts give standard solutions. This simple scheme has been used in deterministic and stochastic settings to solve open problems, and also, using Loeb measures, to provide statistical solutions in each case.

Acknowledgments

This volume, and the Advanced Study Institute from which it stems, would not have been possible without the generous support of the Scientific and Environmental Affairs Division of NATO. For this, heartfelt

thanks are offered by all who participated in the ASI – the twelve lecturers (whose articles appear in this volume), and sixty-one 'students' from nineteen countries, who entered wholeheartedly into two weeks of intense formal and informal scientific activity.

The International Centre for Mathematical Sciences, Edinburgh was the other pillar of support which made the ASI possible. From the original conception of *NSA96*, and through all the planning stages, the encouragement and assistance of two people in particular was invaluable: Professor Angus Macintyre, FRS, Scientific Director of ICMS, and Professor Elmer Rees, FRSE, Chairman of the Executive Committee of ICMS. We take this opportunity to record our appreciation and gratitude to them.

Our sincere thanks, on behalf of all ASI participants, also go to Margaret Cook of ICMS, who acted magnificently in many rôles – dealing with publicity, applications, accommodation, course materials, social activities, financial administration and generally as mother figure to the whole ASI.

Finally, as organizers of the ASI, we would like to record here our thanks to the lecturers, who are also contributors to this volume. We appreciate not only their work in preparing and delivering courses, and reworking lecture notes into the articles herein; we would also like to thank them for the spirit in which they participated in the ASI – organizing and leading informal tutorial sessions and splinter groups, often late into the evenings, and generally helping to create a family atmosphere for the duration of the ASI. It is our hope that these efforts, and this permanent record of the material that was presented, will lead to a lasting wider enjoyment of NSA.

The Editors, and Organizing Committee of the ASI:

Nigel J. Cutland, *Director*
Leif O. Arkeryd
C. Ward Henson

FOUNDATIONS OF NONSTANDARD ANALYSIS

A Gentle Introduction to Nonstandard Extensions

C. WARD HENSON
University of Illinois at Urbana-Champaign
Department of Mathematics
1409 West Green Street
Urbana, Illinois 61801
USA
email: `henson@math.uiuc.edu`

1. Introduction

There are many introductions to nonstandard analysis, (some of which are listed in the References) so why write another one? All of the existing introductions have one or more of the following features: (A) heavy use of logical formalism right from the start; (B) early introduction of set theoretic apparatus in excess of what is needed for most applications; (C) dependence on an explicit construction of the nonstandard model, usually by means of the ultrapower construction.

All three of these features have negative consequences. The early use of logical formalism or set theoretic structures is often uncomfortable for mathematicians who do not have a background in logic, and this can effectively deter them from using nonstandard methods. The explicit use of a particular nonstandard model makes the foundations too specific and inflexible, and often inhibits the free use of the ideas of nonstandard analysis. In this exposition we intend to avoid these disadvantages. The readers for whom we have written are experienced mathematicians (including advanced students) who do not necessarily have any background in or even tolerance for symbolic logic. We hope to convince such readers that nonstandard methods are accessible and that the small amount of logical notation which turns out to be useful in applying them is actually simple and natural. Of course readers who do have a background in logic may also find this approach useful.

We give a natural, geometric definition of nonstandard extension in Section 2; no logical formulas are used and there are no set theoretic structures.

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 1–49.

In Section 3 we introduce logical notation of the kind that all students of mathematics encounter, and we carefully show how it can be used without difficulty to obtain useful facts about nonstandard extensions. In Section 4 we extend the concept of nonstandard extension to mathematical settings in which there may be several basic sets (such as the vector space setting, where there is a field F and a vector space V). In Sections 5 and 6 we show how these ideas can be used to introduce nonstandard extensions in which sets and other objects of higher type can be handled, as is certainly necessary for applications of nonstandard methods in such areas as abstract analysis and topology. However, we do this in stages; in particular, in Section 5 we indicate how to deal with nonstandard extensions in a simple setting where a limited amount of set theoretic apparatus has been introduced. Such limited frameworks for nonstandard analysis are nonetheless adequate for essentially all applications. Section 6 treats the full superstructure apparatus which has become one of the standard ways of formulating nonstandard analysis and which is frequently used in the literature. In Section 7 we briefly discuss saturation properties of nonstandard extensions.

In several places we introduce specific nonstandard extensions using the ultraproduct construction, and we explore the meaning of certain basic concepts (such as *internal set*) in these concrete settings. (See the last parts of Sections 2, 4, 5, and 7.) Our experience shows that it is often helpful at the beginning to have such explicit nonstandard extensions at hand. As noted above, however, we think it is limiting to become dependent on such a construction and we encourage readers to adopt the more flexible axiomatic approach as quickly as possible.

In writing this exposition we have benefitted greatly from conversations with Lou van den Dries about the best ways to present ideas from model theory to the general mathematical public. His ideas are presented in [5] and, with Chris Miller, in [6], and our treatment obviously depends heavily on that work. We have also freely used many ideas from other expositions of nonstandard analysis (listed among the References) and from the other Chapters in this book. To all these authors we express our sincere appreciation, and we recommend their writings to the reader who finishes this exposition with a desire to learn more about how nonstandard methods can be applied.

2. Nonstandard Extensions

The starting point of nonstandard analysis is the construction and use of an ordered field $^*\mathbb{R}$ which is a proper extension of the usual ordered field $\mathbb{R}$ of real numbers, and which satisfies *all of the properties of* $\mathbb{R}$ (in a sense that we will soon make precise). We refer to $^*\mathbb{R}$ as a field of *nonstandard* *real numbers*, or as a field of *hyperreal numbers*. Because the ordered field $\mathbb{R}$ is Dedekind complete, it follows that the extension field $^*\mathbb{R}$ will necessarily have among its new elements both infinitesimal and infinite numbers. These new numbers play a fundamental role in nonstandard analysis, which was created by Abraham Robinson [14] in order to make reasoning with infinitesimals rigorous. (An element a of $^*\mathbb{R}$ is *finite* if there exists $r \in \mathbb{R}$ such that $-r \leq a \leq r$ in $^*\mathbb{R}$; otherwise a is *infinite*; a is *infinitesimal* if $-r \leq a \leq r$ holds for every positive $r \in \mathbb{R}$. In some places a finite number is called *limited.*)

It is easily seen that a proper extension field $^*\mathbb{R}$ of $\mathbb{R}$ cannot satisfy literally *all* the properties of $\mathbb{R}$. For example it cannot be Dedekind complete. (The set of finite numbers in $^*\mathbb{R}$ cannot have a least upper bound s, because then $s-1$ would be a smaller upper bound.) The challenge was to establish a clear and consistent foundation for reasoning with infinitesimals, that captured the known heuristic arguments as much as possible. This was accomplished by Abraham Robinson in the 1960s. The purpose of this paper is to describe the essential features of the resulting frameworks without getting bogged down in technicalities of formal logic and without becoming dependent on an explicit construction of a specific field $^*\mathbb{R}$.

We usually think of $\mathbb{R}$ as being equipped with additional structure, in the form of distinguished sets, relations, and functions; we include whatever objects play a role in the mathematical problems we are considering. For example, these will normally include the set $\mathbb{N}$ of natural numbers (non-negative integers) and often such functions as the sine, the cosine, exponentiation, and the like. When we say (as we did above) that $^*\mathbb{R}$ *satisfies all of the properties of* $\mathbb{R}$, we mean (in part) that each of these additional sets, relations, and functions on $\mathbb{R}$ will have a counterpart on $^*\mathbb{R}$, and that the entire system of counterpart objects will satisfy an appropriate set of conditions. For example, if we are thinking of $\mathbb{N}$ as one of the given sets, then $^*\mathbb{R}$ contains a discrete set $^*\mathbb{N}$ which is the counterpart to $\mathbb{N}$. The conditions that we impose on $^*\mathbb{R}$ and $^*\mathbb{N}$ will imply that $\mathbb{N}$ is an initial segment of $^*\mathbb{N}$, and that the elements of $^*\mathbb{N} \setminus \mathbb{N}$ are infinite numbers in $^*\mathbb{R}$. Moreover, for every positive number r in $^*\mathbb{R}$ there will be a unique $N \in {}^*\mathbb{N}$ which is the hyperinteger part of r, in the sense that $N \leq r < N+1$. We will refer to $^*\mathbb{N}$ as the set of *nonstandard natural numbers* or as the set of *hypernatural numbers.*

The presence of infinitesimal and infinite numbers allows us to give elegant and useful characterizations of many important mathematical concepts, and this phenomenon is the basis for a large part of the impact of nonstandard analysis. For example, suppose $(s_n)_{n\in\mathbb{N}}$ is a sequence of real numbers and t is a real number. Then one can prove the following characterization of the limit concept:

$$s_n \to t \text{ as } n \to \infty \iff s_N \approx t \text{ for all infinite } N \in {}^*\mathbb{N}.$$

(For any two numbers $s, t \in {}^*\mathbb{R}$ we write $s \approx t$ to mean that the difference $s - t$ is infinitesimal.) Using such a characterization allows many heuristic arguments about limits to be made precise; for example, it becomes easy to give elementary algebraic proofs of the basic facts about the algebra of limits.

Note that in the characterization of the limit condition given above, we used the expression s_N where N was an element of ${}^*\mathbb{N}$. This requires explanation, since s_n was originally given only for $n \in \mathbb{N}$. We think of the sequence as a function $s: \mathbb{N} \to \mathbb{R}$ and we regard this function as part of the basic apparatus with which $\mathbb{R}$ is initially equipped. Therefore, it has a counterpart on ${}^*\mathbb{R}$, which will be a function defined on ${}^*\mathbb{N}$ and having values in ${}^*\mathbb{R}$. It is this function that we have in mind when we write s_N for $N \in {}^*\mathbb{N}$.

We are now ready to give a formal description of the properties we require our nonstandard real field ${}^*\mathbb{R}$ to satisfy. For the moment we will only consider *first order* structure on $\mathbb{R}$. Therefore we will not yet be considering the higher order objects of analysis, such as measures, Banach spaces, and the like. We start out in this limited way for pedagogical reasons, to make the task of mastering the fundamental language of nonstandard analysis easier for beginners. (Later on, in Section 5 and especially in Section 6, we will add the machinery of higher type objects which is needed for the full range of arguments in nonstandard analysis.)

We consider $\mathbb{R}$ as being equipped with *all possible* first order properties (*i.e.* sets and relations) and functions. We do this in order to have a foundation which is as flexible as possible and which provides any object we might need later in handling specific mathematical problems. In order to make our basic definition simpler technically, we handle functions by means of their graphs. Therefore, we take the point of view that our additional structure on $\mathbb{R}$ consists of the collection of all possible subsets of every Cartesian power $\mathbb{R}^n$, as n ranges over the integers ≥ 0.

Next we give the key definition. In it we give a precise description of the properties that must be preserved by passage to the nonstandard extension. The requirements are simple and natural, and they have a distinctly geometric character. (Strictly speaking we are defining here a *first order* concept of nonstandard extension; the definition will be suitably modified below when we add higher order objects to our setting.)

2.1. Definition. [Nonstandard Extension of a Set] *Let* $\mathbb{X}$ *be a nonempty set. A* **nonstandard extension of** $\mathbb{X}$ *consists of a mapping that* *assigns a set* $^*A$ *to each* $A \subseteq \mathbb{X}^m$ *for all* $m \geq 0$*, such that* $^*\mathbb{X}$ *is non-empty and the following conditions are satisfied for all* $m, n \geq 0$*:*

(E1) The mapping preserves Boolean operations on subsets of $\mathbb{X}^m$*: if* $A \subseteq \mathbb{X}^m$*, then* $^*A \subseteq (^*\mathbb{X})^m$*; if* $A, B \subseteq \mathbb{X}^m$*, then* $^*(A \cap B) = (^*A \cap {}^*B)$*,* $^*(A \cup B) = (^*A \cup {}^*B)$*, and* $^*(A \setminus B) = (^*A) \setminus (^*B)$*.*

(E2) The mapping preserves basic diagonals: if $1 \leq i < j \leq m$ *and* $\Delta = \{(x_1, \ldots, x_m) \in \mathbb{X}^m \mid x_i = x_j\}$ *then* $^*\Delta = \{(x_1, \ldots, x_m) \in (^*\mathbb{X})^m \mid x_i = x_j\}$*.*

(E3) The mapping preserves Cartesian products: if $A \subseteq \mathbb{X}^m$ *and* $B \subseteq \mathbb{X}^n$*, then* $^*(A \times B) = {}^*A \times {}^*B$*. (We regard* $A \times B$ *as a subset of* $\mathbb{X}^{m+n}$*.)*

(E4) The mapping preserves projections that omit the final coordinate: let π *denote projection of* $n+1$*-tuples on the first* n *coordinates; if* $A \subseteq \mathbb{X}^{n+1}$ *then* $^*(\pi(A)) = \pi(^*A)$*.*

While this definition is reasonably elegant and can be comprehended rather easily, there is certainly some work to be done before we can exploit it. For example, suppose we have a nonstandard extension of $\mathbb{R}$. How do we prove that the subset $^*\mathbb{N}$ of $^*\mathbb{R}$ has the properties that were claimed above? (Namely, that $\mathbb{N}$ is an initial segment of $^*\mathbb{N}$, that the elements of $^*\mathbb{N} \setminus \mathbb{N}$ are infinite numbers in $^*\mathbb{R}$, and that for every positive number r in $^*\mathbb{R}$ there is a unique $N \in {}^*\mathbb{N}$ which satisfies $N \leq r < N + 1$.) Moreover, $^*\mathbb{R}$ is supposed to be an ordered field extension of $\mathbb{R}$, and even this does not seem to be directly guaranteed by the conditions in the definition.

We first turn to a series of elementary arguments which prove some of the most basic properties of nonstandard extensions, especially those having to do with the handling of functions. Not only are the results important, but the arguments illustrate how one can derive information from conditions (E1) – (E4). Near the end of the Section we continue this theme by means of a set of Exercises.

2.2. Proposition. *For each* $n \geq 0$*,* $^*(\mathbb{X}^n) = (^*\mathbb{X})^n$ *and* $^*\emptyset = \emptyset$*.*

Proof. The first equation follows from repeated use of (E3) and the second equation follows from (E1); note that $^*\emptyset = {}^*(\emptyset \setminus \emptyset) = {}^*\emptyset \setminus {}^*\emptyset = \emptyset$. □

2.3. Proposition. *If* $A \subseteq \mathbb{X}^m$ *is non-empty, then* $^*A$ *is also non-empty. Therefore, for any* $A, B \subseteq \mathbb{X}^m$*,* $^*A = {}^*B \iff A = B$*.*

Proof. For ease of notation we consider only the case $m = 2$. Let π_2 and π_3 be the projections defined by $\pi_2(x, y) = x$ and $\pi_3(x, y, z) = (x, y)$ respectively. If $A \subseteq \mathbb{X}^2$ is non-empty, then $\mathbb{X} = \pi_2(\pi_3(\mathbb{X} \times A))$. Using (E4) we get $^*\mathbb{X} = \pi_2(\pi_3(^*\mathbb{X} \times {}^*A))$. Since $^*\mathbb{X}$ is non-empty, it follows that $^*A$ must also be non-empty. The second statement follows from the first and (E1). □

2.4. Proposition. *For all* $A, B \subseteq \mathbb{X}^m$, $A \subseteq B \Longleftrightarrow {}^*A \subseteq {}^*B$.

Proof. Suppose $A \subseteq B$. Then $A = A \cap B$, so by (E1) we have ${}^*A = {}^*(A \cap B) = {}^*A \cap {}^*B$ and hence ${}^*A \subseteq {}^*B$. The reverse implication follows by a similar argument and Proposition 2.3. □

2.5. Proposition. *For each* $x \in \mathbb{X}$, ${}^*\{x\}$ *has exactly one element.*

Proof. By Proposition 2.3, ${}^*\{x\}$ has at least one element. Let $\Delta = \{(u, u) \mid u \in \mathbb{X}\}$, and note that $\{x\} \times \{x\} = \{(x, x)\} \subseteq \Delta$. Using (E3) and (E2) we get ${}^*\{x\} \times {}^*\{x\} \subseteq {}^*\Delta = \{(u, u) \mid u \in {}^*\mathbb{X}\}$, from which it follows that ${}^*\{x\}$ has exactly one element. □

Propositions 2.3 and 2.5 allow us to introduce an embedding of $\mathbb{X}$ into ${}^*\mathbb{X}$ which is canonically associated with the given nonstandard extension. After introducing this embedding, we prove that it is fully compatible with the operation of forming n-tuples, and hence with Cartesian products.

2.6. Notation. For each $x \in \mathbb{X}$, we let ${}^*x$ denote the unique element of the set ${}^*\{x\}$. For each $x = (x_1, \ldots, x_n) \in \mathbb{X}^n$ we let ${}^*x = ({}^*x_1, \ldots, {}^*x_n)$. Note that this gives two usages for an expression of the form ${}^*\beta$; if β is an element of $\mathbb{X}$, then ${}^*\beta$ is defined in this paragraph, while if β is a subset of some Cartesian power $\mathbb{X}^m$, then ${}^*\beta$ is the subset of $({}^*\mathbb{X})^m$ which is provided by the given nonstandard extension.

2.7. Definition. *An element of* $({}^*\mathbb{X})^n$ *is called* **standard** *if it is of the form* ${}^*x$ *for some* $x \in \mathbb{X}^n$. *It follows that an element of* $({}^*\mathbb{X})^n$ *is standard if and only if all of its coordinates are standard elements of* ${}^*\mathbb{X}$.

2.8. Proposition. *For each* $x_1, \ldots, x_n \in \mathbb{X}$,

$$ {}^*\{(x_1, \ldots, x_n)\} = \{({}^*x_1, \ldots {}^*x_n)\}. $$

Proof. ${}^*\{(x_1, \ldots, x_n)\} = {}^*(\{x_1\} \times \cdots \times \{x_n\}) = {}^*\{x_1\} \times \cdots \times {}^*\{x_n\} = \{{}^*x_1\} \times \cdots \times \{{}^*x_n\} = \{({}^*x_1, \ldots, {}^*x_n)\}$.

2.9. Proposition. *For each* $A \subseteq \mathbb{X}^m$ *and* $x_1, \ldots, x_m \in \mathbb{X}$,

$$ (x_1, \ldots, x_m) \in A \Longleftrightarrow ({}^*x_1, \ldots, {}^*x_m) \in {}^*A. $$

Proof. Using Propositions 2.4 and 2.8 note that $(x_1, \ldots, x_m) \in A \Longleftrightarrow \{(x_1, \ldots, x_m)\} \subseteq A \Longleftrightarrow {}^*\{(x_1, \ldots, x_m)\} \subseteq {}^*A \Longleftrightarrow \{({}^*x_1, \ldots, {}^*x_m)\} \subseteq {}^*A \Longleftrightarrow ({}^*x_1, \ldots, {}^*x_m) \in {}^*A$. □

2.10. Remark. The map taking $x \in \mathbb{X}$ to $^*x$ is an embedding of $\mathbb{X}$ into $^*\mathbb{X}$. Without loss of generality we may assume that $\mathbb{X}$ is a subset of $^*\mathbb{X}$ and that $^*x = x$ for all $x \in \mathbb{X}$. When this additional condition is satisfied, the given nonstandard extension is truly an extension mapping, in the strong sense that for all $A \subseteq \mathbb{X}^m$, $(^*A) \cap \mathbb{X}^m = A$ (and therefore, in particular, $A \subseteq {}^*A$).

Justification. For $x, y \in \mathbb{X}$ we have: $^*x = {}^*y \iff {}^*\{x\} = {}^*\{y\} \iff \{x\} = \{y\} \iff x = y$, so this map is an embedding. Therefore we may follow the conventional practice of "identifying" $^*x$ with x for all $x \in \mathbb{X}$. The precise way to do this is to construct an isomorphic nonstandard extension as follows: let $\mathbb{Y}$ be a set and $h\colon {}^*\mathbb{X} \to \mathbb{Y}$ a bijection, chosen together so that $\mathbb{X} \subseteq \mathbb{Y}$ and $x = h(^*x)$ for all $x \in \mathbb{X}$. For each $m \geq 0$ and each $A \subseteq \mathbb{X}^m$, let $\Theta(A)$ be the subset of $\mathbb{Y}^m$ defined by

$$\Theta(A) = \{(h(u_1), \ldots, h(u_m)) \mid (u_1, \ldots, u_m) \in {}^*A\}.$$

It is a straightforward exercise using the previous Propositions to show that the set mapping Θ is a nonstandard extension. It is easily seen that $\mathbb{X} \subseteq \mathbb{Y} = \Theta(\mathbb{X})$ and $\Theta(\{x\}) = \{x\}$ for all $x \in \mathbb{X}$, from which it follows that Θ has the extra properties we wanted to achieve. The facts given in the second sentence of this Remark follow immediately using Proposition 2.9. Note that Θ is isomorphic to the original nonstandard extension in a natural sense. □

When we established the framework of nonstandard extensions, we stated briskly that we would handle functions by means of their graphs. Now we must show that this actually works.

2.11. Proposition. *Suppose $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, and let $f\colon A \to B$ be a function; take $\Gamma \subseteq \mathbb{X}^{m+n}$ to be the graph of f. Then $^*\Gamma$ is the graph of a function from $^*A$ to $^*B$.*

Proof. For ease of notation we treat only the case $m = n = 1$. Let π denote the projection defined by $\pi(x, y) = x$. The key properties of Γ which reflect the fact that it is the graph of a function from A to B are the following: $\Gamma \subseteq A \times B$; $\pi(\Gamma) = A$; and

$$(\Gamma \times \Gamma) \cap \{(x, y, u, v) \in \mathbb{X}^4 \mid x = u\} \subseteq \{(x, y, u, v) \in \mathbb{X}^4 \mid y = v\}.$$

The first of these statements expresses the fact that the domain of the function is contained in A and the range is contained in B. The second statement expresses that the domain of the function is A. The third (displayed) statement expresses the fact that Γ is the graph of a function.

Using conditions (E1) – (E4) we conclude: ${}^*\Gamma \subseteq {}^*A \times {}^*B$; $\pi({}^*\Gamma) = {}^*A$; and

$$({}^*\Gamma \times {}^*\Gamma) \cap \{(x, y, u, v) \in ({}^*\mathbb{X})^4 \mid x = u\} \subseteq \{(x, y, u, v) \in ({}^*\mathbb{X})^4 \mid y = v\}.$$

From these conditions the desired statements about ${}^*\Gamma$ follow immediately. □

2.12. Notation. Suppose $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, and let $f: A \to B$ be a function; take $\Gamma \subseteq \mathbb{X}^{m+n}$ to be the graph of f. We denote by ${}^*f$ the function from ${}^*A$ to ${}^*B$ whose graph is ${}^*\Gamma$.

2.13. Proposition. *If f is the identity function on $A \subseteq \mathbb{X}^m$, then ${}^*f$ is the identity function on ${}^*A$.*

Proof. If f is the identity function on $A \subseteq \mathbb{X}^m$, then the graph Γ of f is given by the following definition:

$$\Gamma = \{(x_1, \ldots, x_m, y_1, \ldots, y_m) \in \mathbb{X}^{2m} \mid x_1 = y_1, \ldots, x_m = y_m\} \cap (A \times A).$$

This set is the intersection of $A \times A$ with m diagonal subsets of $\mathbb{X}^{2m}$,

$$\Gamma = \Delta_1 \cap \cdots \cap \Delta_m \cap (A \times A),$$

where for each $1 \leq j \leq m$ we define

$$\Delta_j = \{(x_1, \ldots, x_m, y_1, \ldots, y_m) \in \mathbb{X}^{2m} \mid x_j = y_j\}.$$

Therefore, using (E1) – (E3) we have

$${}^*\Gamma = \{(x_1, \ldots, x_m, y_1, \ldots, y_m) \in ({}^*\mathbb{X})^{2m} \mid x_1 = y_1, \ldots, x_m = y_m\} \cap ({}^*A \times {}^*A).$$

Since ${}^*\Gamma$ is the graph of ${}^*f$, this proves the desired result. □

2.14. Proposition. *Suppose $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, and let $f: A \to B$ be a function. For all $(x_1, \ldots, x_n) \in A$,*

$$({}^*f)({}^*x_1, \ldots, {}^*x_n) = {}^*(f(x_1, \ldots, x_n)).$$

Proof. Take $x_1, \ldots, x_n \in A$ and let $y = f(x_1, \ldots, x_n)$, so $(x_1, \ldots, x_n, y) \in \Gamma$ where Γ is the graph of f. From Proposition 2.9 we get $({}^*x_1, \ldots, {}^*x_n, {}^*y) \in {}^*\Gamma$, so that $({}^*f)({}^*x_1, \ldots, {}^*x_n) = {}^*y$. □

2.15. Proposition. [Permuting and Identifying Variables] *Suppose σ is any function from $\{1, \ldots, m\}$ into $\{1, \ldots, n\}$. Given $A \subseteq \mathbb{X}^m$ define*

$$B = \{(x_1, \ldots, x_n) \in \mathbb{X}^n \mid (x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A\}.$$

Then

$${}^*B = \{(x_1, \ldots, x_n) \in ({}^*\mathbb{X})^n \mid (x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in {}^*A\}.$$

Proof. For ease of notation we consider the case where $A \subseteq \mathbb{X}^3$ is given and $B \subseteq \mathbb{X}^2$ is defined by

$$B = \{(x, y) \in \mathbb{X}^2 \mid (y, x, y) \in A\}.$$

Introduce $C \subseteq \mathbb{X}^5$ by the definition

$$C = \{(x, y, u, v, w) \in \mathbb{X}^5 \mid u = y \wedge v = x \wedge w = y \wedge (u, v, w) \in A\}.$$

Evidently B is the result of projecting C onto the first two coordinates. Moreover, C is the intersection of three diagonal subsets of $\mathbb{X}^5$ and the set $\mathbb{X}^2 \times A$. Therefore, it follows using conditions (E1) – (E4) that

$${}^*C = \{(x, y, u, v, w) \in ({}^*\mathbb{X})^5 \mid u = y \wedge v = x \wedge w = y \wedge (u, v, w) \in {}^*A\}$$

and that ${}^*B$ is the result of projecting ${}^*C$ onto the first two coordinates. The desired result follows immediately. □

2.16. Proposition. *Condition (E4) in Definition 2.1 holds for all projections π from m-tuples to n-tuples, where $n \leq m$. (By calling π a projection we mean that there exists a sequence $1 \leq i(1) < \ldots < i(n) \leq m$ such that π is defined by*

$$\pi(x_1, \ldots, x_m) = (x_{i(1)}, \ldots, x_{i(n)}).)$$

That is, if $A \subseteq \mathbb{X}^m$, then ${}^*(\pi(A)) = \pi({}^*A)$.

Proof. Let π be as described in the statement of the Proposition. Let σ be a permutation of $\{1, \ldots, m\}$ so that $\sigma(i(j)) = j$ for all $j = 1, \ldots, n$. Let $A \subseteq \mathbb{X}^m$ be given and define

$$B = \{(x_1, \ldots, x_m) \in \mathbb{X}^m \mid (x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A\}.$$

It is routine to check that $\pi(A)$ is the same as the result of projecting B onto the first n coordinates. Condition (E4) (applied $m - n$ times) therefore implies that ${}^*(\pi(A))$ is the result of projecting ${}^*B$ onto the first n coordinates. Proposition 2.15 implies that

$${}^*B = \{(x_1, \ldots, x_m) \in ({}^*\mathbb{X})^m \mid (x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in {}^*A\}.$$

Hence $\pi({}^*A)$ is the same as the result of projecting ${}^*B$ onto the first n coordinates. Therefore ${}^*(\pi(A)) = \pi({}^*A)$. □

2.17. Proposition. *Let $A \subseteq \mathbb{X}^{m+n}$ and $a = (a_1, \ldots, a_m) \in \mathbb{X}^m$. Define*

$$A(a) = \{(x_1, \ldots, x_n) \in \mathbb{X}^n \mid (a_1, \ldots, a_m, x_1, \ldots, x_n) \in A\}$$

and similarly

$$({}^*A)({}^*a) = \{(x_1, \ldots, x_n) \in ({}^*\mathbb{X})^n \mid ({}^*a_1, \ldots, {}^*a_m, x_1, \ldots, x_n) \in {}^*A\}$$

Then ${}^*(A(a)) = ({}^*A)({}^*a)$.

Proof. For ease of notation we consider only the case $m = n = 1$. Let π denote the projection defined by $\pi(x, y) = y$. Note that

$$A(a) = \pi(A \cap (\{a\} \times \mathbb{X})).$$

Therefore, using conditions (E1) and (E3), and Proposition 2.16 we have

$$\begin{aligned} {}^*(A(a)) &= \pi({}^*A \cap ({}^*\{a\} \times {}^*\mathbb{X})) \\ &= \pi({}^*A \cap (\{{}^*a\} \times {}^*\mathbb{X})) = ({}^*A)({}^*a). \end{aligned}$$

□

2.18. Proposition. *Suppose $A \subseteq \mathbb{X}^m, B \subseteq \mathbb{X}^n$ and $C \subseteq \mathbb{X}^p$; let $f: A \to B$ and $g: B \to C$ be functions. Then ${}^*(g \circ f) = ({}^*g) \circ ({}^*f)$.*

Proof. For ease of notation we treat only the case $m = n = p = 1$. Let Γ_f be the graph of f and Γ_g the graph of g, and let Γ be the graph of the composition $g \circ f$. Let π be the projection defined by letting $\pi(x, y, u, v) = (x, v)$. Consider the set $A \subseteq \mathbb{X}^4$ defined by

$$A = \{(x, y, u, v) \in \mathbb{X}^4 \mid y = u\} \cap (\Gamma_f \times \Gamma_g).$$

Evidently $\Gamma = \pi(A)$. The desired result follows immediately from Proposition 2.16. □

Next we give some Exercises which continue the themes developed above. The reader is advised to solve them, as much as possible using the methods of this Section. They will be easier to solve once the machinery of logical notation is developed, as it will be in the next Section. However, especially for readers who have no previous experience with logic, working these Exercises at this point will bring significant benefits. Most of all, such effort will cause the reader to appreciate the advantages of logical notation and to understand how simple are the few technical ideas that it embodies.

2.19. Exercise. Condition (E2) holds for all diagonal sets $\Delta \subseteq \mathbb{X}^n$. By a diagonal set we mean that there is an equivalence relation E on $\{1, \ldots, n\}$ such that $\Delta = \{(x_1, \ldots, x_n) \in \mathbb{X}^n \mid x_i = x_j \text{ whenever } iEj\}$. For every such Δ,

$${}^*\Delta = \{(x_1, \ldots, x_n) \in ({}^*\mathbb{X})^n \mid x_i = x_j \text{ whenever } iEj\}.$$

If A is a subset of $\mathbb{X}^m$, then $\{({}^*a_1, \ldots, {}^*a_m) \mid (a_1, \ldots, a_m) \in A\}$ is a subset of ${}^*A$, by Proposition 2.9. Indeed, $\{({}^*a_1, \ldots, {}^*a_m) \mid (a_1, \ldots, a_m) \in A\}$ is precisely the set of standard elements of ${}^*A$. The next two Exercises explore the extent to which $\{({}^*a_1, \ldots, {}^*a_m) \mid (a_1, \ldots, a_m) \in A\}$ is a *proper* subset of ${}^*A$.

2.20. Exercise. If A is a finite subset of $\mathbb{X}^m$, then

$$^*A = \{(^*x_1, \ldots, {}^*x_m) \mid (x_1, \ldots, x_m) \in A\}.$$

In particular, if A is finite, then $^*A$ is finite and has the same cardinality as A, and all of its elements are standard.

2.21. Definition. *A nonstandard extension of* $\mathbb{X}$ *is called* **proper** *if for every infinite subset* A *of* $\mathbb{X}$*,* $^*A$ *contains a nonstandard element.*

2.22. Exercise. Suppose our nonstandard extension is proper. Then, for any infinite set $A \subseteq \mathbb{X}^m$, $^*A$ has a nonstandard element.

2.23. Exercise. Let $A \subseteq \mathbb{X}^m$ and suppose $f: A \to \mathbb{X}^n$ is a function.

(a) If $B \subseteq A$, then $^*(f(B)) = (^*f)(^*B)$.
(b) If $C \subseteq \mathbb{X}^n$, then $^*(f^{-1}(C)) = (^*f)^{-1}(^*C)$.
(c) If $B \subseteq A$, then $^*(f|B) = (^*f)|(^*B)$.

2.24. Exercise. For $j = 1, \ldots, n$ let $f_j: \mathbb{X}^m \to \mathbb{X}$ be a function, and let $f = (f_1, \ldots, f_n): \mathbb{X}^m \to \mathbb{X}^n$ be the function with $f_1, \ldots, f_n$ as its coordinates. Then $^*f = (^*f_1, \ldots, {}^*f_n)$.

2.25. Exercise. Suppose $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, and let $f: A \to B$ be a function.

(a) f is injective $\Longleftrightarrow$ $^*f$ is injective.
(b) f is surjective $\Longleftrightarrow$ $^*f$ is surjective.
(c) If f is a bijection and its inverse is g, then $^*g$ is the inverse of $^*f$.

2.26. Exercise. Consider a nonstandard extension of $\mathbb{R}$. The set $^*\mathbb{R}$ is equipped with binary functions $^*+$ and $^*\times$ and with a binary relation $^*<$. Equipped with this additional structure, $^*\mathbb{R}$ is an ordered field.

2.27. Exercise. Expand all proofs in this Section so that they are fully general and cover all cases of the results being proved.

We conclude this Section by using the ultraproduct construction to prove the existence of proper nonstandard extensions. (See Definition 2.21.)

Let J be any infinite set and let $\mathcal{U}$ be an ultrafilter on J. Consider an indexed family $(A_j \mid j \in J)$ of non-empty sets. We define the ultraproduct of the sets $(A_j \mid j \in J)$ with respect to the ultrafilter $\mathcal{U}$. This will be denoted $\Pi_{\mathcal{U}}(A_j \mid j \in J)$ or simply $\Pi_{\mathcal{U}} A_j$. To define the ultraproduct, consider the ordinary Cartesian product ΠA_j of the given family of sets; this is the set of all functions α which are defined on J and which satisfy $\alpha(j) \in A_j$ for all $j \in J$. We define a relation $\sim$ on ΠA_j by

$$\alpha \sim \beta \Longleftrightarrow \{j \in J \mid \alpha(j) = \beta(j)\} \in \mathcal{U}.$$

This is an equivalence relation, as can be proved easily using basic properties of ultrafilters. For each $\alpha \in \Pi A_j$ let $[\alpha]$ denote the equivalence class of α under $\sim$. The *ultraproduct* $\Pi_{\mathcal{U}} A_j$ is then defined to be the set of all equivalence classes $[\alpha]$ as α ranges over ΠA_j:

$$\Pi_{\mathcal{U}}(A_j \mid j \in J) := \{[\alpha] \mid \alpha \in \Pi(A_j \mid j \in J)\}.$$

If the sets $(A_j \mid j \in J)$ are all equal to the same set A, then the ultraproduct $\Pi_{\mathcal{U}}(A_j \mid j \in J)$ is called an *ultrapower* of A and it is denoted $A^J/\mathcal{U}$.

2.28. Theorem. [Existence of Nonstandard Extensions] *Each nonempty set $\mathbb{X}$ has a proper nonstandard extension, in which the set ${}^*\mathbb{X}$ may be taken to be an ultrapower of $\mathbb{X}$ with respect to a countably incomplete ultrafilter.*

Proof. Let J be any infinite index set and let $\mathcal{U}$ be any countably incomplete ultrafilter on J. This means that there exists a sequence $(F_k)_{k\in\mathbb{N}}$ of sets in $\mathcal{U}$ whose intersection $\bigcap\{F_k \mid k \in \mathbb{N}\}$ is empty. There exists such an ultrafilter on each infinite index set J. Moreover if J is countable and $\mathcal{U}$ is any nonprincipal ultrafilter on J, then it is easy to see that $\mathcal{U}$ must be countably incomplete.

The underlying set ${}^*\mathbb{X}$ of our nonstandard extension will be the ultrapower $\mathbb{X}^I/\mathcal{U}$ defined above. Therefore each element of ${}^*\mathbb{X}$ is an equivalence class $[\alpha]$ for some function $\alpha: J \to \mathbb{X}$.

Given $m \geq 0$ and $A \subseteq \mathbb{X}^m$, we define ${}^*A \subseteq ({}^*\mathbb{X})^m$ by:

$${}^*A = \{([\alpha_1], \ldots, [\alpha_m]) \mid \{j \in J \mid (\alpha_1(j), \ldots, \alpha_m(j)) \in A\} \in \mathcal{U}\}.$$

In this definition, $\alpha_1, \ldots, \alpha_m$ range over the set $\mathbb{X}^J$ of all functions from J into $\mathbb{X}$.

We need to show this mapping satisfies conditions (E1) – (E4) in Definition 2.1.

(E1) Fix $m \geq 0$ and let $A, B \subseteq \mathbb{X}^m$. It is immediate that ${}^*A$ is a subset of $({}^*\mathbb{X})^m$. Let $\alpha_1, \ldots, \alpha_m$ be functions from J to $\mathbb{X}$ and set

$$F = \{j \in J \mid (\alpha_1(j), \ldots, \alpha_m(j)) \in A\}$$

$$G = \{j \in J \mid (\alpha_1(j), \ldots, \alpha_m(j)) \in B\}.$$

Using properties of the ultrafilter, it is easy to prove

$$([\alpha_1], \ldots, [\alpha_m]) \in {}^*(A \cap B) \iff (F \cap G) \in \mathcal{U} \iff F \in \mathcal{U} \wedge G \in \mathcal{U}$$

$$\iff ([\alpha_1], \ldots, [\alpha_m]) \in {}^*A \text{ and } ([\alpha_1], \ldots, [\alpha_m]) \in {}^*B.$$

Similarly

$$([\alpha_1],\dots,[\alpha_m]) \in {}^*(\mathbb{X}^m \setminus A) \iff J \setminus F \in \mathcal{U} \iff F \notin \mathcal{U}$$
$$\iff ([\alpha_1],\dots,[\alpha_m]) \in ({}^*\mathbb{X}^m) \setminus {}^*A.$$

This suffices to prove condition (E1).

(E2) For simplicity of notation we consider the basic diagonal subset of $\mathbb{X}^2$ given by

$$\Delta = \{(x,y) \in \mathbb{X}^2 \mid x = y\}.$$

Then

$$([\alpha],[\beta]) \in {}^*\Delta \iff \{j \in J \mid \alpha(j) = \beta(j)\} \in \mathcal{U}$$
$$\iff \alpha \sim \beta \iff [\alpha] = [\beta].$$

This shows that ${}^*\Delta$ is the desired diagonal subset of $({}^*\mathbb{X})^2$.

(E3) The fact that Cartesian products are preserved by this mapping is immediate from the definition and an argument similar to the proof of (E1).

(E4) For simplicity of notation we consider only a subset A of $\mathbb{X}^2$ and the projection $\pi(x,y) = x$ onto the first coordinate. Let B be the projection of A under π. Given functions $\alpha, \beta: J \to \mathbb{X}$, let $F = \{j \in J \mid (\alpha(j),\beta(j)) \in A\}$ and $G = \{j \in J \mid \alpha(j) \in B\}$. If $([\alpha],[\beta]) \in {}^*A$ then $F \in \mathcal{U}$ and $F \subseteq G$, so that also $G \in \mathcal{U}$ and hence $[\beta] \in {}^*B$. Conversely, suppose $[\beta] \in {}^*B$ so that $G \in \mathcal{U}$. Define $\alpha(j)$ for $j \in G$ by choosing it so that $(\alpha(j),\beta(j)) \in A$. For $j \notin G$ define $\alpha(j)$ arbitrarily in $\mathbb{X}$. For this pair of functions α, β we have $G \subseteq F$ so $F \in \mathcal{U}$ and therefore $([\alpha],[\beta]) \in {}^*A$. This proves (E4).

To finish the proof we must prove that this nonstandard extension is proper. Let A be an infinite subset of $\mathbb{X}$. Consider a sequence $(F_k)_{k \in \mathbb{N}}$ of sets in $\mathcal{U}$ whose intersection $\bigcap\{F_k \mid k \in \mathbb{N}\}$ is empty. Without loss of generality we may assume $F_0 = J$ and $F_n \supseteq F_{n+1}$ for all $n \in \mathbb{N}$. This allows us to define $d(j)$ for all $j \in J$ to be the largest $k \in \mathbb{N}$ for which $j \in F_k$. Note that for all $n \in \mathbb{N}$ and all $j \in J$, $d(j) = n$ if and only if $j \in F_n \setminus F_{n+1}$. Choose a sequence $(a_k)_{k \in \mathbb{N}}$ out of A which has no repetitions. Define $\alpha: J \to A$ by setting $\alpha(j) = a_{d(j)}$ for all $j \in J$. It remains to show that $[\alpha]$ is not standard. Indeed, if $[\alpha]$ were equal to the standard element ${}^*a$ for some $a \in \mathbb{X}$, then the set $F = \{j \in J \mid \alpha(j) = a\}$ would be an element of $\mathcal{U}$. (See Exercise 2.29.) However, the construction of α ensures that $F = F_n \setminus F_{n+1}$ for some $n \in \mathbb{N}$, and therefore F is not an element of $\mathcal{U}$. □

We conclude this Section with a few Exercises about ultrapower nonstandard extensions.

2.29. Exercise. Let J be an index set and $\mathcal{U}$ an ultrafilter on J, and consider the ultrapower nonstandard extension of $\mathbb{X}$ that is constructed in the proof of Theorem 2.28.

(a) For each $a \in \mathbb{X}$, $^*a = [\alpha]$, where $\alpha: J \to \mathbb{X}$ is the constant function with $\alpha(j) = a$ for all $j \in J$.

(b) Let $f: \mathbb{X}^m \to \mathbb{X}$ be a function. Let $\alpha_1, \dots, \alpha_m$ be elements of $\mathbb{X}^J$ and define $\beta \in \mathbb{X}^J$ by setting $\beta(j) = f(\alpha_1(j), \dots, \alpha_m(j))$ for all $j \in J$. Then $^*f([\alpha_1], \dots, [\alpha_m]) = [\beta]$. Give a similar description of $^*f$ where $f: A \to B$ is any function, with $A \subseteq \mathbb{X}^m$, and $B \subseteq \mathbb{X}^n$.

2.30. Exercise. Let J be an index set and $\mathcal{U}$ an ultrafilter on J. Suppose $A \subseteq \mathbb{X}^m$ and consider the set $^*A$ defined in the proof of Theorem 2.28 above. There is a natural way to identify $^*A$ with the ultrapower $A^J/\mathcal{U}$. (Hint: if $\alpha_1, \dots, \alpha_m$ are functions from J to $\mathbb{X}$, then $\alpha = (\alpha_1, \dots, \alpha_m)$ may be regarded as a function from J into $\mathbb{X}^m$, and every such function arises in this way. If $\{j \in J \mid \alpha(j) \in A\} \in \mathcal{U}$, then there exist functions $\beta_1, \dots, \beta_m$ from J into $\mathbb{X}$ such that (i) for all $j \in J$, $(\beta_1(j), \dots, \beta_m(j)) \in A$ and (ii) $\beta_i \sim \alpha_i$ for all $i = 1, \dots, m$.)

2.31. Exercise. Let $\mathcal{U}$ be a nonprincipal ultrafilter on the index set $\mathbb{N}$ and let $^*\mathbb{R}$ be the ultrapower nonstandard extension of $\mathbb{R}$ that is defined in the proof of Theorem 2.28 above.

(a) Let $\alpha: \mathbb{N} \to \mathbb{R}$ be a sequence which converges to $+\infty$; the element $[\alpha]$ of $^*\mathbb{R}$ is positive infinite. Give some specific examples of such infinite elements of $^*\mathbb{R}$ and compare them with respect to the ordering $^*<$.

(b) Let $\beta: \mathbb{N} \to \mathbb{R}$ be a sequence which converges to 0 from above; the element $[\beta]$ is a positive infinitesimal in $^*\mathbb{R}$. Give some specific examples of such infinitesimal elements of $^*\mathbb{R}$ and compare them with respect to the ordering $^*<$.

3. Logical Formulas

In this Section we discuss how to use informal and familiar logical notation to streamline our reasoning about nonstandard extensions. Logical notation is often suggestive and transparent when defining or describing sets and functions. Using logical formulas permits us to take advantage of our natural linguistic and logical abilities. Moreover, this turns out to be an ideal framework for bringing out the main properties of nonstandard extensions.

To illustrate this use of formulas, let x, y be variables ranging over nonempty sets A, B respectively, and let $\varphi(x, y)$ and $\psi(x, y)$ denote conditions (formulas) on (x, y) defining subsets Φ and Ψ (respectively) of $A \times B$. We consider certain *logical formulas* that can be built up from $\varphi(x, y)$ and $\psi(x, y)$ (on the left below) and the *sets* that are defined by them (on the right):

$$
\begin{array}{rll}
\neg\varphi(x,y) & \text{defines} & \text{the complement of } \Phi \text{ in } A\times B,\\
\varphi(x,y)\vee\psi(x,y) & \text{defines} & \text{the union } \Phi\cup\Psi,\\
\varphi(x,y)\wedge\psi(x,y) & \text{defines} & \text{the intersection } \Phi\cap\Psi,\\
\exists x\,\varphi(x,y) & \text{defines} & \text{the projection } \pi(\Phi)\text{, where } \pi(x,y)=y\\
 & & \text{is the projection onto the second coordinate,}\\
\forall y\,\varphi(x,y) & \text{defines} & \{x\in A \mid \{x\}\times B\subseteq\Phi\}.
\end{array}
$$

Here we are using familiar logical symbols, which have the following meanings:

$\neg$	stands for	the negation, "not",
$\vee$	stands for	the disjunction, "or"
$\wedge$	stands for	the conjunction, "and",
$\exists$	stands for	the existential quantifier, "there exists", and
$\forall$	stands for	the universal quantifier, "for all."

In our use of the quantifiers above, we followed the given restriction that x ranges over A and y ranges over B. This can be made explicit by writing $\exists x \in A\,\varphi(x,y)$ or $\forall y \in B\,\psi(x,y)$ instead of what is written above. In our use of logical formulas we will always have an explicit or implicit understanding about the set over which a given variable ranges.

To illustrate the usefulness of these simple ideas, consider a given function $f\colon A \to B$. The range of f, namely the set $f(A)$, can be defined by the equivalence

$$y \in f(A) \iff \exists x \in A\,[f(x) = y].$$

Let Γ be the graph of f, which is defined as a subset of $A\times B$ by the formula $f(x) = y$. Therefore, this equivalence exhibits the fact that $f(A)$ is the projection of Γ under the projection map π onto the second coordinate. This reduction of *arbitrary* functions to *projections* is used frequently; indeed, we have used it already several times in Section 2, when we used condition (E4) of Definition 2.1 to prove results about functions.

Simple and familiar logical equivalences often capture mathematical facts that seem complicated when they are viewed directly without the use of logical formulas. For example, the familiar equivalence

$$\forall y\,\varphi(x,y) \iff \neg\exists y\,\neg\varphi(x,y)$$

shows that the set defined by $\forall y\,\varphi(x,y)$ can be obtained from Φ by first taking the complement in $A\times B$, then projecting onto the first coordinate, and then taking the complement of that set in A. This technique is particularly useful when dealing with logically complicated notions, such as continuity or differentiability, which we express in the usual way with ϵ's and δ's and quantifiers over them. In such cases we often deal with formulas having more than two variables and with repeated quantifiers.

We will use several additional notational conventions. A formula $\varphi(x, y)$ defining a subset of $A \times B$ will also be viewed sometimes as defining a condition on triples (x, y, z), where z ranges over a non-empty set C; in that case $\varphi(x, y)$ defines a subset of $A \times B \times C$. In such a situation we will indicate the formula also as $\varphi(x, y, z)$ to show that we are thinking of this formula as defining a subset of $A \times B \times C$ instead of just $A \times B$. Here we are making a distinction between the appearance of the formula itself (in which the variable z does not occur) and the notation we use for referring to the formula in a proof or other discussion. This is similar to the situation in algebra where one routinely regards a polynomial $p(x, y)$ as a polynomial in three variables x, y, z in which all monomials containing z are taken to have coefficient 0. For logical formulas the general convention is that when we use notation such as $\varphi(x_1, \ldots, x_n)$ to refer to a formula, then the variables $x_1, \ldots, x_n$ must be distinct and they must *include* all of the variables that occur in the formula in a way that makes them free for substitution. The other variables in the formula, all of which are bound by quantifiers, need not be included in this list. We also sometime denote a formula by writing φ or ψ without any list of variables, when it is not important to name the variables that may be free for substitution. The context will determine which notation we are using.

We use the implication sign $\rightarrow$, as in $\varphi(x, y) \rightarrow \psi(x, y)$, to abbreviate the formula $(\neg\varphi(x, y)) \vee \psi(x, y)$. We use the equivalence symbol $\leftrightarrow$, as in $\varphi(x, y) \leftrightarrow \psi(x, y)$, to abbreviate $[\varphi(x, y) \rightarrow \psi(x, y)] \wedge [\psi(x, y) \rightarrow \varphi(x, y)]$.

Now let us consider the particular logical formulas that we will use in working with nonstandard extensions. For the moment, all of our variables will range over $\mathbb{X}$. For each set $A \subseteq \mathbb{X}^m$ we will regard $(x_1, \ldots, x_m) \in A$ as a formula, in which $x_1, \ldots, x_m$ are variables ranging over $\mathbb{X}$; we do not require these variables to be distinct. Sometimes we will write this formula in the equivalent form $A(x_1, \ldots, x_m)$, if this fits more smoothly with the usual mathematical role of the set A. In a few situations this formula is written in other ways: for example, if A corresponds to a linear ordering $<$, in the sense that A is the set of pairs (a, b) which satisfy the ordering condition $a < b$, then it is natural to write the formula $x < y$ as synonymous with $(x, y) \in A$. All of this is quite familiar usage in mathematics. Moreover, in the formula $(x_1, \ldots, x_n) \in A$ we can replace some or all of the variables x_j by specific elements of $\mathbb{X}$.

If $f\colon A \rightarrow B$ is a function, where $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, then we will also make use of the formulas $f(x_1, \ldots, x_m) = (y_1, \ldots, y_n)$. If Γ is the graph of f, then this formula is equivalent to the formula $(x_1, \ldots, x_m, y_1, \ldots, y_n) \in \Gamma$.

This corresponds to what we did in the previous Section, handling functions by means of their graphs.

Building on the basic formulas discussed in the previous paragraph, we construct more complicated formulas using quantifiers (with variables ranging over the set $\mathbb{X}$) and the logical connectives that are discussed above: $\neg, \vee, \wedge, \rightarrow$, and $\leftrightarrow$. We will refer to these logical formulas as *formulas over* $\mathbb{X}$. (To be precise, the logical formulas we are using here are *first order formulas.* This reflects the fact that the quantifiers we use range over elements of $\mathbb{X}$, and we do not have any quantifiers ranging over subsets of $\mathbb{X}$ or other higher type objects based on $\mathbb{X}$.) To be precise, we have the following definition by induction:

3.1. Definition. **[Formulas Over $\mathbb{X}$]** *Let $\mathbb{X}$ be a non-empty set. The set of* **formulas over** $\mathbb{X}$ *is the smallest set of logical formulas which satisfies the following closure conditions. (We let x and $x_1, \ldots, x_m, y_1, \ldots, y_n$ stand for arbitrary variables, which need not be distinct.)*

(i) For each set $A \subseteq \mathbb{X}^m$, $(x_1, \ldots, x_m) \in A$ is a formula over $\mathbb{X}$;

(ii) for each function $f\colon A \rightarrow B$, where $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$,

$$f(x_1, \ldots, x_m) = (y_1, \ldots, y_n)$$

is a formula over $\mathbb{X}$;

(iii) if $\varphi(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is a formula over $\mathbb{X}$ and $a_1, \ldots, a_n \in \mathbb{X}$, then $\varphi(x_1, \ldots, x_m, a_1, \ldots, a_n)$ is a formula over $\mathbb{X}$;

(iv) if φ and ψ are formulas over $\mathbb{X}$, then $\neg\varphi$, $\varphi \vee \psi$, $\varphi \wedge \psi$, $\varphi \rightarrow \psi$, $\varphi \leftrightarrow \psi$, $\exists x\, \varphi$, and $\forall x\, \varphi$ are formulas over $\mathbb{X}$.

Note that functions appear in formulas over $\mathbb{X}$ only through their graphs. This is less restrictive than it may seem at first. For example, suppose f, g, and h are functions from $\mathbb{X}$ into itself, and we want to express the condition that h is the composition of f and g. This can be done using the following formula

$$\forall x \forall y\, [h(x) = y \leftrightarrow \exists z\, (g(x) = z \wedge f(z) = y)]$$

which is a formula over $\mathbb{X}$. Suppose $< \,\subseteq \mathbb{X}^2$ is an ordering relation on $\mathbb{X}$, and we want to express the condition that $f(x) < g(x)$ holds for all elements x of $\mathbb{X}$. This can be done using the following formula over $\mathbb{X}$:

$$\forall x \forall y \forall z\, [(f(x) = y \wedge g(x) = z) \rightarrow y < z].$$

In this way we see how statements involving the composition of functions and the substitution of functions in predicates can be expressed using formulas over $\mathbb{X}$.

If we consider a formula over $\mathbb{X}$ syntactically, as a string of symbols, then it is important to distinguish two different ways in which variables can be used. The *free* variables are the ones for which values can be substituted; all other occurrences of variables are *bound*, meaning that their use is controlled by the occurrence of quantifiers in the formula. The best way to make this precise is to give the following inductive definition of free variables in a formula φ:

(i) all variables occurring within a basic formula $(x_1, \ldots, x_m) \in A$ or $f(x_1, \ldots, x_m) = (y_1, \ldots, y_n)$ are free variables;

(ii) the free variables in $\neg\varphi$ are the same as the free variables in φ;

(iii) the free variables in $\varphi \vee \psi$, $\varphi \wedge \psi$, $\varphi \rightarrow \psi$, or $\varphi \leftrightarrow \psi$ are the free variables in φ together with the free variables in ψ;

(iv) the free variables in $\exists x\, \varphi$, and $\forall x\, \varphi$ are the free variables in φ that are distinct from x.

If φ is a formula whose free variables are among $x_1, \ldots, x_m$, we indicate this fact by writing the formula as $\varphi(x_1, \ldots, x_m)$; when establishing this notation for the first time we require that $x_1, \ldots, x_m$ be distinct. We then will indicate the result of substituting other variables or functional expressions $t_1, \ldots, t_m$ for $x_1, \ldots, x_m$ respectively by writing the result of the substitutions in the form $\varphi(t_1, \ldots, t_m)$; in such a situation we do not require that the substituted expressions $t_1, \ldots, t_m$ be distinct.

A *sentence* is a logical formula with no free variables. It makes a definite true-or-false statement about the structures to which it refers.

Now we discuss how formulas over $\mathbb{X}$ can be used in connection with nonstandard extensions of $\mathbb{X}$. Consider a specific nonstandard extension of $\mathbb{X}$, based on the set $^*\mathbb{X}$. We will regard this nonstandard extension as fixed for the rest of this Section. Since $^*\mathbb{X}$ is also a (non-empty) set, we also have the class of formulas over $^*\mathbb{X}$. We will now see that there is an important connection between the formulas over $\mathbb{X}$ and (some of the) formulas over $^*\mathbb{X}$. We will normally use the convention that lower case variables such as $x_1, \ldots, x_n$ range over $\mathbb{X}$ and upper case variables such as $X_1, \ldots, X_n$ range over $^*\mathbb{X}$. We will *never* mix the two types of variables in the same formula. More generally, all of the formulas we consider will either be formulas over $\mathbb{X}$ or they will be formulas over $^*\mathbb{X}$.

3.2. Definition. **[$*$-Transform of a Formula Over $\mathbb{X}$]** *Fix a nonstandard extension of the set $\mathbb{X}$. Let $\varphi(x_1, \ldots, x_n)$ be a formula over $\mathbb{X}$. The* **$*$-transform** *of $\varphi(x_1, \ldots, x_n)$ is a formula over $^*\mathbb{X}$, written $^*\varphi(X_1, \ldots, X_n)$, which is defined inductively by the following conditions:*

Basis cases:

(i) Let $\{\sigma(1), \ldots, \sigma(m)\} \subseteq \{1, \ldots, n\}$ and $A \subseteq \mathbb{X}^m$; the $*$-transform of the basic formula

$$(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A$$

is the formula

$$(X_{\sigma(1)}, \ldots, X_{\sigma(m)}) \in {}^*A;$$

similarly, the $*$-transform of

$$f(x_{\sigma(1)},\dots,x_{\sigma(k)}) = (x_{\sigma(k+1)},\dots,x_{\sigma(m)})$$

is

$${}^*f(X_{\sigma(1)},\dots,X_{\sigma(k)}) = (X_{\sigma(k+1)},\dots,X_{\sigma(m)});$$

(ii) if $\varphi(x_1,\dots,x_n,y_1,\dots,y_p)$ is a basic formula (as treated in (i)) and $a_1,\dots,a_p \in \mathbb{X}$, then the $*$-transform of $\varphi(x_1,\dots,x_n,a_1,\dots,a_p)$ is ${}^*\varphi(X_1,\dots,X_n,{}^*a_1,\dots,{}^*a_p)$;

Induction cases: Let φ and ψ be formulas over $\mathbb{X}$;

(iii) the $*$-transform of the negation $\neg\varphi$ is $\neg{}^*\varphi$;

(iv) the $*$-transform of the disjunction $\varphi \vee \psi$ is ${}^*\varphi \vee {}^*\psi$;

(v) the $*$-transform of the conjunction $\varphi \wedge \psi$ is ${}^*\varphi \wedge {}^*\psi$;

(vi) if x is a variable ranging over $\mathbb{X}$, then the $*$-transform of the quantified formula $\exists x\, \varphi$ is $\exists X\, {}^*\varphi$;

(vii) if x is a variable ranging over $\mathbb{X}$, then the $*$-transform of the quantified formula $\forall x\, \varphi$ is $\forall X\, {}^*\varphi$.

While this definition may look complicated, it is merely the precise formulation of a simple idea: constructing the $*$-transform of a formula $\varphi(x_1,\dots,x_n)$ over $\mathbb{X}$ requires the following steps:

(a) Find all of the sets $A \subseteq \mathbb{X}^m$ that occur in $\varphi(x_1,\dots,x_n)$ in basic formulas, and replace each such set by its counterpart ${}^*A$ over ${}^*\mathbb{X}$; similarly, replace each function $f\colon A \to B$ by ${}^*f$ and replace each element a of $\mathbb{X}$ by ${}^*a$, and

(b) replace every variable x in $\varphi(x_1,\dots,x_n)$, including the ones that are used with quantifiers, by a corresponding variable X which ranges over ${}^*\mathbb{X}$.

For example, suppose Γ is a subset of $\mathbb{X}^2$. The sentence over $\mathbb{X}$ given by

$$\forall x \forall y \forall z\, [[(x,y) \in \Gamma \wedge (x,z) \in \Gamma] \to y = z] \wedge \forall x \exists y\, [(x,y) \in \Gamma]$$

expresses the condition that Γ is the graph of a function from $\mathbb{X}$ to $\mathbb{X}$. The $*$-transform of this sentence is given by

$$\forall X \forall Y \forall Z\, [[(X,Y) \in {}^*\Gamma \wedge (X,Z) \in {}^*\Gamma] \to Y = Z] \wedge \forall X \exists Y\, [(X,Y) \in {}^*\Gamma].$$

This is a sentence over ${}^*\mathbb{X}$, meaning in particular that the variables X, Y, Z range over ${}^*\mathbb{X}$. This sentence expresses the condition that ${}^*\Gamma$ is the graph of a function from ${}^*\mathbb{X}$ to ${}^*\mathbb{X}$.

From Proposition 2.11 we know that these two sentences are equivalent, and this is no accident. This is an instance of the Transfer Principle, which we prove next. The Transfer Principle is a flexible and useful result which expresses nearly everything that one needs to know about nonstandard extensions. In particular, it gives precise meaning to the statement "*the nonstandard extension of $\mathbb{X}$ possesses all of the properties that $\mathbb{X}$ does.*"

3.3. Theorem. **[Transfer Principle]** *Let $\mathbb{X}$ be a non-empty set and consider a fixed nonstandard extension of $\mathbb{X}$.*

(a) Let $\varphi(x_1, \ldots, x_m)$ be a formula over $\mathbb{X}$ and let ${}^*\varphi(X_1, \ldots, X_m)$ be its $*$-transform. Suppose $B \subseteq \mathbb{X}^m$ is the set defined by $\varphi(x_1, \ldots, x_m)$:

$$B = \{(x_1, \ldots, x_m) \in \mathbb{X}^m \mid \varphi(x_1, \ldots, x_m) \text{ is true in } \mathbb{X}\}.$$

Then ${}^*B$ is the set defined by ${}^*\varphi(X_1, \ldots, X_m)$:

$${}^*B = \{(X_1, \ldots, X_m) \in ({}^*\mathbb{X})^m \mid {}^*\varphi(X_1, \ldots, X_m) \text{ is true in } {}^*\mathbb{X}\}.$$

(b) Let φ be any sentence over $\mathbb{X}$, and let ${}^*\varphi$ be its $*$-transform. Then

$$\varphi \text{ is true in } \mathbb{X} \Longleftrightarrow {}^*\varphi \text{ is true in } {}^*\mathbb{X}.$$

Proof. We prove (a) by induction on the syntactic complexity of formulas over $\mathbb{X}$. In other words we structure our proof so that it follows the same path as the inductive definition of the $*$-transform.

Before giving the inductive proof, we prove that (a) implies (b). Suppose φ is a sentence over $\mathbb{X}$ and ${}^*\varphi$ is its $*$-transform, a sentence over ${}^*\mathbb{X}$. Let A be the set defined by φ, so that ${}^*A$ is the set defined by ${}^*\varphi$, according to the statement above (which we are using in the case $n = 0$, where the formula treated does not have any variables that are free for substitution). Evidently A is either $\mathbb{X}^0$ or $\emptyset$, according to whether φ is true in $\mathbb{X}$ or not. Similarly ${}^*A$ is either $({}^*\mathbb{X})^0$ or $\emptyset$, according to whether or not ${}^*\varphi$ is true in ${}^*\mathbb{X}$ or not. Proposition 2.2 implies that either $A = \mathbb{X}^0$ and ${}^*A = ({}^*\mathbb{X})^0$ must both hold, or $A = \emptyset$ and ${}^*A = \emptyset$ must both hold. The equivalence of φ and ${}^*\varphi$ follows immediately.

Now we turn to the inductive proof of (a). For the basis step we must consider formulas $\varphi(x_1, \ldots, x_n)$ of the form $(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A$, where $\{\sigma(1), \ldots, \sigma(m)\} \subseteq \{1, \ldots, n\}$ and $A \subseteq \mathbb{X}^m$. Then ${}^*\varphi(X_1, \ldots, X_n)$ is $(X_{\sigma(1)}, \ldots, X_{\sigma(m)}) \in {}^*A$. Let B be the set of all $(x_1, \ldots, x_n) \in \mathbb{X}^n$ for which $(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A$ is true. Proposition 2.15 states that ${}^*B$ is the set of all $(X_1, \ldots, X_n) \in ({}^*\mathbb{X})^n$ for which $(X_{\sigma(1)}, \ldots, X_{\sigma(m)}) \in {}^*A$ is true. This is what we needed to prove.

More generally, in the basis step we must also take into account the possibility that one or more variables in $(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A$ are replaced by specific elements of $\mathbb{X}$. This is handled using Propositions 2.17 and 2.15.

The induction steps are handled using the conditions in Definition 2.1 directly. The logical connectives are handled using condition (E1). Existential quantifiers are handled using the strengthening of (E4) that is given in Proposition 2.16; the stronger form of (E4) is needed in case the existentially quantified variable is not the last variable in the given list. Finally, the duality between universal and existential quantifiers means that universal quantifiers can be handled as a combination of negations and existential quantifiers. □

We remark that the Transfer Principle exactly captures the content of the definition of nonstandard extension. That is, if the Transfer Principle holds and if the equality relation $=$ is given its usual interpretation in the nonstandard extension, then conditions (E1) – (E4) must be true. Proving this is an exercise in the use of logical formulas. Usually the Transfer Principle is explicitly included in the definition of nonstandard extension. We have delayed our discussion of the Transfer Principle in order to avoid heavy use of logical formulas at the beginning of the exposition and to permit introducing logical notation in a natural and convincing way.

To illustrate the usefulness of this result, let us treat some of the Exercises from Section 2. First consider Exercise 2.23. For ease of notation, assume $m = n = 1$. Let A, B, C and f be as given there. The set $f(B)$ is defined by the equivalence

$$x \in f(B) \Longleftrightarrow \exists y\,[y \in B \wedge f(y) = x].$$

Therefore, the Transfer Principle gives us that the equivalence

$$X \in {}^*(f(B)) \Longleftrightarrow \exists Y\,[Y \in {}^*B \wedge {}^*f(Y) = X]$$

holds in ${}^*\mathbb{X}$. But the formula on the right side of this equivalence defines $({}^*f)({}^*B)$, so we have the equality needed for part (a) of the Exercise.

For part (b) we use the equivalence

$$x \in f^{-1}(C) \Longleftrightarrow \exists y\,[y \in C \wedge f(x) = y]$$

and for part (c) we use the equivalence

$$(f|B)(x) = y \Longleftrightarrow [f(x) = y \wedge x \in B].$$

In both cases the Transfer Principle gives us immediately what is needed.

Now consider Exercise 2.24. For simplicity take $m = n = 2$. The function f is characterized by the equivalence

$$[f(u,v) = (x,y)] \Longleftrightarrow [f_1(u,v) = x \wedge f_2(u,v) = y]$$

where u, v, x, y are variables ranging over $\mathbb{X}$. The Transfer Principle yields that the equivalence

$$[({}^*f)(U,V) = (X,Y)] \Longleftrightarrow [({}^*f_1)(U,V) = X \wedge ({}^*f_2)(U,V) = Y]$$

holds in $^*\mathbb{X}$. This implies $^*f = (^*f_1, ^*f_2)$ as desired.

Next we treat Exercise 2.25. For ease of notation we consider only the case $m = n = 1$. Suppose $A \subseteq \mathbb{X}$ and $B \subseteq \mathbb{X}$, and let $f: A \to B$ be a function. For part (a), we note that f is injective if and only if the sentence

$$\forall x \forall y \forall z\, [[f(x) = z \wedge f(y) = z] \to x = y]$$

is true in $\mathbb{X}$. The $*$-transform of this sentence is

$$\forall X \forall Y \forall Z\, [[^*f(X) = Z \wedge {}^*f(Y) = Z] \to X = Y].$$

This sentence holds in $^*\mathbb{X}$ if and only if the function $^*f$ is injective. Therefore the Transfer Principle gives the desired result immediately. Similar arguments using other simple sentences will easily give parts (b) and (c).

Finally we treat Exercise 2.26. Each of the axioms for ordered fields can be expressed as a first-order sentence in which the quantifiers range over the underlying set of the field. For example, the statement that every non-zero element of $\mathbb{R}$ has a multiplicative inverse is expressed by the following sentence over $\mathbb{R}$:

$$\forall x \exists y\, [\neg x = 0 \to x \times y = 1].$$

The $*$-transform of this sentence is

$$\forall X \exists Y\, [\neg X = {}^*0 \to X \,{}^*\!\times\, Y = {}^*1].$$

By the Transfer Principle, this sentence is true in $^*\mathbb{R}$. Since ^{*0} is the additive identity and ^{*1} is the multiplicative identity in $^*\mathbb{R}$, as is shown in a similar way using other sentences over $\mathbb{R}$, it follows that every non-zero element of $^*\mathbb{R}$ has a multiplicative inverse. Similar arguments complete the Exercise.

4. Nonstandard Extensions of Multisets

In many parts of mathematics it is customary to encounter not just a single set, but several sets which are interacting in some way. For example, a vector space over the real field consists of the set $\mathbb{R}$ together with the underlying set $\mathbb{W}$ of the vector space. Among the objects which are included in this vector space setting is the operation of scalar multiplication, which is a function from $\mathbb{R} \times \mathbb{W}$ to $\mathbb{W}$. If $\mathbb{W}$ is a normed space, it is convenient to add the dual space $\mathbb{W}'$ as a third set. One operation which involves all three of these sets is the pairing $\langle w, f \rangle := f(w)$, considered as a function from $\mathbb{W} \times \mathbb{W}'$ into $\mathbb{R}$.

It is therefore natural to extend our concept of nonstandard extension to this kind of setting. Fortunately it is easy to do, requiring nothing more than a more elaborate notation. We lay out the details in this Section, but we omit proofs since they are so close to the ones which we gave in Sections 2 and 3. This material will be required in the next two Sections when we develop frameworks for introducing higher type objects into the foundations of nonstandard analysis.

Fix a non-empty index set I. The objects we consider here consist of families of (non-empty) sets indexed over I. We will refer to them as *multisets* or as *many sorted sets*, when the specific reference to I is omitted, and as *I-sets* when I needs to be mentioned.

4.1. Definition. *An* **I-set** $\mathbb{X}_I$ *is an indexed family* $(\mathbb{X}_i)_{i\in I}$ *of sets. A* **sort** *of the I-set $\mathbb{X}_I$ is one of the sets $\mathbb{X}_i$, where $i \in I$. We say that $\mathbb{X}_I$ is* **non-empty** *if $\mathbb{X}_i$ is non-empty for every $i \in I$.*

We now establish some notation for dealing with I-sets. We will let letters such as α, β, γ stand for finite sequences taken from I. Usually we will write α for the sequence $\alpha(1), \ldots, \alpha(m)$; m will be called the *length* of α and we will also denote the length by $|\alpha|$. Similarly we will normally understand that n is the length of the sequence β and p is the length of γ.

Given such a finite sequence α from I and given an I-set $\mathbb{X}_I$, we consider the Cartesian product of the sorts of $\mathbb{X}_I$ that are indexed by the coordinates of α; our notation for this Cartesian product is the following:

$$\mathbb{X}^\alpha = \mathbb{X}_{\alpha(1)} \times \cdots \times \mathbb{X}_{\alpha(m)}.$$

We consider the I-set $\mathbb{X}_I$ as equipped with all possible subsets of every Cartesian product $\mathbb{X}^\alpha$, where α ranges over all finite sequences from the index set I. In particular, this includes the graph of every function from one such Cartesian product $\mathbb{X}^\alpha$ to another Cartesian product $\mathbb{X}^\beta$. If $f: \mathbb{X}^\alpha \to \mathbb{X}^\beta$ is such a function, then its graph is a subset of $\mathbb{X}^\alpha \times \mathbb{X}^\beta$. Note that this product is also a Cartesian product of sorts. Indeed, $\mathbb{X}^\alpha \times \mathbb{X}^\beta = \mathbb{X}^\gamma$, where γ is the concatenation of α and β: $\gamma = \alpha(1), \ldots, \alpha(m), \beta(1), \ldots, \beta(n)$; $|\gamma| = m + n$.

Now we are ready to give the definition of *nonstandard extension* for I-sets. This results from a straightforward modification of the concept of nonstandard extension introduced Definition 2.1 for single sets. A nonstandard extension of an I-set $\mathbb{X}_I = (\mathbb{X}_i)_{i\in I}$ will be another non-empty I-set $({}^*\mathbb{X}_i)_{i\in I}$. For each finite sequence α from I, we will use the notation $({}^*\mathbb{X})^\alpha$ for the Cartesian product

$${}^*\mathbb{X}_{\alpha(1)} \times \cdots \times {}^*\mathbb{X}_{\alpha(m)}.$$

4.2. Definition. **[Nonstandard Extension of a Multiset]** *Let $\mathbb{X}_I$ be a non-empty I-set. A* **nonstandard extension of** $\mathbb{X}_I$ *is a mapping which assigns a set ${}^*A$ to each $A \subseteq \mathbb{X}^\alpha$ for all finite sequences α from I, such that*

${}^*\mathbb{X}_i$ is non-empty for all $i \in I$ and the following conditions are satisfied for all finite sequences α, β from I:

(M1) The mapping preserves Boolean operations on subsets of $\mathbb{X}^\alpha$: if $A \subseteq \mathbb{X}^\alpha$, then ${}^*A \subseteq ({}^*\mathbb{X})^\alpha$; if $A, B \subseteq \mathbb{X}^\alpha$, then ${}^*(A \cap B) = ({}^*A \cap {}^*B)$, ${}^*(A \cup B) = ({}^*A \cup {}^*B)$, and ${}^*(A \setminus B) = ({}^*A) \setminus ({}^*B)$.

(M2) The mapping preserves basic diagonals: suppose $1 \leq i < j \leq m = |\alpha|$ and suppose $\alpha(i) = \alpha(j)$; if

$$\Delta = \{(x_1, \ldots, x_m) \in \mathbb{X}^\alpha \mid x_i = x_j\}$$

then ${}^*\Delta = \{(x_1, \ldots, x_n) \in ({}^*\mathbb{X})^\alpha \mid x_i = x_j\}$.

(M3) The mapping preserves Cartesian products: if $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, then ${}^*(A \times B) = {}^*A \times {}^*B$.

(M4) The mapping preserves projections that omit the final coordinate: suppose α has length $n+1$ and let π be projection of $n+1$-tuples on the first n coordinates; if $A \subseteq \mathbb{X}^\alpha$, then ${}^*(\pi(A)) = \pi({}^*A)$.

For the rest of this Section we fix a nonstandard extension of $\mathbb{X}_I$, based on the non-empty I-set $({}^*\mathbb{X}_i)_{i \in I}$.

We now follow exactly the same path of Propositions and Exercises as in Sections 2 and 3. In order to be clear about what is intended, we give the results in a precisely worded form, modified appropriately for I-sets. It is routine to modify the arguments given in Sections 2 and 3 for this new setting, and we therefore omit all proofs here.

4.3. Proposition. *For each finite sequence α of elements of I, ${}^*(\mathbb{X}^\alpha) = ({}^*\mathbb{X})^\alpha$ and ${}^*\emptyset = \emptyset$.*

4.4. Proposition. *If $A \subseteq \mathbb{X}^\alpha$ is non-empty, then ${}^*A$ is also non-empty. Therefore, for any $A, B \subseteq \mathbb{X}^\alpha$, ${}^*A = {}^*B \iff A = B$.*

4.5. Proposition. *For all $A, B \subseteq \mathbb{X}^\alpha$, $A \subseteq B \iff {}^*A \subseteq {}^*B$.*

4.6. Proposition. *For each $i \in I$ and each $x \in \mathbb{X}_i$, ${}^*\{x\}$ has exactly one element.*

4.7. Notation. For each $i \in I$ and each $x \in \mathbb{X}_i$, we let ${}^*x$ denote the unique element of the set ${}^*\{x\}$. For each $x = (x_1, \ldots, x_m) \in \mathbb{X}^\alpha$ we let ${}^*x = ({}^*x_1, \ldots, {}^*x_m)$.

4.8. Definition. *An element of $({}^*\mathbb{X})^\alpha$ is called* **standard** *if it is of the form ${}^*x$ for some $x \in \mathbb{X}^\alpha$. It follows that an element of $({}^*\mathbb{X})^\alpha$ is standard if and only if all of its coordinates are standard elements of the appropriate sorts ${}^*\mathbb{X}_{\alpha(j)}$.*

4.9. Proposition. *For each $(x_1, \ldots, x_m) \in \mathbb{X}^\alpha$,*

$${}^*\{(x_1, \ldots, x_m)\} = \{({}^*x_1, \ldots {}^*x_m)\}.$$

4.10. Proposition. *For each $A \subseteq \mathbb{X}^\alpha$ and $(x_1, \dots, x_m) \in \mathbb{X}^\alpha$,*

$$(x_1, \dots, x_m) \in A \Longleftrightarrow ({}^*x_1, \dots, {}^*x_m) \in {}^*A.$$

4.11. Proposition. *Suppose $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, and let $f: A \to B$ be a function; take $\Gamma \subseteq \mathbb{X}^\alpha \times \mathbb{X}^\beta$ to be the graph of f. Then ${}^*\Gamma$ is the graph of a function from ${}^*A$ to ${}^*B$.*

4.12. Notation. Suppose $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, and let $f: A \to B$ be a function; take Γ to be the graph of f. We denote by ${}^*f$ the function from ${}^*A$ to ${}^*B$ whose graph is ${}^*\Gamma$.

4.13. Proposition. *If f is the identity function on $A \subseteq \mathbb{X}^\alpha$, then ${}^*f$ is the identity function on ${}^*A$.*

4.14. Proposition. *Suppose $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, and let $f: A \to B$ be a function. For all $(x_1, \dots, x_m) \in A$,*

$$({}^*f)({}^*x_1, \dots, {}^*x_m) = {}^*(f(x_1, \dots, x_m)).$$

4.15. Proposition. [Permuting and Identifying Variables] *Suppose α, β are finite sequences from I, with $m = |\alpha|$ and $n = |\beta|$. Suppose σ is any function from $\{1, \dots, m\}$ into $\{1, \dots, n\}$. Assume $\beta(\sigma(j)) = \alpha(j)$ for all $j = 1, \dots, m$. Given $A \subseteq \mathbb{X}^\alpha$ define*

$$B = \{(x_1, \dots, x_n) \in \mathbb{X}^\beta \mid (x_{\sigma(1)}, \dots, x_{\sigma(m)}) \in A\}.$$

Then

$${}^*B = \{(x_1, \dots, x_n) \in ({}^*\mathbb{X})^\beta \mid (x_{\sigma(1)}, \dots, x_{\sigma(m)}) \in {}^*A\}.$$

4.16. Proposition. *Condition (M4) in Definition 4.2 holds for all projections π.*

4.17. Proposition. *Let $A \subseteq \mathbb{X}^\gamma$ and $a = (a_1, \dots, a_m) \in \mathbb{X}^\alpha$, where γ is the sequence obtained by putting β after α. Define*

$$A(a) = \{(x_1, \dots, x_n) \in \mathbb{X}^\beta \mid (a_1, \dots, a_m, x_1, \dots, x_n) \in A\}$$

and similarly

$$({}^*A)({}^*a) = \{(x_1, \dots, x_n) \in ({}^*\mathbb{X})^\beta \mid ({}^*a_1, \dots, {}^*a_m, x_1, \dots, x_n) \in {}^*A\}$$

Then ${}^*(A(a)) = ({}^*A)({}^*a)$.

4.18. Proposition. *Suppose $A \subseteq \mathbb{X}^\alpha, B \subseteq \mathbb{X}^\beta$ and $C \subseteq \mathbb{X}^\gamma$; let $f: A \to B$ and $g: B \to C$ be functions. Then $^*(g \circ f) = (^*g) \circ (^*f)$.*

4.19. Exercise. Condition (M2) holds for all diagonal sets $\Delta \subseteq \mathbb{X}^\alpha$.

4.20. Exercise. If A is a finite subset of $\mathbb{X}^\alpha$, then $^*A = \{(^*x_1, \ldots, {}^*x_m) \mid (x_1, \ldots, x_m) \in A\}$. In particular, $^*A$ is finite and has the same cardinality as A, and all of its elements are standard.

4.21. Definition. *A nonstandard extension of $\mathbb{X}_I$ is called* **proper** *if for every $i \in I$ and every infinite subset A of $\mathbb{X}_i$, $^*A$ contains a nonstandard element.*

4.22. Exercise. Suppose our nonstandard extension is proper. Then, for any infinite set $A \subseteq \mathbb{X}^\alpha$, $^*A$ has a nonstandard element.

4.23. Exercise. Let $A \subseteq \mathbb{X}^\alpha$ and suppose $f: A \to \mathbb{X}^\beta$ is a function.
(a) If $B \subseteq A$, then $^*(f(B)) = (^*f)(^*B)$.
(b) If $C \subseteq \mathbb{X}^\beta$, then $^*(f^{-1}(C)) = (^*f)^{-1}(^*C)$.
(c) If $B \subseteq A$, then $^*(f|B) = (^*f)|(^*B)$.

4.24. Exercise. For $j = 1, \ldots, n$ let $f_j: \mathbb{X}^\alpha \to \mathbb{X}_{\beta(j)}$ be a function, and let $f = (f_1, \ldots, f_n): \mathbb{X}^\alpha \to \mathbb{X}^\beta$ be the function with $f_1, \ldots, f_n$ as its coordinates. Then $^*f = (^*f_1, \ldots, {}^*f_n)$.

4.25. Exercise. Suppose $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, and let $f: A \to B$ be a function.
(a) f is injective $\Longleftrightarrow$ $^*f$ is injective.
(b) f is surjective $\Longleftrightarrow$ $^*f$ is surjective.
(c) if f is a bijection and its inverse is g, then $^*g$ is the inverse of $^*f$.

Next we introduce logical formulas in order to state the Transfer Principle for nonstandard extensions of I-sets. Let $\mathbb{X}_I$ be a fixed I-set. For each $i \in I$ we will make use of variables that range over the sort $\mathbb{X}_i$; no other variables will be used in formulas over the I-set $\mathbb{X}_I$. If necessary, we will indicate that a variable ranges over the sort $\mathbb{X}_i$ by including i as a superscript in the name of the variable; thus x^i, y^i, x^i_j all denote variables that range over $\mathbb{X}_i$. However, we will usually omit such superscripts and let the context determine the sort over which a given variable ranges.

For each finite sequence α from I and for each set $A \subseteq \mathbb{X}^\alpha$ we will regard $(x_1, \ldots, x_m) \in A$ as a formula; $x_1, \ldots, x_m$ are variables with the property that for each $j = 1, \ldots, m$ the variable x_j ranges over the sort $\mathbb{X}_{\alpha(j)}$. As before, we do not require these variables to be distinct. If $f: A \to B$ is a function, where $A \subseteq \mathbb{X}^\alpha$ and $B \subseteq \mathbb{X}^\beta$, then we also take $f(x_1, \ldots, x_m) = (y_1, \ldots, y_n)$ to be a formula, where each variable x_i ranges over the sort $\mathbb{X}_{\alpha(i)}$ and each y_j ranges over $\mathbb{X}_{\beta(j)}$. Moreover, in these basic formulas we can replace some or all of the variables by specific elements of the sorts over which they range.

We construct more complicated formulas using quantifiers (with variables ranging over the sorts of $\mathbb{X}_I$) and the logical connectives $\neg, \vee, \wedge, \rightarrow$, and $\leftrightarrow$. We will refer to these logical formulas as *formulas over* $\mathbb{X}_I$. It is left to the reader to formulate a precise definition of this set of formulas similar to Definition 3.1.

Now we discuss how formulas over $\mathbb{X}_I$ can be used in connection with nonstandard extensions of $\mathbb{X}_I$. Consider a specific nonstandard extension of $\mathbb{X}_I$, based on the I-set $({}^*\mathbb{X}_i)_{i\in I}$. We will regard this nonstandard extension as fixed for the rest of this Section. Since $({}^*\mathbb{X}_i)_{i\in I}$ is also an I-set, we also have the class of formulas over $({}^*\mathbb{X}_i)_{i\in I}$. As before, we will see that there is an important connection between the formulas over $\mathbb{X}_I$ and (some of the) formulas over $({}^*\mathbb{X}_i)_{i\in I}$. We will again use the convention that lower case variables such as x^i range over specific sorts $\mathbb{X}_i$ and the corresponding upper case variables X^i range over the corresponding sort ${}^*\mathbb{X}_i$ of the nonstandard extension. As noted above, however, we will not always include the superscript and will let the natural context determine the role of the variables as much as possible.

4.26. Definition. **[$*$-Transform of a Formula Over $\mathbb{X}_I$]** *Consider a given nonstandard extension of the I-set $\mathbb{X}_I$. Let $\varphi(x_1, \ldots, x_n)$ be a formula over $\mathbb{X}_I$. The* **$*$-transform** *of $\varphi(x_1, \ldots, x_n)$ is a formula over $({}^*\mathbb{X}_i)_{i\in I}$, written ${}^*\varphi(X_1, \ldots, X_n)$, which is defined inductively by the following conditions:*

Basis cases:

(i) Let $\{\sigma(1), \ldots, \sigma(m)\} \subseteq \{1, \ldots, n\}$ and $A \subseteq \mathbb{X}^\alpha$, with $|\alpha| = m$; the $*$-transform of the basic formula

$$(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \in A$$

is the formula

$$(X_{\sigma(1)}, \ldots, X_{\sigma(m)}) \in {}^*A;$$

similarly, the $*$-transform of

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(k)}) = (x_{\sigma(k+1)}, \ldots, x_{\sigma(m)})$$

is

$${}^*f(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) = (X_{\sigma(k+1)}, \ldots, X_{\sigma(m)});$$

(ii) suppose $\varphi(x_1, \ldots, x_n, y_1, \ldots, y_p)$ is a basic formula (as treated in (i)), where the variables x_i range over sort $\mathbb{X}_{\beta(i)}$ for all $i = 1, \ldots, n$ and the variables y_j range over sort $\mathbb{X}_{\gamma(j)}$ for all $j = 1, \ldots, p$; if $a_j \in \mathbb{X}_{\gamma(j)}$

for all $j = 1, \ldots, p$, *then the* $*$*-transform of* $\varphi(x_1, \ldots, x_n, a_1, \ldots, a_p)$ *is* ${}^*\varphi(X_1, \ldots, X_n, {}^*a_1, \ldots, {}^*a_p)$;

Induction cases: Let φ *and* ψ *be formulas over* $\mathbb{X}$;

(iii) the $*$*-transform of the negation* $\neg\varphi$ *is* $\neg{}^*\varphi$;

(iv) the $*$*-transform of the disjunction* $\varphi \vee \psi$ *is* ${}^*\varphi \vee {}^*\psi$;

(v) the $*$*-transform of the conjunction* $\varphi \wedge \psi$ *is* ${}^*\varphi \wedge {}^*\psi$;

(vi) if x *is a variable ranging over a sort* $\mathbb{X}_i$, *then the* $*$*-transform of the quantified formula* $\exists x\, \varphi$ *is* $\exists X\, {}^*\varphi$, *in which* X *ranges over* ${}^*\mathbb{X}_i$;

(vii) if x *is a variable ranging over a sort* $\mathbb{X}_i$, *then the* $*$*-transform of the quantified formula* $\forall x\, \varphi$ *is* $\forall X\, {}^*\varphi$, *in which* X *ranges over* ${}^*\mathbb{X}_i$.

As before, this definition captures a simple idea: constructing the $*$-transform of a formula $\varphi(x_1, \ldots, x_n)$ over $\mathbb{X}_I$ requires the following steps:

(a) Find all of the sets $A \subseteq \mathbb{X}^\alpha$ that occur in $\varphi(x_1, \ldots, x_n)$ in basic formulas, and replace each such set by its counterpart ${}^*A$ over $({}^*\mathbb{X}_i)_{i \in I}$; similarly, replace each function $f\colon A \to B$ by ${}^*f$ and replace each element a of a sort $\mathbb{X}_i$ by ${}^*a$, and

(b) replace every variable x^i in $\varphi(x_1, \ldots, x_n)$, including the ones that are used with quantifiers, by a corresponding variable X^i which ranges over ${}^*\mathbb{X}_i$.

4.27. Theorem. **[Transfer Principle for** I**-sets]** *Let* $\mathbb{X}_I$ *be an* I*-set and consider a fixed nonstandard extension of* $\mathbb{X}_I$, *based on the* I*-set* $({}^*\mathbb{X}_i)_{i \in I}$.

(a) Let $\varphi(x_1, \ldots, x_m)$ *be a formula over* $\mathbb{X}_I$; *let* $\alpha(j)$ *be the index of the sort over which* x_j *ranges, for each* $j = 1, \ldots, m$; *let* ${}^*\varphi(X_1, \ldots, X_m)$ *be the* $*$*-transform of this formula. Suppose* $B \subseteq \mathbb{X}^\alpha$ *is the set defined by* $\varphi(x_1, \ldots, x_m)$:

$$B = \{(x_1, \ldots, x_m) \in \mathbb{X}^\alpha \mid \varphi(x_1, \ldots, x_m) \text{ is true in } \mathbb{X}_I\}.$$

Then ${}^*B$ *is the set defined by* ${}^*\varphi(X_1, \ldots, X_m)$:

$${}^*B = \{(X_1, \ldots, X_m) \in ({}^*\mathbb{X})^\alpha \mid \varphi(X_1, \ldots, X_m) \text{ is true in } ({}^*\mathbb{X}_i)_{i \in I}\}.$$

(b) Let φ *be any sentence over* $\mathbb{X}_I$, *and let* ${}^*\varphi$ *be its* $*$*-transform. Then*

$$\varphi \text{ is true in } \mathbb{X}_I \iff {}^*\varphi \text{ is true in } ({}^*\mathbb{X}_i)_{i \in I}.$$

4.28. Theorem. **[Existence of Nonstandard Extensions]** *Each nonempty* I*-set* $\mathbb{X}_I$ *has a proper nonstandard extension, in which the sets* $({}^*\mathbb{X}_i)_{i \in I}$ *may be taken to be ultrapowers of the sets* $(\mathbb{X}_i)_{i \in I}$ *with respect to a fixed countably incomplete ultrafilter.*

Proof. Let J be any infinite index set and let $\mathcal{U}$ be any countably incomplete ultrafilter on J. For each $i \in I$ let $^*\mathbb{X}_i$ be the ultrapower $\mathbb{X}_i^J/\mathcal{U}$. For each finite sequence α from I and each set $A \subseteq \mathbb{X}^\alpha$, define $^*A$ by

$$^*A = \{([\gamma_1], \ldots, [\gamma_m]) \mid \{j \in J \mid (\gamma_1(j), \ldots, \gamma_m(j)) \in A\} \in \mathcal{U}\}.$$

In this definition, for each $k = 1, \ldots m$ we let γ_k range over the set $\mathbb{X}^J_{\alpha(k)}$ of all functions from J into $\mathbb{X}_{\alpha(k)}$, so that $[\gamma_k]$ denotes a typical element of the ultrapower $\mathbb{X}^J_{\alpha(k)}/\mathcal{U}$ The proof that this defines a proper nonstandard extension of $\mathbb{X}_I$ is similar to the proof of Theorem 2.28, and we leave the details to the reader as an Exercise. □

5. Nonstandard Extensions of the Multiset $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$

In this Section we will use the methods developed in Section 4 to give an indication of how to introduce higher type objects into the framework of nonstandard analysis. First we consider a non-empty set $\mathbb{X}$ and the collection $\mathcal{P}(\mathbb{X})$ of all subsets of $\mathbb{X}$. We regard this as a multiset $(\mathbb{X}_0, \mathbb{X}_1)$ indexed over a set of two elements, with $\mathbb{X}_0 = \mathbb{X}$ and $\mathbb{X}_1 = \mathcal{P}(\mathbb{X})$.

Consider an arbitrary nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$, which we denote as $(^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))$. Let E be the restriction of the membership relation $\in$ to $\mathbb{X}$ and $\mathcal{P}(\mathbb{X})$:

$$E = \{(x, A) \in \mathbb{X} \times \mathcal{P}(\mathbb{X}) \mid x \in A\}.$$

As usual, write $\mathcal{P}(^*\mathbb{X})$ for the collection of all subsets of $^*\mathbb{X}$.

5.1. Remark. Without loss of generality we may assume the given nonstandard extension satisfies the following conditions:

(a) $\mathbb{X} \subseteq {}^*\mathbb{X}$ and $^*x = x$ for all $x \in \mathbb{X}$;

(b) $^*\mathcal{P}(\mathbb{X}) \subseteq \mathcal{P}(^*\mathbb{X})$ and $^*E$ is the restriction of the usual membership relation to $^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X})$:

$$^*E = \{(x, Y) \in {}^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X}) \mid x \in Y\}.$$

Justification. We show that every nonstandard extension of the multiset $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ is isomorphic to a nonstandard extension which satisfies (a) and (b). First we carry out a step like the one in the justification of Remark 2.10. As done there, let $\mathbb{Y}$ be a suitable set and $h: {}^*\mathbb{X} \to \mathbb{Y}$ a bijection, chosen so that $\mathbb{X} \subseteq \mathbb{Y}$ and $h(^*x) = x$ for all $x \in \mathbb{X}$. Moreover, given $Y \in {}^*\mathcal{P}(\mathbb{X})$, define

$$\Phi(Y) = \{h(x) \mid x \in {}^*\mathbb{X} \text{ and } (x, Y) \in {}^*E\},$$

which is a subset of $\mathbb{Y}$. It is easy to check that Φ is a 1-1 map on $^*\mathcal{P}(\mathbb{X})$. Finally, we define the new nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ to ensure that the pair (h, Φ) of bijections is an isomorphism of nonstandard extensions. That is, in the new nonstandard extension we map each set $A \subseteq \mathbb{X}^m \times \mathcal{P}(\mathbb{X})^n$ to the set

$$\{(h(x_1), \ldots, h(x_m), \Phi(Y_1), \ldots, \Phi(Y_n)) \mid (x_1, \ldots, x_m, Y_1, \ldots, Y_n) \in {}^*A\}.$$

It is routine to check that this new mapping is a nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ and that it satisfies conditions (a) and (b). $\square$

In this kind of situation it is often convenient to suppress the explicit use of the maps h and Φ; rather we may follow a customary abuse of notation and "identify" $^*x$ with x for each $x \in \mathbb{X}$. With this understanding, the definition of $\Phi(Y)$ for each $Y \in {}^*\mathcal{P}(\mathbb{X})$ becomes

$$\Phi(Y) = \{x \in {}^*\mathbb{X} \mid (x, Y) \in {}^*E\}.$$

We may then identify Y with the subset $\Phi(Y)$ of $^*\mathbb{X}$ defined in this way. This is particularly convenient when (as later in this Section) the nonstandard extension has been constructed in an explicit way, such as we do here using the ultrapower construction.

For the rest of this Section we assume that we have a nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ which satisfies (a) and (b) in Remark 5.1.

Condition (b) in Remark 5.1 ensures that the elements of $^*\mathcal{P}(\mathbb{X})$ are ordinary subsets of $^*\mathbb{X}$, and that the $*$-transform of any formula φ over $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ is well behaved with respect to the membership relation. Suppose x is a variable ranging over $\mathbb{X}$ and y is a variable ranging over $\mathcal{P}(\mathbb{X})$, and suppose $x \in y$ occurs in φ. Recall that $x \in y$ is equivalent to the basic formula $(x, y) \in E$; the process of forming the $*$-transform will replace this basic formula by $(X, Y) \in {}^*E$, which is equivalent to $X \in Y$ according to condition (b). In other words, in forming the $*$-transform we may simply replace basic formulas of the form $x \in y$ by $X \in Y$, when the nonstandard extension satisfies (b).

Consider a subset A of $\mathbb{X}$. It can be considered either as a subset of $\mathbb{X}$ or as an element of $\mathcal{P}(\mathbb{X})$. Accordingly, there are two possible interpretations of the expression $^*A$: let us temporarily write $*(A)$ for the set which the given nonstandard extension assigns to A, and reserve the notation $^*A$ (as in paragraph 4.7) to denote the unique element of the set $*(\{A\})$ which this nonstandard extension assigns to $\{A\}$. Both of these are subsets of $^*\mathbb{X}$. Fortunately they are equal when we adopt the normalization described in Remark 5.1, as we now prove.

5.2. Proposition. *For each $A \subseteq \mathbb{X}$, we have $*(A) = {}^*A$.*

Proof. Fix $A \subseteq \mathbb{X}$ and let E be the restriction of the membership relation as above. Evidently we have that the sentence

$$\forall x \in \mathbb{X}\,[(x, A) \in E \leftrightarrow x \in A]$$

is true in our basic structure $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$. By the Transfer Principle (Theorem 4.27), we conclude that

$$\forall X \in {}^*\mathbb{X}\,[(X, {}^*A) \in *(E) \leftrightarrow X \in *(A)]$$

holds in the nonstandard extension. Using condition (b) in Remark 5.1, we see that

$$\forall X \in {}^*\mathbb{X}\,[X \in {}^*A \leftrightarrow X \in *(A)]$$

holds in the nonstandard extension. This proves $*(A) = {}^*A$ since both are subsets of ${}^*\mathbb{X}$. □

Next we introduce one of the key distinctions in nonstandard analysis: the distinction between *internal* and *external* subsets of ${}^*\mathbb{X}$.

5.3. Definition. [Internal Subset of ${}^*\mathbb{X}$] *A subset A of ${}^*\mathbb{X}$ is* **internal** *if it is an element of ${}^*\mathcal{P}(\mathbb{X})$; A is* **external** *if it is not internal.*

We note that it is only internal subsets of ${}^*\mathbb{X}$ that are referred to within the $*$-transform of a logical formula over $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$. That is, all variables in such a formula either range over ${}^*\mathbb{X}$ itself, or they range over ${}^*\mathcal{P}(\mathbb{X})$. If φ is a logical sentence over $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ and ${}^*\varphi$ is its $*$-transform, it follows that we get the same truth value for ${}^*\varphi$ in the nonstandard extension $({}^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))$ as in the multiset $({}^*\mathbb{X}, \mathcal{P}({}^*\mathbb{X}))$. The same is true for logical formulas over $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ into which we have substituted elements of ${}^*\mathbb{X}$ for all the free first order variables and *internal* subsets of ${}^*\mathbb{X}$ for all the free set variables. (This need *not* be true if we substitute *external* subsets of ${}^*\mathbb{X}$ for one or more of the free set variables in ${}^*\varphi$ and interpret it in $({}^*\mathbb{X}, \mathcal{P}({}^*\mathbb{X}))$.)

The next result is an easy consequence of the Transfer Principle, but it is a key tool for handling internal sets.

5.4. Theorem. [Internal Definition Principle] *Let $\varphi(x, x_1, \ldots, x_m, y_1, \ldots, y_n)$ be a formula over the multiset $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$. Suppose the variables x and x_j range over $\mathbb{X}$ for each j and the variable y_k ranges over $\mathcal{P}(\mathbb{X})$ for each k. Let $a_1, \ldots, a_m \in {}^*\mathbb{X}$ and let $A_1, \ldots, A_n$ be internal subsets of ${}^*\mathbb{X}$. Let B be the subset of ${}^*\mathbb{X}$ defined by ${}^*\varphi(X, a_1, \ldots, a_m, A_1, \ldots, A_n)$:*

$$B = \{X \in {}^*\mathbb{X} \mid {}^*\varphi(X, a_1, \ldots, a_m, A_1, \ldots, A_n) \text{ is true in } ({}^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))\}.$$

Then B is internal.

Proof. Apply the Transfer Principle (Theorem 4.27) to the sentence

$$\forall x_1 \ldots \forall x_m \forall y_1 \ldots \forall y_n \exists z \forall x\, [x \in z \leftrightarrow \varphi(x, x_1, \ldots, x_m, y_1, \ldots, y_n)],$$

which is true in the $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$; therefore the sentence

$$\forall X_1 \ldots \forall X_m \forall Y_1 \ldots \forall Y_n \exists Z \forall X\, [X \in Z \leftrightarrow {}^*\varphi(X, X_1, \ldots, X_m, Y_1, \ldots, Y_n)]$$

is true in the nonstandard extension $({}^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))$. Note that the variables X and $X_1, \ldots, X_m$ range over ${}^*\mathbb{X}$ and $Y_1, \ldots, Y_n$ are restricted to range over ${}^*\mathcal{P}(X)$. Substituting a_j for X_j for each $j = 1, \ldots, m$ and A_k for Y_k for each $k = 1, \ldots, n$ gives the desired result. Note that the substitution of A_k for Y_k is permitted only because A_k is assumed to be an internal subset of ${}^*\mathbb{X}$. This is a key aspect of the Internal Definition Principle. □

5.5. Exercise. Let A, B be internal subsets of ${}^*\mathbb{X}$.

(i) Every Boolean combination of A, B is internal.

(ii) If $f \colon \mathbb{X} \to \mathbb{X}$ is any function, then the sets $({}^*f)(A)$ and $({}^*f)^{-1}(A)$ are internal.

(iii) Every standard element of ${}^*\mathcal{P}(\mathbb{X})$ is an internal subset of ${}^*\mathbb{X}$.

Now we return to the setting in which $\mathbb{X} = \mathbb{R}$. Consider the linear ordering $<$ as a subset of $\mathbb{R}^2$ and the graphs Γ_+ and $\Gamma_\times$ of the functions $+$ and $\times$ as subsets of $\mathbb{R}^3$. By Proposition 4.11, ${}^*\Gamma_+$ and ${}^*\Gamma_\times$ are subsets of $({}^*\mathbb{R})^3$ which are graphs of functions from $({}^*\mathbb{R})^2$ to ${}^*\mathbb{R}$. For ease of notation, we will follow the customary practice of dropping the $*$ and denoting these functions as $+$ and $\times$. Using condition (a) in Remark 5.1 and Proposition 4.14, it follows that $+$ and $\times$ on $({}^*\mathbb{R})^2$ are extensions of the original functions $+$ and $\times$ on $\mathbb{R}^2$. Similarly, the relation $<$ on ${}^*\mathbb{R}$ is an extension of the given linear ordering $<$ on $\mathbb{R}$. Also, for any $A \subseteq \mathbb{R}$, A is easily seen to be a subset of ${}^*A$ using similar reasoning. The following Exercise can be fairly easily solved using the ideas above, especially including the Transfer Principle (Theorem 4.27) and the Internal Definition Principle (Theorem 5.4).

5.6. Exercise. (i) $({}^*\mathbb{R}, +, \times, <)$ is an ordered field extension of the ordered field $(\mathbb{R}, +, \times, <)$.

(ii) $\mathbb{N}$ is an initial segment of ${}^*\mathbb{N}$ and the elements of ${}^*\mathbb{N} \setminus \mathbb{N}$ are infinite numbers in ${}^*\mathbb{R}$.

(iii) For every positive $r \in {}^*\mathbb{R}$ there exists a unique $N \in {}^*\mathbb{N}$ such that $N \leq r < N + 1$.

(iv) If A is a non-empty internal subset of ${}^*\mathbb{R}$ which is bounded above in ${}^*\mathbb{R}$, then A has a least upper bound in ${}^*\mathbb{R}$; this need not be true if A is external.

(v) For each $N \in {}^*\mathbb{N}$, let $\{0, 1, \ldots, N\}$ denote the set of $M \in {}^*\mathbb{N}$ which satisfy $0 \leq M \leq N$; the set $\{0, 1, \ldots, N\}$ is internal.

(vi) For each $r < s$ in $^*\mathbb{R}$ let $[r, s]$ denote the set of all $t \in {}^*\mathbb{R}$ such that $r \leq t \leq s$; the set $[r, s]$ is internal.

(vii) The set $^*\mathbb{N} \setminus \mathbb{N}$ is **not** an internal subset of $^*\mathbb{R}$.

(viii) The set of infinitesimal elements of $^*\mathbb{R}$ is **not** an internal subset of $^*\mathbb{R}$.

(ix) The set of finite elements of $^*\mathbb{R}$ is **not** an internal subset of $^*\mathbb{R}$.

(x) (Overspill Principle) Let A be an internal subset of $^*\mathbb{R}$; if A contains arbitrarily large finite numbers, then it also contains an infinite positive number.

(xi) (Underspill Principle) Let A be an internal subset of $^*\mathbb{R}$; if A contains arbitrarily small positive infinite numbers, then it also contains a positive finite number.

We briefly indicate an expansion of this approach which allows treatment of internal functions between internal subsets of $^*\mathbb{X}$. To introduce such functions we consider the multiset $(\mathbb{X}, \mathcal{P}(\mathbb{X}), \mathcal{P}(\mathbb{X} \times \mathbb{X}))$; if $A, B \subseteq \mathbb{X}$ and $f: A \to B$ is a function, then we regard f as an element of this multiset by considering its graph Γ_f, which is an element of the third sort $\mathcal{P}(\mathbb{X} \times \mathbb{X})$).

Consider an arbitrary nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}), \mathcal{P}(\mathbb{X} \times \mathbb{X}))$, which we denote as $(^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}), {}^*\mathcal{P}(\mathbb{X} \times \mathbb{X}))$. Expanding on the discussion in Remark 5.1, we may pass to an isomorphic nonstandard extension which satisfies the following three conditions. We use the notation

$$E_1 = \{(x, A) \in \mathbb{X} \times \mathcal{P}(\mathbb{X}) \mid x \in A\};$$

$$E_2 = \{(x, y, A) \in \mathbb{X} \times \mathbb{X} \times \mathcal{P}(\mathbb{X} \times \mathbb{X}) \mid (x, y) \in A\}.$$

(a) $\mathbb{X} \subseteq {}^*\mathbb{X}$ and $^*x = x$ for all $x \in \mathbb{X}$;

(b) $^*\mathcal{P}(\mathbb{X}) \subseteq \mathcal{P}(^*\mathbb{X})$ and $^*E_1$ is the restriction of the usual membership relation to $^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X})$:

$$^*E_1 = \{(x, Y) \in {}^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X}) \mid x \in Y\}.$$

(c) $^*\mathcal{P}(\mathbb{X} \times \mathbb{X}) \subseteq \mathcal{P}(^*\mathbb{X} \times {}^*\mathbb{X})$ and $^*E_2$ is the restriction of the usual ordered pairs membership relation to $^*\mathbb{X} \times {}^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X} \times \mathbb{X})$:

$$^*E_2 = \{(x, y, Y) \in {}^*\mathbb{X} \times {}^*\mathbb{X} \times {}^*\mathcal{P}(\mathbb{X} \times \mathbb{X}) \mid (x, y) \in Y\}.$$

Internal subsets of $^*\mathbb{X}$ are handled as was done earlier in this Section. Similarly, we call a subset of $^*\mathbb{X} \times {}^*\mathbb{X}$ *internal* if it is an element of $^*\mathcal{P}(\mathbb{X} \times \mathbb{X})$. If A, B are subsets of $^*\mathbb{X}$ and $f: A \to B$ is any function, we say that f is an *internal* function if its graph Γ_f is an internal subset of $^*\mathbb{X} \times {}^*\mathbb{X}$.

5.7. Exercise. Consider the multiset $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R} \times \mathbb{R})$ and a nonstandard extension of it $(^*\mathbb{R}, {}^*\mathcal{P}(\mathbb{R}), {}^*\mathcal{P}(\mathbb{R} \times \mathbb{R}))$, which has been normalized so that conditions (a), (b), and (c) above are satisfied. We adopt the notation described just before Exercise 5.6.

(i) If f is an internal function between subsets of ${}^*\mathbb{R}$, then the domain and range of f are internal subsets of ${}^*\mathbb{R}$.

(ii) If f is an internal function between subsets of ${}^*\mathbb{R}$ and A is an internal set contained in the domain of f, then the restriction of f to A is internal.

(iii) If f, g are internal functions between subsets of ${}^*\mathbb{R}$ and the domain of g contains the range of f, then the composition $g \circ f$ is internal.

(iv) Suppose $f: {}^*\mathbb{N} \to {}^*\mathbb{R}$ is an internal function; there exists a unique internal function $F: {}^*\mathbb{N} \to {}^*\mathbb{R}$ such that $F(0) = f(0)$ and for all $n \in {}^*\mathbb{N}$, $F(n+1) = F(n) + f(n+1)$.

5.8. Remark. Consider the setting of part (iv) in the previous Exercise. The function F can be viewed as the result of summing the values of f over initial segments of ${}^*\mathbb{N}$, and this is a useful idea in many applications of nonstandard analysis. For obvious reasons, it is customary to denote $F(n)$ for all $n \in {}^*\mathbb{N}$ (including nonstandard n) by

$$\sum_{i=0}^{n} f(i).$$

Such hyperfinite sums appear, for example, in the nonstandard approach to measure and integration.

5.9. Exercise. Suppose $f: {}^*\mathbb{N} \to {}^*\mathbb{R}$ and $g: {}^*\mathbb{N} \to {}^*\mathbb{R}$ are internal functions, and $c \in {}^*\mathbb{R}$. Consider the notation introduced in the previous Remark.

(i) For all $n \in {}^*\mathbb{N}$, $\sum_{i=0}^{n}(f(i) + g(i)) = \sum_{i=0}^{n} f(i) + \sum_{i=0}^{n} g(i)$.

(ii) For all $n \in {}^*\mathbb{N}$, $\sum_{i=0}^{n} c \cdot f(i) = c \cdot \sum_{i=0}^{n} f(i)$.

5.10. Exercise. Consider the multiset which has three sorts, $\mathbb{X}, \mathcal{P}(\mathbb{X})$, and $\mathcal{P}(\mathcal{P}(\mathbb{X}))$, and develop the ideas of this Section in that context. The nonstandard extension should be modified so that not only is ${}^*\mathcal{P}(\mathbb{X})$ a subset of $\mathcal{P}({}^*\mathbb{X})$, but also ${}^*\mathcal{P}(\mathcal{P}(\mathbb{X}))$ is a subset of $\mathcal{P}(\mathcal{P}({}^*\mathbb{X}))$, and so that the modified nonstandard extension preserves the restriction of the membership relation between $\mathbb{X}$ and $\mathcal{P}(\mathbb{X})$, as well as between $\mathcal{P}(\mathbb{X})$ and $\mathcal{P}(\mathcal{P}(\mathbb{X}))$. In this way all elements of ${}^*\mathcal{P}(\mathbb{X})$ and ${}^*\mathcal{P}(\mathcal{P}(\mathbb{X}))$ can be handled as sets in a canonical way.

The setting described in Exercise 5.10 provides a framework in which nonstandard methods can be applied to subsets of $\mathbb{X}$ as well as to collections of subsets of $\mathbb{X}$. This would be a suitable framework for applying nonstandard methods to the study of topologies on $\mathbb{X}$, for example, with the collection of open sets being an element of the third sort. This would also allow us to consider "internal topologies" on ${}^*\mathbb{X}$. These are internal collections $\mathcal{T}$

of (necessarily internal) subsets of $^*\mathbb{X}$ which satisfy the $*$-transform of the formula expressing the familiar defining conditions satisfied by topologies.

If we take the point of view of this Section to its natural limit, we get the type theoretic formulation of nonstandard analysis that Abraham Robinson used in his book [14]. Although this framework did not catch on at the time, that is likely due to the heavily formal presentation in [14] rather than to any essential disadvantages of this point of view. For an example of the use of such a framework for an important application of nonstandard methods, see the nonstandard proof due to van den Dries and Wilkie of Gromov's Theorem about groups of polynomial growth. (See [7], pages 356–363; they present their nonstandard extension explicitly as an ultrapower.)

We conclude this Section by discussing the nature of internal sets in the ultrapower nonstandard extensions that are constructed in the proofs of Theorems 2.28 and 4.28. We restrict our attention to nonstandard extensions of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$, where $\mathbb{X}$ is any nonempty set. Let J be any infinite index set and let $\mathcal{U}$ be an ultrafilter on J. Let $(^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))$ be the ultrapower nonstandard extension of $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ that is constructed in the proof of Theorem 4.28.

Let a be any element of $\mathbb{X}$. We follow the customary practice of identifying a with the corresponding standard element $^*a$ of $^*\mathbb{X}$. As discussed in Exercise 2.29, this means we are identifying a with the equivalence class $[\alpha]$, where α is the constant function defined by $\alpha(j) = a$ for all $j \in J$.

Consider the effect of the normalization that is discussed in Remark 5.1. This mainly hinges on the behavior of the mapping Φ that is defined there. Let Y be an arbitrary element of $^*\mathcal{P}(\mathbb{X})$ and consider

$$\Phi(Y) = \{x \in {}^*\mathbb{X} \mid (x, Y) \in {}^*E\}.$$

In this setting Y is an equivalence class $[F]$ where F is a function from J into $\mathcal{P}(\mathbb{X})$; in other words, F is an indexed family of subsets of $\mathbb{X}$. Taking into account the definition of $^*E$ leads to the equation

$$\Phi([F]) = \{[\alpha] \mid \{j \in J \mid \alpha(j) \in F(j)\} \in \mathcal{U}\}.$$

Therefore, a subset A of $^*\mathbb{X}$ is internal (in the normalized version of the ultrapower nonstandard extension $(^*\mathbb{X}, {}^*\mathcal{P}(\mathbb{X}))$ if and only if there is an indexed family $F\colon J \to \mathcal{P}(\mathbb{X})$ of subsets of $\mathbb{X}$ such that for all $[\alpha] \in {}^*\mathbb{X}$:

$$[\alpha] \in A \iff \{j \in J \mid \alpha(j) \in F(j)\} \in \mathcal{U}.$$

5.11. Exercise. Consider the ultrapower nonstandard extension discussed in the preceding paragraphs.

(a) Let $F\colon J \to \mathcal{P}(\mathbb{X})$ be an indexed family of subsets of $\mathbb{X}$ with the property that $F(j)$ is nonempty for each $j \in J$. The internal subset of $^*\mathbb{X}$

determined as above by F can be identified with the ultraproduct $\Pi_{\mathcal{U}}(F(j) \mid j \in J)$.

(b) Every non-empty internal subset of ${}^*\mathbb{X}$ can be represented as described in (a).

6. Superstructures

In this Section we will explain a setting for nonstandard analysis which was introduced in [13] by Robinson and Zakon. This framework gives a convenient way to apply nonstandard methods to essentially any part of mathematics. Much of the research literature of nonstandard analysis is expressed in terms of the framework that is explained here. The essential ideas in this Section are just an easy elaboration of what was done in the previous Section.

In order to use nonstandard extensions effectively, they must be applicable to mathematical systems which contain objects of higher type, such as spaces of functions, collections of sets (such as filters), systems of open sets in a topological space, and the like. Such objects occur in essentially every part of mathematics, and our framework must accommodate them in a smooth way. Experience has shown that a convenient way to accomplish this is to introduce the *superstructure* based on a given set S of elementary mathematical objects. (In most applications it is natural to take $S = \mathbb{R}$ or $S = \mathbb{N}$. We will always assume that S contains $\mathbb{N}$ as a subset. The choice of S is otherwise somewhat arbitrary and depends on the mathematical problems that are being considered.) The elements of this superstructure are precisely the mathematical objects that can be obtained from S in a finite number of steps, where in each step we form all sets of the previously constructed objects and add each of these sets as a new object in its own right.

If T is a set, we write $\mathcal{P}(T)$ for the *power set* of T, which is the collection of all subsets of T.

6.1. Definition. [Superstructure] *Fix a set S such that $\mathbb{N} \subseteq S$. The* **superstructure based on** *S is the family of sets $(\mathbb{V}_k(S))_{k\in\mathbb{N}}$ defined by the following induction on k:*

$$\mathbb{V}_0(S) = S; \qquad \mathbb{V}_{k+1}(S) = \mathbb{V}_k(S) \cup \mathcal{P}(\mathbb{V}_k(S)).$$

This system is an $\mathbb{N}$-set in the terminology of Section 4; we denote it by $\mathbb{V}(S)$. An element of the union $\bigcup_{k=0}^{\infty} \mathbb{V}_k(S)$ is called an **object in** $\mathbb{V}(S)$. *The* **rank** *of an object a in $\mathbb{V}(S)$ is the smallest k for which $a \in \mathbb{V}_k(S)$.*

Note that

$$S = \mathbb{V}_0(S) \subseteq \mathbb{V}_1(S) \subseteq \mathbb{V}_2(S) \subseteq \dots$$

and hence also

$$\mathbb{V}_j(S) \in \mathbb{V}_k(S) \quad \text{whenever } j < k.$$

When we interpret the membership relation $\in$ in $\mathbb{V}(S)$, we treat members of S as having no elements. Note that the objects of rank ≥ 1 in $\mathbb{V}(S)$ are precisely the sets in $\mathbb{V}(S)$, and the basic objects (elements of S) are the objects of rank 0. The empty set $\emptyset$ has rank 1. If b is an object in $\mathbb{V}(S)$ of rank ≥ 1 and a is an element of b, then a is also an object in $\mathbb{V}(S)$. Note also that when a, b are objects in $\mathbb{V}(S)$, $a \in b$ always implies that the rank of a is strictly less than the rank of b.

We assume that the reader is familiar with a small amount of naive set theory. In particular, we form basic pairs within $\mathbb{V}(S)$ using the familiar definition $\langle x, y\rangle = \{\{x, y\}, \{x\}\}$. Note that the rank of $\langle x, y\rangle$ is $r + 2$ where r is the larger of the ranks of x, y. For each $n \geq 2$ we define the ordered n-tuple $(x_1, \ldots, x_n)$ to be the set $\{\langle i, x_i\rangle \mid i = 1, \ldots, m\}$. Recall that we require $\mathbb{N}$ to be a subset of the basic set S of the superstructure $\mathbb{V}(S)$; therefore $(x_1, \ldots, x_n)$ is an object in $\mathbb{V}(S)$ whenever $x_1, \ldots, x_n$ are objects in $\mathbb{V}(S)$. Moreover, the rank of $(x_1, \ldots, x_n)$ is $r + 3$ if r is the maximum of the ranks of $x_1, \ldots, x_n$. If A is a set of rank k in $\mathbb{V}(S)$ and $n \geq 2$, then A^n, taken to be the set of ordered n-tuples of elements of A, will be a set in $\mathbb{V}(S)$ and its rank will be $k + 3$. Note that this is independent of n. Similar remarks can be made about mixed Cartesian products.

We now want to develop a suitable concept of *nonstandard extension* for superstructures, based on regarding a superstructure as an $\mathbb{N}$-set and using the tools from Section 4. However, as in Section 5 some additional considerations arise from the fact that the sorts $\mathbb{V}_k(S)$ are not just independent sets but rather have a high degree of interrelation. We will work with nonstandard extensions that have been normalized in a way similar to that discussed in Remark 5.1. In the superstructure setting it is convenient to change perspective slightly and to work with rank preserving embeddings between superstructures.

Suppose $\mathbb{V}(S)$ and $\mathbb{V}(T)$ are superstructures and $F\colon \mathbb{V}(S) \to \mathbb{V}(T)$ is any rank preserving function. Let α be a finite sequence from $\mathbb{N}$ and suppose $A \subseteq \mathbb{V}(S)^\alpha = \mathbb{V}_{\alpha(1)}(S) \times \ldots \times \mathbb{V}_{\alpha(m)}(S)$. For large enough $k \in \mathbb{N}$ the set A is a set in $\mathbb{V}_k(S)$ so $F(A)$ is a well defined set in $\mathbb{V}(T)$. (Here $F(A)$ is the value of the function F at A, not to be confused with $\{F(a) \mid a \in A\}$.) It turns out that a good approach to defining nonstandard extensions of superstructures is to define $^*A$ to be $F(A)$ for every such A.

6.2. Definition. [Nonstandard Extension of a Superstructure] *Let $\mathbb{V}(S), \mathbb{V}(T)$ be superstructures and let $F\colon \mathbb{V}(S) \to \mathbb{V}(T)$ be a rank preserving function. Consider the mapping defined by letting $^*A = F(A)$ for each $A \subseteq \mathbb{V}(S)^\alpha$, where α is any finite sequence from $\mathbb{N}$. We say F is a* **nonstandard**

extension of $\mathbb{V}(S)$ (as a superstructure) *if* $T = {}^*S$ *and the following conditions are satisfied: (Note that* ${}^*\mathbb{V}_k(S) \subseteq \mathbb{V}_k({}^*S)$ *for each* $k \in \mathbb{N}$*, because* $T = {}^*S$ *and* F *is rank preserving.)*

(S1) This mapping is a nonstandard extension of $\mathbb{V}(S)$ *(considered as the multiset* $(\mathbb{V}_k(S))_{k\in\mathbb{N}}$ *indexed by* $\mathbb{N}$*) in the sense of Definition 4.2.*

(S2) If $a \in S$*, then* ${}^*a = a$*; in particular* $S \subseteq {}^*S$*.*

(S3) This mapping preserves the membership relation:

For each $k \in \mathbb{N}$ *let* E_k *be the usual membership relation restricted to* $\mathbb{V}_k(S)$*,*

$$E_k = \{(a,b) \in \mathbb{V}_k(S)^2 \mid a \in b\};$$

we require that ${}^*E_k$ *is the restriction of the usual membership relation to* ${}^*\mathbb{V}_k(S)$*,*

$${}^*E_k = \{(x,y) \in ({}^*\mathbb{V}_k(S))^2 \mid x \in y\}.$$

(S4) The nonstandard universe is transitive:

For each $k \in \mathbb{N}$*, if* $a \in {}^*\mathbb{V}_{k+1}(S)$ *and* $b \in a$*, then* $b \in {}^*\mathbb{V}_k(S)$*.*

The extra conditions (S2) – (S4) have a normalizing effect on the nonstandard extension and make it easier to work with. Moreover, if $F\colon \mathbb{V}(S) \to \mathbb{V}(T)$ satisfies only condition (S1), then the extra conditions (S2) – (S4) can be achieved by a series of simple modifications to F which are like those used in the justification of Remark 5.1.

For the remainder of this Section assume that $F\colon \mathbb{V}(S) \to \mathbb{V}({}^*S)$ satisfies the conditions in Definition 6.2. We will explore a few consequences of the Definition and then proceed to introduce some of the main ideas through which nonstandard extensions of superstructures are applied.

We will use the notation ${}^*\mathbb{V}(S)$ for the $\mathbb{N}$-set $({}^*\mathbb{V}_k(S))_{k\in\mathbb{N}}$. Note that ${}^*\mathbb{V}(S)$ is contained in the superstructure $\mathbb{V}({}^*S)$. This means that every object in ${}^*\mathbb{V}(S)$ is a mathematical object of the usual kind, and that it lies at some finite level of higher type objects over the set ${}^*S$. When working with the sets in ${}^*\mathbb{V}(S)$ from the *outside* so to speak, this means that we can regard them as ordinary mathematical objects, to which all of the usual mathematical concepts can be applied. In particular, we can speak of the cardinality (finite, infinite, countable, uncountable, etc) of each set A in ${}^*\mathbb{V}(S)$. This plays a useful role in many applications of nonstandard analysis. When needed for clarity, we will refer to the *external* cardinality of A when we are making use of this point of view.

The set theoretic nature of superstructures means we need to be careful when interpreting the definition of nonstandard extension and when applying the Transfer Principle. An element of $\mathbb{V}_k(S) \setminus V_0(S)$ is simultaneously (1) an element of the sort $V_k(S)$ and (2) a subset of possibly many different Cartesian products of sorts $\mathbb{V}(S)^\alpha$. Under each of these interpretations there is a separate definition of the expression ${}^*a$. In all of the cases under

(2), $^*a$ is taken to be the unique element $F(a)$ by definition. In case (1), we know that $^*a$ is the unique element of $F(\{a\})$, as defined in paragraph 4.7 and justified by Proposition 4.6. But in fact there is no ambiguity, as follows from condition (S3) and a proof like that given for Proposition 5.2. All of these interpretations of the notation $^*a$ refer to the same object.

Suppose x, y are variables ranging over $\mathbb{V}_k(S)$ of the kind that occur in applications of the Transfer Principle. In this context it is permissible to use $x \in y$ as a basic formula, since it is equivalent to the basic formula $(x, y) \in E_k$. The $*$-transform of $(x, y) \in E_k$ is defined to be $(X, Y) \in {}^*E_k$, where X, Y are variables ranging over ${}^*\mathbb{V}_k(S)$. However, condition (S3) implies that this is equivalent to $X \in Y$. In other words, if we use $x \in y$ in a formula φ over $\mathbb{V}(S)$, and we want to construct the $*$-transform of φ in order to apply the Transfer Principle, then we simply modify the formula $x \in y$ to the formula $X \in Y$. Precisely the same thing is true when the variables x, y do not necessarily have the same rank.

When applying the Transfer Principle to nonstandard extensions of superstructures, it is common to use *bounded* quantifiers. These are relativized quantifiers of the form $\forall x \in a$ and $\exists x \in a$, where a is a set in $\mathbb{V}(S)$. Here we can take x to be a variable ranging over $\mathbb{V}_k(S)$, where k is chosen to be at least as large as the ranks of the elements of a. It is easy to interpret these quantifiers in terms of the ones we have been using, and thus determine what to do with them when applying the Transfer Principle. For example, consider a formula of the form $\forall x \in a\, \varphi$; this is equivalent to $\forall x \in \mathbb{V}_k(S)\, [x \in a \rightarrow \varphi]$. The $*$-transform of this formula is $\forall X \in {}^*\mathbb{V}_k(S)\, [X \in {}^*a \rightarrow {}^*\varphi]$, which is in turn equivalent to $\forall X \in {}^*a\, [{}^*\varphi]$. (Here we used the fact that all elements of $^*a$ are in ${}^*\mathbb{V}_k(S)$.) Similarly we can take the $*$-transform of $\exists x \in a\, \varphi$ to be $\exists X \in {}^*a\, [{}^*\varphi]$.

It is also possible to use bounded quantifiers in which two variables appear: $\forall x \in y$ and $\exists x \in y$. Recall that x and y must be variables that range over specific levels of $\mathbb{V}(S)$; say x ranges over $\mathbb{V}_k(S)$. Then $\exists x \in y\, \varphi$ is equivalent to $\exists x \in \mathbb{V}_k(S)\, [x \in y \wedge \varphi]$, which is a formula we already know how to handle. Similarly we rewrite $\forall x \in y\, \varphi$ as $\forall x \in \mathbb{V}_k(S)\, [x \in y \rightarrow \varphi]$.

Formulas such as $z = \{\{x, y\}, \{x\}\}$ and $z = (x_1, \ldots, x_n)$ can easily be expressed in superstructures using simple logical formulas in which only the membership relation $\in$ and bounded quantifiers occur. Therefore, when constructing the $*$-transform of a formula in which such basic formulas occur, they are unchanged except for the fact that the variables $x, y, z, x_1, \ldots, x_n$ are modified to range over the sorts ${}^*\mathbb{V}_k(S)$ for appropriate k.

Finally, if a, b are subsets of $\mathbb{V}_k(S)$ and $f\colon a \rightarrow b$ is a function, the condition $f(x) = y$ can also be expressed using a formula over $\mathbb{V}(S)$ in which only the membership relation $\in$ and bounded quantifiers are used. (As in the previous paragraph, x and y can also be replaced by ordered tuples.) Indeed, if we take x, y to be variables ranging over the sort $\mathbb{V}_k(S)$, then $f(x) = y$ is expressed by

$$\exists z \in \mathbb{V}_{k+3}(S)\,[z = (x, y) \wedge z \in f]$$

which is a bounded formula over $\mathbb{V}(S)$. The $*$-transform of this formula is equivalent to

$$\exists Z \in {}^*\mathbb{V}_{k+3}(S)\,[Z = (X, Y) \wedge Z \in {}^*f]$$

and this is a formula over ${}^*\mathbb{V}(S)$ which expresses the condition ${}^*f(X) = Y$.

6.3. Exercise. Let $a_1, \ldots, a_n$ be in $\mathbb{V}(S)$.

(i) ${}^*\{a_1, \ldots, a_n\} = \{{}^*a_1, \ldots, {}^*a_n\}$.

(ii) ${}^*(a_1, \ldots, a_n) = ({}^*a_1, \ldots, {}^*a_n)$.

6.4. Exercise. Let $a, b, c, d, a_1, \ldots, a_n, f, r$ be sets in $\mathbb{V}(S)$.

(i) $a \in b \iff {}^*a \in {}^*b$.

(ii) $a = b \iff {}^*a = {}^*b$.

(iii) $a \subseteq b \iff {}^*a \subseteq {}^*b$.

(iv) ${}^*(a \cup b) = {}^*a \cup {}^*b$, ${}^*(a \cap b) = {}^*a \cap {}^*b$, and ${}^*(a \setminus b) = ({}^*a) \setminus ({}^*b)$.

(v) ${}^*(a_1 \times \ldots \times a_n) = {}^*a_1 \times \ldots \times {}^*a_n$.

(vi) f is a function from a to $b \iff {}^*f$ is a function from ${}^*a$ to ${}^*b$.

(vii) r is a relation on $a \times b \iff {}^*r$ is a relation on ${}^*a \times {}^*b$; if these conditions are true, and if c is the domain of r (projection on the first coordinate) and d is the range of r (projection on the second coordinate), then ${}^*c$ is the domain of ${}^*r$ and ${}^*d$ is the range of ${}^*r$.

The arguments needed to solve these Exercises are simple applications of the Transfer Principle (Theorem 4.27).

In the next definition we introduce one of the most important concepts in the superstructure framework; this is a natural extension of what was done in Section 5 (Definition 5.3):

6.5. Definition. [Internal Object in $\mathbb{V}({}^*S)$] *An object in $\mathbb{V}({}^*S)$ is* **internal** *if there exists $k \in \mathbb{N}$ such that $a \in {}^*\mathbb{V}_k(S)$. Therefore, the collection of internal objects is transitive: a internal and $b \in a$ implies b internal. An object in $\mathbb{V}({}^*S)$ is* **external** *if it is not internal.*

6.6. Proposition. *Let a be in $\mathbb{V}({}^*S)$; a is internal if and only if it is an element of some standard set in $\mathbb{V}({}^*S)$. That is, a is internal if there exists b in $\mathbb{V}(S)$ such that $a \in {}^*b$.*

Proof. If a is internal, then $a \in {}^*\mathbb{V}_k(S)$ by definition, so a is an element of a standard set. Conversely, suppose $b \in \mathbb{V}(S)$ and $a \in {}^*b$. This implies b is a non-empty set so there exists $k \in \mathbb{N}$ with $b \subseteq \mathbb{V}_k(S)$. But then $a \in {}^*b \subseteq {}^*\mathbb{V}_k(S)$, and we are done. □

Suppose φ is a logical formula over $\mathbb{V}(S)$ and let ${}^*\varphi$ be its $*$-transform. Observe that each quantified variable in ${}^*\varphi$ is restricted to range over the elements of ${}^*\mathbb{V}_k(S)$ for some k. Therefore the quantified variables in ${}^*\varphi$ range only over *internal* elements. This means that if the free variables of ${}^*\varphi$ are taken to stand for internal objects in $\mathbb{V}({}^*S)$, then we will get the same truth value if we interpret ${}^*\varphi$ in the full superstructure $\mathbb{V}({}^*S)$ as if we evaluate it in the nonstandard extension ${}^*\mathbb{V}(S)$. External objects simply do not enter into the picture when we evaluate whether or not ${}^*\varphi$ is true in $\mathbb{V}({}^*S)$.

The following result is an easy consequence of the Transfer Principle (Theorem 4.27) applied to nonstandard extensions of superstructures. Nonetheless, it is an important tool in applications of nonstandard methods.

6.7. Theorem. [Internal Definition Principle] *Let $\varphi(x_1, \ldots, x_m, y_1, \ldots, y_n)$ be a formula over $\mathbb{V}(S)$. Suppose the variable x_j ranges over $\mathbb{V}_{\alpha(j)}(S)$ for each j and the variable y_k ranges over $\mathbb{V}_{\beta(k)}(S)$ for each k. Let $a_1, \ldots, a_n$ be internal objects in $\mathbb{V}({}^*S)$, with $a_j \in {}^*\mathbb{V}_{\alpha(j)}(S)$ for each j. Let b be the set in $\mathbb{V}({}^*S)$ defined by ${}^*\varphi(X_1, \ldots, X_m, a_1, \ldots, a_n)$:*

$$b = \{(X_1, \ldots, X_m) \in {}^*\mathbb{V}(S)^\alpha \mid {}^*\varphi(X_1, \ldots, X_m, a_1, \ldots, a_n) \text{ holds in } {}^*\mathbb{V}(S)\}.$$

Then b is internal.

Proof. This proof follows the same line of argument as the proof of Theorem 5.4. □

6.8. Remark. Note the requirement in the Internal Definition Principle that the objects $a_1, \ldots, a_n$ are internal. This is very important. A very common mistake when using nonstandard analysis is to misapply the Internal Definition Principle in a situation where some of the objects $a_1, \ldots, a_n$ are external.

6.9. Exercise. Let $a_1, \ldots, a_m$ be internal objects in $\mathbb{V}({}^*S)$.

(i) $\{a_1, \ldots, a_n\}$ is internal.

(ii) $(a_1, \ldots, a_n\rangle)$ is internal.

(iii) Every standard set in $\mathbb{V}({}^*S)$ is internal.

6.10. Exercise. Let $a, b, a_1, \ldots, a_n, f, r$ be internal sets in $\mathbb{V}({}^*S)$.

(i) Every Boolean combination of a, b is internal.

(ii) Every element of a is internal.

(iii) $a_1 \times \ldots \times a_n$ is internal.

(iv) If r is a relation on $a \times b$, then the domain of r and the range of r are internal.

(v) The union of all members of a and the intersection of all members of a are both internal.

(vi) The collection of all internal subsets of a is internal.

An important example of an internal concept is the notion of *hyperfinite* set. These are internal sets in $\mathbb{V}({}^*S)$ which obey all the formally expressible properties of finite sets. As a result, they can be handled using ideas of combinatorial and discrete mathematics. However, when viewed from outside, they may be infinite sets, and may share many qualitative features of continuous objects of mathematics. Many important applications of nonstandard analysis depend on the use of hyperfinite sets. The definition of *hyperfinite* is also a model for the introduction of many interesting concepts for internal sets.

6.11. Definition. [Hyperfinite Sets] *Let F_k be the collection of all finite sets in $\mathbb{V}_k(S)$. A set a in $\mathbb{V}({}^*S)$ is* **hyperfinite** *(equivalently* **$*$-finite***) if $a \in {}^*F_k$ for some $k \in \mathbb{N}$.*

6.12. Remark. Note that according to Proposition 6.6, every hyperfinite set is internal, since ${}^*F_k$ is a standard set for each k.

6.13. Notation. See the discussion before Exercise 5.6 for an explanation of the meaning of the relation $<$ and the functions $+$ and $\times$ on ${}^*\mathbb{R}$. Given $N \in {}^*\mathbb{N}$, we write $\{0, 1, \ldots, N\}$ for the set $\{a \in {}^*\mathbb{N} \mid 0 \leq a \leq N\}$. Note that $\{0, 1, \ldots, N\}$ is an infinite set if N is an infinite number in ${}^*\mathbb{N}$.

6.14. Exercise. (i) Every finite set in ${}^*\mathbb{V}(S)$ is hyperfinite.

(ii) For each $N \in \mathbb{N}$ the set $\{0, 1, \ldots, N\}$ is hyperfinite; ${}^*\mathbb{N}$ is not hyperfinite.

(iii) If a is a set in $\mathbb{V}(S)$ and ${}^*a$ is hyperfinite, then a is a finite set and ${}^*a = \{{}^*b \mid b \in a\}$.

6.15. Exercise. A set a in $\mathbb{V}({}^*S)$ is hyperfinite if and only if there exists an internal bijection between a and $\{0, 1, \ldots, N-1\}$ for some $N \in \mathbb{N}$. This N, if it exists, is unique.

6.16. Definition. *If a is a hyperfinite set in $\mathbb{V}({}^*S)$, the unique $N \in \mathbb{N}$ such that there exists an internal bijection between a and $\{0, 1, \ldots, N-1\}$ is called* **the internal cardinality of** *a.*

6.17. Exercise. Let $a, b, a_1, \ldots, a_n, f, r$ be hyperfinite sets in $\mathbb{V}({}^*S)$.

(i) Every Boolean combination of a, b is hyperfinite.

(ii) $a_1 \times \ldots \times a_n$ is hyperfinite.

(iii) If r is a relation on $a \times b$, then the domain of r and the range of r are hyperfinite.

(iv) If every member of a is a hyperfinite set, then the union of all members of a and the intersection of all members of a are both hyperfinite.

(v) The collection of all internal subsets of a is hyperfinite.

(vi) Every internal subset of a is hyperfinite, and its internal cardinality is $\leq$ the internal cardinality of a.

(vii) Suppose S contains $\mathbb{R}$; if a is a hyperfinite subset of $^*\mathbb{R}$ and N is the internal cardinality of a, then there is an internal increasing bijection from $\{0, 1, \ldots, N-1\}$ onto a.

6.18. Exercise. Suppose S contains $\mathbb{R}$. Let A be the set of all (α, N) where $N \in {}^*\mathbb{N}$ and α is an internal function from $\{0, 1, \ldots, N\}$ into $^*\mathbb{R}$. See part (iv) of Exercise 5.7 and Exercise 5.9.

(i) A is an internal set;

(ii) there is a unique internal function $\Sigma\colon A \to {}^*\mathbb{R}$ such that for all $(\alpha, N) \in A$

$$\Sigma(\alpha, N) = \sum_{k=0}^{N} \alpha(k).$$

6.19. Remark. If a is an internal set in $^*\mathbb{V}(S)$, we let $^*\mathcal{P}(a)$ denote the set of all internal subsets of a. By part (vi) of Exercise 6.10 we know that $^*\mathcal{P}(a)$ is an internal set in $^*\mathbb{V}(S)$; it is called the *internal power set of a*. Using part (i) of the same Exercise, we see that $^*\mathcal{P}(a)$ is closed under finite Boolean operations. Since it is an internal set, this implies that $^*\mathcal{P}(a)$ is actually closed under hyperfinite unions and intersections. Such internal Boolean algebras of sets are very important in the construction of Loeb measures.

6.20. Exercise. Let A be a set in $\mathbb{V}(S)$. Then $(^*A, {}^*\mathcal{P}(A), {}^*\mathcal{P}(A \times A))$ is a nonstandard extension of $(A, \mathcal{P}(A), \mathcal{P}(A \times A))$, and it satisfies the normalizing assumptions (b) and (c) given above just before Exercise 5.7.

For a more complete discussion of superstructures and more complete proofs of many facts about their nonstandard extensions, the reader may consult the textbook [8]; see also [1] [2] [4] [9] [11] [13] and [15].

A completely different set theoretic foundation for nonstandard analysis, *Internal Set Theory (IST)*, was introduced by Nelson in [12]. It is based on nonstandard models for the full ZFC axioms for the foundations of mathematics. (ZFC = Zermelo Fraenkel axioms for set theory with the Axiom of Choice.)

7. Saturation

For most applications, especially those in topology and abstract analysis, it is necessary to work with nonstandard extensions which satisfy richness conditions stronger than nontriviality or properness. (See Definitions 2.21

and 4.21.) The most useful of the extra hypotheses are the *saturation* conditions, which were carried over from model theory to nonstandard analysis by Luxemburg [11].

For this Section we fix a superstructure $\mathbb{V}(S)$ and a nonstandard extension $^*\mathbb{V}(S)$ of it. Recall that a family F of sets is said to have the *finite intersection property* if each intersection of a finite subcollection of F is non-empty. We let κ stand for an uncountable cardinal number.

7.1. Definition. *The given nonstandard extension is* κ-**saturated** *if it satisfies the following condition: let $\mathcal{F}$ be a (possibly external) family of internal sets; if $\mathcal{F}$ has (external) cardinality strictly less than κ and $\mathcal{F}$ has the finite intersection property, then the total intersection of $\mathcal{F}$ is non-empty. (The total intersection of $\mathcal{F}$ is the set $\{a \in \mathbb{V}(^*S) \mid a \in b$ for all $b \in \mathcal{F}\}$. Of course this set may be external.)*

The following result gives an alternate formulation of κ-saturation. It is expressed in terms of simultaneous satisfiability of conditions, each expressed by formulas over $^*\mathbb{V}(S)$, in which only internal objects are allowed.

7.2. Theorem. *Let $^*\mathbb{V}(S)$ be a κ-saturated nonstandard extension of the superstructure $\mathbb{V}(S)$, where κ is an uncountable cardinal number. Let J be an index set of cardinality $< \kappa$. Let a be an internal set in $^*\mathbb{V}(S)$. For each $j \in J$, let $\varphi_j(X)$ be a formula over $^*\mathbb{V}(S)$, so all objects mentioned in $\varphi_j(X)$ are internal. Further, suppose that the set of formulas $\{\varphi_j(X) \mid j \in J\}$ is finitely satisfied in a; this means that for every finite subset α of J there exists some $c \in a$ (which may depend on α) such that $\varphi_j(c)$ holds in $^*\mathbb{V}(S)$ for all $j \in \alpha$. Then there exists $c \in a$ such that $\varphi_j(c)$ holds in $^*\mathbb{V}(S)$ simultaneously for all $j \in J$.*

Proof. For each $j \in J$, let f_j be the subset of a that is defined by $\varphi_j(X)$; that is,

$$f_j = \{c \in a \mid \varphi_j(c) \text{ is true in } ^*\mathbb{V}(S)\}.$$

The Internal Definition Principle implies that each f_j is an internal subset of a. The hypotheses imply that the collection $\{f_j \mid j \in J\}$ has the finite intersection property. Since the cardinality of J is $< \kappa$, the fact that our nonstandard extension is κ-saturated implies that the total intersection $\cap\{f_j \mid j \in J\}$ is nonempty. Any element of this intersection satisfies the conclusion of the Theorem. □

Of special importance for most applications is $\aleph_1$-saturation, where $\aleph_1$ denotes the first uncountable cardinal number. This means that whenever $\mathcal{F}$ is a countable collection of internal sets and $\mathcal{F}$ has the finite intersection property, then the total intersection of $\mathcal{F}$ is non-empty. It is customary in nearly all research articles in nonstandard analysis to assume that the nonstandard extensions being used are at least $\aleph_1$-saturated. It is this hypothesis that ensures, for example, that Loeb measures are σ-additive and that nonstandard hulls of metric spaces are complete. In some areas, especially in topology, an even stronger hypothesis of κ-saturation is needed for many applications. For example, in order to give a smooth treatment of a topological space T using the methods of nonstandard analysis, it is usually necessary to assume that the nonstandard extension is κ-saturated where κ is strictly larger than the number of open subsets of T.

Note that the saturation hypotheses can also be applied to the simpler nonstandard extensions treated in Section 5.

7.3. Proposition. *Assume that the nonstandard extension is κ-saturated. Every infinite internal set in $\mathbb{V}(^*S)$ has (external) cardinality $\geq \kappa$.*

Proof. Suppose otherwise, that a is an infinite internal set of cardinality strictly less than κ. Let $\mathcal{F}$ be the collection of all sets of the form $a \setminus \{x\}$ as x ranges over a. Then $\mathcal{F}$ is a collection of internal sets, and the cardinality of $\mathcal{F}$ is less than κ. Moreover, $\mathcal{F}$ obviously has the finite intersection property, since a is infinite. But the total intersection of $\mathcal{F}$ is obviously empty; this contradicts the hypothesis that the nonstandard extension is κ-saturated. □

7.4. Proposition. *Assume that the nonstandard extension is κ-saturated. Let a be an internal set in $\mathbb{V}(^*S)$. Let A be a (possibly external) subset of a such that A has cardinality strictly less than κ. Then there exists a hyperfinite subset b of a such that b contains A as a subset.*

Proof. For each $x \in A$, let F_x denote the set of all hyperfinite subsets of a which contain x as an element. The Internal Definition Principle (Theorem 6.7) yields that each F_x is an internal set in $\mathbb{V}(^*S)$. Let $\mathcal{F}$ be the collection of all the sets F_x as x ranges over A. Obviously $\mathcal{F}$ has cardinality strictly less than κ. Moreover, $\mathcal{F}$ has the finite intersection property: given finitely many elements $x_1, \ldots, x_n$ from A, the set $\{x_1, \ldots, x_n\}$ is hyperfinite and is an element of F_{x_j} for all $j = 1, \ldots, n$. Since our nonstandard extension is κ-saturated, there exists an object b which is an element of F_x for every $x \in A$. This b is therefore the desired hyperfinite set. □

7.5. Theorem. [Saturated Extensions are Comprehensive] *Assume that the nonstandard extension is κ-saturated. Let a and b be internal sets in $\mathbb{V}(^*S)$. Let A be a (possibly external) subset of a such that A has cardinality strictly less than κ and suppose that $f: A \to b$ is a (possibly external) function. Then there exists an internal function $g: a \to b$ such that g is an extension of f. In particular, if $\{c_k \mid k \in \mathbb{N}\}$ is a (possibly external) sequence of elements of b, then there exists an internal function $g: {}^*\mathbb{N} \to b$ such that $g(k) = c_k$ for all $k \in \mathbb{N}$.*

Proof. For each $x \in A$ let F_x be the set of all internal functions $g: a \to b$ which satisfy $g(x) = f(x)$. The Internal Definition Principle (Theorem 6.5) implies that each F_x is internal. Let $\mathcal{F}$ be the collection of all F_x as x ranges over A. Obviously $\mathcal{F}$ has cardinality strictly less than κ. Moreover, $\mathcal{F}$ has the finite intersection property: given finitely many elements $x_1, \ldots, x_n$ from A, consider the function $g: a \to b$ which takes x_j to $f(x_j)$ for $j = 1, \ldots, n$ and which takes all other elements of a to (say) $f(x_1)$. The Internal Definition Principle implies that this function is internal, and it is obviously an element of F_{x_j} for all $j = 1, \ldots, n$. From the fact that our nonstandard model is assumed to be κ-saturated, it follows that there is an object g which is an element of F_x for all $x \in A$. This g is the desired internal function from a to b. □

7.6. Exercise. The following conditions are equivalent:

(a) The nonstandard extension is $\aleph_1$-saturated.

(b) (Countable Comprehension Property) Whenever b is an internal set in $\mathbb{V}(^*S)$ and $(c_k)_{k \in \mathbb{N}}$ is a (possibly external) sequence of elements of b, then there exists an internal function $g: {}^*\mathbb{N} \to b$ such that $g(k) = c_k$ for all $k \in \mathbb{N}$.

7.7. Exercise. Assume the nonstandard extension is $\aleph_1$-saturated and let a be an internal set in $\mathbb{V}(^*S)$. A (possibly external) subset b of a is a Σ_1^0 set if there exists a sequence $\{c_k \mid k \in \mathbb{N}\}$ of internal subsets of a such that $b = \bigcup\{c_k \mid k \in \mathbb{N}\}$. Similarly, b is a Π_1^0 set if there exists a sequence $\{c_k \mid k \in \mathbb{N}\}$ of internal subsets of a such that $b = \bigcap\{c_k \mid k \in \mathbb{N}\}$. A Σ_1^0 set is sometimes called a *galaxy*, and a Π_1^0 set is sometimes called a *monad* or a *halo*.

(a) Suppose $\{c_k \mid k \in \mathbb{N}\}$ is a sequence of internal subsets of a such that b_1 is the Σ_1^0 set $\bigcup\{c_k \mid k \in \mathbb{N}\}$. If b_1 is internal, then there exists $N \in \mathbb{N}$ such that $b_1 = c_1 \cup \ldots \cup c_N$. (Hint: without loss of generality the sequence $\{c_k \mid k \in \mathbb{N}\}$ is the restriction to $\mathbb{N}$ of an internal increasing sequence $\{c_k \mid k \in {}^*\mathbb{N}\}$ of subsets of a. If $b_1 = \bigcup\{c_k \mid k \in \mathbb{N}\}$ is internal, then the set of $N \in {}^*\mathbb{N}$ such that $b_1 \subseteq c_1 \cup \ldots \cup c_N$ is internal and contains all infinite N. By the Underspill Principle, Exercise 5.6, there exists a finite $N \in \mathbb{N}$ such that $b_1 \subseteq c_1 \cup \ldots \cup c_N$.)

(b) Suppose $\{d_k \mid k \in \mathbb{N}\}$ is a sequence of internal subsets of a such that b_2 is the Π_1^0 set $\bigcap\{d_k \mid k \in \mathbb{N}\}$. If b_2 is internal, then there exists $N \in \mathbb{N}$ such that $b_2 = d_1 \cap \ldots \cap d_N$.

(c) Suppose b_1 is a Σ_1^0 subset of a represented as in (a) and b_2 is a Π_1^0 subset of a represented as in (b). If $b_1 \subseteq b_2$ then there is an internal set e such that $b_1 \subseteq e \subseteq b_2$. In particular, if $b_1 = b_2$, then $b_1 (= b_2)$ is an internal set.

7.8. Exercise. This improves on Proposition 7.3 when $\kappa = \aleph_1$. Assume the nonstandard extension is $\aleph_1$-saturated. Every infinite internal set has (external) cardinality $\geq 2^{\aleph_0}$. (Hint: without loss of generality the infinite internal set is of the form $\{0, 1, \ldots, N\}$ where N is an infinite element of $^*\mathbb{N}$. For each standard real number r in the interval $0 < r < 1$ show that there is a smallest element k of $^*\mathbb{N}$ such that $N \cdot {}^*r \leq k$. Obviously $1 \leq k \leq N$. Moreover, k is uniquely determined by r.)

It is useful to introduce two additional richness properties of nonstandard extensions:

7.9. Definition. *(a) A nonstandard extension of* $\mathbb{V}(S)$ *is* **polysaturated** *if it is* κ*-saturated for some* κ *greater than or equal to the number of objects in* $\mathbb{V}(S)$.

(b) A nonstandard extension of $\mathbb{V}(S)$ *is an* **enlargement** *if for each set* a *in* $\mathbb{V}(S)$ *there exists a hyperfinite set* $b \subseteq {}^*a$ *such that* b *contains every standard element of* $^*a$*; that is, we require* $\{^*c \mid c \in a\} \subseteq b \subseteq {}^*a$.

7.10. Exercise. The following conditions are equivalent:

(a) The nonstandard extension is an enlargement.

(b) If $k \in \mathbb{N}$ and $\mathcal{F}$ is a collection of subsets of $\mathbb{V}_k(S)$ with the finite intersection property, then $\bigcap\{^*a \mid a \in \mathcal{F}\}$ is nonempty.

(c) If $k \in \mathbb{N}$ and $(L, \leq)$ is a partially ordered set in $\mathbb{V}_k(S)$ which is directed upwards, then there exists $b \in {}^*L$ such that for all $a \in L$, $^*a \;{}^*\!\!\leq b$.

7.11. Exercise. Every polysaturated nonstandard extension of $\mathbb{V}(S)$ is an enlargement.

Next we give a proof using ultrapowers of the existence of enlargements.

7.12. Theorem. *Every superstructure* $\mathbb{V}(S)$ *has a nonstandard extension which is an enlargement; it can be taken to be an ultrapower extension of* $\mathbb{V}(S)$ *with respect to a suitably chosen ultrafilter.*

Proof. Let J be the collection of all nonempty finite sets of objects from $\mathbb{V}(S)$. Let $\mathcal{U}$ be an ultrafilter on J such that for each object a in $\mathbb{V}(S)$ the set

$$\{j \mid j \text{ is a finite set of objects from } \mathbb{V}(S) \text{ and } a \in j\}$$

is in $\mathcal{U}$. Such an ultrafilter exists because the collection of all these subsets of J has the finite intersection property. Consider the nonstandard extension of $\mathbb{V}(S)$ constructed as an ultrapower using $\mathcal{U}$ as in the proof of Theorem 4.28. We will show that this is an enlargement of $\mathbb{V}(S)$.

Fix a set a from $\mathbb{V}(S)$. For each $j \in J$ let $F(j) = a \cap j$ and let $b = [F]$ be the element of the nonstandard extension that is determined by F. Note that if a has rank k, then all of the values of F are in $\mathbb{V}_k(S)$ so b is an element of $^*\mathbb{V}_k(S)$. Since every $F(j)$ is a finite subset of a, it follows that b

is a hyperfinite subset of $^*a$. It remains to show that every standard element of $^*a$ is an element of b. Let $c \in a$. As in Exercise 2.29 $^*c$ is the equivalence class $[\alpha]$ where α is the constant function with value c at each argument in J. To show that $^*c \in b$ we must show that the set $\{j \in J \mid c \in F(j)\}$ is in $\mathcal{U}$. But this set equals $\{j \in J \mid c \in j\}$, which is in $\mathcal{U}$ by construction. □

7.13. Theorem. *Let $\mathbb{V}(S)$ be a superstructure and let κ be an uncountable cardinal number. There exists a κ-saturated nonstandard extension of $\mathbb{V}(S)$. In particular, there exists a polysaturated nonstandard extension of $\mathbb{V}(S)$.*

Proof. We begin by proving the important fact that if $\mathcal{U}$ is any countably incomplete ultrafilter on an infinite index set, then the ultrapower nonstandard extension of $\mathbb{V}(S)$ constructed as in Theorem 4.28 is necessarily $\aleph_1$-saturated. Since $\mathcal{U}$ is countably incomplete, we may suppose $(F_k)_{k \in \mathbb{N}}$ is a decreasing sequence in $\mathcal{U}$ whose intersection is empty and with $F_0 = J$. Let $\{a_k \mid k \in \mathbb{N}\}$ be a set of internal sets in $\mathbb{V}(^*S)$ with the finite intersection property. We must show that $\bigcap\{a_k \mid k \in \mathbb{N}\}$ is non-empty. Without loss of generality we may suppose that $a_{k+1} \subseteq a_k$ for all $k \in \mathbb{N}$. Therefore there exists $r \in \mathbb{N}$ so that a_k has rank at most r for all $k \in \mathbb{N}$. Since $^*\mathbb{V}(S)$ was obtained by the ultrapower construction using the ultrafilter $\mathcal{U}$, for each $k \in \mathbb{N}$ there is a set function $A_k \colon J \to \mathbb{V}_r(S)$ such that a_k is the equivalence class $[A_k]$. Since $\{a_k \mid k \in \mathbb{N}\}$ has the finite intersection property, for each $k \in \mathbb{N}$ we have

$$\{j \in J \mid A_0(j) \cap \ldots \cap A_k(j) \neq \emptyset\} \in \mathcal{U}.$$

Define G_k for $k \in \mathbb{N}$ as follows: $G_0 = J$ and for $k \geq 1$

$$G_k = F_k \cap \{j \in J \mid A_0(j) \cap \ldots \cap A_k(j) \neq \emptyset\}.$$

Therefore $J = G_0 \supseteq G_1 \supseteq \ldots \supseteq G_k \supseteq \ldots$, $G_k \in \mathcal{U}$ for all $k \in \mathbb{N}$, and $\bigcap\{G_k \mid k \in \mathbb{N}\} = \emptyset$. Therefore we may define $d(j)$ for each $j \in J$ to be the largest $k \in \mathbb{N}$ for which $j \in G_k$. Now we construct $[\alpha]$ in $^*\mathbb{V}(S)$ which is an element of $\bigcap\{a_k \mid k \in \mathbb{N}\}$. Fix $j \in J$ and define $\alpha(j)$ as follows. If $d(j) = 0$ let $\alpha(j)$ be an arbitrary element of $\mathbb{V}_r(S)$. If $d(j) \geq 1$, choose $\alpha(j)$ to be an element of $A_0(j) \cap \ldots \cap A_{d(j)}(j)$, which is guaranteed to be non-empty by the definition of $d(j)$. It is obvious that for each $k \in \mathbb{N}$, $\alpha(j) \in A_k(j)$ holds whenever $d(j) \geq k$ and $d(j) \geq 1$. Therefore $\{j \in J \mid \alpha(j) \in A_k(j)\} \supseteq G_k \in \mathcal{U}$ for $k \geq 1$ and $\{j \in J \mid \alpha(j) \in A_0(j)\} \supseteq G_1 \in \mathcal{U}$. This completes the proof that $[\alpha]$ is an element of $\bigcap\{a_k \mid k \in \mathbb{N}\}$.

We will not give the details of a proof of the general case. The easiest construction of a κ-saturated nonstandard extension of $\mathbb{V}(S)$ for $\kappa > \aleph_1$ is to take the direct limit of a well ordered chain of successive enlargements. The length of the chain should be a regular cardinal number $\geq \kappa$; a chain of length κ^+ will suffice, where κ^+ is the next cardinal number larger than κ. It is also possible, but rather intricate in the case where $\kappa > \aleph_1$, to construct a κ-saturated nonstandard extension in one step as an ultrapower, by choosing the ultrafilter carefully. Details may be found in [8] [9] [11] and [15].

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., and Lindstrøm, T., (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Capiński, M. and Cutland, N. J., (1995) *Nonstandard Methods for Stochastic Fluid Mechanics.* World Scientific, Singapore.
3. Cutland, N. J., Editor, (1988) *Nonstandard Analysis and its Applications.* Cambridge University Press, Cambridge.
4. Davis, M., (1977) *Applied Nonstandard Analysis.* John Wiley & Sons, New York.
5. Van den Dries, L., *Tame Topology and o-minimal Structures.* (Monograph in preparation).
6. Van den Dries, L. and Miller, C., Geometric categories and o-minimal structures, to appear in *Duke Mathematical Journal.*
7. Van den Dries, L. and Wilkie, A.J., (1984) Gromov's Theorem on groups of polynomial growth and elementary logic. *Journal of Algebra,* **89**, pp. 349–374.
8. Hurd, A. and Loeb, P. A., *An Introduction to Nonstandard Real Analysis.* Academic Press, New York.
9. Lindstrøm, T., (1988) An invitation to nonstandard analysis. In Cutland (1988), pp. 1–105.
10. Luxemburg, W.A.J.,(1969a) *Applications of Model Theory to Algebra, Analysis, and Probability.*, Holt, Rinehart and Winston, New York.
11. Luxemburg, W.A.J.,(1969b) A general theory of monads. In Luxemburg (1969a). Pages 18–86.
12. Nelson, E., (1977) Internal set theory. *Bulletin of the American Mathematical Society* **83**, pp. 1165–1193.
13. Robinson, A. and Zakon, E., (1969) A set-theoretical characterization of enlargements. In Luxemburg (1969a), pp. 109–122.
14. Robinson, A., (1966) *Nonstandard Analysis.* North-Holland, Amsterdam. (Second, revised edition, 1974).
15. Stroyan, K. and Luxemburg, W. A. J., (1976) *Introduction to the Theory of Infinitesimals.* Academic Press, New York.

NONSTANDARD REAL ANALYSIS

NIGEL J. CUTLAND
School of Mathematics
University of Hull
Hull HU6 7RX
England
email: n.j.cutland@maths.hull.ac.uk

1. Introduction

In this article we show how a nonstandard extension $^*\mathbb{R}$ of $\mathbb{R}$ can be used to formulate the fundamental ideas of infinitesimal calculus in a natural and intuitive way, and thereby develop real analysis rigorously based on these ideas. We include a number of exercises (which include proofs of results that are only slight developments of the theory) and encourage the reader who is new to this subject to work through as many of these as possible – for it is only by *doing* it that one can become fluent in the ideas and techniques of nonstandard analysis.

We assume given a fixed proper nonstandard extension $^*\mathbb{R}$ of $\mathbb{R}$ as defined in Ward Henson's article [6] in this volume; we shall see below that $^*\mathbb{R}$ is a field (and much more), and refer to it as a field of *hyperreals* (or *nonstandard real numbers*). We assume (see Henson [6], Remark 2.10) that $^*\mathbb{R} \supset \mathbb{R}$. Recall that such an extension carries with it an extension $^*A \supset A$ for each $A \subset \mathbb{R}^n$, for which $A = {}^*A \cap \mathbb{R}^n$. Similarly, for each function $f : A \to B$ where $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ there is an extension $^*f : {}^*A \to {}^*B$ (whose graph is $^*\Gamma$ where Γ is the graph of f). This really *is* an extension in the usual sense that $^*f|A = f$.

The key tool throughout the development below will be the Transfer Principle (Henson [6] Theorem 3.3), which summarizes the properties of the mapping * in an easily usable form.

A fuller treatment of the topics we touch on in this article may be found in the references [1, 2, 3, 4, 5, 7, 9, 10, 13, 14] – and the reader should consult these for discussion of more advanced topics that we do not cover.

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 51–76.

2. Basic Properties of $^*\mathbb{R}$

Theorem 2.1 *$^*\mathbb{R}$ is an ordered field (under the operations $^*+$, $^*\times$ and the relation $^*<$).*

Proof This is proved simply by applying the Transfer Principle to each of the axioms for an ordered field. For example, each of the following is true in $\mathbb{R}$:

$$\forall x \forall y [x + y = y + x]$$

$$\forall x [x \neq 0 \rightarrow \exists y [x \times y = 1]]$$

$$\forall x \forall y [x = y \vee x < y \vee y < x]$$

The *-transforms of these statements (which are true in $^*\mathbb{R}$ by the Transfer Principle) are as follows:

$$\forall \mathrm{X} \forall \mathrm{Y} [\mathrm{X} \ {}^*\!+ \mathrm{Y} = \mathrm{Y} \ {}^*\!+ \mathrm{X}]$$

$$\forall \mathrm{X} [\mathrm{X} \neq 0 \rightarrow \exists \mathrm{Y} [\mathrm{X} \ {}^*\!\times \mathrm{Y} = 1]]$$

$$\forall \mathrm{X} \forall \mathrm{Y} [\mathrm{X} = \mathrm{Y} \ \vee \ \mathrm{X} \ {}^*\!< \mathrm{Y} \ \vee \ \mathrm{Y} \ {}^*\!< \mathrm{X}]$$

Note that the variables X, Y in these statements range over the *elements* of $^*\mathbb{R}$.

Treating all the other axioms for an ordered field in the same way (which is possible because they are all *first order* axioms - i.e. they are expressed without reference to subsets of $\mathbb{R}$ or other higher order entities) we see that the relation $^*<$ is a linear order on $^*\mathbb{R}$ and that $^*\mathbb{R}$ is an ordered field. □

Applying the Transfer Principle to the appropriate statements about $\mathbb{Z}$ we have, in the same way:

Theorem 2.2 *$^*\mathbb{Z}$ is an integral domain.*

Proof Exercise. □

The Transfer Principle tells us that *all* first order properties of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ (and functions and relations on them) are true also in $^*\mathbb{N}$, $^*\mathbb{Z}$, $^*\mathbb{Q}$ and $^*\mathbb{R}$ (and the *-extensions of the functions and relations). We say that such properties hold *by transfer* – for example the following.

- For every $r \in {}^*\mathbb{R}$ there is a unique $n \in {}^*\mathbb{Z}$ with

$$n \leq r \ {}^*\!< n \ {}^*\!+ 1$$

(and this $n = {}^*[r]$, where $^*[\cdot]$ is the extension of the integer part function $[\cdot]$).

- For all $x \in {}^*\mathbb{R}$

$$^*\!\cos^2 x \ {}^*\!+ \ {}^*\!\sin^2 x = 1.$$

- ${}^*\mathbb{N} = \{n \in {}^*\mathbb{Z} : \ n \ {}^*\!\geq 0\}$
- Every pair of nonzero elements $m, n \in {}^*\mathbb{N}$ has a unique positive highest common factor d that can be expressed in the form $d = sm \ {}^*\!\!+ tn$ for some $s, t \in {}^*\mathbb{Z}$.
- ${}^*\mathbb{Q}$ is an ordered field that is not algebraically closed, and ${}^*\mathbb{Q}$ is dense in ${}^*\mathbb{R}$.

Exercise 2.3 Carefully formulate statements to which the Transfer Principle can be applied to give the above properties of ${}^*\mathbb{N}$, ${}^*\mathbb{Z}$, ${}^*\mathbb{Q}$ and ${}^*\mathbb{R}$.

Remark 2.4 In the above we have written sm to mean $s \ {}^*\!\!\times m$. It is common to drop the prefix * from many of the common functions and relations on ${}^*\mathbb{R}$ when there is no ambiguity. So for example we will write

$$\begin{array}{lll} r+s & \text{for} & r \ {}^*\!\!+ s \\ rs & \text{for} & r \ {}^*\!\!\times s \\ r<s & \text{for} & r \ {}^*\!\!< s \\ |r| & \text{for} & {}^*|r| \\ \cos x & \text{for} & {}^*\cos x \end{array}$$

for all $r, s, x, y \in {}^*\mathbb{R}$. This seldom gives any ambiguity because for any relation A and function f the nonstandard extensions ${}^*A$ and ${}^*f$ are extensions in the conventional sense.

Clearly we could continue indefinitely listing properties of ${}^*\mathbb{R}$ that hold by transfer; it is instructive for the beginner to write down (and understand the significance of) further examples for him/herself, to gain familiarity with these new structures.

Exercise 2.5 Formulate further examples of properties of ${}^*\mathbb{N}$, ${}^*\mathbb{Z}$, ${}^*\mathbb{Q}$ and ${}^*\mathbb{R}$ that are true by virtue of the Transfer Principle.

As is pointed out in Henson's article [6], there are properties of $\mathbb{R}$ that are *not* transferred to ${}^*\mathbb{R}$ – for example Dedekind completeness and the Archimedean property. These are *second* order properties, whereas the Transfer Principle we are currently using ([6] Theorem 3.3) deals only with *first* order statements. (There *is* a more general Transfer Principle that is applicable to higher order objects, which will be needed in Section 8. This generalized Transfer Principle is introduced in Section 4 (Theorem 4.27) of Henson's article [6], and gives *Dedekind completeness and *Archimedean properties; however some care is needed to understand what these mean for ${}^*\mathbb{R}$.)

The next definitions are basic for understanding the structure of ${}^*\mathbb{R}$ and its relation to $\mathbb{R}$.

Definition 2.6 Let $x \in {}^*\mathbb{R}$. We say that

(a) x is *infinitesimal* if $|x| < \varepsilon$ for all $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$;
(b) x is *finite* if $|x| < r$ for some $r \in \mathbb{R}$;
(c) x is *infinite* if $|x| > r$ for all $r \in \mathbb{R}$.

We say that x and y are *infinitely close*, denoted by $x \approx y$, if $x - y$ is infinitesimal. So $x \approx 0$ is another way to say that x is infinitesimal.

Remarks (1) It is immediate from the definition that x is infinite if and only if it is not finite (as expected).

(2) By $|x|$ we mean ${}^*|x|$ where ${}^*|\cdot|$ is the extension of $|\cdot|$ to ${}^*\mathbb{R}$. Thus we are using the convention mentioned in Remark 2.4. Clearly, x is infinitesimal if and only if $-\varepsilon < x < \varepsilon$ for all positive $\varepsilon \in \mathbb{R}$; equivalently, if $-n^{-1} < x < n^{-1}$ for all $n \in \mathbb{N}$.

Note that 0 is an infinitesimal (the only real infinitesimal) and all infinitesimals are finite. We have yet not shown that ${}^*\mathbb{R}$ actually has any nonzero infinitesimals – but this will be a consequence of the Standard Part Theorem which follows shortly. First we note the following facts, which agree with our intuition about infinitesimal, finite, and infinite elements.

Proposition 2.7 *Let γ, δ be infinitesimal, x, y finite, and α, β infinite (all in ${}^*\mathbb{R}$). Then*

(a) *the following are infinitesimal: $\gamma + \delta$, $\gamma\delta$, δx, α^{-1};*
(b) *the following are finite: $\delta + x$, $x + y$, xy, x^{-1} (if $x \not\approx 0$);*
(c) *the following are infinite: δ^{-1} (if $\delta \neq 0$), $\alpha + \beta$ (if both positive or both negative), $\alpha\beta$.*

The proof is left as an exercise, and we also suggest the following.

Exercise 2.8 Prove, or find counterexamples to, other general statements that might seem plausible. For instance, show that there is no general way to categorize the product of an infinitesimal with an infinite hyperreal, by finding infinitesimals δ and infinite hyperreals α illustrating each of the possibilities:

(a) $\delta\alpha$ is infinitesimal,
(b) $\delta\alpha$ is finite and not infinitesimal,
(c) $\delta\alpha$ is infinite.

Corollary 2.9 *Let $x, y, u, v \in {}^*\mathbb{R}$ with $x \approx y$ and $u \approx v$. Then*

(a) $x + u \approx y + v$
(b) *$xu \approx yv$ provided x, u are finite.*

Proof Exercise, using Proposition 2.7. □

The next result is fundamental for relating constructions in $^*\mathbb{R}$ to those in $\mathbb{R}$.

Theorem 2.10 (Standard Part Theorem) *If $x \in {}^*\mathbb{R}$ is finite, then there is a unique $r \in \mathbb{R}$ such that $x \approx r$; i.e. x is uniquely expressible as $x = r + \delta$ with r a* standard *real and δ infinitesimal.*

Proof Put $r = \sup\{a \in \mathbb{R} : a \leq x\} = \sup A$, say. The set A is bounded above (in $\mathbb{R}$) since x is finite, and so the least upper bound r exists.

Take any real $\varepsilon > 0$. Then $r - \varepsilon$ is not an upper bound of A, so there is $a \in A$ with $r - \varepsilon < a \leq x$. On the other hand, $r + \varepsilon \notin A$ (because r is an upper bound of A), and thus $x < r + \varepsilon$. Consequently

$$-\varepsilon < x - r < \varepsilon$$

for all ε, and so $r \approx x$.

For uniqueness, suppose that $x = r_1 + \delta_1 = r_2 + \delta_2$. Then $r_1 - r_2 = \delta_2 - \delta_1$ is both real and infinitesimal, hence equal to 0. □

Definition 2.11 The real number r such that $r \approx x$ is called the *standard part* of x, written $r = \mathrm{st}(x) = {}^\circ x$ (both notations are common and useful).

Another way to state the Standard Part Theorem is to say that in $^*\mathbb{R}$ the finite elements are the *nearstandard* elements, according to the following definition.

Definition 2.12 Let $x \in {}^*\mathbb{R}$. Then x is *nearstandard* if $x \approx r$ for some $r \in \mathbb{R}$. We write $\mathrm{ns}({}^*\mathbb{R})$ to denote the nearstandard points in $^*\mathbb{R}$.

The following is easily deduced from Corollary 2.9.

Proposition 2.13 *For finite hyperreals x and y:*

(a) ${}^\circ(x + y) = {}^\circ x + {}^\circ y$,
(b) ${}^\circ(xy) = {}^\circ x \, {}^\circ y$,
(c) *if $x \leq y$ then ${}^\circ x \leq {}^\circ y$.*

Proof Exercise □

Exercise 2.14 Show that it is not generally true that if $x < y$ then ${}^\circ x < {}^\circ y$.

We can now see that infinitesimals and infinite elements exist.

Corollary 2.15 $^*\mathbb{R}$ *has nonzero infinitesimal and infinite elements.*

Proof Let $x \in {}^*\mathbb{R} \setminus \mathbb{R}$. If x is finite, let $\delta = x - {}^\circ x$. Then δ is a nonzero infinitesimal and δ^{-1} is infinite. If, on the other hand, x is infinite, then x^{-1} is a nonzero infinitesimal. □

The following notation and terminology is very useful.

Notation

(a) $\mathrm{Fin}({}^*\mathbb{R}) =$ the finite elements in ${}^*\mathbb{R} = \mathrm{ns}\ ({}^*\mathbb{R})$, by the Standard Part Theorem;
(b) $\mathrm{monad}(x) = \{y : \ y \approx x\}$; so $\mathrm{monad}(0) =$ the infinitesimals in ${}^*\mathbb{R}$.

Exercise 2.16 Show that

(a) $\mathrm{monad}(0)$ and $\mathrm{Fin}({}^*\mathbb{R})$ are both subrings of ${}^*\mathbb{R}$
(b) $\mathrm{monad}(0)$ is an ideal in $\mathrm{Fin}({}^*\mathbb{R})$ and

$$\mathrm{Fin}({}^*\mathbb{R})/\mathrm{monad}(0) \cong \mathbb{R}$$

3. Sequences and Series

The intuitive idea of a sequence s_n converging to a limit a is that when n is infinitely large, then s_n is infinitely close to a. With infinitesimal and infinite numbers we can express this idea precisely, as in the following theorem that characterizes convergence.

First, by way of explanation, consider a real sequence (s_n). This is really a function $s : \mathbb{N} \to \mathbb{R}$, which then has an extension ${}^*s : {}^*\mathbb{N} \to {}^*\mathbb{R}$. If K is now an infinite natural number then we write ${}^*s_K$ to mean ${}^*s(K)$.

The theorem below is typical of results that characterize convergence using the additional framework and expressive power provided by the existence of the hyperreals ${}^*\mathbb{R}$. The reasoning involved in the proof also typifies arguments that use the Transfer Principle to move back and forth between the standard and the nonstandard worlds, so we provide full details on this occasion.

Theorem 3.1 *Let (s_n) be a sequence of real numbers and let $a \in \mathbb{R}$. Then*

$$s_n \to a \text{ as } n \to \infty \quad \Longleftrightarrow \quad {}^*s_K \approx a \text{ for all infinite } K \in {}^*\mathbb{N}.$$

Proof Suppose first that $s_n \to a$, and fix infinite $K \in {}^*\mathbb{N}$. We have to show that $|{}^*s_K - a| < \varepsilon$ for all real $\varepsilon > 0$.

For any such ε there is a number $n_0 \in \mathbb{N}$ such that the following holds in $\mathbb{R}$:

$$\forall n \in \mathbb{N}[n \geq n_0 \to |s_n - a| < \varepsilon]$$

The Transfer Principle now tells us that

$$\forall N \in {}^*\mathbb{N}[N \geq n_0 \to |{}^*s_N - a| < \varepsilon]$$

is true in ${}^*\mathbb{R}$. In particular taking $N = K$ we see that $|{}^*s_K - a| < \varepsilon$ as required.

Conversely, suppose that ${}^*s_K \approx a$ for all infinite $K \in {}^*\mathbb{N}$. Let a real $\varepsilon > 0$ be given. Then for any given infinite $K \in {}^*\mathbb{R}$ the following holds (in ${}^*\mathbb{R}$):

$$\forall N \in {}^*\mathbb{N}[N \geq K \to |{}^*s_N - a| < \varepsilon].$$

Thus, taking one particular infinite K we have

$$\exists K \in {}^*\mathbb{N}\ \forall N \in {}^*\mathbb{N}[N \geq K \to |{}^*s_N - a| < \varepsilon]$$

The Transfer Principle applied to this sentence shows that in $\mathbb{R}$:

$$\exists k \in \mathbb{N} \forall n \in \mathbb{N}[n \geq k \to |s_n - a| < \varepsilon]$$

Taking n_0 to be any such k proves that $s_n \to a$. □

Remarks (**a**) In the above proof we have used the following abbreviations:

$$\begin{array}{lll} \exists k \in \mathbb{N} \ldots & \text{for} & \exists k[k \in \mathbb{N} \wedge \ldots\ ] \\ \forall k \in \mathbb{N} \ldots & \text{for} & \forall k[k \in \mathbb{N} \to \ldots\ ] \end{array}$$

and similarly for the $*$-transforms of these. In many situations, we will go further and use $\exists k$ and $\forall k$ to mean $\exists k \in \mathbb{N}$ and $\forall k \in \mathbb{N}$ as is customary, whenever this is unambiguous – and similarly in ${}^*\mathbb{N}$.

(**b**) In this proof and others below, the symbol $\to$ is used to mean logical implication (when occuring in a statement) *and* to indicate convergence of a sequence. There should be no confusion since the context makes it clear which meaning is intended. (A third common use of $\to$ is in indicating the domain and range of a function as in $f : A \to B$.)

Corollary 3.2 *If $s_n \to a$ and $t_n \to b$ then*

(**a**) *$s_n + t_n \to a + b$,*

(**b**) *$s_n t_n \to ab$,*

(**c**) *$s_n / t_n \to a/b$, provided that $b \neq 0$.*

Proof (**a**) For infinite K, we have

$${}^*s_K + {}^*t_K \approx a + b$$

using Proposition 2.9; proofs of (**b**) and (**c**) are similar, and are left as exercises. □

Further characterizations along the lines of Theorem 3.1 are gathered together in the next Theorem.

Theorem 3.3 *Let* (s_n) *be a real sequence and* $r \in \mathbb{R}$.

(a) $s_n \to \infty \iff {}^*s_K$ *is positive infinite for all infinite* K,

(b) (s_n) *is bounded* $\iff {}^*s_K$ *is finite for all infinite* K,

(c) r *is a limit point of* $(s_n) \iff r \approx {}^*s_K$ *for some infinite* K.

Proof (a) is similar to the proof of Theorem 3.1.

(b) If (s_n) is bounded by $b \in \mathbb{R}$ then the following holds in $\mathbb{R}$:

$$\forall k[|s_k| \leq b]$$

Its $*$-transform is

$$\forall K[|{}^*s_K| \leq b]$$

which holds in ${}^*\mathbb{R}$ and so in particular ${}^*s_K$ is finite for all infinite K.

Conversely, suppose that ${}^*s_K$ is finite for all infinite K. We already know that ${}^*s_k$ is finite for finite k (since ${}^*s_k = s_k$ for finite k). Thus, taking B to be any infinite hyperreal, the statement

$$\exists B\, \forall K[|{}^*s_K| \leq B]$$

holds in ${}^*\mathbb{R}$. This is the $*$-transform of the statement that (s_n) is bounded, which thus holds by the Transfer Principle.

(c) Exercise. □

A slight modification of the proof of Theorem 3.1 gives

Theorem 3.4 (s_n) *is Cauchy* $\iff {}^*s_K \approx {}^*s_M$ *for all infinite* K, M.

Proof Exercise □

Corollary 3.5 *A sequence is Cauchy if and only if it is convergent.*

Proof From Theorems 3.1 and 3.4 it is clear that a convergent sequence is Cauchy. Conversely, a Cauchy sequence is bounded (**Exercise**: Prove this from the above characterization of Cauchy sequences (Theorem 3.4)). So, putting $a = {}^\circ({}^*s_M)$ for any infinite M we see that $s_n \to a$. □

We leave it as an exercise to show that the above characterizations can be used to establish well known (standard) results about sequences, such as the following.

Exercise 3.6 Use the above characterizations to prove:

(a) Suppose that (s_n), (t_n) and (r_n) are real sequences such that $s_n \leq t_n \leq r_n$ for all $n \geq n_0$, and $s_n \to a$ and $r_n \to a$ as $n \to \infty$. Then $t_n \to a$ as $n \to \infty$.

(b) An increasing bounded sequence (s_n) is convergent.
(**Hint**: Let $a = {}^\circ{}^*s_N$ for some infinite N and then $s_n \leq a$ for all finite n. By transfer ${}^*s_M \leq a$ for *all* $M \in {}^*\mathbb{N}$.)

(c) Every bounded sequence has a limit point.

As a final exercise we have the following characterisation of the upper limit $\overline{\lim} s_n = \limsup s_n$ of a sequence – with obvious modification for the lower limit.

Exercise 3.7 Let (s_n) be a bounded real sequence and let $l = \overline{\lim} s_n$. Let $t_n = \sup_{m \geq n} s_m$, so that $l = \lim_{n\to\infty} t_n$. Show that

(a) $t_n \approx \max\{{}^*s_m : n \leq m \leq N\}$ for every finite n and infinite N,
(b) ${}^*t_N \approx l$ for every infinite N,
(c) for every infinite N there is $M \geq N$ with ${}^*s_M \approx l$;
(d) $t_n \geq {}^*s_N$ for all finite n and infinite N;
(e) $l \geq {}^{\circ *}s_N$ for all infinite N;
(f) $l = \max\{{}^{\circ *}s_N : N \text{ is infinite }\}$.

3.1. SERIES

The results above for sequences have natural counterparts for series, and we leave the reader to formulate these for the most part. First let us make a remark about notation.

We write $\sum_{n=1}^{K} {}^*s_n$ to denote ${}^*t_K$, where $t_m = \sum_{n=1}^{m} s_n$. Similarly $\sum_{n=M}^{K} {}^*s_n$ denotes ${}^*t_K - {}^*t_{M-1}$.

Then, translating results above for sequences we have:

Corollary 3.8 (a) $\sum_{n=1}^{\infty} s_n = s \iff \sum_{n=1}^{N} {}^*s_n \approx s$ *for all infinite* N.
(b) $\sum s_n$ *converges* $\iff \sum_{n=N}^{M} {}^*s_n \approx 0$ *for all infinite* M, N.
(c) *If* $\sum s_n$ *is convergent then* $s_n \to 0$ *as* $n \to \infty$.

4. Continuity

In this section we assume initially that f is a real function defined on an interval $I =]a, b[$. The ideas discussed can be extended easily to more general functions.

Continuity and differentiation are expressed in terms of limits, and so the fundamental result here is the following.

Theorem 4.1 *Let* $c \in]a, b[$ *and* $r \in \mathbb{R}$. *Then*

$$\lim_{x \to c} f(x) = r \quad \iff \quad {}^*f(z) \approx r \text{ for all } z \approx c \text{ in } {}^*\mathbb{R} \text{ with } z \neq c$$

Proof The proof is very similar to that of Theorem 3.1, but we will give it in full.

Suppose first that $f(x) \to r$ as $x \to c$, and fix a hyperreal $z \approx c$, $z \neq c$. We have to show that $|{}^*f(z) - r| < \varepsilon$ for all real $\varepsilon > 0$.

For any such ε there is a number $0 < \delta \in \mathbb{R}$ such that the following holds in $\mathbb{R}$:

$$\forall x[0 < |x - c| < \delta \to |f(x) - r| < \varepsilon]$$

The Transfer Principle now tells us that

$$\forall \mathrm{X}[0 < |\mathrm{X} - c| < \delta \to |{}^*f(\mathrm{X}) - r| < \varepsilon]$$

is true in ${}^*\mathbb{R}$. In particular taking $X = z$ we see that $|{}^*f(z) - r| < \varepsilon$ as required.

Conversely, suppose that $|{}^*f(z) - r| \approx 0$ for all $z \approx c$ in ${}^*\mathbb{R}$ with $z \neq c$. Let a real $\varepsilon > 0$ be given. Then taking Y to be any positive infinitesimal the following holds in ${}^*\mathbb{R}$:

$$\exists Y \forall \mathrm{X}[0 < |\mathrm{X} - c| < Y \to |{}^*f(\mathrm{X}) - r| < \varepsilon]$$

The Transfer Principle applied to this sentence gives, in $\mathbb{R}$:

$$\exists y \forall x[0 < |x - c| < y \to |f(x) - r| < \varepsilon]$$

Taking δ to be any such y shows that $f(x) \to r$ as required. □

From the usual definition of continuity (i.e. f is continuous at c if $f(x) \to f(c)$ as $x \to c$) we have the following infinitesimal characterization of continuity at a point.

Theorem 4.2 *Let $c \in]a, b[$. Then*

$$f \text{ is continuous at } c \iff {}^*f(z) \approx f(c) \text{ whenever } z \approx c \text{ in } {}^*\mathbb{R}.$$

Proof Exercise, using Theorem 4.1. □

From the above two results we can easily deduce the usual algebra of limits and continuity (as in Corollary 3.2 for sequences).

It is obvious how to adapt the above characterizations to right and left limits (i.e. $f(x) \to r$ as $x \to c+$ etc), and right and left continuity, so we can discuss continuity on a closed interval.

Exercise 4.3 Use the above characterization to show that the composition of continuous functions is continuous.

Let us now use the above characterization to give intuitive proofs of the Intermediate Value Theorem and the Extreme Value Theorem.

Theorem 4.4 (Extreme Value Theorem) *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous. Then f is bounded and attains its bounds.*

Proof Define a sequence $d_n = \max\{f(a+k(b-a)/n) : k = 0, 1, \cdots, n\} = f(c_n)$ say, where $a \le c_n \le b$. Fix an infinite $N \in {}^*\mathbb{N}$. Let $C = {}^*c_N$ and let $D = {}^*d_N = {}^*f(C)$. By transfer of appropriate statements (**Exercise:** Write down suitable statements for this) we have

$$a \le C \le b$$

and

$$^*f(a_K) \le D \quad \text{for all } K = 0, 1, \cdots, N \tag{1}$$

where $a_K = a + K(b-a)/N$.

Let $c = {}^\circ C$. Then by continuity $D = {}^*f(C) \approx f(c)$, so D is finite and putting $d = {}^\circ D$ we have $f(c) = d$.

We now show that $f(x) \le d$ for all $x \in [a,b]$. For any such x, by the Transfer Principle there is K with

$$a_K \le x \le a_{K+1}$$

so that $a_K \approx x \approx a_{K+1}$. By continuity $f(x) \approx {}^*f(a_K) \le D$ (by (1)) and so $f(x) \le d$.□

Theorem 4.5 (Intermediate Value Theorem) *Suppose that f is continuous on $[a,b]$ and $f(a) < 0 < f(b)$. Then there is $c \in [a,b]$ with $f(c) = 0$.*

Proof (Sketch) Fix an infinite $N \in {}^*\mathbb{N}$. By transfer of an appropriate statement there is a $K \in {}^*\mathbb{N}$, with $0 \le K \le N$ such that

$$^*f(a_K) < 0 \le {}^*f(a_{K+1})$$

(where $a_K = a+K(b-a)/N$ as in the proof of the Extreme Value Theorem). Let $C = a_K$ and let $c = {}^\circ C \approx a_K \approx a_{K+1}$. Then by continuity

$$f(c) \approx {}^*f(a_K) < 0 \le {}^*f(a_{K+1}) \approx f(c).$$

Taking standard parts gives $f(c) = 0$. □

Exercise 4.6 Suppose that $f : [a,b] \to \mathbb{R}$ is strictly increasing and maps $[a,b]$ onto $[f(a), f(b)]$. Use the infinitesimal characterization of continuity to show that f and f^{-1} are both continuous. (**Hint:** Suppose $x \approx c$ and $f(x) \not\approx f(c)$; consider $f^{-1}(r)$ where r is a *real* lying between $f(x)$ and $f(c)$.)

More General Domains Suppose that f is now a function with domain $A \subseteq \mathbb{R}$. Then it is routine to extend Theorem 4.2 to obtain:

Theorem 4.7 *Let $c \in A$. Then*

$$f \text{ is continuous on } A \text{ at } c \iff {}^*f(z) \approx f(c) \text{ whenever } z \in {}^*A \text{ and } z \approx c.$$

Proof Exercise. □

Uniform Continuity There is a pleasant characterization of uniform continuity of a function f on a set $A \subseteq \mathbb{R}$, which should be compared with the characterization above (Theorem 4.7) of continuity of f in A at each point of A.

Theorem 4.8 *Suppose that f is defined on the set A. The following are equivalent:*

(a) f is uniformly continuous on A,
(b) ${}^*f(y) \approx {}^*f(z)$ whenever $y \approx z$ and $y, z \in {}^*A$.

Proof Exercise. □

5. Differentiation

We again assume initially that f is a real function defined on an interval $I =]a, b[$.

Theorem 5.1 *Let $c \in I$. Then the following are equivalent:*

(a) *f is differentiable at c with derivative $D = Df(c) \in \mathbb{R}$,*
(b) $\dfrac{{}^*f(c+\delta) - f(c)}{\delta} \approx D$ *for all nonzero $\delta \approx 0$.*

Thus the following are equivalent:

(c) *f is differentiable at c,*
(d) *for all nonzero $\delta \approx 0$, $[{}^*f(c+\delta) - f(c)]/\delta = \Delta f(\delta)/\delta$, say, is finite and ${}^\circ(\Delta f(\delta)/\delta)$ is independent of δ.*

Proof This is an immediate application of Theorem 4.1 □

Remark This result can be rephrased by saying that f is differentiable at c with derivative $Df(c)$ if and only if for all $\delta \approx 0$

$$ {}^*f(c+\delta) = f(c) + \delta.Df(c) + \delta.\varepsilon(\delta) \tag{2}$$

where $\varepsilon(\delta) \approx 0$.

Proposition 5.2 *If f is differentiable at c then f is continuous at c.*

Proof Elementary from (2), using Theorem 4.2.□

Exercise 5.3 Use Theorem 5.1 to establish the algebra of differentiability; i.e. if f, g are differentiable at c then so are $f+g$, $f.g$ and f/g (provided $Dg(c) \neq 0$ and

(a) $D(f+g)(c) = Df(c) + Dg(c)$
(b) $D(f.g)(c) = Df(c).g(c) + f(c).Dg(c)$
(c) $D(f/g)(c) = [Df(c).g(c) - f(c).Dg(c)]/g^2(c)$

For the following, Exercise 4.6 may be helpful.

Exercise 5.4 Suppose that f is continuous and strictly increasing on $[a, b]$ and f is differentiable at $c \in]a, b[$ with $Df(c) \neq 0$. Let $g = f^{-1}$ (defined on the interval $[f(a), f(b)]$). Show that g is differentiable at $f(c)$ with derivative $Dg(f(c)) = Df(c)^{-1}$.

The above infinitesimal characterization of differentiability (Theorem 5.1) gives a straightforward proof of the chain rule.

Theorem 5.5 (Chain Rule) *Let f be differentiable at c and suppose that g is differentiable at $f(c)$. Then $g \circ f$ is differentiable at c and*

$$D(g \circ f)(c) = Dg(f(c)).Df(c)$$

Proof Let $h = g \circ f$.We have to prove that for all nonzero $\delta \approx 0$

$$\frac{{}^*h(c+\delta) - h(c)}{\delta} \approx Dg(f(c)).Df(c) \tag{3}$$

There are two cases.

Case (i). ${}^*f(c+\delta) = f(c)$. Then from Theorem 5.1, $Df(c) = 0$ and both sides of (3) are equal (to 0).

Case (ii). ${}^*f(c+\delta) \neq f(c)$. Putting $\Delta = {}^*f(c+\delta) - f(c)$, which is infinitesimal (by Proposition 5.2), we have

$$\begin{aligned} \frac{{}^*h(c+\delta) - h(c)}{\delta} &= \frac{{}^*g({}^*f(c+\delta)) - g(f(c))}{\delta} \\ &= \frac{{}^*g(f(c)+\Delta) - g(f(c))}{\Delta} \cdot \frac{{}^*f(c+\delta) - f(c)}{\delta} \\ &\approx Dg(f(c)).Df(c) \end{aligned}$$

using Theorem 5.1 for g at $f(c)$ and f at c. □

To conclude this section we prove the following basic result using our intuitive ideas about derivatives.

Theorem 5.6 *Suppose that f is defined and differentiable on $I =]a,b[$ and achieves a (local) maximum at $c \in I$. Then $Df(c) = 0$.*

Proof Taking an infinitesimal $\delta > 0$ we have

$$Df(c) \approx \frac{{}^*f(c+\delta) - f(c)}{\delta} \leq 0$$

since ${}^*f(c+\delta) - f(c) \leq 0$. So $Df(c) \leq 0$. Similarly, taking $0 \approx \delta < 0$ we see that $Df(c) \geq 0$. □

Rolle's Theorem and the Mean Value Theorem now follow easily in the usual (standard) way.

6. Riemann Integration

To give a characterization of the Riemann integral for continuous functions using infinitesimals it is necessary to first develop some notation. Suppose that f is defined and continuous on a closed interval [a,b]. For any real number $0 < d < b - a$ define the *partition based on d* as follows.

$$a = a_0 < a_1 < \cdots < a_n = b$$

where $a_k - a_{k-1} = d$ for $k = 1, \cdots, n-1$ and $a_n - a_{n-1} \leq d$. For such a partition define the upper and lower Riemann sums in the usual way:

$$\overline{S}_a^b(f,d) = \sum_{k=0}^{n-1} M_k(a_{k+1} - a_k)$$

$$\underline{S}_a^b(f,d) = \sum_{k=0}^{n-1} m_k(a_{k+1} - a_k)$$

where M_k and m_k are the maximum and minimum of f on $[a_k, a_{k+1}]$. By considering common refinements, it is clear that for any d, d' we have

$$\underline{S}_a^b(f,d) \leq \overline{S}_a^b(f,d') \tag{4}$$

and the same holds for more general partitions of $[a,b]$.

We now observe the following.

Theorem 6.1 *If f is continuous on $[a,b]$ and $0 < \delta \approx 0$ then*

$${}^*\underline{S}_a^b(f,\delta) \approx {}^*\overline{S}_a^b(f,\delta)$$

Proof For any real $d > 0$ it is routine to see that

$$0 \leq \overline{S}_a^b(f,d) - \underline{S}_a^b(f,d) \leq K(b-a)$$

where $K = \max\{M_k - m_k : k = 0, 1, \cdots, n-1\}$. Of course K depends on d, so we write $K = K(d)$; and there are points $x, y \in [a,b]$ such that $|x-y| \leq d$ and $K(d) = |f(x) - f(y)|$.

By transfer of an appropriate sentence (**Exercise:** Write down such sentence) we see that for infinitesimal $\delta > 0$

$$0 \leq {}^*\overline{S}_a^b(f,\delta) - {}^*\underline{S}_a^b(f,\delta) \leq {}^*K(\delta)(b-a) \tag{5}$$

with ${}^*K(\delta) = |{}^*f(x) - {}^*f(y)|$ for some x, y with $|x-y| \leq \delta$. This means that ${}^*K(\delta) \approx 0$, so the result follows from (5).□

If we now define the Riemann sum (corresponding to the partition based on d) by

$$S_a^b(f,d) = \sum_{k=0}^{n-1} f(a_k)(a_{k+1} - a_k)$$

then it is clear that

$$\underline{S}_a^b(f,d) \leq S_a^b(f,d) \leq \overline{S}_a^b(f,d) \tag{6}$$

for all d. Thus we have

Corollary 6.2 *If f is continuous on $[a,b]$ then f is Riemann integrable and*

$$\int_a^b f = {}^\circ\left({}^*S_a^b(f,\delta)\right) = {}^\circ\left({}^*\underline{S}_a^b(f,\delta)\right) = {}^\circ\left({}^*\overline{S}_a^b(f,\delta)\right) \tag{7}$$

for any infinitesimal $\delta > 0$.

Proof Theorem 6.1 and the transfer of (4) gives that for any nonzero infinitesimals $\delta, \delta' > 0$ we have

$${}^*\underline{S}_a^b(f,\delta) \leq {}^*\overline{S}_a^b(f,\delta') \approx {}^*\underline{S}_a^b(f,\delta') \leq {}^*\overline{S}_a^b(f,\delta) \approx {}^*\underline{S}_a^b(f,\delta)$$

and so

$${}^*\underline{S}_a^b(f,\delta) \approx {}^*\underline{S}_a^b(f,\delta')$$

Using Theorem 4.1, (6) and Theorem 6.1 this shows that as $d \to 0+$ the sums $\underline{S}_a^b(f,d)$, $\overline{S}_a^b(f,d)$ and $S_a^b(f,d)$ have a common limit L, say, where

$$L \approx {}^*\underline{S}_a^b(f,\delta) \approx {}^*\overline{S}_a^b(f,\delta) \approx {}^*S_a^b(f,\delta)$$

for any nonzero $\delta \approx 0$. So f is Riemann integrable with $L = \int_a^b f$ given by (7). □

The elementary properties of the Riemann integral follow – we leave these as an exercise.

We will sketch the proof of the following theorem.

Theorem 6.3 *Suppose that f is continuous on $[a,b]$ and let $F(x) = \int_a^x f$ for $x \in [a,b]$. Then F is differentiable on $[a,b]$ with $DF(c) = f(c)$ for each $c \in [a,b]$. (For $c = a$ or b we mean the right or left derivative, defined and characterized in the obvious way.)*

Proof From the above characterization of the integral it is easy to see that for any $a_1, a_2 \in [a,b]$ with $a_1 < a_2$ we have

$$m(a_2 - a_1) \le \int_{a_1}^{a_2} f = F(a_2) - F(a_1) \le M(a_2 - a_1)$$

where m, M are respectively the minimum and maximum values of f on $[a_1, a_2]$. From the Extreme Value Theorem we know that $m = f(y)$ and $M = f(z)$ for some $y, z \in [a_1, a_2]$.

Consider now $c \in]a,b[$ and nonzero $\delta \approx 0$. By transfer of an appropriate sentence (**Exercise:** Formulate such a sentence!) we see that

$$^*f(y).\delta \le {}^*F(c+\delta) - F(c) \le {}^*f(z).\delta \tag{8}$$

for some y, z between c and $c + \delta$ (note that δ may be negative). Dividing (8) by δ we have

$$^*f(y) \le \frac{^*F(c+\delta) - F(c)}{\delta} \le {}^*f(z)$$

if $\delta > 0$, and the reverse inequality if $\delta < 0$. But $^*f(y) \approx f(c) \approx {}^*f(z)$ and so in either case

$$\frac{^*F(c+\delta) - F(c)}{\delta} \approx f(c)$$

as required. □

We will conclude this section with an infinitesimal proof of the Fundamental Theorem of the Calculus.

Theorem 6.4 (Fundamental Theorem of the Calculus) *Suppose that F defined on $[a,b]$ has a continuous derivative $DF = f$ say. Then*

$$\int_a^b f = F(b) - F(a)$$

Proof Take positive d and the partition $a = a_0 < a_1 < \cdots a_n = b$ based on d as above. For each k, by the Mean Value Theorem, there is $x_k \in [a_{k+1}, a_k]$ such that $F(a_{k+1}) - F(a_k) = (a_{k+1} - a_k).f(x_k)$. So, using the earlier notation, we have

$$(a_{k+1} - a_k).m_k \leq F(a_{k+1}) - F(a_k) \leq (a_{k+1} - a_k).M_k$$

for each k. Summing over k this gives

$$\underline{S}_a^b(f,d) \leq F(b) - F(a) \leq \overline{S}_a^b(f,d).$$

Now by transfer this holds for any positive infinitesimal δ in place of d. Thus, by Corollary 6.2 we see that $\int_a^b f = F(b) - F(a)$ as required.□

7. Topology on $\mathbb{R}$

Let us now briefly indicate how topological notions can be characterized using the nonstandard extension ${}^*\mathbb{R}$. This will pave the way to a more general treatment of topologies, including non-metric topologies, in Peter Loeb's article in this volume [8].

Theorem 7.1 *Let $A \subseteq \mathbb{R}$. Then*

(a) *A is open $\Longleftrightarrow$ $\mathrm{monad}(a) \subseteq {}^*A$ for every $a \in A$;*
(b) *The following are equivalent:*
(i) A is closed;
(ii) for all $a \in \mathbb{R}$, if $\mathrm{monad}(a) \cap {}^*A \neq \emptyset$ then $a \in A$;
(iii) $\mathrm{monad}(a) \cap {}^*A = \emptyset$ for all $a \notin A$;
(iv) $A = \mathrm{st}({}^*A)$
where $\mathrm{st}(X) = \{{}^\circ x : x \in X \cap \mathrm{ns}({}^*\mathbb{R})\}$ for any set $X \subseteq {}^*\mathbb{R}$.

Proof Exercise. (The proof has some similarities with the proof of Theorem 4.2.) □

Remark The set $\mathrm{st}(X) = \{{}^\circ x : x \in X \cap \mathrm{ns}({}^*\mathbb{R})\}$ is called the *standard part* of the set X.

One of the most useful results of nonstandard analysis is the following, and its extension to more general topological settings.

Theorem 7.2 *Let $A \subseteq \mathbb{R}$. Then A is compact $\Longleftrightarrow$ every $z \in {}^*A$ is finite (i.e.* nearstandard*) and has ${}^\circ z \in A$.*

Proof Suppose first that the condition fails for some $z \in {}^*A$. There are two cases. If z is infinite, then for every finite $n \in \mathbb{N}$ we have

$$\exists X \in {}^*A[|X| > n]$$

is true in $^*\mathbb{R}$. So by transfer

$$\exists x \in A[|x| > n]$$

Thus A is not bounded, and so is not compact (the cover by the intervals $]-n, n[$ $(n \in \mathbb{N})$ has no finite subcover).

If, on the other hand, there is $z \in {}^*A$ that is finite but $c = {}^\circ z \notin A$, then z demonstrates that for each $n \in \mathbb{N}$,

$$\exists X \in {}^*A[|X - c| < n^{-1}]$$

holds in $^*\mathbb{R}$. For each $n \in \mathbb{N}$, this is the transfer of the statement that for each $n \in \mathbb{N}$,

$$A \not\subseteq\ ]-\infty, c - n^{-1}[\ \cup\ ]c + n^{-1}, \infty[\ = A_n$$

say. The family $(A_n)_{n\in\mathbb{N}}$ is thus an open cover of A that has no finite subcover – again showing that A is not compact.

Conversely, suppose that A is not compact. Then there is a countable family of intervals $I_n =]a_n, b_n[$, say, that covers A but for which there is no finite subcover.

For each $n \in \mathbb{N}$ choose $z_n \in A$ with $z_n \notin I_m$ for all $m \leq n$, and let $z = z_N$ for any (fixed) infinite $N \in {}^*\mathbb{N}$. By transfer, $z \in {}^*A$; and we also have

$$\forall M \leq N[z \leq a_M \vee b_M \leq z] \tag{9}$$

If z is infinite, we are done. Otherwise z is finite, and let $c = {}^\circ z$. We must show that c $\notin A$. If, to the contrary, c $\in A$ then c $\in I_n$ for some $n \in \mathbb{N}$; i.e. $a_n <$ c $< b_n$. But then $a_n < z < b_n$, contradicting (9). So indeed $c \notin A$. □

Exercise 7.3 Use the above characterization of compactness to show that:

(a) For any a, b the closed interval $[a, b]$ is compact but $]a, b[$ is not compact.
(b) $\mathbb{R}$ is not compact.
(c) If $K \subset \mathbb{R}$ is compact and f is continuous then $f(K)$ is compact and f is uniformly continuous on K. (Recall the characterization of uniform continuity Theorem 4.8.)

It is an instructive exercise to prove the Heine–Borel Theorem using infinitesimal characterizations of compact, closed and bounded:

Exercise 7.4 Let $A \subseteq \mathbb{R}$.

(a) Prove that A is bounded $\iff {}^*A \subseteq \mathrm{Fin}({}^*\mathbb{R})$).
(b) **(Heine–Borel)** Show that A is compact $\iff A$ is closed and bounded.

8. Using Internal Subsets of ${}^*\mathbb{R}$

So far we have deliberately developed the beginnings of classical calculus using *only* the nonstandard extension ${}^*\mathbb{R}$ and the Transfer Principle. For the rest of this section and the next we illustrate the extra facility that is available if we work with a nonstandard extension $({}^*\mathbb{R}, {}^*\mathcal{P}(\mathbb{R}), {}^*\mathcal{P}(\mathbb{R}\times\mathbb{R}))$ of $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}\times\mathbb{R}))$ – as discussed in Section 5 of Henson's paper [6] (see particularly Exercise 5.7). Note that ${}^*\mathcal{P}(\mathbb{R}) \subset \mathcal{P}({}^*\mathbb{R})$ – and these are **NOT** equal; similarly we regard ${}^*\mathcal{P}(\mathbb{R}\times\mathbb{R}) \subset \mathcal{P}({}^*\mathbb{R}\times{}^*\mathbb{R})$ and again we have inequality.

Recall the very important definitions of *internal subset* of ${}^*\mathbb{R}$ and *internal function* on ${}^*\mathbb{R}$.

Definition 8.1 A subset X of ${}^*\mathbb{R}$ is *internal* if it is an element of ${}^*\mathcal{P}(\mathbb{R})$. A function $F : X \to {}^*\mathbb{R}$ is *internal* if its graph is an element of ${}^*\mathcal{P}(\mathbb{R}\times\mathbb{R})$.

Internal sets and functions are important because, as members of ${}^*\mathcal{P}(\mathbb{R})$ or ${}^*\mathcal{P}(\mathbb{R}\times\mathbb{R})$, we know how they behave from the more general Transfer Principle ([6], Theorem 4.27). This is exemplified in the proof of the following extremely useful properties.

Proposition 8.2 *Let $A \subseteq {}^*\mathbb{R}$ be an internal set.*

(a) **(Overflow)** *If A contains arbitrarily large finite positive numbers, then it also contains an infinite number.*

(b) **(Underflow)** *If A contains arbitrarily small positive infinite numbers, then it also contains a finite number.*

Proof (a) If A is unbounded in ${}^*\mathbb{R}$ then the assertion is obvious. Otherwise, let $X = \{m \in {}^*\mathbb{N} : m \text{ is an upper bound of } A\}$. Then A is a nonempty internal subset of ${}^*\mathbb{N}$, with a least element N, say. (This follows by transfer, which tells us that properties of subsets of $\mathbb{N}$ are inherited by *internal* subsets of ${}^*\mathbb{N}$.) From the hypothesis, N must be infinite. So $N - 1$ is also infinite and is *not* an upper bound of A. Thus there is $a \in A$ with $a > N-1$ and so a is infinite.

(b) If A does not contain a finite number, simply apply **(a)** to the set of lower bounds of A^+, where A^+ is the set of positive elements of A. □

Remark Overflow and underflow are also known as the *overspill* and *underspill* principles. There are, of course, obvious versions for negative numbers.

As an immediate consequence of this Proposition we have

Corollary 8.3 (Infinitesimal Overflow and Underflow) *Let $A \subseteq {}^*\mathbb{R}$ be an internal set. If A contains arbitrary large infinitesimal numbers, then it also contains a non-infinitesimal number. If A contains arbitrarily small* *non-infinitesimal positive numbers, then it also contains a (positive) infinitesimal number.*

Proof Apply Proposition 8.2 to the set $\{x^{-1} : x \in A\}$.□

The next result, which we prove here using overflow, has a generalization to general topological spaces, and is extremely useful in constructions in analysis – for example in proving existence results. First recall the notation $\mathrm{st}(X)$ introduced above for $\{{}^\circ x : x \in X \cap \mathrm{ns}({}^*\mathbb{R})\}$, the standard part of the set X.

Theorem 8.4 *Suppose that $X \subseteq {}^*\mathbb{R}$ is internal and $B \subseteq \mathbb{R}$. Then*

(a) $\mathrm{st}(X)$ *is closed,*
(b) *if* $X \subseteq \mathrm{Fin}({}^*\mathbb{R})$ *then* $\mathrm{st}(X)$ *is compact,*
(c) $\mathrm{st}({}^*B) = \overline{B}$ *(the closure of B),*

Proof Let $A = \mathrm{st}(X)$.

(a) Suppose that $(a_n)_{n\in\mathbb{N}}$ is a sequence in A and $\lim_{n\to\infty} a_n = r$. We may assume that $|a_n - r| < n^{-1}$. We have to show that $r \in A$.

For each $n \in \mathbb{N}$ there is $x_n \in X$ with $x_n \approx a_n$, so $|x_n - r| < n^{-1}$. Now consider the set

$$Y = \{n \in {}^*\mathbb{N} : \exists x \in X[|x - r| < n^{-1}\}$$

This set is internal (by the Internal Definition Principle) and contains all finite natural numbers. So by overflow there is an infinite $N \in Y$, giving an $x \in X$ such that $|x - r| < N^{-1}$. Clearly $r = {}^\circ x$ and so $r \in A$.

(b) From **(a)**, A is closed. By overflow (Proposition 8.2) there is a finite $n \in \mathbb{N}$ such that $X \subset {}^*[-n, n]$ (otherwise X contains infinite numbers). Thus A is bounded, and the result follows immediately from the Heine–Borel theorem (Exercise 7.4).

Alternatively, we can give a direct proof as follows. Suppose that A is not compact. Then there is a countable family of intervals $I_n =]a_n, b_n[$, say, that covers A but for which there is no finite subcover.

For each $n \in \mathbb{N}$ choose $c_n \in A$ with $c_n \notin I_m$ for all $m \leq n$. For each n there is $x_n \in X$ with $x_n \approx c_n$ and so

$$\forall m \leq n[x_n \leq a_m + \frac{1}{n} \vee b_m - \frac{1}{n} \leq x_n] \tag{10}$$

Consider now the set

$$Y = \{n \in {}^*\mathbb{N} : \exists x \in X \forall m \leq n[x \leq a_m + \frac{1}{n} \vee b_m - \frac{1}{n} \leq x]\}$$

which is internal by the Internal Definition Principle (Henson [6], Theorem 5.4). From (10) we know that Y contains all finite natural numbers. So by overflow there is an infinite number $N \in Y$, giving $x \in X$ with

$$x \leq a_m + \frac{1}{N} \text{ or } b_m - \frac{1}{N} \leq x \text{ for all } m \leq N \tag{11}$$

Putting $c = {}^\circ x \in A$ we must have $c \in I_m$ for some finite m. On the other hand, (11) means that $c \leq a_m$ or $b_m \leq c$ – i.e. $c \notin I_m$. This contradiction establishes the result.

(c) Exercise (Use **(a)** together with Theorem 7.1(b) and the characterization of $\overline{B}$ as the smallest closed set containing B.) □

Another useful consequence of overflow is

Lemma 8.5 (Robinson's Sequential Lemma) *Let $(x_n)_{n\in {}^*\mathbb{N}}$ be an internal sequence of hyperreals. If $x_n \approx 0$ for all finite n, then there is infinite N such that $x_m \approx 0$ for all $m \leq N$.*

Proof Let $A = \{n \in {}^*\mathbb{N} : x_m < m^{-1} \text{ for all } m \leq n\}$. This set is internal and contains arbitrarily large finite numbers; hence by overflow it contains an infinite number. (Note that the naïve attempt to consider $A = \{n \in {}^*\mathbb{N} : x_n \approx 0\}$ does not work since A is not necessarily internal.)□

The following is an exercise in the use of Robinson's Lemma.

Exercise 8.6 Let $(x_n)_{n\in {}^*\mathbb{N}}$ be an internal sequence such that x_n is finite for all finite n. Let $a_n = {}^\circ x_n$. Show that

(a) $\lim_{n\to\infty} a_n = a \iff x_N \approx a$ for all sufficiently small infinite N;
(By *sufficiently small infinite* N we mean that there is an infinite K such that this holds for all infinite $M \leq N$.)

(b) (a_n) is convergent $\iff x_N \approx x_M$ for all sufficiently small infinite N, M.

We now turn to an important notion known as *S-continuity.*

Definition 8.7 An internal function

$$G : {}^*\mathbb{R} \to {}^*\mathbb{R}$$

is S-*continuous* on a set $X \subseteq {}^*\mathbb{R}$ if for all $x,\ y \in X$

$$x \approx y \Longrightarrow G(x) \approx G(y).$$

Examples (1) If $G(x) = \int_0^x F(\tau)d\tau$ where F is *integrable (for the integral to make sense) and bounded by a finite constant, $|F(\tau)| \leq c < \infty$, then G is S-continuous since

$$|G(x) - G(y)| \leq c|x - y|$$

by the transfer of properties of the integral. The condition that is really necessary here is

$$\int_x^y F(\tau)d\tau \approx 0 \quad \text{for} \quad x \approx y.$$

(2) Let $F(x) = \sin Nx$ with $N \in {}^*\mathbb{N}\backslash\mathbb{N}$. Then F is *continuous but is not S-continuous. For instance $\frac{\pi}{2N} \approx 0$ but $F(\frac{\pi}{2N}) = 1$, $F(0) = 0$.

Remarks We used the notions *integrable and *continuous above, to mean the *-extensions of the sets $\mathcal{C}$ and $\mathcal{L}$ of continuous and integrable real functions (both subsets of $\mathcal{P}(\mathbb{R}\times\mathbb{R})$). Integration is a function (or operator) $I : \mathcal{L} \to \mathcal{C}$, so the integral G in example (1) above is most easily defined as $G = {}^*I(F)$.

Theorem 8.8 *If $F : {}^*\mathbb{R} \to {}^*\mathbb{R}$ is* S-*continuous on an interval* ${}^*[a, b]$ *where* a, $b \in \mathbb{R}$ *and* $F(x_0)$ *is finite for some* $x_0 \in {}^*[a, b]$, *then* $F(x)$ *is finite for all* $x \in {}^*[a, b]$ *and the standard function defined by*

$$f(x) = {}^\circ F(x) \quad \text{for} \quad x \in [a, b]$$

is continuous.

Proof Take any $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$. Consider the set

$$A = \{\alpha \in {}^*\mathbb{R} : \alpha > 0, |x-y| < \alpha \Longrightarrow |F(x)-F(y)| < \varepsilon \text{ for all } a \le x, y \le b\}$$

which is internal by the Internal Definition Principle. By S-continuity of F the set A contains all positive infinitesimals; hence by Proposition 8.2, A contains a real number δ.

Thus

$$|x - y| < \delta \Longrightarrow |F(x) - F(y)| < \varepsilon$$

which proves that F is finite for all $x \in {}^*[a, b]$, and that f is continuous. □

Remark The function f as defined in this Theorem can also be defined (or described) by

$$f({}^\circ x) = {}^\circ F(x) \quad \text{for all} \quad x \in {}^*[a, b].$$

(In a sense that can be made precise (see Loeb's paper [8]) f is the *standard part* of F in the uniform topology.)

Using internal sets and functions we can usefully extend the characterization of the Riemann Integral (Theorem 6.1) as follows.

Theorem 8.9 *Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and suppose that*

(i) $F : {}^*[a, b] \to {}^*\mathbb{R}$ is internal and

$$F(x) \approx {}^*f(x)$$

for all $x \in {}^*[a, b]$ (hence F is S-continuous, with ${}^\circ F = f$);

(ii) $(a_k)_{0 \le k \le N}$ is an infinite internal partition [1] of ${}^*[a, b]$ with $a_0 = a$ and $a_N = b$, such that $|a_{k+1} - a_k|$ is infinitesimal for all k;

(iii) $(x_k)_{0 \le k < N}$ is an internal sequence with $a_k \le x_k \le a_{k+1}$ for each k.

Then

$$\int_a^b f = {}^\circ\left(\sum_{k=0}^{N-1} F(x_k)\Delta_k\right) \tag{12}$$

where $\Delta_k = a_{k+1} - a_k$.

Proof Suppose that P is any partition of $[a, b]$. We extend the notation used earlier by writing $\overline{S}_a^b(f, P)$ and $\underline{S}_a^b(f, P)$ for the upper and lower sums formed from the partition P. Elementary considerations show that if a partition Q is a refinement of P we have

$$\underline{S}_a^b(f, P) \le \underline{S}_a^b(f, Q) \le \overline{S}_a^b(f, Q) \le \overline{S}_a^b(f, P) \tag{13}$$

and so by taking a common refinement we have

$$\underline{S}_a^b(f, P) \le \overline{S}_a^b(f, P') \tag{14}$$

for any partitions P, P'.

Now let P be the internal partition as indicated in the statement of the theorem. Let $S = \sum_{k=0}^{N-1} F(x_k)\Delta_k$, the sum on the right in the theorem.

Since $F(x_k) \approx {}^*f(x_k)$ for all k, there is an infinitesimal ε such that $|F(x_k) - {}^*f(x_k)| \le \varepsilon$ for all k. (**Exercise:** Prove this.) Thus, putting

$$T = \sum_{k=0}^{N-1} {}^*f(x_k)\Delta_k$$

[1] By this we mean that N is an infinite hypernatural number, and there is an internal function $\alpha : \{0, 1, \ldots, N\} \to {}^*[a, b]$ with $a_k = \alpha(k)$ for each k.

we have, by transfer of standard properties of sums

$$\begin{aligned}|S-T| &= \Big|\sum_{k=0}^{N-1} F(x_k)\Delta_k - \sum_{k=0}^{N-1} {}^*\!f(x_k)\Delta_k\Big| \\ &= \Big|\sum_{k=0}^{N-1} \Big(F(x_k) - {}^*\!f(x_k)\Big)\Delta_k\Big| \\ &\le \varepsilon \sum_{k=0}^{N-1} \Delta_k = \varepsilon(b-a) \approx 0.\end{aligned}$$

Now, by transfer of further statements true in $\mathbb{R}$ we have

$$ {}^*\underline{S}_a^b({}^*\!f, P) \le T \le {}^*\overline{S}_a^b({}^*\!f, P) \tag{15}$$

and by the same arguments as in the proof of Theorem 6.1 and Corollary 6.2, using (14) with P' as any partition based on a non-zero infinitesimal δ, we see that

$$ {}^*\underline{S}_a^b({}^*\!f, P) \approx {}^*\overline{S}_a^b({}^*\!f, P) \approx {}^*\underline{S}_a^b({}^*\!f, \delta) \approx \int_a^b f.$$

Combining this with (15) and since $S \approx T$ we have the result. □

Discussion In the statement of the above result, we did not explain carefully the meaning of the internal sum on the right in (12). This sum has the form

$$\sum_{k=0}^{M} G(k)$$

where G is an internal sequence $G : {}^*\mathbb{N} \to {}^*\mathbb{R}$. A natural way to make sense of this in a general way is to regard $\sum$ here as the nonstandard extension of the standard operator $\sum : \mathcal{S} \times \mathbb{N} \to \mathbb{R}$ such that

$$\sum(g, m) = \sum_{k=0}^{m} g(k)$$

(where we are writing $\mathcal{S}$ for the set of real sequences). In terms of the set theoretic hierarchy we have

$$\sum \subset \mathcal{P}(\mathbb{R}\times\mathbb{R}) \times (\mathbb{R}\times\mathbb{R})$$

and so to discuss $\sum$ and its nonstandard extension ${}^*\sum$ we need the multiset $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}\times\mathbb{R}))$ and a nonstandard extension of it that we have referred to above. An alternative (and mathematically equivalent) way to define the sum $\sum_{k=0}^{M} G(k)$ is discussed in Henson's article ([6] Remark 5.8).

Exercise 8.10 Show that the above definition of $\sum_{k=0}^{M} G(k)$ is equivalent to that of Henson's Remark 5.8.

In the next section we will need the following slight extension of Theorem 8.9.

Corollary 8.11 *The conclusion of Theorem 8.9 holds if the internal sequence* (a_k) *has* $a \leq a_0$ *and* $a_N \leq b$ *with* $a \approx a_0$ *and* $a_N \approx b$.

Proof The difference in the sums involved is infinitesimal.□

9. An Application: Differential Equations

We conclude with an application to the solution of ordinary differential equations. The example we give is almost the simplest possible, but it is a prototype for much more general equations – including partial differential equations (PDEs), stochastic differential equations (SDEs) and even stochastic PDEs.

Consider the differential equation

$$\begin{cases} \dfrac{dy}{dx} &= g(x, y(x)) \qquad (0 \leq x \leq 1) \\ y(0) &= a \end{cases} \tag{16}$$

where g is continuous in both variables and bounded. Then we have

Theorem 9.1 (Peano's Existence Theorem) *The equation* (16) *has a solution; i.e. there is a differentiable function* $y(x)$ *on* $[0, 1]$ *such that* (16) *holds for all* $x \in [0, 1]$ *(where we are writing* dy/dx *for* $Dy(x)$*).*

Proof Pick an infinite $N \in {}^*\mathbb{N}$ and let $\Delta x = N^{-1}$. Let $x_k = k\Delta x$ for $k = 0, 1, \cdots, N$ and let $\mathbf{X} = \{0 = x_0, x_1, x_2, \cdots, x_N = 1\}$. Define an internal function $Y : \mathbf{X} \to {}^*\mathbb{R}$ by

$$\begin{cases} Y(0) &= a \\ Y(x_{k+1}) &= Y(x_k) + {}^*g(x_k, Y(x_k))\Delta x \end{cases}$$

and extend Y to ${}^*[0, 1]$ by making it linear on each $[x_k, x_{k+1}]$.

(It is worth thinking hard about why there is a function Y satisfying the above property and why it is internal – see the exercise below.)

If g is bounded by $B \in \mathbb{R}$ then it easy to see that for all x, z we have $|Y(x) - Y(z)| \leq B|x - z|$. Hence Y is S-continuous, and by Theorem 8.8 we can define the continuous function $y : [0, 1] \to \mathbb{R}$ by $y(°x) = °Y(x)$.

Now let $f(x) = g(x, y(x))$, which is continuous (since g and y are), and let $F(x) = {}^*g(x, Y(x))$. Then since g is continuous we have

$$F(x) \approx g(°x, °Y(x)) = g(°x, y(°x)) = f(°x) \approx {}^*f(x)$$

for all $x \in {}^*[0,1]$. Thus we can apply Corollary 8.11 as follows.

Let $b \in [0,1]$ and let $x_M \leq b \leq x_{M+1}$, so that $x_M \approx b \approx x_{M+1}$. From the definition of Y we have

$$\begin{aligned} y(b) \approx Y(x_M) &= a + \sum_{k=0}^{M-1} F(x_k)\Delta x \\ &\approx a + \int_0^b f(x)dx \end{aligned}$$

by Corollary 8.11. That is

$$y(b) = a + \int_0^b g(x, y(x))dx$$

which proves that y is differentiable and solves the equation. □

Exercise 9.2 Why there is a function F as in the above proof, and why is it internal? (**Hint:** Consider the corresponding construction for a (finite) $n \in \mathbb{N}$. Is there a statement that can be transferred? Alternatively, consider the standard function $Y(n,x) = Y_n(x)$, say, where Y_n is the standard function defined for $\Delta x = n^{-1}$.)

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., and Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Capiński, M. and Cutland, N.J., (1995) *Nonstandard Methods for Stochastic Fluid Mechanics.* World Scientific, Singapore.
3. Cutland, N.J. (Editor), (1988) *Nonstandard Analysis and its Applications.* Cambridge University Press, Cambridge.
4. Davis, M., (1977) *Applied Nonstandard Analysis.* John Wiley & Sons, New York.
5. Henle, J.M. and Kleinberg, E.M., (1979) *Infinitesimal Calculus.* MIT Press, Cambridge, Massachusetts.
6. Henson, C.W., (1997) Foundations of nonstandard analysis: a gentle introduction to nonstandard extensions, *this volume.*
7. Hurd, A. and Loeb, P.A., (1985) *An Introduction to Nonstandard Real Analysis.* Academic Press, New York.
8. Loeb, P.A., (1997) Nonstandard analysis and topology, *this volume.*
9. Keisler, H.J., (1976) *Foundations of Infinitesimal Calculus.* Prindle, Weber & Schmidt, Boston.
10. Lindstrøm, T., (1988) An invitation to nonstandard analysis, in Cutland (1988), pp. 1–105.
11. Luxemburg, W.A.J.,(1969a) *Applications of Model Theory to Algebra, Analysis, and Probability.*, Holt, Rinehart and Winston, New York.
12. Luxemburg, W.A.J., (1969b) A general theory of monads, in Luxemburg (1969a), pp. 18–86.
13. Robinson, A., (1966) *Nonstandard Analysis.* North-Holland, Amsterdam. (Second, revised edition, 1974).
14. Stroyan, K. and Luxemburg, W.A.J., (1976) *Introduction to the Theory of Infinitesimals.* Academic Press, New York.

NONSTANDARD ANALYSIS AND TOPOLOGY

PETER A. LOEB
University of Illinois at Urbana-Champaign
Department of Mathematics
1409 West Green Street
Urbana, Illinois 61801
USA
email: loeb@math.uiuc.edu

1. Metric and Topological Spaces

In what follows, I will use the phrase "In a nonstandard extension" to mean we employ a superstructure which is a nonstandard extension of the structure with which we are working. That structure will at least contain the real numbers. We will write ${}^*\mathbb{N}_\infty$ for ${}^*\mathbb{N} \setminus \mathbb{N}$.

We want to generalize the notion of "closeness" that we have for the real and complex number system. The first generalization is given by the notion of a metric space $\langle X, \rho \rangle$ where X is a nonempty set of elements we think of as points and $\rho : X \times X \to \mathbb{R}^+$ is a function we think of as a distance. For example, we can use the absolute value on $\mathbb{R}$ or a norm $\|\cdot\|$ on a normed space, and set $\rho(x, y) = |x - y|$ or $\rho(x, y) = \|x - y\|$. By an open ball with center x and radius $r > 0$, we mean the set

$$B(x, r) = \{y \in X : \rho(x, y) < r\}.$$

Notice that for a given point y, the more of these balls centered at x which also contain y, the closer y is to x.

In a nonstandard extension, we define the monad of a point $x \in X$ by setting

$$\text{monad}(x) = \mu(x) := \bigcap {}^*B(x, r) = \{y \in {}^*X : {}^*\rho(x, y) \simeq 0\},$$

where the intersection is over all positive standard values of r. We use this monad in the same way we use monads on the real line. For example, a standard set O is called *open* if for each $x \in O$, $\mu(x) \subset {}^*O$.

There are settings where a metric will not capture the notion we want; we need a topological space. We can talk about topological spaces using a

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 77–89.

base at each point in essentially the same way that we talk about metric spaces using balls.

1.1. Definition. *Fix a nonempty set X. A* **base at a point** $x \in X$ *is a nonempty collection $\mathcal{B}_x$ of nonempty subsets of X such that*

$$\forall U, V \in \mathcal{B}_x,\ \exists W \in \mathcal{B}_x \text{ such that } x \in W \subseteq U \cap V.$$

We will assume in what follows that X and a base $\mathcal{B}_x$ at each point $x \in X$ is given.

EXAMPLE: An example of a base not given by balls in a metric space is the base for pointwise convergence of real valued functions on $[0,1]$. Here, each point is actually a function f, and an element of the base specifies a finite number of points $r_1, \cdots, r_n$ in the interval $[0,1]$ and an $\varepsilon > 0$. A function g is in the basic set given by these parameters if for $1 \leq i \leq n$, $|g(r_i) - f(r_i)| < \varepsilon$. To see that the condition for a base is met, simply take two such sets for a given f, take the union of the two sets of points in $[0,1]$ and the smaller of the two ε's. This gives a basic set contained in the two initial ones.

1.2. Definition. *Given $x \in X$, the* **monad**

$$\operatorname{monad}(x) = \mu(x) := \bigcap_{U \in \mathcal{B}_x} {}^*U.$$

As with balls in a metric space, we will indicate that $y \in \mu(x)$ by writing $y \simeq x$. The **nearstandard points** *of ${}^*X$ are the points in the monad of some standard point.*

Note: Since any finite intersection of elements of $\mathcal{B}_x$ contains another element of $\mathcal{B}_x$, there is a $W \in {}^*\mathcal{B}_x$ with $W \subset \mu(x)$. In a metric space, one would take a ball of infinitesimal radius.

EXAMPLE: For pointwise convergence on $[0,1]$, the monad of a real valued function f would consist of all ${}^*\mathbb{R}$-valued functions g on ${}^*[0,1]$ such that at each standard x, $g(x) \simeq f(x)$.

1.3. Definition. *A set $O \subseteq X$ is called an* **open set** *if*

$$\forall x \in O,\ \exists U \in \mathcal{B}_x \ \text{ with } U \subseteq O,$$

or what is the same thing,

$$\forall x \in O,\ \mu(x) \subseteq {}^*O.$$

To see these are the same thing, we note first that for each $U \in \mathcal{B}_x$, $\mu(x) \subset {}^*U$. On the other hand, if $\mu(x) \subseteq {}^*O$, then $\exists W \in {}^*\mathcal{B}_x$ with $W \subseteq$

$\mu(x) \subseteq {}^*O$, and so "$\exists W \in \mathcal{B}_x$ with $W \subseteq O$" must also be true for the standard superstructure.

If for each $x \in X$ and each $U \in \mathcal{B}_x$, U is open, then we say that we are given a **base of open sets** at each $x \in X$ or an open base at each $x \in X$. In what follows, we will assume that our bases always consist of open sets.

Let $\mathcal{T}$ be the collection of all open sets in X. It is easy to see that $\mathcal{T}$ has the following properties:

i) The space X and the empty set $\emptyset$ are open.
ii) Finite intersections and *arbitrary* unions of open sets are open.

A set X and a collection of subsets $\mathcal{T}$ with these properties is called a **topological space**, and $\mathcal{T}$ is called a **topology** on X. Different bases, such as open balls or open rectangles in the plane, can give rise to the same topology just as they can give rise to the same monads of standard points. Given a topology, one can define a maximal base and the monad for the topology at each $x \in X$ by setting

$$\mathcal{B}_x = \{U \in \mathcal{T} : x \in U\}, \quad \mu(x) = \bigcap_{U \in \mathcal{T},\ x \in U} {}^*U.$$

Fix an open base $\mathcal{B}_x$ at each $x \in X$. Given $A \subseteq X$, and $x \in X$, we say that x is a **point of closure** of A if for each $U \in \mathcal{B}_x$, $U \cap A \neq \emptyset$, or what is the same thing, $\mu(x) \cap {}^*A \neq \emptyset$. We write $\bar{A}$ for the set of points of closure of A. Clearly, $A \subseteq \bar{A}$. A set A is called **closed** if $A = \bar{A}$. It is easy to establish the following result.

1.4. Proposition. *A set A is closed if and only if its complement $X \setminus A$ is open.*

1.5. Proposition. *The set X and the empty set $\emptyset$ are closed. Moreover, finite unions and arbitrary intersections of closed sets are closed.*

Proof. Use De Morgan's Law. □

1.6. Definition. *If $D \subseteq X$ and $\bar{D} = X$, then we say that D is a* **dense subset** *of X. If X contains a countable dense subset, we say that X is* **separable**.

1.7. Example. *The rational numbers are dense in $\mathbb{R}$.*

1.8. Definition. *The* **interior of a set** *E as the set of all x for which $\mu(x) \subset {}^*E$. We write E° for the interior.*

2. Continuous mappings

Now assume that $(X, \mathcal{S})$ and $(Y, \mathcal{T})$ are two topological spaces.

2.1. Definition. *A function f from X into Y is* **continuous** *at $x \in X$ if for each $V \in \mathcal{T}$ with $f(x) \in V$, there is a $U \in \mathcal{S}$ with $x \in U$ such that $f[U] \subseteq V$. Equivalently, we want*

$$^*f\,[\text{monad}_{\mathcal{S}}(x)] \subseteq \text{monad}_{\mathcal{T}}\,[f(x)]\,,$$

That is, $y \simeq x \Rightarrow {}^*f(y) \simeq f(x)$. We say that f is continuous or continuous on X if it is continuous at each $x \in X$.

2.2. Exercise. *Show these standard and nonstandard conditions are equivalent.*

2.3. Theorem. *A function f from X into Y is continuous on X if and only if for each open set O contained in Y, $f^{-1}[O]$ is open in X. A function f from $A \subseteq X$ into Y is continuous at each point of A iff $\forall V \in \mathcal{T}$, $\exists U \in \mathcal{S}$ such that $f^{-1}[V] \cap A = U \cap A$.*

Proof. EXERCISE.

3. Convergence

Assume we are working with a topological space X and an open base $\mathcal{B}_x$ at each $x \in X$. (For a metric space, the base at x is the set of open balls centered at x.)

3.1. Definition. *A* **sequence** x_n **converges** *to a point $x \in X$ if it is eventually in each $U \in \mathcal{B}_x$. That is, $\exists m \in \mathbb{N}$ such that for all $n \geq m$, $x_n \in U$.*

3.2. Exercise. *Show this condition is equivalent to $\forall H \in {}^*\mathbb{N}_\infty,\quad x_H \simeq x$.*

3.3. Definition. *A point x is a* **cluster point** *of a sequence x_n if the sequence x_n is frequently in every $U \in \mathcal{B}_x$. That is, $\forall m \in \mathbb{N}$, $\exists n \geq m$ with $x_n \in U$.*

3.4. Exercise. *Show this condition is equivalent to $\exists H \in {}^*\mathbb{N}_\infty$ with $x_H \simeq x$.*

A generalization of sequential convergence, used even in such simple settings as Riemann integration theory, replaces the natural numbers with a directed set. A **directed set** D is a set supplied with a transitive relation $\preceq$ such that for each a and b in D, $a \preceq a$, $b \preceq b$, and there is a $c \in D$ with $a \preceq c$ and $b \preceq c$. Partitions play the role of the directed set in Riemann integration theory. The generalization of a sequence is given by a **net**. This is a function x_a from a directed set into a given topological space X. A point x is a **limit of a net** x_a if for all $U \in \mathcal{B}_x$, there is a $c \in D$ such that for all $a \succeq c$, $x_a \in U$. That is, the net is eventually in U. A point x is a

cluster point of a net x_a if for each $U \in \mathcal{B}_x$ and each $c \in D$, there is an $a \succeq c$ with $x_a \in U$. That is, the net is frequently in U. The notion of a "subnet" is fairly complicated.

3.5. Exercise. *Give the nonstandard criterion for a net to converge and for a point to be a cluster point of a net.*

4. More on Topologies

One often has two topologies, $\mathcal{T}_1$ and $\mathcal{T}_2$ where every set open for the first is open for the second; i.e., $\mathcal{T}_1 \subseteq \mathcal{T}_2$. We say $\mathcal{T}_1$ is weaker or coarser than $\mathcal{T}_2$ and $\mathcal{T}_2$ is stronger or finer than $\mathcal{T}_1$. It is clear that the weaker the topology, i.e., the fewer open sets there are, the larger the monads will be.

The collection consisting of just X and the empty set is the weakest possible topology on X. The full power set of X is the strongest possible topology, and this contains any given collection of sets. It follows that any collection $\mathcal{C}$ of subsets of X generates a topology, namely, the intersection of all topologies containing the collection. This is the weakest topology containing $\mathcal{C}$.

We also have the notion of a **base for a topology** rather than a base at each point. A base $\mathcal{B}$ for a topology is a collection of open sets such that for each $x \in X$, $\{U \in \mathcal{B} : x \in U\}$ is a base at x. This base at x may not be the base you would choose. For example, the disks in the plane with rational centers and rational radii form a base, but some points are no longer centers of balls in this base. Given a base, it is easy to see that a set is open if and only if it is a union of sets from the base.

There are two topologies on the nonstandard extension $^*X$ of a topological space which are important in the literature. The first, called the Q-**topology** by Robinson, has a base consisting of the sets in $^*\mathcal{T}$. That is, each internal open set is open in this topology, but in general, there are external sets which are also open in this topology. The second important topology, called the S-**topology** by Robinson, has base $\mathcal{B} = \{^*U : U \in \mathcal{T}\}$. Notice that if x and y are in a monad, then every S-open set which contains x also contains y. That is, the S-topology is not "Hausdorff."

4.1. Definition. *A topological space is called a* **Hausdorff space** *if distinct points are contained in disjoint open sets, or what is the same thing, distinct standard points have disjoint monads. A topological space is said to satisfy the* **first axiom of countability** *if each point has a countable base* $\mathcal{B}_x$. *A topological space is said to satisfy the* **second axiom of countability** *if there is a countable base for the topology.*

Example: In a metric space, balls of radii $1/n$ form a base at each point. There may, however, be no countable base for the topology. In Euclidean

n-space, balls with rational centers and rational radii form a countable base for the topology.

For a Hausdorff space, we have the important standard part map st : $\text{ns}({}^*X) \to X$ defined as follows: For each $x \in X$ and each $y \in \mu(x)$, set $\text{st}(y) = {}^\circ y = x$. Even for a non-Hausdorff space, we can define the standard part of a set $B \subseteq {}^*X$, by setting

$$\text{st}(B) := \{x \in X : \mu(x) \cap B \neq \emptyset\}.$$

4.2. Theorem. **[Luxemburg]** *Assume that* $\text{card}(\mathcal{T}) < \kappa$ *and we are working with a κ-saturated enlargement. Then for each internal set $B \subseteq {}^*X$,* $\text{st}(B)$ *is closed.*

Proof. Assume that $z \in X$ is a point of closure of $\text{st}(B)$; we must show that $z \in \text{st}(B)$, i.e., $\mu(z) \cap B \neq \emptyset$. Given $U \in \mathcal{T}$ with $z \in U$, $\exists x \in U \cap \text{st}(B)$. Since $x \in \text{st}(B)$, $\exists y \in \mu(x) \cap B \subseteq {}^*U \cap B$. Since ${}^*U \cap B \neq \emptyset$ for each open U containing z, it follows from saturation that $\mu(z) \cap B \neq \emptyset$. □

Note that for the last proof, one could get away with just $\aleph_1$-saturation if one assumed the existence of a countable base at each point of X.

5. Compact Spaces

An open covering of a set $A \subseteq X$ is a collection of open sets in X such that each point of A is in at least one of the open sets. A subset $A \subseteq X$ is called a **compact set** if every open cover has a finite subcover. Of course, A may be all of X. By DeMorgan's law, X is compact iff every family of closed sets with the finite intersection property [this means that every *finite* subset has a nonempty intersection] itself has a nonempty intersection.

The nonstandard extension of a finite set $\{a_1, \cdots, a_n\}$ is just the equivalent finite set $\{{}^*a_1, \cdots, {}^*a_n\}$. Compactness generalizes finiteness. As Oswald has noted (private communication), Robinson's criterion for compactness makes this clear. One does not retain the points in a nonstandard extension, but one does not change them by much.

5.1. Theorem. **[Robinson]** *A set $A \subseteq X$ is compact iff for each $y \in {}^*A$, there is an $x \in A$ with $y \in \mu(x)$. In particular, X is compact if every point of ${}^*X$ is nearstandard.*

Proof. Assume A is compact but $\exists y \in {}^*A$ not in the monad of any $x \in A$. Then each $x \in A$ is contained in an open set O_x with $y \notin {}^*O_x$. The family $\{O_x : x \in A\}$ covers A and therefore has a finite subcover $\{O_1, \cdots, O_n\}$. Now since $A \subseteq \cup_{i=1}^n O_i$, ${}^*A \subseteq \cup_{i=1}^n {}^*O_i$. Since $y \notin {}^*O_i$, for $1 \leq i \leq n$, $y \notin {}^*A$. Contradiction.

Now assume that $\mathcal{U} = \{O_\alpha : \alpha \in \mathcal{A}\}$ is an collection of open sets no finite subset of which covers A. Let $\mathcal{B}$ be a hyperfinite collection in ${}^*\mathcal{U}$ with

${}^*O_\alpha \in \mathcal{B}$ for each $\alpha \in \mathcal{A}$. Then there is a $y \in {}^*A$ such that $y \notin U\ \forall U \in \mathcal{B}$. For each $x \in A$, there is an α with $x \in O_\alpha$. Since $y \notin {}^*O_\alpha$, $y \notin \mu(x)$. □

In a metric space, a set is **bounded** if it is contained in a ball $B(a, R)$ about some point of the space.

5.2. Theorem. **[Heine-Borel]** *A compact subset A of a metric space is closed and bounded. A closed and bounded subset of $\mathbb{R}^n$ is compact.*

Proof. EXERCISE.

Warning: It is not true in an infinite dimensional space that a closed and bounded set is compact. For example, the closed unit ball in l^2 is not compact.

A topological space $(X, \mathcal{T})$ is called **regular** if for each $x \in X$, the singleton set $\{x\}$ is closed, and for each closed set C with $x \notin C$, there are disjoint open sets U and V with $x \in U$ and $C \subseteq V$. That is, singleton sets are closed, and for each x and open U containing x there is an open set W with $x \in W \subseteq \overline{W} \subseteq U$.

5.3. Theorem. **[Luxemburg]** *Assume $(X, \mathcal{T})$ is regular. Also assume that* $\mathrm{card}(\mathcal{T}) < \kappa$, *and we are working in a κ-saturated enlargement. Then if B is an internal set of nearstandard points in ${}^*X$,* $\mathrm{st}(B)$ *is compact.*

Proof. EXERCISE.

5.4. Theorem. *If $K \subseteq X$ is compact and $f : K \to Y$ is continuous at each point of K, then the image $f[K]$ is compact in Y.*

Proof. EXERCISE.

5.5. Theorem. *A real-valued function continuous on a compact set K achieves a maximum and a minimum value at points of K.*

5.6. Theorem. *Let X be a topological space. If X is compact then every sequence in X has a cluster point. The converse holds if X satisfies the second axiom of countability, that is, if X has a countable base for the topology.*

Proof. EXERCISE

Let $(X, \mathcal{T})$ be a topological space and $\mathcal{B}$ a base for the topology. To show X is compact, it is enough to show that any covering by sets from $\mathcal{B}$ has a finite subcovering (exercise).

5.7. Theorem. *The S-topology makes ${}^*X$ compact.*

Proof. EXERCISE.

6. Product spaces

The topology of pointwise convergence on $[0,1]$ can be generalized as follows. Instead of $[0,1]$, we take an arbitrary index set $\mathcal{I}$. Instead of associating the real line with each $\alpha \in \mathcal{I}$, we let X_α be a topological space. Now the point set $\Pi_{\alpha\in\mathcal{I}}X_\alpha$ is the set of all functions f on $\mathcal{I}$ with $f(\alpha) \in X_\alpha$ for each $\alpha \in \mathcal{I}$. The monad of such an element f consists of all internal $g \in {}^*\Pi_{\alpha\in\mathcal{I}}X_\alpha$ with $g(\alpha) \simeq f(\alpha)$ for each standard $\alpha \in \mathcal{I}$. Such a g is a mapping on ${}^*\mathcal{I}$ with $g(\beta) \in X_\beta$ for each $\beta \in {}^*\mathcal{I}$, but the values of g at nonstandard indices are not relevant here. The space $\Pi_{\alpha\in\mathcal{I}}X_\alpha$ is called a **product space**, and the topology is called the **product topology**.

6.1. Theorem. *The product of Hausdorff spaces is Hausdorff.*

Proof. EXERCISE. □

6.2. Theorem. [Tychonoff] *The product of compact spaces is compact.*

Proof. If $X = \Pi_{\alpha\in\mathcal{I}}X_\alpha$ and $g \in {}^*X$, then for each standard $\alpha \in \mathcal{I}$, there is an $x_\alpha \in X_\alpha$ with $g(\alpha) \simeq x_\alpha$. (The x_α's are unique if the spaces X_α are Hausdorff.) The element $f \in X$ with $f(\alpha) = x_\alpha$ already given is in X and $g \in \mu(f)$. □

7. Restricted or relative topologies

Let X be a space with a topology, and let A be a subset of X. One can restrict the topology on X to A by calling a subset of A open if it is the intersection of A with an open subset of X. Thus, if $\mathcal{B}_x$ is a base of open sets for the original topology and $x \in A$, the collection $\{U \cap A : U \in \mathcal{B}_x\}$ is a base at x for the restricted topology. If X is a metric space, the restriction of the metric to pairs of points in A gives the appropriate metric on A. In all of this, one just ignores points outside of A. One does have to recall that an open subset of A in the restricted topology, for example A itself, may not be open in all of X. One also speaks of the **relative topology** on A.

7.1. Exercise. *Describe the relatively open subsets of A in terms of monads for the topology on X.*

8. Uniform continuity on metric spaces

8.1. Definition. *A map f from a set A contained in a metric space (X,d) into a metric space (Y,ρ) is a* **uniformly continuous function** *on A iff $\forall\varepsilon > 0$, $\exists\delta > 0$ such that if x and y are in A and $d(x,y) < \delta$, then $\rho(f(x), f(y)) < \varepsilon$.*

8.2. Theorem. *A map f from a set A contained in a metric space (X,d) into a metric space (Y,ρ) is uniformly continuous on A iff $\forall x,\ y \in {}^*A$, with $x \simeq y$, ${}^*f(x) \simeq {}^*f(y)$.*

Proof. Assume $x \simeq y \Rightarrow {}^*f(x) \simeq^* f(y)$. Pick $\varepsilon > 0$ in $\mathbb{R}$. then the sentence

$$(\exists \delta \in \mathbb{R}^+)(\forall x, y \in A)[d(x,y) < \delta \Rightarrow \rho(f(x), f(y)) < \varepsilon]$$

holds for the extension and therefore for the original structure. The converse is similar to the proof for $\mathbb{R}$. □

8.3. Theorem. *A continuous function on a compact set A is uniformly continuous.*

Proof. EXERCISE.

9. Nonstandard Hulls

If one takes the finite nonstandard rational numbers modulo the infinitesimal ones, i.e., $[{}^*\mathbb{Q} \cap \mathrm{Fin}({}^*\mathbb{R})]/[{}^*\mathbb{Q} \cap \mu(0)]$, one obtains $\mathbb{R}$.

Let (X,ρ) be a metric space and take an $\aleph_1$-saturated enlargement. The set $\mathrm{Fin}({}^*X)$ of finite elements of ${}^*X$ are the elements y for which ${}^*\rho(y,x)$ is finite for some $x \in X$. The relation $\simeq$ is an equivalence relation on $\mathrm{Fin}({}^*X)$. Let $\hat{X}$ denote the set of equivalence classes $\{\mu(x) : x \in \mathrm{Fin}({}^*X)\}$. Then $\hat{X}$ is a metric space with respect to $\hat{\rho}$, where $\hat{\rho}(\mu(x), \mu(y)) = \mathrm{st}({}^*\rho(x,y))$. It is easy to see that $\hat{\rho}$ is well defined and a metric. For example, $\hat{\rho}(\mu(x), \mu(y)) = 0$ iff $x \simeq y$ iff $\mu(x) = \mu(y)$.

9.1. Definition. *The space $(\hat{X}, \hat{\rho})$ is called the* **nonstandard hull** *of (X,ρ).*

9.2. Theorem. *If follows from $\aleph_1$-saturation that $(\hat{X}, \hat{\rho})$ is complete.*

Proof. Let $\langle \mu(a_i), i \in \mathbb{N}\rangle$ be a Cauchy sequence in $\hat{X}$, and pick a sequence $\langle a_i : i \in \mathbb{N}\rangle$ of representatives. For each $k \in \mathbb{N}$, $\exists n(k)$ so that if $i \geq n(k)$ and $j \geq n(k)$, then $\hat{\rho}(\mu(a_i), \mu(a_j)) < 1/k$, whence ${}^*\rho(a_i, a_j) < 1/k$. We may assume that $n(k+1) > n(k)\,\forall k$. Use $\aleph_1$-saturation to extend to an internal sequence $\langle a_i : i \in {}^*\mathbb{N}\rangle$. It follows from the Spillover Principle that for each $k \in \mathbb{N}$, we may pick an unlimited integer $\eta(k)$ such that for $n(k) \leq i \leq \eta(k)$ and $n(k) \leq j \leq \eta(k)$, ${}^*\rho(a_i, a_j) < 1/k$. We may assume the $\eta(k)$ are decreasing. Extend the sequence $\{\eta(k) : k \in \mathbb{N}\}$ to ${}^*\mathbb{N}$. Again, by Spillover, there is an unlimited integer γ with $n(k) \leq \gamma \leq \eta(k)$ for all $k \in \mathbb{N}$. It follows that ${}^*\rho(a_i, a_\gamma) < 1/k$ for every $k \in \mathbb{N}$, whence $\mu(a_i) \to \mu(a_\gamma)$. □

The mapping $x \to \mu({}^*x)$ is clearly an isometry (a distance preserving map which must, therefore, be one-to-one) of (X,ρ) *into* $(\hat{X}, \hat{\rho})$. If we start with a Banach space $(X, \|\cdot\|)$, the resulting standard Banach space $\hat{X}$, built from $^*X$ is called the nonstandard hull of X. Here, $\|\cdot\|^\wedge$ is defined by setting $\|\mu(x)\|^\wedge = \mathrm{st}(\|^*x\|)$. This construction is equivalent to the "Banach Space Ultrapower" construction, but one has the internal structure to help in the development.

In general, there are new elements in $\hat{X}$. For example, if $X = \ell^2$, then the function on $^*\mathbb{N}$ which is 1 at some fixed $\eta \in {}^*\mathbb{N}_\infty$ and 0 elsewhere has norm 1 and is not in the monad of any standard sequence. If $X = \ell^1$, then the function which is $1/\eta$ on $\{n \in {}^*\mathbb{N} : 1 \le n \le \eta\}$ and 0 elsewhere has norm 1 and is not in the monad of any standard sequence.

10. Compactifications

Our treatment of general compactifications will follow the ideas of Salbany and Todorov [8], though related ideas have been in the nonstandard analysis literature since the initial work of Robinson [7]. Given a topological space $(X, \mathcal{T})$, we would like to take the nonstandard extension $^*X$ with the S-topology, and form equivalence classes compatible with the S-topology. The space of equivalence classes Y with open sets formed from the S-open sets is then compact. We hope to do this so that a homeomorphic image of the space X is recaptured as a dense subset. Here is one way to do this.

Let $(X, \mathcal{T})$ be a Hausdorff space, and let Q be a family of bounded, continuous, real-valued functions on X such that for each closed set $A \subset X$ and each $x \notin A$, there is an $f \in Q$ with $f(x) \notin \overline{f[A]}$. (A space which admits such a family is called **completely regular**.) One can imbed X as a dense subset of a compact Hausdorff space $(\bar{X}, \bar{\mathcal{T}})$ such that each $f \in Q$ has a continuous extension $\bar{f}$ to $\bar{X}$ and for $x \neq y$ on the boundary $\bar{X} \setminus X$, $\bar{f}(x) \neq \bar{f}(y)$.

EXAMPLE. let $X = (0, 1]$, and let Q consist of all functions with compact support on Q together with the function $\sin(\pi/x)$. The whole line on the y-axis from -1 to 1 is adjoined as a boundary.

To form what is called the **Q-compactification**, we call any two points x and y in $^*X$ equivalent if $^*f(x) \simeq {}^*f(y)$ for all $f \in Q$. The space of equivalence classes is the compactification $\bar{X}$. The extension of any $f \in Q$ is given by $\bar{f}([x]) = \mathrm{st}(^*f(x))$; this is well defined. A neighborhood base for an $[x]$ in $\bar{X}$ is formed by sets of the form $\{[y] : |\bar{f}_i([y]) - \bar{f}_i([x])| < \varepsilon,$ $1 \le i \le n\}$; that is, it is the weakest topology making all of the extensions $\bar{f}$ continuous.

If one starts with Q being the set of all bounded, continuous real-valued functions on a completely regular Hausdorff space X, one gets the **Stone-Cech compactification** in this way. The Stone-Cech compactification $\beta\mathbb{N}$ of the natural numbers $\mathbb{N}$ is formed from the space Q consisting of bounded sequences.

11. More Exercises

EXERCISE: Let p_n be a sequence of polynomials and x_n a sequence of variables so that $\forall n$, p_n is a function of $x_1, \cdots, x_n$. Let I_n be a sequence of closed and bounded intervals in $\mathbb{R}$. Assume $\forall n \in \mathbb{N}$, $\exists a_i^n \in I_i$ for $1 \leq i \leq n$ such that $\forall i \leq n$, $p_i(a_1^n, \cdots, a_i^n) = 0$. Show that there are values $a_i \in I_i$ $\forall i \in \mathbb{N}$ such that $\forall n \in \mathbb{N}$, $p_n(a_1, \cdots, a_n) = 0$.

Solution: Going to the enlargement, we may pick $\eta \in {}^*\mathbb{N}_\infty$ and a sequence $a_i^\eta \in {}^*I_i$ for $1 \leq i \leq \eta$ so that for all $n \in \mathbb{N}$, ${}^*p_n(a_1^\eta, \cdots, a_n^\eta) = 0$. Since I_i is compact for $i \in \mathbb{N}$, we may choose $a_i = \mathrm{st}(a_i^\eta)$ $\forall i \in \mathbb{N}$. Now by continuity, $\forall n \in \mathbb{N}$, $p_n(a_1, \cdots, a_n) = \mathrm{st}({}^*p_n(a_1^\eta, \cdots, a_n^\eta)) = 0$.

EXERCISE given above on the standard part of nearstandard set: Let $(X, \mathcal{T})$ be a regular Hausdorff space in a κ-saturated enlargement with $\kappa > \mathrm{Card}(\mathcal{T})$. Assume A is an internal set of near-standard points. Show $E = \mathrm{st}(A)$ is compact.

Solution: Fix $y \in {}^*E$. If U is a standard open set with $y \in {}^*U$, then $U \cap E \neq \emptyset$, because if $U \cap E = \emptyset$, then ${}^*U \cap {}^*E = \emptyset$. Given $x \in E \cap U$, by definition of E, $\exists a \in A$ with $a \in \mu(x) \subset {}^*U$. Thus, for each standard open U, if $y \in {}^*U$, $\exists a \in A \cap {}^*U$. By saturation, $\exists a_0 \in A$ with $a_0 \in A \cap {}^*U$ for each standard open U with $y \in {}^*U$. Let $x = \mathrm{st}(a_0)$. We must show that $y \in \mu(x)$, whence $\mathrm{st}(y) = x \in E$. If $y \notin \mu(x)$, then there is an open set V with $x \in V$ and $y \notin {}^*V$; by regularity there is an open set U with $x \in U \subseteq \overline{U} \subseteq V$. It follows that $x \in U$ and $y \in X \setminus {}^*\overline{U}$, whence, $a_0 \in {}^*U$ and $a_0 \in X \setminus {}^*\overline{U}$. This is a contradiction.

EXAMPLE: The **mushroom space** is an example of a nonregular space where things go wrong. Here, X is the unit square $\{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$. The topology is generated by Euclidean neighborhoods except for points $(x, 0)$. Here a typical neighborhood consists of $(x, 0)$ together with the set $\{(\zeta, y) : (x - \zeta)^2 + y^2 < r^2\}$ for some $r > 0$. The restriction of this topology to the set $L = \{(x, 0) : 0 \leq x \leq 1\}$ is the discrete topology. The point $(1/2, 0)$ can not be separated from its closed complement $\{(x, 0) : x \neq 1/2, 0 \leq x \leq 1\}$ in L. The set $A = \{(\zeta, \varepsilon) : \zeta \in {}^*[0, 1]\}$, where ε is a positive infinitesimal, is internal and nearstandard, but its standard part L is not compact.

EXERCISE. Here is a construction of an ultrapower formed from a standard superstructure so that the ultrapower is an **enlargement**. That is, given any standard set A, there is a hyperfinite set $F \subseteq {}^*A$ such that for each $a \in A$, ${}^*a \in F$.

Let J be the set of all nonempty finite subsets of the standard superstructure. For each $a \in J$, set $J_a = \{b \in J : a \subseteq b\}$. Let $\mathcal{F} = \{A \subseteq J : \exists J_a \subseteq A\}$.

(1) Show $\mathcal{F}$ is a free filter on J.

(2) Show that for any ultrafilter $\mathcal{V}$ on J containing $\mathcal{F}$, the corresponding ultrapower constructed from the standard superstructure is an enlargement.

EXERCISE. The setting of this problem is a d-dimensional normed vector space $\mathbb{R}^d$ supplied with a norm which is not necessarily the Euclidean norm. Let $B(\mathbf{0}, 2)$ denote the closed ball of radius 2 about the origin $\mathbf{0}$ in $\mathbb{R}^d$. Let $K(s)$ denote the number of points one can "pack" into $B(\mathbf{0}, 2)$ when one point is at $\mathbf{0}$ and the distance between pairs of distinct points is at least s. It is known, for example, that for the Euclidean norm in the plane $\mathbb{R}^2$, $K(1) = 19$. Prove or disprove the following conjecture: For any $\delta > 0$ in $\mathbb{R}$, $K(1 - \delta) \geq K(1) + 1$.

EXERCISE (From [4]). The setting of this problem is a metric space (X, ρ); we write $B(c, r)$ to denote a closed metric ball $\{x \in X : \rho(c, x) \leq r\}$. We will call the same ball but with a strict inequality the inside of $B(c, r)$. We write $\overline{S}$ to denote the closure of a set S. Let A be an arbitrary subset of X. Fix $R > 0$. Assume that at each point $a \in A$ there is centered a closed ball $B(a, r(a))$ with positive radius $r \leq R$. Also assume that all closed balls in X of radius less than or equal to R are compact. With each point $p \in \overline{A}$ we associate the set

$$L(p) = \{t > 0 : \forall \varepsilon > 0, \exists a \in A \quad \text{with} \quad \rho(p, a) < \varepsilon \quad \text{and} \quad |t - r(a)| < \varepsilon\}.$$

Show that there exists a collection of closed balls $\mathcal{B}$ with

$$\mathcal{B} \subseteq \{B(p, t) : p \in \overline{A} \text{ and } \quad t \in L(p)\}$$

such that no center of any ball in $\mathcal{B}$ is in the inside of any other ball in $\mathcal{B}$, and each $a \in A$ is in a ball $B(p, t) \in \mathcal{B}$ with $t \geq r(a)$. Note that if the radius function r is continuous on A and $\lim_{a \to p} r(a) = 0$ at every point $p \in \overline{A} \setminus A$, then $\mathcal{B} \subseteq \{B(a, r(a)) : a \in A\}$.

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., and Lindstrøm, T., (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Hurd, A. and Loeb, P. A., (1985) *An Introduction to Nonstandard Real Analysis.* Academic Press, New York.
3. Lindstrøm, T., (1988) An invitation to nonstandard analysis. In Cutland (1988), pp. 1–105.
4. Loeb, P. A., (1993) An optimization of the Besicovitch covering, *Proc. Amer. Math. Soc.*, **118**, 715–716.
5. Luxemburg, W.A.J.,(1969) A general theory of monads, in *Applications of Model Theory to Algebra, Analysis, and Probability* (ed. W.A.J. Luxemburg). Holt, Rinehart and Winston, New York, pp. 18–86.
6. Robinson, A., (1966) *Nonstandard Analysis.* North-Holland, Amsterdam. (Second, revised edition, 1974).
7. Robinson, A., (1969) Compactification of groups and rings and nonstandard analysis, *Jour. of Symbolic Logic* **34**, pp. 576–588.

8. Salbany, S. and Todorov, T. D., Nonstandard and Standard Compactifications, Preprint.
9. Stroyan, K. and Luxemburg, W.A.J., (1976) *Introduction to the Theory of Infinitesimals.* Academic Press, New York.

LOEB MEASURE AND PROBABILITY

DAVID A. ROSS
Department of Mathematics
University of Hawaii
Honolulu, HI 96822
USA
email: ross@kahuna.math.hawaii.edu

1. Introduction

In these notes I give an introduction to nonstandard measure theory and probability theory. As with any short introduction, the number of topics that can be covered is a tiny subset of all that one would really like to discuss; in particular, I don't always present theorems in their strongest form. However, this development should be adequate for all but an infinitesimal number of applications.

On the other hand, I *have* made some effort not to give precisely the same introduction that appears in all the other (very good) surveys of the area. For example, I include some results (e.g., Theorem 5.2) which are much stronger than those which are usually given, and give some applications (Haar measure, Skorokhod's Theorem) which haven't appeared elsewhere, at least not in this form.

Despite the title of these notes, I say very little about probability theory. It shows up here only in the form of a couple of examples, the very brief discussion in Section 7, my tendency to assume that all measures have total mass one, and a corresponding use of the locution 'almost surely' instead of the more general 'almost everywhere' (see Section 1.3). Fortunately, there are two other lecturers at this Institute lecturing on nonstandard probability, so the reader should come through the course well-versed in the subject.

I will focus on results which appeal to a measure construction first formulated by Peter Loeb [21] over twenty years ago. While this construction can no longer be called 'recent', it will nevertheless be unfamiliar to anyone whose knowledge of nonstandard analysis comes from the books by Robinson [26], Stroyan and Luxemburg [33], Davis [8], or (more recently)

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 91–120.

Nelson [24]. Over these last two decades the Loeb measure construction has been the source of virtually all of the interesting applications of nonstandard analysis to measure theory, and results in the area which don't use it seem quaint by comparison.

1.1. PREHISTORY

Despite my remark at the end of the last section, there *has* been *some* interesting nonstandard measure theory done without Loeb measures.

Nonstandard measures first appeared in Robinson [26], where some standard concepts were given slightly simpler nonstandard formulations; see, for example, Exercise 1.

Some of the early nonstandard results in measure theory are certainly worth a look. I recommend two here. In [5], Bernstein and Wattenberg give an extremely simple proof of a theorem of Banach, which states that Lebesgue measure can be extended to a finitely-additive motion-invariant measure defined for *all* subsets of $\mathbb{R}$ (a highly nontrivial result which, for example, fails for $\mathbb{R}^n$, $n > 2$).

Henson uses a similar argument in [9] to prove a result (from harmonic analysis) about group amenability, that a property called Folner's condition on a group G of motions of a set S suffices for the existence of a (finitely-additive) invariant measure on S.

The reader should try Exercises 3 and 2 for the flavor of these results.

1.2. LOEB MEASURES

Peter Loeb introduced his measure construction in the paper [21]. There is now a large number of very good introductions to the theory, including [13],[6],[7], [18],[20], [1], [34],[16].

Some of the ideas behind this construction were 'in the air' before Loeb's papers. In one intriguing example, in Rao [25], introduced the notion of a *pure* measure. The proof (below) that the Loeb measure is countably additive is almost exactly the proof that pure measures are countably additive. Without any further explanation, Rao remarks that his definition is "suggested by certain considerations in Nonstandard Analysis" (see section 4 of [25]).

Since Loeb's papers, the construction has been used to obtain results in many areas. Besides probability theory (of which you'll see plenty later, I promise), Loeb measures have appeared in Control Theory, Mathematical Economics, Mathematical Physics, Ergodic Theory, Harmonic Analysis, Statistics, and probably several other fields which I cannot recall right now.

Most recently, many people in the field (especially Loeb himself; see, for example, [13]) have tended towards an operator-theoretic development of measure theory, obtaining the Loeb measure as a special case of a more general linear operator on the lattice of measurable functions. Ultimately this approach is probably better than the more traditional one, in the sense that it produces more general results with somewhat less effort. However, it is less intuitive to most mathematicians, and the generalization is largely unimportant for the applications to probability theory, so I will stick with the more traditional approach here.

Another fairly recent development is the adaptation of the Loeb construction to the creation of *capacities*. Capacity is a more general notion of content than measure; standard capacities constructed from nonstandard ones have been extraordinarily useful since their introduction. For example, one is essential in the proof of Theorem 5.2 below; a similar capacity was the key to a difficult problem in [11]. In his Ph.D. thesis (and several subsequent papers), Boško Živaljević has used capacities to prove an abundance of results in nonstandard descriptive set theory.

1.3. ASSUMPTIONS

I will make a large number of assumptions in these notes, many of which are unnecessary but all of which make the exposition more pleasant:

1. All topological spaces are Hausdorff. (This assumption ensures that st exists; it can often be eliminated by replacing arguments using the standard part *function* by arguments using the standard part *relation.*)
2. All measures are probability measures (i.e., total measure one) unless otherwise specified. When measures are allowed to be infinite, many aspects of the theory become substantially messier. (Of course, everything generalizes with no problem to arbitrary *finite* measure spaces, and often with little problem to σ-finite spaces.) A good reference on infinite Loeb measures is [34].
3. The nonstandard model is 'as saturated as it needs to be'. The reader will by now have seen several examples of saturation arguments (and perhaps related arguments involving 'overspill' or 'underspill'), though probably most have just used ω_1-saturation. Here it will be convenient to assume that the model is very saturated indeed, at least as saturated as every standard entity we are likely to meet. I will then often prove things "by saturation", by which I really mean by κ-saturation, where the model is at least κ-saturated.
4. Sometimes I will leave stars off objects which should really have them. For example, sometimes I'll write $\int FdP$ for ${}^*\!\int FdP$ (and sometimes I'll write both in the same equation!). The actual meaning should always be clear from the context.

I will also assume that the reader has graduate-level training in (standard) measure theory, say enough to know what I mean by 'Lebesgue's Dominated Convergence Theorem'. Very occasionally I will assume some exposure to probability theory, at a very low level. Other lecturers will likely require much more probability, though hopefully we are all amenable to filling in any mismatches in backgrounds.

1.4. REMARKS ON EXERCISES

These notes are followed by a rather large number of exercises, which vary widely in difficulty. I don't expect the reader to work *all* the exercise during the course of my lectures. However, he or she should work as thick a cross-section as is possible, especially since many of the proofs of important theorems have been moved to the exercises. The problems are listed roughly in the same order as the material to which they are relevant, and enought of them are referenced in the text proper that the reader should be able to coordinate the reading with problem solving.

2. Finite Loeb Measure

In this section I describe the construction of standard, finite measure spaces from nonstandard ones. As mentioned above, I will for simplicity always work with probability (i.e., total measure one) spaces, unless otherwise indicated.

2.1. CONSTRUCTION

Start with an *internal, finitely-additive probability space*, which is to say a triple $(\Omega, \mathcal{A}, P)$ such that

i. Ω is an internal set;
ii. $\mathcal{A}$ is an internal subalgebra of $\mathcal{P}(\Omega)$; and
iii. $P : \mathcal{A} \longrightarrow {}^*R$ is an internal function such that
   a) $P(\emptyset) = 0$;
   b) $P(\Omega) = 1$; and
   c) P is *finitely − additive, that is, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ whenever $A, B \in \mathcal{A}$

Note that we needn't worry in (ii) above whether $\mathcal{A}$ is an algebra or an *algebra, since for internal sets the notions algebra and *algebra coincide. (Verify this!) However, there are two other aspects of this definition which are well worth worrying about. First, P is not *externally* a finitely-additive measure, since it takes values in ${}^*\mathbb{R}$ instead of $\mathbb{R}$. It *is* the case, however, that $(\Omega, \mathcal{A}, {}^\circ\!P)$ is a genuine, external finitely-additive probability space; see Exercise 4

Second, $\mathcal{A}$ is in general not (externally) a σ − algebra, even if $\mathcal{A}$ is a ${}^*\sigma$-algebra (see Exercise 5). It follows that $(\Omega, \mathcal{A}, {}^\circ\!P)$ will only be a genuine probability space in the most trivial cases.

The central theorem in modern nonstandard measure theory is the following:

Theorem 2.1 *Let $(\Omega, \mathcal{A}, P)$ be an internal finitely-additive probability space; then there is a standard (σ-additive) probability space $(\Omega, \mathcal{A}_L, P_L)$ such that:*

1. *$\mathcal{A}_L$ is a σ − algebra with $\mathcal{A} \subseteq \mathcal{A}_L \subseteq \mathcal{P}(\Omega)$*
2. *$P_L = {}^\circ\!P$ on $\mathcal{A}$.*
3. *For every $A \in \mathcal{A}_L$ and standard $\epsilon > 0$ there are $A_i, A_o \in \mathcal{A}$ such that $A_i \subseteq A \subseteq A_o$ and $P(A_o \setminus A_i) < \epsilon$.*
4. *For every $A \in \mathcal{A}_L$ there is a $B \in \mathcal{A}$ such that $P_L(A \triangle B) = 0$.*

The space $(\Omega, \mathcal{A}_L, P_L)$ is called a *Loeb (probability) space.* This is such an important result that I'll give two different proofs.

PROOF NUMBER ONE. As mentioned above (and in Exercise 4), $(\Omega, \mathcal{A}, {}^\circ\!P)$ is a standard finitely-additive probability space. Let $A_0 \supseteq A_1 \supseteq A_2 \cdots \supseteq A_n \supseteq \cdots$ be a countable nonincreasing chain of elements of $\mathcal{A}$, and suppose $\cap_{n=1}^{\infty} A_n = \emptyset$. By ω_1-saturation, $A_N = \emptyset$ for some $N \in \mathbb{N}$. It follows that $\lim_{n \to 0} {}^\circ\!P(A_n) = {}^\circ\!P(A_N) = 0$. Conclusions (1)–(3) are now immediate consequences of the Caratheodory Extension Theorem.

Conclusion (4) now follows by a saturation argument: fix $A \in \mathcal{A}_L$, and observe by (3) that for each $n \in \mathbb{N}$ there are $A_i^n, A_o^n \in \mathcal{A}$ such that $A_i^n \subseteq A \subseteq A_o^n$ and $P(A_o^n \setminus A_i^n) < \frac{1}{n+1}$. Without loss of generality, the sequence $\{A_i^n\}_{n \in \mathbb{N}}$ (respectively, $\{A_o^n\}_{n \in \mathbb{N}}$) is increasing (respectively, decreasing) in n. By ω_1-saturation, there is a $B \in \mathcal{A}$ such that $A_i^n \subseteq B \subseteq A_o^n$ for all $n \in \mathbb{N}$. (See Exercise 6.) For each $n \in \mathbb{N}$, $P_L(A \triangle B) \leq P_L(A_o^n \setminus A_i^n) + P_L(A_o^n \setminus A_i^n) \approx P(A_o^n \setminus A_i^n) + P(A_o^n \setminus A_i^n) \leq \frac{2}{n+1}$; it follows that $P_L(A \triangle B) = 0$. ⊣

The next proof uses property (3) to *define* both $\mathcal{A}$ and P_L simultaneously, and requires no standard theory. Most of the details are left as an exercise, which the reader is strongly encouraged to work.

PROOF NUMBER TWO (SKETCH). Put $\mathcal{A}_L = \{A \subseteq \Omega \mid \forall \epsilon > 0 \exists A_i, A_o \in \mathcal{A}$ such that $A_i \subseteq A \subseteq A_o$ and $P(A_o \setminus A_i) < \epsilon\}$. Clearly $\mathcal{A} \subseteq \mathcal{A}_L$, and $A \in \mathcal{A}_L \iff \overline{P}(A) = \underline{P}(A)$ (where $\overline{P}(A) = \inf\{{}^\circ\!P(A_o) \mid A \subseteq A_o,\ A_o \in \mathcal{A}\}$ and $\underline{P}(A) = \sup\{{}^\circ\!P(A_i) \mid A_i \subseteq A,\ A_i \in \mathcal{A}\}$). Put $P_L = \overline{P}$ on $\mathcal{A}_L$. The triple $(\Omega, \mathcal{A}_L, P_L)$ trivially satisfies conclusions (2) and (3), and (4) follows from (3) using the argument of the first proof. It remains to verify that:

a) $A \in \mathcal{A}_L \Rightarrow A^c \in \mathcal{A}_L$ and $P_L(A^c) = 1 - P_L(A)$;
b) $A, B \in \mathcal{A}_L \Rightarrow A \cup B,\ A \cap B \in \mathcal{A}_L$ and$P_L(A \cup B) = P_L(A) + P_L(B) - P_L(A \cap B)$; and
c) $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_L \Rightarrow \cup_{n \in \mathbb{N}} A_n \in \mathcal{A}_L$ and $P_L(\cup_{n \in \mathbb{N}} A_n) = \lim_{N \to \infty} P_L(\cup_{n \leq N} A_n)$.

(a) and (b) are left for the reader (Exercise 7). For (c), put $A_\infty = \cup_{n \in \mathbb{N}} A_n$, and fix an arbitrary standard $\epsilon > 0$. Let $r = \underline{P}(A_\infty) + \epsilon$. For $n \in \mathbb{N}$ find $A_i^n, A_o^n \in \mathcal{A}$ such that $A_i^n \subseteq A \subseteq A_o^n$ and $P(A_o^n \setminus A_i^n) < \frac{\epsilon}{2^{n+1}}$. Consider the following countable list of conditions on (new sets) B and C:

i) $B \in \mathcal{A}, \quad P(B) < r$
ii) $C \in \mathcal{A}, \quad P(C) < \epsilon$
iii) $A_i^n \subseteq B, \qquad n = 0, 1, 2, \cdots$
iv) $A_o^n \setminus A_i^n \subseteq C, \qquad n = 0, 1, 2, \cdots$

Suppose in fact that we have sets B, C satisfying these conditions; since for $n \in \mathbb{N}$ $A_n \subseteq A_o^n \subseteq B \cup C$, it follows that $\overline{P}(A_\infty) < P(B \cup C) + \epsilon \leq P(B) + P(C) + \epsilon < r + 3\epsilon = \underline{P}(A_\infty) + 4\epsilon$. Since ϵ is arbitrary, $\overline{P}(A_\infty) \leq \underline{P}(A_\infty)$, which proves (c).

To find internal B, C satisfying all these conditions, it suffices by ω_1-saturation to show that any finite subset of these conditions is satisfiable. So, let $n_1, n_2, \cdots, n_k \in \mathbb{N}$, put $B = A_i^{n_1} \cup A_i^{n_2} \cup \cdots \cup A_i^{n_k}$, and put $C = (A_o^{n_1} \setminus A_i^{n_1}) \cup (A_o^{n_2} \setminus A_i^{n_2}) \cup \cdots \cup (A_o^{n_k} \setminus A_i^{n_k})$; it remains to show that (i) and (ii) hold for this B and C. B and C are both in $\mathcal{A}$since $\mathcal{A}$is an algebra. Since $B \subseteq A_\infty$, ${}^\circ P(B) \leq \underline{P}(A_\infty)$, so $P(B) \leq r$. By internal finite-additivity of P on $\mathcal{A}$, $P(C) \leq \sum_{j=1}^k A_o^{n_j} \setminus A_i^{n_j} \leq \sum_{j=1}^k \frac{\epsilon}{2^{n_j+1}} < \epsilon$. This completes the proof. $\dashv$

The astute reader will have noticed that while the condition $\underline{P}(A) = \overline{P}(A)\ \forall A \in \mathcal{A}$ (in the notation of the second proof) guarantees that P_L is unique on $\mathcal{A}_L$, it doesn't guarantee uniqueness of $\mathcal{A}_L$ itself. As it happens, both the above proofs produce *complete* measures $(\Omega, \mathcal{A}_L, P_L)$, in fact $\mathcal{A}_L$ is the P_L-completion of Borel($\mathcal{A}$) = the smallest $\sigma -$ algebra containing $\mathcal{A}$. In these lectures I will always assume that this is true of $\mathcal{A}_L$. The reader is warned, however, that many writers assume that $\mathcal{A}_L$ is in fact just Borel($\mathcal{A}$), then write something like $\overline{\mathcal{A}_L}$ or $\overline{\mathrm{L}(\mathcal{A})}$ for the completion.

2.2. REMARKS

There are several ways in which the hypotheses of Theorem 2.1 can be weakened without too much damage to the conclusions; for example:

1. P can be *any* internal set function – not necessarily an internal finitely-additive measure – as long as $°P$ is a finitely-additive probability on $(\Omega, \mathcal{A})$.
2. If the algebra $\mathcal{A}$ is not internal (but still comprised of internal sets), or if P is not internal (but $°P$ is still an externally finitely additive probability on $\mathcal{A}$) then at least conclusions (1)–(3) still hold.

2.3. EXAMPLES

Example 2.1 *Let $(X, \mathcal{B}, \mu)$ be any (standard) probability measure; then $({}^*X\ , {}^*\mathcal{B}_L\ , {}^*\mu_L)$ is sometimes called the* nonstandard hull *of $(X, \mathcal{B}, \mu)$.*

Example 2.2 *Let Ω be hyperfinite, $\mathcal{A}$ be any internal subalgebra of $\mathcal{P}(\Omega)$, and let $p : \Omega \longrightarrow {}^*[0,1]$ be any internal function satisfying $\sum\limits_{\omega\in\Omega} p(\omega) \approx 1$. Define $P : \mathcal{A} \longrightarrow {}^*[0,1]$ by $P(A) = \sum\limits_{\omega\in A} p(\omega)$. Then the probability space $(\Omega, \mathcal{A}_L\ , P_L)$ is in this case called a* hyperfinite Loeb probability space. *When $\mathcal{A} = \mathcal{P}(\Omega)$, and $p(\omega) = \frac{1}{\|\Omega\|}$ for $\omega \in \Omega$ (where $\|\Omega\|$ is the internal cardinality of Ω), then P_L is called* uniform *hyperfinite Loeb probability space. If when defining a hyperfinite Loeb space the specification of $\mathcal{A}$ is omitted, it will be assumed that $\mathcal{A} = \mathcal{P}(\Omega)$. Similarly, if specification of p is omitted, then the space is assumed uniform.*

Example 2.3 *Let $n \in \mathbb{N}$, $H \in {}^*\mathbb{N}$, $\Gamma = \{1, 2, \ldots, H\}$, and let $\rho : \Gamma \rightarrow \{0,1\}$ be internal. (The reader might want to think of Γ as a set of red and blue balls in an urn, and of ρ as an indicator of which of the balls are red.) Put $\Omega = \Gamma^n$, and let P_L be the uniform Loeb measure on Ω. For $\omega = (\gamma_1, \ldots, \gamma_n) \in \Omega$, put $x(\omega) = \sum\limits_{i=1}^{n} \rho(\gamma_i)$.*

I leave it to the reader (Exercise 8) to confirm that x is $\mathbb{N}$-valued, measurable, and has a Binomial(n,p) distribution, where $p = {}^\circ\!\left(\frac{\sum_{\gamma\in\Gamma}\rho(\gamma)}{H}\right)$. (Of course, when H is standard then this is an entirely standard example. For more interesting variants, see Exercise 9 and Exercise 10.)

Example 2.4 *Let $H \in {}^*\mathbb{N}\backslash\mathbb{N}$, $\Delta t = \frac{1}{H}$, and let $\Omega = \{0, \Delta t, 2\Delta t, \ldots, (H-1)\Delta t\}$. The reader is invited to show that the standard part map* st $: \Omega \longrightarrow {}^\circ[0,1]$ *is a measure-preserving function from the uniform Loeb measure on Ω to Lebesgue measure on $[0,1]$. (Exercise 11).*

2.4. NONMEASURABLE SETS

In this section – which the reader is invited to skip – I'll contruct a couple of interesting nonmeasurable sets.

First, I'll use Lebesgue measure λ on ${}^*[0,1]$ to construct a Loeb-nonmeasurable set. Let $(\Omega, \mathcal{A}_L, P_L)$ be the Loeb space from Example 2.4. (Alternately, we could let $(\Omega, \mathcal{A}_L, P_L)$ be the Loeb space constructed from ${}^*\lambda$ on ${}^*[0,1]$.) Put $A = \{\omega \in \Omega \ : \ \omega \geq {}^\circ\omega\}$. (Note that if we take intervals of form $[a,b)$ as the basis for a topology on $\mathbb{R}$ – the "half-open" (or Sorgenfrey) topology – then A is just $NS({}^*[0,1])$ with respect to this topology.) I leave it as an exercise to show that A is not P_L-measurable (Exercise 12). Incidentally, this example will be very useful later.

In the other direction, let I^- be the result of eliminating all dyadic rationals from the unit interval $I = (0,1)$. $\lambda(I^-) = 1$, and any $x \in I^-$ has a unique infinite dyadic expansion $x = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$, where each x_i is 0 or 1. Let $k \in {}^*\mathbb{N} \setminus \mathbb{N}$, and put $B = \{x \in I^- \ : \ {}^*x_k = 1\}$. This is the intersection of the standard reals with the internal set $\{x \in {}^*I^- \ : \ {}^*x_k = 1\}$, and so one might guess that B is Lebesgue measurable; however, it is in fact nonmeasurable (Exercise 13).

Each of these sets illustrates an important principle in nonstandard measure theory. Nonmeasurability of A implies that the standard part map from ${}^*X$ to X is not inevitably a measurable function. Sets like B prevent us from adopting an otherwise attractive method of "pushing down" nonstandard measures into standard ones.

3. Constructing standard measures

3.1. MEASURABILITY OF THE STANDARD PART MAP

Suppose we want to construct a measure on a set X. A natural approach is to first construct a nonstandard measure, say on some $\Omega \subseteq {}^*X$, then somehow push it down to X.

For example, if $(\Omega, \mathcal{A}_L, P_L)$ is a Loeb measure with $\Omega \subseteq {}^*X$, then we might try to define a measure μ on X by $\mu({}^\sigma E) = P_L(E)$, where ${}^\sigma E$ is the set of standard elements of E. The set B from the last example shows that this approach might add too many sets to the σ – algebra on X. In fact, for many measures P_L the measure μ will be degenerate.

Another natural attempt would be to define ν on X by $\nu(E) = P_L({}^*E \cap \Omega)$ for E in some reasonable σ – algebra on X. Unfortunately, as with μ, ν could well end up degenerate. Moreover, ν might fail to be countably additive. (Exercise 14.)

In fact, the most common way the measure P_L is pushed down to X is by fixing a σ – algebra on X, a measurable function ϕ from Ω to X, and letting the measure on X be the image of P_L under ϕ, that is, $\eta(E) = P_L(\phi^{-1}E)$. Usually X is a topological space, and the σ – algebra we fix on X is either $\mathcal{B}[X]$ (the Borel subsets of X) or $\mathcal{B}_a[X]$ (the Baire subsets of X – see below for definition). For the map ϕ we take the standard part map on X.

To make this work we need to address two questions:

1. Is $\Omega \cap NS({}^*X)$ measurable?
2. Is $\mathrm{st} : \Omega \cap NS({}^*X) \longrightarrow X$ measurable?

These questions are, of course, related: if st is measurable then so is $NS({}^*X) = \mathrm{st}^{-1}X$. The converse is less obvious but also true for spaces X of interest; see Theorem 3.2 below.

When X is $\mathbb{R}$ and Ω is S-dense in ${}^*\mathbb{R}$ then there is a well-established literature relating the descriptive set theory of Ω to that of $\mathbb{R}$; see [17] and the references given there. In this case, it generally holds that a set A is at some level of the Borel hierarchy over $\mathbb{R}$ if and only if $\mathrm{st}^{-1}(A)$ is at the same level of Ω.

The situation is less straightforward for more general spaces X. Write $\mathcal{C}\,[X,\mathbb{R}]$ for the set of continuous functions from X to $\mathbb{R}$. For $f \in \mathcal{C}\,[X,\mathbb{R}]$ let $\mathcal{Z}\,[f]$ be the zero set of f, that is, $\mathcal{Z}\,[f] = \{x \in X \ : \ f(x) = 0\}$, and let $\mathcal{Z}\,[X]$ be the collection of all such zero sets, $\mathcal{Z}\,[X] = \{\mathcal{Z}\,[f] \ : \ f \in \mathcal{C}\,[X,\mathbb{R}]\}$. Let $\mathcal{B}_a[X]$ denote the *Baire* σ *–* *algebra*, the smallest σ – algebra generated containing $\mathcal{Z}\,[X]$.

The following result of Henson [10] has not been much improved on:

Theorem 3.1 *Let X be a completely regular topological space, $E \subseteq X$, and suppose for some compactification Y of X, $E \in \mathcal{B}_a[Y]$. Then $\mathrm{st}_X^{-1}(E) \in$ Borel$({}^*\mathcal{B}[X])$.*

(Recall that X is *completely regular* provided singletons are closed and points can be separated from closed sets by continuous functions – in other words, if $E \subseteq X$ is closed and $x \in X \setminus E$ then for some $f \in \mathcal{C}\,[X,\mathbb{R}]$, $f(x) = 1$ and $f = 0$ on E. The requirement that singletons be closed is really only there to ensure that X is Hausdorff, so that st_X exists; the reader is invited to try emulating Aldaz [2] by dropping this requirement and finding sensible reformulations of the results in this section.)

In particular, the standard part map is Baire measurable from compact Hausdorff spaces, a result already known to Loeb and Anderson. This special case of the above is Exercise 15.

This result is adequate for building Baire measures on all reasonable topological spaces, and of course Borel=Baire for most of the spaces we run into on an everyday basis (for example, metric spaces). However, one of the real strenths of nonstandard measure theory is its ability to easily build measures in exotic situations. While there is in general no reason to expect $\mathrm{st}^{-1}(A)$ to be in Borel(Λ) for $A \subseteq X$ open – even with $\Lambda = {}^*\mathcal{P}(\Omega)$ – this isn't actually necessary. All we need is that $\mathrm{st}^{-1}(A)$ be in the *completion* of Borel(Λ) for the measure P_L that we build on Ω. The following result is therefore of interest.

Theorem 3.2 *Let X be a regular topological space, let $\mathcal{A} = {}^*\mathcal{B}[X]$, let P be an internal, finitely-additive *probability measure on $({}^*X, \mathcal{A})$, and suppose that $NS({}^*X) \in \mathcal{A}_L$; then* st *is Borel measurable (i.e., from $({}^*X, \mathcal{A}_L)$ to $(X, \mathcal{B}[X])$). If X is completely regular, then $\mathcal{A}$ need only be ${}^*\mathcal{B}_a[X]$.*

3.2. REMARKS

1. When X is locally compact Hausdorff this is due to Peter Loeb [23]. The general case is due to Landers and Rogge [19].
2. Landers and Rogge point out that in the regular case that P need only be defined on the star of the σ – algebra generated by a *basis* for the topology; this is useful when X is not second countable, and will be evident from the proof below.
3. In an interesting paper Aldaz [2] shows that P need only be a finitely-additive *content*, and that it need only be defined on ${}^*\mathcal{Z}\,[X]$ (instead of all of $\mathcal{B}_a[X]$). This latter is evident from the proof I give below of Theorem 3.2, which is an adaptation of Aldaz's proof.
4. The hypothesis that ${}^*P$ be defined on ${}^*\mathcal{B}[X]$ is really quite weak; for example, this includes measures with hyperfinite support. In the latter case, it will often be convenient to restrict the standard part map to the support. In particular, if $(\Omega, \mathcal{A}_L\,, P_L)$ is a hyperfinite probability space with $\Omega \subseteq {}^*X$, and if $P_L(\Omega \cap NS({}^*X)) = 1$, then the restriction st_Ω of st to Ω is measurable.
5. Conversely, given $({}^*X, \mathcal{A}, P)$ as in the theorem, there exists a hyperfinite Loeb space $(\Omega, \mathcal{P}(\Omega), P')$ with $\Omega \subseteq {}^*X$ such that the image of P'_L under st_Ω is the same as the image of P_L under st_X; see Exercise 22. This doesn't actually use the fact that X is regular, only that st is measurable.
6. The conclusion of Theorem 3.2 (that st is measurable from $\mathcal{B}_a[X]_L$) extends to st being measurable from $\mathcal{A}_L$ for *any* internal algebra $\mathcal{A}$ on ${}^*X$, as long as $\mathcal{A}$ contains the internal Borel sets (or - by an earlier remark - just the internal zero sets, when X is completely regular).
7. For most spaces X, verification that $NS({}^*X)$ is completion-measurable will require some work, depending on the particular internal measure one constructs. For X compact, however, this is automatic (since $NS({}^*X) = {}^*X$ is the zero set of the internal zero function). For some extensions, the reader is referred to Exercises 19 and 20.
8. For an example of a relatively nice space where $NS({}^*X)$ is not measurable, see Example 4.1.

PROOF. I will prove it for the case where X is completely regular; the regular case is Exercise 21. Let C be a closed subset of X; the demonstration that $\mathrm{st}^{-1}C \in \mathcal{B}_a[X]_L$ proceeds in three steps:

Step 1. Suppose U is an open set such that $\mathrm{st}^{-1}C \subseteq {}^*U$; one can find (see Exercise 17) a set $Z_U \in {}^*\mathcal{Z}\,[X]$ such that $\mathrm{st}^{-1}C \subseteq Z_U^{\mathsf{c}} \subseteq {}^*U$. Put $E = \cap\{Z_U^{\mathsf{c}} \;:\; U \subseteq X \text{ open},\ \mathrm{st}^{-1}C \subseteq {}^*U\}$.

Step 2. Show that $\mathrm{st}^{-1}C = E \cap NS({}^*X)$. The $\subseteq$ direction is clear. For $\supseteq$, let x be a nearstandard element of E, and suppose (for a contradiction) that ${}^\circ x \notin C$. Since X is completely regular, it is regular, which means that we can find disjoint open sets U and V such that $C \subseteq U$ and ${}^\circ x \in V$. Note that $\mathrm{st}^{-1}C \subseteq \mathrm{st}^{-1}U \subseteq {}^*U$, so $x \in Z_U^{\mathsf{c}} \subseteq {}^*U \subseteq {}^*V^{\mathsf{c}} \subseteq {}^*X \setminus \{x\}$, a contradiction.

Step 3. It remains to show that $E \in \mathcal{B}_a[X]_L$; but this follows from a stronger result, for which the reader is directed to Exercise 18.

$\dashv$

3.4. APPLICATION: EXISTENCE OF HAAR MEASURE

In this section I use the results above to give a very simple proof of the existence of Haar Measure.

Let X be a compact Hausdorff *topological group*; this means that in addition to its topological structure X has a group structure (say with multiplication, and identity e), and that the functions $x \mapsto x^{-1}$ and $x \mapsto gx$ are continuous in x (for any $g \in X$). Sometimes it will be useful to write multiplication in functional form – that is, $g(x)$ instead of gx - to emphasize the functional nature of the operation.

A Borel probability measure $(X, \mathcal{B}, \mu)$ is called *Haar measure* provided it is invariant under the group operations, that is, for every $g \in X$ and $B \in \mathcal{B}$, $\mu(B) = \mu(g^{-1}B)$. It is easy to see that Haar measure, if it exists, must be unique.

Theorem 3.3 *Haar measure exists.*

PROOF. There exists an internal *neighborhood u of e such that $e \in u \subseteq \mathrm{monad}(e)$. (Why?) Evidently $u^{{}^*X} = \{g(u) \;:\; g \in {}^*X\}$ is an internal open cover of the *compact set ${}^*X$, so there is a hyperfinite subcover u^{Ω}. Choose $\Omega = \{\omega_1, \ldots, \omega_H\}$ so that H is *minimal. Let ${}^*P$ be the uniform *probability measure on $(\Omega, \mathcal{P}(\Omega))$. By Theorem 3.2 and the subsequent remarks, the restriction st_Ω of the standard part map to Ω is Borel measurable. Let μ be the corresponding Borel image measure (under st_Ω) on $(X, \mathcal{B})$.

Clearly μ is a probability measure (why?); it remains to show that μ is invariant. Pick $g \in X$ and $B \in \mathcal{B}$. Let A be any internal subset of $\mathrm{st}^{-1}B$, and let $C = \bigcup_{a \in A} \{\omega \in \Omega \ : \ \omega(u) \cap g^{-1}(a(u)) \neq \emptyset\}$. I'll show (i) that $C \subseteq \mathrm{st}^{-1}(g^{-1}B)$, and (ii) that $\|C\| \geq \|A\|$; this implies that $\mu(g^{-1}B) \geq \mu(B)$ (why?). Since g and B were arbitrary, μ must be invariant.

For (i), let $\omega \in C$. For some $x \in {}^*X$ and $a \in A$, $x \in \omega(u) \cap g^{-1}(a(u))$. Then $\omega \approx x \approx g^{-1}(a)$ (see Exercise 24). It follows from continuity of g^{-1} that ${}^\circ\omega \approx g^{-1}({}^\circ a)$, which of course means ${}^\circ\omega = g^{-1}({}^\circ a)$ (why?), which proves (i).

Now, note that $u^{(\Omega \setminus A) \cup gC}$ is an internal open cover of ${}^*X$ (see Exercise 25), so by minimality of H $\|\Omega\| \leq \|(\Omega \setminus A) \cup gC\| \leq \|\Omega\| - \|A\| + \|C\|$, which verifies (ii). $\dashv$

3.5. REMARKS

1. This is a special case of a much more general (standard) existence theorem for invariant measures; see Ross [27]. It is worth noting that the more general result was previously unknown, required very little work beyond what appears here, and was discovered using the nonstandard machinery. To my knowledge a standard proof of that result has not yet appeared.
2. The technique used for distributing points in the above proof (namely, by choosing them as 'centers' of a minimal cover) is adapted from theory of dynamical systems. This is a common phenomenon in nonstandard proofs — techniques from one field lead very naturally to results in an apparently different field.
3. In the above proof, an appeal to an internal form of Hall's *Marriage Theorem* will show that every $g \in X$ is P_L–almost surely the standard part (in the uniform topology) of an internal permutation of Ω. (Exercise 26.) In fact, it is not difficult to show that this is both necessary and sufficient for μ to be g-invariant.
4. In particular, this gives another way to confirm that the construction in Example 2.4 really leads to Lebesgue measure.

4. Representing standard measures

In Section 3.4 we used Loeb measure to construct a new measure on a given topological space. We've also seen (Example 2.4) that Loeb measure can be used to represent a known measure, in this case Lebesgue measure. In both cases, the standard measure was the measurable *image* of the Loeb measure under the standard part map.

In this section I discuss the general question of which standard measures arise in this way as the image of Loeb measures.

4.1. RADON SPACES

The first major result along these lines was due to Robert Anderson [4]. Recall that a probability space $(X, \mathcal{B}, \mu)$ is *Radon* provided X is a Hausdorff topological space, $\mathcal{B}$ is the Borel $\sigma -$ algebra over X, and μ is *compact-inner-regular*, that is, $\mu(E) = \sup\{K \subseteq E \ : \ K \text{ compact}\}$ for $E \in \mathcal{B}$. Note that such a measure is open-outer-regular.

Theorem 4.1 *(Anderson) Let* $(X, \mathcal{B}, \mu)$ *be a Radon probability measure; then* st *is measure-preserving from* $({}^*X, {}^*\mathcal{B}_L, {}^*\mu_L)$ *to* $(X, \mathcal{B}, \mu)$.

PROOF. Let $E \in \mathcal{B}$, $\epsilon > 0$, and choose K compact, U open with $K \subseteq E \subseteq U$ and $\mu(U) - \mu(K) < \epsilon$. Note that ${}^*K \subseteq \text{st}^{-1}K \subseteq \text{st}^{-1}E \subseteq \text{st}^{-1}U \subseteq {}^*U$, and ${}^*\mu_L({}^*U) - {}^*\mu_L({}^*K) < \epsilon$. Since ϵ was arbitrary, $\text{st}^{-1}E \in {}^*\mathcal{B}_L$, and $\mu(E) = {}^*\mu_L(\text{st}^{-1}E)$. $\dashv$

This shows that Loeb spaces can be used to represent most standard spaces of interest. For example, it is well-known that every Borel probability measure on a Polish space is Radon.

A partial converse to Anderson's result exists. Suppose that X, $\mathcal{A}$, and P are as in Theorem 3.2, and let μ be the image Borel measure on X under the standard part map, i.e., $\mu(E) = P_L(\text{st}^{-1}E)$ for E Borel. For each $n \in \mathbb{N}$ there is an internal set $A_n \subseteq \text{st}^{-1}E$ such that $P(A_n) > P_L(\text{st}^{-1}E) - \frac{1}{n}$. Put $E_n = \text{st}(A_n)$. Since X is regular, E_n is compact (Exercise 27), and $\mu(E_n) = {}^*\mu_L(\text{st}^{-1}E_n) \geq {}^*\mu_L(A_n) \longrightarrow \mu(E)$ as $n \to \infty$. This proves that μ is Radon. (A slight generalization is Theorem 5.1, below.)

4.2. COMPACT SPACES

Less is known about representations of more general spaces. For example, [30] examines the case of so-called *compact* (or *completely pure*) probability spaces. These are (possibly nontopological) spaces which are inner-regular with respect to a compact family $\mathcal{K}$ of measurable sets. (A family $\mathcal{K}$ of sets is *compact* provided that for every subfamily $\mathcal{K}'$ with the finite intersection property, $\bigcap\mathcal{K}' \neq \emptyset$. A family has the *finite intersection property* provided every finite subfamily has nonempty intersection.)

It turns out that like Radon spaces, in the presence of sufficient saturation compact spaces are the measurable images of Loeb spaces. However, the proof reveals that if the measurable σ – algebra of such a space is fattened up a bit, then there is a topology for which the space is Radon. This makes this result a little less significant.

However, it *does* raise the interesting question of whether Loeb spaces are themselves compact. If so, then one can obtain Loeb spaces as the image of other, more saturated Loeb spaces. This ought to be useful for, well, something. (The question itself is interesting for other, more concrete reasons; see, for example, [31] or [32].) Jin and Shelah [14] have recently almost completely resolved the question of whether Loeb spaces are compact; the answer depends largely on one's underlying set theory.

4.3. OUTER MEASURES AND CONTENTS

As mentioned in Section 4.1, if X is regular and μ is the image under the standard part map of a Loeb measure $({}^*X, \mathcal{B}[X]_L, P_L)$, then μ is Radon. This is a consequence of the inner-regularity of P_L with respect to internal sets. If the representation is weakened so that this inner-regularity somehow disappears, then the image need not be Radon.

I'll begin with a motivating example.

Example 4.1 *There is a separable metric space X and a Loeb space $({}^*X, {}^*\mathcal{B}[X]_L, P_L)$ such that ${}^*X$ has inner measure* 0 *and outer measure* 1 *with respect to P_L.*

PROOF. Let X be a subset of $[0,1]$ such that for any uncountable closed subset A of $[0,1]$, both $A \cap X$ and $A \cap X^{\complement}$ are nonempty. (Such an X is easily constructed using transfinite induction.) X inherits a separable metric structure from $[0,1]$, and it is easy to verify that $B \subseteq X$ is Borel in X if and only if $B = X \cap B'$ for some Borel subset B' of $[0,1]$. Let m be any finitely-additive extension of Lebesgue measure to all subsets of $[0,1]$ (see Exercise 3). Let P be the restriction of ${}^*m$ to ${}^*\mathcal{B}[X]$.

If $A \in {}^*\mathcal{B}[X]$ is a set of nearstandard points of ${}^*X$, then $A' = \mathrm{st}_X(A)$ is compact, hence countable, so for any $\epsilon > 0$ is contained in an open subset U_ϵ of Lebesgue measure less than ϵ. Evidently $A \subseteq {}^*X \cap {}^*U_\epsilon \subseteq {}^*U_\epsilon$, so $P(A) \leq {}^*m(U_\epsilon) < \epsilon$. It follows that the P_L-inner measure of $NS({}^*X)$ is 0. The same argument applied to $[0,1] \setminus X$ shows that the P_L-outer measure of $NS({}^*X)$ is 1 $\dashv$

What makes the space X so badly behaved (when its parent space $[0,1]$ is well-behaved) is its Radon-resistance: it contains no uncountable compact sets, so no atomless measure on X is Radon. Remarkably, it is still possible to use the space $({}^*X, {}^*\mathcal{B}[X]_L, P_L)$ to put a nontrivial Borel measure on X, by appealing to the following theorem:

Theorem 4.2 *(Landers and Rogge)Let X be a regular Hausdorff space, and let P be an internal finitely-additive probability measure on $({}^*X, {}^*\mathcal{B}[X])$. Then $\mu = \overline{P_L} \circ \mathrm{st}^{-1}$ is a Borel probability measure X.*

(Here $\overline{P_L}$ is Loeb outer measure, defined in the obvious way.)

The reader is referred to Landers and Rogge [19] for the proof, which is very much like that of Theorem 3.2. That important paper contains quite a few extensions and related results, including applications to standard measure theorey, and is strongly recommended to the interested reader.

In the Example 4.1 above, this theorem guarantees a Borel measure on X which won't be Radon, but (as the reader can verify) *will* be nontrivial.

5. Measurable Functions

Suppose that F is an internal function from a Loeb space $(\Omega, \mathcal{A}_L, P_L)$ to the star of a topological space X, that F is internally *Borel measurable, and that $F(\omega)$ is nearstandard for almost every ω. If X is one of the spaces with st_X Borel measurable (see Section 3.1) then $f = {}^\circ F$ will be a measurable function, since it will then be the composition of a Borel($\mathcal{A}$)–Borel(${}^*\mathcal{B}[X]$) function and a Borel(${}^*\mathcal{B}[X]$)–$\mathcal{B}[X]$ function.

However, by Theorem 3.2 all that one really needs for (completion) measurability of f is that X be regular. This is because $(\Omega, \mathcal{A}_L, P_L)$ induces an internal image measure $({}^*X, {}^*\mathcal{B}[X], P')$ via F, and the condition that $F(\omega)$ is nearstandard for almost every ω means that $NS({}^*X)$ is measurable (with P'_L-measure one).

It follows that f induces an image probability measure μ on $(X, \mathcal{B}[X])$; moreover, by the remarks at the end of Theorem 4.1 μ will be Radon. This proves the following:

Theorem 5.1 *Let $(\Omega, \mathcal{A}_L, P_L)$ be a Loeb probability space, X a regular topological space, and suppose F is an internal $\mathcal{A}$–${}^*\mathcal{B}[X]$ measurable function with $F(\omega)$ nearstandard for almost all ω. Then $f = {}^\circ F$ is $\mathcal{A}_L$–$\mathcal{B}[X]$ measurable, and the induced Borel measure $\mu = P_L \circ f^{-1}$ is Radon.*

A natural question is whether there is a converse, that is, whether every measurable function from the Loeb space $(\Omega, \mathcal{A}_L, P_L)$ is the standard part of an internal function. This kind of internal approximation of a measurable function will be called a *lifting*; there are two kinds, depending on whether the Loeb space in question exists on its own, or as a representation of a standard measure space.

5.1. UNIPEDAL LIFTINGS

Suppose that $(\Omega, \mathcal{A}_L, P_L)$ is a Loeb space, that X is a Hausdorff topological space, and that f is a measurable function from Ω to X. An internal function $F : \Omega \longrightarrow X$ is a *lifting* of f provided $f = {}^\circ\!F$ P_L-almost surely. (Usually we will require that F be $\mathcal{A}$–Borel(${}^*X$) measurable.)

Theorem 5.1 asserts that for reasonable X, if f has a lifting then f is measurable. We will see below that for second-countable X, if f is measurable then f as a lifting. One of the first results in the theory, this result led to speculation that if X has a basis of cardinality less than the saturation of the model, then all measurable X-valued functions would have liftings. The following example shows that this is not the case.

5.1.1. *An important example*

In this section I give an example of a Loeb measurable function with no lifting.

Let $(\Omega, \mathcal{A}_L, P_L)$ be any Loeb preimage of Lebesgue measure m on $X = [0,1]$. Give X the "half-open" (or Sorgenfrey) topology mentioned above, generated by sets of form $[a,b)$ with $a,b \in [0,1]$. Note that the Borel sets for this topology are the same as those for the usual topology. Let f be the standard part map with respect to the usual topology; since f is measurable with respect to the usual topology, it is measurable with respect to the half-open topology. Write st for the usual standard part, and st_1 for the standard part with respect to the half-open topology (which unlike st is not defined on all of X).

Suppose F is a lifting of f, then there is an internal E of positive measure with $f = \mathrm{st}_1 \circ F$ on E. In particular, $F(E) \subseteq NS_1({}^*X) = \{x \in {}^*X \ : \ x \geq \mathrm{st}(x)\}$. Put $E' = f(E) = \mathrm{st}_1(E)$. For any standard $x \in (0,1]$ and any $z < x$ with $x - z \approx 0$, $(z,y) \cap F(E) = \emptyset$. By saturation (or overspill, if you prefer) there is a standard $z = z_x \in [0,x)$ such that ${}^*(z,x) \cap F(E) = \emptyset$. It follows that $(z,x) \cap E' = \emptyset$, and $\{(z_x, x) \ : \ x \in E'\}$ is a collection of disjoint open intervals. This means that E' is countable, so $P_L(E) \leq P_L(\mathrm{st}_1^{-1} E') = m(E') = 0$, a contradiction.

5.1.2. *Existence results*

The question of which Loeb measurable functions have liftings was largely settled in [29]. Statement of the full result requires some descriptive set theory, in particular the notion of a $\mathcal{K}$-analytic set (where $\mathcal{K}$ is a family of subsets of a set X). Rather than go into the details here, readers unfamiliar with this notion should just think of $\mathcal{S}(\mathcal{K})$ (the class of $\mathcal{K}$-analytic sets) as a slight extension of Borel($\mathcal{K}$). (In fact, a set E is $\mathcal{K}$-Borel precisely when both E and E^{c} are $\mathcal{K}$-analytic.)

Fix a Loeb space $(\Omega, \mathcal{A}_L, P_L)$ and a Hausdorff topological space X.

If $f : \Omega \longrightarrow X$, let $\Gamma(f)$ be the graph of f, $\Gamma(f) = \{(\omega, x) \ : \ f(\omega) = x\}$, and let $\Gamma^\circ(f) = \{(\omega, x) \ : \ f(\omega) \approx x\}$ (a kind of thickened graph of f).

Put $\mathcal{K}_0 = \{E \in {}^*\mathcal{P}(\Omega \times {}^*X) : \text{proj}_\Omega(E) \in \mathcal{A}\}$, $\mathcal{K}_1 = \{E \in \mathcal{K}_0 : E \subseteq \Omega \times NS({}^*X)\}$, and $\mathcal{K}_2 = \{\phi(E) : E \in \mathcal{K}_1\}$ (where $\phi((\omega, x)) = (\omega, {}^\circ x))$.

Call a set E *almost* $\mathcal{S}(\mathcal{K}_i)$ provided that for some $D \in \mathcal{S}(\mathcal{K}_i)$, $P_L(D \triangle E) = 0$. E has a $\mathcal{S}(\mathcal{K}_i)$*-almost section* if for some $D \in \mathcal{S}(\mathcal{K}_i)$, $P_L(\text{proj}_\Omega D) = P_L(\text{proj}_\Omega E)$ and $P_L(\text{proj}_\Omega(D \setminus E)) = 0$.

The following is a special case of Theorem 3.1 in [29]; the proof is beyond the scope of these notes.

Theorem 5.2 *Let $(\Omega, \mathcal{A}_L, P_L)$ be a Loeb space, X a Hausdorff space, and $f : \Omega \longrightarrow X$ be measurable. The following are equivalent:*

(a) f has a lifting

(b) $\Gamma(f)$ is almost $\mathcal{S}(\mathcal{K}_2)$

(c) $\Gamma^\circ(f)$ has an $\mathcal{S}(\mathcal{K}_0)$-almost section.

5.1.3. *Applications*

The first application is a result mentioned above, due to Anderson [4] and Loeb [23]. A direct proof – not appealing to Theorem 5.2 – is outlined in Exercise 28

Corollary 5.1 *Every Loeb-measurable function into a second-countable topological space has a lifting.*

PROOF. Let $f : \Omega \longrightarrow X$ be Loeb measurable, and let $\{u_n : n \in \mathbb{N}\}$ be a countable basis for X. Let $B_n \in \mathcal{A}$ such that $A_n = (f^{-1}(u_n) \triangle B_n)$ has P_L-measure 0. Put $D = \bigcap_{n \in \mathbb{N}} (B_n \times {}^*u_n) \cup (B_n^{\complement} \times {}^*X)$, which is evidently in $\mathcal{S}(\mathcal{K}_0)$. Let $A = \bigcup_n A_n$, which has P_L-measure zero.

If $\omega \notin A$, then $\emptyset \neq \bigcap\{{}^*u_n : f(\omega) \in u_n\} \subseteq \bigcap\{{}^*u_n : \omega \in B_n\} \subseteq \{\omega\} \times D$. If in addition $9\omega, x) \in D$ and $f(\omega) \in u_n$ then $\omega \in B_n$ whence $x \in {}^*u_n$. It follows that D is an $\mathcal{S}(\mathcal{K}_0)$-almost section of $\Gamma^\circ(f)$, so by Theorem 5.2 f has a lifting. ⊣

This result guarantees liftings into separable metric spaces. It is interesting to note that it actually extends to *arbitrary* metric spaces. The reason is the following difficult theorem (which requires a bit more saturation than normal — for example, the so-called *Special Model Axiom* suffices; see Renling Jin's lectures for more details).

Theorem 5.3 *Let f be a Loeb measurable function into a metric space X; then there is a separable subspace X' such that $f(\omega) \in X'$ for almost all ω.*

The reader is referred to [31] and [32] for the proof. The extension of Corollary 5.1 to arbitrary metric spaces is an immediate consequence.

For more applications of 5.2, the reader should see [29]. That paper also contains some generalizations, for example to infinite measures (in fact, to arbitrary capacities), and to functions taking values in nonstandard hulls of Banach spaces.

5.2. BIPEDAL LIFTINGS

The above liftings might be called *unipedal* liftings, since the corresponding commutative diagram only has one leg. Another notion of lifting, a *bipedal* lifting, applies when the Loeb space is a representation of a previously given standard measure space.

In particular, suppose that $(X, \mathcal{B}[X], \mu)$ is a Radon probability space, that $(\Omega, \mathcal{A}_L, P_L)$ is a Loeb preimage of μ under the standard part map, that Y is another Hausdorff topological space, and that f is a (standard) measurable function from X to Y. An internal ${}^*\mathcal{B}[X]$–${}^*\mathcal{B}[Y]$ measurable function F is a (two-legged, or bipedal) *lifting* of f provided $f \circ \mathrm{st}_X = \mathrm{st}_Y \circ F$ P_L-almost surely.

Suppose, for example, that f is a continuous function from X to Y; then the restriction of ${}^*f$ to Ω is a bipedal lifting of f (see Exercise 30).

In fact, this is more-or-less the only possible case. Call a measurable function $f : X \longrightarrow Y$ *Lusin measurable* provided that for every $\epsilon > 0$ there is a compact $K \subseteq X$ such that $\mu(K) > 1 - \epsilon$ and such that f is continuous on K. f is *strongly* Lusin measurable if for every $\epsilon > 0$ there is a continuous g from X to Y such that $f = g$ except on a set of measure at most ϵ. Of course, in situations where the Tietze Extension Theorem holds (for example, X normal and $Y = \mathbb{R}$) these two notions agree.

Call Ω *well-distributed* in X provided that whenever K_n is a sequence of compact subsets of X with $\mu(K_n) \longrightarrow 1$, $P_L(\Omega \cap {}^*K_n) \longrightarrow 1$. Every Radon space has at least one well-distributed Loeb preimage, namely its nonstandard hull; on the other hand, many interesting preimages are *not* well-distributed (see Exercise 32).

Theorem 5.4 *Suppose that $(X, \mathcal{B}[X], \mu)$ is a Radon probability space, that $(\Omega, \mathcal{A}_L, P_L)$ is a Loeb preimage of μ under the standard part map, that Y is another Hausdorff topological space, and that f is a (standard) measurable function from X to Y. Consider the following three statements:*

(a) f is strongly Lusin measurable
(b) f has a lifting
(c) f is Lusin measurable

Then (a)$\Rightarrow$(b)$\Rightarrow$(c). Moreover, if Ω is well-distributed in X then (c)$\Rightarrow$(b)

PROOF. (a)$\Rightarrow$(b) and (c)$\Rightarrow$(b) are Exercise 34. For (b)$\Rightarrow$(c) suppose $f = {}^\circ F$ on a set Ω' with $P_L(\Omega') = 1$. Let $K_0 \subseteq K_1 \subseteq \ldots$ be a sequence of compact subsets with $\mu(K_n) > 1 - \frac{1}{n}$, let $A_n \subseteq \Omega' \cap \mathrm{st}^{-1} K_n$ be internal with $P_L(A_n) > 1 - \frac{1}{n}$, let F_n be the restriction of (the graph of) F to A_n, and note that $F_n \subseteq NS({}^*X \times Y)$. Put $G_n = \mathrm{st}_{X\times Y}(F_n)$, which is closed as a subset of $X \times Y$. Therefore $E_n = \mathrm{proj}_X G_n = \mathrm{st}(A_n)$ is closed and a subset of K_n, so is compact. Since $G_n \subseteq \Gamma(f)$, G_n must be compact as well. It follows (Exercise 33) that the restriction of f to E_n is continuous. Moreover, $\mu(E_n) = P_L(\mathrm{st}^{-1}E_n) \geq P_L(A_n) > 1 - \frac{1}{n}$; this proves (b). ⊣

Corollary 5.2 *(Lusin's Theorem) Let $(X, \mathcal{B}[X], \mu)$ be Radon and Y second countable; then every measurable function $f : X \longrightarrow Y$ is Lusin measurable.*

PROOF. Let $(\Omega, \mathcal{A}_L, P_L)$ be a Loeb preimage of μ under the standard part map. By Corollary 5.1 the Loeb measurable function $f \circ \mathrm{st}$ has a (unipedal) lifting $F : \Omega \longrightarrow {}^*Y$. F is evidently a bipedal lifting of f, and the corollary follows from Theorem 5.4. ⊣

Corollary 5.3 *Let $(X, \mathcal{B}[X], \mu)$ be Radon, $(\Omega, \mathcal{A}_L, P_L)$ be a Loeb preimage of μ under the standard part map, and let Y be Hausdorff. If $f : X \longrightarrow Y$ is strongly Lusin measurable then ${}^*f$ is a lifting of f. In particular, this holds (Anderson) when Y is second countable and $(\Omega, \mathcal{A}_L, P_L)$ is the nonstandard hull of $(X, \mathcal{B}[X], \mu)$.*

PROOF. The first part is contained in the proof of (a)⇒(b) of Theorem 5.4. The second part follows from Lusin's Theorem and the fact that the nonstandard hull of $(X, \mathcal{B}[X], \mu)$ is well-distributed in X. ⊣

6. Integration Theory

6.1. S-INTEGRABILITY

Suppose $(\Omega, \mathcal{A}, P)$ is an internal probability space, and $F : \Omega \longrightarrow {}^*\mathbb{R}$ is an internally integrable function such that ${}^\circ F$ exists P_L-almost surely. In this section I'll discuss the relationship between ${}^*\!\int F dP$ and $\int {}^\circ F dP_L$.

If F is bounded, then the relationship is the natural one:

Theorem 6.1 *Suppose $(\Omega, \mathcal{A}, P)$ is an internal probability space, and $F : \Omega \longrightarrow {}^*\mathbb{R}$ is an internal measurable function such that F has a finite bound. Then ${}^\circ F$ is integrable and ${}^*\!\int F dP \approx \int {}^\circ F dP_L$*

PROOF. Since ${}^\circ F$ is a bounded measurable function and P_L is a probability measure, ${}^\circ F$ is Loeb integrable. Fix $n \in \mathbb{N}^+$; by boundedness of F, the collection of sets of the form $F^{-1}([\frac{k}{n}, \frac{k+1}{n}))$ (where k runs over ${}^*\mathbb{Z}$)

is (standardly) finite. It follows that there is an internal *-simple function $s_n = \sum r_i \chi_{A_i}$ taking only finitely many values, with $s_n \le F \le s_n + \frac{1}{n}$. Thus $\int {}^\circ s_n dP_L = \sum {}^\circ r_i P_L A_i \approx \sum r_i P(A_i) = \int s_n dP \le \int F dP \le \int s_n + \frac{1}{n} dP = \frac{1}{n} + \sum r_i P(A_i) \approx \frac{1}{n} + \sum {}^\circ r_i P_L A_i = \frac{1}{n} + \int {}^\circ s_n dP_L$. Also, $\int {}^\circ s_n dP_L \le \underline{\int} {}^\circ F dP_L \le \overline{\int} {}^\circ F dP_L \le \int {}^\circ s_n + \frac{1}{n} dP_L \le \frac{1}{n} + \int {}^\circ s_n dP_L$. Since n was arbitrary, the theorem follows. $\dashv$

Observe (by overspill, or saturation) that if F is finite everywhere then in fact F is standardly bounded; it follows:

Corollary 6.1 *Suppose $(\Omega, \mathcal{A}, P)$ is an internal probability space, and $F : \Omega \longrightarrow {}^*\mathbb{R}$ is an internal measurable function such that ${}^\circ F$ exists everywhere. Then ${}^\circ F$ is integrable and ${}^*\int F dP \approx \int {}^\circ F dP_L$*

The situation is more difficult when ${}^\circ F$ exists almost surely, but not everywhere. The following theorem lists a few results about this case.

Theorem 6.2 *Suppose $(\Omega, \mathcal{A}, P)$ is an internal probability space, and $F : \Omega \longrightarrow {}^*\mathbb{R}$ is an internally integrable function such that ${}^\circ F$ exists P_L-almost surely. The following are equivalent:*

1. *${}^\circ\int |F| dP$ exists and $= \lim_{n\to\infty} {}^\circ\int |F_n| dP$ (where for $n \in \mathbb{N}$, $F_n = \min\{F, n\}$ when $F \ge 0$ and $F_n = \max\{F, -n\}$ when $F \le 0$)*
2. *For every infinite $K > 0$, $\int_{|F|>K} |F| dP \approx 0$*
3. *${}^\circ\int |F| dP$ exists, and for every B with $P(B) \approx 0$, $\int_B |F| dP \approx 0$*
4. *${}^\circ F$ is P_L-integrable, and ${}^*\int F dP \approx \int {}^\circ F dP_L$*

PROOF. (1)$\Rightarrow$(3) Let $r = {}^\circ\int_B |F| dP$ (which exists by (1)) and let $n \in \mathbb{N}^+$. Then $r + \int {}^\circ|F_n| dP_L = r + \int_{B^{\mathsf{c}}} {}^\circ|F_n| dP_L \approx r + \int_{B^{\mathsf{c}}} |F_n| dP \le r + \int_{B^{\mathsf{c}}} |F| dP \approx \int_B |F| dP + \int_{B^{\mathsf{c}}} |F| dP = \int |F| dP$. Let $n \to \infty$ and apply (1) to see that $r = 0$.

(3)$\Rightarrow$(2)Let $B = \{\omega : |F(\omega)| > K\}$. Then $KP(B) \le \int_{|F|>K} |F| dP \le \int |F| dP$, which by (3) is finite; therefore, $P(B) \approx 0$. (2) now follows from (3).

(2)$\Rightarrow$(1)Note ${}^\circ\int |F_n| dP \le {}^\circ\int |F| dP$. Suppose (1) fails, that is there is an $r \in \mathbb{R}$ such that for arbitrarily large $n \in \mathbb{N}$, ${}^\circ\int |F_n| dP < r < {}^\circ\int |F| dP$. For such n, $\int |F_n| dP < r < \int |F| dP$. There is then an infinite $K \in {}^*\mathbb{N}$ such that $\int |F_K| dP < r$. It follows that $\int |F| dP = \int_{|F|>K} |F| dp + \int_{|F|\le K} |F| dP \approx \int_{|F|\le K} |F| dP$ (by (2)) $\approx \int_{|F|\le K} |F_K| dP \le \int |F_K| dP < r$, a contradiction.

(4)$\Rightarrow$(1) $\lim_{n\to\infty} {}^\circ\int |F_n| dP = \lim_{n\to\infty} \int {}^\circ|F_n| dP_L$ (by Theorem 6.1) $= \int {}^\circ|F| dP_L$ (by definition of the integral for unbounded functions) $= {}^\circ\int |F| dP$ by (4).

(1)⇒(4)${}^\circ\!\int |F|dP = \lim_{n\to\infty} {}^\circ\!\int |F_n|dP = \lim_{n\to\infty} \int {}^\circ|F_n|dP_L$ (by Theorem 6.1). Since this limit exists, it equals $\int {}^\circ|F|dP_L$ (by definition of the integral for unbounded functions). ⊣

An internal function F satisfying any of the conditions (1)–(4) in this theorem is called *S-integrable.*

Another criterion for S-integrability, often useful in stochastic analysis, is the following.

Theorem 6.3 *(Lindstrøm) Suppose $(\Omega, \mathcal{A}, P)$ is an internal probability space, $p \in (1, \infty)$, and $F : \Omega \longrightarrow {}^*\mathbb{R}$ is an internally measurable function. If $\int |F|^p dP$ is finite then F is S-integrable.*

PROOF. Exercise 35. ⊣

Now, suppose that $(\Omega, \mathcal{A}_L\ , P_L)$ is a Loeb space, and that $f : \Omega \to \mathbb{R}$ is a Loeb integrable function. We know that f has a lifting F; must F be S-integrable? In general, it need not be; see Exercise 36. However, an S-integrable lifting always exists:

Theorem 6.4 *Let $(\Omega, \mathcal{A}_L\ , P_L)$ is a Loeb space, and let $f : \Omega \to \mathbb{R}$ be measurable. Then f is integrable if and only if it has an S-integrable lifting.*

PROOF. If F is an S-integrable lifting of f then ${}^\circ F$ is integrable by Theorem 6.2, so f is integrable as well. Conversely, suppose that f is integrable. Let F be any lifting of f. Note that F_n lifts f_n for finite $n \in \mathbb{N}$, and F_n lifts f for $n \in {}^*\mathbb{N}$ infinite. Let $a = \int f dP_L$; by the Dominated Convergence Theorem, for every $m \in \mathbb{N}^+$ there is an $n_m \in \mathbb{N}$ such that $|a - \int f_{n_m} dP_L| < \frac{1}{m}$. Without loss of generality the sequence $\{n_m\}_{m\in\mathbb{N}}$ is increasing. Since F_{n_m} lifts the bounded f_{n_m}, $|a - \int F_{n_m} dP| < \frac{2}{m}$. This is true for every finite m, so is true for some infinite m, and by Theorem 6.2 F_{n_m} is an S-integrable lifting of f. ⊣

6.2. PRODUCT MEASURES

Consider two Loeb spaces $(\Omega, \mathcal{A}_L\ , P_L)$ and $(\Lambda, \mathcal{D}_L, Q_L)$. The usual product measure $P_L \times Q_L$ is formed by putting $P_L \times Q_L(A \times B) = P_L(A)Q_L(B)$ for measurable rectangles $A \times B \in \mathcal{A}_L \times \mathcal{D}_L$, then extending to the smallest σ – algebra $\mathcal{A}_L \otimes \mathcal{D}_L$ containing these measurable rectangles.

It often happens that the internal P and Q from which these spaces are constructed are internally *measures, so that an internal product measure $P \times Q$ exists. (For example, if P and Q live on hyperfinite sets Ω and Λ, and if $\mathcal{A}$ and $\mathcal{D}$ are the corresponding internal power sets, then $P \times Q$ will be defined on all of $^*\mathcal{P}(\Omega \times \Lambda)$ (why?).) The Loeb construction can be applied to this internal product measure, giving a measure $(P \times Q)_L$. It is easy to see that $(P \times Q)_L = P_L \times Q_L$ on $\mathcal{A}_L \otimes \mathcal{D}_L$; however, as Example 6.1 below shows, $\mathcal{A}_L \otimes \mathcal{D}_L$ will in general be a smaller $\sigma-$ algebra than $(\mathcal{A} \otimes \mathcal{D})_L$.

These products show up frequently in the nonstandard theory of stochastic processes.

6.2.1. *Hoover's Example*

The following result is due to Doug Hoover. It is usually proved using fairly powerful machinery, e.g. the law of large numbers and/or the monotone class theorem; for example, see Keisler [16].

Example 6.1 *Let $(\Omega, \mathcal{A}_L, P_L)$ be a *finite infinite set, let $\Lambda = {}^*\mathcal{P}(\Omega)$, and let $(\Omega, \mathcal{A}_L, P_L)$, $(\Lambda, \mathcal{D}_L, Q_L)$ be the uniform hyperfinite Loeb probability spaces over the respective sets. Put $E = \{(\omega, \lambda) : \omega \in \lambda \in \Lambda\}$. Then $E \notin \mathcal{A}_L \otimes \mathcal{D}_L$*

PROOF. (Sketch; the details are left as Exercise 37.) Let $A \in {}^*\mathcal{P}(\Omega)$ and $B \in {}^*\mathcal{P}(\Lambda)$, with $A \times B \subseteq E$. It suffices (why) to show that $P_L(A)Q_L(B) = 0$.

Suppose that $P_L(A) > 0$; in particular, $|A| \in {}^*\mathbb{N} \setminus \mathbb{N}$. Note that $B \subseteq \{\lambda \in \Lambda : A \subseteq \lambda\}$ (why?), so without loss of generality we can let B be this latter set. Evidently $|B| = 2^{|\Omega| - |A|}$ (why?), so $Q_L(B) \approx Q(B) = 2^{-|A|} \approx 0$, which completes the proof. $\dashv$

6.2.2. *Keisler's Fubini Theorem*

In view of the example in the last section, it is natural to ask to what extent standard results true for product measures hold for the measure $(P \times Q)_L$. The answer is that such results often do hold. One example is the following Fubini-type theorem, due to Keisler:

Theorem 6.5 *Let $(\Omega, \mathcal{A}_L, P_L)$ and $(\Lambda, \mathcal{D}_L, Q_L)$ be loeb spaces, and suppose $f : \Omega \times \Lambda \longrightarrow \mathbb{R}$ is $(P \times Q)_L$-integrable. Then:*

i) For P_L-almost all ω, the function $\lambda \mapsto f(\omega, \lambda)$ is Q_L-integrable;
ii) The function $\omega \mapsto \int_\Lambda f(\omega, \lambda) dQ_L$ is P_L-integrable; and
iii) $\int f d(P \times Q)_L = \int (\int f dQ_L) dP_L$

The proof is Exercises 38 and 39.

7. Probability Theory

I'll conclude these notes with two specific applications to probability theory; there will of course be more in subsequent lectures.

7.1. RANDOM VARIABLES AND CUMULATIVE DISTRIBUTION FUNCTIONS

Recall that a *random variable* is a measurable function x from some probability space $(\Omega, \mathcal{A}, P)$ to a measurable space X. In these lectures X will usually be $\mathbb{R}^n$, endowed with the Borel $\sigma -$ algebra; however, the reader should know that some of the most interesting applications of nonstandard measure theory are when X is a more exotic topological space, especially since traditional methods often break down in such cases.

Probability is concerned mainly with those properties of random variables which depend only on the image measure on X under x. For this reason reference to the domain of x is usually omitted, and shorthand expressions like $P[x \in A]$ are used in place of the more accurate $P(\{\omega \in \Omega : x(\omega) \in A\})$. When $X = \mathbb{R}^n$ the image measure is entirely determined by the *cumulative distribution function*(cdf), that is, the function $F_x : \mathbb{R}^n \longrightarrow [0,1]$ defined by $F_x(r_1, \ldots, r_n) = Pr[x_1 \leq r_1, x_2 \leq r_2, \ldots, x_n \leq r_n]$ (where $x = (x_1, \ldots, x_n)$).

7.1.1. *Representing the cumulative distribution function*

If x is an $\mathbb{R}$-valued random variable, then its cdf F_x is clearly nondecreasing, right-continuous, and satisfies $\lim_{r \to -\infty} F_x(r) = 0$, $\lim_{r \to \infty} F_x(r) = 1$. It is not too difficult to give a standard proof of the following representation theorem, but the nonstandard proof is of independent interest:

Theorem 7.1 *Let $F : \mathbb{R} \longrightarrow [0,1]$ be nondecreasing, right-continuous, and suppose F satisfies $\lim_{r \to -\infty} F(r) = 0$ and $\lim_{r \to \infty} F(r) = 1$. Then $F = F_x$ for some random variable x.*

PROOF. Let $\Omega = \{0, 1, \ldots, H-1\}$ for any $H \in {}^*\mathbb{N} \setminus \mathbb{N}$, and let Λ be an S-dense *finite subset of $\mathbb{R}$. Without loss of generality there is a $\lambda \in \Lambda$ such that ${}^*F(\lambda) \geq \frac{H-1}{H}$ (since $\lim_{r \to \infty} F(r) = 1$). For $\omega \in \Omega$ let $\tilde{x}$ be the *least $\lambda \in \Lambda$ such that $\omega \leq H{}^*F(\lambda)$. Note that $\tilde{x}$ is internal and nondecreasing.

I claim that $P_L[\tilde{x} \lessapprox r] = F(r)$ for every standard real r. Observe that if this claim is true, then $\tilde{x}$ is nearstandard almost surely; it follows that $x = {}^\circ\tilde{x}$ is defined almost surely, and $F = F_x$. (Of course, x is measurable; why?)

To prove the claim, let $r < s$ be standard real numbers. For any ω with $\tilde{x}(\omega) < s$ there is a $\lambda \in \Lambda$ with $\lambda < s$ and $\frac{\omega}{H} \leq {}^*F(\lambda)$, so $\omega \leq {}^*F(s)$. This means that $P[\tilde{x} < s] \leq {}^*F(s) + \frac{1}{H}$, so $P_L[\tilde{x} \lessapprox r] \leq F(s)$. Since F is right-continuous, we can let $s \downarrow r$ and get $P_L[\tilde{x} \lessapprox r] \leq F(r)$.

In the other direction, suppose $\frac{\omega}{H} \leq {}^*F(r)$; if we take $\lambda \in \Lambda$ with $r < \lambda < s$ (which exists since Λ is S-dense), $\tilde{x}(\omega) \leq {}^*F(\lambda) < s$, so ${}^*F(r) - \frac{1}{H} \leq$

$P[\tilde{x} < s]$, whereupon $F(r) \leq P_L[\tilde{x} < s]$. Let $s \downarrow r$, and outer-continuity of P_L ensures that $F(r) \leq P_L[\tilde{x} \lessapprox r]$. $\dashv$

7.1.2. *Remarks*

1. Since every F satisfying the hypotheses of Theorem 7.1 is the cdf for a random variable, it is reasonable to call such a function (*whatever* its origin) a cumulative distribution function.
2. Call the internal random variable $\tilde{x}$ constructed in this proof a *canonical internal version* for cdf F. Of course, $\tilde{x}$ depends on the choice of Ω and Λ
3. In the construction of $\tilde{x}$, the requirement that there be a $\lambda \in \Lambda$ such that ${}^*F(\lambda) \geq \frac{H-1}{H}$ is only a convenience, so that $\tilde{x}$ is defined everywhere. In an application it might happen that Ω and Λ are given independently of F, in which case this requirement might be violated. In such a case one can just let $\tilde{x}(\omega)$ be $\lambda_{\text{sup}} = \sup \Lambda$ whenever $\frac{\omega}{H} > F(\lambda_{\text{sup}})$; the proof of Theorem 7.1 still holds.
4. The reader should think about generalizing this construction to $\mathbb{R}^n$.

7.1.3. *Skorokhod's Theorem*

If a sequence $x_1, x_2, \cdots$ of random variables converges almost surely to a random variable x, then it is easy to see that the corresponding cdfs $F_{x_1}, F_{x_2}, \cdots$ must converge to F_x at any point of continuity of F_x. The converse is problematic, since the variables x_n need not even have the same domain. However, the following theorem does provide a useful converse of sorts:

Theorem 7.2 *(Skorokhod)Suppose $F_1, F_2, \cdots$ is a sequence of cdfs, that F is a cdf, and that $\lim_{n\to\infty} F_n(r) = F(r)$ for every r at which F is continuous; then there are random variables $x, x_1, x_2, \ldots,$ defined on a common probability space, such that $x_n \to x$ almost surely.*

PROOF. Let Ω and Λ be as in the proof of Theorem 7.1. Let $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots$ be canonical versions for $F_1, F_2, F_3, \ldots,$ and let $x_1 = {}^\circ\tilde{x}_1, x_2 = {}^\circ\tilde{x}_2, \cdots$. It remains to show that $x = \lim_{n\to\infty} x_n$ exists almost surely.

Let $r < s$ be points of continuity for F, and let $E_{r,s}$ be the set of ω where $\liminf_{n\to\infty} x_n(\omega) < r < s < \limsup_{n\to\infty} x_n(\omega)$. Letting r and s range over a countable dense set of points, it suffices to show that each $E_{r,s}$ is a nullset.

Suppose then that $\omega \in E_{r,s}$. If $x_n(\omega) < r$ then $\frac{\omega}{H} < {}^*F_n(r)$; since this is true for infinitely many n and $F_n(r) \longrightarrow F(r)$, ${}^\circ(\frac{\omega}{H}) \leq F(r)$. A similar argument shows that ${}^\circ(\frac{\omega}{H}) \geq F(s)$. It follows that $F(r) \leq F(s) \leq {}^\circ(\frac{\omega}{H}) \leq F(r)$, so $E_{r,s} = \{\omega \ : \ {}^\circ(\frac{\omega}{H}) = F(r)\}$, which is clearly a nullset. $\dashv$

7.2. A POISSON RANDOM SET

Let $(X, \mathcal{B}, \mu)$ be a Radon probability measure, and let $\lambda_0 \in \mathbb{R}^+$. Suppose $(X, \mathcal{B}, \mu)$ has a uniform hyperfinite Loeb space $(\Omega, \mathcal{A}_L, Q_L)$ as a preimage under the standard part map (by now you should know when this is possible!); it will be convenient to write $\Omega = \{1, 2, \ldots, H\}$. Put $p = \frac{\lambda_0}{H}$, $q = 1 - p$. Let $x_1, \ldots, x_H$ be an internal sequence of independent, identically-distributed random variables where $x_i = 1$ with probability p, $x_i = 0$ with probability q. (Note the probability space which is the domain of the functions x_i is *not* Q, but some other internal probability P.)

Put $Y = \{\omega \in \Omega : x_\Omega = 1\}$; Y is an internal *set*-valued random variable, a random subset of Ω.

Suppose $E \subseteq \Omega$ is internal, and $k \in \mathbb{N}$. Evidently $P[|Y \cap E| = k] = \frac{|E|!}{k!(|E|-k)!}p^k q^{|E|-k} = \frac{1}{k!}\frac{|E|(|E|-1)\cdots(|E|-k+1)}{|E|^k}(\lambda_0\frac{|E|}{H})^k(1 - \frac{\lambda_0\frac{|E|}{H}}{|E|})^{|E|}$. (The reader should fill in the missing algebra justifying the last equality.) Since $\frac{|E|}{H} = Q(E) \approx Q_L(E)$, $P_L[|Y \cap E| = k] = \frac{1}{k!}\lambda^k e^-\lambda$, where $\lambda = \lambda_0 Q_L(E)$. (Show!)

Under reasonable conditions (for example, if $X = \mathbb{R}^n$ and μ is absolutely continous with respect to Lebesgue measure) $\mathcal{Y}$ has the same property as a random closed subset of X, i.e. that for any measurable $E \subseteq X$ the cardinality of $\mathcal{Y} \cap E$ has a Poisson distribution with rate parameter $\lambda_0\mu(E)$; I leave the details to the interested reader.

8. Advertisement

This seems a good place to advertise a few papers/results which I haven't discussed above, and likely will not be mentioned elsewhere at this Institute:

1. A paper with some interesting ideas that haven't yet been exploited by others in nonstandard measure theory is Kamae's pretty proof [15] of the Birkhoff's Ergodic Theorem.
2. In [12], Henson and Wattenberg pare down the usual Loeb measure to prove Egoroff's Theorem.
3. The first new standard result proved using Loeb measures was Loeb's construction [22] of ideal boundaries in potential theory.
4. Readers looking for further examples of 'simple' applications of Loeb measures (beyond what appears in these notes) might find [28] interesting, and can even improve the result there using the tools in these notes.

9. Exercises

1. (Robinson)Show that if m is Lebesgue measure on $[0,1]$, and if $A \subseteq [0,1]$, then A is Lebesgue measurable if and only if there is a *-open set U and a *-compact set K such that $K \subseteq A \subseteq U$ and $m(K) \approx m(A) \approx m(U)$.
2. (Henson)Let $(X, \mathcal{B}, \mu)$ be a finitely-additive probability measure such that $\mu(\{x\}) = 0$ for every $x \in X$. Show that there exists a hyperfinite $\Omega \subseteq {}^*X$ such that for any $B \in \mathcal{B}$, $\mu(B) = {}^\circ(\frac{\|{}^*B\|}{\|\Omega\|})$ (where $\|B\|$ denotes the internal cardinality of B).
3. (Bernstein and Wattenberg; Banach)Show that Lebesgue measure on $\mathbb{R}$ extends to a finitely-additive measure on all subsets of $\mathbb{R}$. (Note that for simplicity I'm not requiring translation-invariance for the extension.) (Hint: Exercise 2.)
4. Suppose that $\mathcal{A}$ is a (possibly external) algebra of sets on an internal set Ω, and suppose that $P : \mathcal{A} \longrightarrow {}^*R$ is a (possibly external) function such that

   a) $P(\emptyset) \approx 0$;

   b) $P(\Omega) \approx 1$; and

   c) P is nearly finitely-additive, that is, $P(A \cup B) \approx P(A) + P(B) - P(A \cap B)$ whenever $A, B \in \mathcal{A}$

   Show that $(\Omega, \mathcal{A}, {}^\circ P)$ is a standard finitely-additive probability space.
5. Let $\mathcal{A}$ be an internal algebra on Ω; show that $\mathcal{A}$ is a $\sigma -$ algebraif and only if it is (externally) finite.
6. Suppose $A_0^o \supseteq A_1^o \supseteq A_2^o \cdots A_n^o \supseteq \cdots \supseteq A_n^i \cdots \supseteq A_2^i \supseteq A_1^i \supseteq A_0^i$ are elements of an internal set $\mathcal{A}$; show that there is an internal $B \in \mathcal{A}$ with $A_n^i \subseteq A_n^o$ for all $n \in \mathbb{N}$.
7. Verify statements (a) and (b) in the second proof of Theorem 2.1.
8. Complete the details in Example 2.3. Recall that x is said to have a Binomial(n,p) distribution when for any $k \in \mathbb{N}$ with $0 \leq k \leq n$, $Pr[x = k] = \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)}$.
9. In Example 2.3, let n be infinite, and for $i \in \mathbb{N}$ let $x_i(\gamma_1, \ldots, \gamma_n) = \rho(\gamma_i)$. Show that the sequence of random variables $x_1, x_2, \cdots$ is independent, identically distributed with $P[x_i = 1] = {}^\circ(\frac{\sum_{\gamma \in \Gamma} \rho(\gamma)}{H})$.
10. In Example 2.3, let n be infinite, and let $y(\gamma_1, \ldots, \gamma_n)$ be the least i such that $\rho(\gamma_i) = 1$. Show that if $p = {}^\circ(\frac{\sum_{\gamma \in \Gamma} \rho(\gamma)}{H})$ is strictly between 0 and 1, then y is finite P_L-almost surely. (The resulting distribution is called a *geometric* distribution.)
11. (Fisher)Complete the details in Example 2.4.
12. Show that the set A in Section 2.4 is not P_L-measurable.

13. Show that the set B in Section 2.4 is not Lebesgue measurable.
14. (Refer to Section 3.1.)Let $(\Omega, \mathcal{A}_L\ , P_L)$ a Loeb space with $\Omega \subseteq {}^*\mathbb{R}$, and put $\nu(E) = P_L({}^*E \cap \Omega)$ for Borel E. (a) Show that P can be chseon to make ν identically 0. (b) Give an example of a P for which ν is not countably additive.
15. Prove that if X is compact Hausdorff then st_X is Baire measurable.
16. A measure space is *separable* if its $\sigma-$ algebrais the completion under the measure of a countable subalgebra. Show that the uniform hyperfinite Loeb space is not separable.
17. (Refer to the proof of Theorem 3.2.) Suppose that X is completely regular, that C is a closed subset of X, that U is an open subset of X, and that $\mathrm{st}^{-1}C \subseteq {}^*U$. Find a set $Z \in {}^*\mathcal{Z}\ [X]$ such that $\mathrm{st}^{-1}C \subseteq Z^{\mathbf{c}} \subseteq {}^*U$. (Hint: For $x \in C$, let $f_x \in \mathcal{C}\ [X, \mathbb{R}]$ satisfy $f_x(x) = 1, \quad f = 0$ on $({}^{\mathbf{c}}U)$. Find a hyperfinite subset $\mathcal{F}$ of ${}^*\mathcal{C}\ [0,1]$ such ${}^*f_x \in \mathcal{F}$ for every $x \in C$, and such that for every $f \in \mathcal{F}\ , \quad f = 0$ on $U^{\mathbf{c}}$. Put $F = \sup \mathcal{F}\ $, and verify that the internal zero set $\mathcal{Z}_{\ F}$ of F works for Z.
18. (Landers and Rogge [19])Suppose that $(\Omega, \mathcal{A}, P)$is an internal finitely-additive probability space with corresponding Loeb space $(\Omega, \mathcal{A}_L\ , P_L)$, and suppose that $\mathcal{C}$ is a subset of $\mathcal{A}$such that the nonstandard model is more saturated than the external cardinality of $\mathcal{C}$. Show that $\cap\mathcal{C} \in \mathcal{A}_L$. (Hint: Without loss of generality $\mathcal{C}$ is closed under finite intersections. Let $r = \inf\{P_L(A)\ \ :\ A \in \mathcal{C}\ \}$. Fix a standard $\epsilon > 0$, find $A_o \in \mathcal{C}\ \subseteq \mathcal{A}$ such that $P(A_0) < r + \epsilon$, and use saturation to find a set $A_i \in \mathcal{A}$ such that $P(A_i) > r - \epsilon$ and $A \subseteq \cap\mathcal{C}$.)
19. Let X be a σ-compact or locally compact Hausdorff space and $(X, \mathcal{B}[X]_L, P_L)$ be a Loeb space. Show that $NS({}^*X) \in \mathcal{B}[X]$. (Hint: apply Execise 18.)
20. Let X be a complete metric space and $(X, \mathcal{B}[X]_L, P_L)$ be a Loeb space. Show that $NS({}^*X) \in \mathcal{B}[X]$. (Hint: apply Exercise 18.)
21. Prove Theorem 3.2 in the case when X is only regular and $\mathcal{A} = \mathcal{B}[X]$.
22. Suppose that $(X, \mathcal{B}[X], \mu)$ is the image of a Loeb space $(\Omega, \mathcal{A}_L\ , P_L)$ on a topological space X under the standard part map. Show that here exists a *hyperfinite* Loeb space $(\Omega, \mathcal{P}(\Omega), P')$ with $\Omega \subseteq {}^*X$ such that the image of P'_L under st_Ω is also $(X, \mathcal{B}, \mu)$. (Hint: use $\mathrm{card}(\mathcal{B})^+$-saturation.)
23. (Refer to Section 3.4.) Let $x, y \in {}^*X$ where X is a compact group; show that ${}^\circ xy = {}^\circ x {}^\circ y$.
24. (Refer to Section 3.4 and the last exercise.)Prove the assertion $\omega \approx x \approx g^{-1}(a)$ which appeared in the construction of Haar measure (in the verification of assertion (i)).
25. (Refer to Section 3.4.) In the notation of the construction of Haar measure, verify that $u^{(\Omega \backslash A) \cup gC}$ is an internal open cover of ${}^*X$.

26. (Refer to Section 3.4.) In the notation of the construction of Haar measure, show that for every $g \in X$ there is an internal permutation G of Ω such that for P_L-almost every $\omega \in \Omega$, $g(\omega) \approx G(\omega)$. (Hint: use Hall's Marriage Theorem.)
27. If X a Hausdorff regular space and E is a subset of $NS({}^*X)$ then $\mathrm{st}(E)$ is compact.
28. Give a direct proof of Corollary 5.1. (Hint: Let u_n, B_n, A_n be as in the given proof of Corollary 5.1. Use saturation to find an internal F such that for every $n \in \mathbb{N}$, $F(B_n \setminus \cup_{n\in\mathbb{N}} A_n) \subseteq {}^*u_n$.)
29. Let f be a measurable function from a Loeb space $(\Omega, \mathcal{A}_L, P_L)$ to $\mathbb{R}$, and suppose that f is bounded (above, below, or both). Show that f has a lifting F with the same bound(s).
30. Suppose that $(X, \mathcal{B}[X], \mu)$ is a Radon probability space, that $(\Omega, \mathcal{A}_L, P_L)$ is a Loeb preimage of μ under the standard part map, that Y is another Hausdorff topological space, and that f is a continuous function from X to Y. Show that the restriction of ${}^*f$ to Ω is a bipedal lifting of f
31. Let $(\Omega, \mathcal{A}_L, P_L)$ be a uniform hyperfinite Loeb space, and let ν be a measure on Ω which is absolutely continuous with respect to P_L. Show that there is an internal $p : \Omega \to {}^*[0,1]$ such that $\nu(A) \approx \sum_{\omega\in A} p(\omega)$ for all $A \in \mathcal{A}$.
32. (Refer to Section 5.2.) Show that the Loeb space in Example 2.4 is not well-distributed in $[0,1]$.
33. Let X and Y be Hausdorff spaces, and let f be a function from X to Y with $\Gamma(f)$ compact in $x \times Y$. Show that f is continuous.
34. Prove the (a)$\Rightarrow$(b) and the (c)$\Rightarrow$(b) cases of Theorem 5.4. (Hint: Combine Exercise 30 with a saturation argument.)
35. Prove Theorem 6.3. (Hint: If A is a Loeb nullset, apply Hölder's inequality to the product of $|F|$ and the characteristic function of A.)
36. Let $(\Omega, \mathcal{A}_L, P_L)$ be a Loeb space; find an internal $F : \Omega \to {}^*\mathbb{R}$ which is not S-integrable, but which lifts a Loeb-integrable function.
37. Fill in the details in the proof of Example 6.1
38. (Refer to Theorem 6.5) Let E be a $(P\times Q)_L$-nullset in $\Omega\times\Lambda$. Show that $\{\omega \in \Omega : Q_L(E^\Omega) = 0\}$ has P_L-measure 1, where $E^\Omega = \{\lambda : (\omega, \lambda) \in E\}$. Note that this is Theorem 6.5 for the characteristic function of E. (Hint: For $n \in \mathbb{N}^+$, let E_n internal with measure less than $\frac{1}{n^2}$, show that the set $\{\omega : Q(E_n^\Omega) > \frac{1}{n}\}$ has P-measure less than $\frac{1}{n}$ and contains $\{\omega \in \Omega : E^\Omega$ has Q_L-outer measure $> 0\}$.)
39. Prove Theorem 6.5. (By a standard limit argument, it suffices to assume that f is the characteristic function of a $(P \times Q)_L$-measurable set E. Let $\hat{E}$ be internal differing from E by a nullset, apply the last

exercise to show that $\omega \mapsto \hat{E}^{\Omega}$ lifts $\omega \mapsto E^{\Omega}$, then apply the internal Fubini theorem to the lifting.)

40. In the proof of Skorokhod's Theorem (Theorem 7.2), let $\tilde{y}$ be a canonical internal version of F, and show that $x_n \longrightarrow y$ almost surely.

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., and Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Aldaz, J.M. (1992) A characterization of universal Loeb measurability for completely regular Hausdorff spaces, *Can. J. Math*, **44**, pp. 673–690.
3. Anderson, R.M., (1976) A nonstandard representation for Brownian motion and Itô integration, *Israel Math. Journal* **25**, pp. 15–46.
4. Anderson, R.M., (1982) Star-finite representations of measure spaces, *Trans. Amer. Math. Soc.* **271**, pp. 667–687.
5. Bernstein, A.R., and Wattenberg, F., (1969) Nonstandard measure theory, *International Symposium on Applications of Model Theory to Algebra, Analysis, and Probability*, Holt, Rinehart and Winston, New York, pp. 171-185.
6. Cutland, N.J., (1983) Nonstandard measure theory and its applications, *Bull.London Math. Soc.*, **15**, pp. 529–589.
7. Cutland, N.J., (1995) Loeb measure theory, *Developments in Nonstandard Mathematics*, (ed. Cutland, N.J., Neves, V., Oliveira, F., and Sousa-Pinto, J.), Longman, Harlow, pp. 151–177.
8. Davis, M. (1977) *Applied Nonstandard Analysis* Wiley, New York.
9. Henson, C.W. (1972) On the nonstandard representation of measures, *Trans. Amer. Math. Soc.* **172**, pp. 437–446.
10. Henson, C.W. (1979) Analytic sets, Baire sets, and the standard part map, *Can. J. Math.* **31**, pp. 663–672.
11. Henson, C.W. and Ross, D. (1993) Analytic mappings on hyperfinite sets, *Proc. Amer. Math. Soc.* **118**, pp. 587–596.
12. Henson, C.W. and Wattenberg, F. (1981) Egoroff's theorem and the distribution of standard points in a nonstandard model, *Proc. Amer. Math. Soc.* **81**, pp. 455–461.
13. Hurd, A.E. and Loeb, P.A. (1985) *An Introduction to Nonstandard Real Analysis* Academic Press, New York.
14. Jin, R. and Shelah, S. (1996) Compactness of Loeb spaces, to appear.
15. Kamae, T. (1982) A simple proof of the ergodic theorem using nonstandard analysis, *Israel J. Math* **42**, pp. 284–290.
16. Keiser, H.J. (1984) An infinitesimal approach to stochastic analysis, *Memoirs Amer. Math. Soc.* **297**.
17. Keisler, H.J., Kunen, K., Miller, A., Leth, S. (1989) Descriptive set theory over hyperfinite sets, *J. Symbolic Logic* **54**, pp. 1167–1180.
18. Landers, D. and Rogge, L. (1985) *An introduction to Nonstandard Real Analysis* Academic Press, New York.
19. Landers, D. and Rogge, L. (1987) Universal Loeb-measurability of sets and of the standard part map with applications, *Trans. Amer. Math. Soc.* **304**, pp. 229–243.
20. Lindstrøm, T.L., (1988) An invitation to nonstandard analysis, *Nonstandard Analysis and its Applications*, (ed. Cutland, N.J.), Cambridge University Press, Cambridge, pp. 1–105.
21. Loeb, P.A. (1975) Conversion from nonstandard to standard measure spaces and applications in probability theory, *Trans. Amer. Math. Soc.* **211**, pp. 113-122.
22. Loeb, P.A. (1976) Applications of nonstandard analysis to ideal boundaries in potential theory, *Israel J. Math.* **25**, pp. 154–187.

23. Loeb, P.A. (1979) Weak limits of measures and the standard part map, *Proc. Amer. Math. Soc.* **77**, pp. 128–135.
24. Nelson, E. (1987) *Radically Elementary Probability Theory* Princeton, N.J., Princeton University Press.
25. Rao, M.M. (1971) Projective limits of probability saces, *J. Multivariate Analysis* **1**, pp. 28–57.
26. Robinson, A. (1966) *Nonstandard Analysis* North Holland, Amsterdam.
27. Ross, D.A. (1988) Measures invariant under local homeomorphisms, *Proc. Amer. Math. Soc.* **102**, pp. 901–905.
28. Ross, D.A. (1989) Yet another short proof of the Riesz representation theorem, *Math. Proc. Camb. Phil. Soc.* **105**, pp. 139-140.
29. Ross, D.A. (1990) Lifting theorems in nonstandard measure theory, *Proc. Amer. Math. Soc.* **109**, pp. 809–822.
30. Ross, D.A. (1992) Compact measures have Loeb preimages, *Proc. Amer. Math. Soc.* **115**, pp. 365–370.
31. Ross, D.A. (1996) Unions of Loeb nullsets, *Proc. Amer. Math. Soc.*, **124**, pp.1883–1888.
32. Ross, D.A. (1995) Unions of Loeb nullsets: the context, *Developments in Nonstandard Mathematics* (ed. Cutland, N.J., Neves, V., Oliveira, F., and Sousa-Pinto, J.), Longman, Harlow, pp. 178–185.
33. Stroyan, K.D. and Luxemburg, W.A.J. (1976) *Introduction to the Theory of Infinitesimals.* Academic Press, New York.
34. Stroyan, K.D. and Bayod, J.M., (1986) *Foundations of Infinitesimal Stochastic Analysis.* North Holland, Amsterdam.

AN INTRODUCTION TO NONSTANDARD FUNCTIONAL ANALYSIS

MANFRED P. H. WOLFF
Mathematisches Institut d. Universität
Auf der Morgenstelle 10
D – 72076 Tübingen
Germany
email: manfred.wolff@uni-tuebingen.de

1. Elementary nonstandard analysis of normed linear spaces

1.1. INTRODUCTION

In the following let $\mathcal{V}(X)$ be the full superstructure over an appropriate infinite set X containing $\mathbb{C}$ and containing also the normed linear spaces we want to consider. Since sometimes we also have to look at Banach spaces which are not a priori in our superstructure we first of all prove the following helpful lemma. For an explicit application see 2.6.

Lemma 1.1 *Let $\mathcal{V}(X)$ be the full superstructure over an appropriate infinite set X containing $\mathbb{C}$. Then to every separable Banach space E there exists an isometric copy H, say, in $\mathcal{V}(X)$.*

Notice that with a little more effort this assertion can be generalized to Banach spaces of density character κ strictly less than the cardinality of $\mathcal{V}(X)$).

Proof: Since E is separable there exists an increasing sequence (G_n) of subspaces G_n of dimension n such that $\bigcup_{n\in\mathbb{N}} G_n =: G_\infty$ is dense in E. To every n there exists a norm p_n on $\mathbb{C}^n$ and an isometry J_n from G_n onto $(\mathbb{C}^n, p_n) =: V_n$. By hypothesis $V_n \in \mathcal{V}(X)$. But then the following spaces are also in $\mathcal{V}(X)$.
(i) $F := \{f : f : \mathbb{N} \to \bigcup_{n\in\mathbb{N}} V_n,\ \forall n[f(n) \in V_n], (p_n(f(n)))$ bounded$\}$.
(ii) $F_0 := \{f \in F : \lim_{n\to\infty} p_n(f(n)) = 0\}$.
(iii) $H := F/F_0$.
On H we define the norm by $\|\tilde{f}\| := \limsup_{n\to\infty} p_n(f(n))$. H is a Banach

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 121–151.

space in $\mathcal{V}(X)$. Now we define $J : G_\infty \to F$ by $J(x) = \begin{cases} 0 & x \notin G_n \\ J_n(x) & x \in G_n \end{cases}$
It is easy to prove that $\tilde{J} : x \to J(x) + F_0$ is an isometry. So its continuous extension to all of E is also an isometry onto a subspace of H which therefore is also in $\mathcal{V}(X)$. □

We always use nonstandard extensions $\mathcal{V}(^*X)$ of $\mathcal{V}(X)$ which are polysaturated or at least $\aleph_1$–saturated. [15].

If E is an internal normed linear space in $\mathcal{V}(^*X)$ then we consider its finite part $\mathrm{Fin}(E) = \{x \in E : \|x\| \in \mathrm{Fin}(^*\mathbb{R})\}$ where $\mathrm{Fin}(^*\mathbb{R}) = \{t \in {}^*\mathbb{R} : t \text{ is nearstandard }\}$. $\mathrm{Fin}(E)$ is obviously an external vectorspace over $\mathbb{C}$ and $E_0 = \{x \in E : \|x\| \approx 0\}$ is a subspace. By $x \approx y$ *iff* $x - y \in E_0$ (*iff* $\|x - y\| \approx 0$) there is defined an equivalence relation on the whole space E which is compatible with the external linear operations on E viewed as a vector space over $\mathbb{C}$.

Definition 1.2 *The quotient space* $\mathrm{Fin}(E)/E_0 = \widehat{E}$ *is called the* nonstandard hull of E. *By* $q(\widehat{x}) := {}^\circ\|x\|$ *there is uniquely defined a norm on* $\widehat{E}$.

Proposition 1.3 *The nonstandard hull is always complete with respect to* q.

Though in [22], Theorem 9.2, this statement is only proved for the nonstandard hull $\widehat{^*E}$ of a *standard* space E the arguments used there are also valid in our case. So we suppress the proof.

Remark 1.4 *Let E be a standard normed linear space of our superstructure. Then by abuse of our previous definition we often denote the nonstandard hull* $\widehat{^*E}$ *of the internal space* $^*E$ *by* $\widehat{E}$ *and call it the* nonstandard hull *of E if no confusion is to be feared.*

From now on we shall write $\|\widehat{x}\|$ in place of $q(\widehat{x})$ hoping that no confusion will be generated hereby.

Now if F is a standard normed space then by $F \to {}^*F_{fin} \to \widehat{F}$ (defined by $x \to {}^*x \to \widehat{^*x}$) the space F is isometrically embedded into $\widehat{F}$ in a canonical manner. So F can be identified with a subspace of $\widehat{F}$. Its closure $\overline{F}$ then is the completion of F.

Before we start with some examples we will prove first of all a lemma on the *–linear dependance of vectors and secondly the theorem of F. Riesz on nearly orthogonal elements.

Lemma 1.5 *Let E be an internal Banach space and let* $\widehat{y_1}, \ldots, \widehat{y_n}$ *be linear independent in* $\widehat{E}$. *Then* $y_1, \ldots, y_n$ *are internally linear independent in* E.

Proof: Suppose that $\sum_1^n \alpha_k y_k = 0$. Set $\beta := \max(|\alpha_1|, \ldots, |\alpha_n|)$. (Notice that this number exists by the Transfer Principle.) Then $\sum_1^n \gamma_k y_k = 0$ where $\gamma_k := \alpha_k/\beta$ are of absolute value less than or equal to 1 and at least one of them is of absolute value 1. But then $\sum_1^n {}^\circ\gamma_k \widehat{y_k} = 0$, a contradiction. □

Now let E be an internal normed space and $A \subset E$ an internal set $\neq \emptyset$. Then by the Transfer Principle we can define $d(y, A) = \inf\{\|y-x\| : x \in A\}$ and we conclude
(i) $d(y, A) = 0$ iff $y \in \bar{A}$ (the internal closure of A in E)
(ii) For every $\varepsilon > 0$ ($\varepsilon \approx 0$ is allowed) there exists $z \in A$ such that $d(y, A) \leq \|y - z\| < d(y, A) + \varepsilon$.
So we obtain

Proposition 1.6 (F. Riesz) *Let H be a closed internal subspace of the internal normed space E, and assume $H \neq E$. Then to every $y \in E \setminus H$ there exists x in the internal span of H and y such that $\|x\| = 1$ and $d(x, H) \approx 1$.*

Proof: $y \in E \setminus H$ implies $\frac{y}{\|y\|} = z \notin H$, hence $1 \geq \alpha := d(z, H) \neq 0$. Choose $\eta \approx 0$, $\eta > 0$ arbitrarily. Then there exists $u \in H$ satisfying $\alpha \leq \|z - u\| < \alpha(1 + \eta)$. But then $x = \frac{z-u}{\|z-u\|}$ has the desired properties. For if $v \in H$ then

$$\|x - v\| = \frac{1}{\|z - u\|}\|z - \underbrace{u - \|z - u\|v}_{\in H}\| \geq \frac{\alpha}{\|z - u\|} \geq \frac{1}{1 + \eta}$$

hence $1 \geq d(x, H) \geq \frac{1}{1+\eta} \approx 1$. □

We formulate the standard version of this proposition which is proved by the Transfer Principle:

Corollary 1.7 (standard) *Let F be a standard normed linear space and let H be a closed linear subspace. Then to every (standard) ε with $0 < \varepsilon < 1$ and to every $y \in F \setminus H$ there exists x in the span of H and y such that $\|x\| = 1$ and $d(x, H) > \varepsilon$.*

Corollary 1.8 *Let E be an internal normed linear space. Then $\widehat{E}$ is locally compact iff E is internally linear isomorphic to ${}^*\mathbb{C}^n$ for some standard n.*

Proof: Assume first of all that E is not internally linear isomorphic to ${}^*\mathbb{C}^n$ for all standard n. By induction on n we construct a sequence (y_n) such that $\|y_n\| = 1$, $d(y_n, H_{n-1}) > 1/2$ where H_{n-1} is the internal vector space spanned by $y_1, \ldots, y_{n-1}$. The construction is possible by 1.3. and it does not stop by assumption. But then $\|\widehat{y}_m - \widehat{y}_n\| \geq \frac{1}{2}$ for $m \neq n$, and thus $\widehat{E}$ cannot be locally compact.
Now assume that E is internally linear isomorphic to ${}^*\mathbb{C}^n$ for some standard n. We apply the Transfer Principle to 1.7 in order to get normalized vectors $y_1, \ldots, y_n$ satisfying $d(y_k, H_{k-1}) > 1/2$, where H_k is the internal linear hull of $y_1, \ldots, y_k$ (set $H_0 = \{0\}$). Then $d(\widehat{y_k}, \widehat{H_{k-1}}) \geq 1/2$, so that $\widehat{E}$ has dimension at least n. But if it would be greater than n then by 1.7 there

exists $\widehat{z}$ of norm 1 satisfying $d(\widehat{z}, \widehat{H_n}) > 1/2$ which is impossible since $\widehat{H_n} = \widehat{E}$. So $\widehat{E}$ is a normed linear space of dimension n, hence norm isomorphic to $\mathbb{C}^n$ equipped with an appropriate norm which is known to be locally compact. □

The next lemma is a general useful statement about the cardinality of hyperfinite sets.

Lemma 1.9 *Let M be a hyperfinite not standard finite internal set of internal cardinality N. Then M is (externally) not countable.*

Proof: By Transfer there exists an internal bijection f, say, from M onto $P := \{k/N : 0 \leq k \leq N-1\} \subset {}^*[0,1]$ where $[0,1]$ denotes the unit interval of $\mathbb{R}$. Also by Transfer the mapping $g : {}^*[0,1] \to P$ defined by $g(x) = \max\{k/N \in P : k/N < x\}$ is internal. Its restriction to the external subset $[0,1]$ is injective and the assertion follows. □

Proposition 1.10 *Let E be an internal normed linear space. If E has internal dimension n standard then $\widehat{E} \cong \mathbb{C}^n$; otherwise $\widehat{E}$ is nonseparable.*

Proof: The first part of our assertion follows directly from 1.8 (see also the proof of that corollary). If E does not have internal dimension n standard then there exists an infinitely large $N \in {}^*\mathbb{N}$ with $\dim(E) \geq N$. By induction on n and Transfer we construct an internal sequence (y_n) of normalized vectors satisfying $d(y_n, y_m) > \frac{1}{2}$ for $m \neq n,\ m, n \leq N$. But then $M := \{\widehat{y}_n : n \leq N\}$ is contained in the unit sphere and $\|\widehat{y}_n - \widehat{y}_m\| \geq 1/2$ holds for all $n \neq m \leq N$. By 1.9 M is not countable. □

We now give some surprising examples.

Example 1.11 *(1) Let $E = c_0 = \{x \in \mathbb{C}^{\mathbb{N}} : \lim x_n = 0\}$, equipped with the supremum norm $\|x\| = \sup\{|x_n| : n \in \mathbb{N}\}$. We denote by e_n the sequence $(\delta_{kn})_{k\in\mathbb{N}}$ where δ_{kn} is the Kronecker symbol. Let now E_n be the subspace spanned by $\{e_1, \ldots, e_n\}$. By the Transfer Principle we obtain internal hyperfinite dimensional subspaces E_N of ${}^*E$ for each hyperfinite integer N. If x is a standard element, and $y(x) = \sum_{k=1}^{N} x_k e_k$ (an internal sum) then ${}^*x \approx y(x)$ because to every standard $\varepsilon > 0$ there exists a standard n_0 such that for all $n \geq n_0$ $\|x - \sum_{k=1}^{n} x_k e_k\| < \varepsilon$.*

So the map $c_0 \ni x \to y(x) \to \widehat{y(x)}$ is an isometric embedding into $\widehat{E_N}$. In other words: the standard space E is almost contained in the hyperfinite dimensional space E_N.

(2) In the same manner we obtain for $E = l^p = \{x \in \mathbb{C} : (\sum_{k=1}^{\infty} |x_k|^p)^{1/p} = \|x\|_p < \infty\}$:
E is almost contained in the corresponding hyperfinite dimensional space

$E_N = \{x \in \ {}^*E : x_k = 0$ *for all* $k \geq N+1\}$ *which is isometrically isomorphic to* ${}^*\mathbb{C}^N$ *equipped with the norm* $\|(y_1, \ldots, y_N)\| = \left(\sum_{k=1}^{N} |y_k|^p\right)^{1/p}$

(3) l^∞ *is not almost contained in the corresponding space* $E_N = \{x \in {}^*(l^\infty) : x_k = 0$ *for all* $k \geq N+1\}$ *(consider the sequence* $(1, 1, \ldots)$*).*

Though in the last example the standard space was not "almost contained" in the hyperfinite dimensional space under consideration a stronger result is even true varying the subspace. In order to prove it we need a general lemma on the domination of upwards directed families. Notice that we always assume our extension $\mathcal{V}({}^*X)$ to be polysaturated.

Lemma 1.12 *Let* $\mathcal{A} \subset \mathcal{D} \in \mathcal{V}_n({}^*X)$ *be a partially ordered upwards directed family of internal objects and assume that the cardinality* $\kappa(\mathcal{A})$ *is strictly less than the cardinality of our standard superstructure* $\mathcal{V}(X)$*. Then there exists an element* $A \in \mathcal{D}$ *satisfying* $B \leq A$ *for all* $B \in \mathcal{A}$*.*

Proof: For all $B \in \mathcal{A}$ set $\mathcal{A}_B = \{C \in \mathcal{A} : C \geq B\}$. Since $\mathcal{A}$ is upwards directed the set $\{\mathcal{A}_B : B \in \mathcal{A}\}$ has the finite intersection property and has the same cardinality as $\mathcal{A}$. Hence polysaturation yields $\bigcap_{B \in \mathcal{A}} \mathcal{A}_B \neq \emptyset$. □

Theorem 1.13 *Let* E *be an arbitrary (standard) Banach space of our superstructure. Then there exists an internal hyperfinite dimensional subspace* F *of* ${}^*E$*, so that* E *is externally contained in* F*, that means* ${}^*x \in F$ *for all* $x \in E$*.*

Proof: Let $\mathcal{A}$ be the set of all finite dimensional linear subspaces of E. Then $\mathcal{A}$ is upwards directed by inclusion and contained in ${}^*\mathcal{A}$. Thus the assertion follows from 1.12. □

1.2. S–CONTINUOUS LINEAR OPERATORS

Let E, F be internal normed linear spaces. A map $f : D \subset E \to F$ is called S–continuous at $x \in D$ if for all $y \in D$ with $y \approx x$ we have $f(y) \approx f(x)$. f is called uniformly S–continuous if $u \approx v$ implies $f(u) \approx f(v)$ for all u, $v \in D$ (see [22]). S–continuous internal linear maps are very nice as the following proposition shows.

Proposition 1.14 *Let* E, F *be as above and let* T *be an internal linear map from* E *to* F*. The following assertions are equivalent:*
(1) T *is (uniformly)* S*–continuous.*
(2) T *is* S*–continuous at* 0*.*
(3) $T(\mathrm{Fin}(E)) \subset \mathrm{Fin}(F)$
(4) There exists a standard real number M *such that* $\|Tx\| \leq M\|x\|$ *for all* $x \in E$
(5) T *is (internally) continuous and* $\|T\| := \sup\{\|Tx\| : \|x\| = 1\}$ *is near-standard.*

Note: the supremum of an internal internally bounded subset of ${}^*\mathbb{R}$ exists always by the Transfer Principle. The proof of the proposition follows the corresponding standard proof on the continuity of standard linear maps and is left as an exercise.

Corollary 1.15 *Let $T : E \to F$ be an internal S-continuous linear map. Then by $\widehat{T} : \widehat{x} \to \widehat{T}(\widehat{x}) := \widehat{(Tx)}$ there is uniquely defined a bounded linear operator from $\widehat{E}$ to $\widehat{F}$, called the nonstandard hull of T. Its norm is given by $\|\widehat{T}\| = {}^\circ\|T\|$.*

Proof: Exercise. □

Remark 1.16 *Note that this definition is consistent with that one of the nonstandard hull of an internal normed space. Namely the graph of $\widehat{T}$ is nothing else than the nonstandard hull of the graph of T.*

Let now E, F be standard Banach spaces (in $\mathcal{V}(X)$) and denote by $\mathcal{L}(E,F)$ the Banach space of all bounded linear operators from E to F, equipped with the operator norm $\|T\| = \sup\{\|Tx\| : \|x\| = 1\}$

Corollary 1.17 *Let T be a bounded linear operator from E to F. Then ${}^*T$ is an S-continuous linear operator from ${}^*E$ to ${}^*F$ and so we can build $\widehat{{}^*T}$ as above. The map $\mathcal{L}(E,F) \ni T \to {}^*T \to \widehat{{}^*T}(=: \widehat{T})$ is a linear isometry into $\mathcal{L}(\widehat{E}, \widehat{F})$. This embedding satisfies also $\widehat{(ST)} = \widehat{S}\widehat{T}$ for $T \in \mathcal{L}(E,F)$, $S \in \mathcal{L}(F,G)$.*

Proof: Exercise. □

Remark 1.18 *If in 1.17 we set $F = \mathbb{C}$ then we get an isometric embedding of the nonstandard hull $\widehat{E'}$ of the dual space E' of E into the dual space $(\widehat{E})'$ of the nonstandard hull $\widehat{E}$ of E. More precisely:*

Corollary 1.19 *By $E' \ni \varphi \to \widehat{\varphi} \in (\widehat{E})'$ E' is isometrically imbedded into the dual space $(\widehat{E})'$ of $\widehat{E}$.*

Warning: In general $(\widehat{E})' \overset{\supset}{\neq} \widehat{E'}$ (see section 2 below)

1.3. SPECIAL BANACH SPACES AND THEIR NONSTANDARD HULL

Our first assertion says in particular that the nonstandard hull of every standard Hilbert space is a Hilbert space.

Proposition 1.20 *If H is an internal Hilbert space then $\widehat{H}$ is a Hilbert space.*

Proof: If H is an internal Hilbert space, then in particular the parallelogram law $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$ holds. But this law then holds also in $\widehat{H}$ proving the assertion. □

Let E be a Banach space over $\mathbb{C}$ and assume that there is an idempotent function $|.| : E \to E$ satisfying the following equations:

(1) $|\alpha x| = |\alpha| x|$ for $\alpha \in \mathbb{C}$, where $|\alpha|$ is the usual absolute value in $\mathbb{C}$

(2) $|\ |x| + |y| - |x+y|\ | = |x| + |y| - |x+y|$

(3) $|\ |x| - |y|\ | = |x| - |y| \Rightarrow ||y|| \leq ||x||$

(4) E is the linear hull of $\{|x| : x \in E\}$.

Then we call $(E, |.|)$ a *Banach lattice over* $\mathbb{C}$ and the function $|.|$ the *absolute value.*

One can prove [26] that $E_+ = \{x \in E : |x| = x\}$ is a cone such that $E_+ \cap (-E_+) = \{0\}$.
Moreover $E_{\mathbb{R}} = E_+ - E_+$, equipped with the order induced by E_+ is a real Banach lattice in the usual sense (see [32]) and $E = E_{\mathbb{R}} \oplus iE_{\mathbb{R}}$ is the complexification of $E_{\mathbb{R}}$, in particular $|x + iy| = \sup\{x\cos\theta + y\sin\theta : \theta \in [0, 2\pi]\}$ and $|||x + iy||| = ||x + iy||$ holds.

The way we have introduced Banach lattices enables us to show easily the following result:

Proposition 1.21 *If E is an internal Banach lattice with absolute value $|.|$ then $\widehat{E}$ is a Banach lattice with absolute value $|\widehat{x}| := \widehat{|x|}$*

Proof: This is an easy application of the Transfer Principle and is left as an Exercise. □

Corollary 1.22 *Let E be a standard Banach lattice. Then the embedding J of E into $\widehat{E}$ satisfies $|J(x)| = J(|x|)$, so E can be viewed as a Banach sublattice of $\widehat{E}$.*

Our next concrete examples are based on two well-known theorems, due to S. Kakutani, H. Bohnenblust, resp., for special cases and generalized further by others, see [21], p 135.

Theorem 1.23 *Let $(E, |.|)$ be a complex Banach lattice.*
(1) Assume that there exists $1 \leq p < \infty$ such that $||f + g||^p = ||f||^p + ||g||^p$ for all $f, g \in E$ with $\inf(|f|, |g|) = 0$. Then there exists an appropriate measure space (X, Σ, μ) and a linear positive isometric bijection from E onto $L^p(X, \Sigma, \mu)$.
(2) Assume that (a) $||\sup(|f|, |g|)|| = \sup(||f||, ||g||)$ and (b) $|f| \leq ||f|| \cdot u$ for some $u > 0$.
Then there exists a compact space K and a linear positive bijection from E onto the space $C(K)$ of all complex valued continuous functions on K.

Example 1.24 *(1) Let (X, Σ, μ) be an arbitrary measure space and $1 \leq p < \infty$. Then $E = L^p(X, \Sigma, \mu)$ is a Banach lattice and $\widehat{{}^*E} = L^p(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$, where $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ is an appropriate measure space.*
Proof: $F := \widehat{{}^*E}$ is a Banach lattice by 1.21. Let $\widehat{x}$ and $\widehat{y}$ be arbitrary with

$\inf(|\hat{x}|, |\hat{x}|) = 0$. Then $z := \inf(|x|, |y|) \approx 0$ as follows from 1.21. Now set $u := |x| - z$ and $v := |y| - z$. Then one can show that $\inf(u, v) = 0$, hence $||u||^p + ||v||^p = ||u + v||^p$ holds by Transfer and by the fact that such a formula holds in E. Apply now the quotient mapping from $\mathrm{Fin}(E)$ onto $\widehat{{}^*E}$ to this formula. □

In particular for $X = \mathbb{N}, \mu$ *the counting measure we obtain* $\widehat{l^p(\mathbb{N})} = L^p(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$. *This space contains a sublattice isomorphic to* $L^p([0,1])$.

Proof: Let $N \in {}^*\mathbb{N}$ be an infinite integer. For f continuous on $[0,1]$ we define

$$\tilde{f} \in {}^*l^p(\mathbb{N}) \text{ by } \tilde{f}(k) = \begin{cases} 0 & k > N \\ {}^*f(\frac{k}{N})/N^{1/p} & k \leq N. \end{cases}$$

Obviously $\tilde{f}$ is an internal element. We set $Uf = \hat{\tilde{f}} \in \widehat{l^p(\mathbb{N})}$. Obviously $||Uf||_p = ||f||_p$, so U can be extended to all of $L^p([0,1])$. Moreover $|Uf| = U|f|$ holds, so $U(L^p([0,1]))$ is a sublattice of $L^p(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$. □

Example 1.25 *Let K be compact and let $E = C(K)$ be the Banach lattice of all continuous complex valued functions on E. Then equipped with the usual absolute value E is a complex Banach lattice with the two additional properties: (a) $||\sup(|f|, |g|)|| = \sup(||f||, ||g||)$ and (b) $|f| \leq ||f|| \cdot u$ for some $u > 0$.*
Property (b) says that u is a strong order unit. Then by 1.23(2) there exists another compact space $\tilde{K}$ such that $\widehat{E} = C(\tilde{K})$.

The proof is based on the facts that ${}^*E$ and hence $\widehat{{}^*E}$ satisfy (1) and (2) and 1.23(2).

Let us recall the notion of a *Banach algebra.* Let E be a Banach space equipped with an associative and distributive multiplication satisfying $||xy|| \leq ||x|| ||y||$. Then E is called a Banach algebra. If in addition it possesses a multiplicative unit it is called *unital.* If E is a Banach algebra and if moreover there is an antilinear involution $^\star : x \to x^\star$ with $(xy)^\star = y^\star x^\star$ and $||x||^2 = ||x^\star x||$ for all $x, y \in E$, then E is called a *$C^\star$-algebra.* By the Gelfand–Naimark–Segal Theorem every $C^\star$–algebra is isomorphic to a subalgebra of the $C^\star$–algebra $\mathcal{L}(H)$ of all bounded operators on some Hilbert space H, equipped with the usual involution $T \to T^\star$ (adjoint of T). Moreover by a theorem of Gelfand every commutative $C^\star$–algebra is isomorphic to the algebra $C_0(K)$ of all continuous functions on the locally compact space K vanishing at infinity (if the algebra is unital, then it is isomorphic to $C(K)$ for some compact space K).

Notice that the mapping $(x, y) \to xy$ is obviously S–continuous from $\mathrm{Fin}({}^*E) \times \mathrm{Fin}({}^*E)$ into $\mathrm{Fin}({}^*E)$. But then the following result is not hard to show:

Proposition 1.26 *The nonstandard hull of an internal Banach algebra or $C^\star$-algebra E is of the same type. Moreover $\widehat{E}$ is unital iff E is unital and $\widehat{E}$ is commutative if E is.*

The converse to the last assertion is not true. More precisely there exists an internal Banach algebra E such that E is noncommutative but $\widehat{E}$ is.

Exercise: Construct such an example.

Hint Take $E = {}^*\mathbb{C}^N$ where N is infinitely large. Moreover define the multiplication by $e_ie_j = sg(i-j)\cdot\exp(-N)e_1$, where $\{e_1,\ldots,e_N\}$ denotes the arithmetical base of E and $sg(x)$ is the signum of x. Now find a norm such that E becomes a Banach algebra.

1.4. THE DUAL SPACE OF A BANACH SPACE

Let us recall the notion of the weak and weak* topology. Let E denote a normed space, and E' its dual space. Then the weak topology $\sigma(E,E')$ is given by the set of seminorms $x \to |<x,x'>|$ where x' runs through E'. Likewise the weak* topology $\sigma(E',E)$ on E' is given by the seminorms $x' \to |<x,x'>|$ where now x runs through E. The monad of 0 in E then is

$$\mu_\sigma(0) = \{x \in {}^*E : <x,x'> \approx 0 \text{ for all standard } x' \in E'\}$$

Likewise

$$\mu_{\sigma^\blacksquare}(0) = \{x' \in {}^*(E') : <x,x'> \approx 0 \text{ for all standard } x \in E\}$$

The following theorem of Banach and Alaoglu is easy to prove:

Theorem 1.27 *Let E be a normed space. Then the unit ball of the dual space E' is weak* compact.*

Proof: Let X be the unit ball of E', and let $x' \in {}^*X$ be arbitrary. $\|x'\| \leq 1$ implies that $<x,x'>$ is nearstandard for all $x \in \mathrm{Fin}({}^*E)$, in particular for all x standard.
Then y', defined by $<x,y'> = {}^\circ(<x,x'>)$ for all standard x, is linear and bounded, hence norm continuous. But by definition if $x \in E$ is standard then $<x,y'> \approx <x,x'>$, so $x' \approx y'$ with respect to the weak* topology. □

The bidual E'' of E is by definition the dual space of E'. The mapping $x \to \varphi_x$ from E to E'', where $\varphi_x(x') = <x,x'>$ is the canonical isometric linear embedding, which by definition is onto iff E is reflexive. By what we have seen above if it is reflexive then the unit ball of E is weakly compact, since the weak* topology of E'' restricted to E gives the weak topology. That the converse is also true is due to R.C. James.

1.5. NOTES

Section 1 contains results which are needed in every advanced nonstandard functional analysis. (see e.g. [23, 34, 6]). More recent introductions are to be found in e.g. [17]. Proposition 1.3 is due to Luxemburg [23].

Proposition 1.5 is a reformulation of Riesz' result and its standard proof (see e.g. [42],III.2) within the frame work of nonstandard analysis. Corollary 1.7 holds also for topological vector spaces.

An extensive study of examples like the ones presented in 1.10 is to be found in [14]. Theorem 1.12 is nothing else than a special application of saturation and Proposition 1.13 is orientated to standard facts. The consequences serve as basic facts in all advanced applications. Theorem 1.20 is also known within the frame work of ultraproducts of Banach spaces and traces back to results of Dahuna–Castelle and Krivin in the late sixties.

For a comprehensive representation of these and other results see [12, 33]. Corresponding results within the frame work of nonstandard analysis may be found in [14].

Ultraproducts of Banach algebras seem to be considered already in the late fiftieth (see [29]) and came up again in 1970 on a conference on non-standard analysis. In fact Janssen seems to be the first one who has applied such a construction to $C^\star$–algebras (see [19]) (almost identical results where proved apparently independent of [19] by Hinokuma and Ozawa [16]). Only a little later Golodets [9] and others have used ultrapower techniques e.g. in order to contribute to the classification of $W^\star$–algebras of type III. Groh [11] proved (within the context of the theory of operators on $W^\star$–algebras) that the nonstandard hull of the predual of a $W^\star$–algebra is again the predual of such an algebra .

The standard proof of the Theorem of Banach and Alaoglu (Theorem 1.22) which uses Tychonov's Theorem on products of compact spaces is almost as easy as the one we have given here. (cf. [22]).

2. Advanced theory of Banach spaces

2.1. GENERAL BANACH SPACES

Because of lack of space we only can give very few results. We refer the interested reader to [14], which in our opinion is the best reference to non-standard analysis and Banach spaces (operator theory is not treated there).

Definition 2.1 *The Banach space F is* finitely represented in E *if to each positive real number $\varepsilon > 0$ and to each finite dimensional subspace G of F there exists a linear mapping T from G into E such that $(1-\varepsilon)||x|| \leq ||Tx|| \leq (1+\varepsilon)||x||$ for all x in G.*

Using a nonstandard hull of E this notion becomes very transparent:

Proposition 2.2 *Let E, F be standard Banach spaces of our full superstructure $\mathcal{V}(X)$. The following assertions are equivalent:*
(1) F is finitely representable in E.
(2) F is finitely representable in $\widehat{E}$.
(3) F is embeddable into $\widehat{E}$, that means there exists a linear isometry from F into $\widehat{E}$.

Proof: (1) $\Rightarrow$ (3) Let F be finitely represented in E. Let $G \subset {}^*F$ be an internal closed internally finite dimensional subspace containing F as an external subset (see 1.13). Let $\eta \approx 0$ be arbitrary. Then by the Transfer Principle there exists a linear map $T : G \to {}^*E$ such that $(1-\eta)\|x\| \leq \|Tx\| \leq (1+\eta)\|x\|$ for all $x \in G$. The embedding then is given by $F \ni x \to \widehat{Tx}$.
(3) $\Rightarrow$ (1) Let U be a linear isometry from F into $\widehat{E}$, and let $\varepsilon > 0$ (standard) and $G \subset F$ be given with $\dim(G) = n$. Let $B := \{y_1, \ldots, y_n\}$ be a basis of G of normalized vectors. For each j choose $x_j \in {}^*E$ such that $\widehat{x}_j = U(y_j)$, and $\|x_j\| = 1$.
Now we consider the internal linear map
$V : {}^*G \to {}^*E$, given by $V(\sum_1^n \alpha_j y_j) = \sum \alpha_j x_j$. Then $\widehat{V} = U_{|G}$ since both mappings agree on B, hence $\|V(x)\| \approx 1$ for all x of norm 1. Since G is finite dimensional there exists by the Transfer Principle a normalized vector x_0 such that $\alpha = \|V(x_0)\| = \inf\{\|V(x)\| : \|x\| = 1\}$. Obviously $\alpha \approx 1$. Now set $\gamma = min(\alpha, 1)$ and $W = \frac{1}{\gamma}V$. Then for all x of norm 1 $1 \leq \|W(x)\| \approx 1$. Hence if $\varepsilon > 0$ is an arbitrary standard number then $(1-\varepsilon) \leq \|W(x)\| \leq (1+\varepsilon)$ for all normalized x. So the following theorem holds in $\mathcal{V}({}^*X)$, hence also in the standard world:
There exists a linear mapping W from ${}^*G$ into ${}^*E$ satisfying $(1-\varepsilon)\|x\| \leq \|W(x)\| \leq (1+\varepsilon)\|x\|$ for all $x \in {}^*G$.
Since G and $\varepsilon > 0$ were arbitrary the assertion follows.
(1) $\Rightarrow$ (2) is obvious and (2) $\Rightarrow$ (1) is left as an exercise (Hint: modify the proof of (3) $\Rightarrow$ (1)). □

In what follows we see that nonstandard hulls behave very well concerning the property of finitely representability.

Proposition 2.3 ([14], Theorem 3.2) *Let E be an internal Banach space and let F be a separable Banach space not necessarily contained in our superstructure $\mathcal{V}(X)$. Then F is finitely representable in $\widehat{E}$ iff F is embeddable into $\widehat{E}$.*

Proof: (I) By 1.1 we may assume that F is contained in $\mathcal{V}(X)$. Let F be finitely representable in $\widehat{E}$ and let G be a subspace of F of dimension n. Then to $\varepsilon = 1/n$ there exists a linear mapping T from G into $\widehat{E}$ satisfying $(1-1/(2n)) \leq \|Tx\| \leq (1+1/(2n))$ for all $x \in G$ of norm 1. Let $B :=$

$\{y_1, \ldots, y_n\}$ be a basis of G of normalized vectors. For each k choose $x_k \in E$ such that $\widehat{x_k} = Ty_k$. Then by $V(y_k) := x_k$ and internal linear extension there is defined an internal linear map V from $^*G$ into E (notice that we identify the standard elements of G with their images in $^*G$). Since obviously $\widehat{V} = T$ we have

$$\text{(P)} \quad (1 - 1/n) \leq \|Vx\| \leq (1 + 1/n) \text{ for all } x \in {}^*G \text{ of norm } 1.$$

(II) Now since F is separable there exists an increasing sequence $(G(n))$ of subspaces $G(n)$ of F of dimension n whose union G_∞ is dense in F. Then the set $A := \{n \in {}^*\mathbb{N} :$ there exists $V : {}^*G(n) \to E$ with property (P)$\}$ is internal and contains all standard n hence some N infinitely large. Denote by V_N the corresponding linear mapping. Then the continuous extension to F of the restriction to G_∞ of $\widehat{V_N}$ is the desired embedding. □

The next theorem does not follow from the proposition above since $\widehat{E}$ is not in the full superstructure we have started with.

Theorem 2.4 *Let E be a (standard) Banach space of the full superstructure $\mathcal{V}(X)$. Then $\widehat{E}$ is finitely represented in E.*

Proof: Let F be a subspace of $\widehat{E}$ of dimension n. Then by a standard result there exists an isometry J from F onto $\mathbb{C}^n =: V$, equipped with an appropriate norm. Notice that V is in $\mathcal{V}(X)$ as follows immediately from our hypothesis that $\mathbb{C}$ is contained in $\mathcal{V}(X)$. Let $\{e_1, \ldots, e_n\}$ be a basis of normalized vectors of V and set $y_k = J^{-1}(e_k)$. Then there exist $x_1, \ldots, x_n$ in $^*E$ with $y_k = \widehat{x_k}$ for each k. By 1.5 $B := \{^*e_1, \ldots, {}^*e_n\}$ is internally linear independent. Denote by H the *-linear hull of $\{x_1, \ldots, x_n\}$ in $^*E$. Then by the Transfer Principle the linear map W, defined by $W(\sum_1^n \alpha_k {}^*e_k) = \sum_1^n \alpha_k x_k$ is internal. Moreover it satisfies $\|W(x)\| \approx \|x\|$ for all finite $x \in V$ since $\widehat{W} = J^{-1}$ is an isometry. Let $\varepsilon > 0$ be an arbitrary standard real number. Then

$$\exists S \forall x \in {}^*V[\|x\| = 1 \Rightarrow (1 - \varepsilon) \leq \|S(x)\| \leq (1 + \varepsilon)]$$

holds in $\mathcal{V}(^*X)$ hence by the Transfer Principle also in $\mathcal{V}(X)$. For such an S the map $T := S \circ J$ is the desired one. □

Recall that a Banach space E is called *superreflexive* if F finitely representable in E implies F is reflexive. In order to apply our results within this context we have to recall results of R. C. James on reflexive Banach spaces. The equivalent assertion (d) below goes back to Dunford and Schwartz [7].

Theorem 2.5 (R. C. James [18]) *For a Banach space E the following assertions are equivalent:*
(a) E is reflexive.
(b) Every separable subspace of E is reflexive.

(c) For every linear functional y' in E' there exists x in E such that $\|x\| = 1$ and $< x, y' >= \|y'\|$.
(d) There exists a closed subspace F such that F as well as E/F are reflexive.

Theorem 2.6 *For a Banach space E the following assertions are equivalent:*
(1) E is superreflexive
(2) $\widehat{E}$ is reflexive
(3) $\widehat{E}$ is superreflexive
(4) $\widehat{(E')} = (\widehat{E})'$

Proof: (1) $\Rightarrow$ (3): Let F be finitely representable in $\widehat{E}$. By 2.5 we may assume without loss of generality that F is separable, hence in our superstructure $\mathcal{V}(X)$ by 1.1. Then F is finitely representable in E as follows from 2.4. So F is reflexive by (1), and since F was arbitrary (3) follows.
(3) $\Rightarrow$ (2): obvious.
(2) $\Rightarrow$ (1): If F is finitely representable in E and separable then F is embeddable into $\widehat{E}$ by 2.3, hence reflexive and (1) follows.
(2) $\Rightarrow$ (4): Suppose that $\widehat{E'}$ is a proper closed subspace of $(\widehat{E})'$. Since $\widehat{E}$ is reflexive there exists $\widehat{x} \in \widehat{E}$, $\|\widehat{x}\| = 1$ and $< \widehat{x}, \widehat{y'} >= 0$ for all $\widehat{y'} \in \widehat{E'}$. But by the Hahn–Banach Theorem and the Transfer Principle there exists $z' \in \mathrm{Fin}(E')$ satisfying $< x, z' >= 1$ and $\|z'\| \approx 1$, hence $1 =< \widehat{x}, \widehat{z'} >$, a contradiction.
(4) $\Rightarrow$ (2): Let $x' \in (\widehat{E})' = \widehat{E'}$ be arbitrary. Then $x' = \widehat{y'}$ for some $y' \in {}^*(E')$. But
$\|y'\| = \sup\{| < x, y' > | : \|x\| = 1, \ x \in {}^*E\}$. Hence by the Transfer Principle to $0 < \varepsilon \approx 0$ there exists $x \in {}^*E$ $\|x\| = 1$ such that $\|y'\| - \varepsilon < | < x, y' > | \leq \|y'\|$. This gives $\|x'\| = {}^\circ(\|y'\|) \leq {}^\circ(| < x', y' > |) = | < \widehat{x}, x' > | \leq \|x'\|$, that means 2.5,(c) is satisfied and the assertion follows. □

Using assertion (d) of this result we obtain the easy proof of Rakov (see [12] of the following theorem of Enflo, Lindenstrauss, and Pisier.

Corollary 2.7 *Let E be a Banach space and let F be a closed subspace of E. If F and the quotient space E/F are superreflexive then E is superreflexive.*

Proof: Since the quotient mapping $Q : E \to E/F$ is open we get easily $\widehat{E/F} \cong \widehat{E}/\widehat{F}$. So by the Theorem $\widehat{F}$ as well as $\widehat{E}/\widehat{F}$ are reflexive. But by 2.5,(d) $\widehat{E}$ is reflexive, and the theorem gives the desired result. □

2.2. BANACH LATTICES

There is a theory of finitely representable Banach lattices similar to the theory for Banach spaces sketched so far. Let us denote by c_0 the space of sequence $x = (x_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ with $\lim x_n = 0$, $\|x\| = \sup(|x_n|)$ (see 1.6 (1)). With $|x| = (|x_n|)_{n\in\mathbb{N}}$ it becomes a Banach lattice. Its dual space can be identified with $l^1(\mathbb{N}) = l^1 = \{x \in \mathbb{C}^{\mathbb{N}} : \Sigma|x_k| =: \|x\|_1 < \infty\}$, and it is also a Banach lattice under the canonical order.

Let us recall the notion of a lattice homomorphism: the linear operator T from the Banach lattice E into another one, F, say, is called a lattice homomorphism if $|Tx| = T|x|$ holds for all $x \in E$. Obviously the nonstandard extension $^*T$ as well as $\widehat{T}$ are lattice homomorphisms if T is such one. In full correspondence to 2.1 we define

Definition 2.8 *The Banach lattice F is* finitely lattice representable *in the Banach lattice E, if for each finite-dimensional vector sublattice $G \subset F$ and each $\varepsilon > 0$ there exists a lattice isomorphism T from G into F with $(1-\varepsilon)\|x\| \leq \|Tx\| \leq (1+\varepsilon)\|x\|$ for all $x \in G$.*

We give one example for the usefulness of this concept.

Theorem 2.9 *Let E be a Banach lattice. Then the following assertions are equivalent:*
(1) E is superreflexive.
(2) Neither c_0 nor l^1 are finitely lattice representable in E.
(3) Neither c_0 nor l^1 are lattice embeddable in $\widehat{E}$.

Proof: First of all you may prove (2) $\Leftrightarrow$ (3) generalizing 2.2 to the case of finitely lattice representable spaces (exercise). Then a famous result of Meyer–Nieberg (and others) says that the Banach lattice F is reflexive iff neither c_0 nor l^1 are lattice embeddable in F. 2.3 now yields the result. □

Examples are spaces L^p $(1 < p < \infty)$ and many other Banach function spaces.

2.3. NOTES

Almost all results of section 2 are taken from [14]. Corresponding results within the frame work of ultraproducts may be found in [12, 33] where one also will find further references to the history of the results.

C. W. Henson (see [14]) has developed a special logical language which allows to express problems concerning relations which can be approximately satisfied. Together with nonstandard analysis this gives new and deep insight into various properties of special Banach spaces and their nonstandard hulls. We recommend in particular the papers [14, 13].

Another aspect – the combination of Loeb measure theory with functional analysis – gives also interesting new results. For example a Banach space in which a ball is contained in the range of some countably additive measure is superreflexive (see [35]).

A third aspect is the use of nonstandard analysis for infinite constructions, e. g. infinite tensor products of $C^{\star}$-algebras. This interesting field has applications in quantum physics, see [19, 16, 40].

3. Elementary theory of linear operators

3.1. COMPACT OPERATORS

We have already considered internal S–continuous operators in 1.14 (operator means always linear mapping). The simplest ones are those of standard finite rank. Let T be such an operator. Then $\dim(T(E)) = n$ standard, and its nonstandard hull (see 1.15) is also of finite rank (Exercise: notice that T is of the form$T = \sum_{k=1}^{n} \varphi_k \otimes x_k$).

Another class of simple operators is the class of compact operators.

Definition 3.1 *Let E, F be a standard Banach spaces. The linear operator T from E to F is* compact *iff T maps bounded sets onto norm relatively compact sets.*

Since a relatively compact set A is characterized by the property that all elements of $^*A$ are nearstandard (see [22]) we obtain the following easy characterization of compact operators. We denote the unit ball of a Banach space by$B(0,1)$. Let us recall that in a Banach space E the set A is relatively compact iff it is precompact; that means to every $\varepsilon > 0$ there exists a finite subset $M \subset A$ with $\sup_{x \in A} d(x, M) \leq \varepsilon$.

Proposition 3.2 *Let E, F be (standard) Banach spaces and let$T : E \to F$ be a linear operator. The following assertions are equivalent:*
(1) T is compact.
(2) $^*T(\,^*B(0,1)) \subset ns(^*F)$ (the set of nearstandard points).
(3) $\widehat{T}$ maps $\widehat{E}$ into F (identified with a subset of $\widehat{F}$).
(4) $\widehat{T}$ is compact.

Proof: (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) as well as (4) $\Rightarrow$ (1) are obvious.
(1) $\Rightarrow$ (4): Let $\varepsilon > 0$ be given. Then there exists a finite set $M = \{y_1, \ldots, y_n\} \subset B(0,1)$ with $\sup_{x \in B(0,1)} d(Tx, T(M)) < \varepsilon$. By Transfer the same holds for $^*B(0,1)$. (Notice that $^*M = \{\,^*y_1, \ldots, \,^*y_n\}$). But then $d(\widehat{T}\widehat{x}, \widehat{T}(\widehat{M})) < \varepsilon$ for all $\widehat{x}$ in $\widehat{B}(0,1)$ (the unit ball now in $\widehat{F}$). Since $\varepsilon > 0$ was arbitrary, $\widehat{T}(\widehat{B}(0,1))$ is precompact. □

The uniform topology or operator norm topology on the space $\mathcal{L}(E, F)$ of all bounded linear operators from E to F is given by the operator norm $||T|| = \sup\{||Tx|| : ||x|| \leq 1\}$. Its corresponding monad is given by $\mu_{op}(\mathcal{L}(E,F)) = \{T \in {}^*\mathcal{L}(E,F) : Tx \approx 0 \text{ for all } x \in \text{Fin}({}^*E)\}$. The strong operator topology is given by the monad $\mu_{stop} = \{T \in {}^*\mathcal{L}(E,F) : Tx \approx 0 \text{ for all standard } x \in E\}$. Finally the weak operator topology is given by the monad $\mu_{wop} = \{T \in {}^*\mathcal{L}(E,F) :< Tx, y' >\approx 0 \text{ for all standard } x \in E \text{ and all standard } y' \in E'\}$.

Concerning the uniform convergence we have the following useful lemma:

Lemma 3.3 *Let E, F be standard Banach spaces, and let $(T_\alpha)_{\alpha\in A}$ be a net of bounded linear operators from E to F. Moreover let T be also a bounded operator from E to F. The following assertions are equivalent.*
(1) $(T_\alpha) \to T$ uniformly.
(2) For all $x \in \text{Fin}({}^*E)$ and all infinitely large $\alpha \in {}^*A\backslash A$ we have $T_\alpha x \approx x$.
(3) $(\widehat{T}_\alpha) \to \widehat{T}$ uniformly.

Proof: (1) $\Rightarrow$ (3): For all $\varepsilon > 0$ there exists $\alpha_0 \in A$ such that for all $\alpha \geq \alpha_0$ and all normalized x $||T_\alpha x - Tx|| < \varepsilon$ holds. By Transfer this is true in ${}^*E$ hence also in the quotient, that means we have $||\widehat{T}_\alpha \widehat{x} - \widehat{T}\widehat{x}|| \leq \varepsilon$ for all normalized $\widehat{x}$ and all $\alpha \geq \alpha_0$.
(3) $\Rightarrow$ (1): obvious.
(1) $\Rightarrow$ (2): By Transfer $||T_\alpha x - {}^*Tx|| < \varepsilon$ for all normalized x, all α infinitely large and each standard $\varepsilon > 0$. So (2) follows.
(2) $\Rightarrow$ (1): Let $\varepsilon > 0$ standard be fixed, and choose α_0 infinitely large. Then this α_0 satisfies the formula $\exists\alpha\forall\beta\forall x[||x|| = 1 \text{ and } \beta \geq \alpha \text{ implies } ||T_\beta x - Tx|| < \varepsilon]$. The Transfer Principle yields the assertion. □

Recall that the Banach space E has the bounded approximation property if the identity I is the limit in the strong operator topology of a norm bounded net of operators of finite rank. Almost all classical Banach spaces have this property, but the algebra $\mathcal{L}(\mathcal{H})$ of all bounded operators on the separable Hilbert space $\mathcal{H}$ fails to have this property (a result due to Szankowski).

Proposition 3.4 *Let E, F be arbitrary Banach spaces.*
a) The uniform limit T of a sequence (T_n) of compact operators is compact.
b) If F possesses the bounded approximation property then every compact operator T is the uniform limit of operators of finite rank.

Proof: a) All $\widehat{T}_n$ map $\widehat{E}$ into F and $(\widehat{T}_n)$ converges uniformly to $\widehat{T}$ by 3.3. So $\widehat{T}(\widehat{E}) \subset F$, and T is compact by 3.2.
b) Let $(P_\alpha)_{\alpha\in A}$ be a norm bounded net of operators of finite rank converging strongly to the identity I on F. Then $T_\alpha = P_\alpha T$ is of finite rank for all $\alpha \in A$. If $\alpha \in {}^*A\backslash A$ is infinitely large and $x \in \text{Fin}({}^*E)$ then Tx is nearstandard, and if $y = {}^\circ Tx$ is its standard part then $P_\alpha y \approx y$. But since P_α is S–bounded by hypothesis we obtain $P_\alpha Tx \approx Tx$. Since $x \in \text{Fin}({}^*E)$ and α infinitely large where arbitrary the assertion follows by 3.3.

Before we sketch some elements of Fredholm theory we prove the theorem of Schauder (sometimes also attributed to Gantmacher), that the adjoint or dual operator of a compact operator is also compact.

Theorem 3.5 *The operator T from E to F is compact iff its adjoint T' is compact.*

Proof: Let T be compact, and let $\varphi \in {}^*(F')$ be of finite norm. Then its standard part with respect to the weak $*$-topology is given by $\psi(x) = {}^\circ(\varphi(x))$ for all standard x. Let $y \in \mathrm{Fin}({}^*E)$ be arbitrary. Then ${}^*T'\varphi(y) = \varphi({}^*(Ty))$. Now ${}^*Ty$ is nearstandard, since T is compact. Hence

$$\begin{aligned} {}^*T'\varphi(y) = \varphi({}^*Ty) \quad &\approx \quad \varphi({}^\circ({}^*Ty)) \approx \psi({}^\circ({}^*Ty)) \\ &\underbrace{\approx}_{\psi \text{ S-continuous}} \quad {}^*\psi({}^*Ty) = [({}^*T)'\,{}^*\psi](y). \end{aligned}$$

Since $y \in \mathrm{Fin}({}^*E)$ was arbitrary ${}^*(T)'\varphi \approx ({}^*T)'\,{}^*\psi = {}^*(T'\psi)$ with respect to the norm topology. Hence ${}^*T'$ maps finite elements to nearstandard ones, and 3.2 gives that T' is compact.
Conversely if T' is compact so is T'' by the first part of our proof. But $T = T''|_E$, where E is identified with its canonical image in E''. □

3.2. FREDHOLM OPERATORS

Our next class of operators, Fredholm operators, play an important role in the theory of partial differential equations. We start by considering the quotient space $\mathcal{L}(E,F)/\mathcal{K}(E,F)$ where $\mathcal{K}(E,F)$ denotes the norm closed subspace of compact operators. Let Q denote the quotient map from $\widehat{F} \to \widehat{F}/F := G$. For $T \in \mathcal{L}(E,F)$ set $C(T) = Q\widehat{T} \in \mathcal{L}(\widehat{E},G)$. Then by 3.2 its kernel is $Ker(C) = \mathcal{K}(E,F)$. The norm $\|C(T)\|$ is a *particular measure of noncompactness.*

Let us consider as an example the case where $E = F =: H$ is a Hilbert space. Since $\widehat{H}$ is also a Hilbert space (see 1.20) it splits into $H \oplus H^\perp$

Proposition 3.6 ([27]) *Let $T \in \mathcal{L}(H)$ be arbitrary. Then $\widehat{T}(H^\perp) \subset H^\perp$, hence $\widehat{T}$ splits into $T \oplus T^\perp$ where $T^\perp = \widehat{T}|_{H^\perp}$.*

Proof: We know already $\widehat{T}|_H = T$. Let P be the orthogonal projection of $\widehat{H}$ onto H, and $Q = I - P$. Then $P\widehat{T} = \widehat{T}P = T$, hence $Q\widehat{T}P = 0$. But one proves easily $\widehat{(T^\star)} = \widehat{T}^\star$ (Exercise: use $\widehat{H'} = (\widehat{H})'$, cf. 2.6). So $P\widehat{T}Q = (Q\widehat{T^\star}P)^\star = 0$, and the assertion follows. □

Corollary 3.7 *The map $T + \mathcal{K}(H) \to T^\perp$ is a $*$-isomorphism from the so called Calkin algebra $\mathcal{L}(H)/\mathcal{K}(H)$ into $\mathcal{L}(H^\perp)$. (More precisely it is an algebra isomorphism which preserves adjoints).*

Returning to the general case we recall the definitions of upper and lower Fredholm operators from the Banach space E to the Banach space F.

Definition 3.8 *$T \in \mathcal{L}(E,F)$ is called an* upper Fredholm operator *if its kernel $Ker(T) = T^{-1}(\{0\})$ is finite dimensional and its range is closed. It is called a* lower Fredholm operator *if its range has finite codimension. It is called a* Fredholm operator *if it is an upper and lower Fredholm operator. The difference* $\dim(Ker(T)) - codim(T(E)) = ind(T)$ *is called the* index of T.

Lemma 3.9 (standard) *1) A lower Fredholm operator T has always a closed range.*
2) Let $T : E \to F$ be continuous. The following assertions are equivalent:
a) T has closed range.
b)

$$\inf\{\|Tx\| : x \in E \text{ and } \inf_{Ty=0}(\|x+y\|) = 1\} =: \alpha(T) > 0.$$

c)

$$\sup_{\|u\|=1}(\inf\{\|x\| : Tx = u\}) =: \gamma(T) < \infty.$$

Proof: 1) Without loss of generality we may assume T to be injective. Let $\{y_1, \ldots, y_n\}$ be elements the images of which form a basis in the quotient space $F/T(E)$. Then the mapping
$S : \mathbb{C}^n \times E \to F$ given by $S(\lambda_1, \ldots, \lambda_n, x) = \sum_1^n \lambda_j y_j + Tx$ is continuous and bijective, hence a homeomorphism by the open mapping theorem. So $T(E) = S(\{0\} \times E)$ is closed.
2) Assume that a) holds and let $\tilde{T}$ be the induced mapping on $E/Ker(T)$. Then $\tilde{T}$ is injective and bicontinuous onto $T(E)$ by the open mapping theorem. But then $\alpha(T) = 1/\|\tilde{T}^{-1}\| > 0$.
Now assume that b) holds. We have $\gamma(T) = \|\tilde{T}^{-1}\| = 1/\alpha(T)$, hence c) holds.
Finally assume that c) holds. Then $\tilde{T}^{-1}$ is bounded, hence $\tilde{T}$ is open and therefore its range is closed. □

The following nonstandard lemma is basic for the theory of Fredholm operators and the spectral theory of compact operators:

Lemma 3.10 *Let $(H_n)_{n\in\mathbb{N}}$ be a strictly monotone sequence of standard closed subspaces of the standard Banach space E. Assume that N is infinitely large and that $y \in H_N$ is a vector of finite norm such that $d(y, H_{N-1}) \geq \frac{1}{2}$ if (H_n) is increasing, $d(y, H_{N+1}) \geq \frac{1}{2}$ if (H_n) is decreasing, resp. Then $\widehat{y} - \widehat{g} \notin E$ for all $g \in \mathrm{Fin}(H_{N-1})$, $g \in \mathrm{Fin}(H_{N+1})$, respectively.*

Proof: (I) Assume that (H_n) is increasing and suppose $y - g \approx x$ where x is standard and $g \in \mathrm{Fin}(H_{N-1})$. Then $d(y-g, H_{N-1}) \approx d(x, H_{N-1}) \geq \frac{1}{4}$ since $g \in \mathrm{Fin}(H_{N-1})$, hence $d(x, {}^*H_n) \geq \frac{1}{4}$ for all standard n. So by Transfer this

holds for all $n \in \mathbb{N}$, in particular $d(x, \bigcup H_n) \geq \frac{1}{4}$. Using again the Transfer Principle we obtain $d(x, H_N) \geq \frac{1}{4}$, a contradiction to $y - g \in H_N$.
(II) If (H_n) is decreasing then $d(y, H_{N+k}) \geq \frac{1}{2}$ for all $k \in {}^*\mathbb{N}$. Assume that $y - g \approx x$ where x is standard and $g \in \mathrm{Fin}(H_{N+k_0})$ for some k_0. Then $\exists m \forall k d(x, H_{m+k}) \geq \frac{1}{4}$ holds in ${}^*\mathcal{V}(X)$ since $g \in \mathrm{Fin}(H_{N+k_0})$. Hence this formula holds also in $\mathcal{V}(X)$. So there exists m_0 standard such that $d(x, H_n) \geq \frac{1}{4}$ for all $n \geq m_0$. The Transfer Principle yields now that in particular $d(x, y - g) \geq \frac{1}{4}$ since $y \in H_N$, a contradiction. □

Exercise: Prove an analogous assertion for arbitrary directed families of closed subspaces and use this to give a new proof of 1.4.

Let E, F be standard Banach spaces. If $T : E \to F$ is bounded then $\tilde{T} : \widehat{E}/E \to \widehat{E}/F$ given by $\tilde{T}(\widehat{x} + E) = \widehat{T}\widehat{x} + F$ is well defined.
We obtain the following useful criteria for the operator T to be of one of the types under consideration. A related theorem was proved first by Sadovskii [31]. A similar form was rediscovered by Buoni, Harte and Wickstead [4, 5].

Theorem 3.11 *Let T be an arbitrary bounded linear operator from E to F. Then the following assertions hold:*
a) T is upper Fredholm iff $\tilde{T}$ is injective.
b) T is lower Fredholm iff $\tilde{T}$ is surjective and T has closed range.
c) T is Fredholm iff $\tilde{T}$ is bijective.

Proof: a) Let $Ker(T) = E_0$ be finite dimensional and let P be a projection of E onto E_0, $Q = I - P$. Then $T_0 := TP = 0$, and $\alpha(T_1) > 0$ by 3.9 and our assumption that $T(E)$ is closed. But $\widehat{P} : \widehat{E} \to E_0$ since $\dim E_0 < \infty$ and $\widehat{T}\widehat{Q} = \widehat{T}_1$ is injective. For $\widehat{T}_1\widehat{x} = 0$ implies ${}^*T_1 x \approx 0$, hence $\| {}^*T_1 x\| \approx 0$, hence $\|x\| \approx 0$ since the standard number $\alpha(T_1) > 0$. So $\tilde{T}$ is injective.
Conversely if T is not upper Fredholm then $\dim E_0 = \infty$ or $T(E)$ is not closed. This latter case implies by 3.9 that there exists a sequence of normalized vectors x_n such that $\|x_m - x_n\| \geq \frac{1}{2}$ and $\lim Tx_n = 0$. In both cases $Ker(\widehat{T}) \not\subset E$ and the assertion follows.
b) Assume first of all that T is lower Fredholm. By 3.9 $T(E)$ is closed. But then $E = T(E) \oplus N$ where N is a finite dimensional subspace and the sum is topologically direct. So $\widehat{F} = \widehat{T(E)} \oplus N = \widehat{T}(\widehat{E}) \oplus N$, where $N \subset F$ since N is finite dimensional, i. e. $\widehat{{}^*N} = N$ (see 1.8) and the assertion follows.
Assume on the other hand that $\tilde{T}$ is surjective and that $T(E)$ is closed and moreover that codim $T(E) = \infty$. Since $T(E) =: G$ is closed, there exists by 1.6 a sequence (y_n) of normalized vectors satisfying $d(y_n, H_n) > 1/2$ where H_n is the (closed) linear span of G and $\{y_1, \ldots, y_{n-1}\}$. Take $N \in {}^*\mathbb{N}\backslash\mathbb{N}$. Since $\widehat{G}/F = \widehat{F}/F$ by hypothesis there exists $\widehat{g} \in \widehat{G}$ and $z \in F$ with $\widehat{g} = \widehat{y_N} + z$. This in turn implies $g \approx y_N + z$ which is impossible by 3.10.
c) follows from a) and b). □

3.3. NOTES

The nonstandard analysis of compact operators was already initiated by A. Robinson and A.R. Bernstein [3] who solved the invariant subspace problem for polynomially compact operators. The easy proof of Theorem 3.5 is taken from [30], cf. also [24].

As was pointed out already the first treatment of the theory of Fredholm operators by means of Fréchet products was given by Sadovskii [31]. Subsection 3.2 is to some extent within the spirit of this paper. These ideas came up again apparently independent of [31] in [4, 5]. Lemma 3.10 is due to the author. It helps to facilitate the known proofs of 3.11 as well as of Theorem 5.2.

4. Spectral theory of bounded operators

4.1. BASIC FACTS AND DEFINITIONS

1) Let E be a Banach space and let T be a bounded linear operator on E. Then its *resolvent set* is $\rho(T) = \{z \in \mathbb{C} : (z - T) \text{ is bijective } \}$ and on $\rho(T)$ the *resolvent* $R(z,T)$ is defined by $R(z,T) = (z - T)^{-1}$. Notice that $R(z,T)$ is continuous by the closed graph theorem. $\rho(T)$ is open, and $R(.,T)$ is holomorphic satisfying the famous resolvent equation
$R(z,T) - R(y,T) = (y - z)R(z,T)R(y,T)$.

2) The complement of $\rho(T)$ is called the *spectrum* $\sigma(T)$. It is compact since for $|z| > ||T||$ the *Neumann - series* $\Sigma T^n z^{-(n+1)}$ converges to $R(z,T)$. This fact also implies $\lim_{|z|\to\infty} R(z,T) = 0$. $r(T) = \sup\{|z| : z \in \sigma(T)\}$ is called the *spectral radius* of T. The spectrum is never empty.

3) $z \in \sigma(T)$ is called a *Riesz point* of T if it is a pole of $R(z,T)$ the residue of which is of finite rank.

4) z is an *approximate eigenvalue* if $\inf\{||(z-T)x|| : ||x|| = 1\} = 0$. The set of all approximate eigenvalues is denoted by $\sigma_a(T)$. It is closed in $\sigma(T)$. z is called an *eigenvalue* if the kernel $Ker(z-T) \neq \{0\}$. This space then is the set of *eigenvectors.*

5) A very recent notion was introduced by L. Trefethen [36]: For $\varepsilon > 0$ we define $\rho_\varepsilon(T) = \{z \in \rho(T) : ||R(z,T)|| < \frac{1}{\varepsilon}\}$. $\sigma_\varepsilon(T) = \mathbb{C}\backslash\rho_\varepsilon(T)$ forms the so–called *ε-pseudospectrum.* It begins to play a fundamental role in modern numerical analysis.

4.2. THE SPECTRUM OF AN S-BOUNDED INTERNAL OPERATOR

Let E denote an internal Banach space and let T be an S–bounded operator on E (see section 1.2). By Transfer we may define all the notions above also for T, and we want to consider the connection between $\sigma(T)$ and $\sigma(\widehat{T})$.

To this end we introduce also the external sets $\rho_b(T) = \{z \in \rho(T) : \|R(z,T)\| \text{ finite}\}$ and $\rho_\infty(T) = \{z \in \rho(T) : \|R(z,T)\| \text{ infinitely large}\}$.

By assumption the operator norm of T is finite. The formula for the Neumann-series (see 4.1) then yields for $|z| > \|T\|$: $\|R(z,T)\| \leq \frac{1}{|z|-\|T\|}$. Hence for every $\varepsilon > 0$ the set $\{z : |z| > \|T\| + \varepsilon\}$ is contained in $\rho_\varepsilon(T)$.

Lemma 4.1 *Let S be an internal, S-bounded operator on the internal Banach space E. The following assertions are equivalent:*
(1) $\widehat{S}$ is bijective
(2) S is bijective and $\|S^{-1}\|$ is finite.
If one of these conditions are satisfied then $\|\widehat{S}^{-1}\| = {}^\circ\|S^{-1}\|$.

Proof: (1) $\Rightarrow$ (2): Since (1) holds $\|\widehat{S}\widehat{x}\| \geq \delta > 0$ for all $\widehat{x}$ of norm 1 and $\delta = \|\widehat{S}^{-1}\|^{-1}$. Then obviously $\|Sx\| > \delta/2$ for all x of norm 1. Suppose S is not onto. By the open mapping theorem (and the Transfer Principle) $S(E)$ is (internally) closed. If $S(E) \neq E$ then by 1.6 there exists y such that $\|y\| = 1$ and $d(y, S(E)) \approx 1$. But then $\widehat{y} \notin \widehat{S}(\widehat{E}) = \widehat{S(E)}$. So S is bijective and $\|S^{-1}\| \leq 2/\delta$.
(2) $\Rightarrow$ (1): obvious. □

Theorem 4.2 *Let $\widehat{T}$ denote the nonstandard hull on $\widehat{E}$ of the S-bounded internal operator T on E (see 1.15). Then the following assertions hold:*
(1) $\sigma_a(\widehat{T}) = \{{}^\circ z : \inf\{\|(z-T)x\| : \|x\| = 1\} \approx 0\}$ and $\sigma_a(\widehat{T})$ consists only of eigenvalues.
(2) $\rho(\widehat{T}) = \{{}^\circ z : z \in \rho_b(T)\}$.
(3) Let $0 < \varepsilon' < \varepsilon$ and both numbers be standard. Then $\sigma_{\varepsilon'}(\widehat{T}) \subset \{{}^\circ z : z \in \sigma_\varepsilon(T)\} \subset \sigma_\varepsilon(\widehat{T})\}$.

Proof: (1) (I) Assume that $\inf\{\|(z-T)x\| : \|x\| = 1\} =: \alpha \approx 0$. Then by Transfer to $0 < \eta \approx 0$ there exists x of norm 1 satisfying $\|(z-T)x\| \leq \alpha + \eta$. But then $\widehat{T}\widehat{x} = {}^\circ z\,\widehat{x}$.
(II) Assume now that $\alpha \not\approx 0$. Then $\|({}^\circ z - \widehat{T})\widehat{x}\| > {}^\circ\alpha/2$ for all x with $\|\widehat{x}\| = 1$, which implies by definition that ${}^\circ z \notin \sigma_a(\widehat{T})$.
(2) follows from 4.1.
(3) (I) Let $z \in \sigma_\varepsilon(T)$ be arbitrary. If $z \in \sigma(T)$ then ${}^\circ z \in \sigma(\widehat{T})$ by 4.1, hence ${}^\circ z \in \sigma_\varepsilon(\widehat{S})$. So assume $z \in \rho(T)$. But then $\|(z-T)^{-1}\| \geq \frac{1}{\varepsilon}$ and 4.1 gives $\|({}^\circ z - \widehat{T})^{-1}\| \geq \frac{1}{\varepsilon}$, since ε is standard. So $\{{}^\circ z : z \in \sigma_\varepsilon(T)\} \subset \sigma_\varepsilon(\widehat{T})$.
(II) Assume that z is standard and $z \notin \{{}^\circ v : v \in \sigma_\varepsilon(T)\}$. Then $z \notin \sigma_\varepsilon(T)$, hence $\|(z-\widehat{T})^{-1}\| \leq \frac{1}{\varepsilon} < \frac{1}{\varepsilon'}$. Since ε' is standard 4.1 yields the assertion. □

We denote the set of eigenvalues of T by $\sigma_p(T)$.

Corollary 4.3 *a) If $z \in \sigma(T)$ and $|z| = r(T)$ then ${}^\circ z$ is an eigenvalue of $\widehat{T}$. In particular $r(\widehat{T}) \geq^\circ r(T)$.*

b) Let T be a standard bounded operator on the standard Banach space E. Then
(i) $\sigma(T) = \sigma(\widehat{{}^*T})$.
(ii) $\sigma_a(T) = \sigma_p(\widehat{{}^*T})$.
(iii) $\sigma_\varepsilon(T) = \sigma_\varepsilon(\widehat{{}^*T})$.

Proof: If $z \in \sigma(T)$ and $|z| = r(T)$ then $R(v,T)$ is unbounded near z (i.e. for $|v| > r(T)$ and $v \approx z$) since else z would not be a singularity of $R(.,T)$. So there exists $v \approx z$ with $v \in \rho_\infty(T)$, and the assertion follows.

Example 4.4 *Here is an example that $r(\widehat{T}) > {}^\circ r(T)$ may happen: Consider the Hilbert space $E = {}^*\mathbb{C}^N$ where N is infinitely large and $\|x\| = \sum_{j=1}^{N} |x_j|^2)^{1/2}$. Set*

$$Te_j = \begin{cases} e_{j+1} & j \leq N-1 \\ 0 & j = N \end{cases}$$

where $\{e_1, \ldots, e_N\}$ is the canonical base. Then $r(T) = 0$; in fact $T^N = 0$. But $r(\widehat{T}) = 1$, and $\sigma(\widehat{T}) = \{z \in \mathbb{C} : |z| \leq 1\}$. The proof of these assertions is left as an exercise. ($\widehat{T}$ induces the left shift on a suitable closed subspace of $\widehat{E}$.)

Let T be a bounded linear internal operator on the internal Banach space E. $z \in {}^*\mathbb{C}$ is called an S–Riesz point if it is a Riesz point with residue of standard finite rank. We have the following theorem:

Theorem 4.5 *Let T be S-bounded and moreover let $z \in \mathbb{C}$ be a Riesz point of $\widehat{T}$ with residue of rank r. Then there exists a (standard) δ such that the set $\sigma(S) \cap {}^*B(z,\delta)$ is not empty and consists of at most r S-Riesz points each of which has a residue of rank at most r.*

Proof: (I) Let $\delta = \inf\{|z-v| : v \in \sigma(\widehat{T})\backslash\{z\}\}/2$. Set $\eta = \delta/2$. Consider the annulus $K = \{v \in \mathbb{C} : \eta \leq |v-z| \leq \delta\}$. If $a = \sup\{\|(v-\widehat{T})^{-1}\| : v \in K\}$ and $M = 2a$ then ${}^*K \subset \rho_{1/M}(T)$. For assume that this does not hold. Then there exists $v \in {}^*K \cap \sigma_{1/M}(T)$. Hence ${}^\circ v \in \sigma_{1/M}(\widehat{T})$, a contradiction to the choice of M.
(II) By the Transfer Principle the spectral projection
$Q = \frac{1}{2\pi i} \oint_{|v-z|=\delta} R(v,T)dv$ exists and $\|Q\| < \delta M$ is finite.
Claim: $\widehat{Q} = \frac{1}{2\pi i} \oint_{|v-z|=\delta} R(v,\widehat{T})dv = Res(R(z,\widehat{T}))$.
Proof of the claim: By 4.1 $R(v,\widehat{T}) = \widehat{R(v,T)}$ for all $v \in K$. Moreover since $K \subset \rho_{1/M}(\widehat{T})$ the resolvent equation (see section 4.1) yields $\|R(v,T) -$

$R(w,T)\| \leq |v-w|M^2$, in particular $R(.,T)$ is S–uniformly continuous. Hence the Riemann sums

$$R_m = \frac{1}{2\pi i}\sum_{k=0}^{m-1} R(z+\delta\exp(2\pi ik/m),T)\exp(2\pi ik/m)\frac{\delta}{m}$$

satisfy $R_m \approx Q$ for all infinitely large m. On the other hand $\widehat{R}_m \to \frac{1}{2\pi i} \oint_{|v-z|=\delta} R(v,\widehat{T})dv$. This proves the claim.

Now $\dim \widehat{Q}(\widehat{E}) = r < \infty$. But 1.8 then implies $\dim Q(E) = r$, and the assertion follows. □

4.3. NOTES

The spectral theory of internal S–bounded operators as presented here is due to the author (cf also [28, 41]). Corollary 4.3 is partly new. 4.3 b) (i) and (ii) however are very well–known and trace back (within the frame work of Fréchet–products) to Quigley (see [29]). These facts were rediscovered by Berberian [2], Lotz (cf. [32] V.1), and others and have been used extensively since then. Theorem 4.5 is new. The corresponding result within the frame work of ultraproducts may be found in [28].

5. Applications of nonstandard spectral theory

5.1. THE SPECTRUM OF COMPACT OPERATORS

The following lemma plays a key role in the spectral theory of compact operators.

Lemma 5.1 *Let T be a compact operator on the standard Banach space E. Then the following assertions hold:*
(a) If $0 \neq z \in \mathbb{C}$ then $(z-T)$ is a Fredholm operator
(b) If $z \neq 0$ is not an eigenvalue then $z \in \rho(T)$.

Proof: (a) With the notions and notations of 3.11 $(z-\tilde{T}) = z\tilde{I}$ by 3.2. So 3.11 yields the assertion.
(b) Assume that $0 \neq z$ is not an eigenvalue of T. Then $U = z-T$ is injective. Since U is Fredholm by (a) its range $U(E) = H_1$ is closed. If $H_1 \neq E$ then $H_n = U^n(E)$ build a strictly decreasing sequence of closed spaces. Let N be infinitely large. Then by 1.6 there exists a normalized vector y in H_N satisfying $d(y, H_{N+1}) \geq \frac{1}{2}$. Now $Ty/z = (T-z)y/z + y = y + w$ where $w \in H_{N+1}$. So $y + w \in H_N$ is nearstandard since T is compact, a contradiction to 3.10.

Theorem 5.2 *Let T be a compact operator on the Banach space E. Then $\sigma(T)\backslash\{0\}$ consists of Riesz points only.*

Proof: In view of 5.1 we only have to show that all points in $\sigma(T)\backslash\{0\}$ are isolated. So let $z \neq 0$ be in $\sigma(T)$ and assume that $z = \lim_{n\to\infty} z_n$, where all $z_n \in \sigma(T)$ are pairwise different. Then each z_n is an eigenvalue by 5.1. Let x_n be a normalized eigenvector corresponding to z_n and let H_n be the span of $\{x_1, \ldots, x_n\}$. Choose an infinitely large N, and $y \in H_N$ of norm 1 satisfying $d(y, H_{N-1}) \geq \frac{1}{2}$.
Now $y_N = \sum_{k=1}^{N} \beta_k x_k$, hence $u = y_N - \frac{1}{z_N} T y_N = \sum_{k=1}^{N} \beta_k (1 - \frac{z_k}{z_N}) x_k = \sum_{k=1}^{N-1} \beta_k (1 - \frac{z_k}{z_N}) x_k \in H_{N-1}$. So $\frac{1}{z_N} T y_N = y_N - u$. Since T is compact $\frac{1}{z_N} T y_N$ is nearstandard, a contradiction to 3.10. □

5.2. APPROXIMATION OF SPECTRA

Another field on which nonstandard functional analysis works well is approximation theory. We give an example concerning the approximation of spectra (for another application see [28]). To this end define the distance from the set A to the set B in $\mathbb{C}$ by $\operatorname{dist}(A, B) = \sup_{a \in A}(\inf\{|b-a| : b \in B\})$

Lemma 5.3 *Let (S_n) be a sequence of bounded linear operators on the Banach space E, converging uniformly to the operator T. Then* $\lim_{n\to\infty} \operatorname{dist}(\sigma(S_n), \sigma(T)) = 0$

Proof: If not then there exists $\delta > 0$ standard and N infinitely large, and moreover a $z \in \sigma(S_N)$ such that $\inf\{|b - z| : b \in {}^*\sigma(T)\} \geq \delta$. But then $B(z, \delta) = \{v \in {}^*\mathbb{C} : |v - z| < \delta\} \subset {}^*\rho(T)$, and since ${}^\circ z \in \rho(T)$, $(z - {}^*T)^{-1}$ is S–bounded. Since the invertible operators form an open set in $\mathcal{L}(E)$ $(z - S_N)^{-1} \approx (z - {}^*T)^{-1}$ is S–bounded, a contradiction to $z \in \sigma(S_N)$. □

It is known that $\limsup \operatorname{dist}(\sigma(T), \sigma(S_n)) \neq 0$ may happen. But $\lim \operatorname{dist}(\sigma_a(T), \sigma_\varepsilon(S_n)) = 0$ holds in much more general situations (recall: $\sigma_a(T)$ is the approximate point spectrum of T, $\sigma_\varepsilon(S_n)$ is the ε–pseudospectrum of (S_n), see section 5.1).

First of all let us recall the notion of discrete convergence from approximation theory (cf. [37]): Let E and F_n be Banach spaces and for each n let $P_n : E \to F_n$ be a bounded linear operator. The sequence (F_n, P_n) approximates E if $\lim \|P_n u\|_n = \|u\|$ holds for all $u \in E$. By the Uniform Boundedness Principle the sequence (P_n) is uniformly bounded.

A sequence $(u_n)_n$ with $u_n \in F_n$ converges discretely to $u \in E$ if $\lim \|u_n - P_n u\|_n = 0$. A sequence $(S_n)_n$ of operators S_n on F_n converges discretely to T on E if $\lim_n \|P_n T u - S_n P_n u\|_n = 0$ for all $u \in E$. If this happens then (S_n) is bounded. Obviously strong convergence is a special case of this notion.

Exercises: (1) Let $E = l^2(\mathbb{N})$ be the usual Hilbert space and let T be the shift given by $(Tf)(k) = f(k+1)$. Then $\sigma(T) = \sigma_a(T) = \{z \in \mathbb{C} : |z| \leq 1\}$.

Set $F_n = E,\ P_n = I$ and

$$(S_n f)(k) = \begin{cases} f(k+1) & k \leq n-1 \\ 0 & \text{else} \end{cases}$$

Then (S_n) converges strongly to T and $\sigma(S_n) = \{0\}$ but for all $\varepsilon > 0$ we have
$\lim \operatorname{dist}(\sigma(T), \sigma_\varepsilon(S_n) = 0$

(2) Let $S = T^\star$. Then $\sigma_a(S) = \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Set $F_n = \mathbb{C}^n$, take $S_n(x_1, \ldots, x_n) = (x_n, x_1, \ldots, x_{n-1})$ and $P_n f = (f(1), \ldots, f(n))$. Obviously (S_n) converges discretely to S but $\limsup_n \operatorname{dist}(\sigma(S), \sigma_\varepsilon(S_n)) = 1 - \varepsilon$. However $\lim_n \operatorname{dist}(\sigma_a(S), \sigma_\varepsilon(S_n)) = 0$.

Theorem 5.4 ([41]) *Let (F_n, P_n) approximate the Banach space E. Let (S_n) be a sequence of bounded operators on F_n which converges discretely to the operator T on E. Then for every $\varepsilon > 0$ $\lim\limits_n \operatorname{dist}(\sigma_a(T), \sigma_\varepsilon(S_n)) = 0$.*

Corollary 5.5 *If all F_n are Hilbert spaces and if moreover all S_n are unitary operators then $\sigma_a(T) \subset \{z \in \mathbb{C} : |z| = 1\}$.*

Corollary 5.6 *If E is a Hilbert space and if in addition T is normal then $\lim(\sigma(T), \sigma_\varepsilon(S_n)) = 0$.*

The corollaries are left as an exercise.

Proof of 5.4: (I) Take N infinitely large. Then $P_N : {}^*E \to F_N$ satisfies $\|P_N x\| \approx x$ for all x standard in E. Hence $\widehat{P}_N|_E$ is an isometry into $\widehat{F}_N$. Since $(S_n) \to T$ discretely $P_N\ {}^*Tu \approx S_N P_N u$ for all $u \in E$, hence $\widehat{P}_N Tu = \widehat{S}_N \widehat{P}_N u$ for all $u \in E$.
Now let $z \in \sigma_a(T)$ be arbitrary. Then there exists a sequence (u_n) of normalized vectors in E satisfying $\lim \|zu_n - Tu_n\| = 0$. Since $\widehat{P}_N|_E$ is an isometry $(\widehat{P}_N u_n)$ is a sequence of normalized vectors in $\widehat{F}_n$, and $\|z\widehat{P}_N u_n - \widehat{S}_N \widehat{P}_N u_n\| = \|\widehat{P}_N(z - T)u_n\| \to 0$ for $n \to \infty$. So z is in $\sigma_a(\widehat{S}_N) \subset \sigma(\widehat{S}_N) \subset {}^\circ\sigma_\varepsilon(S_N)$ by 4.2.
(II) Assume now that the theorem does not hold. Then there exists a standard $\varepsilon > 0$ and $\delta > 0$ such that $\limsup\ \operatorname{dist}(\sigma_a(T), \sigma_\varepsilon(S_n)) > \delta$. This in turn implies, that there exists $N \approx \infty$ and $z \in {}^*\sigma_a(T)$ such that $\inf\{|z - v| : v \in \sigma_\varepsilon(S_N)\} > \delta/2$. But then $w := {}^\circ z$ satisfies $|w - v| \geq \delta/2$ for all $v \in \sigma_\varepsilon(S_N)$. But $w \in \sigma(\widehat{S}_N)$ by (I), so we obtain a contradiction to (I).

Exercise: Prove the following generalization of a theorem of V. Lyantse on compact operators on Hilbert spaces (see [25], cf. also [24]):

Proposition 5.7 *Let T be a standard compact operator on the standard Banach space E. Assume that S is an internal operator, for which there*

exists an infinitesimal operator R and another internal S-bounded operator P satisfying
(1) $S = P^*TP + R$ and (2) $Px \approx x$ for all nearstandard elements x.
Then every $0 \not\approx z \in \sigma(S)$ is an S-Riesz point.

5.3. NOTES

Subsection 5.1 is well–known. The proofs seem to be new though they are based on the standard ones (cf. [42], X.5, for a nonstandard treatment see also [30]). Subsection 5.2 is taken from [41]. Similar ideas are already used in [28]. Obviously nonstandard analysis is quite useful in approximation theory see also [39] as well as Korovkin's theory of approximation by positive operators as developed by e.g. Scheffold and also by the author (see [32], V.2).

6. Closed Operators

6.1. INTRODUCTION

Let $(A, \mathcal{D}(A))$ be a closed densely defined operator on the Banach space E. Its resolvent set $\rho(A)$ is defined as $\{z \in \mathbb{C} : (z - A) \text{ is bijective onto } E\}$ If $z \in \rho(A)$ then $(z - A)^{-1} =: R(z, A)$ is continuous by the closed graph theorem. $\rho(A)$ is open (but it might be empty). As in the case of a bounded operator the resolvent $R(., A) : \ \rho(A) \ni z \to R(z, A)$ is holomorphic and satisfies the resolvent equation (see section 4.1).

The complement of $\rho(A)$ is the spectrum $\sigma(A)$. The approximate point spectrum $\sigma_a(A)$ is defined as previously.

In general it is quite difficult to define the nonstandard hull of A in $\widehat{E}$ for if $G(A)$ denotes the graph of A then $\widehat{G(A)}$ is no longer the graph of a mapping in $\widehat{E}$. But if we assume that $\rho(A) \neq \phi$ and that in addition there exists a sequence $(z_n) \subset \rho(A)$ with $\lim |z_n| = \infty$ and $(z_n R(z_n, A))_n$ is bounded then the situation is quite better as we shall see later on.

6.2. STRONGLY CONTINUOUS SEMIGROUPS

A typical example of the preceding notions is the *generator* $(A, \mathcal{D}(A))$ of a bounded *strongly continuous semigroup* $\mathcal{T} = (T_t)_{t \geq 0}$ of operators T_t on E. Let us recall this notion a little more detailed: $\mathcal{T}$ is called a strongly continuous semigroup if $T_{s+t} = T_s T_t$ for all $s, t \geq 0$ and if moreover for all $x \in E$ $\lim_{t \to 0} T_t x = x$. Then its generator $(A, \mathcal{D}(A))$ is defined by $x \in \mathcal{D}(A)$ iff $Ax := \lim_{t \to 0} \frac{1}{t}(T_t x - x)$ exists.

The semigroup property implies that $t \to T_t x$ is norm continuous for every $x \in E$. Moreover $x \in \mathcal{D}(A) \Rightarrow T_t x \in \mathcal{D}(A)$ and $(T_t x)' = AT_t x = T_t Ax$.

Since $\mathcal{T}$ is bounded by $R(z, A)(x) := \int\limits_0^\infty e^{-tz} T_t x dt$ there is defined a bounded linear operator for $z \in \mathbb{C}$ with $Re(z) > 0$ which turns out to be the resolvent of A. This proves A to be closed and since $\lim\limits_{u \to \infty} uR(u, A)x = x$ (as is easily seen) $\mathcal{D}(A)$ is dense. Moreover the map $t \to T_t R(z, A)$ is continuous with respect to the operator norm.

Here is the first result on strongly continuous semigroups:

Theorem 6.1 *The following assertions are equivalent:*
(i) A is everywhere defined
(ii) $t \to T_t$ is continuous with respect to the operator norm
(iii) $t \to \widehat{T}_t$ is continuous with respect to the operator norm

Proof: (i) $\Rightarrow$ (ii): The Closed Graph Theorem implies that A is bounded. Hence $S_t = e^{tA}$ satisfies $(S_t x) = AS_t x$. Thus $H(t) = S_{-t} T_t$ is constant $= H(0) = I$.
(ii) $\Rightarrow$ (iii) and (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) are left as an exercise.
(*Hint:* For $0 < t \approx 0$ we have $t^{-1} \int_0^t T_s \, ds \approx I$ with respect to the operator norm. This implies that $t^{-1} \int_0^t T_s \, ds$, hence $\int_0^t T_s \, ds$ are invertible. Conclude from this fact for $0 < h \approx 0$ and $h < t$

$$h^{-1}(T_h - I) = (h^{-1} \int_t^{t+h} T_s ds - h^{-1} \int_0^h T_s ds)(\int_0^t T_s ds)^{-1}.)$$

6.3. THE NONSTANDARD HULL OF A CLOSED OPERATOR

We come now to the problem how to define $(\widehat{A}, \mathcal{D}(\widehat{A}))$ in the case when A is not bounded. For the following section cf. [20], where this theory is developed within the framework of ultraproducts. To solve the question we recapitulate the notion of a pseudo–resolvent introduced by E. Hille (see [42]):

Let $D \subset \mathbb{C}$ be not empty and $R : D \to \mathcal{L}(E)$ be a function satisfying the resolvent equation $R(u) - R(v) = (v - u)R(u)R(v)$. Then all operators $R(u)$ have a common null space denoted by $N(R)$ and a common range, denoted by $R(E)$. Moreover $R(u)R(v) = R(v)R(u)$ holds for all $u, v, \in D$.

Theorem 6.2 (standard, see [42], p. 216) *(1) A pseudo–resolvent is the resolvent of a closed densely defined linear operator A iff $N(R) = \{0\}$. Then $R(E)$ is the domain of definition of A and $A = uI - R(u)^{-1}$.*
(2) Moreover if there is a sequence $(z_n) \subset D$ with $\lim |z_n| = \infty$ such that $(z_n R(z_n))$ is bounded then $\overline{R(E)} = \{x \in E : \lim z_n R(z_n)x = x\}$ and $N(R) \cap \overline{R(E)} = \{0\}$.

In the following let us assume that $(A, \mathcal{D}(A))$ is closed, densely defined and that $\rho(A) \neq \phi$. Moreover let us assume that there is a sequence $(z_n) \subset \rho(A)$ such that $\lim |z_n| = \infty$ and $(z_n R(z_n, A))$ is bounded. Notice that this happens for generators of strongly continuous semigroups for in that case $u\, R(u, A)$ converges strongly to the identity for $u \to \infty$. Then by the Transfer Principle $\rho(A) \ni z \to \widehat{R}(z, A) =: \widehat{R}(z)$ defines a pseudo–resolvent on $\widehat{E}$ for which $(z_n \widehat{R}(z_n))$ is bounded.

So we define $\widehat{E}_R = \overline{\widehat{R}(\widehat{E})}$. The space $\widehat{E}_R$ is invariant under $\widehat{R}(u)$ for all u and $\widehat{R}(u)|_{\widehat{E}_R}$ is injective with dense range $\mathcal{D}(\widehat{A}) := \widehat{R}(u)(\widehat{E})$. We then set $\widehat{A} = uI - (\widehat{R}(u)|_{\widehat{E}_R})^{-1}$, and call this the *nonstandard hull of the closed operator* $(A, \mathcal{D}(A))$.

Remark 6.3 *If A has compact resolvent then by 3.2, $\widehat{E}_R = E$ and $\widehat{A} = A$.*

Theorem 6.4 *We adhere to the preceding hypotheses. Then the following assertions hold:*
(1) $\sigma(A) = \sigma(\widehat{A})$
(2) $\sigma_a(A) = \{z \in \sigma(\widehat{A}) : z$ is an eigenvalue of $\widehat{A}\}$.

Proof: (I) Fix $u \in \rho(A)$. Then $z \in \rho(A)\backslash\{u\} \to \frac{1}{u-z}$ is a bijection of $\rho(A)\backslash\{u\}$ onto $\rho(R(u, A))\backslash\{0\}$ For $\frac{1}{u-z} - R(u, A) = \frac{1}{u-z}(z - A)R(u, A)$ is invertible iff $z \in \rho(A)$ and $z \neq u$. Then its inverse is $(u - z)(u - A)R(z, A)$. The inverse mapping is $v \in \rho(R(u, A))\backslash\{0\} \to z = \frac{vu-a}{v} \in \rho(A)\backslash\{0\}$.
(II) Now if $z \in \sigma(A)$ then choose $u \notin \sigma(A)$. Then $\frac{1}{u-z} \in \sigma(R(u, A) \Leftrightarrow \frac{1}{u-z} \in \sigma(\widehat{R}(u)) \Leftrightarrow z \in \sigma(\widehat{A})$ and the assertion follows e.g. from 4.2.
(2) is proven similarly and is left as an exercise. □

6.4. APPLICATION TO ONE-PARAMETER SEMIGROUPS

If $(A, \mathcal{D}(A))$ is the generator of a bounded strongly continuous semigroup, $\widehat{A}$ can be characterized in another manner.

Theorem 6.5 (cf. [38]) *Let $(A, \mathcal{D}(A))$ be the generator of a strongly continuous semigroup $\mathcal{T} = (T_t)_{t \geq 0}$. Then the following assertions hold:*
(1) $\widehat{E}_R = \{\widehat{x} : t \to {}^*T_t x$ is $\underline{S}$ – continuous $\}$
(2) $\mathcal{D}(\widehat{A}) = \{\widehat{x} : \exists x \in \widehat{x}[x \in {}^*\mathcal{D}(A) \wedge {}^*Ax$ is finite $]\}$

Proof: (1) Since $t \to T_t R(u, A)$ is continuous with respect to the operator norm, so is $t \to \widehat{T}_t \widehat{R}(u)$. So $t \to \widehat{T}_t \widehat{y}$ is continuous for every $\widehat{y} \in \widehat{R}(\widehat{E})$, hence for $\widehat{y} \in \widehat{E}_R$. So if $\widehat{x} \in \widehat{E}_R$ then $t \to {}^*T_t x$ is S–continuous. If conversely $t \to {}^*T_t$ is S–continuous then for every $u > 0$ nearstandard ${}^*R(u, A)x = \int_0^\infty e^{-ut}\ {}^*T_t x dt$ satisfies $\widehat{R}(u)\widehat{x} \in \widehat{R}(\widehat{E})$, hence $\widehat{x} = \lim u\widehat{R}(u)\widehat{x} \in \widehat{E}_R$.

(2) By definition $\mathcal{D}(\widehat{A}) = \widehat{R}(u)(\widehat{E}_R)$ for some $u > 0$. By (1) $\widehat{x} \in \mathcal{D}(\widehat{A}) \Rightarrow x = \int_0^\infty e^{-ut} \; {}^*T_t y dt = {}^*R(u, A)y$ for some y such that $t \to {}^*T_t y$ is S-continuous. But then ${}^*Ax = {}^*Ax - ux + ux = ux - y$ is finite.
Conversely assume that $x \in {}^*\mathcal{D}(A)$ and ${}^*Ax$ is finite. Since $x \in {}^*\mathcal{D}(A)$ there exists $y \in {}^*E$ such that $x = {}^*R(u, A)y$ (where u is standard > 0). But then $y = ux - {}^*Ax$ is finite hence $\widehat{x} = \widehat{R}(u)\widehat{y} \in \mathcal{D}(\widehat{A})$. □

6.5. NOTES

Section 6 goes back to Krupa (within the frame work of ultraproducts). Subsection 6.4 is due to the author (cf. [38], and also [41, 28]). This section may also serve as a base for the spectral theory of closed operators. The trick to use pseudoresolvents instead of the operator itself is due to Greiner [10] – at least in the context of generators of strongly continuous semigroups. An extensive nonstandard analytical treatment of concrete closed operators, e.g. of differential operators is to be found in [1].

References

1. Albeverio, S., Fenstad, J.E., Høegh-Krohn, R. and Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, Orlando etc.
2. Berberian, S. K., (1962) Approximate proper vectors, *Proc. AMS* **13**, pp. 111–114.
3. Bernstein, A.R., Robinson, A., (1966) Solution of an invariant subspace problem of K.T. Smith and P.R. Halmos, *Pacific J. Math.* **16**, pp. 421–431.
4. Buoni, S., Harte, R. Wickstead, A.W., (1977) Upper and lower Fredholm spectra I, *Proc. AMS* **66**, pp. 309–314.
5. Chadwick, J.J.M., and Wickstead, A.W. (1977), A quotient of ultrapowers of Banach spaces and semi–Fredholm operators, *Bull. London Math. Soc.* **9**, pp. 321–325.
6. Davis, M. (1977) *Applied Nonstandard Analysis*, Wiley, New York.
7. Dunford, N., Schwartz, J. (1958) *Linear Operators Part I*, Interscience Publishers, New York.
8. Enflo, P., Lindenstrauss, J. Pisier, G., (1975) On the "three space problem", *Math. Scand.* **36**, pp. 199–210.
9. Golodets, V. Ya., (1978) Modular operators and asymptotic commutativity in von Neumann algebras, *Russian Mathem. Surveys* **33**, pp. 47–106.
10. Greiner, G., (1981) Zur Perron–Frobenius Theorie stark stetiger Halbgruppen, *Math. Z* **177**, pp. 401–423.
11. Groh, U., (1984) Uniformly ergodic theorems for identity preserving Schwarz maps on $W^\star$-algebras, *J. Operator Theory* **11**, pp. 395–402.
12. Heinrich, S., (1980) Ultraproducts in Banach space theory, *J. Reine Angew. Math.* **313**, pp. 72–104.
13. Heinrich, S., Henson, C. W., Moore, L. C., (1987) A note on elementary equivalence of $C(K)$ spaces, *J. Symbolic Logic* **52**, pp. 368–373.
14. Henson, C.W., Moore, L. (1983) Nonstandard Analysis and the theory of Banach spaces, in: Hurd, A.E. (ed), *Nonstandard Analysis – Recent Developments*, Springer, Berlin Heidelberg New York, pp. 27–112.
15. Henson, C.W. (1997), Foundations of nonstandard analysis: a gentle introduction to nonstandard extensions, *this volume.*

16. Hinokuma, T., Ozawa, M. (1993), Conversion from nonstandard matrix algebras to standard factors of type II_1, *Illinois Math. J.* **37**, pp. 1–13.
17. Hurd, A.E., Loeb, P.A. (1985) *An Introduction to Nonstandard Real Analysis*, Academic Press, Orlando etc.
18. James, R.C. (1963/64), Characterizations of reflexivity, *Studia Math.* **23**, pp. 205–216
19. Janssen, G. (1972), Restricted ultraproducts of finite von Neumann algebras, in: A. Robinson (ed.) *Contributions to non-standard analysis*, Studies in Logic and Found. Math., **69**, North Holland Amsterdam, pp. 101–114.
20. Krupa, A. (1990), On various generalizations of the notion of an $\mathcal{F}$-power to the case of unbounded operators, *Bull. Pol. Acad. Sci. Math.* **38**, pp. 159–166.
21. Lacey, H.E. (1974) *The Isometric Theory of Classical Banach spaces* Springer New–York Heidelberg.
22. Loeb, P.A. (1997), Nonstandard analysis and topology, *this volume.*
23. Luxemburg, W.A.J. (1969), A general theory of monads, in Luxemburg, W.A.J. (ed.) *Applications of Model Theory to Algebra, Analysis, and Probability*, Holt, Rinehart, and Winston, New York, pp. 18–86.
24. Luxemburg, W.A.J. (1995), Near–standard compact internal linear operators, in Cutland, N.J. (ed.) et al. *Developments in Nonstandard Mathematics,* Pitman Res. Notes Math. Ser. 336, London pp. 91–98
25. Lyantse, V. E., (1989) On a perturbation that is infinitely small in the strong operator topology, *Ukrainian Math. J.* **41**, pp. 845–847.
26. Mittelmeyer, G., Wolff, M., (1974) Über den Absolutbetrag auf komplexen Vektorverbänden, *Math.Z.* **137**, pp. 87–92.
27. Moore, L.C., (1976) Hyperfinite extensions of bounded operators on a separable Hilbert space *Trans. Amer. Math. Soc.* **218**, pp. 285–295
28. Räbiger, F., Wolff, M.P.H. (1995) On the approximation of positive operators and the behaviour of the spectra of the approximants, to appear in *Integral Equations and Op. Theory.*
29. Rickart, C. E. (1960) *General Theory of Banach Algebras*, Van Nostrand Princeton, N.Y.
30. Robert, A. ,(1995) Functional analysis and NSA in Cutland, N.J. (ed.) et al. *Developments in Nonstandard Mathematics,* Pitman Res. Notes Math. Ser. 336, London, pp. 73–90.
31. Sadovskii, B.N.(1972), Limit-compact and condensing operators, *Uspehi Math. Nauk* **27**, pp. 81–146.
32. Schaefer, H.H. (1974) *Banach Lattices and Positive Operators,* Springer, Berlin Heidelberg New York.
33. Sims, B. (1982) *"Ultra"-techniques in Banach Space Theory,* Queen papers in pure and applied mathematics 60, Queen's Univ., Kingston, Ontario, Canada.
34. Stroyan, K. D., Luxemburg, W. A. J. (1976), *Introduction to the Theory of Infinitesimals*, Academic Press, New York.
35. Sun, Yeneng, (1990) A Banach space in which a ball is contained in the range of some countable additive measure is superreflexive, *Canad. Math. Bull.* **33**, pp. 45–49.
36. Trefethen, L.N., (1993) Pseudospectra of matrices, in: Griffiths, D.F. (ed), *Numerical Analysis,* Proc. of the 14th Dundee Conference 1991, Pitman, London, pp. 234–264
37. Vainikko, G. (1976) *Funktionalanalysis der Diskretisierungsmethoden,* Teubner, Leipzig
38. Wolff, M.P.H., (1984) Spectral theory of group representations and their nonstandard hull, *Israel J. Math.* **48**, pp. 205–224.
39. Wolff, M., (1992) An application of spectral calculus to the problem of saturation in approximation theory, *Note di Matematica* **12**, pp. 291–300.
40. Wolff, M.P.H., (1993) A nonstandard analysis approach to the theory of quan-

tum meanfield systems, in Albeverio, S., Luxemburg, W. A. J., Wolff, M. P. H. (eds): *Advances in Analysis, Probability and Mathematical Physics – Contributions of Nonstandard Analysis.* Kluwer, Dordrecht, pp. 228–246.

41. Wolff, M.P.H.; (1996) On the approximation of operators and the convergence of the spectra of the approximants, to appear in R. Mennicken, Ch. Tretter (eds): *IWOTA 95 Proceedings*, Birkhäuser Basel.
42. Yosida, K., (1968) *Functional Analysis*, Springer, Berlin Heidelberg New York, 2nd. ed.

APPLICATIONS OF NONSTANDARD ANALYSIS IN ORDINARY DIFFERENTIAL EQUATIONS

E. BENOIT
Université de La Rochelle
Avenue Marillac
17042 La Rochelle
France
email: ebenoit@math.univ-lr.fr

1. Introduction

The purpose of my lectures is to show you that the formalism of nonstandard analysis is a good way for an intuitive understanding *and* for proving some results in ordinary differential equations.

First, I will give you some definitions and properties of functions from $({}^*\mathbb{R})^p$ to $({}^*\mathbb{R})^q$. They have already been seen in Cutland's article in this volume [7]. After that, I want to show you the interest of the comparison between a solution of an ODE and a recursive sequence.

In a second part, I will study the one parameter family of ODE with a parameter ε which tends to zero (in classical language) or is a fixed infinitesimal real number (with NSA language). These problems are called *perturbations* of an ODE.

2. Tools in NSA

2.1. NELSON'S LANGUAGE

In France, people who are interested in NSA and specially in ODE are speaking with Nelson's language. It will be difficult for me to translate the Nelson-french into star-english, but I will try.

There are many differences between Nelson's and Robinson's presentation of NSA. I want to emphasize two of them: the first is a difference of point of view. In Nelson's presentation, the natural set of physical numbers is not $\mathbb{R}$, but ${}^*\mathbb{R}$. The set of reals you have used for a long time is ${}^*\mathbb{R}$, but until now, you couldn't distinguish standard and nonstandard elements. So you worked with infinitesimals and with unlimited numbers without any

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 153–182.

knowledge of their properties. Nelson's formalism allows us to distinguish two kinds of objects, the standard and the nonstandard, and it gives rules to use this distinction in mathematics. The consequence of this is that you stay always in the nonstandard extension, and that you see the standard objects as particular objects included in the nonstandard extension. So I should always put a star before all my standard objects, but I will forget this star in the notation.

The second, more important, difference is in the foundations: Nelson's language has a new predicate with one place, the predicate *standard*. One can make quantifications with this predicate, and we will note

$\forall^{\mathrm{st}} x \in X\ \mathcal{F}(x)$ will be an abbreviation of $\forall x \in X$ (x is standard $\Rightarrow\ \mathcal{F}(x)$)

$\exists^{\mathrm{st}} x \in X\ \mathcal{F}(x)$ will be an abbreviation of $\exists x \in X$ (x is standard and $\mathcal{F}(x)$)

The basis of Nelson's theory is three axioms to use the predicate *standard*. The three axioms are the Transfer principle, Idealization (which implies the existence of nonstandard elements and some properties of saturation), and Standardization (which implies the existence of the standard part of some objects). I will always try to avoid this difference between the two languages; my purpose is not to make lectures on foundations of NSA with Nelson's approach, and in the applications, the choice of foundations of NSA is not too important.

I will also avoid the words *finite* and *infinite* for the real numbers; I think that they are confusing, and I prefer the words *limited* and *unlimited*.

2.2. INTERNAL SETS

A standard subset A of ${}^*\mathbb{R}^N$ is the image of a subset of $\mathbb{R}^N$ by the $*$-transform. I should note ${}^*A$ but I will forget the $*$. Take care: an infinite standard set contains nonstandard elements.

When I do not specify, a subset will be internal. Furthermore, the internal sets which are not standard in these notes are generally constructed from a standard one and a non standard parameter with the following procedure:

Let $\mathcal{A}$ be a standard subset of the standard product $X \times Y$. Let y be a nonstandard element of Y, and consider for instance

$$A = \{x \in X \quad \text{such that} \quad (x, y) \in \mathcal{A}\}$$

The same occurs for functions: an (internal) function will often be a partial function of a standard function, where one argument is a fixed nonstandard element: if $\mathcal{F} : \mathcal{D} \to Z$ is standard, with $\mathcal{D}$ a standard subset of $X \times Y$, and $\pi_1 : X \times Y \to Y$ the canonical projection, we define $F : D = \pi_1 (\mathcal{D} \cap (X \times \{y\})) \to Z$ by $F(x) = \mathcal{F}(x, y)$.

The notion of standard part is very important throughout my lectures. It was introduced by Cutland (see [7] Theorem 8.4). I will, however, say *shadow* instead of *standard part* but these words are equivalent.

Definition 2.1 *Let A be a subset of a standard set X. The* shadow $\mathrm{st}A$ *of A is the standard subset of X defined by*

$$\forall^{\mathrm{st}} x \quad (x \in \mathrm{st}A \Leftrightarrow \exists y \in A \; y \approx x)$$

You see that, in the definition above, only the standard elements of $\mathrm{st}A$ are characterized. You should know that it is enough to characterize a standard set.

Example 2.2 Let ε be non-zero infinitesimal.

(a) If γ is the graph of the function $x \to \varepsilon x$, then $\mathrm{st}\gamma$ is the set $\mathbb{R} \times \{0\}$.
(b) If γ is the graph of the function $x \to \arctan \frac{x}{\varepsilon}$, then $\mathrm{st}\gamma$ is the union of the sets $]-\infty, 0] \times \{-\frac{\pi}{2}\}$, $\{0\} \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[0, +\infty[\times\{\frac{\pi}{2}\}$.
(c) If γ is the graph of the function $x \to \sin \frac{x}{\varepsilon}$, then $\mathrm{st}\gamma$ is the strip $\mathbb{R} \times [-1, 1]$.

In the definition of the shadow, the set A could be external, but, in that case, the Theorem 8.4 of Cutland [7] is not valid. For example the shadow of the set $\{x$ such that $°x \in]0, 1[\}$ is the open interval $]0, 1[$.

2.3. EXTERNAL SETS

Sometimes, I will consider external subsets. Most of them will be *halos* and *galaxies*:

Definition 2.3 *An external set H (resp. G) is a* halo *(resp. a* galaxy*) if one can find an (internal) sequence $(A_n)_{n\in\mathbb{N}}$ of (internal) sets such that*

$$H = \bigcap_{n\in\mathbb{N}, n \text{ standard}} A_n \quad \left(resp. \;\; G = \bigcup_{n\in\mathbb{N}, n \text{ standard}} A_n \right)$$

Example 2.4 – The monad of 0 in $\mathbb{R}$ is a halo.

- The external set of all limited real numbers is a galaxy.
- If f and g are two functions, the set $\{x \mid f(x) \approx g(x)\}$ is a halo or is internal.
- The external set of all standard real numbers is neither a halo, nor a galaxy.

A generalization of the Robinson's lemma and of the overflow principle is the following:

Theorem 2.5 (Permanence principle) *A halo is never a galaxy.*

Proof Suppose that a halo H and a galaxy G coincide:

$$H = \bigcap_{n\in\mathbb{N}, n \text{ standard}} A_n = \bigcup_{n\in\mathbb{N}, n \text{ standard}} B_n = G$$

Then we have

$$\forall^{\text{st}} p \in \mathbb{N} \quad \bigcap_{n<p} A_n \supset \bigcup_{n<p} B_n$$

The overflow principle shows that the same formula is also true for some unlimited p. Thus we have the inclusions

$$H \supset I = \bigcap_{n<p} A_n \supset \bigcup_{n<p} B_n \supset G$$

If H and G coincide, they coincide also with the internal set I, and they are internal. This is impossible by definition of a halo. □

2.4. S-CONTINUITY

You have already seen such a lemma in Cutland's lecture (Theorem 8.8), but I need a slightly more general lemma. Before that, I will give some more precise definitions and some exercises to make precise the notion of S-continuity.

Definition 2.6 *A function $f : E \to F$ is called* S-continuous, *if*

$$\forall^{\text{st}} x_0 \in E \quad \forall x \in E \quad x \approx x_0 \Rightarrow f(x) \approx f(x_0)$$

Definition 2.7 *A function $f : E \to F$ is called* uniformly S-continuous, *if*

$$\forall x \in E \quad \forall y \in E \quad x \approx y \Rightarrow f(x) \approx f(y)$$

In almost all books, uniform S-continuity is called S-continuity, and I will also use this wrong terminology. To understand the difference between the two notions, you should prove the following propositions

Example 2.8 Let ε be nonzero infinitesimal.

(a) $f(x) = \varepsilon \text{ int}\left(\frac{x}{\varepsilon}\right)$ is piecewise constant, non continuous, and S-continuous on $\mathbb{R}$.
(b) $f(x) = x^2$ is S-continuous, but not uniformly S-continuous on $\mathbb{R}$.
(c) $f(x) = \frac{\varepsilon}{x+\varepsilon}$ is S-continuous on $]0,1]$, but not uniformly. The same function is uniformly S-continuous on $[\varepsilon^{1/2}, 1]$.

Remark The definition of (non uniform)-S-continuity looks strange when the set E is not a "good" set : for example, if E doesn't contain any standard points, then any function on E is S-continuous. Uniform-S-continuity avoid this problem.

Proposition 2.9 *A uniformly S-continuous function is S-continuous.*

Proposition 2.10 *If each point of E has a standard part in E, an S-continuous function on E is also uniformly S-continuous.*

Proposition 2.11 *Let t_0 be a point of a limited interval $[a,b]$, such that $^\circ t_0 \in [a,b]$. If f is S-continuous on $[a,b]$, and if $f(t_0)$ is limited, then $f(t)$ is limited for all t in $[a,b] \cap [\,^\circ a,\ ^\circ b]$.*

Proof Left to the reader. □

Exercise Let f be an S-continuous function. Let g be a translation. Is $f \circ g$ S-continuous? The same question with uniform S-continuity.

2.5. THE CONTINUOUS SHADOW LEMMA

Theorem 2.12 (Continuous shadow, [9]) *Let D be a (internal) subset of $\mathbb{R}^p$ (p is a standard integer). Let $F : D \to \mathbb{R}^q$ a (internal) function. We suppose that F is (uniformly)-S-continuous and limited for all limited points of D. Then, there exists a standard uniformly continuous function $f : \mathrm{st}(D) \to \mathbb{R}^q$ such that*

$$\forall X \in D \quad X \text{ limited} \;\Rightarrow\; {}^\circ F(X) \;=\; f({}^\circ X)$$

Proof Let Γ be the graph of F :

$$\Gamma \;=\; \{(X,Y) \in \mathbb{R}^p \times \mathbb{R}^q \;\mid\; X \in D \text{ and } Y = F(X)\}$$

Let γ be the shadow of Γ. First, we want to check that γ is the graph of a function on $\mathrm{st}(D)$. For that purpose, we have to prove that

1. $\forall x \in \mathrm{st}(D)\ \exists y\ (x,y) \in \gamma$
2. $\forall x\ \forall y_1\ \forall y_2\ \ (x,y_1) \in \gamma$ and $(x,y_2) \in \gamma \;\Rightarrow\; x \in \mathrm{st}(D)$ and $y_1 = y_2$

By the axiom of transfer it suffices to prove 1) and 2) for standard elements x, y_1, and y_2, because all the formulas and constants are standard. Now we have

1.
$$\forall^{\mathrm{st}} x \in \mathrm{st}(D)\ \exists X \in D\ \ X \approx x$$

Then X is limited, hence $F(X)$ too, and

$$^\circ(X, F(X)) \;=\; (x,\ {}^\circ F(X)) \;\in\; \gamma$$

2.
$$\forall^{\text{st}} x\, \forall^{\text{st}} y_1\, \forall^{\text{st}} y_2 \;\; (x, y_1) \in \gamma \text{ and } (x, y_2) \in \gamma \;\Rightarrow$$
$$\Rightarrow\; \exists (X, Y_1) \in \Gamma\; \exists (X, Y_2) \in \Gamma \;\; X_1 \approx x \;\; X_2 \approx x \;\; Y_1 \approx y_1 \;\; Y_2 \approx y_2$$

Then we have $X_1 \approx X_2$ and $Y_1 = F(X_1) \approx F(X_2) = Y_2$. Consequently, the two standard real numbers y_1 and y_2 are equivalent, and thus coincide.

Now, write f for the function which has graph γ. It is standard, defined on $\text{st}(D)$ and it satisfies the required properties:

1. If X is in D, limited, it has a standard part x in $\text{st}(D)$ and the standard point $(x, f(x))$ is in the shadow of Γ. Thus there exists (X_1, Y_1) in Γ such that $x \approx X_1$ and $f(x) \approx Y_1$. By the S-continuity of F, the equivalence $X \approx X_1$ implies $F(X) \approx F(X_1) = Y_1 \approx f(x)$. We deduce the equality of the standard parts.
2. We know nothing about the nonstandard points of γ, therefore, to check the uniform continuity of f, we need to use the classical definition, which, by transfer, is

$$\forall^{\text{st}} \varepsilon > 0\; \exists^{\text{st}} \alpha > 0\; \forall^{\text{st}} x_1\, \forall^{\text{st}} x_2 \;\; \|x_1 - x_2\| < \alpha \;\Rightarrow\; \|f(x_1) - f(x_2)\| < \varepsilon$$

But we have $f(x_1) - f(x_2) \approx F(x_1) - F(x_2)$ hence the uniform continuity is a consequence of

$$\forall^{\text{st}} \varepsilon > 0\; \exists^{\text{st}} \alpha > 0\; \forall^{\text{st}} x_1\, \forall^{\text{st}} x_2 \;\; \|x_1 - x_2\| < \alpha \;\Rightarrow\; \|F(x_1) - F(x_2)\| < \frac{\varepsilon}{2}$$

And this is true because we know that for all given standard ε, the set

$$\{\alpha > 0 \mid \forall x_1\, \forall x_2 \;\; \|x_1 - x_2\| < \alpha \;\Rightarrow\; \|F(x_1) - F(x_2)\| < \frac{\varepsilon}{2}\}$$

contains all the infinitesimal numbers and also (by infinitesimal overflow, see N. CUTLAND [7] Corollary 8.3) a standard positive real number.

□

Proposition 2.13 *Let F be a S continuous function on D. If* $\text{st}(D)$ *is path-connected and if $F(X_0)$ is limited for some limited X_0, then $F(X)$ is limited for all limited X in D.*

Proof Let ε be a standard positive real number. Let A be the set

$$A \;=\; \{\alpha > 0 \mid \forall X, Y \in D \;\; \|X - Y\| < \alpha \;\Rightarrow\; \|F(X) - F(Y)\| < \varepsilon\}$$

By S-continuity, the set A contains all infinitesimal real numbers. Then by the permanence principle, it contains a non infinitesimal real number δ.

Let X be a limited point of D. There exists a path connecting ${}^\circ X_0$ to ${}^\circ X$ in $\mathrm{st}(D)$. This path can be chosen standard (use transfer). We can find a standard number of standard points x_0, x_1, $...x_n$ on this path such that

$$x_0 = {}^\circ X_0 \qquad x_n = {}^\circ X \qquad ||x_{i+1} - x_i|| < \frac{\delta}{2}$$

Then there exists X_i in D with $X_i \approx x_i$, and we have $||F(X_{i+1}) - F(X_i)|| < \varepsilon$, then $||F(X) - F(X_0)|| < n\varepsilon$ which is standard. □

Exercise Find a path-connected set D such that $\mathrm{st}(D)$ is not path connected. Find a S-continuous function on this set, limited for some limited X_0 but not for all limited X.

3. Differential Equations and Recursive Sequences

3.1. INTRODUCTION

The use of nonstandard analysis (NSA) to study ordinary differential equations (ODE) has some advantages: the first one is to replace a solution of an ODE by a recursively defined sequence.

I have to go back to Leibniz: at this time a solution of an ODE was given by a relation such as

$$\delta x \; = \; f(x)\,\delta t$$

where δx and δt were *small* increments of the variables.

I will return to this idea, with *infinitely small* increments. NSA gives tools to make the approximation rigorous.

But we have to avoid the following mistake: from a qualitative point of view, a solution of an ODE is easier to study than a recursive sequence. When the ODE has an explicit solution, given by usual functions, it is much easier to find it with the classical methods of solution than to use sequences. On the other hand, the recursive sequences will give a good way to prove theorems about ODEs.

3.2. EXAMPLE

The following example is too easy, but we can see the different points of view. The ODE is

$$\begin{cases} \frac{dx}{dt} & = \; x \\ x(0) & = \; 1 \end{cases} \tag{1}$$

The corresponding easiest recursive equation is

$$\begin{cases} \dfrac{\xi_{i+1} - \xi_i}{\varepsilon} & = \xi_i \\ \xi(0) & = 1 \end{cases} \tag{2}$$

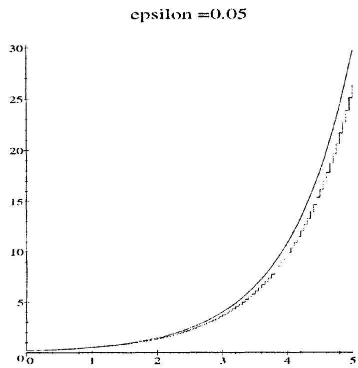


Figure 1. The recursive sequence and the solution of the ODE.

The solution of the recursive equation here is obvious:

$$\xi_i = (1 + \varepsilon)^i$$

It is now convenient to introduce the piecewise constant function (in Cutland's article [7] the corresponding function is piecewise linear ; it doesn't matter).

$$\xi(t) = \xi_i \quad \text{for} \quad \varepsilon i \leq t < \varepsilon(i + 1)$$

and using the rules of calculus in NSA, we can compute for all limited t:

$$\xi(t) = \xi_i = (1 + \varepsilon)^i = \left((1 + \varepsilon)^{\frac{1}{\varepsilon}}\right)^{\varepsilon i} \approx e^t$$

We found here the *exact* solution of the ODE. We did not solve the differential equation, we only had to know the formula $(1 + \varepsilon)^{\frac{1}{\varepsilon}} \approx e$.

3.3. THE MAIN THEOREM (EASY CASE)

The theorem we will use many times in this lecture is the stroboscopy theorem. It allows us to find a good approximation of some recursive sequences by continuous functions. It is a bridge between discrete dynamical systems and continuous ones.

I will give you a complete version of this theorem, but, for a better understanding, it is preferable to see first an easy case, where all the technical difficulties do not appear.

The theorems are given only for the positive part of the solutions, the same holds for the negative part.

Theorem 3.1 (Stroboscopy, [5]) *Let $f : \mathbb{R}^N \to \mathbb{R}^N$ be a standard continuous bounded function.*

Let $(t_i)_{i\in\{0,1,\ldots,\omega\}}$ be a real sequence such that

$$t_0 < t_1 < \ldots < t_\omega \quad , \quad t_i \approx t_{i+1} \quad , \quad t_0 \text{ limited} \quad , \quad t_\omega - t_0 \not\approx 0$$

Let $(\xi_i)_{i\in\{0,1,\ldots,\omega\}}$ be a sequence in $\mathbb{R}^N$ such that ξ_0 is limited and

$$\forall i \qquad \frac{\xi_{i+1} - \xi_i}{t_{i+1} - t_i} \approx f(\xi_i) \tag{3}$$

Then the function ξ defined on $[t_0, t_\omega[$ by

$$\xi(t) = \xi_i \qquad \textit{for} \qquad t_i \leq t < t_{i+1}$$

is S-continuous. It has a shadow x, which is a solution of the Cauchy problem

$$\begin{cases} \frac{dx}{dt} & = & f(x) \\ & & \\ x({}^\circ t_0) & = & {}^\circ\xi_0 \end{cases} \tag{4}$$

Proof First, we prove that the function ξ is S-continuous. Let s and t be two elements of $[t_0, t_\omega]$ with $s \approx t$. We can suppose that $s > t$, and we have to prove that $\xi(s) \approx \xi(t)$. By definition, there exist some integers m and n such that

$$\xi(s) = \xi_m \qquad \xi(t) = \xi_n \qquad t_m \leq s < t_{m+1} \qquad t_n \leq t < t_{n+1}$$

Consequently, we have

$$\xi(s) - \xi(t) \; = \; \xi_m - \xi_n \; = \; \sum_{i=n}^{m-1} (\xi_{i+1} - \xi_i) \; \approx \; \sum_{i=n}^{m-1} (t_{i+1} - t_i) f(\xi_i)$$

The function f is bounded and standard, hence bounded by a limited M and

$$||\xi(s) - \xi(t)|| \leq \sum_{i=n}^{m-1} (t_{i+1} - t_i) ||f(\xi_i)|| \leq (t_m - t_n) M \approx 0$$

Furthermore, $\xi(t_0)$ is limited, and the domain is an interval, thus the *continuous shadow theorem* (with Proposition 2.13) gives the function x.

To check that the derivative of x is $f(x)$ we approximate the integral of $f(x)$ between two standard real numbers s and t.

$$\int_t^s f(x(\tau))d\tau \approx \sum(t_{i+1}-t_i)f(x(t_i)) \approx$$

$$\approx \sum(t_{i+1}-t_i)f(\xi_i) \approx \xi_m - \xi_n \approx x(s)-x(t)$$

Since two equivalent standard real numbers are equal, we have

$$\int_t^s f(x(\tau))d\tau = x(s)-x(t)$$

3.4. THE MAIN THEOREM (GENERAL CASE)

In the general case, we will take a more general function f, defined only on a standard open set D, and not necessarily bounded. The first part of the conclusion gives a local description of the phenomenon. The second part shows how the recursive sequence and the solution of the ODE can leave the domain D.

Theorem 3.2 (Stroboscopy, [5]) *Let D be a standard open set in $\mathbb{R}^N$. Let $f: D \to \mathbb{R}^N$ be a standard continuous function.*

Let $(t_i)_{i\in\{0,1,\ldots,\omega\}}$ be a real sequence such that

$$t_0 < t_1 < \ldots < t_\omega \quad , \quad t_i \approx t_{i+1} \quad , \quad t_0 \text{ limited} \quad , \quad t_\omega - t_0 \not\approx 0$$

Let $(\xi_i)_{i\in\{0,1,\ldots,\omega\}}$ be a sequence in $\mathbb{R}^N$ such that ξ_0 is near-standard in D and

$$\forall i \quad {}^o\xi_i \in D \Rightarrow \frac{\xi_{i+1}-\xi_i}{t_{i+1}-t_i} \approx f(\xi_i) \tag{5}$$

Let $\xi: [t_0, t_\omega[\to \mathbb{R}^N$ be the piecewise constant function defined by

$$\xi(t) = \xi_i \quad \text{for} \quad t_i \leq t < t_{i+1}$$

Then there exist T standard (T can be $+\infty$), $T > {}^o t_0$, and $x: [{}^o t_0, T[\to \mathbb{R}^N$ a standard continuous function such that

1. *x is a solution of the Cauchy problem $\frac{dx}{dt} = f(x)$, $x({}^o t_0) = {}^o\xi_0$ and*

$$^o t < T \Rightarrow {}^o\xi(t) = x({}^o t) \in D$$

2. *Either $T \geq {}^o t_\omega$ or*

$$\exists t_e \approx T \quad (\forall s \leq t_e \;\; \xi(s) \approx x(s)) \text{ and } \xi(t_e) \text{ is not near standard in } D.$$

Proof of the first part Let x_0 be the standard part of ξ_0. It is in D. Then its monad is also in D (which is an open standard set) and f is equivalent to $f(x_0)$ in this monad. Hence, by overflow, there exists a standard neighborhood V of x_0 and a standard real number M such that

$$x \in V \quad \Rightarrow \quad ||f(x)|| < M$$

Let t_n be the time defined by

$$n \;=\; \min\{\, i \;\mid\; \xi_i \notin V \,\}$$

For all i less than n, we have $||\xi_{i+1} - \xi_0|| < 2M(t_{i+1} - t_0)$. Consequently, $t_n - t_0$ is not infinitesimal, and we can apply the easy case of the stroboscopy theorem on the sequence $\xi_0, ..., \xi_n$. The existence of a standard function x defined on $[\,{}^{o}t_0,\; {}^{o}t_n[$ which has the required properties follows. □

Proof of the second part Let Γ be the graph of ξ. Let γ be it's shadow in $\mathbb{R} \times \mathbb{R}^N$. Let T be the standard real number (may be $+\infty$) defined by

$$T \;=\; \sup\{\, t \;\mid\; \gamma \bigcap ([0,t] \times \mathbb{R}^N) \;\text{ is the graph of a function } x : [0,t] \to D\}$$

For all standard $t < T$, the function x is uniquely defined by $x(t) = {}^{o}\xi(t)$. It is a continuous standard function on $[\,{}^{o}t_0, T[$. For this T, it is obvious that the first part of the conclusion of the theorem holds.

Let t_i be an element of $[0, T[$, and suppose that ${}^{o}t_i < {}^{o}t_\omega$. If ξ_i is near-standard in D, we can apply the first part of the theorem to the sequence beginning at time t_i. It follows that ξ is S-continuous and near-standard in D on a standard neighborhood of ${}^{o}t_i$. Then ${}^{o}t_i < T$.

Then, suppose that $T < {}^{o}t_\omega$. The external set

$$\{\, t < t_\omega \;\mid\; \xi(t) \approx x(t)\}$$

contains all the times t such that ${}^{o}t_0 < {}^{o}t < T$. Using the permanence principle, it contains also an interval $[t_b, t_e]$, with $t_e \approx T$. And we proved above that $\xi(t_e)$ cannot be near-standard in D. Then all the conclusions of the theorem hold. □

Example 3.3 To obtain a better understanding of part two of the theorem, you can study the sequence

$$\xi_0 = 1 \qquad \xi_{i+1} = \xi_i + \varepsilon\xi_i^2$$

It is defined for all positive integer i. We have the equality

$$\frac{\xi_{i+1} - \xi_i}{t_{i+1} - t_i} \;=\; \xi^2 \;=:\; f(\xi_i) \qquad \text{with} \qquad t_i = \varepsilon i$$

The function f is defined on $\mathbb{R}$. The solution of the Cauchy problem is $x(t) = \frac{1}{1-t}$. It is defined on $[0, 1[$. Hence we have $T = 1$. The theorem claims that the function x and the sequence ξ become unlimited together: they are equivalent until a time t_e where they are both non limited.

3.5. EXISTENCE OF A SOLUTION

The theorem above is very general. It has many applications. The first one is an elementary proof of the existence of solutions of an ODE.

Corollary 3.4 *Let $\dot{x} = f(x)$ be an ODE where f is a continuous function from D to $\mathbb{R}^N$. Let x_0 be an initial condition. Then, there exists a function $x(t)$ defined on a maximal interval I such that $x(t_0) = x_0$, and $\dot{x}(t) = f(x(t))$.*

Proof By the transfer principle, we can suppose that all the given objects are standard. Let ε be an infinitesimal real number (positive to define x on $[t_0, +\infty[$, negative to define x on $]-\infty, t_0]$). We denote

$$\begin{aligned} t_i &= t_0 + \varepsilon i \\ \xi_0 &= x_0 \\ \xi_{i+1} &= \xi_i + \varepsilon f(\xi_i) \end{aligned}$$

and we apply the main theorem to the sequence ξ_i. This yields a maximal solution of the ODE. □

Remark In the proof, we defined ξ_i by Euler's method of numerical analysis. We could define ξ_i with any formula coming from a numerical algorithm for solution of an ODE step by step. Because we are not interested in the order of convergence, Euler's method is the simplest one, and thus the best. (See the paragraph "Ghosts" to expand this remark).

3.6. UNICITY OF A SOLUTION

I will not prove here the theorem on the unicity of the solution of the Cauchy problem. I will only give an exercise which explains what we can expect from the study of recursive sequences.

Exercise Let us study the well known ODE

$$\dot{x} = 3x^{2/3}$$

which does not have the property of unicity of solutions. For any given limited real number (standard or not) ξ_0, we define the recursive sequence

$$\xi_{i+1} = \xi_i + 3\varepsilon\xi_i^{2/3}$$

and the standard function x such that

$$x(\varepsilon i) \approx \xi_i$$

This function is a solution of the differential equation, with $x(0) = {}^{\circ}\xi_0$.

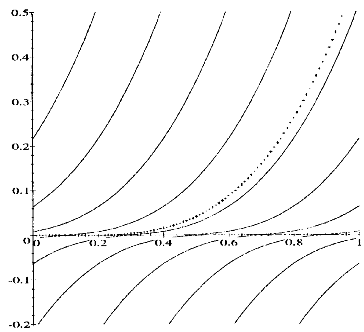


Figure 2. Solutions of $\dot{x}=3x^{2/3}$ and the recursive sequence for $\xi_0=10^{-301}$ and $\varepsilon=10^{-2}$.

Let k be a given real standard positive number. Prove that there exists ξ_0 such that

$$x(t) = \begin{cases} 0 & \text{for } t \in [0,k] \\ (t-k)^3 & \text{for } t \in [k,+\infty[\end{cases}$$

3.7. GHOSTS

Applying stroboscopy theorem to numerical methods, we can generally prove the convergence of the method. In some cases, one can explain more complicated behaviours. I will here study an example with a two-steps method:

$$\xi_{n+2} = \xi_n + 2\varepsilon f(\xi_{n+1}) \quad \text{with} \quad f(x) = x(1-x)$$

You see in figure 3 one trajectory of this recursive sequence. At the beginning, it goes along the trajectory of the differential equation with initial condition 0.2. It reaches the halo of the stable equilibrium point $x = 1$. Then, the odd-indexed and the even-indexed sequences differ, and each one goes to the unstable equilibrium point $x = 0$.

We will now explain this behaviour. Consider in $\mathbb{R}^2$ the point

$$(\xi_n, \eta_n) = (\xi_n, \xi_{n+1})$$

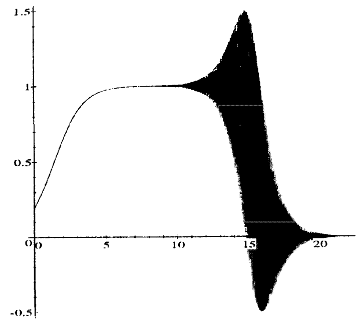


Figure 3. The "ghost" trajectory of $\dot{x} = x(1-x)$.

We have the following recursive definition of the sequence (ξ_n, η_n):

$$\left\{ \begin{array}{lcl} \xi_{n+1} & = & \eta_n \\ \eta_{n+1} & = & \xi_n + 2\varepsilon f(\eta_n) \end{array} \right.$$

It is not possible to apply the stroboscopy theorem to this sequence, but the second iteration gives

$$\left\{ \begin{array}{lcl} \xi_{n+2} & = & \xi_n + 2\varepsilon f(\eta_n) \\ \eta_{n+2} & = & \eta_n + 2\varepsilon \left[f\left(\xi_n + 2\varepsilon f(\eta_n)\right)\right] \end{array} \right.$$

Now, we can write

$$\frac{1}{\varepsilon}\left((\xi_{n+2}, \eta_{n+2}) \ - \ (\xi_n, \eta_n)\right) \ \approx \ 2F(\xi_n, \eta_n)$$

$$\text{with} \quad F(\xi, \eta) \ = \ (f(\eta), f(\xi))$$

Choose now an initial condition (ξ_0, η_0) such that ξ_0 and η_0 are equivalent. We have to study the differential equation

$$\left\{ \begin{array}{lcl} x' & = & f(y) \\ y' & = & f(x) \end{array} \right. \tag{6}$$

We draw the phase portrait for this equation in $\mathbb{R}^2$ in figure 4. (For that, you have to note that the function $H(x, y) = \int f(y)dy - \int f(x)dx$ is constant on the trajectories).

For limited t, we check that the solution $x(t)$ and $y(t)$ with initial condition $({}^{o}\xi_0, {}^{o}\xi_0)$ remains equivalent to the solution of the initial equation. Thus, the stroboscopy theorem shows that the recursive sequence satisfies

$$\forall n \quad \varepsilon n \text{ limited} \ \Rightarrow \ \xi_n \ \approx \ \eta_n \ \approx \ x(\varepsilon n)$$

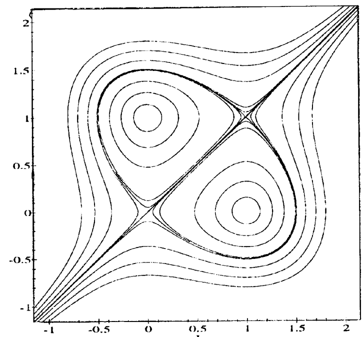


Figure 4. The phaseportrait of $\dot{x} = y(1-y)$, $\dot{y} = x(1-x)$.

By overspill, there exists n_1 such that εn_1 is unlimited and such that the property above holds for all positive n smaller than n_1. Therefore,

$$\forall n < n_1 \quad \varepsilon n \text{ unlimited} \quad \Rightarrow \quad \xi_n \approx \eta_n \approx 1$$

Assume that the recursive sequence eventually leaves the halo of $(1,1)$. Then for a standard small positive number a, the sequence also leaves the disc of radius a. Let $n_2(a)$ be the smallest index bigger than n_1 such that $||(\xi_n - 1, \eta_n - 1)|| \geq a$. The stroboscopy theorem shows that the shadow of $\{(\xi_{n_2+n}, \eta_{n_2+n}), \varepsilon n \text{ limited}\}$ is the trajectory of the differential equation (6) with initial condition $({}^o\xi_{n_2}, {}^o\eta_{n_2})$. For negative limited t, this trajectory must be in the disc of radius a, therefore, it is the unstable manifold of the saddle point $(-1,1)$ of (6).

4. Regular perturbations

The classical problem is the following: let (E_ε) a family of ODEs, where ε is a (real) parameter. What is the behaviour of the solutions of E_ε when ε tends to zero? Is it the behaviour of the solutions of E_0?

Using nonstandard analysis, we can first suppose that the family (E_ε) is standard and then we take a fixed infinitesimal ε. We now just have to study the solutions of the equation (E_ε), or, better the shadows of the solutions. To emphasize this point of view, we will delete ε in the notations: we have two differential equations E and E_0. The second one is standard, and both are equivalents (we will make precise what is meant by "equivalent" in different contexts).

In this section, we will study regular perturbations, and the main tool is the "short shadow lemma" below.

SHORT SHADOW LEMMA

There are two versions of this lemma, the first one is easier, but the second one is more useful.

Theorem 4.1 (Short shadow lemma (weak), [9]) *Let f and F be two functions. Let us suppose that f is standard and continuous. Let H be the external subset of $\mathbb{R}^N$ defined by*

$$x \in H \quad \Leftrightarrow \quad f(x) \approx F(x)$$

Let $\Phi : [0, T[\to \mathbb{R}^N$ be a solution of the equation $\dot{x} = F(x)$, with T limited. Let us suppose that

$$\forall t \in [0, T[\quad (\quad \Phi(t) \in H, \quad \Phi(t) \text{ is limited}, \quad {}^{\circ}\Phi(t) \in H \ )$$

Then, Φ is S-continuous and its shadow is a solution of the equation $\dot{x} = f(x)$, defined on the interval $[0, {}^{\circ}T[$.

Proof Let us choose an infinitesimal positive real number ε. For convenience, we take ε such that T is a multiple of ε.

Let us denote $\xi_i = \Phi(\varepsilon i)$ and $t_i = \varepsilon i$ for all integer i in $[0, T/\varepsilon]$.

For all t in $[t_i, t_{i+1}[$, we have $\Phi(t) \in H$ and ${}^{\circ}\Phi(t) \in H$, hence

$$\dot{\Phi}(t) \; = \; F(\Phi(t)) \; \approx \; f(\Phi(t)) \; \approx \; f({}^{\circ}\Phi(t))$$

Then $\dot{\Phi}(t)$ is quasi constant on this interval.

Now, we have

$$\frac{\xi_{i+1} - \xi_i}{t_{i+1} - t_i} \; = \; \frac{\Phi(t_{i+1}) - \Phi(t_i)}{t_{i+1} - t_i} \; \approx \; f({}^{\circ}\Phi(t_i)) \; \approx \; f(\xi_i)$$

Then, applying the stroboscopy lemma, we define the required function φ on $[0, {}^{\circ}T[$. $\square$

Theorem 4.2 (Short shadow lemma (strong), [9]) *Let f and F be two functions. Let us suppose that f is standard and continuous. Let H be the external subset of $\mathbb{R}^N$ defined by*

$$x \in H \quad \Leftrightarrow \quad f(x) \approx F(x)$$

Let us suppose that the two equations $\dot{x} = f(x)$ and $\dot{x} = F(x)$ have the properties of existence and unicity of solutions. Let $\varphi : [0, T] \to \mathbb{R}^N$ be a standard solution of the equation $\dot{x} = f(x)$. Let us suppose that there exists a (internal) set K such that

$$\text{halo}(\varphi([0, T])) \; \subset \; K \; \subset \; H$$

Then the solution Φ of $\dot{X} = F(X)$ with initial condition $\Phi(0) = X_0 \approx \varphi(0)$ is defined on $[0, T]$, and φ is its shadow.

Proof First, we will prove that K could be chosen compact. For that purpose, we define

$$K_\varepsilon = \{x \in \mathbb{R}^N \mid \exists y \in \varphi([0,T]) \quad \|y - x\| \leq \varepsilon\}$$

For all ε infinitesimal, this set K_ε is included in halo$(\varphi([0,T]))$ then in K. By overspill there exists a non infinitesimal real number ε_0 such that K_{ε_0} is included in K. This set K_{ε_0} is compact, included in K and it contains halo$(\varphi([0,T]))$.

Let us denote

$$t_1 = \sup\{t \leq T \mid \Phi([0,t]) \subset K_{\varepsilon_0}\}$$

By compactness, $\Phi([0,t_1]) \subset K_{\varepsilon_0}$. Applying the weak short shadow lemma, the function Φ has a shadow which is the restriction of φ on $[0, {}^{o}t_1]$. We have now

$$\Phi(t_1) \approx \varphi({}^{o}t_1) \quad \text{then} \quad \Phi(t_1) \in \text{halo}(\varphi([0,T])) \subset \text{interior}(K_{\varepsilon_0})$$

But, if t_1 is not equal to T, then we have $\Phi(t_1) \in \partial K_{\varepsilon_0}$. Therefore $t_1 = T$ and the theorem is proved.□

5. Example

In many ODEs, the dynamic mixes fast oscillations and slow trends. In this situation, the main stroboscopy theorem is needed. I will give an example, the equation

$$\dot{x} = \sin\left(\frac{tx}{\varepsilon}\right) \qquad \varepsilon > 0 \quad \varepsilon \approx 0\,.$$

First, you have to look at the two pictures, the first one with $\varepsilon = 0.2$, the second one with $\varepsilon = 0.01$. The second one is a caricature of the first one. I will explain the pictures.

The symmetries allow to study the solutions only for $x > 0$ and $t > 0$.

First region: $t \leq x$ *In this region, the shadows of the trajectories are hyperbolas.*

Proof Along the hyperbolas $tx = (2k\pi - \frac{\pi}{2})\varepsilon$, we have $\dot{x} = -1$. Moreover, the tangent vector of the hyperbolas is $(1, -\frac{x}{t})$. The slope is bigger than -1, so the trajectories of the equation can cross this hyperbolas only from high to low. Along the hyperbolas $tx = (2k\pi + \frac{\pi}{2})\varepsilon$, we have $\dot{x} = +1$, and the same arguments prove that the trajectories can cross the hyperbolas only from low to high.

Then, a given trajectory is caught in a trap, between two such hyperbolas, which have the same shadows. □

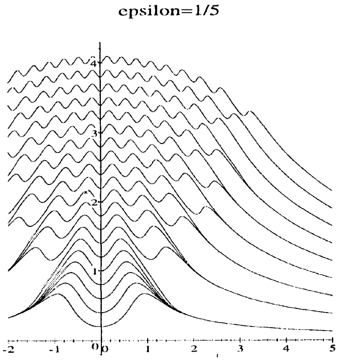


Figure 5. Phase portrait of equation for $\varepsilon = 0.2$.

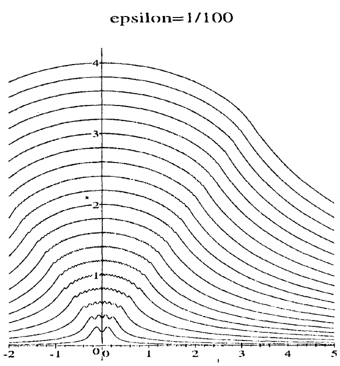


Figure 6. Phase portrait of equation for $\varepsilon = 0.01$.

Second region: $t < x$, $t \not\simeq x$ *In this region, the shadows of the trajectories are trajectories of the equation*

$$\dot{x} = \frac{-t}{\sqrt{x^2 - t^2} + x}$$

.

Proof In this region, we have

$$\frac{d}{dt}(tx) = x + t\dot{x} = x + t\sin\left(\frac{tx}{\varepsilon}\right) \geq x - t > 0$$

Then, the function $\frac{tx}{\varepsilon}$ is increasing and it is infinitely often a multiple of 2π. We will restrict our attention to these points.

Let us denote by (t_n, x_n) the points of a trajectory, such that $t_n x_n = 2n\pi\varepsilon$. We will apply the stroboscopy theorem, and, for that purpose, we will compute an approximation of $\frac{x_{n+1}-x_n}{t_{n+1}-t_n}$.

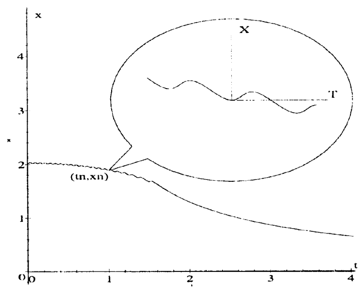


Figure 7. Magnifying glass around the point (t_n, x_n).

We now use a magnifying glass to see better what happens in the halo of the point (t_n, x_n) (see figure 7).

$$x = x_n + \varepsilon X \qquad\qquad t = t_n + \varepsilon T$$

$$\left\{\begin{array}{rcl} \dot{x} & = & \sin\left(\frac{tx}{\varepsilon}\right) \\ \dot{t} & = & 1 \end{array}\right. \qquad \text{yields} \qquad \left\{\begin{array}{rcl} \varepsilon\dot{X} & = & \sin\left(t_n X + x_n T + \varepsilon X T\right) \\ \varepsilon\dot{T} & = & 1 \end{array}\right.$$

After a change of time, we obtain the equation

$$\left\{\begin{array}{rcl} X' & = & \sin\left(t_n X + x_n T + \varepsilon X T\right) \\ T' & = & 1 \end{array}\right.$$

with the initial condition $T = 0$, $X = 0$.

Using the short shadow lemma, we know that the (short) shadow of the trajectory is the trajectory of the standard equation

$$\left\{\begin{array}{rcl} X' & = & \sin\left({}^\circ t_n X + {}^\circ x_n T\right) \\ T' & = & 1 \end{array}\right.$$

with the initial condition $T = 0$, $X = 0$.

We want to check the point (t_{n+1}, x_{n+1}), where ${}^\circ t_n X + {}^\circ x_n T = 2\pi$. A very boring computation, using the change of variables

$$u = \tan\frac{{}^\circ t_n X + {}^\circ x_n T - \pi}{2}$$

gives the coordinates of this point:

$$(X,T) \;=\; \left(\frac{-2\pi\,{}^{o}t_n}{{}^{o}x_n^2 - {}^{o}t_n^2 + {}^{o}x_n\sqrt{{}^{o}x_n^2 - {}^{o}t_n^2}}\,,\; \frac{2\pi}{\sqrt{{}^{o}x_n^2 - {}^{o}t_n^2}}\right)$$

Then, returning to the initial equation, we have proved that $t_{n+1} - t_n$ is infinitesimal, and

$$\frac{x_{n+1} - x_n}{t_{n+1} - t_n} \;\approx\; \frac{-t_n}{\sqrt{x_n^2 - t_n^2} + x_n}$$

We apply now the stroboscopy lemma with the function

$$f(x,t) \;=\; \left(\frac{-t}{\sqrt{x^2 - t^2} + x}\,,\; 1\right)$$

□

Third region: $x \approx t$ In this region, one can easily prove that, most of time, $\dot{x} < 0$. Then, a trajectory cannot go along the diagonal, and the two first regions collapse without difficulties.

6. Dynamical Systems: Notions of Stability

In the study of ordinary differential equations, or dynamical systems, the notion of stability is very important. I will give here the nonstandard definitions, but, I prefer to give them with the language of flows and vector fields.

Definition 6.1 *If $\dot{x} = f(x)$ is a vector field on $\mathbb{R}^N$ (the function f is always supposed to be C^1), the* flow *is the function $\Phi : \mathbb{R} \times \mathbb{R}^N \mapsto \mathbb{R}^N$ such that $\Phi(t, x_0)$ is the position, at time t of the solution of the differential equation, with initial condition x_0.*

The two following properties are obvious:

$$\frac{\partial}{\partial t}\Phi(t,x_0) \;=\; f\left(\Phi(t,x_0)\right) \qquad \Phi(t_1 + t_2, x_0) \;=\; \Phi\left(t_1, \Phi(t_2, x_0)\right)$$

Examples:

$\dot{x} = x$ gives the flow $\Phi(t,x_0) = x_0 e^t$

$\left\{\begin{array}{l} \dot{x} = -y \\ \dot{y} = x \end{array}\right.$ gives the flow $\Phi(t,x_0,y_0) = \begin{pmatrix} x_0\cos t - y_0 \sin t \\ x_0 \sin t + y_0 \cos t \end{pmatrix}$

$\dot{x} = x^2$ gives the flow $\Phi(t,x_0) = \frac{x_0}{1 - tx_0}$ if $tx_0 < 1$

Definition 6.2 *A vector field is* complete *if the flow $\Phi(t,x_0)$ is defined for all (t,x_0) in $\mathbb{R} \times \mathbb{R}^N$.*

Proposition 6.3 *If f is bounded, the vector field $\dot{x} = f(x)$ is complete.*

Proof Exercise (of standard mathematics). □

Remark The vector fields $\dot{x} = f(x)$ and $\dot{x} = \frac{f(x)}{1+\|f(x)\|}$ have the same trajectories. Moreover, the second one is always complete. Hence, in order to study the trajectories (we forget the time law) of a vector field, one can suppose that it is complete.

Definition 6.4 *A point x_0 is an* equilibrium point *of $\dot{x} = f(x)$ if for all t, we have $\Phi(t, x_0) = x_0$.*

This property is obviously equivalent to $f(x_0) = 0$.

Definition 6.5 *Let $\dot{x} = f(x)$ be a standard vector field. Let x_0 be an equilibrium point of this vector field.*

(a) The point x_0 is Lyapunov-stable *if*

$$\forall x \approx x_0 \ , \ \forall t > 0 \ , \ \Phi(t, x) \approx x_0$$

(b) The point x_0 is attractive *if there exists a standard open set U containing x_0 such that*

$$\forall^{\mathrm{st}} x \in U \ , \ \forall t > 0 \textit{ unlimited} \ , \ \Phi(t, x) \approx x_0$$

(c) The point x_0 is asymptotically-Lyapunov-stable *if there exists a standard open set U containing x_0 such that*

$$\forall x \in U \ , \ \forall t > 0 \textit{ unlimited} \ , \ \Phi(t, x) \approx x_0$$

(d) The attraction basin *of x_0 is the shadow of the external set*

$$\{x \ \ \forall t > 0 \textit{ unlimited} \ , \ \Phi(t, x) \approx x_0\}$$

Exercises

(a) Using formal transformations, write classical definitions of these notions.

(b) Prove that asymptotic-Lyapunov-stability is equivalent to Lyapunov-stability and attractivity.

(c) On each picture of the figures 8 and 9, is the origin Lyapunov-stable? attractive?

Definition 6.6 *Let $\dot{x} = f(x, y)$ be a standard family of vector fields, with one parameter y. Let us suppose that there exists a standard function $\xi : D_y \mapsto \mathbb{R}^N$ such that $\xi(y)$ is an equilibrium point of $\dot{x} = f(x, y)$. Then, $\xi(y)$ is* uniformly asymptotically Lyapunov *stable on D_y if there exists an open set U, standard, containing $(\xi(y), y)$ for all y in D_y such that*

$$\forall (x, y) \in U \quad \forall t > 0 \text{ unlimited} \quad \Phi_y(t, x) \approx \xi(y)$$

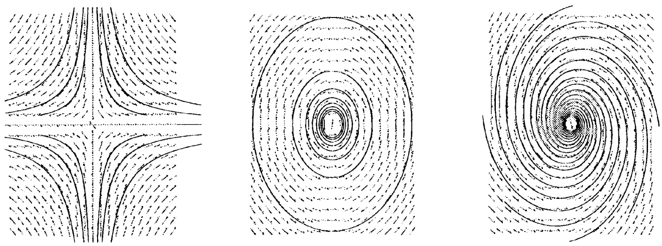


Figure 8.

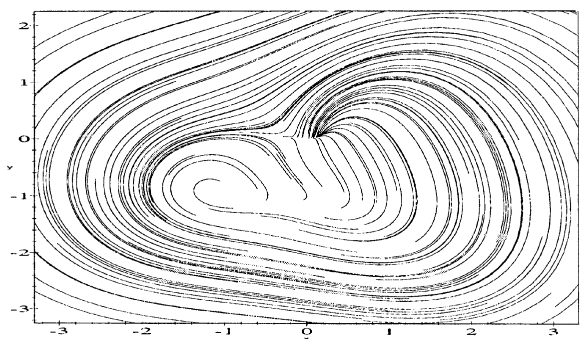


Figure 9. Is the origin attractive? Lyapunov-stable?

Exercise Study the uniform asymptotic Lyapunov stability of

$$\dot{x} = -x(x-y)(x-2y) \qquad y \in [-1,1]$$

7. Singular perturbations

The most classical singular perturbations are the slow-fast equations:

$$\begin{cases} \varepsilon\dot{x} = f(x,y) \\ \dot{y} = g(x,y) \end{cases}$$

where $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$. The functions f and g are supposed to be S-continuous, with shadows f_0 and g_0. Often, these functions are in fact standard functions of x, y and ε.

A very classical example is the VAN DER POL equation

$$\varepsilon\ddot{x} + (x^2-1)\,\dot{x} + x = 0$$

which we prefer write as a first order system:

$$\begin{cases} \varepsilon\dot{x} &= y - \left(\frac{x^3}{3} - x\right) \\ \dot{y} &= -x \end{cases}$$

7.1. GEOMETRIC DESCRIPTION

In this paragraph, I will not prove anything. I only want to describe a trajectory of the vector field. It is recommended to have always in mind the figure 10, although it is too particular (mainly because it is two-dimensional).

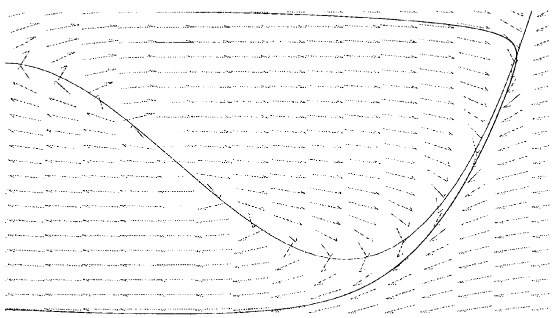


Figure 10. A slow-fast vector field.

At a point where $f(x, y) \not\simeq 0$, the vector field is quasi-horizontal and infinitely large. More precisely, its norm is unlimited, and the y-coordinate of the shadow of its direction is zero. Then, a trajectory will follow a horizontal curve, rapidly, until something happens. In the case of the picture, what happens is that $f_0(x, y)$ becomes small. After that, the trajectory moves with limited speed in the halo of the curve $f_0(x, y) = 0$ until it reaches an extremum of the cubic curve.

If you check the arrows indicating the direction of the vector field, you will see on the picture that the increasing segment of the curve $f_0(x, y) = 0$ is "attractive" and the decreasing one is "repulsive".

7.2. DEFINITIONS

To make precise the geometric description in the general case, and to be able to give theorems, we need some definitions. All the objects we define below are standard. They are determined by f_0 and g_0. Remember that the aim of our study is to describe as well as possible the trajectories of a slow-fast vector field only with the shadows of the given functions.

Definition 7.1 *For a slow-fast vector field*

$$\begin{cases} \varepsilon \dot{x} &= f(x,y) \\ \dot{y} &= g(x,y) \end{cases} \qquad \textit{where} \qquad \begin{cases} x \in \mathbb{R}^p \\ y \in \mathbb{R}^q \end{cases}$$

we define the following standard objects:

(a) The slow surface *is defined by its equation* $f_0(x,y) = 0$.

(b) The fast vector field *is the following standard family of vector fields on* $\mathbb{R}^p$ *indexed on* $\mathbb{R}^q$:

$$\dot{x} = f_0(x,y)$$

(c) The fast time *is* $\tau = t/\varepsilon$. *The derivative with respect to* τ *will be indicated by the symbol* $'$.

(d) If the slow surface contains the graph of a function $\varphi : D \subset \mathbb{R}^q \to \mathbb{R}^p$, *we define the* slow vector field *on this part of the slow surface by*

$$\begin{cases} \dot{y} &= g_0(\varphi(y), y) \\ \dot{x} &= \frac{d\varphi}{dy}(y)\, g_0(\varphi(y), y) \quad (\textit{ or } \quad x = \varphi(y)) \end{cases}$$

7.3. TIKHONOV'S THEOREM

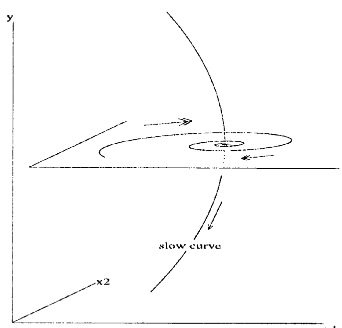


Figure 11. $p = 2$, $q = 1$.

Theorem 7.2 *Let* D *be a standard open subset of* $\mathbb{R}^q$ *and* $\varphi : D \to \mathbb{R}^p$ *such that* $\{(\varphi(y), y)\}$ *is a part of the slow surface. Let us suppose that the equilibrium point* $\varphi(y)$ *is uniformly asymptotically Lyapunov stable on* D *for the fast vector field.*

Let (x_0, y_0) *be a point with* ${}^\circ y_0 \in D$ *and* ${}^\circ x_0$ *in the attraction basin of* $\varphi({}^\circ y_0)$. *Let* $\gamma(t)$ *be the positive half-solution with initial condition* (x_0, y_0), *restricted to* ${}^\circ y \in D$ *and* t *limited.*

Then the shadow of the trajectory γ *is the union of*

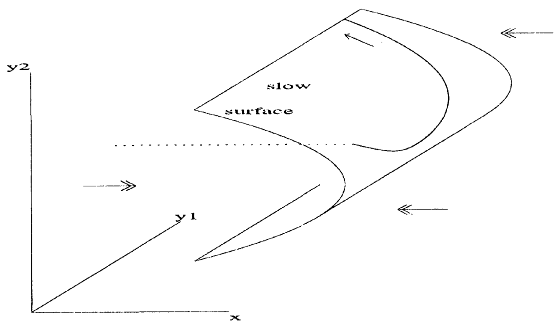


Figure 12. $p = 1$, $q = 2$.

1. *the positive half-trajectory of the fast vector field, with initial condition* $({}^\circ x_0, {}^\circ y_0)$,
2. *the positive half-trajectory of the slow vector field, with initial condition* $(\varphi({}^\circ y_0), {}^\circ y_0)$, *restricted to* ${}^\circ y \in D$.

Proof

1. The change of time $t = \varepsilon\tau$ yields the system

$$\begin{cases} x' &= f(x,y) \\ y' &= \varepsilon g(x,y) \end{cases}$$

which is a regular perturbation of

$$\begin{cases} x' &= f_0(x,y) \\ y' &= 0 \end{cases}$$

Let us denote by $\gamma_f(\tau)$ the positive half-solution of the fast vector field with initial condition $({}^\circ x_0, {}^\circ y_0)$. The short shadow lemma shows that

$$\forall \tau \text{ limited} \quad , \quad \gamma(\varepsilon\tau) \approx \gamma_f(\tau)$$

Using permanence principle, we find an unlimited $\tau_1 = t_1/\varepsilon$ such that

$$\forall \tau < \tau_1 \quad , \quad \gamma(\varepsilon\tau) \approx \gamma_f(\tau)$$

Using the asymptotic Lyapunov stability of $\varphi({}^\circ y_0)$, we proved that the shadow of $\gamma([0,t_1])$ is exactly the positive half-trajectory γ_f with its limit point $(\varphi({}^\circ y_0), {}^\circ y_0)$.

2. If we know that $\gamma([t_1,t_2]) \subset \text{halo}(x = \varphi(y))$, we have $\dot{y} \approx g_0(\varphi(y), y)$ and the short shadow lemma allows us to conclude.

3. Assume that there exists t_2 such that $\gamma(t_2) \notin \text{halo}(x = \varphi(y))$.
   (a) If U is the standard set given by the definition of uniform asymptotic Lyapunov stability, we can find a compact set K such that
   $$\text{halo}(x = \varphi(y), y \in D) \subset K \subset U$$
   (we already proved such a lemma in the proof of theorem 4.2.) Let t_3 be defined by
   $$t_3 = \inf\left(\{t > t_1 \mid \gamma(t) \notin K\} \cup \{t_2\}\right).$$
   Then, $t_3 \notin \text{halo}(x = \varphi(y))$ (prove it!, with a permanence principle) and $\gamma([t_1, t_3]) \in K$.
   (b) Applying the short shadow lemma to the negative half trajectory with initial condition $\gamma(t_3)$, we find a trajectory $\bar{\gamma}$ of the fast vector field such that
   $$\forall \tau < 0 \text{ limited}\ , \ \bar{\gamma}(\tau) \approx \gamma(t_3 + \varepsilon\tau)$$
   By permanence, it remains true for an unlimited τ_4. Then,
   $$\bar{\gamma}(\tau_4) \approx \gamma(t_4) \qquad t_4 = t_3 + \varepsilon\tau_4 \qquad \tau_4 < 0 \qquad \tau_4 \text{ unlimited.}$$
   (c) Moreover, $t_4 < t_3$ then $\gamma(t_4) \in K$ then $\bar{\gamma}(\tau_4) \in U$. The trajectory of the fast vector field with initial condition $\bar{\gamma}(\tau_4)$ goes, after an unlimited positive time $-\tau_4$ at point $\bar{\gamma}(0)$, outside the halo of $(x = \varphi(y))$. Then the hypothesis of uniform asymptotic stability couldn't be satisfied.

□•

Exercise

1. Show that the system
   $$\begin{cases} \varepsilon\dot{x} &= -x(x-y)(x-2y) \\ \dot{y} &= 1 \end{cases}$$
   doesn't satisfy the hypothesis of Tikhonov's theorem, when one takes $\varphi(y) = 0$ and $y \in [-1, 1]$.
2. Determine the shadow of the positive half solution with initial condition $(0, -1)$ of the system
   $$\begin{cases} \varepsilon\dot{x} &= -x(x-y)(x-2y) \\ \dot{y} &= 1 \end{cases}$$
   or of the system
   $$\begin{cases} \varepsilon\dot{x} &= -x(x-y)(x-2y) + 2\varepsilon \\ \dot{y} &= 1 \end{cases}$$
   Hint : $(2t, t)$ is a solution.

3. Deduce that, for some systems, the knowledge of f_0 and g_0 is not sufficient to determine the shadows of the trajectories of the system.

7.4. CANARDS

In this paragraph, I will not give the general theory of canards (see [1, 2, 3, 4, 8, 10, 11]). I will only explain the phenomenon by one example, the VAN DER POL equation which we write as a system.

$$\begin{cases} \varepsilon\dot{x} &= y - \left(\frac{x^3}{3} - x\right) \\ \dot{y} &= a - x \end{cases}$$

It is a well known family (with two parameters a and ε) of vector fields on $\mathbb{R}^2$. For us, ε is positive infinitesimal fixed. We suppose that $a > 0$.

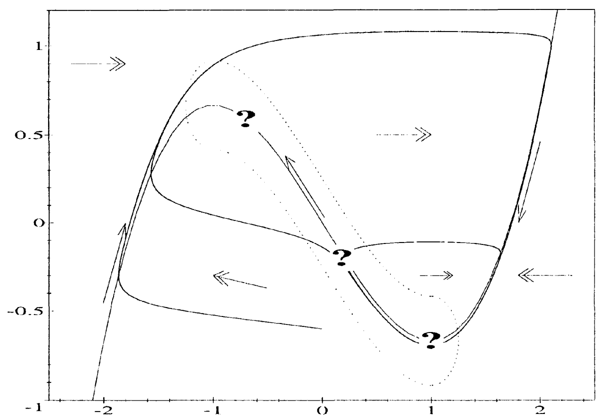


Figure 13. van der Pol equation.

The behaviour of the shadows of the positive half-trajectories is well determined by the Tikhonov's theorem, except in the "Danger area" $\mathcal{A}$:

$$\mathcal{A} = \left\{(x, y) \mid y \approx \frac{x^3}{3} - x \quad -1 \lesssim x \lesssim 1\right\}$$

Let us study a positive half trajectory with an initial condition (x_0, y_0) such that $x_0 > 1$, $y_0 > 0$. By Tikhonov's theorem, it goes fast into the halo of the right branch of the cubic, after that, it goes along the right branch, with $\dot{y} \approx a - x$.

If $°a > 1$, the trajectory remains in the halo of the equilibrium point $(a, \frac{a^3}{3} - a)$.

If $°a \leq 1$, it enters in the halo of $(1, -\frac{2}{3})$ and here, we are in the danger area $\mathcal{A}$.

If $°a < 1$, the vertical component of the vector field is always negative non infinitesimal in this halo, then the trajectory must go outside $\mathcal{A}$; it goes left, with the fast vector field.

Definition 7.3 *A value a of the parameter and a trajectory γ of the vector field will be called a* canard-value *and a* canard *if the shadow of γ contains first an attractive segment of the slow curve, and after that, a repulsive segment of the slow curve.*

Theorem 7.4 *There exists a canard.*

Idea of the proof The solution γ with fixed initial condition (x_0, y_0) is continuous with respect the parameter a. Then, the intermediate value theorem gives the proof.

In higher dimensions, this problem is more difficult and until now, there are many results when $q = 1$ (and p more than 1), but very few for greater q.□

In classical analysis, this result on the existence of canards is more difficult to explain, because the value a must be written as a function of ε. The following theorem gives more details on the canard-value:

Theorem 7.5 *If a is a canard-value, there exists a standard sequence a_n such that*

$$\forall^{\mathrm{st}} N \quad a = \sum_{i=0}^{N} a_i \varepsilon^i + \eta \varepsilon^N \qquad \eta \approx 0$$

Except in some very particular cases, the series $\sum a_i \varepsilon^i$ is divergent, but of Gevrey-type ([6]).

Idea of the proof First I have to say that the hint below is not the best way to prove this theorem; the proof of M. Canalis gives easier computations and more information on the series. But the following method is more geometric. It is a recursive proof on N, and I will explain the two first steps.

1. Use a magnifying glass around the cubic:

$$y = \left(\frac{x^3}{3} - x\right) + \varepsilon v_1$$

It gives a new equation with variables (x, v_1):

$$\begin{cases} \dot{x} &= v_1 \\ \varepsilon \dot{v}_1 &= a - x - (x^2 - 1) v_1 \end{cases}$$

2. It is a new slow-fast vector field, with slow curve

$$v_1 = \frac{{}^\circ a - x}{x^2 - 1}$$

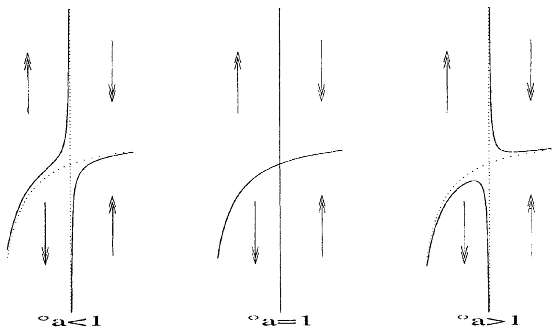


Figure 14. slow-fast vector field in (x, v_1).

Now, it is not too difficult to prove that (a, γ) can be a canard only if it goes along the slow curve of the system in (x, v_1). Then, the hypothesis ${}^\circ a = 1$ is needed. (We knew already this fact, but it is the first step of the recursive proof).

3. Let us use a magnifying glass around the curve $v_1 = \frac{1-x}{x^2-1}$:

$$v_1 = \frac{-1}{1+x} + \varepsilon v_2$$

It gives a new equation (with variables x and v_2):

$$\left\{ \begin{array}{rcl} \dot{x} & = & \frac{-1}{1+x} + \varepsilon v_2 \\ \varepsilon \dot{v_2} & = & \frac{a-1}{\varepsilon} + \frac{1}{(1+x)^3} - (x^2-1)v_2 - \varepsilon \frac{v_2}{(1+x)^2} \end{array} \right.$$

4. Iterating recursively (with standard indexes) steps 2 and 3, we prove the theorem.

References

1. Benoît, E. (1983), Systèmes lents-rapides dans $\mathbb{R}^3$ et leurs canards. In *Troisième rencontre du Schnepfenried, Astérisque* **109-110**(2), pp. 159–191. Société Mathématique de France.
2. Benoît, E. (1990), Canards et enlacements, *Publications de l'Inst. des Hautes Etudes Scient.* **72**, pp.63–91.

3. Benoît, E, (editor) (1991) *Dynamic Bifurcations.* Lecture Notes in Mathematics, **1493**, Springer Verlag.
4. Benoît, E., Callot, J.L., Diener, F. and Diener, M. (1981) Chasse au canard, *Collectanea Mathematica*, **31–32**(1–3), pp. 37–119.
5. Callot J.L. and Sari, T. (1983) Stroboscopie infinitésimale et moyennisation dans les systèmes d'équations différentielles à solution rapidement oscillante. In I.D. Landau (editor), *Outils et modèles mathématiques pour l'automatique, l'analyse des systèmes et le traitement du signal*, Editions du C.N.R.S. **3**.
6. Canalis-Durand, M. (1991) Formal expansion of van der Pol equation canard solutions are Gevrey. In E. Benoît (editor), *Dynamic Bifurcations* Lecture Notes in Mathematics **1493**, Springer Verlag, pp. 29–39.
7. Cutland, N.J. (1997) Nonstandard real analysis, *this volume.*
8. Diener F. and Diener, M. (1983), Sept formules relatives aux canards, *Comptes-Rendus de l'Académie des Sciences de Paris, série I*, **297**, pp. 577–580.
9. Diener, F. and Reeb, G. (1989) *Analyse Non Standard.* Collection Enseignement des Sciences. Hermann, Paris.
10. Troesch, A. (1985) Lorsque les canards naissent dans les tourbillons, in M. Diener and G. Wallet (editors), *Mathématiques finitaires et Analyse Non Standard*, Publications mathématiques de l'Université Paris 7, **31**(1) pp. 67–90 (publié en 1989).
11. Zvonkin, A.K., and Shubin, M.A. (1984) Non-standard analysis and singular perturbations of ordinary differential equations, *Russian Mathematical Surveys, 39*(2), pp. 69–131 (*Uspekhi Math. Nauk pp. 77–127*).

BETTER NONSTANDARD UNIVERSES WITH APPLICATIONS

R. JIN
Rutgers University
New Brunswick, NJ 08903
and
College of Charleston
Charleston, SC 29424
USA
email: rjin@math.rutgers.edu

1. Introduction

There are various reasons why some nonstandard universes are considered to be better than others. Different people may have different opinions and may choose different nonstandard universes to work within for different purposes. We think it is reasonable to consider that a nonstandard universe which possesses stronger power for deriving results in either standard or nonstandard mathematics, or which supplies more convenient tools so that, in practice, some complicated derivation procedures admit significant simplifications, is better than the one which doesn't.

In the early time of nonstandard analysis what people needed from a nonstandard universe was basically the existence of an infinitesimal. Since the introduction of κ-saturation (κ-saturation was first singled out in [24]), nonstandard analysis has experienced great prosperity. Given an infinite regular cardinal κ, a nonstandard universe $\mathcal{V}$ is said to be κ-saturated if every family of less then κ internal sets in $\mathcal{V}$ with the finite intersection property has non-empty intersection. $\aleph_1$-saturation is also called countable saturation. Countable saturation is one of the most commonly used properties in nonstandard analysis. Countable saturation gives us great convenience in handling countable sequences of internal sets or countable constructions with each step internal. For example, countable saturation laid down a foundation for the invention of Loeb measure construction by P. Loeb in [23], and guarantees the completeness of the nonstandard hull of

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 183–208.

a metric space. Part of the reason why countable saturation offers so much was made clear by C. W. Henson and H. J. Keisler in [11], where they showed that: second-order nonstandard arithmetic with countable saturation implies exactly the same sentences of standard second-order arithmetic as are implied by the system of standard third-order arithmetic.[1] Bearing those reasons in mind we consider that a nonstandard universe with countable saturation is better than the one without it.

Could we find an even better property for nonstandard universes? By that we mean to find a saturation-like property essentially stronger and evidently more useful than countable saturation for dealing with countable sequences of internal objects.

In §2 we introduce a property called *the isomorphism property* suggested by C. W. Henson in [7]. Henson's isomorphism property is elegant, stronger than countable saturation and very useful. In [8] Henson pointed out that a nonstandard universe satisfying the isomorphism property could be an ultrapower[2] of a standard superstructure by a result of S. Shelah [31]. In both [7] and [8] Henson studied nonstandard hulls of Banach spaces in a nonstandard universe satisfying the isomorphism property. The reader should see there that the isomorphism property makes the subject simple and clear. With three further sophisticated applications we illustrate more of the strength of the isomorphism property.

In §3 we push the issue further in the direction of §2. We introduce a property called *the special model axiom* suggested by D. Ross in [27] and a property called *full saturation.* The special model axiom is even stronger than the isomorphism property and full saturation is the strongest among them. Ross mentioned in [27] that the nonstandard universe satisfying the special model axiom could be an ultralimit of the standard superstructure. But it couldn't be an ultrapower without extra set theoretic assumptions beyond ZFC. The existence of a fully saturated nonstandard universe is even questionable. In fact, the existence of a fully saturated nonstandard universe is undecidable in ZFC. In §3 we present one application of the special model axiom and one application of full saturation. With those two applications the reader should have a taste of the power of those two properties.

Besides the reasons we mentioned above, there might be some other reasons why some nonstandard universes are better than others. For ex-

[1]The complete result in [11] is: Let $k \geqslant 2, n \in \mathbb{N}$ and $1 \leqslant m \leqslant \infty$. Then ${}^*PA^{(k)} + \Pi_m^{k-1} - CA + \beth_n^+$-saturation and $PA^{(k+n+1)} + \Pi_m^{k+n} - CA$ have the same consequences in the language of $PA^{(k)}$.

[2]By an ultrapower or an ultralimit in this chapter we always mean a nonstandard universe of the standard superstructure obtained by a bounded ultrapower or a bounded ultralimit construction, respectively.

ample, one may consider that a nonstandard universe is better if the hyperreal field in it bears a stronger resemblance to the standard real field. In §4 we introduce such a property called the λ-Bolzano-Weierstrass property for some uncountable regular cardinal λ. By a result of H. J. Keisler and J. H. Schmerl in [22] one can construct a countably saturated nonstandard universe satisfying the λ-Bolzano-Weierstrass property for some λ *e.g.* $\lambda = (2^{\aleph_0})^+$. But the λ-Bolzano-Weierstrass property is inconsistent with the isomorphism property. This shows that the better-ness of nonstandard universes in §4 takes a different direction from the better-ness in the sense of §§2,3.

Finally, we would like to mention, but not to include the details, one more kind of nonstandard universe called *Minimal Nonstandard Universes* . A minimal nonstandard universe is an ultrapower of the standard superstructure modulo a selective ultrafilter on a countable set. Note that such nonstandard universes are countably saturated. Some interesting applications of minimal nonstandard universes to measure theory were obtained by C. W. Henson and B. Wattenberg in [12] and by M. Benedikt in [2] and [3]. But the existence of a selective ultrafilter is undecidable in ZFC. For further study on this subject the reader should consult [12],[2] and [3].

In the end of each section we include some exercises. Star symbols $*$ are added to indicate difficulty.

Remark The definitions of the isomorphism property and the special model axiom could be trivially generalized to the κ-isomorphism property and the κ-special model axiom for any infinite regular cardinal κ. In fact, it was the κ-forms of those properties which were originally defined. In this chapter we need only the forms with $\kappa = \aleph_1$ for all applications. Hence we drop the cardinal κ in the definitions for simplicity. Similar to the generalization of countable saturation to κ-saturation the reader should be able to generalize those properties, without any difficulties, to the κ-forms for future applications.

We assume the reader has a basic training in model theory. A one-semester graduate level model theory course should be more than enough. Our notation in model theory is consistent with that in [5]. By a nonstandard universe in this chapter we mean always a triple $(\mathbb{V}(\mathbb{N}), \mathbb{V}({}^*\mathbb{N}), *)$, where $\mathbb{V}(\mathbb{N})$ is the superstructure on $\mathbb{N}$, the set of all standard natural numbers as urelements, $\mathbb{V}({}^*\mathbb{N})$ is the superstructure on ${}^*\mathbb{N}$, and $*$ is a nonstandard extension from $\mathbb{V}(\mathbb{N})$ to $\mathbb{V}({}^*\mathbb{N})$ defined in the first chapter by C. W. Henson in this volume [10]. We call $\mathbb{V}(\mathbb{N})$ the standard superstructure . Our restriction on the set of urelements for the standard superstructure to be $\mathbb{N}$ instead of an arbitrary set S is just for simplicity. Note that the standard superstructure $\mathbb{V}(\mathbb{N})$ contains a copy of the standard real field. In order to avoid writing the whole expression $(\mathbb{V}(\mathbb{N}), \mathbb{V}(^*\mathbb{N}), *)$ we always denote $\mathcal{V}$ for a nonstandard universe. We write $\alpha, \beta, \gamma, \ldots$ for ordinal numbers and denote ω for the first infinite ordinal. For any set S we denote $card(S)$ for the (external) cardinality of S. For a model $\mathfrak{A}$ with base set A we often write $card(\mathfrak{A})$ for the cardinality of A. We reserve the notion $|A|$ for the internal cardinality of A when A is an internal set. For each nonstandard universe $\mathcal{V}$ we denote $\Xi_{\mathcal{V}}$ for the cardinality of the family of all internal sets in $\mathcal{V}$. Note that $\Xi_{\mathcal{V}} = \bigcup_{n\in\omega} card(^*\mathbb{V}_n(\mathbb{N}))$. By a language we always mean a first-order language.

In order to be coherent we would like to give equivalent forms of countable saturation and κ-saturation in the next exercise, in terms of internally presented structures. First we define internally presented structures.

Definition 1.1 *Given a nonstandard universe $\mathcal{V}$ and given a language $\mathcal{L}$, an $\mathcal{L}$-structure $\mathfrak{A}$ is called* internally presented *in $\mathcal{V}$ iff the base set and every $\mathcal{L}$-relation or $\mathcal{L}$-function in $\mathfrak{A}$ is internal in $\mathcal{V}$.*

Note that an internally presented structure itself may not be internal when $\mathcal{L}$ contains infinitely many symbols.

1.1. EXERCISES

Exercise 1.2 *Show that*

(1) a nonstandard universe $\mathcal{V}$ is countably saturated iff for any countable language $\mathcal{L}$, for every internally presented $\mathcal{L}$-structure $\mathfrak{A}$, and for any set of first-order $\mathcal{L}$-formulas $\Gamma(x)$ with one free variable x consistent with $Th(\mathfrak{A})$, the set $\Gamma(x)$ is realizable (or satisfiable) in $\mathfrak{A}$.

(2) for any infinite regular cardinal κ a nonstandard universe $\mathcal{V}$ is κ-saturated iff for any language $\mathcal{L}$ with $card(\mathcal{L}) < \kappa$, for every internally presented $\mathcal{L}$-structure $\mathfrak{A}$, and for any set of first-order $\mathcal{L}$-formulas $\Gamma(x)$ with one free variable x consistent with $Th(\mathfrak{A})$, the set $\Gamma(x)$ is realizable in $\mathfrak{A}$.

2. The Isomorphism Property

Definition 2.1 *A nonstandard universe $\mathcal{V}$ is said to satisfy the* isomorphism property *iff any two elementarily equivalent, internally presented $\mathcal{L}$-structures for some countable language $\mathcal{L}$, are isomorphic.*

Remark Given two structures of some language, people might check the elementary equivalence between them by playing an Ehrenfeucht-Fraissé game or constructing a set of partial isomorphisms with the back-and-forth property [1]. So if one wants to show that two internally presented structures of some countable language in a nonstandard universe satisfying the isomorphism property are isomorphic, he might needs only to show that he could win the related Ehrenfeucht-Fraissé game or construct a set of partial isomorphisms with back-forth property.

From now on we will write *IP* for the isomorphism property.

Proposition 2.2 *(C. W. Henson [7]) If $\mathcal{V}$ satisfies IP, then every infinite internal set has cardinality $\Xi_{\mathcal{V}}$.*

Proof: It suffices to prove that any two infinite internal sets have same cardinality. Let A and B be two internal sets. Then A and B are two elementarily equivalent structures of the empty language. By *IP* there is a bijection between them. □

Proposition 2.3 *(C. W. Henson [7]) If $\mathcal{V}$ satisfies IP, then $\mathcal{V}$ is countably saturated.*

Proposition 2.3 is a trivial corollary of Lemma 2.6. See the remark right after Lemma 2.6.

Proposition 2.4 *(C. W. Henson [7] and S. Shelah [31]) Suppose κ is an infinite cardinal such that $\kappa^{\beth_\omega} = \kappa$.*[3] *Then there is an ultrafilter $\mathcal{F}$ on κ such that the ultrapower of the standard superstructure modulo $\mathcal{F}$ satisfies IP.*

The first two applications need an equivalent form of *IP* in terms of the realizability of a "second-order type" due to S. Shelah and the author [17].

Let $\mathcal{L}$ be a countable language and X be a new n-ary relation symbol not in $\mathcal{L}$. Suppose $\Gamma(X)$ is a set of first-order $\mathcal{L} \cup \{X\}$-sentences and $\mathfrak{A}$ is an $\mathcal{L}$-structure with base set A. We say that $\Gamma(X)$ is realizable in $\mathfrak{A}$ iff there exists $R \subseteq A^n$ such that

$$(\mathfrak{A}, R) \models \varphi(R)$$

for every $\varphi(X) \in \Gamma(X)$.

Definition 2.5 *A nonstandard universe $\mathcal{V}$ is said to satisfy the* resplendency property *iff for any countable language $\mathcal{L}$, for any n-ary new relation symbol X not in $\mathcal{L}$, for any internally presented $\mathcal{L}$-structure $\mathfrak{A}$, and for any set of first-order $\mathcal{L} \cup \{X\}$-sentences $\Gamma(X)$, if $\Gamma(X) \cup Th(\mathfrak{A})$ is consistent, then $\Gamma(X)$ is realizable in $\mathfrak{A}$.*

Lemma 2.6 *(R. Jin and S. Shelah [17]) A nonstandard universe $\mathcal{V}$ satisfies IP iff $\mathcal{V}$ satisfies the resplendency property.*

[3]Note that $\beth_0 = \aleph_0$, $\beth_{n+1} = 2^{\beth_n}$ and $\beth_\omega = \bigcup_{n<\omega} \beth_n$. Note also that the cardinality of the standard superstructure is $\beth_\omega$.

Remark Comparing with Exercise 1.2, it is easy to see that the resplendency property is a natural generalization of countable saturation. To get countable saturation simply view the new relation symbol X in the definition as an individual variable. This justifies why we call *IP* a saturation-like property.

2.1. UNLIMITED LOEB MEASURE SPACES

Suppose we are given an internal *finitely additive measure space $(\Omega, \mathcal{A}, \mu)$ such that $\mathcal{A}$ contains every singleton set, $\mu(\Omega) > n$ for every $n \in \mathbb{N}$, and $\mu(\{x\}) < \frac{1}{n}$ for every $x \in \Omega$ and $n \in \mathbb{N}$. Following the Loeb construction we can extend the standard finitely additive measure ${}^\circ\mu$ on $\mathcal{A}$ to a countably additive measure L_μ on the σ-algebra generated by $\mathcal{A}$. But this measure is not complete. In order to make the measure complete, we have to throw in the subsets of all L_μ-measure zero sets. We need to define a σ-algebra of all "measurable" sets for this. For each $S \subseteq \Omega$ let the outer measure of S be

$$\overline{{}^\circ\mu}(S) = \inf\{{}^\circ\mu(A) : A \in \mathcal{A} \wedge S \subseteq A\}$$

and the inner measure of S be

$$\underline{{}^\circ\mu}(S) = \sup\{{}^\circ\mu(A) : A \in \mathcal{A} \wedge A \subseteq S\}.$$

If $L_\mu(\Omega)$ were finite, then it would be immediate that a set S is measurable iff its outer measure and inner measure coincide. But now we have $L_\mu(\Omega) = \infty$ (this is what the word "unlimited" means). By a conventional method we define the σ-algebra $\mathcal{B}$ of all "measurable" sets as follows:

$$\mathcal{B} = \{S \subseteq \Omega : (\forall A \in \mathcal{A})\ ({}^\circ\mu(A) < \infty \rightarrow \overline{{}^\circ\mu}(S \cap A) = \underline{{}^\circ\mu}(S \cap A))\}.$$

Note that $\mathcal{B}$ is a σ-algebra, $\mathcal{A} \subseteq \mathcal{B}$, and every subset of an L_μ-measure zero set is in $\mathcal{B}$.

When trying to extend L_μ to a complete measure on $\mathcal{B}$ one encounters a problem. For an $S \in \mathcal{B}$ should one let $L_\mu(S) = \overline{{}^\circ\mu}(S)$ or let $L_\mu(S) = \underline{{}^\circ\mu}(S)$? There would have no problem if every $S \in \mathcal{B}$ had same outer measure and inner measure. But this may not be the case. So one has at least two different ways to extend L_μ to a complete countably additive measure on $\mathcal{B}$ if there are some sets in $\mathcal{B}$ having different outer measure and inner measure.

Theorem 2.7 *(R. Jin and S. Shelah [17]) Suppose $\mathcal{V}$ satisfies IP. Then every unlimited, non-atomic Loeb measure space has a measurable subset with infinite outer measure and zero inner measure.*

Proof: Given an unlimited, non-atomic Loeb measure space $(\Omega, \mathcal{B}, L_\mu)$ generated by an internal space $(\Omega, \mathcal{A}, \mu)$ as above. We need to find a set $S \subseteq \Omega$ such that for any $A \in \mathcal{A}$, if $\mu(A)$ is finite, then $\overline{{}^\circ\mu}(A \cap S) = 0$, and if $S \subseteq A$, then $L_\mu(A) = \infty$.

We will use the resplendency property in the proof. The main idea is the following: First we write a set of sentences $\Gamma(X)$ to express that X is the desired set, *i.e.* a measurable set with infinite outer measure and zero inner measure. Then we prove that $\Gamma(X)$ is consistent with our internally presented structure. By the resplendency property one could find a set realizing $\Gamma(X)$ in our structure. Therefore, the desired set exists.

We first form an internally presented structure $\mathfrak{A}$ of a finite language $\mathcal{L}_\mathfrak{A}$. Let

$$\mathfrak{A} = (\Omega \cup \mathcal{A} \cup {}^*\mathbb{R}; \Omega, \mathcal{A}, {}^*\mathbb{R}, \in, \mu, \cap, \smallsetminus, +, \cdot, <, 0, 1),$$

where $\Omega, \mathcal{A}$ and ${}^*\mathbb{R}$ are unary relations, $\in$ is a membership relation between Ω and $\mathcal{A}$, μ is the measure function from $\mathcal{A}$ to ${}^*\mathbb{R}$, $\cap$ and $\smallsetminus$ are Boolean operators on $\mathcal{A}$, and $({}^*\mathbb{R}; +, \cdot, <, 0, 1)$ is the hyperreal field in $\mathcal{V}$. Notice that we abuse the notation here. Rigorously, we should use ${}^*\in, {}^*\cap, {}^*+$, etc. We often omit $*$ when the meaning is obvious. Let X be a new unary relation symbol not in $\mathcal{L}_\mathfrak{A}$. We now express X as a desired set by a set of $\mathcal{L} \cup \{X\}$-sentences $\Gamma(X)$. Let

$$\Gamma(X) = \{\phi(X)\} \cup \{\psi_n(X) : n \in \mathbb{N}\} \cup \{\chi_n(X) : n \in \mathbb{N}\},$$

where

$$\phi(X) =: \forall x (X(x) \to \Omega(x)),$$

$$\begin{aligned}\psi_n(X) =: \;& \forall U (\mathcal{A}(U) \wedge \mu(U) < n \to \\ & \exists V (\mathcal{A}(V) \wedge \forall x (X(x) \wedge x \in U \to x \in V) \wedge \mu(V) < \tfrac{1}{n})),\end{aligned}$$

$$\chi_n(X) =: \forall U (\mathcal{A}(U) \wedge \forall x (X(x) \to x \in U) \to \mu(U) > n).$$

Note that in $\mathfrak{A}$, the elements n and $\frac{1}{n}$ are definable. The sentence $\phi(X)$ says that X is a subset of Ω, the sentence $\psi_n(X)$ says that if $U \in \mathcal{A}$ has measure $< n$, then the outer measure of $U \cap X$ is $< \frac{1}{n}$, and $\chi_n(X)$ says that if $U \in \mathcal{A}$ and $X \subseteq U$, then U has measure $> n$. So if the set $\Gamma(X)$ is realized by some set S in $\mathfrak{A}$, then S is clearly the desired set.

By the resplendency property we need only to prove that $\Gamma(X)$ is consistent with $Th(\mathfrak{A})$.

Claim $\Gamma(X) \cup Th(\mathfrak{A})$ is consistent.

Proof of Claim: By the Downward Löwenheim-Skolem theorem we can find a countable $\mathcal{L}_\mathfrak{A}$-structure $\mathfrak{A}_0$ such that $\mathfrak{A}_0 \models Th(\mathfrak{A})$. Let $\Omega_0 \cup \mathcal{A}_0 \cup \mathbb{R}_0$ be the base set of $\mathfrak{A}_0$. It suffices to prove that $\Gamma(X)$ is realizable in $\mathfrak{A}_0$.

So we want to find a set $S_0 \subseteq \Omega_0$ such that S_0 realizes $\Gamma(X)$ in $\mathfrak{A}_0$. Let $\{A_n : n \in \mathbb{N}\}$ be an enumeration of the set

$$\{A \in \mathcal{A}_0 : \exists n \in \mathbb{N}(\mu(A) < n)\}.$$

We choose x_n inductively such that

$$x_n \in \Omega_0 \smallsetminus ((\bigcup_{k=0}^{n-1} A_k) \cup \{x_k : k < n\}).$$

The set $\Omega_0 \smallsetminus ((\bigcup_{k=0}^{n-1} A_k) \cup \{x_k : k < n\})$ is non-empty because it has unlimited measure. Let $S_0 = \{x_n : n \in \mathbb{N}\}$. It is easy to check that $(\mathfrak{A}_0, S_0)$ realizes $\Gamma(X)$. □

Remark: The result in the theorem was first proved by Henson [9] in a nonstandard universe called a polyenlargement. Then Ross in [27] proved the result by assuming the special model axiom. It is still open if one could settle the result by countable saturation or κ-saturation for any infinite regular cardinal κ.

2.2. THE EXISTENCE OF BAD CUTS

Let H be a hyperfinite integer and let $\mathcal{H} = \{0, 1, \ldots, H-1\}$. $\mathcal{H}$ is called a hyperfinite time line in [21]. Let L_μ be the Loeb probability measure on $\mathcal{H}$ generated by the internal normalized uniform counting measure μ. Through a standard part map $x \mapsto {}^\circ(\frac{x}{H})$ the hyperfinite time line $\mathcal{H}$ together with L_μ could be closely associated with the standard unit interval $[0,1]$ together with Lebesgue measure on it. What about the order topology? Could one define a topology on $\mathcal{H}$ which resembles the usual order topology on $[0,1]$? Note that the natural order topology on $\mathcal{H}$ is discrete, hence not interesting. H. J. Keisler and S. C. Leth in [21] gave the definition of a U-topology on $\mathcal{H}$ for each cut $U \subseteq \mathcal{H}$.

An infinite initial segment U of ${}^*\mathbb{N}$ is called a cut (or additive cut in some literature) if $a + b \in U$ for any $a, b \in U$. Given a cut $U \subseteq \mathcal{H}$, a set $O \subseteq \mathcal{H}$ is U-open iff for any $a \in O$ there exists a $d \in \mathcal{H} \smallsetminus U$ such that $[a-d, a+d] \subseteq O$, where $[a-d, a+d] = \{x \in \mathcal{H} : a - d \leqslant x \leqslant a + d\}$. All U-open sets form a U-topology. Given a point $x \in \mathcal{H}$, by the U-monad of x we mean the set $\{y \in \mathcal{H} : |y - x| \in U\}$. Note that no U-topology could be Hausdorff because it could not separate two points within a U-monad. But if we view each U-monad as a "point" and consider the natural order of those monads, then the resulting "order" topology is just the U-topology.

With a U-topology we can define U-nowhere dense sets and U-meager sets (recall that a meager set is a countable union of nowhere dense sets).

Definition 2.8 *(H. J. Keisler and S. C. Leth [21]) A cut U is called a* good cut *if there exists, in $\mathcal{H}$, a U-meager set of Loeb measure one.*

Recall that the standard unit interval has a meager set of Lebesgue measure one. So a good cut U will make $\mathcal{H}$ together with Loeb measure and the U-topology much like $[0,1]$ together with Lebesgue measure and the order topology.

Keisler and Leth proved in [21] that almost all cuts are good. They were also able to construct a bad cut in some nonstandard universe under an assumption beyond ZFC in set theory, *e.g.* $2^{\aleph_0} < 2^{\aleph_1}$. In next theorem we construct a bad cut in every nonstandard universe satisfying *IP* in ZFC.

Theorem 2.9 *([15]) If $\mathcal{V}$ satisfies IP, then every hyperfinite time line has bad cuts .*

In the proof we need an equivalent form of bad-ness from [21]. Given a cut U in ${}^*\mathbb{N}$, an internal increasing function $f : \{0, 1, \ldots, L_f\} \mapsto {}^*\mathbb{N}$ for some $L_f \in {}^*\mathbb{N}$ is called a U-crossing sequence iff the set $U \cap \{f(n) : n \leqslant L_f\}$ is upper unbounded in U.

Lemma 2.10 *(H. J. Keisler and S. C. Leth [21], Proposition 4.5) For any cut $U \subseteq \mathcal{H}$ the following are equivalent.*

(1) There is a U-meager set of Loeb measure one.

(2) There is a U-crossing sequence f such that

$$\sum_{n<L_f} \frac{f(n)}{f(n+1)} < 1.$$

Proof of Theorem 2.9: By Lemma 2.10 it suffices to construct a cut $U \subseteq \mathcal{H}$ such that for every U-crossing sequence f the sum

$$\sum_{n<L_f} \frac{f(n)}{f(n+1)}$$

is not finite. The main idea of the proof is similar to the idea in the proof of Theorem 2.7. Let $\mathcal{F}$ be the set of all internal increasing functions f from $\{0, 1, \ldots, L_f\}$ for some $L_f \in \mathcal{H}$, to $\mathcal{H}$. Then $\mathcal{F}$ is internal. We form an internally presented structure $\mathfrak{A}$. Let

$$\mathfrak{A} = (\mathcal{H} \cup \mathcal{F} \cup {}^*\mathbb{R}; \mathcal{H}, \mathcal{F}, {}^*\mathbb{R}, R, S, +, \cdot, <, 0, 1),$$

where $\mathcal{H}, \mathcal{F}$ and ${}^*\mathbb{R}$ are unary relations, R is a ternary relation such that $\langle a, b, f\rangle \in R$ iff $f \in \mathcal{F}$, $a \in dom(f)$ and $f(a) = b$, S is a function from $\mathcal{F}$ to ${}^*\mathbb{R}$ such that for any $f \in \mathcal{F}$

$$S(f) = \sum_{n<L_f} \frac{f(n)}{f(n+1)},$$

and $({}^*\mathbb{R};+,\cdot,<,0,1)$ is the hyperreal field in $\mathcal{V}$. Let $\mathcal{L}$ be the language of $\mathfrak{A}$. Note that the following first-order $\mathcal{L}$-sentences are true in $\mathfrak{A}$.

$$\theta_n =: \exists x(\mathcal{H}(x) \wedge x \geqslant n \wedge \forall y(\mathcal{H}(y) \rightarrow y \leqslant x))$$

for each $n \in \mathbb{N}$, and

$$\begin{aligned} \eta =: \quad & \forall f \forall x \forall y(\mathcal{F}(f) \wedge \mathcal{H}(x) \wedge \mathcal{H}(y) \wedge x < y \rightarrow \\ & \exists g(\mathcal{F}(g) \wedge range(g) = range(f) \cap [x,y])), \end{aligned}$$

where the formula $range(g) = range(f) \cap [x,y]$ is an abbreviation of the first-order $\mathcal{L}$-formula

$$\forall z(\exists u R(u,z,g) \leftrightarrow x \leqslant z \wedge z \leqslant y \wedge \exists u R(u,z,f)).$$

Let $X \notin \mathcal{L}$ be a unary predicate symbol. We define $\Gamma(X)$ to be the set of $\mathcal{L} \cup \{X\}$-sentences which contains exactly the following:

$$\varphi_1(X) =: \forall x(X(x) \rightarrow \mathcal{H}(x))$$

$$\varphi_2(X) =: \forall x \forall y(x \leqslant y \wedge \mathcal{H}(x) \wedge X(y) \rightarrow X(x))$$

$$\varphi_3(X) =: \forall x \forall y(X(x) \wedge X(y) \rightarrow X(x+y))$$

$$\begin{aligned} \psi_n =: \quad & \forall f(\mathcal{F}(f) \wedge \forall x(X(x) \rightarrow \\ & \exists y \exists z(R(y,z,f) \wedge X(z) \wedge x \leqslant z)) \rightarrow S(f) \geqslant n) \end{aligned}$$

for each $n \in \mathbb{N}$.

Note that the sentences $\varphi_1(X)$, $\varphi_2(X)$ and $\varphi_3(X)$ say that X is a cut in $\mathcal{H}$. The sentences $\psi_n(X)$ for $n \in \mathbb{N}$ say that if f is a crossing sequence of X, then the internal sum $S(f)$ is not finite. So $\Gamma(X)$ describes that X is a bad cut by Lemma 2.10. So if $\Gamma(X)$ is realizable in $\mathfrak{A}$, then $\mathcal{H}$ must contain a bad cut. By Lemma 2.6 it suffices to show that $\Gamma(X) \cup Th(\mathfrak{A})$ is consistent.

Let $\mathfrak{A}'$ be a countable elementary submodel of $\mathfrak{A}$. Then $Th(\mathfrak{A}') = Th(\mathfrak{A})$. If we show that $\Gamma(X)$ is realizable in $\mathfrak{A}'$, then it is clear that $Th(\mathfrak{A}) \cup \Gamma(X)$ is consistent.

Claim $\Gamma(X)$ is realizable in $\mathfrak{A}'$.

Proof of Claim: Let $A' = \mathcal{H}' \cup \mathcal{F}' \cup \mathbb{R}'$ be the base set of $\mathfrak{A}'$ and let $\mathcal{F}' = \{f_i : i \in \mathbb{N}\}$. We now inductively construct an increasing sequence $\langle a_i : i \in \mathbb{N}\rangle$ and a decreasing sequence $\langle b_i : i \in \mathbb{N}\rangle$ in $\mathcal{H}'$ such that for each $i \in \mathbb{N}$

(a) $a_i < b_i$,

(b) $2a_i < a_{i+1}$,

(c) b_i/a_i is not finite in $\mathbb{R}'$,

(d) if $f \in \mathcal{F}'$ such that

$$range(f) = range(f_i) \cap \{x \in \mathcal{H}' : a_i \leqslant x \leqslant b_i\},$$

if $S(f)$ is finite in $\mathbb{R}'$ and if L_f is not finite, then there is a $k \in \{0, 1, \ldots, L_f - 1\} \cap \mathcal{H}'$ such that $f(k) \leqslant a_{i+1}$ and $f(k+1) \geqslant b_{i+1}$ (*i.e.* f has a jump across the interval (a_{i+1}, b_{i+1})).

We show first that the claim follows from the construction. Let

$$U = \{x \in \mathcal{H}' : (\exists i \in \mathbb{N})\,(x \leqslant a_i)\}.$$

Then $\varphi_1(U)$ and $\varphi_2(U)$ are trivially true in $(\mathfrak{A}', U)$. The sentence $\varphi_3(U)$ is true in $(\mathfrak{A}', U)$ by condition (b). Given any $f_i \in \mathcal{F}'$ such that f_i is a crossing sequence of U. To show that $\psi_n(U)$ is true in $(\mathfrak{A}', U)$ for any $n \in \mathbb{N}$ we need only to show that $S(f_i)$ is not finite. Suppose $S(f_i)$ is finite. By the fact that η is true in $\mathfrak{A}'$ there exists a $g \in \mathcal{F}'$ such that

$$range(g) = range(f_i) \cap \{x \in \mathcal{H}' : a_i \leqslant x \leqslant b_i\}.$$

Then $S(g)$ is also finite because $S(g) \leqslant S(f_i)$. Since f_i is a crossing sequence of U, $a_i \in U$ and $b_i \notin U$, then g is also a crossing sequence of U. Hence L_g is not finite (since no finite sequence could be a crossing sequence of any cut). By condition (d) we know that g has a jump from a_{i+1} to b_{i+1}, *i.e.* $g(k) \leqslant a_{i+1}$ and $g(k+1) \geqslant b_{i+1}$ for some $k \in dom(g)$. So g can't be a crossing sequence of U, a contradiction.

We now do the inductive construction. Choose any a_1 and b_1 in $\mathcal{H}'$ such that b_1/a_1 is not finite (for example, $a_1 = 1$ and $b_1 = H$). Suppose we have found $\langle a_i : i < k\rangle$ and $\langle b_i : i < k\rangle$ for some $k > 1$ such that they satisfy the conditions (a)—(d). We need to find a_k and b_k. Let $g \in \mathcal{F}'$ be such that

$$range(g) = range(f_{k-1}) \cap \{x \in \mathcal{H}' : a_{k-1} \leqslant x \leqslant b_{k-1}\}.$$

Case 1: $S(g)$ is not finite or L_g is finite. Simply let $a'_k = a_{k-1}$ and $b'_k = b_{k-1}$.

Case 2: $S(g)$ is finite and L_g is not finite. Let $m \in \mathbb{N}$ be such that $S(g) < m$. Since g is an element in $\mathfrak{A}'$ and $\mathfrak{A}' \preceq \mathfrak{A}$, then there is a t in $\mathfrak{A}'$ such that

$$t = \min\{\frac{g(n)}{g(n+1)} : n \in \mathcal{H}' \wedge n < L_g\}.$$

Let $n_0 \in \mathcal{H}'$ and $n_0 < L_g$ be such that $t = g(n_0)/g(n_0+1)$. Then

$$tL_g \leqslant \sum_{n<L_g} \frac{g(n)}{g(n+1)} = S(g) < m.$$

So we have $g(n_0+1)/g(n_0) \geqslant L_g/m$. Now let $a'_k = g(n_0)$ and $b'_k = g(n_0+1)$.

Clearly, we have that b'_k/a'_k is not finite. Let $a_k = 2a'_k$ and $b_k = b'_k - 1$. Then it is easy to see that b_k/a_k is still not finite. Now it is obvious that the sequences

$$\langle a_i : i < k+1 \rangle \text{ and } \langle b_i : i < k+1 \rangle$$

satisfy conditions (a)—(d). $\square$

2.3. DEFINABILITY IN CONSTRAINT QUERY LANGUAGES

Although the following application is from database theory, we will present it in a purely model theoretic way. No knowledge on database theory is assumed. The application is due to M. Benedikt *et al.* [4].

Let $\mathcal{L}$ be the language of ordered fields and let $\mathcal{L}' = \{X_1, \ldots, X_n\}$, where X_i is an n_i-ary relation symbol not in $\mathcal{L}$. Let $\mathcal{R}$ be the standard real field.

By an $\mathcal{L}'$-finite expansion of $\mathcal{R}$ we mean an expansion of $\mathcal{R}$ to an $\mathcal{L} \cup \mathcal{L}'$-structure $\mathcal{R}_e = (\mathcal{R}, R_1, \ldots, R_n)$ such that the interpretation $e(X_i) = R_i$ of every symbol X_i in $\mathcal{L}'$ is a finite relation in $\mathcal{R}$.

For an $\mathcal{L}'$-finite expansion $\mathcal{R}_e$ let the *carrier* of $\mathcal{R}_e$ be the set

$$\begin{aligned} C_e = \ & \{r \in \mathbb{R} : \exists X_i \in \mathcal{L}' \ \exists (a_1, \ldots, a_{n_i}) \in \mathbb{R}^{n_i} \\ & (r \in \{a_1, \ldots, a_{n_i}\} \wedge (a_1, \ldots, a_{n_i}) \in e(X_i))\}. \end{aligned}$$

Clearly, C_e is finite.

Two $\mathcal{L} \cup \mathcal{L}'$-sentences ϕ and ψ are equivalent over $\mathcal{R}$ iff for any $\mathcal{L}'$-finite expansion $\mathcal{R}_e$,

$$\mathcal{R}_e \models \phi \text{ iff } \mathcal{R}_e \models \psi.$$

Given an $\mathcal{L}'$-finite expansion $\mathcal{R}_e$ and an order-preserving injection

$$\Phi : C_e \mapsto \mathbb{R},$$

let $\mathcal{R}_{\Phi(e)}$ be another $\mathcal{L}'$-finite expansion such that $C_{\Phi(e)} = \Phi[C_e]$ and

$$\Phi(e)(X_i) = \{(\Phi(a_1), \ldots, \Phi(a_{n_i})) : (a_1, \ldots, a_{n_i}) \in e(X_i)\}.$$

It is easy to see that $\mathcal{R}_e \restriction \mathcal{L}' \cup \{<\}$ and $\mathcal{R}_{\Phi(e)} \restriction \mathcal{L}' \cup \{<\}$, the reducts of $\mathcal{R}_e$ and $\mathcal{R}_{\Phi(e)}$ on $\mathcal{L}' \cup \{<\}$, are isomorphic.

Definition 2.11 *An $\mathcal{L} \cup \mathcal{L}'$-sentence ϕ is* order-invariant *in $\mathcal{R}$ iff for any $\mathcal{L}'$-finite expansion $\mathcal{R}_e$ and for any order-preserving injection $\Phi : C_e \mapsto \mathbb{R}$*

$$\mathcal{R}_e \models \phi \textit{ iff } \mathcal{R}_{\Phi(e)} \models \phi.$$

Theorem 2.12 *(M. Benedikt et al. [4]) For any order-invariant $\mathcal{L} \cup \mathcal{L}'$-sentence ϕ there exists an $\mathcal{L}' \cup \{<\}$-sentence ψ such that ϕ and ψ are equivalent over $\mathcal{R}$.*[4]

The proof of this theorem is long. The main idea of the proof is as follows: By assuming the contrary we first find two $\mathcal{L}'$-*finite expansions $^*\mathcal{R}_e$ and $^*\mathcal{R}_{e'}$ so that they agree on all $\mathcal{L}' \cup \{<\}$-sentences but disagree on ϕ, *i.e.*

$$^*\mathcal{R}_e \models \varphi \text{ iff } ^*\mathcal{R}_{e'} \models \varphi$$

for every $\mathcal{L}' \cup \{<\}$-sentence φ and

$$^*\mathcal{R}_e \models \phi \text{ iff } ^*\mathcal{R}_{e'} \models \neg\phi.$$

Then we show, by the order-invariance of ϕ, that $^*\mathcal{R}_e$ and $^*\mathcal{R}_{e'}$ could be chosen so that C_e and $C_{e'}$ are subsets of an $\mathcal{L}$-indiscernible sequence in $^*\mathbb{R}$. Finally, one derives a contradiction by showing that $^*\mathcal{R}_e$ and $^*\mathcal{R}_{e'}$ agree on ϕ.

Proof of Theorem 2.12 Suppose the theorem is not true and let ϕ be a witness. Let $\{\varphi_n : n \in \mathbb{N}\}$ be an enumeration of all $\mathcal{L}' \cup \{<\}$-sentences.

Claim 1 For any $m \in \mathbb{N}$ there are two $\mathcal{L}'$-finite expansions $\mathcal{R}_e$ and $\mathcal{R}_{e'}$ such that they agree on every φ_n for $n < m$ and disagree on ϕ.

Proof of Claim 1: Otherwise ϕ will be equivalent to some Boolean combination of those φ_n's for $n < m$. $\square$(Claim 1)

Let $\mathcal{V}$ be a nonstandard universe satisfying *IP*.

By Claim 1 and Transfer Principle we can find two $\mathcal{L}'$-*finite expansions $^*\mathcal{R}_e$ and $^*\mathcal{R}_{e'}$ such that they agree on every φ_n for $n < m$ and disagree on ϕ. By countable saturation one can let m be ∞. So we have now two $\mathcal{L}'$-*finite expansions $^*\mathcal{R}_e$ and $^*\mathcal{R}_{e'}$ such that they agree on $\{\varphi_n : n \in \mathbb{N}\}$ and disagree on ϕ. Let $|C_e| = H_e$ and let $|C_{e'}| = H_{e'}$.

Again by Transfer Principle and countable saturation one can find an $\mathcal{L}$-indiscernible increasing internal sequence $\langle r_i : i \in {}^*\mathbb{N}\rangle$ in $^*\mathbb{R}$.

Claim 2 We can assume that C_e and $C_{e'}$ are the sets $\{r_i : i < H_e\}$ and $\{r_i : i < H_{e'}\}$, respectively.

Proof of Claim 2: For $^*\mathcal{R}_e$ there is an internal order-preserving bijection Φ from C_e to $\{r_i : i < H_e\}$. By Transfer Principle and the order-invariance of ϕ one has that $^*\mathcal{R}_e$ and $^*\mathcal{R}_{\Phi(e)}$ agree on ϕ. $^*\mathcal{R}_e$ and $^*\mathcal{R}_{\Phi(e)}$ agree also on $\{\varphi_n : n \in \mathbb{N}\}$ because the reducts of $^*\mathcal{R}_e$ and $^*\mathcal{R}_{\Phi(e)}$ on $\mathcal{L}' \cup \{<\}$ are isomorphic. Same for $^*\mathcal{R}_{e'}$. $\square$(Claim 2)

[4]The original theorem in [4] is more general with $\mathcal{R}$ replaced by any infinite o-minimal structure.

Claim 3 There is an order-preserving bijection I from C_e to $C_{e'}$ such that

$$(a_1, \ldots, a_{n_i}) \in e(X_i) \text{ iff } (I(a_1), \ldots, I(a_{n_i})) \in e'(X_i)$$

for $i = 1, \ldots, n$.

Proof of Claim 3: Note that "$x \in C_e$" could be written as an $\mathcal{L}'$-formula. Because the submodel of ${}^*\mathcal{R}_e \restriction \mathcal{L}' \cup \{<\}$ generated by C_e and the submodel of ${}^*\mathcal{R}_{e'} \restriction \mathcal{L}' \cup \{<\}$ generated by $C_{e'}$ are elementarily equivalent, by *IP*, there is an isomorphism I between them. (Here is the place we use *IP*.) $\square$(Claim 3)

Let $\mathcal{L}_C$ be the language $\mathcal{L} \cup \{c_i : i < H_e\}$, where c_i is a constant symbol not in $\mathcal{L}$. Note that ${}^*\mathcal{R}_C = ({}^*\mathcal{R}, r_i)_{i<H_e}$ and ${}^*\mathcal{R}_{C'} = ({}^*\mathcal{R}, I(r_i))_{i<H_e}$ are two $\mathcal{L}_C$-structures by interpreting c_i as r_i or as $I(r_i)$, respectively.

Claim 4 ${}^*\mathcal{R}_C$ and ${}^*\mathcal{R}_{C'}$ are elementarily equivalent.

Proof of Claim 4: It is a straightforward consequence of indiscernibility of r_i's. $\square$(Claim 4)

Let $\mathcal{L}_C^m$ be the language $\mathcal{L}_C \cup \{d_1 \ldots, d_m\}$ and let $\mathcal{L}^m$ be the language $\mathcal{L} \cup \{d_1 \ldots, d_m\}$, where d_i's are new constant symbols.

We need a property, called o-minimality, of real fields in the next claim. Let $\mathcal{F}$ be either a standard real field or a hyperreal field. for any $\mathcal{L}$-formula $\chi(x, \bar{y})$ and any $\bar{r}$ in $\mathcal{F}$ the set defined by $\chi(x, \bar{r})$ in $\mathcal{F}$ is a finite union of intervals. Let those intervals be maximal and let $E_\chi^{\bar{r}}$ be the set of all endpoints of those intervals. Clearly, $E_\chi^{\bar{r}}$ is finite and every point in $E_\chi^{\bar{r}}$ is definable by some $\mathcal{L}$-formula with parameters from $\bar{r}$.

Claim 5 Suppose $\bar{u} = (u_1, \ldots, u_m)$ and $\bar{v} = (v_1, \ldots, v_m)$ are in ${}^*\mathbb{R}^m$ such that $({}^*\mathcal{R}_C, \bar{u})$ and $({}^*\mathcal{R}_{C'}, \bar{v})$ are elementarily equivalent in $\mathcal{L}_C^m$. Then

(1) for every $u_{m+1} \in {}^*\mathbb{R}$ there is a $v_{m+1} \in {}^*\mathbb{R}$ such that $({}^*\mathcal{R}_C, \bar{u}, u_{m+1})$ and $({}^*\mathcal{R}_{C'}, \bar{v}, v_{m+1})$ are elementarily equivalent in $\mathcal{L}_C^{m+1}$,

(2) for every $v_{m+1} \in {}^*\mathbb{R}$ there is a $u_{m+1} \in {}^*\mathbb{R}$ such that $({}^*\mathcal{R}_C, \bar{u}, u_{m+1})$ and $({}^*\mathcal{R}_{C'}, \bar{v}, v_{m+1})$ are elementarily equivalent in $\mathcal{L}_C^{m+1}$.

Proof of Claim 5: By symmetry we need only to prove (1). Given u_{m+1}, let $\Lambda(x)$ be the set all $\mathcal{L}_C^m$-formulas realized in $({}^*\mathcal{R}_C, \bar{u})$ by u_{m+1}. If $\Lambda(x)$ were countable, then v_{m+1} could be chosen easily by countable saturation. Unfortunately, $\Lambda(x)$ is not countable. To overcome this difficulty we use o-minimality. We want to choose a countable set $\Gamma(x) \subseteq \Lambda(x)$ so that for any $\sigma(x) \in \Lambda(x)$ there is a $\tau(x) \in \Gamma(x)$ such that

$$({}^*\mathcal{R}_C, \bar{u}) \models \forall x(\tau(x) \to \sigma(x)).$$

Given any $\mathcal{L}^m$-formula $\chi(x, \bar{y})$, where $\bar{y} = \{y_1, \ldots, y_n\}$. Let

$$l_\chi = \max(\{a : \exists \bar{r} \in C_e^n(a \in E_\chi^{\bar{r}} \wedge a \leqslant u_{m+1})\} \cup \{-\infty\})$$

and let

$$r_\chi = \min(\{a : \exists \bar{r} \in C_e^n (a \in E_\chi^{\bar{r}} \wedge a \geqslant u_{m+1})\} \cup \{+\infty\}).$$

Note that (1) max and min above exist because the sets are hyperfinite, (2) $l_\chi \leqslant u_{m+1} \leqslant r_\chi$, (3) if $l_\chi < u_{m+1} < r_\chi$, then for any $\bar{r} \in C_e^n$ the formula $\chi(x, \bar{r})$ has a constant truth value in (l_χ, r_χ), (4) l_χ and r_χ are definable from some $\bar{r} \in C_e^n$.

Case 1: $u_{m+1} = l_\chi$ or $u_{m+1} = r_\chi$ for some $\mathcal{L}^m$-formula χ. Then u_{m+1} is definable by some $\mathcal{L}_C^m$-formula $\theta(x)$. Let $\Gamma(x) = \{\theta(x)\}$.

Case 2: $l_\chi < u_{m+1} < r_\chi$ for any $\mathcal{L}^m$-formula χ. Let $\theta_\chi(x)$ be an $\mathcal{L}_C^m$-formula saying that $l_\chi < x < r_\chi$ and let

$$\Gamma(x) = \{\theta_\chi : \chi \text{ is an } \mathcal{L}^m\text{-formula.}\}.$$

Then $\Gamma(x)$ is countable because there are only countably many $\mathcal{L}^m$-formulas. Obviously $\Gamma(x)$ is finitely realizable in $({}^*\mathcal{R}_{C'}, \bar{v})$. By countable saturation $\Gamma(x)$ is realized in $({}^*\mathcal{R}_{C'}, \bar{v})$ by some $v_{m+1} \in {}^*\mathbb{R}$.

It is easy now to check that $({}^*\mathcal{R}_C, \bar{u}, u_{m+1})$ and $({}^*\mathcal{R}_{C'}, \bar{v}, v_{m+1})$ are elementarily equivalent in $\mathcal{L}_C^{m+1}$. $\square$(Claim 5)

Note that in Claim 5 $u_{m+1} \in C_e$ iff $v_{m+1} = I(u_{m+1}) \in C_{e'}$.

Claim 6 Suppose $({}^*\mathcal{R}_C, \bar{u})$ and $({}^*\mathcal{R}_{C'}, \bar{v})$ are elementarily equivalent. Suppose $\psi(\bar{x})$ is a quantifier-free $\mathcal{L} \cup \mathcal{L}'$-formula. Then

$${}^*\mathcal{R}_e \models \psi(\bar{u}) \text{ iff } {}^*\mathcal{R}_{e'} \models \psi(\bar{v}).$$

Proof of Claim 6: It suffices to assume that ψ is an atomic formula.

Case 1: ψ contains no symbol from $\mathcal{L}'$. Then it is trivial by the elementary equivalence.

Case 2: ψ contains symbols from $\mathcal{L}'$. Then $\psi =: X_i(\bar{\tau}(\bar{x}))$ for an $X_i \in \mathcal{L}'$ and an n_i-tuple of $\mathcal{L}$-terms $\bar{\tau}$. Note that $\mathcal{L}'$ contains no function symbols. If ${}^*\mathcal{R}_e \models \psi(\bar{u})$, then $\bar{\tau}(\bar{u}) \in e(X_i)$. Hence $\bar{\tau}(\bar{u}) \in C_e^{n_i}$. Since $({}^*\mathcal{R}_C, \bar{u})$ and $({}^*\mathcal{R}_{C'}, \bar{v})$ are elementarily equivalent, then $I(\bar{\tau}(\bar{u})) = \bar{\tau}(\bar{v})$. Note that I is an isomorphism from the submodel of ${}^*\mathcal{R}_e \restriction \mathcal{L}' \cup \{<\}$ generated by C_e, to the submodel of ${}^*\mathcal{R}_{e'} \restriction \mathcal{L}' \cup \{<\}$ generated by $C_{e'}$. Then $I(\bar{\tau}(\bar{u})) = \bar{\tau}(\bar{v}) \in e'(X_i)$. Hence ${}^*\mathcal{R}_{e'} \models \psi(\bar{v})$. $\square$(Claim 6)

Claim 7 ${}^*\mathcal{R}_e \models \phi$ iff ${}^*\mathcal{R}_{e'} \models \phi$.

Proof of Claim 7: Without loss of generality we assume that ϕ has the form $\forall \bar{x}_n \exists \bar{x}_{n-1} \ldots \psi(\bar{x}_n, \ldots, \bar{x}_1)$, where ψ is a quantifier-free $\mathcal{L} \cup \mathcal{L}'$-formula. Suppose ${}^*\mathcal{R}_e \models \phi$. By a back-and-forth argument using Claim 5 we are able to find $\bar{u}_n, \ldots, \bar{u}_1$ and $\bar{v}_n, \ldots, \bar{v}_1$ such that $({}^*\mathcal{R}_C, \bar{u}_n, \ldots, \bar{u}_1)$ is elementarily equivalent to $({}^*\mathcal{R}_{C'}, \bar{v}_n, \ldots, \bar{v}_1)$, and

$$({}^*\mathcal{R}_e, \bar{u}_n, \ldots, \bar{u}_1) \models \psi(\bar{u}_n, \ldots, \bar{u}_1).$$

By Claim 6 we have

$$({}^*\mathcal{R}_{e'}, \bar{v}_n, \ldots, \bar{v}_1) \models \psi(\bar{v}_n, \ldots, \bar{v}_1).$$

This shows that for any $\bar{v}_n$ there exists $\bar{v}_{n-1}$ such that for any $\bar{v}_{n-2}$ there exists $\bar{v}_{n-3}$...... such that

$$({}^*\mathcal{R}_{e'}, \bar{v}_n, \ldots, \bar{v}_1) \models \psi(\bar{v}_n, \ldots, \bar{v}_1).$$

So it is clear that ${}^*\mathcal{R}_{e'} \models \phi$. □

2.4. EXERCISES

Exercise 2.13 *Let $\mathcal{F} = (F; +, \cdot, <, 0, 1)$ be an ordered field. An upper bounded initial segment I of F is called a regular gap iff it has no least upper bound and for any positive $\epsilon \in F$ there exists an $r \in I$ such that $r + \epsilon \notin I$. $\mathcal{F}$ is called Scott complete iff $\mathcal{F}$ has no regular gap. Show the following:*

(1) Suppose $\mathcal{F}$ is a subfield of $\mathcal{F}'$. If an initial segment I in F has a least upper bound in $F' \setminus F$, then I is a regular gap in $\mathcal{F}$.

(2) If $\mathcal{V}$ satisfies IP, then the hyperreal field in $\mathcal{V}$ is not Scott complete.
Hint: *Write a set of formulas $\Gamma(X)$ to express that X is a regular gap.*

Exercise 2.14 *Show that IP implies the existence of an external set $S \subseteq {}^*\mathbb{N}$ such that $S \cap \{0, 1, \ldots, H\}$ is internal for every $H \in {}^*\mathbb{N}$. This means that the nonstandard universe satisfying IP can't be classless.*

Exercise 2.15 *Show that IP implies the existence of a bijection f between two *infinite sets A and B such that for any *finite sets $a \subseteq A$ and $b \subseteq B$ the restrictions $f \upharpoonright a$ and $f^{-1} \upharpoonright b$ are internal.*

***Exercise 2.16** *Let IP_0 be IP restricted only on finite languages in Definition 2.1. Prove that IP is equivalent to IP_0 plus countable saturation. See [14] for help.*

****Exercise 2.17** *Find a nonstandard universe $\mathcal{V}$ satisfying IP such that*

$$cf({}^*\mathbb{N}) \neq coin({}^*\mathbb{N} \setminus \mathbb{N}),$$

where $cf({}^*\mathbb{N})$ is the smallest cardinality of some set $S \subseteq {}^*\mathbb{N}$ cofinal in ${}^*\mathbb{N}$ and $coin({}^*\mathbb{N} \setminus \mathbb{N})$ is the smallest cardinality of some set $S \subseteq {}^*\mathbb{N} \setminus \mathbb{N}$ coinitial in ${}^*\mathbb{N} \setminus \mathbb{N}$. See [13] for help.

Exercise 2.18 *(M. Benedikt et al. [4]) Let $\mathcal{L}$ be any language including $<$ and $\mathcal{L}'$ be same as in §2.3. Let $\mathfrak{A}$ be any infinite totally-ordered $\mathcal{L}$-structure (not necessarily o-minimal). By an $\mathcal{L}'$-bounded quantifier sentence we mean an $\mathcal{L} \cup \mathcal{L}'$-sentence built up from atomic formulas in $\mathcal{L} \cup \mathcal{L}'$ via the usual* *logical connectives and the quantifications* $\forall \bar{x} \in X_i$ *and* $\exists \bar{x} \in X_j$ *for some* $X_i, X_j \in \mathcal{L}'$.

Let ϕ *be an* $\mathcal{L}'$*-bounded quantifier sentence. Show that if* ϕ *is order-invariant in* $\mathfrak{A}$*, then there is an* $\mathcal{L}' \cup \{<\}$*-sentence* ψ *such that* ϕ *and* ψ *are equivalent over* $\mathfrak{A}$.

3. The Special Model Axiom and Full Saturation

Given a language $\mathcal{L}$, an $\mathcal{L}$-structure $\mathfrak{A}$ is called a special model if there is a sequence $\langle \mathfrak{A}_\alpha : \alpha < card(\mathfrak{A})\rangle$ of $\mathcal{L}$-structures such that

(1) for any $\alpha < \beta < card(\mathfrak{A})$, $\mathfrak{A}_\alpha$ is an elementary submodel of $\mathfrak{A}_\beta$,

(2) $\mathfrak{A} = \bigcup_{\alpha < card(\mathfrak{A})} \mathfrak{A}_\alpha$, and

(3) for any $\alpha < card(\mathfrak{A})$, $\mathfrak{A}_{\alpha+1}$ is $(card(\alpha))^+$-saturated.

The sequence $\langle \mathfrak{A}_\alpha : \alpha < card(\mathfrak{A})\rangle$ is called a specializing chain for $\mathfrak{A}$.

Definition 3.1 *A nonstandard universe* $\mathcal{V}$ *is said to satisfy the* special model axiom *iff any internally presented structure of some countable language is a special model.*

We will write *SMA* for the special model axiom.

Proposition 3.2 *(D. Ross [27]) Suppose* $\mathcal{V}$ *satisfies SMA. Then every infinite internal set in* $\mathcal{V}$ *has cardinality* $\Xi_{\mathcal{V}}$.

Proof: It suffices to show that any two infinite internal sets have the same cardinality. Given two infinite internal sets C and D, we want to derive a contradiction by assuming that $card(C) = \kappa < card(D) = \lambda$. We form an internally presented structure

$$\mathfrak{A} = (C \cup D; C, D),$$

where C, D are unary relations. Then one has $card(\mathfrak{A}) = \lambda$. Suppose $\langle \mathfrak{A}_\alpha : \alpha < \lambda\rangle$ is a specializing chain for $\mathfrak{A}$. Let $C_\alpha \cup D_\alpha$ be the base set of $\mathfrak{A}_\alpha$. Note that $C_\alpha \subseteq C$ for every $\alpha < \lambda$. Since $\mathfrak{A}_{\kappa+1}$ is κ^+-saturated, then $card(C_{\kappa+1}) \geqslant \kappa^+$. This contradicts the fact $C_{\kappa+1} \subseteq C$. So $card(D) \leqslant card(C)$. By symmetry we have $card(C) = card(D)$. □

Proposition 3.3 *(D. Ross [27]) If* $\mathcal{V}$ *satisfies SMA, then* $\mathcal{V}$ *satisfies IP.*

Proof: It is proved in [5] that any two elementarily equivalent special models with the same cardinality are isomorphic. □

Proposition 3.4 *For any strong limit*[5] *cardinal* κ *with* $cf(\kappa) > \aleph_0$ *there is an ultralimit* $\mathcal{V}$ *of the standard superstructure such that* $\mathcal{V}$ *satisfies SMA and* $\Xi_{\mathcal{V}} = \kappa$.

[5]A cardinal κ is called a strong limit iff for every $\lambda < \kappa$ one has $2^\lambda < \kappa$.

Consult [5] for the proof.

Definition 3.5 *A nonstandard universe $\mathcal{V}$ is* fully saturated *iff every internally presented $\mathcal{L}$-structure $\mathfrak{A}$ for some countable language $\mathcal{L}$ is a saturated model,* i.e. *$\mathfrak{A}$ is $card(\mathfrak{A})$-saturated.*

Proposition 3.6 *If $\mathcal{V}$ is fully saturated, then $\mathcal{V}$ satisfies SMA.*

Proof: A saturated model is trivially a special model. □

Proposition 3.7 *For any cardinal κ such that $\kappa > \beth_\omega$ and $\kappa^{<\kappa} = \kappa$ there exists a fully saturated nonstandard universe $\mathcal{V}$ such that $\Xi_\mathcal{V} = \kappa$.*

Consult [5] for the proof.

Next we give one application of *SMA* and one application of full saturation .

3.1. COMPACTNESS OF LOEB PROBABILITY SPACES

Given any probability space $(\Omega, \mathcal{B}, P)$, a family $\mathcal{C} \subseteq \mathcal{B}$ is called compact iff for any $\mathcal{D} \subseteq \mathcal{C}$, $\mathcal{D}$ has f.i.p., *i.e.* if every finite subfamily of $\mathcal{D}$ has non-empty intersection, then $\bigcap \mathcal{D} \neq \emptyset$. A family $\mathcal{C} \subseteq \mathcal{B}$ is called inner-regular iff for any $B \in \mathcal{B}$

$$P(B) = \sup\{P(C) : C \in \mathcal{C} \wedge C \subseteq B\}.$$

Definition 3.8 *(D. Ross [28]) A probability space $(\Omega, \mathcal{B}, P)$ is called* compact *iff there is a compact, inner-regular family $\mathcal{C} \subseteq \mathcal{B}$.*

A Radon space is an example of a compact space. Ross showed in [28] that a compact probability space could be topologized so that the resulting topological measure space is Radon. In [28] the question whether a Loeb probability space is compact, is posed. The following theorem is one of many results in [18] concerning the compactness of Loeb probability spaces.

Theorem 3.9 *(R. Jin and S. Shelah [18]) Assume CH (Continuum Hypothesis). Suppose $\mathcal{V}$ satisfies SMA and $cf(\Xi_\mathcal{V}) = \aleph_1$. Then every non-atomic Loeb probability space is compact.*

We need more notation in the proof. For any set S we write 2^S for the set of all functions from S to $\{0,1\}$. Let $2^{<\mathbb{N}}$ be the set $\bigcup_{n\in\mathbb{N}} 2^n$, where n could be viewed as a set $\{0,\ldots,n-1\}$. A set $t \subseteq 2^{<\mathbb{N}}$ is called a tree iff for any $s, s' \in 2^{<\mathbb{N}}$, $s' \subseteq s$ and $s \in t$ imply $s' \in t$. By a branch of t we mean a maximal totally ordered subset of t with order $\subseteq$. We denote T for the trees without maximal nodes and $[T]$ for all branches of T. Let's consider $2^\mathbb{N}$ as a Cantor space with the usual probability measure ν, *i.e.*

$$\nu(\{f \in 2^\mathbb{N} : f(n) = 0\}) = \frac{1}{2}$$

for every $n \in \mathbb{N}$. Note that every closed subset of $2^{\mathbb{N}}$ could be written as $[T]$ for some tree $T \subseteq 2^{<\mathbb{N}}$. By *CH* we can fix an enumeration $\{f_\gamma : \gamma < \aleph_1\}$ of $2^{\mathbb{N}}$. For each $\beta < \aleph_1$ and $m \in \mathbb{N}$ we choose a tree $T_{\beta,m}$ such that

$$[T_{\beta,m}] \cap \{f_\gamma : \gamma < \beta\} = \emptyset \text{ and } \nu([T_{\beta,m}]) > \frac{m}{m+1}.$$

This can be done because $\nu(\{f_\gamma : \gamma < \beta\}) = 0$.

Given a probability space $(\Omega, \mathcal{B}, P)$ and a sequence of measurable sets $\langle A_n : n \in \mathbb{N}\rangle$. The sequence $\langle A_n : n \in \mathbb{N}\rangle$ is called independent iff for any $m \in \mathbb{N}$ and for any $h \in 2^m$

$$P(\bigcap_{n=0}^{m-1} A_n^{h(n)}) = \prod_{n=0}^{m-1} P(A_n^{h(n)}),$$

where $A_n^0 = A_n$ and $A_n^1 = \Omega \smallsetminus A_n$.

Proof of Theorem 3.9: Given a non-atomic Loeb probability space $(\Omega, \mathcal{B}, L_\mu)$ generated by an internal probability space $(\Omega, \mathcal{A}, \mu)$, choose an independent sequence $\langle A_n : n \in \mathbb{N}\rangle$ in $\mathcal{A}$ such that $L_\mu(A_n) = \frac{1}{2}$ for each $n \in \mathbb{N}$. For any tree $T \subseteq 2^{<\mathbb{N}}$ let

$$A_T = \bigcap_{n\in\mathbb{N}} \bigcup_{h\in 2^n\cap T} \bigcap_{i=0}^{n-1} A_i^{h(i)}.$$

It is easy to check that $L_\mu(A_T) = \nu([T])$. Note that A_T is a countable intersection of internal sets. We are now ready to construct an inner-regular, compact family $\mathcal{C}$. Note that one needs only to deal with the inner-regularity for all sets in $\mathcal{A}$.

Let $\mathfrak{A}$ be the internally presented structure same as the one in the proof of Theorem 2.7, *i.e.*

$$\mathfrak{A} = (\Omega \cup \mathcal{A} \cup {}^*\mathbb{R}; \Omega, \mathcal{A}, {}^*\mathbb{R}, \in, \mu, \cap, \smallsetminus, +, \cdot, <, 0, 1).$$

By *SMA* there is a specializing chain $\langle \mathfrak{A}_\alpha : \alpha < \Xi_{\mathcal{V}}\rangle$ for $\mathfrak{A}$. Let $\{\kappa_\beta : \beta < \aleph_1\}$ be an increasing sequence of regular cardinals cofinal in $\Xi_{\mathcal{V}}$. Let $\mathfrak{B}_\beta = \bigcup_{\alpha<\kappa_\beta} \mathfrak{A}_\alpha$. Suppose the base set of $\mathfrak{B}_\beta$ is $\Omega_\beta \cup \mathcal{A}_\beta \cup \mathbb{R}_\beta$. We can choose an enumeration $\{a_\alpha : \alpha < \Xi_{\mathcal{V}}\}$ of $\mathcal{A}$ such that for every $\beta < \aleph_1$

$$\{a_\alpha : \alpha < \kappa_\beta\} \subseteq \mathcal{A}_\beta.$$

Without loss of generality we assume $A_n \in \mathcal{A}_0$ for every $n \in \mathbb{N}$. For any $\alpha < \Xi_{\mathcal{V}}$ let

$$g(\alpha) = \min\{\beta < \aleph_1 : \alpha < \kappa_\beta\}.$$

For each $a_\alpha \in \mathcal{A}$ and each $m \in \mathbb{N}$ we choose $b_{\alpha,m} \subseteq a_\alpha \cap A_{T_{g(\alpha),m}}$ such that $b_{\alpha,m} \in \mathcal{A}_{g(\alpha)+1}$ and $L_\mu(b_{\alpha,m}) = L_\mu(a_\alpha \cap A_{T_{g(\alpha),m}})$. Note that $b_{\alpha,m}$ exists by countable saturation of $\mathfrak{B}_{g(\alpha)+1}$. Now let

$$\mathcal{C} = \{b_{\alpha,m} : \alpha < \Xi_\mathcal{V} \wedge m \in \mathbb{N}\} \cup \{A_n : n \in \mathbb{N}\}.$$

Claim $\mathcal{C}$ is compact and inner-regular.

Proof of Claim: Clearly, $\mathcal{C}$ is inner-regular. Given any $\mathcal{D} \subseteq \mathcal{C}$ with f.i.p., we want to show that $\bigcap \mathcal{D} \neq \emptyset$. Without loss of generality we assume that $\mathcal{D}$ is maximal. For each $n \in \mathbb{N}$ one has either $A_n^0 \in \mathcal{D}$ or $A_n^1 \in \mathcal{D}$. So there is a function $h' \in 2^\mathbb{N}$ such that $A_n^{h'(n)} \in \mathcal{D}$ for every $n \in \mathbb{N}$. Let

$$\delta = \bigcup \{g(\alpha) : \exists m (b_{\alpha,m} \in \mathcal{D})\}.$$

Case 1: $\delta < \aleph_1$. Then $\mathcal{D} \subseteq \mathcal{A}_{\delta+1}$ and $card(\mathcal{D}) \leqslant \kappa_\delta$. Since $\mathfrak{B}_{\delta+1}$ is $(\kappa_\delta)^+$-saturated, then $\bigcap \mathcal{D} \neq \emptyset$.

Case 2: $\delta = \aleph_1$. For each $b_{\alpha,m} \in \mathcal{D}$ and for every $n \in \mathbb{N}$ we have

$$(\bigcap_{i=0}^{n-1} A_i^{h'(i)}) \cap b_{\alpha,m} \neq \emptyset.$$

But that means

$$(\bigcap_{i=0}^{n-1} A_i^{h'(i)}) \cap A_{T_{g(\alpha),m}} \neq \emptyset.$$

Note that

$$A_{T_{g(\alpha),m}} = \bigcap_{n \in \mathbb{N}} \bigcup_{h \in 2^n \cap T_{g(\alpha),m}} \bigcap_{i=0}^{n-1} A_i^{h(i)}.$$

By a careful check one can see that $h' \upharpoonright n \in T_{g(\alpha),m}$ for every $n \in \mathbb{N}$. So $h' \in [T_{g(\alpha),m}]$. But $h' = f_\gamma$ for some $\gamma < \aleph_1$ (recall that we have a fixed enumeration of $2^\mathbb{N}$). So when $g(\alpha) > \gamma$ one has $h' \notin [T_{g(\alpha),m}]$ because

$$[T_{g(\alpha),m}] \cap \{f_{\gamma'} : \gamma' < g(\alpha)\} = \emptyset.$$

This contradicts that $\delta = \aleph_1$. □

Remark In Theorem 3.9 the assumptions *CH* and $cf(\Xi_\mathcal{V}) = \aleph_1$ couldn't be eliminated. If one replaces $cf(\Xi_\mathcal{V}) = \aleph_1$ by $cf(\Xi_\mathcal{V}) = (2^{\aleph_0})^+$ (with or without *CH*), then the result will be just the opposite. If one replaces *CH* by $\neg$*CH* (we can even weaken the condition $cf(\Xi_\mathcal{V}) = \aleph_1$ to $cf(\Xi_\mathcal{V}) = \kappa$ for any uncountable regular $\kappa \leqslant 2^{\aleph_0}$), then the compactness of Loeb probability spaces is undecidable in ZFC.

3.2. AUTOMORPHISMS OF LOEB MEASURE ALGEBRAS

Let $(\Omega, \mathcal{B}, L_\mu)$ be a Loeb probability space generated by the internal normalized uniform counting measure $(\Omega, \mathcal{A}, \mu)$. We use $\bar{\mathcal{B}}$ to denote the Loeb algebra, *i.e.* the Boolean algebra $\mathcal{B}$ modulo the ideal of L_μ-measure zero sets. For each element $B \in \mathcal{B}$ we denote $\bar{B} \in \bar{\mathcal{B}}$ for the equivalence class containing B. Note that each $\bar{B} \in \bar{\mathcal{B}}$ contains an internal set in $\mathcal{A}$.

Definition 3.10 *An* automorphism *of $\bar{\mathcal{B}}$ is a bijection Φ from $\bar{\mathcal{B}}$ to $\bar{\mathcal{B}}$ such that Φ is a Boolean algebra homomorphism and preserves the measure,* i.e. *for any $A, B \in \mathcal{B}$, $\Phi(\bar{A}) = \bar{B}$ implies $L_\mu(A) = L_\mu(B)$.*

Definition 3.11 *A bijection $T : \Omega \mapsto \Omega$ is called a* point-automorphism *iff both T and T^{-1} are measurable and for any $B \in \mathcal{B}$ one has $L_\mu(B) = L_\mu(T[B])$.*

It is easy to see that a point-automorphism induces, in a natural way, an automorphism of $\bar{\mathcal{B}}$.

Theorem 3.12 *(D. Ross [26]) Suppose $\mathcal{V}$ is fully saturated. Suppose $(\Omega, \mathcal{B}, L_\mu)$ is a Loeb probability space generated by an internal normalized uniform counting measure space $(\Omega, \mathcal{A}, \mu)$. Then every automorphism Φ of $\bar{\mathcal{B}}$ is induced by a point-automorphism T.*

Proof: Let $\mathcal{A} = \{A_\alpha : \alpha < \Xi_\mathcal{V}\}$. We construct two sequences $\langle B_\alpha : \alpha < \Xi_\mathcal{V}\rangle$ and $\langle C_\alpha : \alpha < \Xi_\mathcal{V}\rangle$ such that

(1) $\{B_\alpha : \alpha < \Xi_\mathcal{V}\} = \{C_\alpha : \alpha < \Xi_\mathcal{V}\} = \mathcal{A}$,

(2) for any $\alpha < \Xi_\mathcal{V}$, $\Phi(\bar{B}_\alpha) = \bar{C}_\alpha$,

(3) for any $\alpha_0 < \alpha_1 < \ldots < \alpha_{n-1} < \Xi_\mathcal{V}$ and for any $h \in 2^n$ one has

$$|\bigcap_{i=0}^{n-1} B_{\alpha_i}^{h(i)}| = |\bigcap_{i=0}^{n-1} C_{\alpha_i}^{h(i)}|,$$

where $A^0 = A$, $A^1 = \Omega \smallsetminus A$ and $|\cdot|$ means internal cardinality.

Claim The theorem follows from the construction.

Proof of Claim: For any $x \in \Omega$ there is an α such that $\{x\} = B_\alpha$ by (1). It is easy to see by (3) that C_α is also a singleton $\{y\}$ for some $y \in \Omega$. Let $T(x) = y$. Then it is easy to check again by (3) that T is a well-defined bijection. Also it is not hard to check that $T[B_\alpha] = C_\alpha$. So one has $|A| = |T[A]|$ for every $A \in \mathcal{A}$. This implies that T and T^{-1} are measurable and preserve the measure. So T is a point-automorphism. By (2) and (3) one can easily see that Φ is induced by T. $\square$(Claim)

We now construct B_α and C_α by induction. Let

$$\mathfrak{A} = (\Omega \cup \mathcal{A} \cup {}^*\mathbb{R}; \Omega, \mathcal{A}, {}^*\mathbb{R}, \in, \mu, \cap, \smallsetminus, +, \cdot, <, 0, 1)$$

be the internally presented structure same as in the proof of Theorem 2.7. Suppose we have found $\{B_\beta : \beta < \alpha\}$ and $\{C_\beta : \beta < \alpha\}$ such that (2) and (3) are true up to stage α.

Case 1: α is even. We pick B_α first. Let

$$\gamma = \min\{\delta : A_\delta \notin \{B_\beta : \beta < \alpha\}\}$$

and let $B_\alpha = A_\gamma$. This step guarantees (1). Let $\Phi(\bar{B}_\alpha) = \bar{A}$ for some $A \in \mathcal{A}$. We define a set of formulas $\Gamma_\alpha(x)$ with only one free variable x, which expresses that x is a candidate for C_α. The set $\Gamma_\alpha(x)$ contains exactly the following:

(a) $\mathcal{A}(x)$, *i.e.* x is an internal subset of Ω,

(b) $\mu(x\Delta A) < \frac{1}{m}$ for every $m \in \mathbb{N}$, *i.e.* the symmetric difference of x and A will have Loeb measure zero,

(c)

$$\mu((\bigcap_{i=0}^{n-1} B_{\alpha_i}^{h(i)}) \cap B_\alpha^j) = \mu((\bigcap_{i=0}^{n-1} C_{\alpha_i}^{h(i)}) \cap x^j)$$

for any $\alpha_0 < \alpha_1 < \ldots < \alpha_{n-1} < \alpha$, for any $h \in 2^n$ and for any $j = 0, 1$. Note that for any internal sets $A, B \subseteq \Omega$ one has $\mu(A) = \mu(B)$ iff $|A| = |B|$.

Since $card(\Gamma_\alpha(x)) = card(\alpha) < \Xi_\mathcal{V}$ and $\Gamma_\alpha(x)$ is clearly finitely realizable, then, by full saturation, $\Gamma_\alpha(x)$ is realized by some $C \in \mathcal{A}$. Let $C_\alpha = C$. Clearly, (2) and (3) are true up to stage $\alpha + 1$.

Case 2: α is odd. We pick C_α first and then B_α by symmetry.

Finally, (1) is true because of the way we choose B_α when α is even, and C_α when α is odd. □

3.3. EXERCISES

Exercise 3.13 *(D. Ross [27]) Show that SMA implies that*

$$cf(^*\mathbb{N}) = coin(^*\mathbb{N} \smallsetminus \mathbb{N}) = cf(\Xi_\mathcal{V}).$$

By comparing with Exercise 2.17 this exercise witnesses that IP does not imply SMA.

Exercise 3.14 *Let SMA_0 be SMA restricted only on finite languages in Definition 3.1. Show that SMA is equivalent to SMA_0 plus countable saturation.*

Exercise 3.15 *Assuming CH. Suppose $\mathcal{V}$ is an ultrapower of the standard superstructure modulo an ultrafilter on a countable set. Show that every Loeb probability space generated by an internal normalized uniform counting measure on a hyperfinite set in $\mathcal{V}$ is compact.*

***Exercise 3.16** *Show that Theorem 3.12 is still true if $\mathcal{V}$ satisfies SMA instead of full saturation (see [16] for hint).*

Exercise 3.17 *(D. Ross [26]) Let $\mathcal{V}$ be fully saturated. Let $(\Omega, \mathcal{B}, L_\mu)$ be a Loeb probability space generated by an internal normalized uniform counting measure on a hyperfinite set Ω. Let $\bar{\mathcal{B}}$ be the Loeb algebra. Show the following:*

(1) There exist automorphisms Φ of $\bar{\mathcal{B}}$, which are not induced by any internal point-automorphisms,

(2) Let $\bar{\mathcal{C}} \subseteq \bar{\mathcal{B}}$ be a subalgebra such that $card(\bar{\mathcal{C}}) < \Xi_{\mathcal{V}}$. Then for any automorphism Φ of $\bar{\mathcal{C}}$ there exists an internal point-automorphism T such that Φ is induced by T.

***Exercise 3.18** *If $\mathcal{V}$ satisfies SMA but is not fully saturated, then for any Loeb probability space $(\Omega, \mathcal{B}, L_\mu)$ as in Exercise 3.17 there exists a subalgebra $\bar{\mathcal{C}} \subseteq \bar{\mathcal{B}}$ with $card(\bar{\mathcal{C}}) < \Xi_{\mathcal{V}}$ and there exists an automorphism Φ of $\bar{\mathcal{C}}$ such that Φ is not induced by any internal point-automorphism. (This exercise and Exercise 3.16 are related, in fact, to a question posed in [27]. This exercise witnesses that SMA does not imply full saturation.)* Hint: *Show first that $\Xi_{\mathcal{V}}$ is a singular strong limit cardinal. Then construct a subalgebra of cardinality $cf(\Xi_{\mathcal{V}})$ together with an automorphism by a diagonal method so that the automorphism avoids to be induced by any internal point-automorphisms (see [16] for more help).*

4. The λ-Bolzano-Weierstrass Property

Let $\mathcal{F} = (F; +, \cdot, <, 0, 1)$ be an ordered field. Let λ be an uncountable regular cardinal. By a bounded λ-sequence in $\mathcal{F}$ we mean a sequence $\langle a_\alpha : \alpha < \lambda \rangle$ in $\mathcal{F}$ such that $\{a_\alpha : \alpha < \lambda\} \subseteq [-r, r]$ for some positive $r \in F$. A λ-sequence $\langle a_\alpha : \alpha < \lambda \rangle$ in $\mathcal{F}$ is said to converge in $\mathcal{F}$ iff there is an $r \in F$ such that for any positive $\epsilon \in F$ there is an $\beta < \lambda$ such that $\{a_\alpha : \beta < \alpha < \lambda\} \subseteq [r - \epsilon, r + \epsilon]$.

Definition 4.1 *Given a nonstandard universe $\mathcal{V}$ and let ${}^*\mathcal{R}$ be the hyperreal field in $\mathcal{V}$. Let λ be an uncountable regular cardinal less than or equal to $card({}^*\mathbb{R})$. ${}^*\mathcal{R}$ satisfies the* λ-Bolzano-Weierstrass property *iff every bounded λ-sequence in ${}^*\mathcal{R}$ has a convergent λ-subsequence in ${}^*\mathcal{R}$. $\mathcal{V}$ satisfies the λ-Bolzano-Weierstrass property iff ${}^*\mathcal{R}$ does.*

Clearly, the λ-Bolzano-Weierstrass property is a natural generalization of the Bolzano-Weierstrass property for the standard real field.

Proposition 4.2 *(H. J. Keisler and J. H. Schmerl [22]) There exist countably saturated nonstandard universes satisfying the λ-Bolzano-Weierstrass property for $\lambda = (2^{\aleph_0})^+$.*[6]

Definition 4.3 *A nonstandard universe $\mathcal{V}$ is called* λ-Archimedean *iff $card({}^*\mathbb{N}) = \lambda$ and $card(\{0, 1, \ldots, H\}) < \lambda$ for every $H \in {}^*\mathbb{N}$.*

Proposition 4.4 *(J. Cowles and R. LaGrange [6]) If $\mathcal{V}$ satisfies the λ-Bolzano-Weierstrass property, then $\mathcal{V}$ is λ-Archimedean.*

Proof: Suppose $\mathcal{V}$ is not λ-Archimedean.

Case 1: $card({}^*\mathbb{N}) < \lambda$. Since the set of all *rational numbers has cardinality $card({}^*\mathbb{N}) < \lambda$, then there are no λ-convergent sequences in ${}^*\mathbb{R}$. But there are λ-sequences in ${}^*\mathcal{R}$.

Case 2: $card({}^*\mathbb{N}) \geqslant \lambda$. Because $\mathcal{V}$ is not λ-Archimedean, there exists an $H \in {}^*\mathbb{N}$ and an $S \subseteq \{0, 1, \ldots, H\}$ such that $card(S) = \lambda$. Clearly, the set S could be ordered as a λ-sequence. That sequence is bounded and has no convergent λ-subsequence because it is discrete. $\Box$

Remark Suppose $\mathcal{U}$ is a regular ultrafilter on any κ. Then the ultrapower of the standard superstructure modulo $\mathcal{U}$ could never be λ-Archimedean for any λ. So the existence of an λ-Archimedean ultrapower would imply the existence of some non-regular ultrafilters, which may imply the consistency of some large cardinals. By assuming the consistency of a measurable cardinal it is consistent that there exists a λ-Archimedean ultrapower for some λ [19]. It is still open whether such kind of ultrapowers could satisfy the λ-Bolzano-Weierstrass property.

Proposition 4.5 *(J. Cowles and R. LaGrange [6]) If $\mathcal{V}$ satisfies the λ-Bolzano-Weierstrass property, then the hyperreal field in $\mathcal{V}$ is Scott complete. (See Exercise 2.13 for the definition of Scott completeness.)*

Proof: Suppose ${}^*\mathcal{R}$ is not Scott complete. Then there is an upper bounded regular gap I in ${}^*\mathbb{R}$. It is easy to see that the cofinality of I is same as the cofinality of ${}^*\mathbb{N}$. But $\mathcal{V}$ is λ-Archimedean. So there exists an increasing λ-sequence cofinal in I. Clearly, the sequence is bounded and has no convergent λ-subsequence. $\Box$

Remarks: (1) Since *IP* implies the hyperreal field is not Scott complete, the λ-Bolzano-Weierstrass property is inconsistent with *IP*. (2) The reader who is interested in doing research on this subject should consult the papers [6], [19], [20], [22], [29], [30], [32], [33] and [34].

[6]The exact result in [22] is that: Suppose κ and λ are uncountable cardinals, κ is regular, $\kappa < \lambda$, and $\eta^\delta < \lambda$ whenever $\delta < \kappa$ and $\eta < \lambda$. Then there exists a κ-saturated nonstandard universe satisfying the λ-Bolzano-Weierstrass property.

Exercise 4.6 *(1) Let $\mathcal{F}$ be an ordered field. Show that the unit interval $[0,1]$ in $\mathcal{F}$ is compact iff $\mathcal{F}$ is isomorphic to the standard real field.*

(2) Suppose $\aleph_1 = card({}^*\mathbb{N})$ in $\mathcal{V}$ ($\mathcal{V}$ may not be countably saturated). Show that $\mathcal{V}$ satisfies the $\aleph_1$-Bolzano-Weierstrass property iff the unit interval of the hyperreal field is Lindelöf.

References

1. Barwise, K. J., (1973) Back and forth through infinitary logic, *Studies in Model Theory*, M. Morley ed., MAA Studies **8**, pp. 5–34.
2. Benedikt, M., (1993) *Nonstandard Analysis and Special Ultrafilters*, Ph.D. Thesis, University of Wisconsin.
3. Benedikt, M., (to appear) Ultrafilters which extend measures, *The Journal of Symbolic Logic.*
4. Benedikt, M., Dong, G., Libkin, L., and Wong, L., (to appear) Model theoretic results on constraint queries, *The Journal of the Association for Computing machinery.*
5. Chang, C.C. and Keisler, H.J., (1990) *Model Theory.* 3rd ed., North-Holland, Amsterdam, (2nd ed., 1977).
6. Cowles, J. and LaGrange, R., (1983) Generalized Archimedean fields, *Notre Dame Journal of Formal Logic*, **24**, pp. 133–140.
7. Henson, C.W., (1974) The isomorphism property in nonstandard analysis and its use in the theory of Banach space, *The Journal of Symbolic Logic* **39**, pp. 717–731.
8. Henson, C.W., (1975) When do two Banach spaces have isometrically isomorphic nonstandard hulls?, *Israel Journal of Mathematics* **22**, pp. 57–67.
9. Henson, C.W., (1979) Unbounded Loeb measure, *Proceedings of the American Mathematical Society* **74**, pp. 143–150.
10. Henson, C.W., (1997) Foundations of nonstandard analysis: a gentle introduction to nonstandard extensions, *this volume.*
11. Henson, C.W. and Keisler, H.J., (1986) On the strength of nonstandard analysis, *The Journal of Symbolic Logic* **51**, pp. 377–386.
12. Henson, C.W. and Wattenberg, F., (1981) Egoroff's theorem and the distribution of standard points in a nonstandard model, *Proceedings of American Mathematical Society* **81**, pp. 455–461.
13. Jin, R., (1992) The isomorphism property versus the special model axiom, *The Journal of Symbolic Logic* **57**, pp. 975–987.
14. Jin, R., (1992) A theorem on the isomorphism property, *The Journal of Symbolic Logic* **57**, pp. 1011–1017.
15. Jin, R., (to appear) Type two cuts, bad cuts and very bad cuts, *The Journal of Symbolic Logic.*
16. Jin, R., (in preparation) Distinguishing strong saturation properties in nonstandard analysis.
17. Jin, R. and Shelah, S., (1994) The strength of the isomorphism property, *The Journal of Symbolic Logic* **59**, pp. 292–301.
18. Jin, R. and Shelah, S., (submitted) Compactness of Loeb spaces, *The Journal of Symbolic Logic.*
19. Jin, R. and Shelah, S., (in preparation) Possible size of an ultrapower of ω.
20. Keisler, H.J., (1974) Models with tree structures, *Proceedings of Symposia in Pure Mathematics* **25**, American Mathematical Society, Providence, Rhode Island, pp. 331–348.

21. Keisler, H.J. and Leth, S.C., (1991) Meager sets on the hyperfinite time line, *The Journal of Symbolic Logic* **56**, pp. 71–102.
22. Keisler, H.J. and Schmerl, J.H., (1991) Making the hyperreal line both saturated and complete, *The Journal of Symbolic Logic* **56**, pp. 1016–1025.
23. Loeb, P., (1975) Conversion from nonstandard to standard measure spaces and applications in probability theory, *Transactions of American Mathematical Society* **211**, pp. 113–122.
24. Luxemburg, W.A.J., (1969) A general theory of monads, *Applications of Model Theory to Algebra, Analysis and Probability*, (W.A.J. Luxemburg, editor), Holt, Rinehart and Winston, New York, pp. 18–86.
25. Ross, D., (1983) *Measurable Transformations in Saturated Models of Analysis*, Ph.D. Thesis, University of Wisconsin.
26. Ross, D., (1987) Automorphisms of the Loeb algebra, *Fundamenta Mathematicae* **128**, pp. 29–36.
27. Ross, D., (1990) The special model axiom in nonstandard analysis, *The Journal of Symbolic Logic* **55**, pp. 1233–1242.
28. Ross, D., (1992) Compact measures have Loeb preimages, *Proceedings of American Mathematical Society* **115**, pp. 365–370.
29. Schmerl, J.H., (1985) Peano arithmetic and a question of Sikorski on ordered field, *Israel Journal of Mathematics* **50**, pp. 145–159.
30. Scott, D., (1969) On completing ordered fields, *Applications of Model Theory to Algebra, Analysis and Probability*, (W.A.J. Luxemburg, editor), Holt, Rinehart and Winston, New York, pp. 274–278.
31. Shelah, S., (1971) Every two elementarily equivalent models have isomorphic ultrapowers, *Israel Journal of Mathematics* **10**, pp. 224–233.
32. Shelah, S., (1978) Models with second order properties, II: Trees with no undefined branches, *Annals of Mathematical Logic* **14**, pp. 73–87.
33. Shelah, S., (1983) Models with second order properties, IV: A general method for eliminating diamonds, *Annals of Pure and Applied Logic* **25**, pp. 183–212.
34. Sikorski, R., (1948) On an ordered algebraic field, *Comptes Rendus des Séances de la Sociétés Sciences et des Lettres de Varsovie, Classe III* **41**, pp. 69–96.

INTERNAL MARTINGALES AND STOCHASTIC INTEGRATION

TOM LINDSTRØM
Department of Mathematics
University of Oslo
PO Box 1053, Blindern
N-0316 Oslo
Norway
e-mail: lindstro@math.uio.no

Introduction

These lectures are primarily intended as a bridge connecting David Ross' lectures on measure and probability theory to Jerry Keisler's and Ekkehard Kopp's lectures on stochastic differential equations and mathematical finance later in the volume. Although my task is primarily to provide you with the technical tools necessary to understand these applications, I hope to a certain extent to give you a feeling for stochastic analysis as an interesting subject in its own right.

Nonstandard stochastic analysis started twenty years ago with Robert M. Anderson's seminal paper [3] in which he gave a nonstandard construction of Brownian motion and the Itô integral. A few years later, Keisler used Anderson's results as the starting point for a deep and penetrating study of stochastic differential equations and Markov processes [13]. Nonstandard stochastic integration was extended to cover more general integrators by Hoover and Perkins [11] and Lindstrøm [15], [16]. You will find other expositions of the theory in the books by Stroyan and Bayod [30] and Albeverio, Fenstad, Høegh-Krohn, and Lindstrøm [1].

Everybody who tries to write about stochastic integration is faced with a difficult problem of generality—which class of integrators should one concentrate on? The most general choice is stochastic integration with respect to semi-martingales, and from a theoretical point of view this choice is quite attractive as the theory then becomes, in a very strong sense, the best possible. However, it also becomes quite technical, and the extra generality it provides is rarely needed for applications. An alternative is to concentrate

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 209–258.

on square integrable martingales—this theory has a pleasant geometrical flavour and avoids the heavy technical machinery of the general theory. Another popular choice is only to consider continuous martingales, but in the nonstandard setting little is gained by assuming continuity and something is definitely lost—even those who want to concentrate on the continuous case, need to prove that their martingales are, in fact, continuous, and this requires some knowledge of the discontinuous theory. I have chosen a combination of the latter two approaches; as long as I stay firmly within the nonstandard theory, I work with square integrable martingales, but when I connect the standard and the nonstandard theory, I only treat the continuous case. This may not seem very logical, but it is my experience that this approach will provide you with the tools you are most likely to need with a minimum of technical fuss. If you later realize that you need the full theory of semi-martingales, you should look it up in [11].

In addition to working with square integrable martingales, I have made one more simplifying assumption—I only consider hyperfinite probability spaces. In fact, almost everything in the paper goes through (occasionally with a small change in the phrasing) for nonstandard probability spaces in general, but I feel that hyperfinite spaces provide a very concrete setting where beginners can concentrate on the basic ideas without being overwhelmed by abstract formalism. You should be aware, however, that in some situations it is convenient and advantageous to use nonstandard probability spaces which are not hyperfinite.

Compared to the standard case, very little has been written about nonstandard stochastic integration, and you should look up the standard literature for more insight and inspiration. The books by Chung and Williams [5] and Kopp [14] are relatively short introductions. More comprehensive treatments can be found in the books by Karatzas and Shreve [12], Revuz and Yor [28], and Rogers and Williams [29]. The ultimate text on stochastic integration is Dellacherie and Meyer's multi-volume work [9], but you should probably not turn to this before you have a good grasp of the aim and purpose of stochastic integration.

Throughout the paper we shall work with an $\aleph_1$-saturated model of nonstandard analysis.

1. Hyperfinite Probability Spaces

A hyperfinite probability space is just a hyperfinite set Ω and an internal function

$$P : \Omega \to {}^*\mathbb{R}$$

such that

$$\sum_{\omega\in\Omega} P(\omega) = 1$$

We think of ω as an *event*—e.g. the outcome of a statistical experiment—and we think of $P(\omega)$ as the probability of the event ω happening. The set Ω is the collection of all possible events, and thus $\sum_{\omega\in\Omega} P(\omega) = 1$. If A is a hyperfinite subset of Ω, we shall abuse notation slightly by writing $P(A) = \sum_{\omega\in A} P(\omega)$ and think of this as the probability of the more complex event A. From this point of view, P is an internal, finitely additive measure on the the algebra $\mathcal{A}$ of all internal subsets of Ω. We shall let $L(\mathcal{A})$ and P_L be the Loeb algebra of $\mathcal{A}$ and the Loeb measure of P, respectively.

We shall often refer to an internal map $F : \Omega \to^* \mathbb{R}$ as a *random variable.* The *expectation* (or *expected value*) $E(F)$ of F is just the average

$$E(F) = \sum_{\omega\in\Omega} F(\omega)P(\omega)$$

The *variance* of F is defined as

$$Var(F) = E((F - E(F))^2)$$

and gives a measure of how much the random variable deviates from its mean.

Assume now that $\mathcal{B}$ is another internal algebra of subsets of $\mathcal{A}$ (hence $\mathcal{B}$ is necessarily a subset of $\mathcal{A}$). This algebra $\mathcal{B}$ induces a partition of Ω where the equivalence class of ω is given by

$$[\omega]_{\mathcal{B}} = \bigcap\{B \in \mathcal{B} : \omega \in B\}$$

If $F : \Omega \to^* \mathbb{R}$ is an internal map, we let the *conditional expectation of F with respect to $\mathcal{B}$* be the function $E(F|\mathcal{B}) : \Omega \to^* \mathbb{R}$ defined by

$$E(F|\mathcal{B})(\omega) = \frac{\sum_{\tilde{\omega}\in[\omega]_{\mathcal{B}}} F(\tilde{\omega})P(\tilde{\omega})}{P([\omega]_{\mathcal{B}})}$$

Hence $E(F|\mathcal{B})(\omega)$ is simply the average of F over the equivalence class of ω.

In probability theory the subalgebra $\mathcal{B}$ usually codifies the information you have available in a certain situation (e.g. at a certain time), and the conditional expectation $E(F|\mathcal{B})$ then represents the best estimate you can make of the random variable F on the basis of this information. This picture will become clearer as we proceed.

We shall be interested in random phenomena evolving in time. To model such phenomena, we introduce *hyperfinite timelines*

$$T = \{0, t_1, t_2, t_3, \ldots, t_H\}$$

where H is an infinite integer and $t_{i+1} - t_i$ is infinitesimal for all i. In most cases, all the increments $t_{i+1} - t_i$ will equal the same infinitesimal Δt, but there are situations where it is convenient to let the size of the increments vary. An *internal, stochastic process* is just an internal map

$$X : \Omega \times T \to {}^*\mathbb{R}$$

We think of $X(\omega, t)$ as the position of the process at time t given that ω is the outcome of our statistical experiment.

There is a corresponding set-up for standard probability which I review briefly for readers unfamiliar with the theory. A (standard) *probability space* $(\Omega, \mathcal{F}, P)$ is just a measure space where $P(\Omega) = 1$, and a *random variable* is a measurable map $X : \Omega \to \mathbb{R}$. The *expectation* of an integrable random variable is defined by

$$E(X) = \int X \, dP$$

and the *variance* (of a square integrable random variable) by

$$Var(X) = E((X - E(X))^2)$$

If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, the *conditional expectation* of an integrable random variable X is a $\mathcal{G}$-measurable random variable Y such that

$$\int_G Y \, dP = \int_G X \, dP$$

for all $G \in \mathcal{G}$. As in the nonstandard case we denote Y by $E(X|\mathcal{G})$. A *stochastic process* is a map $X : \Omega \times [0, \infty) \to \mathbb{R}$ such that the map $\omega \mapsto X(\omega, t)$ is measurable for each t.

As probabilistic notation is often a stumbling block to beginners, let me give you a few words of warning. The variable ω is often suppressed in formulas, and hence you will often see X or $X(t)$ where you would expect to find $X(\omega)$ or $X(\omega, t)$. In complicated expressions it is often more convenient to write the time dependence as a subscript, and thus $X(\omega, t)$ will become $X_t(\omega)$ or just X_t. In the nonstandard setting, we shall often use $\Delta X(\omega, t)$ (alias $\Delta X_t(\omega)$ alias ΔX_t) to denote the *forward increment* of a process X. It is defined by $\Delta X(\omega, t_i) = X(\omega, t_{i+1}) - X(\omega, t_i)$ for all $t_i \in T$.

In order to give readers who do not have much experience with probability theory, a better feeling for what is going on, I shall in the next two sections sketch nonstandard constructions of some of the most important and fundamental stochastic processes, *Poisson processes* and *Brownian motion*.

PROBLEMS

1.1 *Show that any internal algebra $\mathcal{B}$ of sets is closed under hyperfinite unions and intersections (i.e, if $\mathcal{F}$ is a hyperfinite subset of $\mathcal{B}$, then $\bigcup\{B \in \mathcal{F}\}$ and $\bigcap\{B \in \mathcal{F}\}$ both belong to $\mathcal{B}$).*

1.2 *Let $(\Omega, \mathcal{A}, P)$ be a hyperfinite probability space, and let $\mathcal{B}$ be an internal subalgebra of $\mathcal{A}$. Show that if $G(\omega) = E(F|\mathcal{B})(\omega)$ is the conditional expectation of an internal function $F : \Omega \to^* \mathbb{R}$, then*

$$\int_B F\,dP = \int_B G\,dP$$

for all $B \in \mathcal{B}$.

1.3 *Assume that $F : \Omega \to^* \mathbb{R}$ is an internal function defined on an internal probability space. Show that if $\int |F|^p\,dP$ is finite for some real, positive number p, then $|F|^q$ is S-integrable for all positive, real numbers $q < p$.*

2. Poisson Processes

In many situations in life, there is a small probability that things will change drastically and a large probability that they will continue the same. If you are driving your car along a highway, there is always a small probability that some kind of accident will force you to stop within the next minute, and a large probability that you will continue as before. If you are playing the same lottery every week, there is a small probability that you will get rich next week, and a large probability that you will remain as poor as you have always been.

Poisson processes are simple models of such phenomena. The basic assumption is that the probability for a drastic change in a short time interval of length t is (almost) proportional to t. Hence the probability that a change will occur in the interval $(s, s+t)$ is approximately equal to αt, where $\alpha (\in \mathbb{R})$ is the *intensity* of the process (the larger α is, the larger the probability for a change).

We shall now take a look at a nonstandard construction of a Poisson process. The construction is due to Peter Loeb and appeared in his fundamental paper [20] on the Loeb measure. We start with a hyperfinite timeline $T = \{0, t_1, t_2, t_3, \ldots, t_H\}$ where $t_{i+1} - t_i = \Delta t$ for all i, and where t_H is infinitely large. The sample space Ω consists of all internal functions $\omega : T \to \{0, 1\}$. Note that each ω represents a possible scenario—if ω takes the value 1 at the points $s_1, s_2, s_3 \ldots \in T$, we interpret this to mean that changes occur in the intervals $(s_1, s_1 + \Delta t]$, $(s_2, s_2 + \Delta t]$, $(s_3, s_3 + \Delta t]$ etc.

We now define an internal process

$$N : \Omega \times T \to^* \mathbb{R}$$

by

$$N(\omega, t) = \sum_{s<t} \omega(s),$$

(here and in all similar sums we are summing over elements s of the time-line T) and observe that $N(\omega, t)$ counts the number of changes that have occurred up to time t (just as your insurance company keeps track of *how many* accidents you have had with your car, the process N counts the *number* of changes).

We next have to introduce a probability measure P on Ω. Intuitively, the probability that the process will make a jump in the interval from t to $t + \Delta t$ is $\alpha \Delta t$ and the probability that it will *not* make a jump is $1 - \alpha \Delta t$. Hence it is natural to define

$$P(\omega) = (\alpha \Delta t)^n (1 - \alpha \Delta t)^{H+1-n},$$

where n is the number of jumps that $t \mapsto N(\omega, t)$ makes.

Assume that we have observed the process up to time t. We then know the value of $\omega(s)$ for all $s < t$, but we know nothing about $\omega(s)$ for $s \geq t$. Put differently, we can identify ω up to the equivalence relation $\sim_t$ defined by

$$\omega \sim_t \tilde{\omega} \Leftrightarrow \forall s < t(\,\omega(s) = \tilde{\omega}(s)\,)$$

If we let $[\omega]_t = \{\tilde{\omega} : \tilde{\omega} \sim_t \omega\}$ be the equivalence class of ω, and $\mathcal{B}_t$ be the internal algebra generated by these equivalence classes, then $\mathcal{B}_t$ in an obvious way codifies the information we have available at time t. It easy to check if ω makes k jumps before time t, then

$$P([\omega]_t) = (\alpha \Delta t)^k (1 - \alpha \Delta t)^{\frac{t}{\Delta t} - k}$$

We can use this formula to compute the probability that the process has made exactly k jumps before time t: Since there are $\binom{t/\Delta t}{k}$ subsets of $\{0, \Delta t, 2\Delta t, \ldots, t - \Delta t\}$ of cardinality k, this probability must be

$$\binom{t/\Delta t}{k} (\alpha \Delta t)^k (1 - \alpha \Delta t)^{\frac{t}{\Delta t} - k}$$

(note that we are really using transfer of finite combinatorics). If we assume that t is not infinitesimal, then $t/\Delta t$ is infinite, and hence

$$\binom{t/\Delta t}{k} \Delta t^k \approx \frac{t^k}{k!}$$

and

$$(1-\alpha\Delta t)^{\frac{t}{\Delta t}-k} = e^{-\alpha t}$$

Hence the probability of making exactly k jumps before time t is approximately

$$(\alpha t)^k e^{-\alpha t}/k!$$

So far our process $N : \Omega \times T \to^* \mathbb{R}$ is a nonstandard object living on a nonstandard measure space $(\Omega, \mathcal{A}, P)$ and a nonstandard timeline T. To get a Poisson process in the usual sense, we must turn N into a standard process $n : \Omega \times [0,\infty) \to \mathbb{N}_0$ living on a standard probability space. Before we do this, let me remind you exactly what a Poisson process is. (Don't despair if you don't understand all the details of this definition—the parts that will be needed in the sequel, will be explained as we go along.) We first define a *Poisson distribution*: A measure μ on the nonnegative integers $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ is Poisson distributed with index γ if

$$\mu\{k\} = \gamma^k e^{-\gamma}/k!$$

for all $k \in \mathbb{N}_0$. Note that our calculations above basically says that $N(\omega, t)$ is Poisson distributed with index αt.

Definition 2.1 *A process $n : \Omega \times [0,\infty) \to \mathbb{R}$ is called a* Poisson process *of intensity α if*
(i) $n(0) = 0$
(ii) n has independent increments, i.e. if $s_1 < t_1 \le s_2 < t_2 \le \ldots \le s_k < t_k$, then the random variables $n(\omega, t_1) - n(\omega, s_1)$, $n(\omega, t_2) - n(\omega, s_2), \ldots$, $n(\omega, t_k) - n(\omega, s_k)$ are stochastically independent
(iii) If $t > s$, the increment $n(t) - n(s)$ is Poisson distributed with index $\alpha(t-s)$
(iv) Almost all paths $t \mapsto n(\omega, t)$ are right continuous

To turn our nonstandard process N into a standard Poisson process, we first apply Loeb's construction to get a standard probability space $(\Omega, L(\mathcal{A}), P_L)$. We then define the process $\tilde{n} : \Omega \times [0,\infty) \to \mathbb{N}_0$ by

$$\tilde{n}(\omega, t) = N(\omega, t^-)$$

where t^- is the element in T to the immediate left of t (note that t itself need not belong to T). The process $\tilde{n}$ satisfies all the conditions of a Poisson process except that it fails to be right continuous. To remedy this flaw, we simply take right limits and define

$$n(\omega, t) = \lim_{s \downarrow t} \tilde{n}(\omega, s)$$

It is a good exercise to show that n really is a Poisson process.

PROBLEMS

2.1 *Above we computed the probability*

$$\binom{t/\Delta t}{k}(\alpha\Delta t)^k(1-\alpha\Delta t)^{t/\Delta t-k}$$

with a vague reference to the transfer principle. Work out the details.

2.2 *Above I claimed that*

$$\binom{t/\Delta t}{k}\Delta t^k \approx \frac{t^k}{k!} \quad \textit{and} \quad (1-\alpha\Delta t)^{t/\Delta t-k} \approx e^{-\alpha t}$$

Work out the details.

2.3 *Prove that n is a Poisson process.*

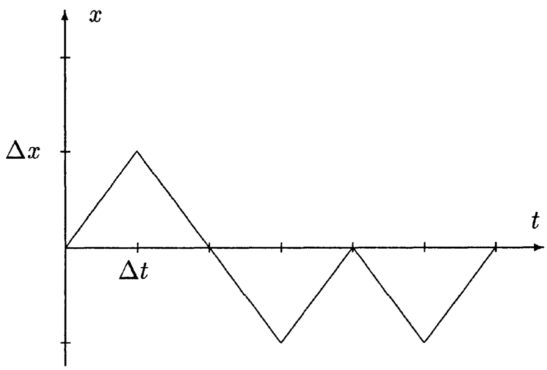
3. Brownian Motion

We now turn to the construction of another fundamental process in probability theory—*Brownian motion.* The name goes back to the Scottish botanist Robert Brown (1773-1858) who in 1827 observed that small particles dispersed in a liquid perform strange and irregular movements. Although others had made the same observation before him, Brown was the first to undertake a systematic study of these movements and to query their cause. Throughout the 19th century the question remained a topic of much discussion, and it was only in 1905-06 that Albert Einstein and the Polish physicist Marian Smoluchowski (1872-1917) independently showed that Brown's movement was caused by collisions with the molecules of the fluid. Einstein's and Smoluchowski's results broke down the final resistance to the molecular theory of matter, and Einstein used his theory to give the first estimate of Avogadro's number.

In 1923 Norbert Wiener used Lebesgue's new-fangled theory of measures to refine Einstein's and Smoluchowski's approach. Although physical Brownian motion today is well understood and no longer a research topic in its own right, Wiener's model has become a cornerstone of modern probability theory. There are several reasons for this—one is that the same laws which made Wiener's construction a natural model for physical Brownian motion, also makes it a natural model in many other contexts, another is that Wiener's model can easily be used as a starting point for more complicated constructions.

Although Brown's particles move in a three-dimensional liquid, we shall first build a one-dimensional model (instead of thinking of a one-dimensional fluid, you should think of this as modelling one component at a time—see Section 13). As above we shall work with a hyperfinite timeline $T = \{0, t_1, t_2, t_3, \ldots, t_H\}$ where $t_{i+1} - t_i = \Delta t$ for all i, and where t_H is infinitely large. We also fix a *space increment* $\Delta x = \sqrt{\Delta t}$.

Let us first give an informal description of *Anderson's random walk* B which shall be our nonstandard model of Brownian motion (see figure). We start at the origin at time 0 and toss a fair coin. If we get 'heads', we move up a distance Δx, if we get 'tails' we move down a distance Δx. At time Δt we toss the coin again and go up Δx if we get 'heads' and down Δx if we get 'tails'. At time $2\Delta t$ we repeat the experiment again, and so on. The figure shows a path starting with the sequence heads, tails, tails, heads, tails, heads.



To formalize this construction, we introduce the sample space

$$\Omega = a = \{\omega : T \to \{-1, 1\} : \omega \text{ is internal } \}$$

Each ω represents a sequence of coin tosses, and the interpretation is that $\omega(t)$ is 1 or -1 according to whether the coin toss at time t results in 'heads' or 'tails'. The process

$$B : \Omega \times T \to^* \mathbb{R}$$

defined by

$$B(\omega, t) = \sum_{s<t} \omega(s)\sqrt{\Delta t}$$

tells us the position at time t given that ω codes the sequence of coin tosses.

It often helps the intuition to think of B as a gambling game; if at each instant $t \in T$ you put an amount $\sqrt{\Delta t}$ on getting 'heads' next time, then

$B(t)$ records what you have won (or lost) up to time t. Note that B is a *fair* game—in average you neither win nor lose.

Just as for our nonstandard Poisson process, we may define an equivalence relation $\sim_t$ by

$$\omega \sim_t \tilde{\omega} \Leftrightarrow \forall s < t(\omega(s) = \tilde{\omega}(s))$$

If we let $[\omega]_t = \{\tilde{\omega} : \tilde{\omega} \sim_t \omega\}$ be the equivalence class of ω and $\mathcal{B}_t$ be the internal algebra generated by these equivalence classes, then $\mathcal{B}_t$ in an obvious way codifies the information we have available at time t.

There are several natural questions to ask at this point. The most obvious is perhaps why the space increment Δx should be the square root of the time increment Δt. To see why, let us replace $\sqrt{\Delta t}$ by Δx and compute the variance of the process at time t:

$$\begin{aligned} E(B(t)^2) &= E((\sum_{s<t} \omega(s)\Delta x)^2) \\ &= E(\sum_{r,s<t} \omega(r)\omega(s)\Delta x^2) \\ &= \Delta x^2 E(\sum_{r\neq s} \omega(r)\omega(s)) + \Delta x^2 E(\sum_{s<t} \omega(s)^2) \end{aligned}$$

Since $\omega(s)^2 = 1$, the last sum equals $\Delta x^2 \frac{t}{\Delta t}$. In the next to last sum, each term equals plus or minus one with probability $1/2$, and hence the expectation is zero. Thus

$$E(B(t)^2) = \Delta x^2 \frac{t}{\Delta t}$$

In order for this quantity not to be infinite or infinitesimal, Δx must be of order of magnitude $\sqrt{\Delta t}$. The most convenient choice is $\Delta x = \sqrt{\Delta t}$.

As for the Poisson process, we have to turn our nonstandard process B into a standard process b. We simply define $b : \Omega \times [0, \infty) \to \mathbb{R}$ by

$$b(\omega, t) = {}^\circ B(\omega, t^-)$$

where t^- is the largest element in T smaller than or equal to t (since we shall later show that b is continuous, there is no need to take limits here as we did for the Poisson process).

It turns out that b is a Brownian motion on the Loeb probability space $(\Omega,\ L(\mathcal{A}), P_L)$. I shall not prove this here as it will follow quite easily from more general results later in the paper, but I should at least explain what it means. Recall first that a (standard) random variable X is *normally distributed with mean zero and variance* t if

$$P\{\omega : c < X(\omega) < d\} = \frac{1}{\sqrt{2\pi t}} \int_c^d e^{-\frac{x^2}{2t}}\, dx$$

for all $c < d$. Recall also that $\sigma\{b(r) : r \leq s\}$ is the σ-algebra generated by the random variables $b(r), r \leq s$, i.e. the smallest σ-algebra such that $b(r)$ is measurable for all $r \leq s$. With these preliminaries we can define Brownian motion as follows:

Definition 3.1 *A* Brownian motion *is a stochastic process $b : \Omega \times [0, \infty) \to \mathbb{R}$ such that:*
(i) $b(0) = 0$
(ii) Almost all paths $t \mapsto b(\omega, t)$ are continuous
(iii) For $s < t$ the increment $b(t) - b(s)$ is independent of the the σ-algebra $\sigma\{b(r) : r \leq s\}$ generated up to time s, and $b(t)-b(s)$ is normally distributed with mean zero and variance t

REMARK: As there are several equivalent ways of defining Brownian motion, the definition above may look slightly different from what you are used to, but it is the most convenient formulation for our purposes.

PROBLEMS

3.1 *(i) Let B be Anderson's random walk and assume that $a = k\Delta x$ for some $k \in^* \mathbb{N}$. Show that*

$$P\{\max_{s \leq t} B(s) \geq a\} = 2P\{B(t) \geq a\} - P\{B(t) = a\}$$

This is called the reflection principle.
(ii) Show that

$$E((B_t - B_s)^4) = 3(t-s)^2 - 2(t-s)\Delta t < 3(t-s)^2$$

(iii) Show that

$$P\{\exists r \in [s,t](|B(r) - B(s)|^4 \leq \epsilon)\} \leq \frac{6}{\epsilon}|t-s|^2$$

(iv) Show that $t \mapsto B(\omega, t)$ is S-continuous for almost all ω.

3.2 *Compute the Fourier transform of the increment $B(t) - B(s)$, $t > s$, of Anderson's random walk by showing that*

$$E[e^{iy(B_t - B_s)}] \approx e^{-y^2(t-s)/2}$$

This shows that ${}^\circ(B_t - B_s)$ is normally distributed with mean zero and variance $t - s$.

3.3 *Use the preceding two problems to show that the standard part of B is a Brownian motion.*

4. Internal Martingales

Having had a look at two important examples of stochastic processes, we should now be ready to take a look at the general theory. Let $(\Omega, \mathcal{A}, P)$ be an internal probability space and let T be a hyperfinite timeline. A *filtration* $\{\mathcal{A}_t\}_{t\in T}$ is an internal and non-decreasing family of subalgebras of $\mathcal{A}$. An internal process

$$X : \Omega \times T \to^* \mathbb{R}$$

is *adapted* to this filtration if the map $\omega \mapsto X(\omega, t)$ is $\mathcal{A}_t$-measurable for each $t \in T$. We get an example of a filtration and an adapted process by starting with a process X and letting $\mathcal{A}_t$ be the internal algebra ${}^*\sigma\{X(s) : s \leq t\}$ generated by $\{X(s) : s \leq t\}$, but in general it is convenient to allow the algebra $\mathcal{A}_t$ to be larger than ${}^*\sigma\{X(s) : s \leq t\}$. If we, for instance, want to study two processes X and Y, it may be better to let $\mathcal{A}_t =^* \sigma\{X(s), Y(s) : s \leq t\}$.

As we have mentioned several times above, it is often convenient to think of the algebra $\mathcal{A}_t$ as the information available at time t. That the process X is adapted, means that $X(t)$ only depends on information available at time t and not on information that will unfold in the future. If we look at the random walk $B(t)$ in the previous section, the information available at time t are the outcomes $\omega(0), \omega(\Delta t), \ldots, \omega(t - \Delta t)$ of the previous coin tosses, and the random walk is adapted to the filtration $\{\mathcal{B}_t\}$ since its value at time t only depends on these coin tosses and not on the ones that will occur later.

We are now ready to define one of the main concepts of this paper.

Definition 4.1 *An internal process $M : \Omega \times T \to^* \mathbb{R}$ is called a* martingale *with respect to the filtration $(\Omega, \{\mathcal{A}_t\}, P)$ if*
(i) M is adapted to $(\Omega, \{\mathcal{A}_t\}, P)$
(ii) If $s, t \in T$ and $s < t$, then $E(M(t)|\mathcal{A}_s) = M(s)$.

If we think of martingales in terms of gambling just as we thought of the random walk B in Section 3, condition (ii) tells us that M is a *fair* game in the sense that in average we neither win nor lose from time s to time t. Even if a game is 'fair' in this sense, it is not necessarily a game you would care to play—it could, for instance, be set up such that with probability almost one you would lose everything you had, and with a tiny probability you would win much more than you would ever need!

There is a slightly different way to define the martingale property (ii) which is often more convenient: We first define equivalence relations $\sim_t$ on Ω by

$$\omega \sim_t \tilde{\omega} \Leftrightarrow \forall A \in \mathcal{A}_t (\omega \in A \Leftrightarrow \tilde{\omega} \in A)$$

and note that the equivalence class $[\omega]_t$ of ω is given by

$$[\omega]_t = \bigcap \{A \in \mathcal{A}_t : \omega \in A\}$$

Condition (ii) above is then equivalent to

(ii') For each $\omega \in \Omega$ and $t \in T$,

$$\int_{[\omega]_t} \Delta M(\tilde{\omega}, t)\, dP(\tilde{\omega}) = 0$$

where $\Delta M(\omega, t)$ is the forward increment *of M defined by $\Delta M(\omega, t_i) = M(\omega, t_{i+1}) - M(\omega, t_i)$ for each $t_i \in T$.*

Observe that also the notion of being adapted can be reformulated in terms of the equivalence classes $[\omega]_t$: A process X is adapted if and only if X_t is constant on each equivalence class $[\omega]_t$.

Example 4.2 *Anderson's random walk is obviously a martingale with respect to the filtration it generates. The nonstandard Poisson process N is not since*

$$\int_{[\omega]_t} \Delta N(\omega, t)\, dP(\omega) = \alpha \Delta t P([\omega]_t)$$

But this means that the compensated Poisson process *$M(\omega, t) = N(\omega, t) - \alpha t$ is a martingale.*

Having defined martingales, let us now take a brief preview of the other main concept of this paper.

Definition 4.3 *Assume that $X : \Omega \times T \to^* \mathbb{R}$ and $Y : \Omega \times T \to^* \mathbb{R}$ are two internal processes. The* stochastic integral *$\int X\, dY$ is the internal process defined by*

$$\int X\, dY\, (\omega, t) = \sum_{s<t} X(\omega, s)\, \Delta Y(\omega, s)$$

In this generality, there is little that can be said about the stochastic integral. Even if the integrand X is finite, the integral may be infinite, as the following example shows.

Example 4.4 *Let B be the hyperfinite random walk in Section 3, and define X by*

$$X(\omega, s) = \omega(s)$$

Then

$$\int X \, dB(\omega, t) = \sum_{s<t} \omega(s)\omega(s)\sqrt{\Delta t} = t/\sqrt{\Delta t}$$

which is infinite for all noninfinitesimal t.

To avoid examples of this kind, we restrict ourselves to situations where the integrand X is adapted and where the integrator Y is a martingale. The next result gives us the first indication that we are on the right track:

Proposition 4.5 *Let $\{\mathcal{A}_t\}$ be an internal filtration. Assume that $X : \Omega \times T \to^* \mathbb{R}$ is adapted and that $M : \Omega \times T \to^* \mathbb{R}$ is a martingale. Then $\int X \, dM$ is a martingale with respect to $\{\mathcal{A}_t\}$.*

Proof: If $Y = \int X \, dM$, we have to prove that

$$\int_{[\omega]_t} \Delta Y(\tilde{\omega}, t) \, dP(\tilde{\omega}) = 0$$

Since X is adapted, X is constant on the equivalence class $[\omega]_t$, and hence

$$\int_{[\omega]_t} \Delta Y(\tilde{\omega}, t) \, dP(\tilde{\omega}) = X(\omega, t) \int_{[\omega]_t} \Delta M(\tilde{\omega}, t) \, dP(\tilde{\omega}) = 0$$

where the last step uses that M is a martingale. □

There is a nice interpretation of this proposition in terms of gambling. If the martingale M represents the possibilities of the game, think of the integrand $X(\omega, t)$ as the bet you make at time t if you are in the situation described by ω (you may prefer only to consider nonnegative integrands X at this stage). The stochastic integral $\int X \, dM$ then represents your win or loss. What the proposition says, is that as long as you can not see into future (i.e. as long as your integrand X is adapted), you can not turn a fair game in your favour by adapting your bets.

Let us remain on the gambling scene for a second. A special kind of gambling strategy is always to play the same bet until you decide to quit the game. If you can not see into the future, the decision of when to stop has to be made on the basis of the games you have already played. In mathematical terminology such a stopping strategy is called a *stopping time.*

Definition 4.6 *An internal map $\tau : \Omega \to^* \mathbb{R}$ is called a* stopping time *if the following holds for all $\omega \in \Omega$: If $\tau(\omega) = t$, then $\tau(\tilde{\omega}) = t$ for all $\tilde{\omega} \in [\omega]_t$*

An equivalent definition would be to demand the process $(\omega, t) \mapsto 1_{\{\tau(\omega)\leq t\}}$ to be adapted.

Given a process X and a stopping time τ, we define the *stopped process* X_τ by

$$X_\tau(\omega, t) = X(\omega, t \wedge \tau(\omega))$$

There are several reasons why we would like to stop processes. One is simply to prevent them from getting larger than we can handle, another is to catch them at a particularly critical time.

PROBLEMS

4.1 *Show that the two formulations (ii) and (ii') of the martingale property really are equivalent.*

4.2 *Show that it is true that a process X is adapted if and only if X_t is constant on each equivalence class $[\omega]_t$.*

4.3 *Assume that $(\Omega, \{\mathcal{A}_t\}, P)$ is a filtration and that $M : \Omega \to^* \mathbb{R}$ is an internal random variable. Show that $M_t := E(M \mid \mathcal{A}_t)$ is a martingale.*

5. Doob's Inequality

Due to the built-in fairness, it is rather obvious that in average the supremum of a martingale over an interval can not be much worse than it s value at the right end point. A famous inequality of Doob's makes this intuition precise, but before we prove it, we introduce two closely related classes of processes.

Definition 5.1 *An internal process $X : \Omega \times T \to^* \mathbb{R}$ is called a* submartingale *(*supermartingale*, respectively) with respect to an internal filtration $\{\mathcal{A}_t\}$ if X is adapted to the filtration and if for all $s < t$*

$$E(X_t|\mathcal{A}_s) \geq X_s$$

$(E(X_t|\mathcal{A}_s) \leq X_s$, respectively).

An alternative way of defining submartingales would be to say that for all ω and all t

$$\int_{[\omega]_t} \Delta X(\omega, t)\, dP(\omega) \geq 0$$

and similarly for supermartingales. In gambling terms, sub- and supermartingales are games which are systematically unfair in one direction—submartingales are in your favour and supermartingales in your disfavour.

Games you find in casinos (such as roulette) are typically supermartingales. Observe that if M is a martingale and ϕ is a convex function, then the process $X_t = \phi(M_t)$ is a submartingale (use Jensen's inequality).

We are now ready to state (the nonstandard version of) Doob's result.

Theorem 5.2 (Doob's inequality) *If $X : \Omega \times T \to^* \mathbb{R}$ is a positive submartingale and $p \in^* \mathbb{R}$ is larger than 1, then*

$$\| \max_{s \leq t} X_s \|_p \leq \frac{p}{p-1} \|X_t\|_p$$

for all t (where $\| \cdot \|_p$ denotes L^p-norm).

Before we can prove Doob's inequality, we need a lemma from real analysis.

Lemma 5.3 *Let U and V be two internal, nonnegative random variables. Assume that $p, \alpha \in^* \mathbb{R}$ with $p > \alpha$, $p > 1$, $\alpha > 0$, and that for all positive $\xi \in^* \mathbb{R}$*

$$\xi^\alpha \, P[U > \xi] \leq \int_{[U > \xi]} V^\alpha \, dP$$

Then

$$E(U^p) \leq \left(\frac{p}{p-\alpha}\right)^{\frac{p}{\alpha}} E(V^p)$$

Proof: Let μ be the distribution of U (i.e. $\mu(A) = P[U \in A]$). Then

$$E(U^p) = \int_0^\infty y^p \, d\mu(y)$$

$$= \int_0^\infty \left(\int_0^y p\xi^{p-1} \, d\xi\right) d\mu(y) = \int_0^\infty \left(\int_\xi^\infty p\xi^{p-1} \, d\mu(y)\right) d\xi$$

$$= \int_0^\infty p\xi^{p-1} \, P[U > \xi] \, d\xi \leq \int_0^\infty p\xi^{p-1-\alpha} \left(\int_{[U>\xi]} V^\alpha \, dP\right) d\xi$$

where the last step uses the assumption. Continuing, we see that

$$E(U^p) \leq \int_0^\infty p\xi^{p-1-\alpha} \left(\int_{[U>\xi]} V^\alpha \, dp\right) d\xi = \int \left(\int_0^U p\xi^{p-1-\alpha} \, d\xi\right) V^\alpha \, dP$$

$$= \int \frac{p}{p-\alpha} U^{p-\alpha} V^\alpha \, dP \leq \frac{p}{p-\alpha} E(U^p)^{1-\alpha/p} E(V^p)^{\alpha/p}$$

by Hölder's inequality. Dividing by $E(U^p)^{1-\alpha/p}$ and raising both sides to the p/α-th power, we prove the lemma. $\square$

Proof of Doob's inequality: To prove Doob's inequality, we apply the lemma for $\alpha = 1$ to the random variables $U = \max_{s\leq t} X_s$ and $V = X_t$. We only have to check that

$$\xi\, P[U > \xi] \leq \int_{[U>\xi]} V\, dP$$

for all $\xi > 0$.

If we define a stopping time τ by

$$\tau(\omega) = \min\{s \in T : X(\omega, s) > \xi\} \wedge t_\infty$$

where $t_\infty > t$, we see that

$$\{\max_{s\leq t} X_t > \xi\} = \{\tau \leq t\}$$

Hence

$$\xi P[\max_{s\leq t} X_s > \xi] = \xi P[\tau \leq t] \leq \int_{[\tau\leq t]} X_\tau\, dP$$

$$= \int (X_{\tau\wedge t} - X_t)\, dP + \int_{[\tau\leq t]} X_t\, dP \leq \int_{[\tau\leq t]} X_t\, dP$$

where the last inequality uses the submartingale property. Hence

$$\xi P[\max_{s\leq t} X_s > \xi] \leq \int_{[\max_{s\leq t} X_s > \xi]} X_t\, dP$$

and Doob's inequality follows. □

PROBLEMS

5.1 *Show that if M is a martingale, then*

$$E(\max_{s\leq t} M_s^2) \leq 4E(M_t^2)$$

This is probably the most commonly used instance of Doob's inequality.

5.2 *Show that if M is a martingale and ϕ is a convex function, then the process $X(\omega, t) = \phi(M(\omega, t))$ is a submartingale.*

5.3 *Show that if M is a martingale such that $E(M_t^2)$ is finite and $M_t \approx 0$ P_L-a.e., then $max_{s\leq t} M_s \approx 0$ P_L-a.e.*

6. Quadratic Variation

A martingale is a complicated object. The paths of a Brownian motion, for instance, are nowhere differentiable and have an intricate structure. Associated with a martingale we have a much simpler process which in many ways reflects the properties of the martingale itself.

Definition 6.1 *Let* $M : \Omega \times T \to^{*} \mathbb{R}$ *be an internal martingale. The* quadratic variation $[M]$ *is the internal process defined by*

$$[M](\omega, t) = \sum_{s<t} \Delta M(\omega, s)^2$$

Example 6.2 *If* B *is the random walk in Section 3, then*

$$[B](\omega, t) = \sum_{s<t} \Delta t = t$$

This example shows that a complicated martingale may have a very simple quadratic variation.

The next result gives us a simple but very useful connection between a martingale and its quadratic variation.

Proposition 6.3 *If* M *is a martingale*

$$M_t^2 = M_0^2 + 2\int_0^t M_s\, dM_s + [M]_t$$

Proof: This is just a simple algebraic identity. Observe that

$$M_{s_{i+1}}^2 - M_{s_i}^2 = (M_{s_i} + \Delta M_{s_i})^2 - M_{s_i}^2 = 2M_{s_i}\Delta M_{s_i} + \Delta M_{s_i}^2$$

Summing this identity over all s_i less than t, we get the result. □

Taking expectations we get an important corollary.

Corollary 6.4 *If* M *is a martingale*

$$E(M_t^2) = E(M_0^2) + E([M]_t)$$

Proof: The process $Y = \int M dM$ is a martingale. Hence $E(Y_t) = E(Y_0) = 0$, and

$$\begin{aligned} E(M_t^2) &= E(M_0^2) + 2E(\int_0^t M_s\, dM_s) + E([M]_t) \\ &= E(M_0^2) + E([M]_t) \end{aligned}$$

which proofs the corollary. □

We also have the following corollary:

Corollary 6.5 *If* M *is an internal martingale and* $\{\tau_n\}_{n\in {}^*\mathbb{N}}$ *is an internal and non-decreasing sequence of stopping times, then*

$$E(M_{\tau_n}^2) = E(M_{\tau_0}^2) + E(\sum_{k<n} (M_{\tau_{k+1}} - M_{\tau_k})^2)$$

where M_{τ_n} *denotes the random variable* $M_{\tau_n}(\omega) = M(\omega, \tau_n(\omega))$.

Proof: Define an internal process $N : \Omega \times^* \mathbb{N} \to^* \mathbb{R}$ by

$$N(\omega, n) = M(\omega, \tau_n(\omega))$$

and observe that N is a martingale (with respect to which filtration?). Then apply the corollary above. $\square$

Here is a final result of the same type which will be quite useful in later sections:

Proposition 6.6 *Assume that M is an internal martingale and that τ and σ are two stopping times such that $\sigma(\omega) \geq \tau(\omega)$ for all ω. Then*

$$E((M_\sigma - M_\tau)^2) = E([M]_\sigma - [M]_\tau)$$

Proof: We may prove this as a corollary of the results above, but it is just as easy to prove from scratch. Note that

$$E((M_\sigma - M_\tau)^2) = E((\sum_{\tau(\omega)\leq t<\sigma(\omega)} \Delta M(\omega, t))^2)$$

$$= E(\sum_{\tau(\omega)\leq t<\sigma(\omega)} \Delta M(\omega, t)^2) + 2E(\sum_{\tau(\omega)\leq s<t<\sigma(\omega)} \Delta M(\omega, s)\Delta M(\omega, t))$$

$$= E([M]_\sigma - [M]_\tau)$$

where the last step uses that

$$E(\sum_{\tau(\omega)\leq s<t<\sigma(\omega)} \Delta M(\omega, s)\Delta M(\omega, t)) = 0$$

by the martingale property and the definition of stopping times. $\square$

As we shall see later, the results in this section give us some very useful relations between the L^2-norm of a martingale M and the L^1-norm of its quadratic variation $[M]$. In particular, Corollary 6.4 tells us that

$$\| M_t \|_2^2 = \| M_0^2 + [M]_t \|_1$$

It is natural to ask if there is a similar relationship for other L^p-norms. A positive answer is provided by the famous *Burkholder-Davis-Gundy inequalities.* Here is one of several versions:

Theorem 6.7 *For all internal martingales and all $p \in^* (0, \infty)$*

$$(10p)^{-1} \| \sqrt{M_0^2 + [M]_t} \|_p \leq \| \max_{s\leq t} M_s \|_p \leq p\sqrt{12} \| \sqrt{M_0^2 + [M]_t} \|_p$$

Proofs of (versions of) these inequalities can be found in most standard texts on stochastic analysis, e.g., [12], [28], [29].

Although I shall avoid the Burkholder-Davis-Gundy inequalities in this paper, you should be aware of their existence—they are indispensable tools in some contexts, and they simplify arguments substantially in other situations.

PROBLEMS

6.1 *In the proof of Corollary 6.5 I claimed that the process N is a martingale. Show that this is correct if we choose the right filtration.*

6.2 *In the proof of Proposition 6.6 I claimed that*

$$E\Big(\sum_{\tau(\omega)\leq s<t<\sigma(\omega)} \Delta M(\omega,s)\Delta M(\omega,t)\Big) = 0$$

by the martingale property. Check the details.

6.3 *If M and N are two internal martingales, the* joint variation $[M,N]$ *is the process defined by*

$$[M,N](t) = \sum_{s<t} \Delta M_s \Delta N_s$$

Show that

$$[M,N] = \frac{1}{2}([M+N] - [M] - [N])$$

7. Standard Parts

In order to get a reasonably behaved theory, we have to put size restrictions on our martingales.

Definition 7.1 *A martingale $M : \Omega \times T \to^* \mathbb{R}$ is called a* λ^2-martingale *if $E(M_t^2)$ is finite for all finite $t \in T$.*

Proposition 7.2 *The following are equivalent:*
(i) M is a λ^2-martingale
(ii) $E(M_0^2 + [M]_t)$ is finite for all finite $t \in T$
(iii) $E(\max_{s\leq t} M_s^2)$ is finite for all finite $t \in T$

Proof: (i) and (ii) are equivalent by Corollary 6.4, and (i) and (iii) are equivalent by Doob's inequality (use $X = |M|$ and $p = 2$). □

Note that according to (iii), almost all the paths of a λ^2-martingale are finite for all finite times.

Many of the properties of λ^2-martingales can be generalized to a larger class in a routine way:

Definition 7.3 *An internal martingale $M : \Omega \times T \to^* \mathbb{R}$ is called a* local λ^2-martingale *if there exists a non-decreasing sequence $\{\tau_n\}_{n\in\mathbb{N}}$ of internal stopping times such that*
(i) For each $n \in \mathbb{N}$, the stopped martingale M_{τ_n} (defined by $M_{\tau_n}(\omega, t) = M(\omega, t \wedge \tau_n(\omega))$) is a λ^2-martingale
(ii) For P_L-almost all ω,

$$\lim_{n\to\infty} {}^\circ\tau_n(\omega) = \infty$$

Such a sequence of stopping times τ_n is called a localizing sequence *for the martingale M.*

We shall now see how we can turn (local) λ^2-martingales into standard processes. First a definition:

Definition 7.4 *Assume that $F : T \to^*\mathbb{R}$ is an internal function, and let b and r be real numbers. We say that b is the* S-right limit *of F at r if for each $\epsilon \in \mathbb{R}_+$, there exists a $\delta \in \mathbb{R}_+$ such that if $t \in T$ satisfies $r <^\circ t < r+\delta$, then $|F(t) - b| < \epsilon$. We write*

$$S - \lim_{t\downarrow r} F(t) = b$$

The S-left limit *$S - \lim_{t\uparrow r} F(t)$ is defined similarly*

Note that the behaviour of F on the monad of r does *not* enter into the definition of $S - \lim_{t\downarrow r} F(t)$ and $S - \lim_{t\uparrow r} F(t)$.

If an internal function F has S-left and S-right limits at all $r \in \mathbb{R}$, we say that it has *one-sided limits*. An internal process $X : \Omega \times T \to^* \mathbb{R}$ has *one-sided limits* if almost all paths $t \mapsto X(\omega, t)$ have one-sided limits.

If a function or a process has one-sided limits, we can turn it into a standard function or process by using a standard part construction:

Definition 7.5 *If $F : T \to^* \mathbb{R}$ has one-sided limits, we define the* standard part ${}^\circ F$ *to be the function ${}^\circ F : [0, \infty) \to \mathbb{R}$ defined by*

$${}^\circ F(r) = S - \lim_{t\downarrow r} F(t)$$

The left standard part ${}^\circ F^-$ *is defined by*

$${}^\circ F^-(r) = S - \lim_{t\uparrow r} F(t)$$

REMARK: It turns out that the standard part map $F \mapsto {}^\circ F$ really corresponds to the standard part map of a topology, namely the *Skorohod topology* on the space of all right continuous functions with left limits (see [11], [30]).

Theorem 7.6 *A local λ^2-martingale has one-sided limits.*

Proof: It is clearly enough to prove the theorem for λ^2-martingales on a finite interval $[0, S]$. Since a λ^2-martingale is finite almost everywhere on such an interval, it can only fail to have one-sided limits by oscillating too much. This means that if the theorem fails, then

$$\bigcup_{a,b\in\mathbb{Q}} \{\omega : M(\omega,\cdot) \text{ crosses the interval } [a,b] \text{ infinitely often before time } S\}$$

must have positive measure. Since there are only countably many pairs of rationals a, b, this means that we can find $a, b \in \mathbb{Q}$, $a < b$, such that M crosses $[a, b]$ infinitely many times with positive probability.

We define a sequence of stopping times τ_n as follows: $\tau_0 = 0$, and for n odd, let

$$\tau_n(\omega) = \min\{t > \tau_{n-1}(\omega) : M(\omega, t) \leq a\} \wedge S$$

Similarly, for n even, let

$$\tau_n(\omega) = \min\{t > \tau_{n-1}(\omega) : M(\omega, t) \geq b\} \wedge S$$

The sequence τ_n is strictly increasing until it reaches S. If N is the number of elements in the timeline smaller than or equal to S, we see by Proposition 6.6 that

$$E(M_S^2) = E(M_{\tau_N}^2) = E(M_0^2 + \sum_{n=0}^{N-1} (M_{\tau_{n+1}} - M_{\tau_n})^2)$$

The expression on the left is finite by assumption, while the sum on the right is infinite on a set of positive measure, and hence has infinite expectation. This gives us our contradiction, and the theorem is proved. $\square$

The result above means that we can always turn a λ^2-martingale into a right continuous process with left limits.

PROBLEMS

7.1 *Show that the nonstandard Poisson process N in Section 2 has one-sided limits.*

7.2 *Show that if the process X has one-sided limits, then for each $t \in (0, \infty)$ there are $t_-, t_+ \in T$ such that ${}^\circ t_- = {}^\circ t_+ = t$ and*

$$^\circ X(\omega, r) = S - \lim_{s\uparrow t} X(\omega, s) \quad P_L\text{-a.s. for all } r \approx t,\ r \leq t_-$$

$$^\circ X(\omega, r) = S - \lim_{s\downarrow t} X(\omega, s) \quad P_L\text{-a.s. for all } r \approx t,\ r \geq t_+$$

8. S-continuity

In the previous section we showed that all λ^2-martingales have standard parts. In practice, most of the martingales we meet are continuous, but this is not always easy to prove from scratch. In this section we shall prove a powerful result which is often helpful. But before we state the theorem, we must describe the notion of continuity that we shall be using.

Definition 8.1 *An internal function $F : T \to^* \mathbb{R}$ is* S-continuous *if for all finite and infinitely close $s, t \in T$*

$$-\infty <^\circ F(s) =^\circ F(t) < \infty$$

An internal process $X : \Omega \times T \to^* \mathbb{R}$ is S-continuous *if almost all paths $t \mapsto X(\omega, t)$ are S-continuous.*

Theorem 8.2 *A local λ^2-martingale is S-continuous if and only if its quadratic variation is S-continuous.*

The aim of this section is to prove the theorem above. We shall need the following definition.

Definition 8.3 *An internal martingale M has* infinitesimal increments almost surely *if for all finite $S \in T$ the set*

$$\{\omega : \exists t < S(\Delta M(\omega, t) \not\approx 0)\}$$

has Loeb measure zero. M has infinitesimal increments *if the set above is empty for all $S \in T$.*

The key observation is that we may replace a martingale with infinitesimal increments almost surely by one with infinitesimal increments without altering the path properties noticeably:

Proposition 8.4 *Assume that M is an λ^2-martingale with infinitesimal increments almost surely. Then there is a λ^2-martingale N with infinitesimal increments and a set $\tilde{\Omega}$ of Loeb-measure one such that*

$$N(\omega, t) \approx M(\omega, t) \text{ and } [N](\omega, t) \approx [M](\omega, t)$$

for all finite t and all $\omega \in \tilde{\Omega}$.

Proof: For each $k \in^* \mathbb{N}$ let

$$\Omega_k = k = \{\omega : \exists t < k(|\Delta M(\omega, t)| > 1/k)\}$$

The set $\{k \in^* \mathbb{N} : P(\Omega_k) < 1/k\}$ is internal and contains $\mathbb{N}$, and hence by *overspill* it has an infinite element K. For each $\omega \in \Omega_K$ let t_ω be the first t such that $|\Delta M(\omega, t)| > 1/K$, and put $t_\omega = K$ if $\omega \notin \Omega_K$.

Define

$$[\omega]_t^+ = \{\tilde{\omega} \in [\omega]_t : t_{\tilde{\omega}} \le t\}$$

and note that if $t > t_{\tilde{\omega}}$ for some $\tilde{\omega} \in [\omega]_t$, then $[\omega]_t^+ = [\omega]_t$.

We first modify M by cutting off those increments which are larger than $1/K$: More precisely, we let $\tilde{M}$ be the process defined by $\tilde{M}(0) = M(0)$ and

$$\Delta\tilde{M}(\omega,t) = \begin{cases} 0 & \text{if } t \ge t_\omega \\ \Delta M(\omega,t) & \text{if } t < t_\omega \end{cases}$$

$\tilde{M}$ is usually not a martingale, but if we add the process Y given by $Y(0) = 0$ and

$$\Delta Y(\omega,t) = \frac{\int_{[\omega]_t^+} \Delta M(t)\,dP}{P([\omega]_t^+)}$$

then $\tilde{N} = \tilde{M} + Y$ is a martingale (if $P([\omega]_t^+) = 0$, we just let $\Delta Y(\omega,t) = 0$).

The key observation is that for all finite S, $\sum_{t<S} |\Delta Y(\omega,t)|$ is infinitesimal almost everywhere. To see why, we perform the following computation:

$$E(\sum_{t<S} |\Delta Y(\omega,t)|) = \sum_{\omega\in\Omega}\sum_{t<S} |\sum_{\tilde{\omega}\in[\omega]_t^+} \frac{\Delta M(\tilde{\omega},t)P(\tilde{\omega})}{P([\omega]_t^+)}| P\{\omega\}$$

$$\le \sum_{\omega\in\Omega}\sum_{t<S}\sum_{\tilde{\omega}\in[\omega]_t^+} \frac{|\Delta M(\tilde{\omega},t)|P(\tilde{\omega})}{P([\omega]_t^+)} P\{\omega\}$$

$$= \sum_{\tilde{\omega}\in\Omega_K}\sum_{\omega\in[\tilde{\omega}]_{t_{\tilde{\omega}}}^+} \frac{|\Delta M(\tilde{\omega},t_{\tilde{\omega}})|P\{\tilde{\omega}\}}{P([\omega]_{t_{\tilde{\omega}}}^+)} P\{\omega\}$$

$$= \int_{\Omega_K} |\Delta M(\tilde{\omega},t_{\tilde{\omega}})|\,dP \le 2\int_{\Omega_K} \max_{t<S} |M(t)|\,dP$$

Since $E(\max_{t<S} |M(t)|^2)$ is finite by Doob's inequality, $\max_{t<S} |M(t)|$ is S-integrable (see, e.g., Problem 1.3), and hence the last integral is infinitesimal since Ω_K has infinitesimal measure.

On the subset of $\Omega - \Omega_K$ where $\sum_{t<S} |\Delta Y(\omega,t)|$ is infinitesimal, we clearly have $M(\omega,t) = \tilde{N}(\omega,t)$ for all $t < S$. Moreover, for $\omega \in \Omega - \Omega_K$:

$$[\tilde{N}](\omega,t) - [M](\omega,t) = \sum_{s<t}(2M(s) + \Delta Y(s))\Delta Y(s)$$

which is infinitesimal for all $t < S$ and almost all ω. Thus $\tilde{N}$ looks like a good candidate for our martingale N, but there is still a problem—$\tilde{N}$ need not have infinitesimal increments since $\Delta Y(\omega,t)$ may be noninfinitesimal. There is a simple remedy. By *overspill* the set

$$\{n \in^* \mathbb{N} : P\{\omega : \exists t < n(|\Delta Y(\omega,t)| > \frac{1}{n})\} < \frac{1}{n}\}$$

has an infinite element γ. Define $\tau : \Omega \to T$ by

$$\tau(\omega) = \min\{t \in T : |\Delta Y(\omega,t)| > \frac{1}{\gamma}\} \wedge \gamma$$

Since $\Delta Y(\omega,t)$ only depends on the equivalence class $[\omega]_t$, τ is a stopping time. Since $\tau(\omega) = \gamma$ almost everywhere, it is trivial to check that the stopped process $N(\omega,t) = \tilde{N}(\omega, t \wedge \tau(\omega))$ satisfies the proposition. □

The proposition above may seem to give us very little in exchange for a lot of hard work. The result is quite useful, however, because it allows us to stop a martingale before it grows too large. If M has infinitesimal increments, and τ_n is the stopping time

$$\tau_n(\omega) = \min\{t \in T : M(\omega,t) \geq n\}$$

then we know that the stopped process $M_\tau(\omega,t) = M(\omega, t \wedge \tau)$ is finite. If we only knew that M had infinitesimal increments almost everywhere, then M_τ could be very large on an infinitely small set, and in many situations this would be enough to ruin our estimates. You can see how this technique works in the proof of the first half of our theorem:

Proposition 8.5 *Let M be an S-continuous local λ^2-martingale. Then the quadratic variation $[M]$ is S-continuous.*

Proof: It is clearly enough to prove the proposition for λ^2-martingales. Since an S-continuous martingale obviously has infinitesimal increments almost everywhere, Proposition 8.4 gives us a martingale N with infinitesimal increments and basically the same path properties. Replacing M by N if necessary, we may assume that M has infinitesimal increments. Using a sequence of stopping times $\{\tau_n\}$ as above, we may assume that M is bounded by a standard number C.

Assume for contradiction that $[M]$ is not S-continuous. Since $[M]$ is increasing, it can only fail to be S-continuous by jumping a noninfinitesimal distance in infinitesimal time. This means that the set

$$\bigcup_{a,b,S} \{\omega : \exists s,t < S(s \approx t \text{ and } M(\omega,s) < a \text{ and } M(\omega,t) > b)\},$$

where the union is over all rational numbers $a < b$ and all natural numbers S, must have positive Loeb measure. Since there are only countably many such triples (a, b, S), we can find two rational numbers $a < b$ and an $S \in \mathbb{N}$ such that with probability larger than some positive, real number p the process $[M]$ crosses $[a, b]$ during some time interval of infinitesimal length. If

$$\Omega_n = n = \{\omega : \exists s, t < S(s < t < s + \tfrac{1}{n} \text{ and } [M](s, \omega) < a \text{ and } [M](\omega, t) > b)\}$$

we see that $P(\Omega_n) > p$ for all finite n, and hence by *overspill*, the internal set

$$\{n \in^* \mathbb{N} : P(\Omega_n) > p\}$$

has an infinite element N. If we set $\epsilon = 1/N$ and $\tilde{\Omega} = \Omega_N$, we get an infinitesimal ϵ and an internal set $\tilde{\Omega}$ of noninfinitesimal measure $> p$ such that for each $\omega \in \tilde{\Omega}$, we can find $s, t < T$ such that $s < t < t + \epsilon$ and $M(\omega, s) < a$, $M(\omega, t) > b$.

We are now ready to complete the proof. Define two stopping times

$$\begin{aligned} \tau(\omega) &= \min\{t \in T : [M](\omega, t) > a\} \\ \sigma(\omega) &= \tau(\omega) + \epsilon \end{aligned}$$

and note that by Proposition 6.6

$$E((M_\sigma - M_\tau)^2) = E([M]_\sigma - [M]_\tau)$$

Since M is S-continuous and S-bounded, the expression on the left is infinitesimal, while the (standard part of the) expression on the left is bounded below by $(b - a)p/2$. Hence we have our contradiction and the proposition is proved. □

Using a more complicated version of the same argument, we may now prove the converse implication.

Proposition 8.6 *Let M be a local λ^2-martingale and assume that the quadratic variation $[M]$ is S-continuous. Then M is also S-continuous.*

Proof: Arguing as in the preceding proof, we may assume that M and $[M]$ are S-bounded and have infinitesimal increments.

Assume for contradiction that M is not S-continuous. Since there are only countably many intervals with rational endpoints, we can find two rational numbers $a < b$ and a finite $S \in T$ such that the set

$$\{\omega : \exists s, t < S(s \approx t \text{ and } M(\omega, s) < a \text{ and } M(\omega, t) > b)\}$$

has positive Loeb measure. Each ω in this set corresponds to a path which either traverses the interval $[a, b]$ infinitely fast on its way up or on its way down. Let us for simplicity assume that the set of upcrossings has positive measure. This means that we can find a positive real number p such that for all $n \in \mathbb{N}$

$$\Omega_n = n = \{\omega : \exists s, t < S(s < t < s + \tfrac{1}{n} \text{ and } M(s, \omega) < a \text{ and } M(\omega, t) > b)\}$$

has internal measure larger than p. By *overspill*, the internal set

$$\{n \in^* \mathbb{N} : P(\Omega_n) > p\}$$

has an infinite element N. If we set $\epsilon = 1/N$ and $\tilde{\Omega} = \Omega_N$, we get an infinitesimal ϵ and an internal set $\tilde{\Omega}$ of noninfinitesimal measure $> p$ such that for each $\omega \in \tilde{\Omega}$, we can find $s, t < T$ such that $s < t < s + \epsilon$ and $M(\omega, s) < a$, $M(\omega, t) > b$.

To show that this leads to a contradiction, we have to be a little more careful than in the previous proof. The problem is that the martingale M may traverse the interval $[a, b]$ many times, and we are only interested in the infinitely fast instances. Let $d = b - a$ and define two sequences $\{\tau_n\}$ and $\{\sigma_n\}$ of stopping times by $\tau_0 = 0$, $\sigma_0 = 0$ and:

$$\begin{aligned}
\tau_{k+1}(\omega) &= \min\{t > \sigma_k(\omega) : M(\omega, t) > a + \frac{d}{3} \\
&\qquad \text{and there exists } s \in (t - \epsilon, t) \text{ such that } M(\omega, s) < a\} \wedge S \\
\sigma_{k+1}(\omega) &= \min\{t > \tau_{k+1}(\omega) : M(\omega, t) > b \text{ or } M(\omega, t) < a\} \wedge (\tau_{k+1} + \epsilon)
\end{aligned}$$

The idea is that each time M tries to traverse $[a, b]$ in less time than ϵ, one of the stopping times τ_n will catch the attempt before it is half way, and σ_n will then register when M reaches b. The τ_n's will also catch some failed attempts to reach b, but these will be relatively few since the martingale can only cross the interval $[a, a + \frac{d}{3}]$ finitely many times before S.

By Proposition 6.6 we now see that for any $K \in \mathbb{N}$,

$$\sum_{k=1}^{K} E((M_{\sigma_k} - M_{\tau_k})^2) = \sum_{k=1}^{K} E([M]_{\sigma_k} - [M]_{\tau_k})$$

Since $[M]$ is S-bounded and S-continuous, the expression on the right is obviously infinitesimal for all finite K. In the expression on left, however, (almost) all paths in $\tilde{\Omega}$ will sooner or later make a noninfinitesimal contribution, and hence this expression has to be noninfinitesimal for sufficiently large $K \in \mathbb{N}$. We have our contradiction and the proposition is proved. □

Combining propositions 8.5 and 8.6, we get Theorem 8.2. As an obvious application, we have:

Corollary 8.7 *Anderson's random walk B is S-continuous.*

Proof: $[B](t) = t$ which is obviously S-continuous. □

There is a modification of Theorem 8.2 which is occasionally more convenient to use in applications. Given a martingale M with respect to a filtration $\{\mathcal{A}_t\}$, the *angular bracket* process $\langle M \rangle$ is defined by

$$\langle M \rangle(\omega, t) = \sum_{s<t} E(\Delta M(\omega, s)^2 | \mathcal{A}_s)$$

The theorem I have in mind is:

Theorem 8.8 *Assume that M is a λ^2-martingale with infinitesimal increments. Then M is S-continuous if and only if $\langle M \rangle$ is S-continuous.*

Note the additional assumption of infinitesimal increments a.s.—without it, the theorem fails (see Problem 8.4 for an example). The observation behind the theorem is quite simple:

Proposition 8.9 *If M is a λ^2-martingale with infinitesimal increments, then for almost all ω*

$$[M](\omega, t) \approx \langle M \rangle(\omega, t)$$

for all finite t.

Proof: The proposition above was given as a problem during the Summer Institute, and variants of the following proof were found by several of the participants.

Let $N(\omega, t) = [M](\omega, t) - \langle M \rangle(\omega, t)$, and observe that N is a martingale since

$$E(\Delta N(t) | \mathcal{A}_t) = E(\Delta M(t)^2 - E(\Delta M(t)^2 | \mathcal{A}_t)) = 0$$

Using a localizing sequence of stopping times if necessary, we may assume that M and $[M]$ are S-bounded. The quadratic variation of N is given by

$$E([N](t)) = E\left(\sum_{s<t} \Delta N(s)^2\right) = E\left(\sum_{s<t} (\Delta M(t) - E(\Delta M(t)^2 | \mathcal{A}_t))^2\right)$$

$$= E\left(\sum_{s<t} (\Delta M(s)^4 - 2\Delta M(s)^2 E(\Delta M(s)^2 | \mathcal{A}_s) + E(\Delta M(s)^2 | \mathcal{A}_s)^2)\right)$$

$$= E\left(\sum_{s<t} (\Delta M(t)^4 - E(\Delta M(s)^2 | \mathcal{A}_s)^2)\right) \leq E\left(\sum_{s<t} \Delta M(s)^4\right)$$

Since M has infinitesimal increments, the last expression is infinitesimally small compared to $E(\sum_{s<t} \Delta M(s)^2) = E([M](t))$, which is finite by assumption. By Doob's inequality,

$$E(\max_{s<t} N(s)^2) \leq 4E([N](t)) \approx 0$$

and hence N is infinitesimal. □

Theorem 8.8 now follows immediately from Theorem 8.2 and the proposition above.

PROBLEMS

8.1 *Let B be Anderson's random walk and let X be an adapted and S-bounded process (i.e. $|X| \leq K$ for some* real *number K). Prove that the stochastic integral $\int X\, dB$ is S-continuous.*

8.2 *(i) Show that if M, N are two S-continuous λ^2-martingales, then the* joint variation $[M,N](t) := \sum_{s<t} \Delta M_s \Delta N_s$ *is S-continuous.*
(ii) Prove a similar result for the process

$$\langle M,N\rangle(t) := \sum_{s<t} E(\Delta M_s \Delta N_s | \mathcal{A}_s)$$

8.3 *Modify Anderson's random walk such that at each t*

$$\Delta B(\omega,t) = \begin{cases} 2\sqrt{\Delta t} & \textit{with probability } 1/3 \\ -\sqrt{\Delta t} & \textit{with probability } 2/3 \end{cases}$$

Show that B is still a continuous martingale.

8.4 *Let $M(\omega,t) = N(\omega,t) - \alpha t$ where N is the nonstandard Poisson process from Section 3. Show that M is a λ^2-martingale, that $\langle M\rangle$ is S-continuous, but that M is not S-continuous.*

8.5 *Describe a λ^2-martingale with infinitesimal increments which is not S-continuous.*

9. Stochastic Integration

It's time to take a serious look at stochastic integration. We first want to decide what kind of integrators X we shall allow in our stochastic integrals $\int X\, dM$. We have already observed that X should be adapted, but we also need to put restrictions on the size of X. First a preliminary definition: If $t \in T$, let

$$T_t = \{s \in T : s < t\}$$

and let $\mathcal{T}_t$ be the set of all internal subsets of T_t. The (internal) *Doleans measure* ν_{M_t} on the space $(\Omega \times T_t, \mathcal{A} \times \mathcal{T}_t)$ is defined by

$$\nu_{M_t}(\omega,s) = P(\omega)\, \Delta M(\omega,s)^2$$

Note that all ν_{M_t} are essentially the same measure only defined on larger and larger parts of the timeline. When it is irrelevant (or clear from the context) which of these measures I have in mind, I shall just write ν_M.

We are now ready to define our space of integrators:

Definition 9.1 *Assume that M is a λ^2-martingale. The space $SL^2(M)$ consists of all adapted processes X such that $X \in SL^2(\nu_{M_t})$ for all finite $t \in T$. (Recall that $X \in SL^2(\nu_{M_t})$ means that X^2 is S-integrable with respect to the measure ν_{M_t}.)*

By using localizing sequences of stopping times, it is possible to extend the theory to larger classes of processes. The following definition should give you the idea.

Definition 9.2 *Assume that M is a local λ^2-martingale. An adapted process X belongs to $SL(M)$ if there is a localizing sequence of stopping times τ_n for M (recall Definition 7.3) such that $X \in SL^2(M_{\tau_n})$ for all $n \in \mathbb{N}$.*

Why are the Doleans-measures ν_{M_t} the right measures to use? The following simple observation shows that they are intimately connected to stochastic integrals:

Proposition 9.3 *If M is a martingale and X is an adapted process, then for all $t \in T$*

$$E((\int_0^t X\,dM)^2) = \int_{\Omega\times T_t} X^2\,d\nu_{M_t}$$

Proof: Since $\int X\,dM$ is a martingale starting at 0,

$$E((\int_0^t X\,dM)^2) = E([\int X\,dM](t))$$

according to Corollary 6.4. Since

$$[\int X\,dM](t) = \sum_{s<t} X(s)^2 \Delta M(s)^2$$

we see that

$$\begin{aligned} E([\int X\,dM](t)) &= E(\sum_{s<t} X(s)^2 \Delta M(s)^2) \\ &= \sum_{\omega\in\Omega}\sum_{s<t} X(\omega,s)^2 \Delta M^2(\omega,s) P(\omega) \\ &= \int_{\Omega\times T_t} X^2\,d\nu_{M_t} \end{aligned}$$

and the proposition is proved. □

Corollary 9.4 *If M is a λ^2-martingale and $X \in SL^2(M)$, then $\int X\,dM$ is a λ^2-martingale. If M is a local λ^2-martingale and $X \in SL(M)$, then $\int X\,dM$ is a local λ^2-martingale.*

The proposition also implies that if $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of adapted processes converging to X in $L^2(\nu_{M_t})$ (in the sense that ${}^\circ\int(X-X_n)^2\,d\nu_{M_t} \to 0$), then the stochastic integrals $\int_0^t X_n\,dM$ converge to $\int_0^t X\,dM$ in $L^2(P)$. Hence $SL^2(M)$ is the natural completion of the set of S-bounded, adapted processes as far as integration with respect to M is concerned.

We may now combine this observation with Theorem 8.2 to obtain a very useful result.

Theorem 9.5 *If M is an S-continuous local λ^2-martingale and X belongs to $SL(M)$, then $\int X\,dM$ is S-continuous.*

Proof: It suffices to prove the theorem when M is a λ^2-martingale and $X \in SL^2(M)$ (why?). Let us first assume that $|X| \leq n$, where $n \in \mathbb{N}$. Since

$$[\int X\,dM](\omega,t) = \sum_{r<t} X(\omega,r)^2\,\Delta M(\omega,r)^2$$

we have

$$[\int X\,dM](\omega,t) - [\int X\,dM](\omega,s) = \sum_{s\leq r<t} X(\omega,r)^2\,\Delta M(\omega,r)^2$$

$$\leq n^2 \sum_{s\leq r<t} \Delta M(\omega,r)^2 = n^2([M](\omega,t) - [M](\omega,s))$$

By assumption M is S-continuous, and hence $[M]$ is S-continuous by Theorem 8.2. By the calculation above, this means that $[\int X\,dM]$ is S-continuous, and applying Theorem 8.2 again, we get the S-continuity of $\int X\,dM$.

Generalizing to an arbitrary $X \in SL^2(M)$ is now just an exercise in measure theory. We first define the truncated integrands

$$X_n(\omega,t) = \begin{cases} n & \text{if } X(\omega,t) > n \\ X(\omega,t) & \text{if } -n \leq X(\omega,t) \leq n \\ -n & \text{if } X(\omega,t) < -n \end{cases}$$

and observe that by what we have just shown, $[\int X_n\,dM]$ is S-continuous. Now

$$0 \leq E(\sup_{s\leq t} {}^\circ([\int_0^s X\,dM] - [\int_0^s X_n\,dM]))$$

$$\leq {}^\circ E(\max_{s\leq t}([\int_0^s X\,dM] - [\int_0^s X_n\,dM]))$$

$$= ^\circ E([\int_0^t X\, dM] - [\int_0^t X_n\, dM])$$

$$= ^\circ E(\sum_{0 \le s < t} (X^2 - X_n^2)\, \Delta M^2) = ^\circ \int_{\Omega \times T_t} (X^2 - X_n^2)\, d\nu_{M_t}$$

Since $X \in SL^2(\nu_{M_t})$, the last integral goes to zero as $n \to \infty$, and consequently

$$E(\sup_{s \le t} {}^\circ([\int_0^s X\, dM] - [\int_0^s X_n\, dM])) \to 0$$

By standard measure theory, there is a subsequence

$$\sup_{s \le t} {}^\circ([\int_0^s X\, dM] - [\int_0^s X_{n_k}\, dM])$$

which converges to zero almost everywhere. Since each $[\int_0^s X_{n_k}\, dM]$ is S-continuous, it is easy to check that the uniform limit $[\int_0^s X\, dM]$ is S-continuous. By a final appeal to Theorem 8.2, we prove that $\int X\, dM$ is S-continuous. □

PROBLEMS

9.1 *Assume that $A \in \mathcal{A}_s$. Show that*

$$\nu_M(A \times [s, t - \Delta t]) = P(A)(M_t - M_s)^2$$

9.2 *Prove the claim about uniform limits of S-continuous functions made toward the end of the proof of Theorem 9.5.*

9.3 *An internal martingale is called an* SL^2-martingale *if M_t^2 is S-integrable for all finite t. Show that an internal martingale M is an SL^2-martingale if and only if $M_0^2 + [M]_t$ is S-integrable for all finite t.*

9.4 *Show that if M is an SL^2-martingale, then* $\max_{s \le t} M_s^2$ *is S-integrable for all finite t.*

10. Itô's Formula

In this section we shall prove a (simple version of) the fundamental theorem of stochastic calculus.

Theorem 10.1 (Itô's Formula) *Let M be an S-continuous, local λ^2-martingale, and assume that $\phi : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable. Then for almost all ω*

$${}^*\phi(M_t) \approx {}^*\phi(M_0) + \int_0^t {}^*\phi'(M_s)\, dM_s + \frac{1}{2}\int_0^t {}^*\phi''(M_s)\, d[M]_s$$

holds for all finite $t \in T$.

Proof: By Taylor's formula there is a point ξ_s between M_s and $M_s + \Delta M_s$ such that

$$^*\phi(M_s + \Delta M_s) -^* \phi(M_s) =^* \phi'(M_s)\, \Delta M_s + \frac{1}{2}\, ^*\phi''(\xi_s)\, \Delta M_s^2$$

$$=^* \phi'(M_s)\, \Delta M_s + \frac{1}{2}\, ^*\phi''(M_s)\, \Delta M_s^2 + \frac{1}{2}(^*\phi''(\xi_s) -^* \phi''(M_s))\Delta M_s^2$$

Since ΔM_s is infinitesimal almost surely and ϕ'' is S-continuous, the difference $^*\phi(\xi_s) -^* \phi(M_s)$ is infinitesimal. Hence

$$^*\phi(M_s + \Delta M_s) -^* \phi(M_s) =$$

$$=^* \phi'(M_s)\Delta M_s + \frac{1}{2}\, ^*\phi''(M_s)\Delta M_s^2 + \sigma(\omega, s)\Delta M_s^2$$

where for almost all ω, $\sigma(\omega, s)$ is infinitesimal for all finite s. Summing over all s less than t, we prove the theorem. □

REMARK: The remarkable feature of Itô's Formula is the second order term

$$\frac{1}{2}\int_0^t {}^*\phi''(M_s)\, d[M]_s$$

which does not occur in deterministic analysis. The reason is that in ordinary Stieltjes' integration, the integrator M is assumed to be of bounded variation and hence the quadratic variation is zero.

There are more general versions of Itô's Formula which are often useful in applications. You may amuse yourself by proving the following.

Theorem 10.2 (Itô's Formula) *Let M be an S-continuous, local λ^2-martingale, and let V be a non-decreasing, S-continuous process. Assume that $\phi(v, x)$ is continuously differentiable in the first variable and twice continuously differentiable in the second variable. Then for almost all ω*

$$^*\phi(V_t, M_t) \approx^* \phi(V_0, M_0) + \int_0^t \frac{\partial^*\phi}{\partial v}(V_s, M_s)\, dV_s$$

$$+ \int_0^t \frac{\partial^*\phi}{\partial x}(V_s, M_s)\, dM_s + \frac{1}{2}\int_0^t \frac{\partial^2\, {}^*\phi}{\partial x^2}(V_s, M_s)\, d[M]_s$$

holds for all finite $t \in T$.

PROBLEMS

10.1 *Prove Theorem 10.2.*

10.2 *(i) Let M be an S-bounded and S-continuous internal martingale with $M_0 = 0$. Show that $N(\omega, t) = \prod_{s<t}(1 + \Delta M_s(\omega))$ is a martingale and that for almost all ω*

$$N_t \approx e^{M_t - \frac{1}{2}[M]_t}$$

for all finite t.
(ii) Use Theorem 10.2 on $\phi(v, x) = e^{x - \frac{1}{2}v}$, $V_t = [M]_t$ and compare the result to (i).

10.3 *(i) Let B be Anderson's random walk, and let θ, α, β be real numbers. Define a process Θ by $\Theta_0 = \theta$ and*

$$\Theta_{t+\Delta t} = \Theta_t(1 + \alpha\Delta t + \beta\Delta B_t)$$

Show that for almost all ω

$$\Theta_t \approx \theta e^{(\alpha - \frac{1}{2}\beta^2)t + \beta B_t}$$

for all finite t.
(ii) Use Theorem 10.2 with $V_t = t$ and $\phi(t, x) = e^{(\alpha - \frac{1}{2}\beta^2)t + \beta x}$ and compare the result to (i).

11. Lévy's Characterization of Brownian Motion

In probability theory it is often useful to be able to recognize a Brownian motion in disguise, and we shall now use Itô's Formula to prove a very useful characterization. Before we begin, I should remind you that if F is a normally distributed random variable with mean zero and variance v, then the Fourier transform $\hat{F}$ of F is given by

$$\hat{F}(y) := \int e^{iyF(\omega)}\, dP(\omega) = e^{-\frac{y^2 v}{2}}$$

(you may check the last equality by using your calculus skills on the integral

$$\int e^{iyF(\omega)}\, dP(\omega) = \int_{-\infty}^{\infty} e^{iyx} \frac{e^{-x^2/2v}}{\sqrt{2\pi v}}\, dx \text{).}$$

Theorem 11.1 *Assume that M is a λ^2-martingale with $M(0) = 0$ such that for almost all ω we have $[M](\omega, t) \approx t$ for all finite $t \in T$. Then the standard part $^\circ M$ of M is a Brownian motion.*

Proof: We have to check that the three conditions in Definition 3.1 are satisfied. Since $[M]$ obviously is S-continuous, the continuity of M follows from Theorem 8.2, and as an immediate consequence we see that $°M(0) = 0$. It remains to show that if $s < t$, then $°M(t) - °M(s)$ is independent of $\sigma\{°M(r) : r \leq s\}$ and that $°M(t) - °M(s)$ is gaussian distributed with mean zero and variance $t - s$.

To prove this, assume that $t > s$ and let $B \in \sigma\{°M(r) : r \leq s\}$. Pick $\tilde{s}, \tilde{t} \in T$ such that $°\tilde{s} = s$ and $°\tilde{t} = t$. Applying Itô's Formula to the function $\phi(x) = e^{ixy}$, we get

$$e^{iy(M_{\tilde{t}} - M_{\tilde{s}})} \approx 1 + iy \int_{\tilde{s}}^{\tilde{t}} e^{iy(M_r - M_{\tilde{s}})}\, dM(r)$$

$$-\frac{y^2}{2} \int_{\tilde{s}}^{\tilde{t}} e^{iy(M_r - M_{\tilde{s}})}\, d[M](r)$$

If we take standard parts, multiply by the indicator function 1_B and integrate with respect to P_L, we get

$$\int_B e^{iy(°M_t - °M_s)}\, dP_L = P_L(B) - \frac{y^2}{2} \int_s^t (\int_B e^{iy(°M_r - °M_s)}\, dP_L) dr$$

This means that $t \mapsto \int_B e^{iy(°M_t - °M_s)}\, dP_L$ is a solution of the integral equation

$$u(t) = P_L(B) - \frac{y^2}{2} \int_s^t u(r)\, dr$$

Since $u(t) = P_L(B)\, e^{-\frac{y^2(t-s)}{2}}$ is the unique solution of this equation, we get

$$\int_B e^{iy(°M_t - °M_s)}\, dP_L = P_L(B)\, e^{-\frac{y^2(t-s)}{2}}$$

This shows that the distribution of $°M_t - °M_s$ over B is normal with variance $t - s$ for all B in $\sigma\{°M(r) : r \leq s\}$. $\square$

As an immediate consequence, we get:

Corollary 11.2 *The standard part of Anderson's random walk is a Brownian motion.*

There is a reformulation of Theorem 11.1 in terms of the angular bracket process which is often easier to use:

Theorem 11.3 *Assume that M is a λ^2-martingale with infinitesimal increments such that $M(0) = 0$. Assume also that for almost all ω we have $\langle M \rangle(\omega, t) \approx t$ for all finite $t \in T$. Then the standard part $°M$ of M is a Brownian motion.*

Proof: Combine the theorem above and Proposition 8.9. □

PROBLEMS

11.1 *Modify the definition of Anderson's random walk such that*

$$\Delta B(\omega, t) = \begin{cases} 4\sqrt{\Delta t} & \textit{with probability } 1/20 \\ 0 & \textit{with probability } 3/4 \\ -\sqrt{\Delta t} & \textit{with probability } 1/5 \end{cases}$$

Show that the standard part of B is a Brownian motion.

12. Connections to Standard Theory

If we want to use nonstandard stochastic integration to prove results about standard objects, we need to know how the nonstandard theory of stochastic integration relates to the standard theory. I'll sketch these connections very briefly—if you are interested in full proofs and more thorough explanations, you should look up the main references [11], [15], [1], [30].

Let us first take a look at the framework of standard stochastic integration. A (standard) *stochastic filtration* consists of a complete probability space $(\Omega, \mathcal{F}, Q)$ and an increasing family $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ of sub-σ-algebras of $\mathcal{F}$. For technical reasons, the filtration is usually assumed to satisfy the so-called *usual conditions* which means that all the null sets of $\mathcal{F}$ are included in $\mathcal{F}_0$ and that $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all t. An *L^2-martingale* with respect to this filtration is a stochastic process $m : \Omega \times [0, \infty) \to \mathbb{R}$ such that

(i) for each $t \in [0, \infty)$, the map $t \mapsto m(\omega, t)$ is $\mathcal{F}_t$-measurable
(ii) $E(m_t^2) < \infty$ for all $t \in [0, \infty)$
(iii) If $s < t$, then $E(m_t|\mathcal{F}_s) = m_s$

Just as in the nonstandard theory, there is a corresponding notion of a *local* L^2-martingale which I leave to the reader to define.

Before we can define stochastic integrals with respect to L^2-martingales, we must find the right class of integrands. We first need to introduce the notion of a predictable set. A *predictable rectangle* is a subset of $\Omega \times [0, \infty)$ of the form

$$B \times (s, t] \text{ where } s < t \text{ and } B \in \mathcal{F}_s$$

or

$$B \times [0, t] \text{ where } B \in \mathcal{F}_0$$

(the last clause is added for technical reasons and need not concern us here). A set is *predictable* if it is in the σ-algebra generated by the predictable rectangles. We get a measure μ_m on the predictable sets by letting

$$\mu_m(B \times (s,t]) = \int_B (m_t - m_s)^2 \, dQ$$

and

$$\mu_m(B \times [0,t]) = \int_B (m_t - m_0)^2 \, dQ$$

on predictable rectangles and then extending to the full σ-algebra. The measure μ_m is called the *Doleans measure* of m.

Definition 12.1 *If m is an L^2-martingale, the class $L^2(m)$ consists of all predictable processes x such that*

$$\int_{\Omega \times [0,t]} x^2 \, d\mu_m < \infty$$

for all $t \in [0,\infty)$.

For local L^2-martingales the natural class of integrands is denoted by $L(m)$. Again the details are left to the reader.

We now want to define the stochastic integral $\int x \, dm$ for all $x \in L^2(m)$. We first do this for *simple* processes of the form

$$x(\omega,t) = \sum_{i=1}^{n} a_i 1_{F_i}(\omega) 1_{(s_i,t_i]}(t)$$

where $a_i \in \mathbb{R}$ and $F_i \in \mathcal{F}_{s_i}$. In this case we simply put

$$\int_0^t x \, dm = \sum_{i=1}^{n} a_i 1_{F_i}(m_{t_i \wedge t} - m_{s_i \wedge t})$$

It turns out that for all t, the simple processes are dense among the predictable elements of $L^2(\Omega \times [0,t], \mu_m)$, and that the map

$$x \mapsto \int_0^t x \, dm$$

is an isometry from $L^2(\Omega \times [0,t], \mu_m)$ to $L^2(Q)$ (by this I just mean that

$$\int_{\Omega \times [0,t]} x^2 \, d\mu_m = E((\int_0^t x \, dm)^2)$$

for all simple processes x). By continuity, we may then extend the stochastic integral to all $x \in L^2(m)$. This definition is a bit convoluted, but since the paths of a martingale are of unbounded variation, there is no simple pathwise definition of the stochastic integral.

Note also that since the stochastic integral is defined as an L^2-limit for each t, I can change it on a null set for each t and still have an equally good representation of the integral. This means that we must be a little careful when we ask for path properties—for instance, it does not really make sense to ask if a stochastic integral is continuous; all we can ask is whether it has a continuous version.

Just as in the nonstandard theory, there is a quadratic variation process $[m]$ going with an L^2-martingale m. It can be defined as

$$[m](t) = L^2 - \lim_{|\mathcal{P}|\to 0} \sum_{i=0}^{n-1} (m(t_{i+1}) - m(t_i))^2$$

where $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ is a partition of the interval $[0, t]$, but it is technically more convenient to take a hint from Proposition 6.3 and define it as

$$[m](t) = m_t^2 - m_0^2 - 2\int_0^t m_s\, dm_s$$

(if the martingale is discontinuous, the integral $\int_0^t m_s\, dm_s$ must be replaced by $\int_0^t m_s^-\, dm_s$, where m^- is the left limit of m).

Having completed this quick survey of standard stochastic integration, we are ready to take a look at the connections between the standard and the nonstandard theory. To avoid some technical complications, we shall assume that our nonstandard λ^2-martingale M is S-continuous and has infinitesimal increments.

We have already seen that M has a standard part $m = {}^\circ M$. It turns out that this standard part process is an L^2-martingale with respect to the filtration $(\Omega, \{\mathcal{F}\}_t, P_L)$, where

$$\mathcal{F}_t = \sigma(\bigcup_{s \gtrsim t} L(\mathcal{A}_s) \cup \mathcal{N})$$

(here $\mathcal{N}$ is the family of P_L-null sets—they are thrown in to guarantee that $\mathcal{F}_t$ satisfies the *usual conditions* mentioned above). If $x \in L^2(m)$, a process $X \in SL^2(M)$ is called a *2-lifting* of x if

$${}^\circ X(\omega, t) = x(\omega, {}^\circ t)$$

for $L(\nu_M)$-almost all (ω, t) with ${}^\circ t < \infty$ (I use $L(\nu_M)$—and not $(\nu_M)_L$—to denote the Loeb measure of ν_M as I do not want too many subscripts). The main result is:

Theorem 12.2 *Assume that M is an S-continuous λ^2-martingale with infinitesimal increments. Then any $x \in L^2(m)$ has a 2-lifting $X \in SL^2(M)$ and the standard part of $\int X\,dM$ is $\int x\,dm$.*

In order to prove this result, we shall need the following lemma:

Lemma 12.3 *For all $t \in [0,\infty)$*

$$\mathcal{F}_t = \bigcup_{s\approx t} \sigma(L(\mathcal{A}_s) \cup \mathcal{N})$$

Proof: It suffices to show that $\mathcal{G}_t := \bigcup_{s\approx t} \sigma(L(\mathcal{A}_s) \cup \mathcal{N})$ is a σ-algebra, and to do so, we only have to check that $\mathcal{G}_t$ is closed under countable unions. If G_n belongs to $\mathcal{G}_t$ for all $n \in \mathbb{N}$, there are $t_n \approx t$ such that $G_n \in \mathcal{G}_{t_n}$. By saturation the set

$$\bigcap_{n\in\mathbb{N}} [t_n, t + 1/n]$$

is nonempty, and hence there is a $\tilde{t} \approx t$ such that $G_n \in \sigma(L(\mathcal{A}_{\tilde{t}}) \cup \mathcal{N})$ for all n. But then

$$\bigcup_{n\in\mathbb{N}} G_n \in \sigma(L(\mathcal{A}_{\tilde{t}}) \cup \mathcal{N}) \subset \mathcal{G}_t$$

and the proof is complete. □

Proof of Theorem 12.2: Assume first that x is a simple function

$$x(\omega,t) = \sum_{i=1}^{n} a_i 1_{F_i}(\omega) 1_{(s_i,t_i]}(t)$$

(the other kind of predictable rectangle, $B \times [0,t]$, is treated similarly). By the lemma, we can find elements $\tilde{t}_i \approx t_i$ and internal sets $A_i \in \mathcal{A}_{\tilde{t}_i}$ such that

$$P_L(F_i \bigtriangleup A_i) = 0$$

Since M is S-continuous, it is easy to see that

$$X(\omega,t) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega) 1_{(s_i,t_i]}(t)$$

is a 2-lifting of x and that $\int x\,dm$ is the standard part of $\int X\,dM$.

Assume now that x is a general element in $L^2(m)$. By lifting theory, we can find a square S-integrable lifting Y of x, the problem is only that there is no reason why such a lifting should be adapted. On the other hand, we can find a sequence x_n of simple processes such that

$$\|x - x_n\|_{L^2(\Omega\times[0,n],\mu_m)} < 1/n$$

for all natural numbers n. By what we have already shown, each x_n has an adapted 2-lifting X_n, and the sequence $\{X_n\}_{n\in\mathbb{N}}$ can be extended to an internal sequence $\{X_n\}_{n\in{}^*\mathbb{N}}$. For each finite n

$$\|Y - X_n\|_{L^2(\Omega\times T_n,\mu_M)} < 1/n$$

(check this, I am using a nontrivial relationship between μ_m and μ_M!) and by *overspill*, there must be an infinite N such that

$$\|Y - X_N\|_{L^2(\Omega\times T_N,\mu_M)} < 1/N$$

Since Y is a square S-integrable lifting of x, so is X_N, and we have hence shown that x has an adapted lifting $X = X_N$. In addition

$$\int_0^t x\,dm = \lim_{n\to\infty}\int_0^t x_n\,dm = \lim_{n\to\infty}\ {}^\circ\!\int_0^t X_n\,dM = {}^\circ\!\int_0^t X\,dM$$

which completes the proof. □

Corollary 12.4 *If M is a λ^2-martingale with infinitesimal increments and m is its standard part, then $[m]$ is the standard part of $[M]$.*

Proof: Since

$$[M]_t = M_t^2 - M_0^2 - 2\int_0^t M_s\,dM_s$$

and

$$[m]_t = m_t^2 - m_0^2 - 2\int_0^t m_s\,dm_s$$

it suffices to show that $\int_0^t m_s\,dm_s$ is the standard part of $\int_0^t M_s\,dM_s$. But this follows from the theorem as M is clearly a lifting of m (use stopping times if necessary to assure that M is sufficiently integrable). □

REMARK: In the results above we have assumed that M is continuous. Although this is convenient for technical reasons and makes the proofs run more smoothly, it is not really necessary for Theorem 12.2. All that is really needed for this theorem is that the martingale is S-continuous at 0 (otherwise the nonstandard martingale M will have a jump at the origin which the standard martingale m can not recapture—see [11], [15], [1], [30] for precise statements and proofs). For the corollary, things are more complicated since a discontinuous martingale M may have several noninfinitesimal jumps inside the same monad. The nonstandard quadratic variation $[M]$ will count these jumps separately, while the standard quadratic variation m can only feel their total effect. For this reasons, $[m]$ is usually not the standard part of $[M]$ in the discontinuous case. However, if we restrict ourselves to martingales which have at most one jump in each monad, then the equality holds (such martingales are called *well-behaved* in [15] and [1] and SDJ in [11] and [30]). It turns out that all processes with S-left and S-right limits can be made well-behaved by coarsening the timeline (see [11] or [30]).

So far it seems that standard and nonstandard stochastic integration are very closely connected. One differences, however, is worth pointing out: The standard filtration $\mathcal{F}_t = \sigma(\bigcup_{s \approx t} L(\mathcal{A}_s) \cup \mathcal{N})$ generated by the internal filtration $\mathcal{A}_t$ and the internal process M, is much larger than the standard filtration $\mathcal{G}_t = \sigma\{m(s) : s \leq t\}$ generated by the standard part $m = {}^\circ M$. In situations where the filtrations are important (e.g. in control theory), this distinction may cause problems (see the paper [6] by Cutland, Kopp and Willinger for a very interesting discussion of a closely related problem). As one example of how much the filtrations may mean, we include the following result from standard stochastic integration.

Theorem 12.5 (Martingale Representation Theorem) *Let $b : \Omega \times [0, \infty) \to \mathbb{R}$ be a Brownian motion on a probability space $(\Omega, \mathcal{G}, Q)$, and let $\{\mathcal{G}_t\}$ be the filtration generated by b. If m is an L^2-martingale adapted to $\{\mathcal{G}_t\}$, then m has a continuous version and can be written as a stochastic integral*

$$m_t = m_0 + \int_0^t g(\omega, s)\, db_s$$

where g is an adapted process which belongs to $L^2(\Omega \times [0, t], Q \times dt)$ for all t.

Proof: See, e.g., [28], [29] or [12]. □

If we let b in the theorem above be the standard part of Anderson's Brownian motion B, but replace the filtration $\{\mathcal{G}_t\}$ by the filtration $\mathcal{F}_t = \sigma(\bigcup_{s \approx t} L(\mathcal{A}_s) \cup \mathcal{N})$, the conclusion fails miserably—for an example, return to the setting of Section 3 (on Anderson's random walk) and let m be the standard part of the internal martingale M defined by $M(\omega, s) = 0$ for $s < 1$ and $M(\omega, s) = \omega(1)$ for $s \geq 1$.

PROBLEMS

12.1 *Show that the map $x \mapsto \int_0^t x\, dm$ is an isometry.*

12.2 *Give an example of an internal martingale M such that ${}^\circ[M] \neq [{}^\circ M]$.*

13. Stochastic Integrals in Higher Dimensions

As long as we are just interested in the *theory* of stochastic integration, there is very little difference between the one dimensional and the higher dimensional case, but when we turn to *applications*, most interesting models are higher dimensional. In this section we shall show that the one dimensional theory can be extended to higher dimensions with very little effort. Beginning with the nonstandard theory, we call an internal process $M : \Omega \times T \to^* \mathbb{R}^n$ a *martingale* if each component M_i is a martingale. If $X : \Omega \times T \to^* \mathbb{R}^m \otimes^* \mathbb{R}^n$ is an internal process taking values in the set ${}^*\mathbb{R}^m \otimes^* \mathbb{R}^n$ of nonstandard $m \times n$ matrices, we define the stochastic integral as

$$Y(t) = \int_0^t X\, dM = \sum_{s<t} X(s) \cdot \Delta M(s)$$

where $\cdot$ indicates matrix multiplication. This means that the the i-th component of Y is a sum of one-dimensional stochastic integrals

$$Y_i(t) = \sum_{j=1}^{n} \int_0^t X_{ij}(s)\, dM_j(s)$$

In the standard case, we start by copying this formula. If $m : \Omega \times [0, \infty) \to \mathbb{R}^n$ is an n-dimensional L^2-martingale (which just means that each component is a L^2-martingale) and x is a predictable process taking values in $\mathbb{R}^m \otimes \mathbb{R}^n$, we define the stochastic integral $y = (y_1, \ldots, y_m)$ by

$$y_i(t) = \sum_{j=1}^{n} \int_0^t x_{ij}(s)\, dm_j(s)$$

This formula makes sense if $x_{ij} \in L^2(m_j)$ (or $L(m_j)$) for each j, in which case we write $x \in L^2(m)$ (or $L(m)$). Similar extensions of the one-dimensional terminology are also used in the nonstandard context.

An *n-dimensional Brownian motion* is simply a process $b : \Omega \times [0, \infty) \to \mathbb{R}^n$ where the components are independent Brownian motions. The easiest way to model such a process in nonstandard terms, is to let independent copies of Anderson's random walk run along orthogonal axes. To be more precise, let $\Omega = ega = \{\omega : T \to \{-1, 1\}^n \mid \omega \text{ internal}\}$, and define an internal process $B : \Omega \times T \to^* \mathbb{R}^n$ by

$$B(\omega, t) = \sum_{i=1}^{n} \Big(\sum_{s<t} \omega_i(s)\sqrt{\Delta t}\Big)\, \mathbf{e}_i$$

where $\mathbf{e}_i$ is the i-th element of the standard basis in $\mathbb{R}^n$.

PROBLEMS

13.1 *(i) Let $M : \Omega \times T \to^* \mathbb{R}^n$ be an S-continuous, n-dimensional λ^2-martingale, and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Prove an Itô formula for ${}^*\phi(M_t)$ (you will need the joint variations*

$[M_i, M_j]$ from Problems 6.3 and 8.2).
(ii) Prove the following multi-dimensional version of Theorem 11.1: Assume that $M : \Omega \times T \to {}^*\mathbb{R}^n$ is a λ^2-martingale such that
a) $M(0) = 0$.
b) $M(\omega, \cdot)$ is S-continuous for almost all ω and all finite t.
c) $[M_i, M_j](t) \approx \delta_{ij} t$ for almost all ω.
Then the standard part of M is a Brownian motion.

14. Stochastic Differential Equations

In the final three sections of this paper, we shall take a look at some applications which, I hope, will give you a better feeling for how the theory of stochastic integration is used in practice. We shall start by taking a brief look at stochastic differential equations, a subject that is covered in much greater depth and detail in Keisler's contribution to this volume.

Just as ordinary and partial differential equations are the main tools for constructing deterministic mathematical models, stochastic differential equations are the most flexible tools for building indeterministic models. You can find a good (standard) introduction with lots of applications in Øksendal's book [24]. As one would expect, there are many kinds of stochastic differential equations, but we shall stick to the most common ones - the *Itô equations*:

$$x_t = x_0 + \int_0^t f(s, x(s))\, ds + \int_0^s g(s, x(s))\, db(s)$$

In such an equation, x_0 is a given point in $\mathbb{R}^m$, $f : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}^m$ and $g : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}^m \otimes \mathbb{R}^n$ are given functions, $b : \Omega \times [0,\infty) \to \mathbb{R}^n$ is an n-dimensional Brownian motion, and $x : \Omega \times [0,\infty) \to \mathbb{R}^m$ is an unknown stochastic process. The fundamental problem is, of course, to show that under reasonable conditions, the equation has a (unique) solution. Itô equations have been used to model a host of problems in physics, electrical engineering, biology, economics, etc.

If we let b be the standard part of an Anderson process B and assume that the coefficients f and g are continuous and bounded, we can prove the existence of a solution very easily. Consider the internal difference equation

$$X(t) = x_0 + \sum_{s<t} {}^*f(s, X(s))\, \Delta t + \sum_{s<t} {}^*g(s, X(s))\, \Delta B(s)$$

By induction, this equation clearly has a unique solution X, and the idea is simply that the standard part $x = {}^\circ X$ will be a solution of the stochastic differential equation. To check this, we first observe that ${}^*g(s, X(\omega, s))$ is a lifting of $g(s, x(s))$, and hence $\int g(s, x(s))\, db(s)$ is the standard part of

$\sum {}^*g(s, X(s))\,\Delta B(s)$. Similarly, one may check that $\int f(s, x(s))\,ds$ is the standard part of $\sum {}^*f(s, X(s))\,\Delta t$. Combining these two observations, we get that x is the standard part of X and we have reached our goal.

Theorem 14.1 *Let b be the standard part of Anderson's random walk and assume that $f : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}^m$ and $g : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}^m \otimes \mathbb{R}^n$ are continuous and bounded functions. For each $x_0 \in \mathbb{R}^n$ the stochastic differential equation*

$$x(t) = x_0 + \int_0^t f(s, x(s))\,ds + \int_0^s g(s, x(s))\,db(s)$$

has a solution.

If the coefficients f and g fail to be continuous, we have to be smarter as we can no longer use ${}^*f$ and ${}^*g$ as nonstandard representations of f and g, but have to turn to lifting theory. See [1] and, especially, [13] for expositions.

PROBLEMS

14.1 *Assume that $f : [0,\infty) \to \mathbb{R}$ and $g : [0,\infty) \to \mathbb{R}$ are continuous and bounded functions, and show that*

$$x(t) = x_0 e^{\int_0^t f(s)\,ds + \int_0^t g(s)\,db_s - \frac{1}{2}\int_0^t g(s)^2\,ds}$$

is a solution of the stochastic differential equation

$$x_t = x_0 + \int_0^t f(s)x(s)\,ds + \int_0^t g(s)x(s)\,db(s)$$

14.2 *Let $B : \Omega \times T \to^* \mathbb{R}$ be Anderson's random walk, and consider an S-bounded, internal function $F : T \times^* \mathbb{R}^n \to^* \mathbb{R}$. We replace the usual probability measure P with a new measure Q defined informally as follows: If the process is at the point x at time t, then*

$$\Delta B(\omega, t) = \begin{cases} \sqrt{\Delta t} & \text{with probability } \frac{1}{2} + \frac{1}{2}F(t,x)\sqrt{\Delta t} \\ -\sqrt{\Delta t} & \text{with probability } \frac{1}{2} - \frac{1}{2}F(t,x)\sqrt{\Delta t} \end{cases}$$

(i) Show that

$$B(\omega, t) = \sum_{s<t} F(s, X(\omega, s))\Delta t + \tilde{B}(\omega, t)$$

where the standard part $\tilde{b}$ of $\tilde{B}$ is a Brownian motion on $(\Omega, \mathcal{A}_L, Q_L)$.
(ii) Show that if $F =^* f$ for a bounded, continuous function f, then the standard part b of B is a solution of the stochastic differential equation

$$x(t) = \int_0^t f(s, x(s))\,ds + \tilde{b}(t)$$

on the space $(\Omega, \mathcal{A}_L, Q_L)$. Hence we may solve some stochastic differential equations by modifying the underlying measure and not the process.
(iii) Show that for almost all ω

$$\frac{Q([\omega]_t)}{P([\omega]_t)} = e^{\int_0^t F(s,B_s)\,dB_s - \frac{1}{2}\int_0^t F(s,B_s)^2\,ds}$$

15. Brownian Local Time

Choose a time $t \in [0,\infty)$ and a point $x \in \mathbb{R}$. How much time has the Brownian path $s \mapsto b(\omega, s)$ spent at x before time t, i.e. how large is the set

$$\{s \leq t \mid b(\omega, s) = x\} \quad ?$$

The obvious answer is that the set has Lebesgue measure zero, but although this answer is correct, it is not very informative. The reason is that we have tried to measure the set on too coarse a scale, and we shall now see how we can get a more useful answer by using a finer scale.

Let us first ask the same question about Anderson's random walk. If $t \in T$ and

$$x \in \Gamma := \{k\sqrt{\Delta t} \mid k \in {}^*\mathbb{R}\,\}$$

how large is the set

$$A(\omega, x, t) = \{s \leq t \mid B(\omega, s) = x\,\} \quad ?$$

In this discrete setting, it is easy to make an estimate of the size of $A(\omega, x, t)$: There are order of magnitude Δt^{-1} points s in T smaller than t, while the corresponding values $b(\omega, s)$ are distributed among order of magnitude Δx^{-1} points in Γ. Hence the number of points in $A(\omega, x, t)$ should be of order $\Delta t^{-1}/\Delta x^{-1} = \Delta t^{-1/2}$. Instead of looking at $A(\omega, x, t)$, we should therefore consider the *internal local time*

$$L(\omega, x, t) = |A(\omega, x, t)|\sqrt{\Delta t} = |\{s \leq t \mid B(\omega, s) = x\,\}|\,\sqrt{\Delta t}$$

In order to get a better handle on L, we first introduce the following version of the *signum function*

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

(the value at the origin is significant as you soon will see). For fixed $x \in \Gamma$, let X be the process $X(\omega, t) = |B(\omega, t) - x|$, and observe that

$$\Delta X(\omega, t) = \begin{cases} sgn(B(\omega, t) - x)\,\Delta B(\omega, t) & \text{if } B(\omega, t) \neq x \\ \sqrt{\Delta t} & \text{if } B(\omega, t) = x \end{cases}$$

This means that

$$\begin{aligned} X(\omega,t) &= X_0(\omega)+\sum_{s<t}\Delta X(\omega,s) \\ &= |-x|+\int_0^t sgn(B(\omega,s)-x)\,dB(\omega,s)+L(\omega,x,t) \end{aligned}$$

Hence

$$L(\omega,x,t)=-|x|+|B(\omega,t)-x|-\int_0^t sgn(B(\omega,s)-x)\,dB(\omega,s)$$

From this equation it follows immediately that for each fixed x, the function $t\mapsto L(\omega,x,t)$ is S-continuous (with some work one may show that L is actually jointly continuous in x and t). If we let b denote the standard part of B, and let l be the standard part of L (what do I mean by this?), then

$$l(\omega,x,t)=-|x|+|b(\omega,t)-x|-\int_0^t sgn(b(\omega,s)-x)\,db(\omega,s)$$

which is known as *Tanaka's formula.* To prove this formula, you only need to check that $sgn(B(s)-x)$ is a lifting of $sgn(b(s)-x)$. In standard texts, Tanaka's formula is often used as a definition of Brownian local time. For more information about the nonstandard theory, consult Perkins' paper [25].

PROBLEMS

15.1 *(i) Show that for all finite $t\in T$ and $x\in\Gamma$ the set*

$$\{(\omega,s)\mid s<t \text{ and } B(\omega,s)\approx x\}$$

has $L(\nu_{B_t})$-measure zero.
(ii) Show that $sgn(B(s)-x)$ is a lifting of $sgn(b(s)-x)$.
(iii) Show that $L(\omega,x,t)$ is jointly S-continuous in x and t for almost all ω.

16. The Infinite Dimensional Ornstein-Uhlenbeck Process

One of the hottest topics in probability theory over the last few years, has been the Malliavin calculus and related theories of infinite dimensional analysis. The most important process in these theories is the infinite dimensional Ornstein-Uhlenbeck process. In this section I shall indicate briefly how this process can be constructed from Anderson's random walk. You will find more details in [19].

The infinite dimensional Ornstein-Uhlenbeck process u is a stochastic process taking values in the space $C([0,1])$ of continuous functions. Intuitively, it looks like a continuous, random modification of Brownian paths which keeps the Wiener measure invariant. Using Anderson's construction we can make this intuition rigorous in the following way.

Let Ω be the probability space in Anderson's construction, but assume for simplicity that the timeline stops at 1, i.e. $T = \{0, \Delta t, 2\Delta t, \dots, 1\}$. Pick an initial element ω_0 in Ω (and do it in such a way that $B(\omega_0, \cdot)$ is S-continuous). At time 0, toss an unfair coin for each $s \in T$ to decide whether you want to reverse the sign of s-th component $\omega_0(s)$ or not; the probability of changing the sign should be $\Delta t/2$ and the probability of keeping it $(1 - \Delta t/2)$. The resulting path is $\omega_{\Delta t}$. At time Δt repeat the procedure; flip independent coins for each s and change the sign of each component with probability $\Delta t/2$ to obtain $\omega_{2\Delta t}$. Continuing in this way, we get a random sequence of elements $\omega_0, \omega_{\Delta t}, \dots, \omega_{n\Delta t}, \dots$ in Ω. Each of these elements corresponds to an Anderson path $B(\omega_{n\Delta t}, \cdot)$, and hence we have constructed a randomly moving sequence of such paths. Taking standard parts, we get a randomly moving process of Brownian paths. The standard part of this random motion turns out to be an infinite dimensional Ornstein-Uhlenbeck process. Note that this process is a combination of our two basic processes, the Poisson process and Brownian motion; the Brownian path is changed by letting independent Poisson processes act on each increment.

To formalize the construction sketched above, we define the space Ξ of all internal function

$$\xi : T \to \{-1, 1\}^T$$

where $\{-1,1\}^T$ denotes the space of all *internal* functions from the timeline T to $\{-1,1\}$. We shall think of $\xi(t)$ as the sequence of coin tosses performed at time t, and $\xi(t)(s) = -1$ will represent the event of switching the s-th component at time t. Let Q be the internal probability measure on Ξ which makes all the events $\{\xi(t)(s)\}_{t,s \in T}$ independent, and which gives the event $\xi(t)(s) = -1$ probability $\Delta t/2$ (and, consequently, the event $\xi(t)(s) = 1$ probability $1 - \Delta t/2$). Given an initial value ω_0, we define a process

$$\Theta^{\omega_0} : \Xi \times T \to \Omega$$

by

$$\Theta^{\omega_0}(\xi, t + \Delta t)(s) = \xi(t)(s)\Theta^{\omega_0}(\xi, t)(s)$$

i.e., the state changes from t to $t + \Delta t$ by switching the s-th coordinate if and only if $\xi(t)(s) = -1$. When the initial value ω_0 is of little interest, we drop the superscript and write Θ for Θ^{ω_0}. We now define the *hyperfinite Ornstein-Uhlenbeck process* U $(= U^{\omega_0})$ by

$$U(\xi, t)(\cdot) = B(\Theta(\xi, t), \cdot)$$

It is quite instructive (and a very good exercise in using the tools developed in this paper) to show that U is a continuous process whose standard part is an infinite dimensional Ornstein-Uhlenbeck process (see [19]). If you are more interested in a nonstandard introduction to Malliavin calculus, you should look up [7] which is set in a slightly different framework.

PROBLEMS

16.1 *Show that for all $s, t \in T$ and $\omega_0 \in \Omega$*

$$Q\{\xi \in \Xi | \Theta^{\omega_0}(\xi, t)(s) = \omega_0(s)\} = [(1 - \Delta t)^{t/\Delta t} + 1]/2 \approx (e^{-t} + 1)/2$$

16.2 *Fix t and ω_0. Show that*

$$\begin{aligned} E_Q[U(\xi, t)(s + \Delta t) - U(\xi, t)(s) | \mathcal{F}_s] &= \omega_0(s)\sqrt{\Delta t}\,(1 - \Delta t)^{t/\Delta t} \\ &= \omega_0(s)\sqrt{\Delta t}\, e^{-t} + o(\sqrt{\Delta t}) \end{aligned}$$

where $o(\sqrt{\Delta t})$ denotes a quantity which is infinitesimal compared to $\sqrt{\Delta t}$.

16.3 *Fix ω_0 and t. Show that*

$$U(\xi, t)(s) = (1 - \Delta t)^{t/\Delta t}(U(\xi, 0)(s) + K^{(\omega_0, t)}(\xi, s))$$

where $(\xi, s) \mapsto K^{(\omega_0,t)}(\xi, s)$ is an S-continuous martingale such that the standard part of $K^{(\omega_0,t)}/\sqrt{e^{2t} - 1}$ is a Brownian motion. Moreover, for all $s, s' \in T$

$$[K^{(\omega_0,t)}](\xi, s) - [K^{(\omega_0,t)}](\xi, s') \leq 4e^2|s - s'|$$

where $[K^{(\omega_0,t)}]$ denotes the quadratic variation of $K^{(\omega_0,t)}$.

16.4 *The proof of the Markov property (part of Theorem 4.2) in [19] is hopelessly wrong. Give a correct proof.*

Suggestions for Further Study

The books [1] and [30] contain a rather complete exposition of nonstandard stochastic analysis up to 1985, but you should also consult [13], [25], [15], and [11] for additional information and insights. Nelson's beautiful little book [21] treats nonstandard analysis from a 'naive' point of view, but at this stage you should have no problems filling in the formal details. As for more recent developments, three of the most important ones are covered by other contributions to this volume: Kopp discusses the fashionable field of mathematical finance, Keisler shows how the methods of neo-compactness can be used to prove existence results for stochastic differential equations, and Capiński gives a wonderful exposition of nonstandard methods in fluid dynamics. If I may add a personal opinion, Capiński's and Cutland's work on the last topic is the most important and exciting contribution to nonstandard analysis in recent years, and you should not fail to look up the full treatment of their ideas in [4]. Another popular field which seems to lend itself naturally to nonstandard techniques, is the theory of *super processes*, see, e.g., Perkins' construction of the *historical process* in [26] (a fuller—standard—treatment is given in [8]). Another very promising development is Ponosov's work on stochastic dynamical systems [27]. Nonstandard Dirichlet forms and Markov processes were first studied in [1] and this work has been continued by Fan (see, e.g., [10]), while Wu has been interested in a number of questions in infinite dimensional analysis (see, e.g., [31] and the joint paper with Albeverio [2]). Let me finally mention work on Brownian motion on fractals by Nyberg [22], [23] and myself [17], [18].

References

1. Albeverio, S., Fenstad, J.E., Høegh-Krohn, R., and Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics*, Academic Press, New York.
2. Albeverio, S. and Wu, J.-L. (1995) Nonstandard flat integral representation of the Euclidean field and a large deviation bound for the exponential interaction, in Cutland, N.J. et al. *Developments in Nonstandard Mathematics*, Longman, Harlow, pp. 198-210.
3. Anderson, R.M. (1976) A nonstandard representation for Brownian motion and Itô integration, *Israel J. Math.* **25**, pp. 15–46.
4. Capiński, M. and Cutland, N.J. (1995) *Nonstandard Methods for Stochastic Fluid Dynamics*, World Scientific, Singapore.
5. Chung, K.-L. and Williams, R.J. (1983) *Introduction to Stochastic Integration*, Birkhäuser, Boston.
6. Cutland, N.J., Kopp, P.E., and Willinger, W. (1995) From discrete to continuous stochastic calculus, *Stochastics and Stochastic Reports* **52**, pp. 173-192.
7. Cutland, N.J. and Ng, S.-A. (1995) A nonstandard approach to the Malliavin Calculus, in Albeverio, S. et al. *Advances in Analysis, Probability and Mathematical Physics*, Kluwer Academic Publishers, Dordrecht, pp. 149-170.
8. Dawson, D.A. and Perkins, E.A. (1991) Historical processes, *Mem. Amer. Math. Soc.* **454**.
9. Dellacherie, C. and Meyer, P.-A. (1976-80-83-87) *Probabilités et Potentiel A,B,C,D*, Hermann, Paris (an English translation, *Probability and Potential*, is published by North-Holland, Amsterdam).
10. Fan, R.-Z. (1996) Potential theory for hyperfinite Dirichlet forms, *Potential Analysis* **5**, pp. 417-462.
11. Hoover, D.N. and Perkins, E.A. (1983) Nonstandard construction of the stochastic integral and applications to stochastic differential equations, *Trans. Amer. Math. Soc.* **286**, pp. 1-58.
12. Karatzas, I. and Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd Edition, Springer-Verlag, New York.
13. Keisler, H.J. (1984) An infinitesimal approach to stochastic analysis, *Mem. Amer. Math. Soc.* **297**.
14. Kopp, P.E. (1984) *Martingales and Stochastic Integrals*, Cambridge University

Press, Cambridge.
15. Lindstrøm, T. (1980) Hyperfinite stochastic integration I-III, *Math. Scand.* **46**, pp. 265-333.
16. Lindstrøm, T. (1983) Stochastic integration in hyperfinite dimensional linear spaces, in Hurd, A.E.(ed.) *Nonstandard Analysis—Recent Developments*, LNM 983, Springer-Verlag, Berlin, pp. 134-161.
17. Lindstrøm, T. (1990) Brownian motion on nested fractals, *Mem. Amer. Math. Soc.* **420**.
18. Lindstrøm, T. (1993) Brownian motion penetrating the Sierpinski gasket, in Elworthy, K.D and Ikeda, N. *Asymptotic Problems in Probability Theory: Stochastic Models and Diffusions on Fractals*, Longman, Harlow, pp. 248-278.
19. Lindstrøm, T. (1995) Anderson's random walk and the infinite dimensional Ornstein-Uhlenbeck process, in Albeverio, S. et al. *Advances in Analysis, Probability and Mathematical Physics*, Kluwer Academic Publishers, Dordrecht, pp. 186-199.
20. Loeb, P.A. (1975) Conversion from nonstandard to standard measure spaces and applications in probability theory, *Trans. Amer. Math. Soc.* **211**, pp. 113-122.
21. Nelson, E. (1987) *Radically Elementary Probability Theory*, Princeton University Press, Princeton.
22. Nyberg, S.O. (1995) Brownian motion on simple fractal spaces, *Stochastics and Stochastics Reports* **55**, pp. 21-45.
23. Nyberg, S.O. (1996) *Brownian Motion on Simple Fractal Spaces*, Doctoral dissertation, University of Oslo.
24. Øksendal, B. (1996) *Stochastic Differential Equations*, 4th Edition, Springer-Verlag, Berlin.
25. Perkins, E.A. (1981) A global intrinsic characterization of local time, *Ann. Prob.* **9**, pp. 800-817.
26. Perkins, E.A. (1988) A space-time property of a class of measure-valued branching diffusions, *Trans. Amer. Math. Soc.* **305**, pp. 743-795.
27. Ponosov, A. (1995) Two applications of NSA in the theory of stochastic dynamical systems, in Albeverio, S. et al. *Advances in Analysis, Probability and Mathematical Physics*, Kluwer Academic Publishers, Dordrecht, pp. 200-211.
28. Revuz, D. and Yor, M. (1991) *Continuous Martingales and Brownian Motion*, Springer-Verlag, Berlin.
29. Rogers, L.G.C. and Williams, D. (1987) *Diffusions, Markov Processes, and Martingales, Volume 2, Itô Calculus*, John Wiley & Sons, Chichester.
30. Stroyan, K.D. and Bayod, J.M. (1986) *Foundations of Infinitesimal Stochastic Analysis*, North-Holland, Amsterdam.
31. Wu, J.-L. (1994) On the regularity of stochastic difference equations in hyperfinite-dimensional vector spaces and applications to $\mathcal{D}'$-valued stochastic differential equations, *Proc. Roy. Soc. Edin.* **124A**, pp. 1089-1117.

STOCHASTIC DIFFERENTIAL EQUATIONS WITH EXTRA PROPERTIES

H. JEROME KEISLER
Department of Mathematics
University of Wisconsin
Madison, WI 53706
USA
email: keisler@math.wisc.edu

1. Introduction

The Loeb measure construction has been a powerful tool in proving existence theorems for stochastic differential equations. There are many strong existence theorems which depend on the richness of the adapted Loeb space and which cannot be proved by classical methods. See, for example, [1].

In these lectures we shall first use the method to show that solutions exist. We shall then exploit the method further to find solutions of stochastic differential equations with additional properties, such as solutions which are optimal in a variety of ways, and solutions which are Markov processes.

In most cases, a nonstandard existence proof shows more than mere existence of a solution—it also gives a characterization of the set of all solutions in terms of liftings. By the monad of a set C of stochastic processes we shall mean the set of all liftings of elements of C. A typical lifting theorem will show that the monad of the set of all solutions of the stochastic differential equation under consideration is a countable intersection of internal sets.

These lifting theorems draw their power from the fact that sets C whose monads are countable intersections of internal sets behave much like compact sets. For this reason, we call a set whose monad is a countable intersection of internal sets a **neocompact set**.

Some of the ideas developed here go back to the monograph [5], where several existence theorems for stochastic differential equations with extra properties were obtained. We are now taking another look at these ideas

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 259–277.

in the light of more recent developments. The notion of a neocompact set captures a common thread which appears in many proofs both in [5] and in the more recent literature.

In these lectures we use neocompact sets in the "conventional" nonstandard setting. In a recent series of papers (see [6] for a survey), the neocompact sets are instead taken as a primitive notion and used to prove existence theorems directly—avoiding the steps of lifting to the nonstandard universe and coming back down to the standard universe.

2. Spaces of Stochastic Processes

We begin by fixing notation and setting up a framework which is appropriate for studying liftings of stochastic processes. For simplicity we shall restrict time to the closed interval $[0,1]$. In the spirit of the previous lectures in this conference, we shall concentrate on square-integrable stochastic processes. We first look at liftings of random variables with values in a metric space $\mathcal{M}$, and then use the fact that a continuous or L^2 stochastic process with values in $\mathcal{M}$ is the same thing as a random variable with values in the metric space $C([0,1],\mathcal{M})$ or $L^2([0,1],\mathcal{M})$.

Let

$$T = \{0, \Delta t, 2\Delta t, \ldots, H\Delta t = 1\}$$

be a hyperfinite time line where H is an infinite hyperinteger and $\Delta t = 1/H$. Our sample space $\Omega = \Omega_0^T$ will be the set of all internal functions from T into Ω_0 where Ω_0 is a *finite set with at least two elements. Let P be the hyperfinite counting measure on Ω, so that every internal set A is P-measurable and $P(A) = |A|/|\Omega|$. P_L will denote the Loeb measure generated by P. For $\omega \in \Omega$ and $t \in T$ let

$$[\omega]_t = \{\alpha \in \Omega : \alpha(s) = \omega(s) \text{ for all } s < t\}.$$

Let $\mathcal{G}_t$ be the *-algebra composed of all internal sets A such that $[\omega]_t \subseteq A$ for all $\omega \in A$, and let $\sigma(\mathcal{G}_t)$ be the P_L-complete σ-algebra generated by $\mathcal{G}_t$. For $t \in [0,1)$ let

$$\mathcal{F}_t = \bigcap\{\sigma(\mathcal{G}_s) : {}^\circ s > t\},$$

and let $\mathcal{F}_1 = \sigma(\mathcal{G}_1)$.

We let $(\mathcal{M},\rho)$, $(\mathcal{N},\pi),\ldots$ be standard complete separable metric spaces. Let us pick out an element $m_0 \in \mathcal{M}$. The metric space $L^2(\Omega,\mathcal{M})$ is the space of all Loeb measurable random variables $x : \Omega \to \mathcal{M}$ such that $(\rho(x(\omega), m_0))^2$ is integrable, with the metric ρ_2 defined by

$$\rho_2(x,y) = \left[\int (\rho(x(\omega), y(\omega))^2 d\omega\right]^{1/2}.$$

We identify each $m \in \mathcal{M}$ with the constant function from Ω to m, so that $\mathcal{M} \subseteq L^2(\Omega, \mathcal{M})$. If $A \subseteq L^2(\Omega, \mathcal{M})$ and $r \in \mathbb{R}$, we let A^r be the set of all x such that $\rho_2(x, y) \le r$ for some $y \in A$.

We also need an internal counterpart of $L^2(\Omega, \mathcal{M})$. To give us some flexibility, we first let $\mathcal{M}'$ be an internal subset of ${}^*\mathcal{M}$ which is S-dense, that is, every point of ${}^*\mathcal{M}$ is infinitely close to some point of $\mathcal{M}'$. We now define $SL^2(\Omega, \mathcal{M})$ as the internal set consisting of all internal functions $X : \Omega \to \mathcal{M}'$, with the internal metric

$$\bar{\rho}_2(X, Y) = \left[\sum({}^*\rho(X(\omega), Y(\omega))^2 \Delta\omega\right]^{1/2}.$$

Let $X \in SL^2(\Omega, \mathcal{M})$ and $x : \Omega \to \mathcal{M}$. We shall say that X is **S^2-integrable** if $({}^*\rho(X(\omega), m_0))^2$ is S-integrable over Ω. We say that X **lifts** x, and that x is the **standard part** of X (in symbols $x = {}^\circ X$), if $X(\omega) \approx x(\omega)$ P_L-almost surely and X is S^2- integrable. If X has a standard part, we say that X is **near-standard** and write $X \in ns^2(\Omega, \mathcal{M})$.

The following proposition, which follows from the fundamental results in the earlier lectures, gives the connection between the standard part map and the spaces $L^2(\Omega, \mathcal{M})$.

2.1. Proposition. *(Loeb [9] and Anderson [2]). $L^2(\Omega, \mathcal{M})$ is the set of all standard parts of elements of $ns^2(\Omega, \mathcal{M})$.* □

We shall extend the standard part terminology to sets. For a set $A \subseteq SL^2(\Omega, \mathcal{M})$, the **standard part** of A is defined as the set

$${}^\circ A = \{{}^\circ X : X \in A \cap ns^2(\Omega, \mathcal{M})\}$$

of standard parts of near-standard elements of A. We say that A is near-standard if every element of A is near-standard. In the upward direction, the **monad** of a set $B \subseteq L^2(\Omega, \mathcal{M})$ is the set of all $X \in ns^2(\Omega, \mathcal{M})$ such that ${}^\circ X \in B$.

We next apply our setup to spaces of L^2 stochastic processes and of continuous stochastic processes.

We first consider L^2 processes. Let $\mathcal{L}(\mathcal{M}) = L^2([0,1], \mathcal{M})$ be the space of L^2 paths in $\mathcal{M}$. Thus $L^2(\Omega, \mathcal{L}(\mathcal{M}))$ is the space of L^2 stochastic processes with values in $\mathcal{M}$. In this case we take the internal set $\mathcal{L}(\mathcal{M})'$ to be the set of all T-step functions induced by internal functions $X : T \to {}^*\mathcal{M}$. This set is S-dense in ${}^*(L^2([0,1], \mathcal{M}))$ as required. Then $ns^2(\Omega, \mathcal{L}(\mathcal{M}))$ turns out to be the set of all X such that $X(\omega, t)$ is S^2-integrable over $\Omega \times T$ and near-standard in ${}^*\mathcal{M}$ almost everywhere in $\Omega \times T$.

We now consider continuous processes. Let

$$\mathcal{C}(\mathcal{M}) = C([0,1], \mathcal{M})$$

be the space of continuous paths in $\mathcal{M}$ with the sup metric, and assume that $\mathcal{M}$ is a linear space. Then $L^2(\Omega, \mathcal{C}(\mathcal{M}))$ is the space of L^2 continuous stochastic processes with values in $\mathcal{M}$. This time we take the internal set $\mathcal{C}(\mathcal{M})'$ to be the set of all polygonal paths induced by internal functions $X : T \to {}^*\mathcal{M}$. This set is again S-dense. $ns^2(\Omega, \mathcal{C}(\mathcal{M}))$ is the set of all X such that $X(\omega)$ is S-continuous P_L-almost surely and is S^2-integrable over Ω.

Another space which is often used for the paths of a stochastic process is the space $D([0,1], \mathbb{R}^d)$ of right continuous functions with left limits and the Skorokhod metric. In the interest of simplicity, we shall avoid that space in these lectures.

We are now ready to study liftings of stochastic processes in a systematic way.

By an **adapted process** in $\mathcal{M}$ we shall mean a stochastic process $x \in L^2(\Omega, \mathcal{L}(\mathcal{M}))$ such that $x(\omega, t)$ is $\mathcal{F}_t$-measurable for each $t \in [0,1]$. A **continuous adapted process** in $\mathcal{M}$ is defined similarly but with $x \in L^2(\Omega, \mathcal{C}(\mathcal{M}))$. A **continuous martingale** in $\mathbb{R}^d$ is a continuous adapted process x in $\mathbb{R}^d$ such that $E[x(\bullet, t)|\mathcal{F}_s] = x(\omega, s)$ whenever $s \leq t$.

An internal stochastic process $X \in SL^2(\Omega, \mathcal{L}(\mathcal{M}))$ or $X \in SL^2(\Omega, \mathcal{C}(\mathcal{M}))$ will be called **adapted after** r if $X(\omega, s)$ is $\mathcal{G}_t$-measurable whenever $s \leq t \in T$ and $r \leq t$, and called **adapted** if it is adapted after $1/n$ for each $n \in \mathbb{N}$. X is called a **martingale after** r if X is adapted after r and $E[X(\bullet, t)|\mathcal{G}_s] = X(\omega, s)$ whenever $r \leq s \leq t$, and a **martingale** if it is a martingale after each $1/n$.

We shall need the following lifting lemma which gives the connection between the standard notions of an adapted process and martingale and the nonstandard counterparts of these notions. We shall leave this lemma as an exercise for the reader, with a warning that the proof is not as easy as one would expect!

2.2. Lemma. *(i) A process $x \in L^2(\Omega, \mathcal{L}(\mathcal{M}))$ is adapted in $\mathcal{M}$ if and only if x has an adapted lifting $X \in SL^2(\Omega, \mathcal{L}(\mathcal{M}))$.*

(ii) A process $x \in L^2(\Omega, \mathcal{C}(\mathcal{M}))$ is continuous adapted in $\mathcal{M}$ if and only if x has an adapted lifting $X \in SL^2(\Omega, \mathcal{C}(\mathcal{M}))$.

(iii) A process $x \in L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$ is a martingale in $\mathbb{R}^d$ if and only if x has an adapted lifting $X \in SL^2(\Omega, \mathcal{C}(\mathbb{R}^d))$ which is a martingale. □

Parts (i) and (ii) are proved in [5], and part (iii) is due to Hoover, Perkins, and Lindstrøm, (see [1]). Going up, the idea in the proof is to start with a lifting and modify it on a set of measure zero to a lifting which is adapted after $1/n$ for each n. Going down, the idea is to start with a standard part and modify it on a set of measure zero to an adapted process.

The following result is a lifting theorem for stochastic integrals. To avoid the complication of introducing $SL^2(w)$ liftings, we restrict our discussion to the case of uniformly bounded integrands.

2.3. Proposition. *(Anderson [2] for Brownian motions; Hoover, Perkins, and Lindstrøm in general). Suppose that $f \in L^2(\Omega, \mathcal{L}(\mathbb{R}^{d\times d}))$ is uniformly bounded and adapted and $w \in L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$ is a continuous martingale. Then for any uniformly bounded adapted lifting F of f and any martingale lifting W of w, the hyperfinite sum*

$$S(\omega,t) = \sum_{s<t} F(\omega,s)\Delta W(\omega,s)$$

is a lifting of the stochastic integral

$$I(\omega,t) = \int_0^t f(\omega,s)dw(\omega,s)$$

in the space $L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$.

Sketch of Proof: The hyperfinite sum $S(\omega,t)$ is S-continuous by Lindstrøm Theorem 9.5 (this volume [8]). Since F is uniformly bounded, one can check that $S(\omega,t)$ is also SL^2, and hence near-standard. By Lindstrøm [8] Theorem 12.2, $S(\omega,t)$ is a lifting of the stochastic integral $I(\omega,t)$.

The idea of the proof of this last fact is as follows. For any sequence of adapted step functions f_n converging to f in $L^2(\Omega, \mathcal{L}(\mathbb{R}^d))$, the stochastic integrals

$$\int_0^t f_n(\omega,s)dw(\omega,s)$$

are defined in the natural way and can be shown to be convergent in $L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$. The limit of this sequence is the standard definition of the stochastic integral

$$\int_0^t f(\omega,s)dw(\omega,s).$$

Taking F_n to be a step function lifting f_n, the hyperfinite sums

$$\sum_{s<t} F_n(\omega,s)\Delta W(\omega,s)$$

S-converge to $S(\omega,t)$, and it follows that $S(\omega,t)$ lifts $I(\omega,t)$. $\square$

3. Solutions of Stochastic Differential Equations

To motivate our approach to solving stochastic differential equations, let us examine the simplest existence theorem for stochastic differential equations in [5]. Let $C(\mathbb{R}^d, \mathbb{R}^{d\times d})$ be the space of all continuous functions from $\mathbb{R}^d$ into $\mathbb{R}^{d\times d}$ with a metric for the topology of uniform convergence on compact sets.

3.1. Theorem. *Let $w(\omega, s)$ be a continuous martingale in $\mathbb{R}^d$ and let*

$$g \in L^2(\Omega, \mathcal{L}(C(\mathbb{R}^d, \mathbb{R}^{d\times d})))$$

be a uniformly bounded adapted process with values in the space $C(\mathbb{R}^d, \mathbb{R}^{d\times d})$. Then there exists a continuous martingale x in $\mathbb{R}^d$ such that

$$x(\omega, t) = \int_0^t g(\omega, s)(x(\omega, s))dw(\omega, s). \tag{1}$$

Proof: By Lemma 2.2, g has an adapted lifting G and w has a martingale lifting W. $W(\omega, \bullet)$ is S-continuous P_L-almost surely. It follows from Lindstrøm's article ([8] Theorem 8.2) that the quadratic variation $[W]$ is S-continuous P_L-almost surely. By truncating we may take G to have the same finite bound as g. Define $X(\omega, t)$ as the unique solution of the hyperfinite difference equation

$$X(\omega, t) = \sum_{s<t} G(\omega, s)(X(\omega, s))\Delta W(\omega, s). \tag{2}$$

X is clearly an internal martingale. Since $[W]$ is S-continuous and G is bounded, $[X]$ is S-continuous, and therefore X is S-continuous. Similarly, since W is S^2-integrable, one can show that X is S^2-integrable. Therefore X is near-standard and has a standard part x which is a continuous martingale in $\mathbb{R}^d$. Furthermore,

$$°G(\omega, t)(X(\omega, t)) = g(\omega, °t)(x(\omega, °t))$$

almost surely in $\Omega \times T$. By Proposition 2.3,

$$\sum_{s<t} G(\omega, s)(X(\omega, s))\Delta W(\omega, s)$$

lifts

$$\int_0^t g(\omega, s)(x(\omega, s))dw(\omega, s).$$

Taking standard parts we see that x is a solution of the original equation (1). □

This proof actually gives a characterization of the set C of all solutions of (1). Let $\hat{C}$ be the set of all $X \in SL^2(\Omega, C(\mathbb{R}^d))$ such that

$$\bar{\rho}_2 \left(X(\omega, t), \sum_{s<t} G(\omega, s)(X(\omega, s))\Delta W(\omega, s) \right) \approx 0,$$

and

$$(\exists Y)[Y \text{ is adapted and } \bar{\rho}_2(X,Y) \approx 0].$$

The set $\hat{C}$ is the intersection of the decreasing chain of internal sets $\hat{C}_n$, where $\hat{C}_n$ is the set of all X such that

$$\bar{\rho}_2\left(X(\omega,t), \sum_{s<t} G(\omega,s)(X(\omega,s))\Delta W(\omega,s)\right) \leq 1/n, \tag{3}$$

and

$$(\exists Y)[Y \text{ is adapted after } 1/n \text{ and } \bar{\rho}_2(X,Y) \leq 1/n]. \tag{4}$$

If $x \in C$ and X lifts x, then $X \in \hat{C}$. Moreover, if $X \in \hat{C}$ then X is near-standard, and taking standard parts we see that ${}^{\circ}X \in C$. Therefore $\hat{C}$ is the monad of C. This shows that the set C of all solutions of equation (1) has the property that the monad of C is a countable intersection of internal sets. In the following definition, we shall call sets with this property neocompact sets. In these lectures we show how to exploit the fact that the set of solutions of a stochastic differential equation in an adapted Loeb space is neocompact

3.2. Definition. *By a Π_1^0 set we mean a countable intersection of internal sets. A set C of random variables or stochastic processes on Ω is* **neocompact** *if the monad of C is a Π_1^0 set. A neocompact relation, i.e. a neocompact set of n-tuples of random variables and/or stochastic processes, is defined similarly.*

3.3. Theorem. *(See [3]) For every neocompact set C of continuous martingales, the set D of all pairs (x,w) such that (x,w) solves equation (1) and $w \in C$ is neocompact.*

Proof: Let the monad of C be $\bigcap_n C_n$ where each set C_n is internal. Let $\hat{D}$ be the monad of D. Let D_n be the internal set consisting of all pairs (X,W) such that $W \in C_n$ and (X,W) satisfies 3 and 4. The proof of Theorem 3.1 shows that $\hat{D} = \bigcap_n D_n$, so $\hat{D}$ is a Π_1^0 set. Therefore D is neocompact. □

Here is an alternative proof of Theorem 3.1, the "delay" proof, which will be easier to generalize to other cases. Let us take $x(\omega,u)$ to be zero when $u<0$. Let h be the delayed stochastic integral function

$$h(x,u)(\omega,t) = \int_0^t g(\omega,s)(x(\omega,s-u))dw(\omega,s).$$

Using the liftings G and W as before, we may form the internal counterpart

$$H(X,U)(\omega,t) = \sum_{s<t} G(\omega,s)(X(\omega,s-U))\Delta W(\omega,s).$$

It follows as before that the set of all pairs (x, u) such that $x = h(x, u)$ is neocompact. For each $u > 0$ we can easily build an x such that $x = h(x, u)$ by first building x on the time interval $[0, u]$, then building x on $[u, 2u]$, and so on. This is done without using the lifting at all. From the lifting we see that the set D of all $u \in [0, 1]$ such that $\exists x\ x = h(x, u)$ is also neocompact. We have $(0, 1] \subseteq D$, so the monad of D contains all noninfinitesimals. By $\aleph_1$-saturation, the monad of D contains an infinitesimal. Therefore $0 \in D$, so there exists x such that $x = h(x, 0)$. This shows that x is a solution of the equation (1). □

There are many other natural examples of neocompact sets. For instance, the set of all Brownian motions w in $L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$ such that $w(\omega, 0) = 0$ is neocompact. Its monad is the Π_1^0 set $\hat{B} = \bigcap_n \hat{B}_n$ where $\hat{B}_n$ is the internal set of all processes W such that $W(\omega, t)$ is within $1/n$ of a process which is adapted after $1/n$, and the law of W is within $1/n$ of the Wiener law (in the Prohorov metric on the set of measures on $C([0, 1], \mathbb{R}^d)$).

Another important example is the set of all stopping times τ in the time interval $[0, 1]$. A random variable $\tau \in L^2(\Omega, [0, 1])$ is a **stopping time** if the stochastic process $\min(t, \tau(\omega))$ is adapted. The corresponding notion of an internal stopping time was introduced in Lindstrøm's article ([8] Definition 4.6). The set of all internal stopping times is itself internal. The monad of the set of stopping times is the Π_1^0 set of all X such that X is infinitely close some internal stopping time.

Lemma 2.2 shows that for every neocompact set C in either $L^2(\Omega, \mathcal{L}(\mathcal{M}))$ or $L^2(\Omega, \mathcal{C}(\mathcal{M}))$, the set of all adapted $x \in C$ is again neocompact. Similarly, for each neocompact set C in $L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$, the set of all continuous martingales in C is neocompact.

The two proofs of Theorem 3.1 illustrate the usefulness of the following notion of a neocontinuous function.

3.4. Definition. *Let $B \subseteq L^2(\Omega, \mathcal{M})$ and $f : B \to L^2(\Omega, \mathcal{N})$. We say that a function $F : \hat{B} \to SL^2(\Omega, \mathcal{N})$ is a* **lifting** *of f if F is internal, $B \subseteq {}^\circ\hat{B}$, and whenever $X \in \hat{B}$ and ${}^\circ X = x \in B$ we have ${}^\circ(F(X)) = f(x)$. We say that f is* **neocontinuous** *if it has a lifting on each neocompact subset of B.*

It is clear that the composition of two neocontinuous functions is again neocontinuous.

Many examples of neocontinuous functions can be found in the earlier lectures. For example, the distance function

$$\rho : L^2(\Omega, \mathcal{M}) \times L^2(\Omega, \mathcal{M}) \to \mathbb{R}$$

is neocontinuous. The projection functions $(x, u) \mapsto x$ and $(x, u) \mapsto u$ are neocontinuous. For each bounded continuous function $\varphi : \mathcal{M} \to \mathbb{R}$, the function $x \mapsto E[\varphi(x(\bullet))]$ is a neocontinuous function from $L^2(\Omega, \mathcal{M})$ to $\mathbb{R}$.

Proposition 2.3 shows that the stochastic integral

$$(f, w) \mapsto \int_0^t f(\omega, s) dw(\omega, x)$$

is a neocontinuous function on the set of pairs (f, w) where f is uniformly bounded and adapted and w is a continuous martingale.

In the proof of Theorem 3.1 the application function

$$(g(\omega, t), x(\omega, t)) \mapsto g(\omega, t)(x(\omega, t))$$

is neocontinuous

$$L^2(\Omega, \mathcal{L}(C(\mathbb{R}^d, \mathbb{R}^{d\times d}))) \times L^2(\Omega, C(\mathbb{R}^d)) \to L^2(\Omega, \mathcal{L}(C(\mathbb{R}^d, \mathbb{R}^{d\times d}))).$$

It follows that the function

$$(g, x, w) \mapsto \int_0^t g(\omega, s)(x(\omega, s)) dw(\omega, s)$$

is neocontinuous on the set of triples (g, x, w) where g is uniformly bounded and adapted, and x, w are continuous martingales. If $(x, y) \mapsto f(x, y)$ is a neocontinuous function of two variables, then $x \mapsto f(x, y_0)$ is neocontinuous in x for each y_0. Thus, for example,

$$\int_0^t g(\omega, s)(x(\omega, s)) dw(\omega, s)$$

is also neocontinuous as a function of x alone. The function $h(x, u)$ from the delay proof is also neocontinuous.

The following proposition about neocompact sets and neocontinuous functions contains the key facts needed in many of the applications. The proofs in a more general abstract setting are in [3] and [4].

3.5. Proposition. *Let C be a neocompact set and let f be a neocontinuous function on C.*

(i) C is closed and bounded.

(ii) f is continuous.

(iii) $f(C)$ is neocompact.

(iv) If D is neocompact then $C \cap f^{-1}(D)$ is neocompact.

(v) Every compact set is neocompact.

(vi) The intersection of any countable chain C_m of nonempty neocompact sets is nonempty.

Proof: (i) Let C be neocompact and let x be a limit of a sequence x_n of points in C. Let the monad of C be $\hat{C} = \bigcap_n \hat{C}_n$. Let X lift x. For each n there exists $Y_n \in C_n$ such that Y_n is within $1/n$ of X. By $\aleph_1$-saturation there exists $Y \in \hat{C}$ such that $Y \approx X$, and therefore $x \in C$. This proves that C is closed.

Suppose C is not bounded. Then for each n there is a pair of points X_n, Y_n in the monad of C such that $\bar{\rho}(X_n, Y_n) \geq n$. By $\aleph_1$-saturation there is a pair of points X, Y in the monad of C such that $\bar{\rho}(X, Y)$ is infinite, which is impossible.

(ii) By definition, f has an S-continuous lifting F.

(iii) Let F be a lifting of f, and let the monad of C be $\hat{C} = \bigcap_n \hat{C}_n$. Let $\hat{B}$ be the domain of F. Then $\hat{B}$ is internal and $C \subseteq {}^\circ(\hat{B})$. Let $\hat{D} = \hat{B} \cap \hat{C}$, and $\hat{D}_n = \hat{B} \cap \hat{C}_n$. Then $C = {}^\circ\hat{D}$ and $f(C) = {}^\circ(F(\hat{D}))$. We have $F(\hat{D}) = \bigcap_n F(\hat{D}_n)$; the nontrivial inclusion follows from $\aleph_1$-saturation. By $\aleph_1$-saturation again,

$$^\circ(F(\hat{D})) = {}^\circ(\bigcap_n F(\hat{D}_n)) = \bigcap_n ({}^\circ F(\hat{D}_n)) = \bigcap_n (({}^\circ F(\hat{D}_n))^{1/n}).$$

It follows that the monad of $f(C)$ is the Π^0_1 set $\bigcap_n (({}^\circ F(\hat{D}_n))^{1/n})$.

The proof of (iv) is similar.

(v) Let E be compact. For each n, there is a finite subset E_n such that $E \subseteq ((E_n)^{1/n})$. Then the monad of E is the Π^0_1 set $\bigcap_n \hat{E}_n$ where

$$\hat{E}_n = \{X : \bar{\rho}_2(X, E_n) \leq 1/n\}.$$

(vi) For each m we may represent the monad of C_m as an intersection $\bigcap_n \hat{C}_{m,n}$ of a decreasing chain of internal sets. Then the intersection of any finite number of the internal sets $\hat{C}_{m,n}$ is nonempty. By $\aleph_1$-saturation, the intersection $\bigcap_m \bigcap_n \hat{C}_{m,n}$ is nonempty. Let X belong to this intersection. Then X is near-standard and ${}^\circ X \in \bigcap_m C_m$. □

As a consequence, we see that if C is a nonempty neocompact set and $f : C \to \mathbb{R}$ is a neocontinuous function, then f has a minimum and a maximum. (Because the range $f(C)$ is a closed bounded set of reals). This allows us to prove that optimal solutions of various kinds exist.

Another consequence is that for any neocompact relation $C \subseteq \mathcal{M} \times \mathcal{N}$, the projection function $f(x, y) = x$ is neocontinuous and hence its range

$$\{x \in \mathcal{M} : (\exists y \in \mathcal{N})(x, y) \in C\}$$

is neocompact.

We can now very quickly get many applications of the result that the set of solutions of the stochastic differential equation (1) is neocompact. Here are several typical examples. In each case, we can conclude that optimal solutions exist and that the set of all optimal solutions is again neocompact.

3.6. Corollary. *(i) Let w be a continuous martingale, and let*

$$f : C([0,1], \mathbb{R}^d) \to \mathbb{R}$$

be a bounded continuous function. Then the set of solutions x of equation (1) such that $E[f(x(\omega))]$ is a minimum is nonempty and neocompact.

(ii) Let C be a nonempty neocompact set of continuous martingales, and let

$$f : C([0,1], \mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R}$$

be a bounded continuous function. Then the set of pairs (x, w) such that $w \in C$, (x, w) solves equation (1), and $E[f(x(\omega), w(\omega))]$ is a minimum, is nonempty and neocompact.

(iii) For every pair of stochastic differential equations of the form (1), the set of pairs of solutions (x_1, x_2) such that $\rho_2(x_1, x_2)$ is a minimum is nonempty and neocompact.

(iv) Let w be a continuous martingale. For any nonempty neocompact set $C \subseteq L^2(\Omega, C(\mathbb{R}^d))$ or $C \subseteq L^2(\Omega, \mathcal{L}(\mathbb{R}^d))$, the set of all $y \in C$, such that

$$\rho_2\left(y, \int_0^t g(\omega, s)(y(\omega, s))dw(\omega, s)\right)$$

is a minimum, is nonempty and neocompact.

3.7. Corollary. *Suppose that we have a sequence of equations*

$$x(\omega, t) = \int_0^t g_n(\omega, s)(x(\omega, s))dw_n(\omega, s)$$

where each g_n is a bounded adapted process with values in the space $C(\mathbb{R}^d, \mathbb{R}^{d\times d})$), and w_n is a continuous martingale with values in $\mathbb{R}^d$. Assume that for each n there exists an x which is a solution of the first n equations. Then there exists an x which is a simultaneous solution of all the equations, and the set of all such x is again neocompact. □

3.8. Corollary. *(Stochastic differential equations with control)*

(i) Let $w(\omega, s)$ be a continuous martingale in $\mathbb{R}^d$, let $h(\omega, s)$ be a uniformly bounded adapted process with values in the space $C(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d\times d})$, and let x be an adapted process with values in $\mathbb{R}^d$. Then the set of all continuous martingales y in $\mathbb{R}^d$ such that

$$y(\omega, t) = \int_0^t h(\omega, s)(x(\omega, s), y(\omega, s))dw(\omega, s) \tag{5}$$

is nonempty and neocompact.

(ii) For any neocompact set C of triples (h, x, w) of the appropriate kind, the set of quadruples (h, x, w, y) such that $(h, x, w) \in C$ and y is a continuous martingale which is a solution of the above equation is neocompact. □

3.9. Corollary. *Let y be a continuous martingale such that for some continuous martingale x, x solves equation (1) and (x, y) solves (5). Let*

$$f : C([0,1], \mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R}$$

be bounded and continuous. Then the set of controls x such that x solves equation (1), (x, y) solves (5), and $E[f(x(\omega), y(\omega))]$ is a minimum, is nonempty and neocompact. □

4. Solutions which are Markov processes

In [5] it was shown that in the case that w is a Brownian motion and the coefficient g is deterministic, equation (1) has a solution with the strong Markov property. In this section we shall give a simpler argument and prove a weaker result—there is a solution with the ordinary Markov property. To find such a solution, we shall use a particular countable sequence of optimal solutions and a lifting theorem from [5] for Markov processes. A continuous stochastic process x in $\mathcal{M}$ is a **Markov process** (with respect to $\mathcal{F}_\bullet$) if it is adapted and for each pair of times $s < t$ in $[0,1]$ and each bounded continuous function $\varphi : \mathcal{M} \to \mathbb{R}$,

$$E[\varphi(x(\bullet, t))|\mathcal{F}_s] = E[\varphi(x(\bullet, t))|x(\bullet, s)]. \tag{6}$$

That is, the value of x at time s gives all information available at time s about the value of x at time t. The strong Markov property is a stronger condition obtained by replacing the time s with a stopping time τ.

We need a lifting theorem for conditional expectations of random variables.

4.1. Lemma. *Let $x \in L^2(\Omega, \mathbb{R}^d)$ be a random variable, let X lift x, and let $\mathcal{A}$ be a countably generated sigma-algebra contained in $\mathcal{F}_t$. Then for all sufficiently large $s \approx t$ in T,*

(i) $E[X|\mathcal{G}_s]$ is a lifting of $E[x|\mathcal{F}_t]$.

(ii) There is an internal algebra $\mathcal{B} \subseteq \mathcal{G}_s$ such that $E[X|\mathcal{B}]$ is a lifting of $E[x|\mathcal{A}]$.

Proof: Part (i) is in [5] and is left as an exercise. (ii) Let $\mathcal{A}_n$ be an increasing chain of finite algebras whose union generates $\mathcal{A}$. Then $E[x|\mathcal{A}] = \lim_{n\to\infty} E[x|\mathcal{A}_n]$. Let $\mathcal{B}_n$ be a finite internal algebra which approximates $\mathcal{A}_n$

within a null set. For each n, $E[X|\mathcal{B}_n]$ lifts $E[x|\mathcal{A}_n]$. Since $\mathcal{A}_n \subseteq \mathcal{F}_t$, we may take $\mathcal{B}_n$ so that $\mathcal{B}_n \subseteq \mathcal{G}_s$ for some $s \approx t$. By $\aleph_1$-saturation we may extend the sequence $\mathcal{B}_n$ to an internal sequence $\mathcal{B}_J, J \in {}^*\mathbb{N}$. By overspill, for all sufficiently small infinite J we have $\mathcal{B}_J \subseteq \mathcal{G}_s$ and $E[X|\mathcal{B}_J]$ lifts $E[x|\mathcal{A}]$. □

4.2. Theorem. *(See [5]) Let w be a Brownian motion with values in $\mathbb{R}^d$ on Ω, and let $g \in \mathcal{L}(C(\mathbb{R}^d, \mathbb{R}^{d\times d}))$ be uniformly bounded. Then the stochastic differential equation*

$$x(\omega, t) = \int_0^t g(s, x(\omega, s))dw(\omega, s) \tag{7}$$

has a solution which is a Markov process.

One cannot expect to have a Markov solution in the case that the coefficient g depends on ω, because the value of x at time t will then depend on ω through g. Similarly, one cannot expect a Markov solution in the case that w is an arbitrary continuous martingale. However, in the case that w is a continuous martingale with the Markov property, the theorem can be improved, with more work, to say that the equation has a solution x such that the joint process (x, w) has the Markov property.

Proof of Theorem 4.2: Let Φ be a countable set of bounded continuous functions from $\mathbb{R}^d$ into $\mathbb{R}$ such that whenever $E[\varphi(x(\omega))] = E[\varphi(y(\omega))]$ for all $\varphi \in \Phi$, x and y have the same distribution. Then for x to be a Markov process it is sufficient that equation (6) hold for all $\varphi \in \Phi$. Since each side of equation 6 is continuous in t, it is even sufficient that (6) holds for all rational t and all $\varphi \in \Phi$. Let (φ_n, t_n), $n \in \mathbb{N}$ be an enumeration of the countable set $\Phi \times (\mathbb{Q} \cap [0,1])$.

Let C_0 be the set of all solutions of equation (7). We inductively define C_{n+1} to be the set of all $x \in C_n$ such that $E[\varphi_n(x(\bullet, t_n))]$ is maximal among all members of C_n. The functions $x \mapsto E[\varphi_n(x(\bullet, t_n))]$ are neocontinuous. Using Corollary 3.6, it follows by induction that for each n, the set C_n is nonempty and neocompact. The sets C_n form a decreasing chain. Then by countable compactness, the intersection $x \in \bigcap_n C_n$ is nonempty and neocompact.

Let $x \in \bigcap_n C_n$. x is a solution of (7) because it belongs to C_0. We shall prove that x is a Markov process. To do this we prove by induction that for all n,

$$E[\varphi_n(x(\bullet, t_n))|\mathcal{F}_s] = E[\varphi_n(x(\bullet, t_n))|x(\bullet, s)] \tag{8}$$

for all $s \leq t_n$. Suppose this holds for all $n < m$, but fails for m and some $s \leq t_m$. Since $\mathbb{R}^d$ is separable, the σ-algebra determined by $x(\bullet, s)$ is countably generated.

We now go up to the hyperfinite world. Let G be a uniformly bounded lifting of g. Let X be a martingale lifting of x. Then for all ω in a set U_0 of Loeb probability one,

$$(\forall t)^{\circ}X(\omega,t) = x(\omega, {}^{\circ}t)$$

and

$$(\forall t)X(\omega,t) \approx \sum_{s<t} G(s, X(\omega,s))\Delta W(\omega,s). \tag{9}$$

Moreover, any X which satisfies (9) is a lifting of an element of C_0.

By Lemma 4.1, there exists $u \approx s$ in T and an internal algebra $\mathcal{B} \subseteq \mathcal{G}_u$ such that $E[\varphi_m(X(\bullet,t_m)|\mathcal{G}_u]$ lifts $E[\varphi_m(x(\bullet,t_m)|\mathcal{F}_s]$ and $E[\varphi_m(X(\bullet,t_m)|\mathcal{B}]$ lifts $E[\varphi_m(x(\bullet,t_m)|x(\bullet,s)]$. Since equation (8) fails, there is a set $U \in \mathcal{F}_s$ of positive Loeb measure and a real $\varepsilon > 0$ such that for all $\omega \in U$,

$$E[\varphi_m(x(\bullet,t_m))|\mathcal{F}_s](\omega) + \varepsilon \le E[\varphi_m(x(\bullet,t_m))|x(\bullet,s)](\omega).$$

U has an internal subset $V \in \mathcal{G}_u$ of positive Loeb measure such that both conditional expectation liftings hold at all $\omega \in V$.

We now form a new internal stochastic process Y as follows. For each equivalence class $[\omega]_u \subseteq V$, internally choose a new equivalence class $[\omega']_u$ such that ω,ω' belong to the same $\mathcal{B}$-equivalence class but

$$E[\varphi_m(X(\bullet,t_m))|\mathcal{G}_u](\omega) + \varepsilon/2 \le E[\varphi_m(X(\bullet,t_m))|\mathcal{G}_u](\omega').$$

Form the process Y from X by exchanging the set of paths in the class $[\omega]_u$ by a copy of the set of paths in the class $[\omega']_u$, for each $\omega \in V$. Then Y is an improvement on X for the function φ_m, because

$$E[\varphi_m(X(\bullet,t_m))] + \varepsilon \cdot P(V)/2 \le E[\varphi_m(Y(\bullet,t_m))].$$

Moreover, Y is near-standard, and we may take $y = {}^{\circ}Y$. Taking standard parts, the corresponding inequality also holds for x and y. We shall show that $y \in C_m$. This will contradict the fact that $x \in C_{m+1}$ and hence that $E[\varphi_m(x(\bullet,t_m))]$ is maximal.

Y still satisfies equation (9) and thus y belongs to the set C_0. By inductive hypothesis, x satisfies (8) for all $n < m$. The exchange procedure will not disturb this property, so y also satisfies (8) for all $n < m$. Therefore

$$E[\varphi_m(X(\bullet,t_m))] \approx E[\varphi_m(Y(\bullet,t_m))]$$

for all $n < m$. Then

$$E[\varphi_m(x(\bullet,t_m))] = E[\varphi_m(y(\bullet,t_m))]$$

for all $n < m$. This shows that $y \in C_m$ and completes the induction. □

The longer proof in [5] uses the same neocompact set $\bigcap_n C_n$ and shows that every $x \in \bigcap_n C_n$ is a strong Markov process.

4.3. Corollary. *Suppose that the solutions of the stochastic differential equation (7) in Theorem 4.2 are unique in distribution, that is, for any two solutions x and y, we have $E[\varphi(x(\bullet,t))] = E[\varphi(y(\bullet,t))]$ for each bounded continuous $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $t \in [0,1]$. Then every solution of (7) is a Markov process with respect to $\mathcal{F}_\bullet$.*

Proof: In the proof of Theorem 4.2, it was shown that every x in the neocompact set $\bigcap_n C_n$ is a Markov process. But in the case that the solutions of (7) are unique in distribution, every solution $x \in C_0$ maximizes $E[\varphi_n(x(\bullet,t_n))]$ for every n, so the sets C_n are all the same. Therefore the set $\bigcap_n C_n$ is equal to the set C_0 of all solutions of (7). □

In the above corollary, the weaker conclusion that every solution x is a Markov process with respect to the filtration generated by the process x itself is well known and easily proved by classical methods. The point of the above result is that all solutions are Markov processes with respect to the filtration $\mathcal{F}_\bullet$ which is given in advance and is rich enough so that the existence theorem holds.

5. A Fixed Point Theorem

We shall now prove a simple but quite general fixed point theorem which can be used to show that for many stochastic differential equations set of all solutions is both nonempty and neocompact.

Let $A^2(\Omega,\mathcal{C}(\mathcal{M}))$ be the set of all adapted processes in $L^2(\Omega,\mathcal{C}(\mathcal{M}))$.

Given a stochastic process $x \in L^2(\Omega,\mathcal{C}(\mathcal{M}))$ and a time $t \in [0,1]$, we let $x[0,t]$ be the restriction of x to the time interval $[0,t]$, that is, $(x[0,t])(\omega) = x(\omega) \cap ([0,t] \times \mathcal{M})$.

For $x \in L^2(\Omega,\mathcal{C}(\mathcal{M}))$ and $u \in [0,1]$, define the **delay function** dl by

$$dl(x,u)(\omega,t) = x(\omega, max(0,t-u)).$$

The delay function has the following properties:

$$dl(x,t+u) = dl(dl(x,t),u),$$

$$x[0,t] = y[0,t] \Rightarrow (dl(x,u))[0,t+u] = (dl(y,u))[0,t+u].$$

One can readily check that the delay function dl is neocontinuous from $L^2(\Omega,\mathcal{C}(\mathcal{M})) \times [0,1]$ to $L^2(\Omega,\mathcal{C}(\mathcal{M}))$, and also maps $A^2(\Omega,\mathcal{C}(\mathcal{M})) \times [0,1]$ to $A^2(\Omega,\mathcal{C}(\mathcal{M}))$.

5.1. Definition. *By an* **adapted function** *on $A^2(\Omega,\mathcal{C}(\mathcal{M}))$ we shall mean a function*

$$I : A^2(\Omega,\mathcal{C}(\mathcal{M})) \to A^2(\Omega,\mathcal{C}(\mathcal{M}))$$

such that for all x, y, t,

$$(I(x))(\omega, 0) = x(\omega, 0),$$

and

$$x[0,t] = y[0,t] \Rightarrow (I(x))[0,t] = (I(y))[0,t].$$

That is, $I(x)$ has initial value $x(\omega, 0)$ and for each t, $I(x)[0,t]$ depends only on $x[0,t]$.

For example, if

$$w \in A^2(\Omega, C(\mathbb{R}^{d\times d}))$$

is a continuous martingale and

$$g \in A^2(\Omega \times [0,1], L^2(\mathbb{R}^d, \mathbb{R}^{d\times d}))$$

is uniformly bounded then the stochastic integral

$$I(x)(\omega, t) = x(\omega, 0) + \int_0^t (g(\omega, s, x(\omega, s))dw(\omega, s)$$

is an adapted function on $A^2(\Omega, \mathcal{C}(\mathcal{M}))$.

5.2. Theorem. *(Fixed Point Theorem) Let $C \subseteq A^2(\Omega, \mathcal{C}(\mathcal{M}))$ be a nonempty neocompact set such that for each $x \in C$ and $t \in [0,1]$, $dl(x,t) \in C$. Let I be an adapted function on $A^2(\Omega, \mathcal{C}(\mathcal{M}))$ such that $I(C) \subseteq C$ and I is neocontinuous. Then there exists a point $x \in C$ such that $I(x) = x$ (a* **fixed point for** *I), and the set of all fixed points for I in C is neocompact.*

Proof: The function $j(x) = \rho(x, I(x))$ is a composition of neocontinuous functions and hence is itself neocontinuous on C. The set $\{0\}$ is neocompact, and therefore the inverse image

$$j^{-1}(\{0\}) = \{x \in C : x = I(x)\},$$

which is the set of all fixed points of I, is neocompact.

The proof that a solution exists is an abstract form of the delay argument.

Let D be the set of all pairs $(y, u) \in C \times [0,1]$ such that $y = I(dl(y, u))$. Then D is a neocompact set. Since the projection function $(y, u) \mapsto u$ is neocontinuous, the set E of all $u \in [0,1]$ such that $(\exists y \in C)(y, u) \in D$ is a neocompact subset of $[0,1]$. We show that $(0,1] \subseteq E$. Once this is done, the proof is completed as follows. Since E is neocompact it is closed, and therefore $0 \in E$. But this means that there exists $y \in C$ such that

$$y = I(dl(y, 0)) = I(y)$$

as required.

We let $u \in (0,1]$ and prove that $u \in E$. Choose an element $y_0 \in C$. Inductively define a sequence y_n by

$$y_{n+1} = I(dl(y_n, u)).$$

We see by induction that each y_n belongs to C.

We now claim that for each n,

$$y_{n+1}[0, nu] = y_n[0, nu].$$

We prove this claim by induction on n. For $n = 0$ we have

$$(y_1)(\omega, 0) = (I(dl(y_0, u)))(\omega, 0) = (dl(y_0, u))(\omega, 0) = (y_0)(\omega, 0),$$

so

$$y_1[0, 0] = y_0[0, 0].$$

Assume that the claim holds for n and let $t = nu$, so that

$$y_{n+1}[0, t] = y_n[0, t].$$

Then

$$(dl(y_{n+1}, u))[0, t+u] = (dl(y_n, u))[0, t+u],$$

and therefore

$$\begin{aligned} & y_{n+2}[0, t+u] \\ &= (I(dl(y_{n+1}, u))[0, t+u] \\ &= (I(dl(y_n, u)))[0, t+u] \\ &= y_{n+1}[0, t+u]. \end{aligned}$$

This completes the induction and proves the claim.

Now take k large enough so that $ku \geq 1$. Then by the claim,

$$y_{k+1} = y_k,$$

and therefore

$$y_k = I(dl(y_k, u)).$$

This shows that $(y_k, u) \in D$ and so $u \in E$ as required. □

As a first illustration let us apply the Fixed Point Theorem to the case of equation (1). Let k be a uniform bound for the adapted continuous function g, and let C be the set of all stochastic integrals $\int_0^t h(\omega, s) dw(\omega, s)$ where w is a continuous martingale of dimension d and h is an adapted process in $L^2(\Omega, \mathbb{R}^{d \times d})$ with bound k. Then C is neocompact because its monad is the set of all X such that for each n, X is within $1/n$ of some hyperfinite sum

$$\sum_{s<t} H(\omega, s)\Delta W(\omega, s)$$

where H is adapted after $1/n$ and bounded by $k + 1/n$. Whenever $x \in C$ and $u \in [0, 1]$, we have $dl(x, u) \in C$. This can be seen by changing the coefficient h to be zero before u.

In this case,

$$I(x)(\omega, t) = \int_0^t g(\omega, s, x(\omega, s))dw(\omega, s).$$

Then I is an adapted function, $I : C \to C$, and I is neocontinuous. By the Fixed Point Theorem, the set of all fixed points $x \in C$ of I is a nonempty neocompact set, and this set is the set of all solutions of equation (1).

6. Stochastic Differential Equations with Nondegenerate Coefficients

In this section we apply the Fixed Point Theorem to give a short proof of a more difficult existence theorem. This is the case of stochastic differential equations where the coefficient is measurable rather than continuous in x, but the determinant of the coefficient is bounded away from zero. This result is from [5], and is an improvement of a weak existence theorem of Krylov [7]. The present proof uses some neocontinuity results from [3].

Let us choose a uniform bound $k > 0$ once and for all, and let J be the compact set of all $d \times d$ matrices A such that the entries of A are bounded by k and $det(AA^T) \geq 1/k$.

We collect the needed facts in a lemma which we state without proof.

6.1. Lemma. *([3]) Let w be a Brownian motion in $\mathbb{R}^d$. There is a neocompact set $C \subseteq L^2(\Omega, \mathcal{C}(\mathbb{R}^d))$ such that:*

(i) For each adapted process $y \in L^2(\Omega, \mathcal{L}(J))$, and $r \in [0, 1]$, the integral $\int_{min(r,t)}^t y(\omega, s)dw(\omega, s)$ belongs to C,

(ii) C is closed under delays,

(iii) For each function $g \in L^2([0, 1] \times \mathbb{R}^d, J)$ where $\mathbb{R}^d$ has the normal measure, the function

$$I(x)(\omega, t) = \int_0^t g(s, x(\omega, s))dw(\omega, s)$$

is neocontinuous on C. □

Here is the existence theorem.

6.2. Theorem. *([5]) Let w be a Brownian motion in $\mathbb{R}^d$. For each function $g \in L^2([0,1] \times \mathbb{R}^d, J)$ where $\mathbb{R}^d$ has the normal measure, the equation*

$$x(\omega, t) = \int_0^t g(s, x(\omega, s)) dw(\omega, s)$$

has a continuous martingale solution, and the set of all solutions is neocompact.

Proof: Let C be the neocompact set from Lemma 6.1 and let $I(x)$ be the stochastic integral function

$$I(x) = \int_0^t g(s, x(\omega, s)) dw(\omega, s).$$

I is neocontinuous on C by Lemma 6.1. By Lemma 2.2, we may take C to be included in the set of adapted processes and may also take C so that $x(\omega, 0) = 0$ for all $x \in C$. Then by part (i) of Lemma 6.1, $I(C) \subseteq C$, and $(I(x))(\omega, 0) = x(\omega, 0)$. Since $(I(x))(\omega, t)$ depends only on (ω, s) and the values of $x(\omega, s)$ for $s \leq t$, I is an adapted function. The conclusion of the theorem now follows from the Fixed Point Theorem. □

It would be interesting to use the Fixed Point Theorem to find additional existence theorems. One candidate to be checked is the equation of Theorem 6.2 with the coefficient g being an adapted function rather than deterministic.

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., and Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Anderson, R.M., (1976) A nonstandard representation for Brownian motion and Itô integration, *Israel Math. Journal* **25**, pp. 15–46.
3. Fajardo, S. and Keisler, H.J., (1996) Existence theorems in probability theory, *Advances in Mathematics* **120**, pp. 191–257.
4. Fajardo, S. and Keisler, H.J., (1996) Neometric spaces, *Advances in Mathematics* **118**, pp. 134–175.
5. Keisler, H.J., (1984) An Infinitesimal Approach to Stochastic Analysis, *Memoirs Amer. Math. Soc.* **297**.
6. Keisler, H.J., (1995) A Neometric survey, in *Developments in Nonstandard Mathematics* (N. Cutland et al Eds.), pp. 233–250, Longman, Harlow.
7. Krylov, N.V., (1980) *Controlled Diffusion Processes,* Springer-Verlag.
8. Lindstrøm, T. (1997) Internal martingales and stochastic integration, *this volume.*
9. Loeb, P. (1975) Conversion from nonstandard to standard measure spaces and applications in probability theory, *Transactions Amer. Math. Society* **211**, pp. 113–122.

HYPERFINITE MATHEMATICAL FINANCE

P. EKKEHARD KOPP
School of Mathematics
University of Hull
Hull HU6 7RX
England
email: p.e.kopp@maths.hull.ac.uk

1. Introduction

Financial markets have provided one of the most remarkable growth industries in the past two decades, and now constitute a major source of employment for graduates with high levels of mathematical expertise. The principal reason for this phenomenon lies in the explosive growth of the market in *derivatives*, whose levels of activity now frequently exceed the underlying markets on which their products are based. The variety and complexity of new financial instruments is often bewildering, and much effort goes into the analysis of the mathematical models on which their existence is predicated.

These lectures provide a brief discussion of the principal stochastic models which underlie much of this activity, without addressing practitioners' questions about the realism of the simplifying assumptions. For ease of exposition we shall concentrate throughout on the more basic *stock options* rather than attempting to discuss more 'exotic' financial instruments or interest rate models.

A *contingent claim* or *derivative security* is an asset whose value is determined by the values of one or more underlying variables (usually securities themselves). The analysis of such claims, and their *pricing* in particular, is the main purpose of mathematical finance. While the construction of mathematical models for this analysis often involves quite technical parts of stochastic analysis or PDEs, the *economic* insights which underlie the modelling are often remarkably simple and transparent.

In order to highlight these insights we first develop rather simplistic mathematical models based on discrete time and finitely generated prob-

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 279–307.

ability spaces, before showing how the analogous concepts can be used in the more widely known continuous-time models based on diffusions and *Itô processes.* Nonstandard techniques will then be used to provide a direct link between the two types of model and to gain insight into their underlying structure. In particular, they will provide a natural convergence concept which preserves the operations of the stochastic calculus - this turns out to be highly desirable for the interpretation of market models.

1.1. BASIC OPTIONS TERMINOLOGY

An *option* on a stock is a contract giving the owner to right, but not the obligation, to trade a given number of shares of a common stock for a fixed price at a future date (the *expiry date* T). A *call* option gives the owner the right to buy stocks, a *put* option confers the right to sell, at the fixed *strike price*, K. The option is *European* if it can only be exercised at the fixed expiry date, T; the option is *American* if the owner can exercise his right to trade at any time up to the expiry date.

The problem of *option pricing* is to determine what value to assign to the option at a given time, e.g. at time 0. It is clear that traders can make a riskless profit (at least in the absence of inflation) unless they have paid an 'entry fee' which allows them the chance of exercising the option favourably at the expiry date. On the other hand, if this 'fee' is too high, there may be no possibility of favourable exercise, and no sensible trader would then buy the option at this price. To take the simplest example: the buyer of a European call option on a stock with price process $(S_t)_{t\in[0,T]}$ will have the opportunity of receiving a pay-out at time T of

$$C = max(S_T - K, 0) = (S_T - K)^+,$$

simply by exercising the option if and only if the final price of the stock S_T is greater than the strike price K previously agreed upon.

1.2. THE BLACK-SCHOLES FORMULA

Of course, seeking a solution to this pricing problem pre-supposes that we have a good understanding of the dynamics which underlies the stock price S. The idea that we could use Brownian motion (BM) as the underlying driving force to model the random behaviour of S goes back to *Bachelier*'s doctoral thesis in 1900, which pre-dates Einstein's 1905 papers and Wiener's rigorous definition of BM in 1923. But it was only in 1964 that *Samuelson* proposed an exponential BM pricing model (which eliminates negative prices and focuses on proportional price changes) and the first 'rational' option pricing formula was the *Black-Scholes formula*, first derived in [6] in 1973.

This pricing model is simple enough, even if the solution is not very transparent at first sight: the riskless interest rate $r > 0$ is assumed constant throughout a time interval $[0, T]$, so that 'money in the bank' will accumulate deterministically, solving the differential equation

$$dS_t^0 = rS_t^0 dt,$$

so that

$$S_t^0 = S_0^0 e^{rt}$$

describes the behaviour of a 'riskless bond' or bank account over time. We also suppose the existence of a risky stock S, governed by the stochastic differential equation

$$dS_t = \sigma S_t db_t + \mu S_t dt$$

for $\sigma > 0$ and $\mu \in \mathbb{R}$. By Itô's formula we obtain the solution on $[0, T]$:

$$S_t = S_0 \exp(\sigma b_t + (\mu - \frac{\sigma^2}{2})t).$$

This models a geometric BM with 'rate of return' $\frac{dS_t}{S_t}$ given by $\sigma db_t + \mu dt$, so that we naturally interpret μ as the (long-term) *drift* in the prices, and σ as the *volatility*. Assuming that in the market there are no costs of nor restrictions on trading, Black and Scholes argued that the *unique* value of a European call option with strike K and expiry T is given at any time $t \in [0, T]$ by

$$V_t = S_t \Phi(d_t) - K e^{-r(T-t)} \Phi(d_t - \sigma\sqrt{T-t})$$

where Φ denotes the normal c.d.f. and

$$d_t = \frac{\log(\frac{S_t}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}.$$

In particular, the *Black-Scholes price* of the call option is then $\pi = V_0$; this is the formula which quickly found its way into the calculators of market traders in the mid-1970s and provided them with 'baseline guidance' in setting prices for traded options.

Black and Scholes arrived at their formula by arguing heuristically that the process describing the value of the option satisfies a PDE which has the above solution, under the boundary condition $V_T = (S_T - K)^+$.

One way of describing the *fair price* for the option, they observed, is as the *current* value of a portfolio which will yield this return by time T. One could thus attempt to calculate the fair price by trying to *replicate* the value of the option throughout the time interval $[0, T]$ by means of a portfolio consisting of stock and a fixed savings account alone. The second problem is then to construct a *hedge portfolio* (η_t, θ_t) such that at each t we have $V_t = \eta_t S_t^0 + \theta_t S_t$. In the Black-Scholes case we can read this off from the above formula: $\theta_t = \Phi(d_t)$ and $\eta_t = -\Phi(d_t - \sigma\sqrt{T-t})$. Thus our hedge position is to have θ_t shares and owe η_t dollars to the bank at time t. At this stage it seems far from clear why this should be so!

1.3. ARBITRAGE AND CALL-PUT PARITY

The solution of the pricing problem for the European *put* option can be read off at once from that for the call option, under the crucial assumption that our market model rules out *arbitrage*, that is, no investor should be able to make riskless profits, in a sense which we will shortly make more precise. This assumption underlies the hedge portfolio approach and is basic to option pricing theory, since there can be no market equilibrium otherwise. (In fact, it has been argued that the very existence of 'arbitrageurs' justifies the assumption: in general, markets will adjust prices so as to eliminate disequilibrium, and hence will move to eliminate arbitrage. How realistic this assumption is in practice is something we shall not pursue here.)

So let C_t (resp. P_t) be the value at time t of the European call (resp. put) option on the stock (S_t). We assume for simplicity that the 'value of money' remains constant throughout (i.e. the riskless interest rate is 0). We claim that, in order to avoid arbitrage, the call and put prices must be related as follows:

$$C_t - P_t = S_t - K$$

To see this, consider the following transactions: at time t, buy a share S_t and a put P_t and sell a call C_t. The balance of these transactions is $C_t - P_t - S_t$. Now consider what happens at time T: if $S_T > K$, then the call will be exercised, yielding a cash sum K in return for the share. We will not exercise the put, since we would be selling the share for less than its value S_T. Hence the total value at time T from these transactions will be K. On the other hand, if $S_T < K$, then we exercise the put (our buyer will not exercise the call) and we gain K for our share, so again the total 'gain from trade' is K. Consequently, buying a share and a put and selling a call provides a sure, riskless way of obtaining an amount K at time T: hence $P_T + S_T - C_T = K$. Interest rates (and thus discount rates) being zero, it follows that if arbitrage is to be avoided, we must have

$$P_t + S_t - C_t = K$$

for all times $t \leq T$. If $C_t - P_t \neq S_t - K$, there is room for riskless profits, accruing either to ourselves or to our trading partner. This verifies the *call-put parity relation* given above.

2. Finite Market Models

Fix a time set $\mathbb{T} = \{0, 1, 2, \ldots, T\}$, where the *trading horizon* T is treated as the terminal date of the economic activity being modelled, and the points of $\mathbb{T}$ are the admissible *trading dates*. We assume given a fixed probability space $(\Omega, \mathcal{F}, P)$ to model all 'possible states of the market'. For the present we shall assume that Ω is a *finite* probability space with $P(\{\omega\}) > 0$ for all $\omega \in \Omega$.

The *information structure* available to the investors is given by an increasing (finite) sequence of sub-σ-fields of $\mathcal{F}$: we assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_T = \mathcal{F}$. An increasing family of σ-fields is called a *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ on $(\Omega, \mathcal{F}, \mathrm{P})$. We think of $\mathcal{F}_t$ as containing the information available to our investors at time t: investors learn without forgetting, but they should not be prescient - insider trading will not be allowed for. Moreover, our investors think of themselves as 'small investors', in that their actions will not change the probabilities they assign to events in the market. Note that in a finite market model each σ-field $\mathcal{F}_t$ is generated by a minimal finite partition $\mathcal{P}_t$ of Ω, and that $\mathcal{P}_0 = \{\Omega\} \subset \mathcal{P}_1 \subset \mathcal{P}_2 \subset \ldots \subset \mathcal{P}_T = \mathcal{P}$. At time t all our investors know which cell of $\mathcal{P}_t$ contains the 'true state of the market', but none of them know more.

We assume given a $(d+1)$-dimensional stochastic process $S = \{S_t^i\colon t \in \mathbb{T}, 0 \le i \le d\}$ to represent the time evolution of our securities price process, which we take generally to consist of a riskless bond S^0 and d stocks $S^1, S^2, \ldots, S^d$. Here each S_t^i is a strictly positive real-valued random variable on Ω. For simplicity, we shall also assume that the filtration $\mathbb{F}$ is that generated by the price process $S = \{S^0, S^1, \ldots, S\}$. Then $\mathcal{F}_t = \sigma(S_u : u \le t\}$ is the smallest σ-field such that all the $\mathbb{R}^{d+1}$-valued random variables $S_u = (S_u^0, S_u^1, \ldots, S_u^d), u \le t$, are $\mathcal{F}_t$-measurable. In other words, at time t the investors know the values of the price vectors $(S_u : u \le t)$, but they have no information about later values of S, or about extraneous factors affecting the prices.

2.1. SELF-FINANCING STRATEGIES

In keeping with tradition we have taken S^0 as a *bond* or riskless security, and without loss of generality we shall assume that $S_0^0 = 1$. The *discount factor* $\beta_t = 1/S_t^o$ is then the sum of money we need to invest in bonds at time 0 in order to have 1 dollar at time t.

The securities $S^0, S^1, S^2, \ldots, S^d$ are traded at times $t \in \mathbb{T}$: an investor's *portfolio* at time t is given by the $\mathbb{R}^{d+1}$-valued random variable $\theta_t = (\theta_t^i)$ $(0 \le i \le d)$, with value $V_t(\theta) = \theta_t \cdot S_t = \sum_{i=0}^d \theta_t^i S_t^i$. Investors can adjust their portfolios at each trading date, using only the information they have up to that date, so that θ is non-anticipating, i.e. θ_t is $\mathcal{F}_t$-measurable. (This may seem an unusual relaxation of the demand that θ is *predictable*, i.e. that θ_t is $\mathcal{F}_{t-1}$-measurable, which is normal in the literature, but it is simply a technical consequence of our use of *forward increments* $\Delta X_t = X_{t+1} - X_t$ for random quantities X).

We shall also assume throughout that we are dealing with a 'frictionless' market: that is, there are no transaction costs, unlimited short sales and borrowing are allowed (the random variables θ_t^i can take any real values) and the securities are perfectly divisible (the S_t^i can take any positive values).

We shall further call the trading strategy θ *self-financing* if any changes in the value $V_t(\theta)$ result entirely from net gains realised on the investments. Thus we require that:

$$\Delta V_t = \theta_t \cdot \Delta S_t$$

i.e. $\Delta V_t = V_{t+1} - V_t$ is the scalar product of the portfolio vector θ_t with the vector $\Delta S_t = S_{t+1} - S_t$ of price increments. Thus, defining the *gains process* associated with θ by setting

$$G_0(\theta) = 0, \qquad G_t(\theta) = \theta_1 \Delta S_1 + \theta_2 \Delta S_2 + \ldots + \theta_{t-1} \Delta S_{t-1}$$

we see at once that θ is self-financing if and only if

$$V_t(\theta) = V_0(\theta) + G_t(\theta)$$

for all $t \in \mathbb{T}$.

Trivially, the above have an equivalent 'discounted' form. Writing $\overline{X}_t = \beta_t X_t$ for the discounted form of the vector X_t in $\mathbb{R}^{d+1}$, we see at once that θ is self-financing if and only if

$$\overline{V}_t(\theta) = V_0(\theta) + \overline{G}_t(\theta) \text{ forall } t \in \mathbb{T}.$$

Note that the definition of $G(\theta)$ does not involve the amount θ_t^0 held in bonds (i.e. in the security S^0) at time t. Hence, if θ is self-financing, the initial investment $V_0(\theta)$ and the processes θ^i $(1 \leq i \leq d)$ completely determine θ^0.

We use the class Θ of all self-financing strategies to define our concept of 'free lunch':

Definition 2.1 *An* arbitrage opportunity *is a strategy $\theta \in \Theta$ such that $V_0(\theta) = 0$, $V_T(\theta) \geq 0$ and $\mathbb{E}(V_T(\theta)) > 0$.*

Definition 2.2 *The market model is* viable *if it does not contain any arbitrage opportunities, i.e. if $\theta \in \Theta$ has $V_0(\theta) = 0$ and $V_T(\theta) \geq 0$ then $V_T(\theta) = 0$ P-a.s..*

A natural question is whether *every* contingent claim C can be replicated in this way; i.e. whether we can find a self-financing strategy $\theta \in \Theta$ for which $V_T(\theta) = C$. Such claims will be called *attainable* and a strategy with this property will be said to *generate* C. A market model in which every contingent claim is attainable by some $\theta \in \Theta$ is said to be a *complete* market model. In such (idealised) models the pricing problem is straightforward to solve, as we shall see.

It is also easy to see that in a viable market any two generating strategies for a given claim C must have the same value process, since otherwise one may construct an arbitrage opportunity: suppose that θ and ψ both generate C, that $V_u(\theta) = V_u(\phi)$ for all $u < t$, and that $A = \{V_t(\theta) > V_t(\phi)\}$ has $P(A) > 0$. Define a new strategy η: set $\eta_u = \theta_u - \phi_u$ for $u < t$, for $u \geq t$ set $\eta_u = \theta_u - \phi_u$ on A^c and $\eta_u^i = 0$ for $i \geq 1$, set $\eta_u^0 = \frac{V_t(\theta) - V_t(\phi)}{S_t^0}$. The strategy η is strictly positive on A and replicates C. Hence it is an arbitrage opportunity.

If a claim C has a generating strategy $\theta \in \Theta$ the uniqueness of the value process suggests a means by which we can determine the (time 0) *price* of the claim, namely by taking the discounted time 0 value of the strategy θ as the price of the claim C. However, in general we have no guarantee that the quantity $\bar{V}_0(\theta)$ will remain positive, even if $\bar{V}_T(\theta) = C > 0$. It is for this reason that we need to restrict ourselves to viable markets, where such difficulties cannot occur.

2.2. MARTINGALE PRICING

Now suppose that the discounted price process $\overline{S^i}$ for the i-th stock happens to be a martingale under some probability measure Q. Then the discounted value process $\overline{V}^i(\theta)$ for the i-th stock is a discrete stochastic integral and is therefore a martingale (see Lindstrøm: Proposition 4.5) with initial (constant) value $V_0(\theta)$. Therefore we have $\mathbb{E}(\overline{V}_T^i(\theta)) = \mathbb{E}(V_0^i(\theta))$. If this condition holds for each $i \in \mathbb{T}$, it precludes the possibility of arbitrage: if $V_0(\theta) = 0$ and $V_T(\theta) \geq 0$ a.s. (Q), but $\mathbb{E}(\overline{V}_T^i(\theta)) = 0$ for each i, it follows that $V_T(\theta) = 0$ a.s. (Q). This remains true a.s. (P) provided that the probability measure Q has the same null sets as P (i.e. if Q and P are *equivalent measures*, which we write $Q \sim P$). If such a measure can be found, then no self-financing strategy θ can lead to arbitrage, i.e. the market is viable. This leads to an important concept:

Definition 2.3 *A probability measure $Q \sim P$ is an* equivalent martingale measure *for S if the discounted price process $\overline{S}$ is a (vector) martingale under Q for the filtration $\mathbb{F}$.*

We have just seen that the existence of an equivalent martingale measure for S is *sufficient* for viability of the securities market model. For finite models it has also been shown to be *necessary* (see e.g. [16]). Mathematically, the search for equivalent measures under which the given process $\overline{S}$ is a martingale is often much more convenient than having to show that no arbitrage opportunities exist for $\overline{S}$. We shall not consider this further; instead we now show that the existence of an equivalent martingale measure provides a general method for pricing contingent claims.

Assume that we are given a viable market model $(\Omega, \mathcal{F}, P, \mathbb{F}, S)$ with equivalent martingale measure Q. Mathematically, a *contingent claim* in this model is a nonnegative ($\mathcal{F}$-measurable) random variable C, representing a contract that pays out $C(\omega)$ dollars at time T under certain conditions, provided that $\omega \in \Omega$ occurs. Its time 0 value or (current) *price* $\pi(C)$ is then the value which the parties to the contract would deem a 'fair price' for this contract to be entered into. In a viable model, an investor could hope to evaluate $\pi(C)$ by constructing a trading strategy $\theta \in \Theta$ which generates C, i.e. exactly replicates the returns (cash-flow) yielded by C at time T. For such a strategy θ, the initial investment $V_0(\theta)$ would represent the price $\pi(C)$ of C. As Q is a martingale measure for S, and by construction $\overline{V}(\theta)$ is a discrete stochastic integral, and hence a martingale, under Q, it follows that for all $t \in \mathbb{T}$, $\overline{V}_t(\theta) = \mathbb{E}_Q(\beta_T C|\mathcal{F}_t)$, and thus

$$V_t(\theta) = \beta_t^{-1}\mathbb{E}_Q(\beta_T C|\mathcal{F}_t)$$

for any $\theta \in \Theta$. In particular, therefore,

$$\pi(C) = \overline{V}_0(\theta) = \mathbb{E}_Q(\beta_T C|\mathcal{F}_0) = \mathbb{E}_Q(\beta_T C).$$

(Note that as $C > 0$, $\pi(C) > 0$ follows automatically.)

Thus, in *complete* models any contingent claim can be priced simply by calculating its (discounted) expectation relative to an equivalent martingale measure for the model.

Remark: Note that $\bar{V}_0(\theta) = \mathbb{E}_Q(\beta_T C)$ holds for *every* equivalent martingale measure (EMM) Q in the model, hence if the claim C is attainable then its price $\pi(C)$ will be independent of the choice of the EMM Q. In a complete model, therefore, if Q and R are two EMMs, and C is any claim, we must have $\mathbb{E}_Q(\beta_T C) = \pi(C) = \mathbb{E}_R(\beta_T C)$. This suggests strongly that in a viable complete model there is a *unique* EMM. The proof is quite simple: C is an arbitrary non-negative random variable, and replacing C by $-C$ and using linearity shows that $\mathbb{E}_Q$ and $\mathbb{E}_R$ also agree on non-positive integrands. Splitting a general random variable X into its positive and negative parts shows that $\mathbb{E}_R(X) = \mathbb{E}_Q(X)$ for all random variables X, hence $Q = R$.

The converse is also true in *finite* markets - the proof is somewhat more involved (see [22]).

2.3. EXAMPLE: THE CRR MODEL

We now specialise to the most basic discrete-time model. The *Cox-Ross-Rubinstein* binomial market model (see [8]) has $d = 1$ that is, there is single stock, S^1, and a riskless bond S^0, which accrues interest at a fixed rate $r > 0$. Taking $S_0^0 = 1$ we have $S_t^0 = (1+r)^t$ for $t \in \mathbb{T}$, and hence $\beta_t = (1+r)^{-t}$. The *ratios* of successive stock values are Bernoulli random variables, i.e. for $0 \leq t < T$, either $S_{t+1}^1 = S_t^1(1+a)$ or $S_{t+1}^1 = S_t^1(1+b)$, where $b > a > -1$ are fixed throughout, while S_0^1 is constant. We can thus conveniently choose the sample space $\Omega = \{1+a, 1+b\}^{\mathbb{T}\setminus T}$, together with the natural filtration $\mathbb{F}$ generated by the stock price values, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_t = \sigma(S_u^1 : u \leq t)$ for $t > 0$. Note that $\mathcal{F}_T = \mathcal{F} = 2^\Omega$ is the σ-field of all subsets of Ω. The measure P on Ω is that induced by the ratios of the stock values. More explicitly, let us write S for S^1 for the rest of this section to simplify the notation, and set $R_t = S_{t+1}/S_t$ for $t > 0$. For $\omega = (\omega_0, \omega_1, \ldots, \omega_{T-1})$ in Ω, define

$$P(\{\omega\}) = P(R_t = \omega_{t+1}, t = 0, 1, 2, \ldots, T-1)$$

For any probability measure Q on $(\Omega, \mathcal{F})$, the relation $\mathbb{E}_Q(\overline{S}_{t+1}|\mathcal{F}_t) = \overline{S}_t$ is equivalent to $\mathbb{E}_Q(R_t|\mathcal{F}_t) = 1 + r$, since $\beta_{t+1}/\beta_t = 1 + r$. Hence if Q is an equivalent martingale measure for S it follows that $\mathbb{E}_Q(R_t) = 1 + r$. On the other hand, R_t only takes the values $1 + a$ and $1 + b$, hence its average value can equal $1 + r$ only if $a < r < b$. We have proved:

(i) For the binomial model to have an EMM we must have $a < r < b$.

When the binomial model is viable, there is a *unique* equivalent martingale measure Q for S. We construct this measure by showing that:

(ii) $\overline{S}$ is a Q-martingale if and only if the random variables (R_t) are i.i.d., with $Q(R_1 = 1 + b) = q$ and $Q(R_1 = 1 + a) = 1 - q$, where $q = \frac{r-a}{b-a}$.

Exercise: Prove this.

Note that $q \in (0, 1)$ if and only if $a < r < b$. We see that a viable binomial market model admits a *unique* equivalent martingale measure, given by Q above. This property characterises *complete models* in the discrete-time set-up.

The CRR pricing formula, which is usually proved inductively by using an explicit hedging argument, can now be deduced directly from our general martingale formulation by calculating the Q-expectation of a European call option on the stock. In fact, the value of the call $C = (S_T - K)^+$ at time

$t \in \mathbb{T}$ is given by

$$V_t(C) = \beta_t^{-1}\mathbb{E}_Q(\beta_T C|\mathcal{F}_t).$$

But since $S_T = S_t \prod_{u=t}^{T+1} R_u$ (by definition of the (R_u)) we can calculate this expectation quite simply, as S_t is $\mathcal{F}_t-$measurable, and each R_u $(u \geq t)$ is independent of $\mathcal{F}_t$, since the choice of the forward increment is independent of the past:

$$\begin{aligned} V_t(C) &= \beta_t^{-1}\beta_T\mathbb{E}_Q((S_t \prod_{u=t}^{T-1} R_u - K)^+) \\ &= (1+r)^{t-T}\,\mathbb{E}_Q((S_t \prod_{u=t}^{T-1} R_u - K)^+) = v(t, S_t) \end{aligned}$$

where

$$\begin{aligned} v(t,x) &= (1+r)^{t-T}\mathbb{E}_Q(x \prod_{u=t}^{T-1} R_u - K)^+) \\ &= (1+r)^{-(T-t)} \\ &\quad \times \sum_{u=0}^{T-t} \binom{T-t}{u} q^u (1-q)^{T-t-u}(x(1+b)^u(1+a)^{T-t-u} - K)^+. \end{aligned}$$

In particular, the price of the European call option $C = (S_T - K)^+$ is given by

$$\pi(C) = v(0, S_0) = (1+r)^{-T} \sum_{u=A}^{T} \binom{T}{u} q^u(1-q)^{T-u}(S_0(1+b)^u(1+a)^{T-u} - K)$$

where A is the first integer ν for which $S_0(1+b)^\nu(1+a)^{T-\nu} > K$. This special case is conveniently rewritten as follows: observe that setting $q' = q\frac{(1+b)}{(1+r)}$ we obtain $q' \in (0,1)$ and $1 - q' = (1-q)\frac{(1+a)}{(1+r)}$, so that

$$\pi(C) = S_0\Phi(A; T, q') - K(1+r)^{-T}\Phi(A; T, q)$$

where Φ is the complementary binomial distribution function, i.e.

$$\Phi(m; n, p) = \sum_{j=m}^{n} \binom{n}{m} p^j(1-p)^{n-j}.$$

This form of the price $\pi(C)$ of the European call is widely known as the *CRR price*.

3. Pricing Options in a Hyperfinite CRR Model

In this section we construct a hyperfinite version of the CRR model in which the parameters $a < r < b$ are chosen to fit the dynamics of the Anderson random walk. Fix any *infinite* integer N. The hyperfinite time set

$$\mathbb{T} = \{0, \Delta t, 2\Delta t, \ldots, N\Delta t\}$$

leads to the internal finitely additive probability space $(\Omega, \mathcal{A}, P)$, where

$$\Omega = \{-1, 1\}^{\mathbb{T} \setminus \{T\}}$$

and P is counting measure on the algebra $\mathcal{A}$ of all internal subsets of Ω. We can construct a multiplicative random walk on Ω (which we call the *hyperfinite CRR stock price)*:

$$S^1(\omega, t) = s_0^1 \prod_{s<t} (1 + \sigma\omega(s)\sqrt{\Delta t} + \mu\Delta t) \tag{1}$$

where $\{\omega(s) : s \in \mathbb{T}\}$ is a $*$-i.i.d. sequence with $P(\omega(0) = -1\} = 1/2 = P(\omega(0) = 1\}$ and $s_0^1, \sigma > 0$, while $\mu \in \mathbb{R}$. Here we have taken $r = 0$ for simplicity. Note that with these choices $b = \sigma\sqrt{\Delta t} + \mu\Delta t > 0 > -\sigma\sqrt{\Delta t} + \mu\Delta t$ will hold in general for infinitesimal Δt.

The behaviour of S^1 is determined by the paths of the *Anderson Brownian Motion* B, defined, as in Lindstrøm [25], section 3, as the internal function $B : \Omega \times \mathbb{T} \mapsto {}^*\mathbb{R}$ which has $B(\omega, 0) = 0$,

$$\Delta B(\omega, t) = B(\omega, t + \Delta t) - B(\omega, t) = \omega(t)\sqrt{\Delta t} \quad (t \in \mathbb{T}).$$

Writing $\mathcal{A}_t$ for the internal algebra generated by the paths up to time t, i.e. by sets of the form $[\omega]_t =]_t = \{\omega \in \Omega : \omega'(s) = \omega(s) \text{ for } s < t\}$, we have a natural filtration generated by B, and under the discounting assumption $S_t^0 \equiv 1$, we can easily find a probability measure $Q \sim P$ under which S^1 is an $(\mathcal{A}_t)$-martingale. In fact:

$$\Delta S^1(\omega, t) \equiv S^1(\omega + \Delta t) - S^1(\omega, t) = S^1(\omega, t)(1 + \sigma\Delta B(\omega, t) + \mu\Delta t - 1)$$

so that we need Q to satisfy

$$\mathbb{E}_Q(\sigma\Delta B_t + \mu\Delta t | \mathcal{A}_t) = 0.$$

Writing $q = Q(\omega(t) = 1 | \mathcal{A}_t)$ we obtain

$$q(\sigma\sqrt{\Delta t} + \mu\Delta t) + (1 - q)(-\sigma\sqrt{\Delta t} + \mu\Delta t) = 0,$$

i.e. $q = \frac{1}{2}(1 - \frac{\mu}{\sigma}\sqrt{\Delta t})$. Hence the EMM is given by:

$$Q(Q(\{\omega\}) = \prod_{t<T} \frac{1}{2}(1 - \omega(t)\frac{\mu}{\sigma}\sqrt{\Delta t}). \tag{2}$$

(This construction is a discrete version of the famous Girsanov theorem, which shows how to change the underlying probability measure in order to remove the drift term in a stochastic differential equation. Internally, this reduces to simple linear algebra.)

Let $B'(\omega, t) = B(\omega, t) + \frac{\mu}{\sigma}t$, then B' has the increments

$$\Delta B'(\omega, t) = \begin{cases} +\sqrt{\Delta} + \frac{\mu}{\sigma}\Delta t & \text{with prob. } \frac{1}{2} \\ -\sqrt{\Delta} + \frac{\mu}{\sigma}\Delta t & \text{with prob. } \frac{1}{2} \end{cases}$$

Hence B' is not an Anderson process under P (but, instead, has an infinitesimal drift $\frac{\mu}{\sigma}\Delta t$), while it is an Anderson Brownian Motion under Q.

Since S^1 is a $(Q, (\mathcal{A}_t))$-martingale, we can calculate the fair price of any claim C simply as $\mathbb{E}_Q(C)$. In particular, for the European call $C = (S^1_T - K)^+$ we obtain as above:

$$\begin{aligned} \mathbb{E}_Q(C) &= \mathbb{E}_Q((S^1(\omega, T) - K)^+) \\ &= \sum_{\omega \in \Omega} (S^1(\omega, T) - K)\, 1_{\{S^1(\omega,T)>K\}}(\omega) Q() Q(\{\omega\}) \\ &= \sum_{\omega \in \Omega} (s^1_0 \prod_{t<T} (1 + \sigma\omega(t)\sqrt{\Delta t} + \mu\Delta t) - K) \\ &\quad \times 1_{\{S^1(\omega,T)>K\}}(\omega) \prod_{t<T} \frac{1}{2}(1 - \omega(t)\frac{\mu}{\sigma}\sqrt{\Delta t}) \\ &= \sum_{j=0}^{N} \binom{N}{j} (s^1_0 u^j d^{N-j} - K)\, 1_{\{s^1_0 u^j d^{N-j} > K\}}(j) q^j (1-q)^{N-j} \end{aligned}$$

with $u = 1 + \sigma\sqrt{\Delta t} + \mu\Delta t$, $d = 1 - \sigma\sqrt{\Delta t} + \mu\Delta t$, $q = \frac{1}{2}(1 - \frac{\mu}{\sigma}\sqrt{\Delta t})$ as above. Again $d(1-q) = 1 - uq$, so that, finally, we have derived the *hyperfinite CRR pricing formula*:

$$\begin{aligned} \Pi(C) &= \mathbb{E}_Q(C) \\ &= s^1_0 \sum_{j=A}^{N} \binom{N}{j} (uq)^j (1-uq)^{N-j} - K \sum_{j=A}^{N} \binom{N}{j} q^j (1-q)^{N-j} \end{aligned} \tag{3}$$

where A is the first integer ν such that $s^1_0 u^\nu d^{N-\nu} > K$.

Exercise: Show that if the assumption $r = 0$ is relaxed, so that the bond price is given as

$$S^0_t = s^0_0(1 + r\Delta t)^{t/\Delta t} \tag{4}$$

for each $t \in \mathbb{T}$, then the measure Q satisfies

$$Q(\{\omega\}) = \prod_{t<T} \frac{1}{2}(1 - \omega(t)\frac{(\mu - r)}{\sigma}\sqrt{\Delta t}). \tag{5}$$

In the (hyperfinite) binomial pricing model we have found the EMM Q and again computed the fair price for a European call option C as $\mathbb{E}_Q(C)$. Rather than repeat the calculations we want to illustrate how *both* the discrete and continuous-time model are already contained in the above hyperfinite one. To do this we simply need to exhibit the Black-Scholes model as the standard part of our hyperfinite CRR model.

For the price process this presents no difficulty: first note that whenever $B(\omega, t)$ is finite, then

$$\prod_{s<t}(1 + \sigma\Delta B(\omega, s) + \mu\Delta t) \approx \exp(\sigma B(\omega, t) + (\mu - \frac{\sigma^2}{2})t). \tag{6}$$

Sketch of Proof: (see [9], Lemma 3.1 for details) Recall that for $|x| < \frac{2}{3}$, $|\log(1 + x) - (x - \frac{x^2}{2})| \leq |x|^3$. Write

$$\begin{aligned}
\log(\prod_{s<t}(1 + \sigma\Delta B_s + \mu\Delta t) &= \sum_{s<t}\log(1 + \sigma\Delta B_s + \mu\Delta t) \\
&= \sum_{s<t}(\sigma\Delta B_s + \mu\Delta t - \frac{1}{2}(\sigma\Delta B_s + \mu\Delta t)^2 + \epsilon_s)
\end{aligned}$$

where $\epsilon_s \leq |\sigma\Delta B(\omega, s) + \mu\Delta t|^3 \leq (\Delta t)^{3/2}(|\sigma| + |\mu|)^3$. Multiplying out the square and using $(\Delta B)^2 = \Delta t$ it is now easy to see that the final sum is infinitely close to $\sigma B(\omega, t) + (\mu - \frac{1}{2}\sigma^2)t$, and the $\mathcal{S}-$continuity of the exponential function does the rest. □

We have therefore shown that

$$S^1(\omega, t) \approx s_0^1 \exp(\sigma B(\omega, t) + (\mu - \frac{\sigma^2}{2})t)$$

for $t \in \mathbb{T}$ and P_L-almost all ω, since B is a.s. finite. Moreover, finite integer powers of the product in (6) have finite P-expectation, since for each finite m,

$$(1 + \sigma\Delta B_s + \mu\Delta t)^m = 1 + \alpha\Delta B_s + \beta\Delta t$$

for some finite α, β, again because $\Delta B_s^2 = \Delta t$ for all $s \in \mathbb{T}$. This ensures that S^1 is $\mathcal{S}L^p$ for each finite p, by the following useful lemma due to Lindstrøm:

Lindstrøm's Lemma If F is an internal random variable and $\mathbb{E}_P(F^2)$ is finite, then F is $\mathcal{S}$-integrable.

Proof: It is clear that $\mathbb{E}(|F|)$ is finite. For $A \in \mathcal{A}$ with $P(A) \approx 0$ the Schwartz inequality yields

$$\int_A |F|dP \leq (P(A))^{\frac{1}{2}}(\mathbb{E}_P(F^2)^{\frac{1}{2}} \approx 0.$$

□

In summary: the continuous process $s^1 : \Omega\times\mathbb{T} \mapsto \mathbb{R}$ defined by $s^1(\omega, {}^\circ t) = {}^\circ(S^1(\omega, t))$ for $t \in \mathbb{T}$ satisfies the identity

$$s^1(\omega, \tau) = s_0^1 \exp(b(\omega, \tau) + (\mu - \frac{\sigma^2}{2})\tau) \tag{7}$$

for all $\tau \in [0, T]$ and P_L–a.a. $\omega \in \Omega$. Trivially, the standard part of S^0 is

$$s^0(\omega, t) = s_0^0 e^{rt}. \tag{8}$$

Hence: the Black-Scholes price process (s^0, s^1) can be defined on $(\Omega, \mathcal{A}_L, P_L)$ (for P_L-a.a. $\omega \in \Omega$) as the standard part of the hyperfinite CRR price process (S^0, S^1). Stating this in terms of liftings we have:

Theorem 3.1 *The hyperfinite CRR price process given on* $(\Omega, \mathcal{A}, P)$ *by* (4, 1) *is an* $\mathcal{S}$*–continuous* $\mathcal{S}L^2_P$*–lifting of the Black-Scholes price process defined on the Loeb space* $(\Omega, \mathcal{A}_L, P_L)$ *by* (8, 7).

This result can be used to give a direct derivation of the Black-Scholes formula, by a careful calculation of standard parts in the CRR formula. We refer to [9] for details - this provides a good exercise in the use of elementary hyperfinite probability and the nonstandard Central Limit Theorem. The consequence is the following:

Theorem 3.2 *Given a European call option* $c = (s^1 - K)^+$ *on the stock in a Black-Scholes model defined on the above Loeb space* $(\Omega, \mathcal{A}_L, P_L)$*, then* $C = (S^1 - K)^+$ *is an* $\mathcal{S}L^2$*–lifting of* c*, the CRR option price* $\Pi(C)$ *is given by the CRR formula* (3) *and* $\pi(c) = {}^\circ(\Pi(C))$ *agrees with the Black-Scholes formula.*

More generally, combining hyperfinite liftings with the equivalent martingale measure (EMM) approach gives us a method for calculating option prices - again we illustrate this for the CRR and BS models: recall that the EMM Q for the CRR model is given by (5), so that its density relative to the uniform counting measure P is

$$\frac{Q}{P} = \prod_{s<T}(1 - (\frac{\mu - r}{\sigma})\Delta B_s) \approx \exp(-(\frac{\mu - r}{\sigma})B_T - (\frac{\mu - r}{\sigma})^2 T).$$

This density is $\mathcal{S}$-integrable, so its standard part is

$$°(\frac{Q}{P}) = \frac{dQ_L}{dP_L} = \exp(-(\frac{\mu - r}{\sigma})b_T - (\frac{\mu - r}{\sigma})^2 T),$$

which is the Girsanov density obtained by standard means for the Black-Scholes price. For a general contingent claim c in the BS model we can therefore use the hyperfinite machinery according to the following recipe:

(i) find an $\mathcal{S}L^2_P$-lifting C of c,
(ii) compute the EMM Q for the hyperfinite price process S
(iii) calculate $\Pi(C) = \mathbb{E}_Q(\overline{C})$ since the current price of C is the Q-expectation of $\overline{C} = (1+r)^{-N}C$,
(iv) find $\pi(c) = °(\Pi(C))$ and note that $\pi(c) = \mathbb{E}_{Q_L}(e^{-rt}c)$.

The final step follows from the identities (we have taken $r = 0$ for simplicity) - note the role of $\mathcal{S}$-integrability:

$$\Pi(C) = °(\mathbb{E}_Q(C)) = °(\mathbb{E}_P(\frac{Q}{P}C)) = \mathbb{E}_{P_L}(°(\frac{Q}{P})°C) = \mathbb{E}_{P_L}(\frac{dQ_L}{dP_L}c) = \mathbb{E}_{Q_L}(c).$$

4. Hyperfinite Trading Strategies

It is easy to describe the unique self-financing trading strategy $\Theta = (\Theta^0, \Theta^1)$ which generates the contingent claim C in the hyperfinite CRR model. We shall assume take $S^0 \equiv 1$ throughout. Recall that a *trading strategy* is a nonanticipating function $\Theta : \Omega \times \mathbb{T} \mapsto {}^*\mathbb{R}^2$ yielding a *value process* $V_t = \Theta_t \cdot S_t$ and *gains process*

$$G_t = \Sigma_{u<t}\Theta_u \cdot \Delta S_u = \left(\int \Theta dS\right)(\cdot, t)$$

The strategy Θ is *self-financing* if changes in value are accounted for solely by gains, i.e. $G_t = V_t - V_0$. As we saw earlier, this is equivalent to: $\Theta^1_t \Delta S^1_t = \Delta V_t$, and (since we use forward increments) it means that $\Delta\Theta_t$ is orthogonal to $S_{t+\Delta t}$ in ${}^*\mathbb{R}^2$, i.e. that

$$\Theta_t \cdot S_{t+\Delta t} = \Theta_{t+\Delta t} \cdot S_{t+\Delta t}.$$

We also say that Θ *generates* the claim C if $C_T = \Theta_T \cdot S_T$. Thus:

Lemma: A trading strategy Θ is self-financing and generates the claim C iff for all $t \in \mathbb{T}$:

$$\mathbb{E}_Q(C|\mathcal{A}_t) = V_t = V_0 + G_t. \tag{9}$$

For any C in a (hyper-)finite market model the unique such trading strategy can easily found in terms of C by simple linear algebra, by calculating conditional expectations w.r.t. the sets

$$A_1 = A_1 = \{\omega' : \omega'\lceil t = \omega\lceil t \text{ and } \omega'(t) = 1\}$$

and

$$A_{-1} = \{\omega' : \omega'\lceil t = \omega\lceil t \text{ and } \omega'(t) = -1\}.$$

To do this, note that since $\Delta S^0 \equiv 0$, (9) can be written as follows (we also write $\mathbb{E}_Q$ as $\mathbb{E}$ for convenience):

$$\Gamma_t = \mathbb{E}(C|\mathcal{A}_t) = V_0 + \sum_{u<t} \Theta_u^1 \Delta S_u^1$$

so that for $i = \pm 1$, and recalling that $A_i \in \mathcal{A}_{t+\Delta t}$,

$$\mathbb{E}(C|A_i) = \Gamma_t + \Theta_t^1 \mathbb{E}(\Delta S_t^1 | A_i).$$

Since $S_t^1 = S_0^1 + \sum_{u<t} \Delta S_u^1$, and S^1 is a martingale it follows that

$$\mathbb{E}(S_T^1 | A_i) = S_t^1 + \mathbb{E}(\Delta S_t^1 | A_i).$$

This leads to the unique solution for Θ^1 and hence Θ^0 :

$$\Theta_t^1 = \frac{\mathbb{E}(C|A_1) - \mathbb{E}(C|A_{-1})}{\mathbb{E}(S_T^1|A_1) - \mathbb{E}(S_T^1|A_{-1})} \qquad \Theta_t^0 = \mathbb{E}(C|\mathcal{A}_t) \ - \ \Theta_t^1 \cdot S_t^1 \tag{10}$$

However, we cannot simply take standard parts: the denominator in Θ_t^1 is infinitesimal, in fact it equals $S_t^1(u-d) = 2S_t^1\sigma\sqrt{\Delta t}$. This suggests that one can construct contingent claims for the hyperfinite model whose generating strategies do *not* have nice integrability properties - in the following example, which is taken from [9], the terminology used below is that used by Lindstrøm [25]:

Example: Consider a contingent claim $C \in \mathcal{S}L_Q^2$. Then $\Gamma_t \in \mathcal{S}L_Q^2$ and the process $(\Gamma_t)_{t\in\mathbb{T}}$ is an $\mathcal{S}L_Q^2$−martingale. This means (by construction of Γ) that $\sum \Theta_t^1 \Delta S_t^1$ defines an $\mathcal{S}L_Q^2$−martingale. Nonetheless Θ^1 need not belong to $\mathcal{S}L^2(S^1)$: taking the claim $C = \sqrt{N}\Delta S_u^1$, where $u \in \mathbb{T}$ is fixed, we can write down a generating strategy Θ with

$$\Theta^1(\omega, t) = \mathbf{1}_{\{t=u\}}(t)\sqrt{N}.$$

Although the claim C belongs to $\mathcal{S}L_Q^2$ (by Lindstrøm's lemma), it is not hard to see that the strategy Θ is not in $\mathcal{S}L^2(S^1)$.

The cause of such pathology lies in the fact that the gains process is defined as a stochastic integral $\int \Theta \cdot dS$, and, under Q, the λ^2-martingale S^1 is used as an integrator. For a general internal claim it is therefore possible to construct a strategy Θ for which S and $\int \Theta \cdot dS$ are $\mathcal{S}L^2_Q$–martingales, but Θ does not belong to the space $\mathcal{S}L^2(S)$. This means that we cannot exhibit such Θ as an $\mathcal{S}L^2$–lifting of a (predictable) process θ such that the integral $\int \theta \cdot ds$ is lifted by $\int \Theta \cdot dS$.

However, if we restrict attention to claims C which *arise* as liftings of claims $c \in L^2(\mathcal{F}_t)$, where we recall that $\mathbb{F} = (\mathcal{F}_t)$ is the *Brownian* filtration, i.e. that generated by the BS price $s = (s^0, s^1)$ on the Loeb space $(\Omega, \mathcal{A}_L, P_L)$, we can connect our hyperfinite model more successfully with the BS model:

Fix $c \in L^2(\mathcal{F}_t)$. By the representation theorem for Brownian martingales, the martingale $\gamma = (\gamma_t)_{t\in[0,T]}$ given by $\gamma_t = \mathbb{E}_{Q_L}(c|\mathcal{F}_t)$ can be expressed as a stochastic integral

$$\gamma_t = \mathbb{E}_{Q_L}(c) + \int_0^t \theta^1_u ds^1_u \tag{11}$$

where $\theta^1 \in L^2(\nu_{s^1})$. Here ν_{s^1} is the *Doleans measure* of s^1; see [21] for the general definition, but note also that in our case, since $d\nu_{s^1} = \sigma^2(s^1)^2 dt$, the integrability condition means simply that $\mathbb{E}_{Q_L \times m}\left[(\theta^1 s^1)^2\right] < \infty$, where m denotes Lebesgue measure on $[0,T]$. Setting $\theta^0_t = \gamma_t - \theta^1_t s^1_t$ yields the unique (predictable) generating strategy $\theta = (\theta^0, \theta^1)$ for the claim c in the BS model on $(\Omega, \mathcal{A}_L, P_L)$. We need to define the right kind of (bipedal) lifting for this process.

Definition 4.1 (cf. Lindstrøm [25], Def. 9.1, Thm 12.2) *For any $\mathcal{S}$-continuous $\mathcal{S}L^2_Q$-martingale X define the internal measure ν_X by*

$$\nu_X(u_X(\{\omega, t\}) = \Delta X(\omega, t)^2 Q(^2Q(\{\omega\}).$$

Let $x = {}^\circ X$ be the standard part of X and let ν_x be its Doleans measure. If $\psi \in L^2(\nu_x)$ is predictable, then a 2-lifting *of ψ is a nonanticipating process $\Psi : \Omega \times \mathbb{T} \mapsto {}^*\mathbb{R}$ which belongs to $\mathcal{S}L^2(X)$ and satisfies*

$${}^\circ(\Psi(\omega, t)) = \psi(\omega, {}^\circ t)$$

for $(\nu_X)_L$-almost all (ω, t).

Theorem 4.2 (see Lindstrøm [25], Thm 12.2) *Under the above hypotheses on x and ψ, 2-liftings exist and $\int \psi dx = {}^\circ \left(\int \Psi dX\right)$.*

We aim to show that the hyperfinite generating strategy Θ constructed as in (10) for any $\mathcal{S}L^2$-lifting C of our given claim $c \in L^2(\mathcal{F}_T)$ qualifies as a 2-lifting of the generating strategy θ for c. First consider Θ^1: we need to show that ${}^\circ(\Theta^1(\omega,t)) = \theta^1(\omega,{}^\circ t)$ $(\nu_{S^1})_L$-a.s. and that $\Theta^1 \in \mathcal{S}L^2(S^1)$. We know that θ^1 has a nonanticipating 2-lifting $\overline{\Theta}^1 \in \mathcal{S}L^2(S^1)$, so that

$$c = \gamma_0 + {}^\circ\left(\int_0^T \overline{\Theta}_t^1 dS_t^1\right)$$

holds Q_L-a.s., i.e. $\gamma_0 + \int_0^T \overline{\Theta}_t^1 dS_t^1$ is an $\mathcal{S}L_Q^2$-lifting of c. But so is $C = \Gamma_0 + \int_0^T \Theta_t^1 dS_t^1$ and Γ_0 lifts γ_0, hence

$$0 \approx \mathbb{E}_Q\left[\left(\int_0^T (\Theta_t^1 - \overline{\Theta}_t^1) dS_t^1\right)^2\right] = \int_{\Omega\times\mathbb{T}} (\overline{\Theta}^1 - \Theta^1)^2 d\nu_{S^1}.$$

This shows that Θ^1 is a 2-lifting of θ, as required. In particular,

$${}^\circ\left(\int_0^t \Theta_u^1 dS_u^1\right) = \int_0^{{}^\circ t} \theta_\tau^1 ds_\tau^1$$

for $t \in \mathbb{T}$, Q_L-a.s. (The corresponding statement for Θ^0 follows easily.)

This proves the essential implication $(i) \to (ii)$ in the following fundamental result ($\Lambda = \Lambda_N$ denotes counting measure on $\mathbb{T}$):

Theorem 4.3 *With the above notation the following are equivalent:*
(i) C is an $\mathcal{S}L_Q^2$-lifting of c
(ii) Θ^0 is an $\mathcal{S}L^2(Q\times\Lambda)$-lifting of θ^0 and Θ^1 is an $\mathcal{S}L^2(\nu_{S^1})$-lifting of θ^1 [alternatively: $\Theta^0, \Theta^1 S^1$ are $\mathcal{S}L^2(Q\times\Lambda)$-liftings of $\theta^0, \theta^1 s^1$ respectively]
(iii) $V(\omega,\cdot)$ is Q_L-a.s. $\mathcal{S}$-continuous and for $t \in \mathbb{T}$, $V(\cdot,t)$ is an $\mathcal{S}L^2$-lifting of $v(\cdot,{}^\circ t)$.

The remaining implications are elementary consequences of the definitions and basic Loeb theory. hence we have verified that for *every* infinite N the $\mathcal{S}L^2$-liftings we have constructed are preserved under the operations of the stochastic calculus - this idea will be made precise and exploited in the next section.

5. Convergence of Prices and Strategies

The content of Theorem 4.3 is to exploit the fact that in the BS model as well as in the CRR model, knowledge of any one member of the triple (c,θ,V) (or (C,Θ,V)) will *determine* the other two, by showing that this property is preserved under $\mathcal{S}L^2$-liftings. Given a sequence of CRR models with (C_n,Θ_n,V_n) defined as above for each n, and a BS model with corresponding triple (c,θ,v), it is natural to ask in what sense these CRR triples converge to their BS counterpart. The underlying dynamics are determined by normalised random walks B_n and a Brownian motion b respectively, so that *Donsker's theorem* (see [21]) immediately supplies the fact that $B_n \to b$ *weakly*, i.e. that the distributions of values of the B_n converge to those of b. Since in general the stochastic processes B_n and b (and hence also the associated price processes S_n and s) are defined on different probability spaces, this seems the only plausible mode of convergence.

However, in discussing the convergence of claims or strategies, we would hope to be able to include information on the functional relation between C_n and B_n and c and b, for example, so that a mode of 'weak convergence along the graphs' appears desirable. This idea can be related to that of liftings, as we shall see.

In a nonstandard setting convergence of C_n to c could be defined as meaning that C_N is an $\mathcal{S}L^2$–lifting of c for every infinite N. In general this does not appear to make sense for an arbitrary Brownian functional c; however, if $c \in L^2(\Omega, \mathcal{F}_T, Q)$, where $(\mathcal{F}_t)$ is the Brownian filtration generated by b, then $c = c(b)$, and we can say that C is an $\mathcal{S}L^2(Q_N)$ – *lifting* of c if C is in $\mathcal{S}L^2(Q_N)$ and ${}^\circ C(\omega) = c({}^\circ B_N(\omega))$ $L(Q_N)$-a.s..

To avoid such notational devices we shall construct our BS model on *Wiener space* from the beginning, i.e. take as our underlying probability the path space

$$\Omega = \mathcal{C} = \{\omega \in C[0,T] : \omega(0) = 0\}$$

and define a sequence of CRR models with $\Omega_n = \mathcal{C}_n$ as the path space of the simple random walk B_n which begins at 0 and has step size $\sqrt{\Delta_n}$ where $\Delta_n = \frac{T}{n}$. To embed $\mathcal{C}_n$ in $\mathcal{C}$ we fill in linearly between points in the graph of B_n. We are then able to use liftings to construct, for a given claim $c \in L^2(\Omega, \mathcal{F}_T, Q)$, a sequence of CRR claims (C_n) which converges to c in the appropriate sense.

A particularly desirable feature of this convergence theory will be the preservation of convergence under the operations of the stochastic calculus, i.e. that the claims converge if and only if their generating strategies (i.e. stochastic integrands) do, and similarly for the associated value processes (stochastic integrals). This is exactly what Theorem 4.3 yields, as well as providing nonstandard formulation of the convergence concept we need: we should require that C_N is an $\mathcal{S}L^2(Q_N)$-lifting of c for each infinite N. We now develop the notation to make these ideas more precise, and at the same time we give two standard formulations of this convergence concept.

In the CRR model we consider the discrete time set $\mathbb{T}_n = \{k\Delta_n : 0 \leq k \leq n\}$ where $\Delta_n = \frac{T}{n}$, and the 'path space' $\mathcal{C}_n = \{B_n(\omega, \cdot) : \omega \in \Omega_n\}$ is derived from and identified with the finite probability space $\Omega_n = \{-1,1\}^{\mathbb{T}_n \setminus \{T\}}$, together with the algebra $\mathcal{A}_n$ of all its subsets and P_n as the counting probability on Ω_n. An element ω of Ω_n is thus an i.i.d. sequence $(\omega(t))_{t\in\mathbb{T}_n}$ with $P_n(\omega(0) = -1) = P_n(\omega(0) = 1) = \frac{1}{2}$, and the normalised random walk B_n with $B_0 = 0$ has increments

$$\Delta B_n(t,\omega) = B_n(t+\Delta_n,\omega) - B_n(t,\omega) = \omega(t)\sqrt{\Delta_n}.$$

To treat $\mathcal{C}_n$ as a subset of $\mathcal{C}$ we join the points of $B_n \in \mathcal{C}_n$ linearly to yield a polygonal path in $\mathcal{C}$.

Again we take the bond price $S_n^0 \equiv 1$ and we define the stock price S_n^1 by

$$S_n^1(t,\omega) = s_0^1 \prod_{s<t} (1 + \mu\Delta_n + \sigma\Delta B_n(t,\omega)).$$

The equivalent martingale measure Q_n (relative to the filtration $\mathbb{A}_n = (\mathcal{A}_{n,t})_{t\in\mathbb{T}_n}$ of the algebras generated by $\{B_n(s,\cdot) : s \le t\}$, or equivalently by $\{\omega(s) : s < t\}$) is then given with density

$$\frac{dQ_n}{dP_n}(\omega) = \prod_{t<T} (1 - \frac{\mu}{\sigma}\Delta B_n(t,\omega)),$$

as we have seen. The Doleans measure ν_n for S_n^1 under Q_n is given by $\nu_n(u_n(\{\omega,t\}) = (\Delta S_n^1(\omega,t))^2 Q_n(n(\{\omega\})$. Clearly P_n and Q_n can be viewed as measures on $\mathcal{C}_n$. We refer to

$$\underline{\Omega_n} = (\Omega_n, \mathcal{A}_n, P_n, B_n, S_n^0, S_n^1, \mathbb{A}_n, Q_n)$$

as the n-th CRR model. A trading strategy in this model is an $\mathbb{A}_n$-adapted process $\Theta_n = (\Theta_n^0, \Theta_n^1)$ generating the value process

$$V_{n,t} = \Theta_{n,t}^0 + \Theta_{n,t}^1 S_n^1(t)$$

and the gains process

$$G_{n,t} = \sum_{u<t} \Theta_{n,u}^1 \Delta S_n^1(u)$$

records the gains from trading at dates $u \in \mathbb{T}_n$. The strategy is self-financing if $V_{n,t} = V_{n,0} + G_{n,t}$ for all $t \in \mathbb{T}_n$. Every L^2–claim C_n on $\underline{\Omega_n}$ is generated by a unique self-financing strategy Θ_n (by a trivial 'martingale representation theorem'), so that in this setting all the models under consideration are *complete*.

5.1. WEAK CONVERGENCE CRITERIA

The finance literature contains much work on criteria for the *weak* convergence of option prices and trading strategies, using heavy machinery from stochastic analysis, such as regularity conditions on integrands and integrators which ensure weak convergence of the corresponding stochastic integrals. However, weak convergence is not normally preserved under these operations, nor under the formation of suprema, such as are needed in the analysis of American options. In proposing the mode of convergence which arises naturally via liftings, we need to examine how it improves on weak convergence, and to do this we first consider nonstandard criteria for the latter, given independently by Loeb [26] and Anderson and Rashid [3].

Recall that 'weak' in this context actually means 'weak-star' (not to be confused with '$*$-weak'!) since we wish to describe a topology on the dual space of $C(X)$ where X a topological space and $C(X)$ denotes the space of bounded continuous real-valued functions on X – for our purposes we can take X as a separable metric space; in fact our applications will always have $X = \mathcal{C}$ or $X = \mathcal{C} \times [0,T]$.

A net (ϕ_α) in the topological dual $C(X)'$ *converges weakly* to ϕ iff $\phi_\alpha(f) \to \phi(f)$ for all $f \in C(X)$. Write $\phi_\alpha \to_w \phi$. Let $\mathcal{F}$ be the Borel σ-field on X. Then can identify each $\phi \in C(X)'$ with a finitely additive measure μ_ϕ on $(X, \mathcal{F})$ via $\phi(f) = \int_X f d\mu_\phi$. It follows in particular that for probability measures, $\mu_\alpha \to_w \mu$ iff $\int_X f d\mu_\alpha \to \int_X f d\mu$ for all $f \in C(X)$.

Let ${}^*X$ be nonstandard extension of X; as usual, the topology $\mathcal{T}$ on X will determine the degree of saturation we require. We shall take $\mathcal{T}$ as the topology with subbase $\{f^{-1}((-\infty, \alpha)) : f \in C(X), \alpha \in \mathbb{R}\}$. For our application to $\mathcal{C}$ or $\mathcal{C} \times [0,T]$ it is clear that ω_1-saturation would suffice, since here $\mathcal{T}$ is countably generated. Define the standard part map $\mathrm{st} : {}^*X \mapsto X$ by $\mathrm{st}(y) = x$ iff y is in the monad $m(x) = \cap_{x \in T \in \mathcal{T}} {}^*T$.

For an internal finitely additive probability ν on ${}^*\mathcal{F}$ we form the Loeb algebra $({}^*\mathcal{F})_L$ and the Loeb measure ν_L. Then ν is *nearstandardly concentrated (nsc)* if $\nu_L({}^*X \setminus \mathrm{ns}({}^*X)) = 0$. For such ν, $\mu(B) = \nu_L(\mathrm{st}^{-1}(B))$ defines a probability measure on $(X, \mathcal{B})$, where $\mathcal{B}$ is the Baire σ-field on X. To see this we must show that $\mathrm{st}^{-1}(B)$ is Loeb-measurable for all $B \in \mathcal{B}$. But if $f \in C(X)$, then

$$\begin{aligned} \mathrm{st}^{-1}(\{x \in X : f(x) < \alpha\}) &= \{z \in \mathrm{ns}({}^*X) : {}^\circ({}^*f(z)) < \alpha\} \\ &= \mathrm{ns}({}^*X) \cap (\bigcup_{n=1}^{\infty} \{z \in {}^*X : {}^*f(z) < \alpha - \frac{1}{n}\}) \in ({}^*\mathcal{F})_L. \end{aligned}$$

Consider a functional $\phi \in {}^*C(X)$. When the internal probability $\nu = \nu_\phi$ is nsc, its counterpart $\mu = \mu_\phi$ is *tight* (i.e. essentially supported on a compact set) and $\phi({}^*f) \approx \int_X f \mu_\phi$, so that ϕ is in the weak-star monad of μ_ϕ. (See [3] and [26] for details of the proof.)

Theorem 5.1 (Loeb, Anderson-Rashid) *For bounded nets in $C(X)'$ weak convergence of $(\phi_\alpha)_{\alpha\in D}$ to a (tight) measure on X is equivalent to the following: for all infinite $\alpha, \beta \in D$, ν_ϕ is nsc and $\mu_\alpha = \mu_\beta$ (in which case the weak limit of (ϕ_α) is their common value).*

We specialise to sequences (see also the comments in section 3 of [29], and recall that Borel and Baire sets coincide in this special case): suppose we are given a Borel probability measure μ defined on the separable metric space X and a sequence (μ_n) of probability measures (each supported on a Borel set $X_n \subset X$) converging weakly to μ, then for any infinite N, the internal probability ν_N satisfies:

$$\mu = (\mu_N)_L \circ \mathrm{st}^{-1}.$$

In particular, $(\mu_N)_L$-almost all points $x \in X_N$ are near-standard.

This provides us with the standard characterisation of our nonstandard convergence concept for claims and strategies in the BS-model on Wiener space:

To phrase the set-up in terms of 'weak convergence along the graphs' we introduce the following notation: let $F_n : X_n \mapsto \mathbb{R}$ and $f : X \mapsto \mathbb{R}$ be (Borel) measurable functions on X; for any Borel set $A \subset X \times \mathbb{R}$ define the measures:

$$\begin{aligned} \nu_n(A) &= \mu_n(\{x \in X_n : (x, F_n(x)) \in A\} \\ \nu(A) &= \mu(\{x \in X : (x, f(x)) \in A\}). \end{aligned}$$

Theorem 5.2 *Assume that $\mu_n \to_w \mu$ and that F_n, f, ν_n, ν are as defined above. The following are equivalent:*
(i) For every infinite N, F_N lifts f, i.e. $F_N(z) \approx f({}^\circ z)$ holds for $(\mu_N)_L$-almost all $z \in X_N$;
(ii) $\nu_n \to_w \nu$. (We also describe this by saying that the pairs $(x, F_n(x))$ converge *to* $(x, f(x))$ weakly *as $n \to \infty$.)*

If either of the above holds then F_N is $SL^2(\mu_N)$ for all infinite N iff $\mathbb{E}_{\mu_n}(F_n^2) \to \mathbb{E}_\mu(f^2)$ as $n \to \infty$.

Proof: If F_N lifts f then we have

$$\begin{aligned} (\mu_N)_L\{z : ({}^\circ z, {}^\circ F_N(z)) \in A\} &= (\mu_N)_L\{z : ({}^\circ z, f({}^\circ z)) \in A\} \\ &= \mu\{x : (x, f(x)) \in A\} \qquad (12) \end{aligned}$$

using the fact that, since $\mu_n \to_w \mu$, we have $\mu = (\mu_N)_L \circ \mathrm{st}^{-1}$ for all infinite N. In particular, the left and right extremes in (12) are equal for each infinite N, which means that $\nu_N(A) \approx \nu(A)$ for all such N, and thus (ν_n) converges weakly to ν.

Conversely, if $\nu_n \to_w \nu$, the extremes in (12) are equal. Since f is Loeb-measurable, it has a lifting F, say. Hence we also have

$$(\mu_N)_L\{z : (^\circ z, {}^\circ F(z)) \in A\} = \mu\{x : (x, f(x)) \in A\}.$$

In other words, under the Loeb measure $P = (\mu_N)_L$, $^\circ F$ and $^\circ F_N$ have the same distribution. This does not yet show that they are almost surely equal. But fix reals $a < b$ and consider $D = \{z \in X_N : F_N(z) \leq a < b \leq F(z)\}$. To show that D is a P-null set, first choose an *internal* set $D_0 \subset D$ with $P(D_0) \geq \frac{1}{2}P(D)$ and $D_0 \subset \{z : {}^\circ F(z) = f(^\circ z)\}$ (the latter has full measure). Let $C = \mathrm{st}(D_0)$, which is closed, as a standard part of a set in $^*X$. Since $^\circ F$ and $^\circ F_N$ have the same distribution, it follows that the sets $S_1 = \{^\circ F_N(z) \leq a, {}^\circ z \in C\}$ and $S_2 = \{^\circ F(z) \leq a, {}^\circ z \in C\}$ have the same P-measure. Now S_1 contains D_0, so that $P(D_0)$ is no greater than the common measure S_1 and S_2. But if $z \in S_2$ then $^\circ z = {}^\circ z'$ for some $z' \in D_0$, which means that $f(^\circ z) = f(^\circ z') = {}^\circ F(z')$, as F lifts f, and $^\circ F \geq b$ on D_0. Hence $S_2 \subset (z : {}^\circ F(z) \neq f(^\circ z)\}$ and so is null. Hence $P(D) = 0$, which means that $^\circ F(z) \leq {}^\circ F_N(z)$ P-a.s.. As they have the same distribution they are now a.s. equal. But then F_N is also a lifting of f. This completes the proof of the theorem. □

Corollary 5.3 *Let (C_n) be a sequence of contingent claims in the CRR models $\underline{\Omega_n}$ and let $c \in L^2(Q, \mathcal{F}_T)$. The following are equivalent:*
(a) For each infinite N, C_N is an $SL^2(Q_N)$-lifting of c.
(b) $(B_n, C_n(B_n)) \to (b, c(b))$ weakly and $\mathbb{E}_{Q_n}(C_n^2) \to \mathbb{E}_Q(c^2)$.

Definition 5.4 *If either of the above statements holds, we say that the sequence (C_n)* D^2-converges *to c. (The reason for the terminology will become evident in the next paragraph.)*

Exercise: Give an example of a sequence of claims (depending only on the final price) which converge weakly but are not D^2-convergent.

5.2. ADAPTED DISCRETISATIONS OF WIENER SPACE

Our definition is stronger than weak convergence, since it keeps track of the functional behaviour of the limiting sequence, i.e. the relationship between C_n and B_n is fixed to 'resemble' that between c and b, at least for large n. It is this which provides some of the stability properties of D^2-convergence, and allows us to formulate the concept in a further, intuitively appealing, way:

Definition 5.5 *A family (d_n) of measurable maps $C \mapsto C_n$ is an* adapted Q-discretisation scheme *if for each n:*
(i) d_n is adapted;
(ii) d_n is measure-preserving;
(iii) $d_n(b) \to b$ in Q-probability,
i.e. $\forall\epsilon > 0 : Q(|d_n(b) - b| < \epsilon) \to 1$ as $n \to \infty$. (Here $|.|$ denotes the sup norm in C.)

The existence of such a scheme can be established by modifying a construction given by Frank Knight in 1962, using polygonal paths approximating Brownian motion. Here we give only the barest outline of the construction, which is given in [10]:

under Q the process $\bar{b}(u) = b(u) + \alpha u$ $(u \in [0,T])$ is a Brownian motion, where $\alpha = \frac{\mu}{\sigma}$, and the stock price is the exponential martingale

$$s^1(u) = s^1_0 \exp(\sigma\bar{b}(u) - \frac{1}{2}\sigma^2 u).$$

In the n-th CRR model approximating the BS model on $\Omega = C$ we have the discrete martingale (for $t \in \mathbb{T}_n$):

$$\overline{B}_n(t) = B_n(t) + \alpha t$$

and setting $p_n = \frac{1}{2}(1 - \alpha\sqrt{\Delta t})$, $q_n = 1 - p_n$, we see that B_n has increments $\Delta B_n(t) = \pm\sqrt{\Delta t}$ w.p. p_n, resp. q_n. Defining the stopping time $e(b) = \min\{u : |\bar{b}(u) - \alpha\Delta t| = \sqrt{\Delta t}\}$ for the Brownian path b, we see by optional sampling that $\bar{b}(e)$ has the same distribution as the increments of $\overline{B}_n$, so that we can define a sequence of stopping times inductively by setting:

$$\begin{aligned} e_0 &= 0 \\ e_{k+1}(b) &= \min\{u > e_k(b) : |\bar{b}(u) - \bar{b}(e_k) - \alpha\Delta t| = \sqrt{\Delta t}\} \end{aligned}$$

Then $e_1 = e$, and the increments $e_{k+1} - e_k$ are independent copies of e. Using Knight's results one may show that

$$Q(\max_{k\leq n}|e_k - t_k| > n^{-1/5}) \leq \kappa n^{-6/5}$$

for some constant κ which is independent of n. The 'Knight scheme' $d^0_n : C \mapsto C_n$ is then given by $(d^0_n(b))(t_k) = \bar{b}(e_k) - \alpha t_k$. This modification of Knight's construction thus takes care of the measure change to Q, and we can again check that $Q((d^0_n)^{-1}(B)) = Q_n(B)$ for $B \in C_n$, and that (as paths in C) we have $|d^0_n(b) - b| \to 0$ with probability 1.

What is left is the need to make the discretisation adapted: first restrict (to $\Delta t + 2n^{-1/5}$) the amount of 'looking ahead' that we permit in searching for the e_k, and for paths where this does not allow us to determine e_k simply ahead in steps of Δt and send our approximating paths up or down by $\sqrt{\Delta t}$, depending on where the increment of the shifted Brownian path is positive or negative between such successive 'stopping times'. The values $d_n^1(b)(t_k)$, for $t_k \in \mathbb{T}_n$, of this modified Knight scheme d_n^1 can then be shown to be determined by the values of the Brownian path b up to $t_k + 3n^{-1/5}$. Thus if σ_n is the smallest element of $\mathbb{T}_n$ after $3n^{-1/5}$ we can finally set $d_n(b)(t) = d_n(b)(\sigma_n) + d_n^1(b)(t - \sigma_n)$ for $t \geq \sigma_n$, and use the $\pm\sqrt{\Delta t}$ increments only on $[0, \sigma_n]$. Using Anderson's $\mathcal{S}$-continuity results for internal Brownian motion (see [25], section 11) we can show that this scheme still has the convergence properties, in addition to adaptedness.

We can use the discretisation scheme to produce our second characterisation of D^2-convergence:

Theorem 5.6 *A sequence of claims (C_n) in CRR models D^2-converges to the claim $c \in L^2(Q)$ in the BS model iff $C_n(d_n(\cdot))$ converges to* c *in $L^2(Q)$-norm.*

Proof: For infinite N, if C_N lifts c, then (as d_N is measure-preserving) $C_N(d_N(\beta)) \approx c({}^\circ d_N(\beta))$ for $L({}^*Q)$-almost all β in ${}^*\mathcal{C}$. On the other hand, the convergence of $(d_n(b))$ to b can be stated as: ${}^\circ d_N(\beta) = {}^\circ\beta$ for such $*$-paths β, so that, in particular, with $\beta = {}^*b$, we see that $C_N(d_N({}^*b))$ lifts c to ${}^*\mathcal{C}$. This lifting is $\mathcal{S}L^2({}^*Q)$, since $\mathbb{E}_Q(c^2)$ is infinitely close to $\mathbb{E}_{Q_N}(C_N(B_N)^2) = \mathbb{E}_{{}^*Q}(C_N(d_N({}^*b))^2)$. But ${}^*c$ is a similar lifting of c (by Anderson's Lusin Theorem – see [29]), hence they must be infinitely close in $L^2({}^*Q)$, and this means that $C_n(d_n(b))$ converges to $c(b)$ in $L^2(Q)$-norm. The converse follows by taking any $\mathcal{S}L^2(Q_N)$-lifting C of c, so that both C and C_N are infinitely close to ${}^*c$ in $L^2({}^*Q)$-norm, and thus to each other, and as d_N preserves measure, $\mathbb{E}_{Q_N}(C - C_N)^2 \approx 0$. Thus C_N also $\mathcal{S}L^2(Q_N)$-lifts c. □

Corollary 5.7 *If $C_n \to c$ in the above sense, then $\pi(C_n) \to \pi(c)$.*

We can construct a convergent sequence of claims very simply – although the proof requires some effort, see [10]: given $c \in L^2(Q)$, let $\overline{C}_n(B) = \mathbb{E}_Q(c|d_n(b) = B)$. This sequence will D^2-converge to c and any other sequence of claims (C_n) which D^2-converges to c will satisfy: $\mathbb{E}_Q((C_n - \overline{C}_n)^2) \to 0$, since both C_N and $\overline{C}_N$ are $SL^2(Q)$-liftings of c for all infinite N.

Our stability properties of D^2-convergent sequences can be reformulated (after the fairly obvious extension of the concepts to functions on $\mathbb{R} \times \mathcal{C}$) as follows:

Theorem 5.8 *If (C_n, Θ_n, V_n) are claims, together with their generating strategies and value processes on CRR models $\underline{\Omega_n}$, and (c, θ, v) are their counterparts in the BS model on Ω, then the following are equivalent:*
(i) C_n D^2-converges to c
(ii) Θ_n D^2-converges to θ
(iii) V_n D^2-converges to v.

This is just a restatement of Theorem 4.3. However, note that we now have three equivalent formulations of D^2-convergence, two of which involve no nonstandard notions. In this context it seems worthwhile emphasise the stability features of this mode of convergence.

The above result is of course meaningful outside the finance theory framework: in essence, what it shows is that our convergence concept is stable under both stochastic differentiation and stochastic integration, since we have the relations:

$$\begin{aligned} C_n &= V_n(0) + \int_0^T \Theta_n(u)dS_n(u) \\ \text{and} \qquad c &= v(0) + \int_0^T \theta_u ds_u. \end{aligned}$$

As weak convergence has no such stability properties under the stochastic calculus, our results suggest that D^2-convergence may well be a more appropriate mode of convergence when adapted processes are considered. We are also tempted to conjecture that D^2-convergence is a special case (in the very restricted situation of functions on the path space $\mathcal{C}$) of the general theory of 'convergence in adapted distribution' which has been developed in much more abstract form by Hoover and Keisler. Nonetheless, its concrete formulation via discretisation schemes is at least potentially useful in the development of numerical approximation techniques.

6. Further Developments

The discussion in these notes has concentrated on the results obtained in [9] and [10]. There have been several other applications of nonstandard methods in option pricing theory - most of these require further preliminaries of the current debates in option pricing, however.

(i) Within the Black-Scholes framework, much current research concerns the approximation of American put options, which have no closed formula solutions. A key role is played by the *Snell envelope*, which is the smallest supermartingale dominating the payoff function of the option. In this Brownian optimisation problem (formulated as a free boundary problem for PDEs) there exists a unique optimal stopping time which gives the first occurrence of the *critical price* at which it becomes optimal to exercise the option. It therefore becomes significant to approximate both this optimal time and the Snell envelope by means of a sequence of discrete models – again this could provide an algorithm with practical applications. In [14] this is done in the context of D^2-convergence of optimal times and discrete Snell envelopes to their Brownian counterpart - the uniqueness of the optimal time in the BS setting is crucial in avoiding difficulties wit the 'correct' standard parts (once again these result from the excessive size of the stochastic filtration). In particular, the lifting theorems imply that the critical prices in an appropriately constructed sequence of CRR models converge uniformly to the critical price in the BS model.

(ii) Self-similar processes such as fractional Brownian motion (FBM) have been proposed as providing more realistic models than the more common semimartingale models. It has long been observed that the time series in actual stock price data do not conform to the BS hypotheses, but have rather 'fatter tails' as well appearing to display some evidence of long-term dependence (though the latter is still hotly debated). FBM models provide an additional parameter and have been shown to be a good limiting models for long-term dependence. However, FBM is not a semimartingale, which makes it impossible to use the apparatus of stochastic calculus in its current form. More seriously for the usual economists' paradigm of market equilibrium and the 'efficient markets hypothesis', the absence of the semimartingale property means that there is no equivalent martingale measure for FBM, so that assumptions about market equilibrium can be violated and arbitrage becomes possible.

A nonstandard definition of FBM, based on a fractional version of the Anderson random walk, was constructed in [13] and arbitrage opportunities were identified in the hyperfinite model. How these might be adjusted to yield a set of fractional brownian paths of positive Loeb measure along which arbitrage is possible, remains an open question. Other (standard) discussions of FBM as a pricing model have not so far displayed an explicit set of 'arbitrage paths'. Papers on a possibly useful 'calculus' with FBM as integrator include [18], [22].

(iii) Term structure models for interest rates form perhaps the major pre-occupation of finance theorists at present. There is no consensus on the 'correct' model to use and competing alternatives abound, based mostly on the insights gained from the Black-Scholes model and its martingale generalisations. To my knowledge the only work in this field using nonstandard methods is the recent thesis by Wellmann [31], who discusses a hyperfinite version of the Heath-Jarrow-Morton model. This area seems to offer much scope for the development of nonstandard approaches.

References

1. Albeverio, S., Fenstad, J-E., Høegh-Krohn, R., Lindstrøm, T. (1986) *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.
2. Anderson, R.M., (1978) A nonstandard representation for Brownian Motion and Itô Integration, *Israel Math. J.* **25**, pp. 15-46.
3. Anderson, R.M., Rashid, S. (1978) A nonstandard characterization of weak convergence, *Proc. Amer. Math. Soc.* **69**, pp. 327-332.
4. Bick, A., Willinger, W. (1994) Dynamic spanning without probabilities, *Stoch. Proc. Appl.* **50**, pp. 349-374.
5. Bielecki, T.R., (1994) On integration with respect to fractional Brownian motion, to appear in *Statistics and Probability Letters.*
6. Black, F., Scholes, M. (1973) The pricing of options and corporate liabilities, *J. Polit. Econom.* **81**, pp. 637-654.
7. Cox, J., Ross, S., Rubinstein, M. (1979) Option pricing: a simplified approach, *J. Financial Econom.* **7**, pp. 229-263.
8. Cox, J., Rubinstein, M., (1985) *Options Markets.* Prentice-Hall, Englewood Cliffs, NJ.
9. Cutland, N.J., Kopp, P.E., Willinger, W. (1991) A nonstandard approach to option pricing, *Math. Finance* **1**(4), pp. 1-38.
10. Cutland, N.J., Kopp, P.E., Willinger, W., (1993) From discrete to continuous financial models: new convergence results for option pricing, *Math. Finance* **3**(2), pp. 101-123.
11. Cutland, N.J., Kopp, P.E., Willinger, W., (1993) A nonstandard treatment of options driven by Poisson processes, *Stochastics and Stoch. Reports* **42**, pp. 115-133.
12. Cutland, N.J., Kopp, P.E., Willinger, W., (1995) From discrete to continuous stochastic calculus, *Stochastics and Stoch. Reports* **52**, pp. 173-192.
13. Cutland, N.J., Kopp, P.E., Willinger, W., Stock price returns and the Joseph effect: a fractional version of the Black-Scholes model, in *Progress in Probability* **36**. Birkhaeuser, Basel.
14. Cutland, N.J., Kopp, P.E., Willinger, W., Wyman, M.C. Convergence of Snell envelopes and critical prices in the American put, to appear in *Mathematics of Derivative Securities*, eds. M.H.A. Dempster and S.R. Pliska. CUP, Cambridge.
15. Duffie, D., (1988) *Security Markets: Stochastic models.* Academic Press, Boston.
16. Duffie, D., (1992) *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, NJ.
17. Duffie, D., Protter, P. (1989) From discrete to continuous finance: weak convergence of the financial gain process, *Technical Report* #89/02, Department of Statistics, Purdue University.
18. Gripenberg, G., Norros, I., (1994) On the prediction of fractional Brownian motion, preprint, University of Helsinki, 11pp.
19. Harrison, J.M., Pliska, S.R. (1981) Martingales, stochastic integrals and continuous trading, *Stoch. Proc. Appl.*, **11**, pp. 215-260.
20. He, H. (1990) Convergence from discrete- to continuous-time contingent claims prices, *Rev. Fin. Stud.*, **3**, pp. 523-546.
21. Kopp, P.E. (1984) *Martingales and Stochastic Integrals.* CUP, Cambridge.
22. Lamberton, D., Lapeyre, B. (1996) *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall, London.
23. Lin, S.J. (1995) Stochastic analysis of fractional Brownian motion, *Stochastics and Stoch. Reports* **55**, pp. 121–140.
24. Lindstrøm, T., (1980), Hyperfinite stochastic integration I-III, *Math. Scand.* **46**, pp. 265-333.
25. Lindstrøm, T. (1996) Internal martingales and stochastic integration, *this volume.*
26. Loeb, P.A. (1979) Weak limits of measures and the standard part map, *Proc. Amer.*

Math. Soc. **77**, pp. 128-135.

27. Merton, R.C. (1973) Theory of rational option pricing, *Bell J. Econ. Man. Sci.* **4**, pp. 141-183.
28. Myneni, R. (1992) The pricing of the American option. *Ann. Appl. Prob.* **2**, pp. 1-23.
29. Ross, D. (1996) Loeb measure and probability, *this volume.*
30. Taqqu, M.S., Willinger, W. (1987) The analysis of finite security markets using martingales, *Adv. Appl. Prob.* **18**, pp. 1-25.
31. Wellmann, V. (1996) Stochastic models for the term structure of interest rates, *MSc thesis*, Hull University.

APPLICATIONS OF NSA TO MATHEMATICAL PHYSICS

LEIF ARKERYD
Department of Mathematics
Chalmers University of Technology
and Göteborg University
S-412 96 Göteborg
Sweden
email: arkeryd@math.chalmers.se

I will here present three applications of nonstandard techniques to the analysis of models in mathematical physics. The first provides an answer to a longstanding (standard) question about the stability of a functional equation arising in kinetic theory. It has relevance for a wide range of problems in the theoretical and numerical studies of rarefied gases.

The second application considers questions of time asymptotics for equations of Boltzmann type. In this area a number of important (standard) convergence results were first obtained by nonstandard techniques.

The third example discusses some properties of the Schrödinger equation discovered by J. Harthong and valid for wave packets, when Planck constant is infinitesimal. This is here applied to (standard) semiclassical limits of the Schrödinger equation in an approach that avoids the Wigner transform and with possible applications to open problems in the area.

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 309–339.

1. A kinetic inequality

This example discusses a kinetic inequality related to the Cauchy equation and with all solutions close to Maxwellians.

Consider a kinetic equation of the type

$$\frac{\partial}{\partial t}f + v \cdot \nabla_x f + E \cdot \nabla_v f = Q(f)$$

with f a nonnegative density, or equivalently,

$$\frac{d}{dt}f^{\#} = Q(f)^{\#},$$

where $\#$ denotes evaluation along the characteristics. For simplicity we work in a 3D space domain, $x \in \mathbb{R}^3$, with velocities $v \in \mathbb{R}^3$ and with t the time variable. This equation models streaming (transport) of f driven by exterior forces (E) and collisions (Q). When there is mass conservation, formally $\int Q(f)dxdv = 0$. If an entropy function such as $\int f \log f dxdv$ makes sense (and E const.), then formally

$$\frac{d}{dt}\int f \log f dxdv = \int Q(f) \log f := D(f).$$

If the entropy dissipation term $D(f)$ is non-positive, then the entropy is decreasing. If, moreover, the mass and energy of the system are bounded in time, then the entropy has a lower bound and so $\int_0^\infty D(f)dt < \infty$. That estimate is fundamental in most proofs of convergence to equilibrium for such equations. In a number of cases the convergence results were pioneered by NSA approaches. In e.g. the Boltzmann equation case a consequence of this entropy dissipation bound is that the factor $f_1'f_2' - f_1f_2$ in the D-integrand in a suitable sense converges to zero when $t \to \infty$. One is thus lead to the infinitesimal relation

$$f_1'f_2' - f_1f_2 \approx 0 \quad \text{Loeb a.e. in } ns{}^*\text{domain.} \tag{IR}$$

Here $1, 2$ indicate two precollisional velocity variables (v_1, v_2), and the prime indicates the corresponding postcollisional velocity (v_1', v_2') in a binary collision.

What does the relation (IR) per se imply about f? I first studied that question already in the 1980's by a combination of geometric and Loeb measure techniques in order to better understand the time asymptotics of the Boltzmann equation. The problem got a renewed actuality by a question from C. Cercignani earlier this year "What can strictly be proved (in the standard context) of the type: $f_1'f_2' - f_1f_2$ small, implies f close to a

Maxwellian". Here f belongs to a family of (say L^1_{loc}) functions, not necessarily having the extra structure of being a 1-parameter family $(f_t)_{t\in\mathbb{R}_+}$ representing the solution of some Cauchy problem for a Boltzmann equation with $\int_0^\infty D(f_t)dt < \infty$.

Starting from the above infinitesimal problem, I shall here give one type of answer to Cercignani's question and base the presentation on a recent approach due to [1].

Theorem 1.1 **([1])** *Suppose* $f : {}^*\mathbb{R}^3 \to {}^*\mathbb{R}_+$ *is *Lebesgue measurable, S-integrable in* $\{|v| \leq \lambda\}$ *for* $\lambda \in ns^*\mathbb{R}_+$*, and satisfying (IR) in* $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$*. Then for* $n \in \mathbb{N}$*, there exist internal functions* $h_n : {}^*\mathbb{R}^3 \to {}^*\mathbb{R}_+$ *such that* $f - h_n \approx 0$ *Loeb a.e.* $ns^*\mathbb{R}^3$*,* h_n *n times *differentiable in* $ns^*\mathbb{R}^3$ *with S-continuous derivatives and* h_n *satisfying (IR). Here* $f_j = f(v_j)$*,* $f'_j = f(v'_j)$*,* $j = 1, 2, v'_1 = v_1 - (\omega, v_2 - v_1)\omega$*,* $v'_2 = v_2 + (\omega, v_2 - v_1)\omega$*,* $\omega \in {}^*S^2$*, the unit sphere in* ${}^*\mathbb{R}^3$*.*

Corollary 1.2 *Under the hypotheses of Theorem 1.1, there is a standard function* $g \in C^\infty$*, such that* $f -^* g \approx 0$ *Loeb a.e.* $ns^*\mathbb{R}^3, g'_1 \cdot g'_2 - g_1 \cdot g_2 \equiv 0$ *in* $\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$*.*

Proof of Corollary 1.2 If $g, \tilde{g}$ are standard continuous and ${}^*g - {}^*\tilde{g} \approx 0$ Loeb a.e. $ns^*\mathbb{R}^3$, then $g \equiv \tilde{g}$ in $\mathbb{R}^3$. Define $g_n(x) = {}^0h_n(x)$ for $x \in \mathbb{R}^3$. Then ${}^*g_n - h_n \approx 0$ in $ns^*\mathbb{R}^3$, g_n is n times differentiable and $g_n = g_1 := g,\ n \in \mathbb{N}$.

Clearly g satisfies ${}^*g'_1\,{}^*g'_2 - {}^*g_1^*g_2 \approx 0$ Loeb a.e $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$. Since g is standard, this implies

$$g'_1 g'_2 - g_1 g_2 \equiv 0. \tag{FE}$$

□

For the proof of Theorem 1.1 we notice that if $\int f^* dv \approx 0$, then the theorem holds with $h \equiv 0$. Otherwise the proof will depend on the following lemma

Lemma 1.3 **([4], [5])** *Suppose* $b(|v|, |w{\cdot}v|) \in C^\infty(\mathbb{R}^3{\times}S^2), g \in L^1(\mathbb{R}^3_v), f \in L^2(\mathbb{R}^3_v)$ *(or* f, g *conversely),* b *vanishes for* v *near 0 and for* v *large, uniformly in* w*, as well as for* $|w \cdot v|$ *near 0 and near* $|v|$ $(w \in S^2)$*. Set*

$$Q^+(f,g)(v_1) = \int_{\mathbb{R}^3} dv_2 \int_{S^2} dw b(|v_1 - v_2|, |w \cdot (v_1 - v_2)|) f(v'_1) g(v'_2), v_1 \in \mathbb{R}^3.$$

Then

$$\|Q^+(f,g)\|_{H^1} \leq C\|f\|_{L^2}\|g\|_{L^1} \quad \textit{(or } f, g \textit{ conversely)}$$

for some C *independent of* f, g*.*

Lemma 1.4 **([2], [4])** *Solutions* $0 \leq f \in L^1_{\text{loc}}(\mathbb{R}^3)$ *of the functional equation (FE) are smooth.*

Proof We give the short proof of [4]. Clearly $g := \sqrt{f} \in L^2_{\text{loc}}(\mathbb{R}^3)$ satisfies (FE). There is nothing to prove if $g \equiv 0$. Otherwise, introduce

$$b_\epsilon(v_1, v_2, \omega) = \varphi_\epsilon^{(1)}(v_1^2 + v_2^2)\varphi_\epsilon^2(|v_1 - v_2| - |\omega \cdot (v_1 - v_2)|) \cdot \varphi_\epsilon^{(3)}(|\omega \cdot |v_1 - v_2|),$$

where $0 \leq \varphi^{(j)}$, $j = 1, 2, 3$,

$$\varphi_\epsilon^{(1)} \in C_0^\infty(\mathbb{R}),\ \varphi_\epsilon^{(1)}(t) \equiv 1 \quad \text{for } 0 \leq t \leq \epsilon^{-1}, \quad \text{and}$$
$$\varphi_\epsilon^{(2,3)} \in C^\infty(\mathbb{R}),\ \varphi_\epsilon^{(2,3)}(t) = 0 \quad \text{for } t \leq \frac{\epsilon}{2}, \varphi_\epsilon^{(2,3)}(t) = 1 \quad \text{for } t \geq \epsilon.$$

Given b_ϵ, define Q_ϵ^+ as in Lemma 1.3, and set

$$\ell_\epsilon(v_1) = \int_{\mathbb{R}^3 \times S^2} b_\epsilon(v_1, v_2, \omega) g(v_2) dv_2 d\omega \in C_0^\infty(\mathbb{R}^3).$$

For given $C > 0$, we have $\ell_\epsilon(v_1) > 0$, when $|v_1| \leq C$ and ϵ is small enough. From $g_1' g_2' \equiv g_1 g_2$ it follows

$$g(v_1)\ell_\epsilon(v_1) = Q_\epsilon^+(g, g)(v_1).$$

So using Lemma 1.3, $g \in H^1_{\text{loc}}(\mathbb{R}^3)$. The proof of Lemma 1.3 actually implies that if $g \in L^1(\mathbb{R}^3)$, $f \in H^s(\mathbb{R}^3)$, then $\|Q^+(f, g)\|_{H^{s+1}} \leq C\|f\|_{H^s}\|g\|_{L^1}$. We conclude that in our case $g \in H^k_{\text{loc}}(\mathbb{R}^3)$, $k \in \mathbb{N}$, hence that $f, g \in C^\infty(\mathbb{R}^3)$. □

Corollary 1.5 *Solutions $0 \leq f \in L^1_{\text{loc}}(\mathbb{R}^3)$ of (FE) are Maxwellians, $f \equiv a \exp(b(v - c)^2)$ for some $a \in \mathbb{R}_+, b \in \mathbb{R}, c \in \mathbb{R}^3$.*

Proof This is a well known result for smooth f's. The proof is there reduced to solving the Cauchy equation

$$\varphi(x) + \varphi(y) = \varphi(x + y), \quad x, y \in \mathbb{R}, \tag{CE}$$

for which it is easy to see that any continuous solution φ is of the type $\varphi =$ constant $\cdot x$. □

Proof of Theorem 1.1 We consider the case ${}^0\!\int f^* dv > 0$. Set $g = \sqrt{f}$. Similarly to the proof of Lemma 1.4

$$g(v) \approx Q_\epsilon^+(g, g)(v)/\ell_\epsilon(v).$$

(Here ϵ is chosen depending on the set $|v| \leq \lambda$, so that $\inf_{|v| \leq \lambda} {}^0\ell_\epsilon(v) > 0$, which is possible since f is S-integrable with ${}^0\!\int f^* dv > 0$.)

The first derivatives of the right-hand side in the above relation are in ${}^*L^2(|v| \leq \lambda)$. Since f is finite Loeb a.e. $ns^*\mathbb{R}^3$, by overspill there is a function $q^1 \in {}^*L^1(|v| \leq \lambda)$ with its first derivatives having finite norms in ${}^*L^1(|v| \leq \lambda)$ for λ finite, $q^1 \approx f$ and (IR) holding for q^1, Loeb a.e. $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$. By iteration, for $n \in \mathbb{N}$ there is a function q^n with all derivatives of order $\leq n$ finite in ${}^*L^1(|v| \leq \lambda)$, when λ is finite, $q^n \approx f$ and (IR) holding for q^n, Loeb a.e. $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$.

It follows that a subsequence $(h^n)_{n\in\mathbb{N}}$ of (q^n) has all derivatives up to order n S-continuous in $ns^*\mathbb{R}$. □

It follows from Corollary 1.2 and Corollary 1.5 that

Theorem 1.6 ([3]) *Suppose $f : {}^*\mathbb{R}^3 \to {}^*\mathbb{R}_+$ is *Lebesgue measurable, is S-integrable on $\{|v| \leq \lambda\}$ for λ finite, satisfies (IR) Loeb a.e. $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$. Then there is a standard Maxwellian M_f such that*

$$f - {}^*M_f \approx 0 \quad \textit{Loeb a.e. } ns^*\mathbb{R}^3.$$

One implication of Theorem 1.6 in the standard context, is the following result.

Theorem 1.7 *Given $C > 0$, consider the set of non-negative functions with $\int f(1 + |\log f|)dv \leq C$. Set*

$$S_\delta = \{(v_1, v_2, \omega) \in \mathbb{R}^3 \times \mathbb{R}^3 \times S^2; |v_1| \leq \delta^{-1}, |v_2| \leq \delta^{-1}\}.$$

Given $\epsilon > 0$, there is $\delta > 0$ such that if $|f_1'f_2' - f_1f_2| < \delta$ in S_δ outside of some (f-dependent) subset of measure bounded by δ, then there is a Maxwellian M_f with

$$|f - M_f| < \epsilon \quad \textit{for } |v| \leq \epsilon^{-1},$$

outside of some (f-dependent) subset of measure bounded by ϵ.

Remark that the condition $f \log f \in L^1(\mathbb{R}^3)$ can be replaced by L^1-conditions involving weaker, strictly convex functions of f than $f \log f$. In fact the theorem and following consequences hold for any locally weakly precompact set of positive L^1 functions.

Proof Consider the set of *Lebesgue measurable functions f with c$\int f(1 + |\log f|)^*dv \leq C$. Such function are S-integrable on $\{|v| \leq \lambda\}$ for λ finite. Let F_δ be the set of such functions with $|f_1'f_2' - f_1f_2| < \delta$ on S_δ outside a subset of measure $\leq \delta$. If this holds for $\delta \approx 0$, then (IR) holds Loeb a.e. $ns^*\mathbb{R}^3 \times \mathbb{R}^3 \times S^2$. (A converse is also true.) It follows by Theorem 1.6 that there is a standard Maxwellian M_f such that $f - {}^*M_f \approx 0$, Loeb a.e. $ns^*\mathbb{R}^3$. In particular given $\epsilon > 0$ and standard, $|f - {}^*M_f| < \epsilon$ for $|v| \leq \epsilon^{-1}$ outside of a (f dependent) set of *Lebesgue measure bounded by ϵ. For each infinitesimal δ this holds for all $f \in F_\delta$. But the set of δ for which the ϵ-property holds for all $f \in F_\delta$, is internal. Hence there is also a standard $\delta > 0$ such that it holds for all $f \in F_\delta$. For ϵ, δ standard, by transfer the statement holds in the standard context. □

Corollary 1.8 *Given $C > 0$, consider the set of non-negative functions f with $\int f(1+v^2+|\log f|)dv \le C$. Given $\epsilon > 0$ there is $\delta > 0$ (only depending on C, ϵ) such that if $|f_1'f_2' - f_1f_2| < \delta$ in S_δ outside of some subset of measure bounded by δ, then for some Maxwellian M_f (depending on f) $\int |f - M_f|dv < \epsilon$.*

Remark that the corollary holds for any weakly precompact set of positive L^1-functions.

Proof In the class of non-negative functions f with

$$\int f(1+v^2+|\log f|)dv < C(<\infty),$$

evidently

$$\int_{|v|\ge\lambda} f dv \le \frac{C}{1+\lambda^2} < \epsilon \quad \text{for} \quad 1+\lambda^2 > \frac{C}{\epsilon}.$$

There is $(\epsilon >>)\epsilon_1 > 0$ (and depending on C) such that for any f in the class and any set S of measure bounded by ϵ_1,

$$\int_S f dv < \epsilon 10^{-3} \quad \text{(equi-integrability)}.$$

Given λ take ϵ_1 so that, moreover, $\int_{|v|\le\lambda} \epsilon_1 dv < \epsilon$.

Recall that by Theorem 1.7, there is $\delta > 0$ such that the following holds. If in S_δ outside of some subset of measure bounded by δ

$$|f_1'f_2' - f_1f_2| < \delta,$$

then there is a Maxwellian M_f such that outside of some set S_1 of measure bounded by ϵ_1,

$$|f - M_f| < \epsilon_1 10^{-3} \quad \text{for} \quad |v| \le \epsilon_1^{-1},$$

in particular for $|v| \le \lambda$.

If $\int f dv < \epsilon$, then the corollary holds with $M_f \equiv 0$. Otherwise, take λ so that

$$\int_{|v|>\lambda 10^{-3}} f dv < \epsilon 10^{-3}.$$

If M_f attains its maximum for $|v| \ge \frac{\lambda}{2} - 1$, then

$$\begin{aligned} \int f dv &\le \int_{A_1} f dv + \int_{A_2} |f - M_f| dv + \int_{A_2} M_f dv + \int_{S_1} f dv \le \\ &\le 3\epsilon 10^{-3} + \int_{A_2} M_f dv. \end{aligned}$$

There

$$\begin{aligned} A_1 &= \{v; |v| \geq \lambda 10^{-3}\}, \\ A_2 &= \{v; |v| \leq \lambda 10^{-3}, v \notin S_1\}. \end{aligned}$$

But in this case

$$\int_{A_2} M_f dv \leq \int_{A_3} M_f dv \leq \int_{A_3} |M_f - f| dv + \int_{A_3} f dv \leq 2\epsilon 10^{-3}$$

with $A_3 = \{v; \lambda \geq |v| \geq \lambda 10^{-3},\ v \notin S_1\}$, and so

$$\int f dv \leq 5\epsilon 10^{-3}.$$

That contradicts the present assumption $\int f dv > \epsilon$, and so M_f attains its maximum for $|v| \leq \frac{\lambda}{2} - 1$.

With $A_4 = \{v; \lambda \geq |v| \geq \frac{\lambda}{2},\ v \notin S_1\}$

$$\int_{A_4} M_f dv \leq \int_{A_4} |M_f - f| dv + \int_{A_4} f dv < 2\epsilon 10^{-3}.$$

Hence

$$\int_{|v| \geq \lambda/2} M_f dv \leq 4\epsilon 10^{-3}.$$

If $\int_{|v| \leq \lambda/2} M_f dv \leq \frac{\epsilon}{2}$, then

$$\int f dv \leq \int_{|v| \geq \lambda/2} f dv + \int_{A_5} |f - M_f| dv + \int_{A_5} M_f dv + \int_{S_1} f dv < \epsilon.$$

Here

$$A_5 = \{v; |v| \leq \frac{\lambda}{2},\ v \notin S_1\}.$$

Since $\int f dv \geq \epsilon$, we conclude that $\int_{|v| \leq \lambda/2} M_f dv > \frac{\epsilon}{2}$.

Consider now the case when S_1 is a sphere with centre at the maximum of M_f. If $\int_{S_1} M_f dv \leq \frac{\epsilon}{2}$, then the corollary holds. If $\int_{S_1} M_f dv > \frac{\epsilon}{2}$, set S_2 as the sphere concentric with S_1 and with ten times its radius. If

$$\int_{S_2 \setminus S_1} M_f dv > \epsilon 10^{-2}$$

then

$$\int_{S_2 \setminus S_1} f dv \geq \int_{S_2 \setminus S_1} M_f dv - \int_{S_2 \setminus S_1} |f - M_f| dv > \epsilon 10^{-3}.$$

This contradiction implies that $\int_{R^3 \setminus S_1} M_f dv \leq \epsilon 10^{-1}$, hence that

$$\int f dv \quad \leq \&int_{|v|\geq\lambda} f dv + \int_{|v|\leq\lambda} f dv \leq \epsilon 10^{-3} + \int_{A_6} |f - M_f| dv + \int_{A_6} M_f dv$$
$$+ \int_{S_1} f dv \leq 2\epsilon 10^{-1}, \quad A_6 = \{v; |v| \leq \lambda,\ v \notin S_1\},$$

which again contradicts our assumption.

Hence the corollary holds provided S_1 is a sphere with centre at the maximum of M_f. Finally if all or part of the bad set S_1 lies outside of the above sphere, then the previous argument still holds with minor changes. □

There are corresponding results in the space-dependent case.

Theorem 1.9 *Given $C > 0$ and $\Omega \subset \mathbb{R}^3$ measurable. Consider the set of nonnegative functions f with $\int_{\Omega\times\mathbb{R}^3} f(1 + |\log f|) dx dv \leq C$. Given $\epsilon > 0$, there is $\delta > 0$ such that if for $x \in \Omega$, outside of some (f dependent) subset $S(f, \epsilon)$ of measure bounded by $\frac{\epsilon}{2}$, it holds that $|f_1' f_2' - f_1 f_2| < \delta$ in S_δ outside some (x, f dependent) subset of measure bounded by δ, then there is a (local) Maxwellian M_f such that outside of a subset $S(f, \epsilon)$ of measure bounded by ϵ in Ω, $|f - M_f| < \epsilon$ for $|v| \leq \epsilon^{-1}$ outside of a v-subset of measure bounded by ϵ.*

Proof This is similar to the proof of Theorem 1.7. □

Corollary 1.10 *Given $C > 0$ and a bounded measurable set Ω in x-space. Consider the set of non-negative functions f with $\int f(1+v^2+|\log f|) dx dv \leq C$. Given $\epsilon > 0$, there is $\epsilon_1 > 0$ and $\delta > 0$, such that if for all x in Ω outside some f-dependent subset of measure $< \epsilon_1$, $|f_1' f_2' - f_1 f_2| < \delta$ in S_δ outside of some (x, f-dependent) subset of measure $< \delta$, then for some local Maxwellian M_f (depending on f)*

$$\int |f - M_f| dx dv < \epsilon.$$

Remark that the Corollary also holds for unbounded measurable sets in x-space when $\int f(1 + v^2 + x^2 + |\log f|) dx dv < C$.

Problem Find standard proofs of Theorem 1.7-10.

Problem Can similar NSA ideas be used in the study of stability questions for classical functional equations?

References

1. Andreasson, H. (1996) A regularity property and strong L^1-convergence to equilibrium for the relativistic Boltzmann equation, *SIAM Journ. Anal.*, to appear.

2. Arkeryd, L. (1972) On the Boltzmann equation, *Arch. Rat. Mechs. Anal.* **45**, pp. 1-34.
3. Arkeryd, L. (1986) On the Boltzmann equation in unbounded space far from equilibrium, *CMP* **105**, pp 205-219.
4. Lions, P.L. (1994) Compactness in Boltzmann's equation via Fourier integral operators and applications I, *Journ. Math. Kyoto* **36**, pp. 391-427.
5. Wennberg, B. (1994) Regularity in the Boltzmann equation, *Comm. PDE* **19**, pp. 2057-2074.

2. The time asymptotic behaviour for certain rarefied gases when the incoming fluxes at the boundary are given

Let Ω be an open, bounded, convex domain in $\mathbb{R}^3$ with C^1 smooth boundary $\partial\Omega$, x the position variable, t respectively v the time and velocity variables, and $f(t,x,v) \geq 0$ the distribution function of a solute gas satisfying the linear Boltzmann equation

$$\partial_t f + v.\nabla_x f = Q(f),\ x \in \Omega, v \in \mathbb{R}^3,\ t > 0, \tag{A1}$$

where

$$\begin{aligned}
&Q(f)(t,x,v) = \int_{\mathbb{R}^3} dv_* \int_0^{\pi/2} d\theta \int_0^{2\pi} d\epsilon B(\theta,\omega)(f'F'_* - fF_*),\\
&f' = f(t,x,v'), F'_* = F(t,x,v'_*), f = f(t,x,v), F_* = F(t,x,v_*),\\
&\omega = |v - v_*|, v' = v - \frac{2}{1+\kappa}((v-v_*)\cdot e)e,\ e \in S^2,\\
&v'_* = v_* + \frac{2\kappa}{1+\kappa}((v-v_*)\cdot e)e.
\end{aligned}$$

Here $0 < F(t,x,v)$ is the given distribution function of a (solvent) gas, with $(1+|v|)^\gamma F \in C([0,\infty); L^1(\Omega \times \mathbb{R}^3))$, κ is the ratio between the molecular masses, m of the solute and m_* of the solvent, S^2 is the unit sphere in $\mathbb{R}^3$ and $du = d\theta d\epsilon$ the angular measure on S^2. We assume

$$B(\theta,\omega) = \omega^\gamma b(\theta),\ 0 \leq \gamma \leq 1,\ 0 < b \in L^1.$$

The equation (A1) is complemented with an initial condition $f(0,x,v) = f_i(x,v)$,
$x \in \Omega$, $v \in \mathbb{R}^3$, and given indata on $\partial\Omega$,

$$f(t,x,v) = f_b(x,v),\ x \in \partial\Omega, v\cdot n(x) > 0,\ t > 0.$$

Here $n(x)$ is the inward normal at x. Let $(\partial\Omega\times\mathbb{R}^3)^+$ respectively $(\partial\Omega\times R^3)^-$ denote the sets of $(x,v) \in \partial\Omega \times \mathbb{R}^3$ such that $v\cdot n(x) > 0$ and $v \cdot n(x) < 0$.

The existence and uniqueness approach of e.g. [1], can be used to prove the following theorem

Theorem 2.1 *If* $(1+|v|)\,f_i$ *and* $(1+|v|^2)f_b$ *belong to* $L^1(\Omega\times\mathbb{R}^3)$ *and* $L^1_{v\cdot n(x)}((\partial\Omega\times R^3)^+)$ *respectively, then there exists a unique* L^1*-solution* f *of (A1) with initial and ingoing boundary data* f_i *respectively* f_b*. Moreover,* f *is non-negative when* f_i *and* f_b *are non-negative.*

Include in B a factor χ with $\chi = 0$ for $|v| \leq \delta$, and for $|v'| \leq \delta$, and $\chi = 1$ otherwise. For $E = E_0 \exp(-\frac{am}{2}v^2 + b\cdot(mv))$, suppose that the integrals

$$\int |f_i|\log\frac{|f_i|}{E} + E - |f_i|dxdv,$$

$$\int_{n(x)\cdot v>0} n(x)\cdot v(|f_b|\log\frac{|f_b|}{E} + E - |f_b|)dvd\sigma$$

converge (with $d\sigma$ the usual Lebesgue surface measure on $\partial\Omega$) and that $F = F(v)$ is independent of x, t. By linearity it is enough to consider the case $f_i, f_b \geq 0$.

Theorem 2.2 *Under the above hypotheses, when* $t \to \infty$ *the solution* $f(t)$ *converges strongly in* $L^1_{\mathrm{loc}}(\Omega\times\mathbb{R}^3)$ *to the unique solution* g *of the linear stationary Boltzmann equation*

$$v\cdot\nabla_x g = Q(g) \tag{A2}$$

with the boundary condition

$$g(x,v) = f_b(x,v), (x,v)\in(\partial\Omega\times\mathbb{R}^3)^+. \tag{A3}$$

Proof Let us first consider the stationary problem (A2-3). In exponential form it becomes

$$f(x,v) = f_b(x-sv,v)\exp(-s\int BF_*dv_*du) +$$

$$+\int_{-s}^0 d\tau\int dv_*duBF'_*f'(x+\tau v)\exp(\tau\int BF_*dv_*du) := Lf.$$

For $(x-sv,v)\in(\partial\Omega\times\mathbb{R}^3)^+$ consider the iteration scheme $f_0 \equiv 0$, $f_j = Lf_{j-1}$ for $j \geq 1$. Green's formula gives

$$\int_{(\partial\Omega\times\mathbb{R}^3)^-} |n(x)\cdot v|f_j d\sigma dv \leq \int_{(\partial\Omega\times\mathbb{R}^3)^+} n(x)\cdot vf_b d\sigma dv,$$

where the inequality is due to the term

$$\int B(F_*f_{j-1} - F_*f_j)dxdvdv_*du \leq 0.$$

The sequence f_j is increasing and the limit f_∞ satisfies

$$\int_{(\partial\Omega\times\mathbb{R}^3)^-} |n(x)\cdot v| f_\infty d\sigma dv \leq \int_{(\partial\Omega\times\mathbb{R}^3)^+} n(x)\cdot v f_b d\sigma dv.$$

From the exponential form with $(x,v)\in(\partial\Omega\times\mathbb{R}^3)^-$, $(x-sv,v)\in\Omega\times\mathbb{R}^3$, it then follows that the limit f_∞ satisfies the problem (A2-3). One may also prove that the mass is bounded.

Let us next discuss the time dependent problem. Due to the linearity of (A1) f can be split into the sum of the solution with initial value $\tilde{f}_i = f_i - f_\infty$ and zero boundary values, and the stationary solution. Again by linearity it is enough to consider nonnegative initial values $\tilde{f}_i$. It only remains to prove that the first part with initial value $\tilde{f}_i$ tends to zero in $L^1(\Omega\times\mathbb{R}^3)$ when $t\to\infty$. In this case integration of (A1) gives

$$\int f(t,x,v)dxdv + \int_0^t \int_{(\partial\Omega\times\mathbb{R}^3)^-} |v\cdot n(x)| f(t) dt d\sigma dv = \int \tilde{f}_i(x,v)dxdv.$$

It follows that $\int f(t,x,v)dxdv$ decreases with time. The hypotheses of the theorem imply that the entropy remains uniformly bounded for $t\geq 0$, and so for $t>0$, $\int_{|v|\leq\eta} f(t,x,v)dxdv$ remains uniformly small provided η is small.

$$f(t,x+(t-s)v,v)|_{(\partial\Omega\times\mathbb{R}^3)^-} \geq f(s,x,v)\exp(-(t-s)\int F_* B dv_* du). \quad \text{(A4)}$$

If (s,x,v) is an interior point and $(t,x+(t-s)v,v)$ is an outgoing boundary point, then

$$\frac{\text{diam }\Omega}{|v|} \geq t-s \quad \text{and} \quad \int_s^t \int BF_* dv_* du d\tau \leq$$
$$\leq \frac{\text{const.}}{|v|}(1+|v|)^\gamma \leq \text{const. for } |v|\geq\eta>0.$$

This together with (A4) and the entropy control imply that

$$\lim_{t\to\infty}\int f(t,x,v)dxdv = 0.$$

□

Remark that the original proof at the Edinburgh meeting used non-standard arguments. Discussion with participants, in particular P. Loeb, A. Nouri and M. Wolff, lead to successive simplifications. The end-product was the above purely standard proof. It should be stressed that one important usage of NSA – seldom noticed in published papers – is for producing a first understanding in an unfamiliar situation. In such cases the power of NSA is a.o. due to its reduction of complexity and its easy separation of scales.

The generalization to the case $F = F(t, x, v)$ with $\lim_{t\to\infty} F(t, \cdot) = F_\infty(\cdot)$ will next be discussed in the particular case

i) $B = 0$ when $|v| \leq \eta$ or $|v'| \leq \eta$,
ii) $\sup_x \int F_\infty(1 + |v|)^\gamma dv < \infty$,
iii) $\lim_{t\to\infty} \int |F(t, x, v) - F_\infty(x, v)|(1 + |v|)^\gamma dv = 0$, uniformly in x.

In this case the previous result still holds.

Theorem 2.2′ *When $t \to \infty$ the solution $f(t)$ of (A1), (A3) converges strongly in $L^1(\Omega \times \mathbb{R}^3)$ to the unique solution g of the stationary problem (A2-3).*

Remark that here a nonstandard approach is transparent and easy (even if a standard proof along the same lines is possible).

Proof Take $f_i, f_b \geq 0$. The existence of the stationary solution f_∞ follows as in the previous proof. Consider (A4) for the full problem under the present hypotheses. It can be used to prove that the mass $\int f(t, x, v) dx dv$ remains uniformly bounded in time. Pick in the nonstandard context an unlimited time $t_\infty \in {}^*\mathbb{R}_+^\infty$. Split the problem starting at $t = t_\infty$ into one with initial value ${}^*f(t_\infty) - {}^*f_\infty$ and ingoing boundary value zero, and another with initial value ${}^*f_\infty$ and ingoing boundary value ${}^*f_b$. By the argument after (A4) in the preceding theorem, the solution of the present first problem decreases exponentially with time with the rate of decrease independent of t_∞. By Green's formula the solution $\tilde{f}$ of the second problem satisfies.

$$\int |\tilde{f} - {}^*f_\infty| dx dv|_t + \int_{t_\infty}^t \int_{(\partial\Omega\times\mathbb{R}^3)^-} |n(x) \cdot v||\tilde{f} - {}^*f_\infty| d\sigma dv$$
$$\leq 2 \int_{t_\infty}^t ds \int dx dv dv_* du B|{}^*F_*(s) - {}^*F_{\infty *}|\, {}^*f_\infty.$$

Here by hypothesis the right-hand side is infinitesimal for $t - t_\infty$ finite, hence up to some unlimited $T = t' - t_\infty$ with T independent of t_∞. It follows that

$$\int |{}^*f(t) - {}^*f_\infty| dx dv \approx 0, \quad t \in {}^*R_+^\infty.$$

This implies in the standard context that

$$\lim_{t\to\infty} \int |f(t) - f_\infty| dx dv = 0.$$

□

Let us also discuss some difficulties with a similar approach for the non-linear Boltzmann equation

$$(\partial_t + v \cdot \nabla_x) f = Q(f, f),$$

where

$$Q(f,f)(v) = \int B(\theta,\omega)(f' f'_* - f f_*) dv_* du.$$

The notations are as in the linear case, but with the additional requirement $m = m_*$. This comes from the colliding molecules now being of the same type.

The problem with zero ingoing boundary flux and positive initial condition f_i has a non-negative solution if $0 \le f_i(1+|v|^2+|\log f_i|) \in L^1(\Omega \times \mathbb{R}^3)$. Green's identity shows that there is a mass flow out of Ω, but none into Ω, hence that the mass $\int_{\Omega\times\mathbb{R}^3} f(t,x,v)dxdv$ is decreasing with time. With a slight extension of the earlier linear proof, one obtains that strongly in L^1 the limit of $f(t)$ when $t \to \infty$ is zero. However, a splitting as before of the full problem with given influx f_b through $\partial\Omega$ and given initial value f_i is not at all an obvious approach in this non-linear case. Nevertheless, a number of results are known related to the problem

$$(\partial_t + v \cdot \nabla_x) f = Q(f,f), t > 0, (x,v) \in \Omega \times \mathbb{R}^3 \tag{A5}$$

$$f(0,x,t) = f_i, (x,v) \in \Omega \times \mathbb{R}^3 \tag{A6}$$

$$f(t,x,v) = f_b(x,v), (x,v) \in (\partial\Omega \times R^3)^+. \tag{A7}$$

From the theory of this initial boundary value problem, we notice that there is a solution satisfying the boundary condition in a weakened form (cf. [3]), if

$$0 \le f_b(1 + v^2 + |\log f_b|) \in L^1_{v\cdot n(x)}((\partial\Omega \times \mathbb{R}^3)^+).$$

Problem Does $\lim_{t\to\infty} f(t)$ exist in some meaningful sense?

The following is e.g. known. Let $f_i(1 + |v|^2 + |\log f_i|) \in L^1_+(\Omega \times \mathbb{R}^3)$.

If f_b is given by Maxwellian diffuse reflection with a normalized x-independent Maxwellian M, then $\lim_{t\to\infty} f|_t = cM$ strongly in $L^1(\Omega \times \mathbb{R}^3)$ (see [4]). Here $c \int M dv \int_\Omega dx = \int f_i dxdv$.

If the boundary condition is specular or direct reflection or periodic, then given $(t_j)_{j\in N}$, $t_j \nearrow \infty$ there is a subsequence $(t_{j'})$ such that $\lim_{t_{j'}\to\infty} f(t_{j'} + t) = cM$ in weak or strong $L^1([0,T] \times \Omega \times \mathbb{R}^3)$ for all $T > 0$ (e.g. [2, 7]]) and for some Maxwellian M (possibly different for different sequences $(t_{j'})$) with

$$\int f_i = \int cM, \int v^2 f_i \ge \int cv^2 M.$$

In our present case it is not known if $\int f(t,x,v)dxdv$ is bounded or not. If the boundedness of mass in time were known, then the time independence of f_b could be taken as a starting point for an attack on the time asymptotics. The proof of boundedness for f in the linear case, used the boundedness of $\int Fdv$. Can this be handled differently in the nonlinear case where in the exponential form the exponent in the equation is

$$\int f_* B dv_* du dt \quad \text{instead of} \quad \int F_* B dv_* du dt.$$

For possible inspiration let us end with a NSA proof of sL^1 convergence in the case of Maxwellian diffuse reflection with x-independent Maxwellian on the boundary, i.e., the boundary condition

$$f(t,x,v) = M(v) \int_{v'\cdot n(x)<0} |v'\cdot n(x)| f(t,x,v') dv', t > 0, x \in \partial\Omega, v\cdot n(x) > 0. \tag{A8}$$

M is a normalized Maxwellian, $\int_{v\cdot n(x)>0} v \cdot n(x) M(v) dv = 1$, with constant temperature $\frac{1}{\theta} > 0$,

$$M(v) = (2\pi)^{-2}\theta^2 \exp\left(-\frac{\theta v^2}{2}\right).$$

The relevant equilibrium solution is $f_s = cM$ with

$$c = \frac{1}{|\Omega|} \int_{\Omega\times\mathbb{R}^3} f_i(x,v) dx dv. \tag{A9}$$

An existence result from [3] is recalled

Theorem 2.3 *There is a function*

$$f \in C(R^+, L^1(\Omega \times \mathbb{R}^3)),\ f \geq 0$$

satisfying (A5), (A6) in DiPerna Lions sense and (A8) for the traces, possibly with inequality, the left-hand side being greater than or equal to the right-hand side.

The following à priori bounds hold [4].

Theorem 2.4 *Let f be a solution of the above problem, (A5-6), (A). Then*

$$\int_{\Omega\times\mathbb{R}^3} |v|^2 f(t,x,v) dx dv, \int_{\Omega\times\mathbb{R}^3} f \log f |(t,x,v) dx dv,$$
$$\int_{[t,t+T]\times\partial\Omega} \int_{v\cdot n(x)>0} |v|^2 v \cdot n(x) f(\tau,x,v) d\tau dx dv,$$

$$\int_{[t,t+T]\times\partial\Omega}\int_{v\cdot n(x)<0}|v|^2|v\cdot n(x)|f(\tau,x,v)d\tau dxdv,$$
$$\int_{[t,t+T]\times\partial\Omega}\int_{v\cdot n(x)>0}v\cdot n(x)f(\tau,x,v)d\tau dxdv,$$
$$\int_{[t,t+T]\times\partial\Omega}\int_{v\cdot n(x)<0}|v\cdot n(x)|f(\tau,x,v)d\tau dxdv$$

are uniformly bounded for $t\in\mathbb{R}_+$.

We recall the beginning of the proof, which will be used later. Formally

$$(\partial_t+v\nabla_x)f\log\frac{f}{M}=Q(f,f)\log\frac{f}{M}+Q(f,f).$$

Integrating this over $[0,t]\times\Omega\times\mathbb{R}^3$ and using the initial-boundary conditions gives

$$\int_{\Omega\times\mathbb{R}^3}(f\log\frac{f}{M}(t,x,v)dxdv-\int_0^t\int_{\partial\Omega\times\mathbb{R}^3}v\cdot n(x)(f\log\frac{f}{M})d\tau dxdv$$
$$+\int_0^t\int_{\Omega\times\mathbb{R}^3}e(f)d\tau dxdv\leq\int_{\Omega\times\mathbb{R}^3}f_0\log\frac{f_0}{M}(x,v|dxdv$$

where

$$e(f)=\frac{1}{4}\int_{\mathbb{R}^3}\int_{B+}B(|v-v_*|,u)(f'f'_*-ff_*)\log\frac{f'f'_*}{ff_*}dv_*du.$$

This holds strictly (see [3]). Since $e(f)\geq 0$ and since by Darrozes-Guiraud's inequality (see [6]) the boundary term (including sign) is positive, it follows that $\int_{\Omega\times R^3}f\log\frac{f}{M}(t,x,v)dxdv<$ const.

$$0\leq\int_0^\infty\int_{\Omega\times\mathbb{R}^3}e(t)(t,x,v)dtdxdv<\infty. \tag{A10}$$

The rest of the proof of Theorem 4 relies on a careful use of Green's formula arguments and the observation that for $\epsilon>0$ (see [5])

$$\int_{\mathbb{R}^3}f|\log^- f|dv=\int_{f<e^{-|v|^\epsilon}}f|\log f|dv+\int_{e^{-|v|^\epsilon}<f\leq 1}f|\log f|dv<$$
$$<\int_{|v|\leq 1}dv+\int_{|v|>1}e^{-|v|^2}|v|^\epsilon dv+\int f|v|^\epsilon dv.$$

The asymptotic behaviour of a solution as in Theorem 2.3 can be derived. For simplicity we assume $B>0$.

Theorem 2.5 *[4] Let f be a solution of Theorem 2.3 for the problem (A5), (A6), (A8). When $t \to \infty$, $f(t,.)$ converges strongly in $L^1(\Omega \times \mathbb{R}^3)$ to the global Maxwellian cM with c given by (A9).*

Proof It is a consequence of the compactness implied by Theorem 2.4, that for every sequence t_k tending to infinity, there is a subsequence $(t_{k'})$ such that
$f((.+t_{k'}) \stackrel{\omega L'}{\rightharpoonup} g(\cdot)$ in $L^1([0,T]\times\Omega\times\mathbb{R}^3)$ for $T > 0$. The proof will demonstrate that the limit g equals the global Maxwellian of the theorem and discuss the strong L^1 convergence. The first part of the proof will show that g equals a time dependent local Maxwellian

$$M(t,x,v) = a(t,x)\exp(-b(t,x)(v - c(t,x))^2).$$

For this, notice that there is a countable sequence $\phi_1, \phi_2, \dots$ of functions with bounded support in $\Omega \times R^3 \times R_+$, such that $g = M$ in L^1, if

$$\int g\phi_j dxdvdt = \int M\phi_j dxdvdt, \quad j \in N.$$

Also let the sequence contain $\chi_{\nu\rho}$, $v\chi_{\nu\rho}$, $v^2\chi_{\nu\rho}$, ν, $\rho \in N$, where $\chi_{\nu\rho}(v,t) = 1$ for $v^2 \leq \nu^2$, $t \leq \rho$, $\chi_{\nu\rho}(v,t) = 0$, otherwise. Set $M_k = \{\phi_1, \dots, \phi_k\}$.

It is a consequence of the existence theory and Theorem 2.4 that g satisfies the Boltzmann equation, and that for some subsequence of $(t_{k'})$, which will from here on be denoted (t_k),

$$\int_0^k ds \int_\Omega dx |\int_{R^3} f(.+t_k)\phi - g(.)\phi dv| < 1/k, \quad \phi \in M_k. \tag{A11}$$

We also assume that (t_k) was so chosen that

$$\int_{t_k}^{t_k+k} ds \int ((f_1f_2)' - (f_1f_2))\log((f_1f_2)'/(f_1f_2))Bdxdv_1dv_2du < 1/k, \tag{A12}$$

which is possible by (A10).

From here the present proof relies on NSA. By transfer, in the non-standard context (A11) and (A12) hold for all $k \in {}^*N$. Given $\kappa \in {}^*\mathrm{N}_\infty$, (A11-12) implies in particular for $k = \kappa$ that

$$\int_{t_k}^{t_k+k} ds \int ((f_1f_2)' - (f_1f_2))\log((f_1f_2)'/(f_1f_2))Bdxdv_1dv_2du \approx 0, \tag{A13}$$

$$\int_0^k ds \int_\Omega dx |\int_{R^3} (f(t_k + .)\phi - *g\phi)dv| \approx 0. \tag{A14}$$

From (A13) it follows that the integrand in the left-hand side is infinitesimal; for Loeb a.a. $(x,t) \in {}^*\Omega \times ns^*R_+$

$$f(x,v_1,t+t_\kappa)f(x,v_2,t+t_\kappa) \approx f(x,v_1',t+t_\kappa)f(x,v_2',t+t_\kappa)$$

for Loeb a.a. $(v_1,v_2,u) \in ns^*R^3 \times R^3 \times S^2$. By Example 1, that implies

Lemma 2.6 *For Loeb a.a. $(x,t) \in {}^*\Omega \times ns^*R_+$, there are*

$$\bar{a}(x,t), \bar{b}(x,t) \in R_+, \quad \bar{c}(x,t) \in R^3,$$

such that

$$f(x,v,t+t_\kappa) \approx \bar{M}(x,v,t) \quad \textit{for Loeb a.a. } v \in ns^*R^3,$$

with

$${}^*M = \bar{M}(x,v,t) = \bar{a}(x,t)\exp(-\bar{b}(x,t)(v-\bar{c}(x,t))^2).$$

This result together with (A14) gives for $\phi \in \cup_{k\in N} M_k (\subset M_\kappa)$ and $T \in R_+$

$$\begin{aligned} 0 &= {}^0\!\int_{{}^*\Omega\times[0,T]} {}^*dxdt| \int_{*R^3} f(x,v,t+t_\kappa) * \phi * dv -{}^* \int_{R^3} g\phi dv| = \\ &= \int_{{}^*\Omega\times[0,T]} Ldxdt| \int_{ns^*R^3} {}^0 f(x,v,t+t_\kappa)^{0*}\phi L dv - {}^{0*}\int_{R^3} g\phi dv| = \\ &= \int Ldxdt| \int {}^0\bar{M}(x,v,t)^{0*}\phi Ldv - {}^0 * \int g\phi dv|. \end{aligned}$$

In particular, for $\phi \in \cup_{k\in N} M_k$

$$\begin{aligned} &| \int_{\Omega\times R^3\times[0,T]} (g\phi - M\phi)dxdvdt| = \\ &= | \int_{ns^*\Omega\times\, {}^*R^3\times\, {}^*[0,T]} ({}^{0*}g^{0*}\phi - {}^{0*}M\, {}^{0*}\phi)Ldxdvdt| = \\ &= | \int_{{}^*\Omega\times {}^*[0,T]} Ldxdt \int_{ns^*R^3} Ldv({}^{0*}g^{0*}\phi - {}^0\bar{M}^{0*}\phi)| \le \\ &\le \int_{{}^*\Omega\times\, {}^*[0,T]} Ldxdt| \int_{ns^*R^3} Ldv({}^{0*}g^{0*}\phi - {}^0\bar{M}^{0*}\phi| = 0, \end{aligned}$$

It follows that $g = M$, i.e., g is a time-dependent local Maxwellian.

By Lemma A6 for $\nu \in N$, $\kappa \in {}^*N^\infty$

$$\int_{{}^*\Omega\times R^3} {}^*\chi_\nu | {}^*f(t+t_\kappa) - {}^*M|^* dxdv \approx 0, \text{ Loeb a.e. } t \in ns^*R^+. \quad \text{(A15)}$$

By the S-continuity of the L^1-mapping $t \to {}^*f(t)$, the relation (A15) holds for all $t \in ns^*R^+$. From here it follows by underspill that in the standard context

$$\lim_{k\to\infty} \int_{\Omega\times R^3\times[0,T]} \chi_\nu |f(t+t_k) - M| dxdvdt = 0, \quad T \in R^+,$$

$$\lim_{k\to\infty} \int_{\Omega\times R^3} \chi_\nu |f(t+t_k) - M| dxdv = 0, \quad t \in R^+.$$

By the energy bound of Theorem 2.4, the same result holds, without χ_ν.

From here it is a routine standard result that M has the form given in the theorem. □

Remark that a standard proof based on the result of Example 1 is also possible. Both proofs (above resp. in Example 1) contain transparent, typically nonstandard arguments. Whether there is any transparent standard counterpart of those arguments, is unknown at the time of this writing.

References

1. Arkeryd, L., (1972) On the Boltzmann equation , *Arch. Rat. Mechs. Anal.* **45**, pp.1-34.
2. Arkeryd, L., (1993) Some examples of NSA methods in kinetic theory, *Lecture notes in Mathematics* **1551**, Springer, Berlin.
3. Arkeryd, L. and Maslova, N., (1994) On diffuse reflection at the boundary for the Boltzmann equation and related equations, *Journ. Stat. Phys.* **77**, pp. 1051–1077.
4. Arkeryd, L. and Nouri, A. Asymptotics of the Boltzmann equation with diffuse reflection boundary conditions, to appear in *Monatsheft für Mathematik.*
5. Carleman, T. (1957) *Théorie Cinétique des Gaz.* Almquist & Wiksel, Uppsala.
6. Cercignani, C. (1988) *The Boltzmann equation and its applications.* Springer, Berlin.
7. Desvillettes, L. (1990) Convergence to equilibrium in large time for Boltzmann and BGK equation, *Arch. Rat. Mechs. Anal.* **110**, pp. 73–91.
8. Pettersson, R (1990) On solutions to the linear Boltzmann equation with general boundary conditions and infinite range forces, *Journ. Stat. Phys.* **59**, pp. 403–440.

3. On semiclassical limits for the Schrödinger equation

The aim of this presentation is to discuss linear and nonlinear Schrödinger equations from the perspective of their infinitesimally supported wave packets and to connect the Schrödinger equation for Planck's constant $\hbar \approx 0$, to semiclassical limits such as the Liouville and Vlasov equations. No new standard results are offered, but some aspects of the approach may not have their standard counterpart in the existing literature. My hope is that the intuitive simplicity of the NSA approach should be helpful also for an attack on related but so far unsolved problems, just as the NSA perspective has already been helpful in a number of cases within nonlinear kinetic theory (e.g. Example 1, 2 above and [1]).

Let H be a Hamiltonian, for simplicity

$$H(x,p) = \frac{p^2}{2m} + V(x), \quad H(x, -i\hbar\nabla_x) = -\frac{\hbar^2}{2m}\Delta_x + V(x).$$

Consider the Schrödinger equation

$$i\hbar\partial_t u = -\frac{\hbar^2}{2m}\Delta_x u + V(x)u,\ t > 0,\ x \in \mathbb{R}^n, \tag{SE}$$

with initial values

$$u(0,x) = u^0(x).$$

We shall first follow J. Harthong [4] in discussing wave packet solutions for $\hbar \simeq 0$, and the appearance of the trajectories of the Hamilton equations

$$\dot{x} = \frac{\partial H}{\partial p}(x,p), \quad \dot{p} = -\frac{\partial H}{\partial x}(x,p). \tag{HE}$$

i.e., the trajectories of the corresponding classical particle system.

Let us start by solving (SE) for a free particle, i.e., in the particular case when $V \equiv 0$,

$$i\hbar\frac{\partial u}{\partial t} = -\frac{\hbar^2}{2m}\Delta u, \quad x \in \mathbb{R}^3, t > 0,$$

with e.g.

$$u(0,x) = k^{-1}\exp\Big(-\frac{(x-x_0)^2}{2\epsilon} + \frac{i}{\hbar}p_0 \cdot x\Big).$$

For $k = (\epsilon\pi)^{3/4}$ the amplitude square $|u(0,x)|^2$ is a Gaussian probability density, $\int |u(0,x)|^2 dx = 1$. Under the *physicist's Fourier transform*

$$\hat{u}(t,p) = (2\pi\hbar)^{-3/2}\int u(t,x)\exp(-\frac{i}{\hbar}p\cdot x)dx$$

this (SE) is transformed into

$$i\hbar\frac{\partial}{\partial t}\hat{u} = \frac{p^2}{2m}\hat{u}.$$

We shall later also need the *mathematician's Fourier transform* $\mathcal{F}f$ which is $\hat{f}$ with $\hbar = 1$.

The Schrödinger equation for a free particle has the solution

$$\hat{u}(t,p) = \hat{u}(0,p)\exp(-\frac{i}{\hbar}\frac{p^2}{2m}t).$$

The easy explicit computation of $\hat{u}(0,p)$ followed by an explicit inverse Fourier transform gives

$$u(t,x) = \left(\frac{\pi\beta^2}{\epsilon}\right)^{-3/4} \exp\left(-\frac{(x - x_0 - \frac{t}{m}p_0)^2}{2\beta} + \frac{i}{\hbar}p_0 \cdot x - \frac{i}{\hbar}\frac{p_0^2 t}{2m}\right),$$

with $\beta = \epsilon + \dfrac{i\hbar t}{m}$. Here $|u(t,x)|^2$ is also a (Gaussian) probability density. The center $x = x_0 + t\dfrac{p_0}{m}$ moves with velocity $v_0 = \frac{p_0}{m}$. At $t = 0$ but not for $t > 0$ the important momentum p_0 is not contained in the macroscopic probability density $|u|^2$, only in the microscopic phase factor $\exp(\frac{i}{\hbar}p_0 \cdot x)$.

The computation so far could equally well have been carried out in a standard and a nonstandard context. But for $\epsilon \approx 0$, the initial value $u(0,x)$ is concentrated in an infinitesimal neighbourhood of $x = x_0$, and for $\frac{\hbar t}{m\epsilon} \simeq 0$ the solution at time t in an infinitesimal neighbourhood of $x = x_0 + tv_0$. This is a suitable mathematical counterpart of the wave packet from physicist's. The physicist's Fourier transform of $u(0,x)$ is

$$\hat{u}(0,p) = (\pi\eta)^{-3/4} \exp\left(-\frac{(p-p_0)^2}{2\eta} - \frac{i}{\hbar}x_0(p-p_0)\right), \quad \eta = \frac{\hbar^2}{\epsilon}.$$

Hence for $\eta = \frac{\hbar^2}{\epsilon} \approx 0$, the physicist's Fourier transform of such wave packets are also wave packets. The center satisfies the (HE), i.e., the equations for the classical particle trajectory.

With more work an analogous result can be proved for more general Hamiltonians and wave packets. For simplicity we take $H(x,p)$ polynomial in p and C^β for β large enough. Let

1) $c \in {}^*L^2(\mathbb{R}^n)$, $\int |c|^2 dx \sim 1$.

2) There exists $x_0 \in ns\,{}^*\mathbb{R}^n$, such that for any standard multi-index $\alpha \neq 0$

$$(i\hbar\frac{\partial}{\partial x})^\alpha c \in {}^*L^2(\mathbb{R}^n), \int |(i\hbar\frac{\partial}{\partial x})^\alpha c|^2 dx \approx 0, \int |(x - x_0)^\alpha c|^2 dx \approx 0.$$

Definition A function $u \in {}^*L^2(\mathbb{R}^n)$ is a *wave packet*, if there is a function c satisfying 1), 2), and $p_0 \in {}^*\mathbb{R}^n$ such that

$$\int |u - c\exp(\frac{i}{\hbar}p_0 \cdot x)|^2 dx \approx 0.$$

The pair (x_0, p_0) is called the *dynamic spectrum* of u.

Wave Packet Theorem [4]. *Suppose that $H(x,p) \in C^{\beta}(\mathbb{R}^{2n})$ for β large enough, and that $H(x, -i\hbar\nabla_x)$ is self-adjoint on ${}^*L^2(\mathbb{R}^n)$. Consider the initial value problem*

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}u &= H(x, -i\hbar\nabla_x)u, \\ u(0,x) &= u_0(x) \end{aligned}$$

with u_0 a wave packet having dynamic spectrum (x_0, p_0).

The solution $u(t,x)$ is a wave packet for ${}^0t < \infty$, and the dynamic spectrum (x_t, p_t) satisfies

$$\dot{x} = \frac{\partial H}{\partial p},\ \dot{p} = -\frac{\partial H}{\partial x},\ x(0) = x_0,\ p(0) = p_0. \tag{HE}$$

This can be proved along the above lines for the free particle (see [4] for a complete proof). Locally the problem is first transformed to canonical coordinates (i.e., new coordinates such that the (HE) still hold) with $H(x,p) = H(x_0, q(x,p)) = K(q(x,p))$. The Schrödinger equation $i\hbar\frac{\partial v}{\partial t} = K(q)v(t,q)$ with initial value $v(0,q)$ a wave packet is then solved. The solution remains a wave packet and its dynamic spectrum propagates along the classical trajectories of the Hamiltonian system.

To return to u and H from v and K, one has to study integrals

$$u(x) = (2\pi\hbar)^{-n/2}\int \exp(\frac{i}{\hbar}\phi(x,q))a(x,q)v(q)dq$$

for which $H(x, -i\hbar\nabla_x)u$ consists of the essential term

$$\begin{aligned} &\int H(x, \frac{\partial\phi}{\partial x}(x,q))a(x,q)\exp(\frac{i}{\hbar}\phi(x,q))v(q)dq = \\ &= \int K(q)v(q)a(x,q)\exp(\frac{i}{\hbar}\phi(x,q)dq, \end{aligned}$$

together with infinitesimal terms. This is the key to how the transformation back to the original coordinates preserves the desired properties of the theorem. Here the step is considerably more involved than in the free particle case. The central ingredient is a kind of Fourier integral operators with the property that for smooth wave packets there is an inversion formula and a Plancherel's theorem. This is an infinitesimal phenomenon which does not hold for functions with extended support.

We now discuss these infinitesimally localized inversion and Plancherel formulas. The Hamilton-Jacobi theory is concerned with canonical transformations χ, $(x,p) = \chi(y,q)$. Here p, q are given through a local generating function ϕ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$, such that $\frac{\partial^2\phi}{\partial x \partial y}$ is nondegenerate, with ϕ a solution to $p = -\frac{\partial\phi}{\partial x}$, $q = \frac{\partial\phi}{\partial y}$

$$H\left(x, \frac{\partial\phi}{\partial x}\right) = \text{constant.} \tag{HJE}$$

We may find such a transformation with $(x_0, p_0) = \chi(y_0, q_0)$, $\phi(x, y) = q_0 y$ on the subspace $x = x_0$. With

$$\hat{u}(x) = \mathcal{F}_{\chi a}u(x) = (2\pi\hbar)^{n/2}\int \exp(-\frac{i}{\hbar}\phi(x,y))a(x,y)u(y)dy,$$

the inversion formula is given by

Inversion Theorem [4]. *Let $u(y) = c(y)\exp(\frac{i}{\hbar}q_0 \cdot y)$ be a wave packet with c satisfying 1) and 2). Consider $\hat{u} = \mathcal{F}_{\chi a}u$ with a standard weight function $a(x,y) \in C_0^\infty$ having compact support in the domain of definition of $\phi(x,y)$. Let b be another weight function of the same type and set*

$$\begin{aligned} v(z) &= (2\pi\hbar)^{-\frac{n}{2}}\int \exp(\frac{i}{\hbar}\phi(x,z))b(x,z)\hat{u}(x)dx, \\ D(z) &= a(x_0,z)b(x_0,z)|\det\frac{\partial^2\phi}{\partial x\partial y}(x_0,z)|^{-1}. \end{aligned}$$

Then $v \approx D(y_0)u$ in ${}^*L^2$.

Under the same hypotheses a Plancherel type result also holds.

Plancherel's Theorem [4] *For a wave packet u*

$$\|\mathcal{F}_{\chi a}u\| \approx |a(x_0,y_0)||\det\frac{\partial^2\phi}{\partial x\partial y}(x_0,y_0)|^{-1/2}\|u\|.$$

Proof of Plancherel's Theorem

$$\begin{aligned} \|\hat{u}\|^2 &= \int \hat{u}\bar{\hat{u}}dx = \\ &= \int[(2\pi\hbar)^{-\frac{n}{2}}\int \exp(-\frac{i}{\hbar}\phi(x,y))a(x,y)\bar{\hat{u}}(x)dx]u(y)dy = \\ &= \int[(2\pi\hbar)^{-\frac{n}{2}}\overline{\int \exp(\frac{i}{\hbar}\phi(x,y))\overline{a(x,y)}\hat{u}(x)dx]\overline{u(y)}}dy \approx \\ &\approx \int |a(x_0,y_0)|^2|\det\frac{\partial^2\phi}{\partial x\partial y}(x_0,y_0)|^{-1}|u(y)|^2dy. \end{aligned}$$

The last step holds for a "smooth" wave packet by the inversion formula and the S-continuity of a, hence by an approximation argument for any wave packet u. □

Proof of the Inversion Theorem

In view of Fubini's theorem we want to prove that

$$\begin{aligned} v(z) &= (2\pi\hbar)^{-\frac{n}{2}} \int \exp(\frac{i}{\hbar}\phi(x,z))b(x,z)\hat{u}(x)dx = \\ &= \int [(2\pi\hbar)^{-n} \int \exp(-\frac{i}{\hbar}(\phi(x,y)-\phi(x,z)))a(x,y)b(x,z)dx]u(y)dy \\ &\approx a(x_0,y_0)b(x_0,y_0)|\det \frac{\partial^2\phi}{\partial x\partial y}(x_0,y_0)|^{-1}u(z) = D(y_0)u(z). \end{aligned}$$

By the choice of ϕ, $\phi(x_0,y) - q_0 y = 0$ and

$$\phi(x,y) - \phi(x,z) = (z-y)(X(x,y,z) - q_0)$$

with the non-degenerate

$$\frac{\partial X}{\partial x}(x_0,z,z) = \frac{\partial^2\phi}{\partial x\partial y}(x_0,z).$$

The assumptions of the theorem imply that supp $a \cdot b$ is contained in a neighbourhood of (x_0,y_0,y_0) where $(x,y,z) \to (X(x,y,z),y,z)$ is a diffeomorphism. Then

$$\begin{aligned} &\exp(-\frac{i}{\hbar}q_0\cdot z)v(z) = \\ &\int [(2\pi\hbar)^{-n} \int \exp(-\frac{i}{\hbar}[\phi(x,y) - q_0 y - \phi(x,z) + q_0 z])abdx]c(y)dy = \\ &= \int [(2\pi\hbar)^{-n} \int \exp(\frac{i}{\hbar}(y-z)\cdot X)J(X,y,z). \\ &\qquad .a(x(X,y,z),y)b(x(X,y,z)z)dXc(y)dy \\ &= \int (2\pi\hbar)^{-n} \int \exp(i\frac{y-z}{\hbar}\cdot X)F(X,y,z)dXc(y)dy \end{aligned}$$

with the Jacobian

$$J = |\det \frac{\partial X}{\partial x}(x,y,z)|^{-1}$$

$$F(X,y,z) = J(X,y,z)a(x(X,y,z),y)b(x(X,y,z),z).$$

Notice that $F(0,z,z) = D(z)$. The change of variables $Z = \frac{-y+z}{\hbar}$ gives

$$\begin{aligned} &\exp(-\frac{i}{\hbar}q_0\cdot z)v(z) \\ &= \int (2\pi)^{-n} \int \exp(-iZ\cdot X)F(X,z-\hbar Z,z)dXc(z-\hbar Z)dZ = \\ &= \int (2\pi)^{-n/2}\mathcal{F}F(Z,z-\hbar Z,z)c(z-\hbar Z)dz. \end{aligned}$$

It can be shown that the ${}^*L^2$ error is infinitesimal when $\mathcal{F}F(Z, z-\hbar Z, z)$ is replaced by $\mathcal{F}F(Z, z, t)$.

Let us end with the details of this last infinitesimal argument. Write

$$\begin{aligned}&\mathcal{F}F(Z, z-\hbar Z, z)c(z-\hbar Z) = \mathcal{F}F(Z,z,z)c(z) + \\ &(\mathcal{F}F(Z, z-\hbar Z, z) - \mathcal{F}F(Z,z,z))c(z-\hbar Z) + \\ &\mathcal{F}F(Z,z,z)(c(z-\hbar Z) - c(Z)) = I_1 + I_2 + I_3.\end{aligned}$$

For I_1 we notice that $(2\pi)^{-\frac{n}{2}} \int \mathcal{F}F(Z,z,z)dZ = F(0,z,z) = D(z)$. By hypotheses $D(z)$ is a smooth standard function and so $|D(y_0) - D(z)| \leq C|y_0 - z|$, where $0 < C \in \mathbb{R}$. Here

$$\begin{aligned}&\int |D(y_0)c(z) - (2\pi)^{\frac{-n}{2}} \int \mathcal{F}F(Z,z,z)c(z)dZ|^2 dz \\ &\leq C \int |y_0 - z|^2 |c(z)|^2 dz \simeq 0,\end{aligned}$$

the last step by 2).

We next consider

$$(2\pi)^{\frac{-n}{2}} \int I_3(Z,z,z)dz = (2\pi)^{-\frac{n}{2}} \int \mathcal{F}F(Z+z)Z \int_0^{\hbar} \frac{\partial c}{\partial z}(z - tZ)dtdz.$$

Now $\mathcal{F}F(Z,z,z)$ is decreasing as $|z|^{-\beta}$, hence for some $\alpha > z/2$

$$\sup \int |\mathcal{F}F(Z,z,z)|^2 (1 + |Z|^{2\alpha})dZ$$

is finite.

$\displaystyle\int \frac{|Z|^2}{1+|Z|^{2\alpha}} dZ$ is also finite. Hence by 2) (and with C denoting finite constants).

$$\begin{aligned}&\int dz |(2\pi)^{-\frac{n}{2}} \int I_3(Z,z,z)dZ|^2 \leq \\ &\leq C \int dz \int dZ \int_0^{\hbar} \frac{\hbar Z^2}{1+|Z|^{2\alpha}} |\frac{\partial c}{\partial z}(z+tZ)|^2 dt \leq \\ &\leq C \int dZ \int_0^{\hbar} dt \int dz |\frac{\partial c}{\partial z}(z+tZ)|^2 \frac{\hbar Z^2}{1+|Z|^{2\alpha}} \leq \\ &\leq C\hbar^2 \int |\frac{\partial c}{\partial z}|^2 dz \approx 0.\end{aligned}$$

The term I_2 can be estimated in a similar manner. $\square$

To study the Schrödinger equation for initial values, $u^0 \in L^2(\mathbb{R}^n)$, we shall introduce a localized frequency distribution together with a natural probability in phase space. The much used Wigner transform has the disadvantage of not being positive, thus not representing a true probability distribution. Instead, we take a measurable family of wave-packets, $c_y(x)\exp(\frac{i}{\hbar}p\cdot x)$, and define the *localized frequency distribution*

$$f_{uc}(y,p) = (2\pi\hbar)^{-\frac{n}{2}} \int u(x)\overline{c_y(x)} \exp\Big(-\frac{i}{\hbar}x\cdot p\Big)dx = \widehat{u(\cdot)\overline{c_y(\cdot)}}(p).$$

Here $\hat{f}$ is the physicist's Fourier transform of u,

$$\hat{f}(\xi) = (2\pi\hbar)^{-\frac{n}{2}} \int \exp\Big(-\frac{i}{\hbar}\xi\cdot x\Big) f(x)dx.$$

It follows that – for $\int |c(y)|^2 dy = 1$ –

$$u(x)\overline{c_y(x)} = (2\pi\hbar)^{-\frac{n}{2}} \int \exp\Big(\frac{i}{\hbar}x\cdot p\Big) f_{uc}(y,p)dp.$$

This gives

$$u(x) = (2\pi\hbar)^{-\frac{n}{2}} \int f_{uc}(y,p)\exp\Big(\frac{i}{\hbar}x\cdot p\Big)c_y(x)dp\,dy.$$

Then with $c_{yp} = c_y(x)\exp\left(\frac{i}{\hbar}x\cdot p\right)$ we can write

$$u(x) = (2\pi\hbar)^{-\frac{n}{2}} \int f_{uc}(y,p)c_{yp}(x)dp\,dy.$$

We thus conclude that the solution of (SE) with initial value $u(0,x) = u^0(x)$ can be written

$$u(t,x) = (2\pi\hbar)^{-\frac{n}{2}} \int f_{u^0c^0}(y,p)c^t_{yp}(x)dp\,dy \tag{S}$$

(for $\hbar \approx 0$ as well as $\hbar \sim 1$ and for $V \in L^2_{\mathrm{loc}}$. The interpretation of (S) is obvious for u^0 having compact support and some regularity). Here $c^t_{yp}(x) = c_{yp}(t,x)$ is the solution of the Wave Packet Theorem with c^0_{yp} a wavepacket with dynamic spectrum (y,p), such that $c^0_y(x) = c^0(y-x)$, $\int |c^0|^2 dx = 1$. In the case $\hbar \approx 0$ and V smooth enough for the Wave Packet Theorem, the solution (S) is a superposition of contributions from individual wavepackets travelling along the trajectories of (HE). The invariance of the measure $dp\,dy$ in phase space implies that the solution can be written

$$u^t(x) = (2\pi\hbar)^{-\frac{n}{2}} \int f_{u^0c^0}(y(y^t,p^t),p(y^t,p^t))c_{y^tp^t}(x)dp^t dy^t, \tag{S'}$$

with $c_{y^t p^t}(x) = c^t_{yp}(x)$ a wavepacket.

We take $|f_{uc}(y,p)|^2$ as a representation of *probability in phase space.*

Properties

P1)[4] $\int |f_{uc}(y,p)|^2 dy\,dp = \int |u(x)|^2 dx$,
if $c_y(x) = c(y-x)$, $\int |c|^2 dx = 1$, $u \in L^2$.

Proof $\int |f_{uc}(y,p)|^2 dy\,dp = \int\int |\widehat{u(\cdot)\overline{c_y(\cdot)}}(p)|^2 dp dy =$
$= \int\int |u(x)\overline{c_y(x)}|^2 dx dy = \int |u(x)|^2 dx \int |c(y)|^2 dy = \int |u(x)|^2 dx.$ □

Remark that the proof implies $\int |f|^2 dp \approx |u(y)|^2$ if $u \in C_0 \cap L^2$ without the condition $c_y(x) = c(y-x)$, and then by L^2 continuity P1) with $\approx$ for $u \in L^2$.

P2)[4] For $\hbar \approx 0$, $u \in L^2(\mathbb{R}^n)$, $\mathcal{F}_p\varphi \in C_c(\mathbb{R}^{2n})$ it holds

$$\int |f_{uc}(x,p)|^2 \varphi(x,p) dx dp \approx (2\pi)^{-\frac{n}{2}} \int\int u(x)\bar{u}(x+\hbar X)\mathcal{F}_p\varphi(x,X) dx dX.$$

Remark that from a standard point of view, this is independent of the (measurable) family c_{yp} of wavepackets. In particular $c_{y^t p^t}(x)$ and $c(y^t - x)\exp\left(\frac{ip^t x}{\hbar}\right)$ with c of P1) give the same result.

Proof

$$(2\pi)^{-\frac{n}{2}} \int |\widehat{u(\cdot)\bar{c}_x(\cdot)}(p)|^2 \exp(iXp) dp = [p = \hbar\xi]$$
$$(2\pi)^{-\frac{n}{2}} \int \hbar^{\frac{n}{2}} \widehat{u(\cdot)\bar{c}_x(\cdot)}(\hbar\xi) \hbar^{\frac{n}{2}} \overline{\widehat{u(\cdot)\bar{c}_x(\cdot)}(\hbar\xi)} \exp(i\hbar X\xi) d\xi$$
$$= (2\pi)^{-\frac{n}{2}} \int u(y)\bar{c}_x(y)\bar{u}(y+\hbar X)c_x(y+\hbar X) dy =$$
$$(2\pi)^{-\frac{n}{2}} \int u(y)\bar{u}(y+\hbar X)\bar{c}_x(y)c_x(y+\hbar X) dy.$$

Hence

$$\int \overline{\mathcal{F}_p |f_{uc}(x,\cdot)|^2} \mathcal{F}_p\varphi dx dX =$$
$$= (2\pi)^{-\frac{n}{2}} \int\int \left(\int u(y)\bar{u}(y+\hbar X)\bar{c}_x(y)c_x(y+\hbar X)\mathcal{F}_p\varphi(x,X) dx dX dy.\right.$$

$\mathcal{F}_p\varphi$ is S-continuous with support in $ns^*\mathbb{R}^n$. By 2) in the wavepacket definition c_x is infinitesimally close in $^*L^2\mathbb{R}^n)$ to a function $\tilde{c}$ with – for $X \in ns^*\mathbb{R}^n$ –

$$\|\hbar X \int_0^1 \partial_z \tilde{c}(z + t\hbar X) dt\|_{L^2_z} \leq |X| \int_0^1 dt \|\hbar\partial_z \tilde{c}(z + \hbar t X)\|_{L^2_z} \approx 0.$$

Since

$$\tilde{c}(y+\hbar X)=\tilde{c}(y)+\hbar X\int_0^1 \partial_y \tilde{c}(y+t\hbar X)dt,$$

it follows that for $u \in L^2(\mathbb{R}^n)$

$$\overline{\int \mathcal{F}_p |f_{uc}(x,\cdot)|^2 \mathcal{F}_p\varphi dx dX} \approx$$
$$(2\pi)^{-\frac{n}{2}}\int\int\Big(\int u(y)\bar{u}(y+\hbar X)|c_x(y)|^2\mathcal{F}_p\varphi(x,X)dy\Big)dX dx \approx$$
$$(2\pi)^{-\frac{n}{2}}\int\int u(x)\bar{u}(x+\hbar X)\mathcal{F}_p\varphi(x,X)dX dx,$$

the last step since u, φ are standard, hence approximately constant on monads. □

P3) For $u^t \in H^2$ with $Vu^t \in L^2$ and $c(y-x)\exp\big(\frac{i}{\hbar}p\cdot x\big)$ a time independent wavepacket

$$i\hbar\partial_t |f_{u^t c}|^2 = f_{Hu^t c}\overline{f_{u^t c}} - f_{u^t c}\overline{f_{Hu^t c}}.$$

Proof

$$i\hbar\partial_t |f_{u^t c}(y,p)|^2 = i\hbar\partial_t f_{u^t c}\overline{f_{u^t c}} - f_{u^t c}\overline{i\hbar\partial_t f_{u^t c}}$$
$$=(2\pi\hbar)^{-n}\int e^{\frac{i}{\hbar}(x-z)p}\overline{c(x)c(z)}\Big[\Big(-\frac{\hbar^2}{2m}\Delta+V\Big)u^t(y-x)\overline{u^t(y-z)}-$$
$$-u^t(y-x)\Big(-\frac{\hbar^2}{2m}\Delta+V\Big)\overline{u^t(y-z)}\Big]dx dz.$$

(For details cf. [2], [7] for smooth functions, then take the limit.) □

P4) If c is a Gaussian, then $(2\pi)^{\frac{n}{2}}|f_{uc}|^2$ is the Husimi transform.

Proof Let W_u denote the Wigner transform

$$W_u(x,p)=(2\pi\hbar)^{-n}\int u\Big(x+\frac{z}{2}\Big)\overline{u\Big(x-\frac{z}{2}\Big)}e^{-\frac{i}{\hbar}p\cdot z}dz.$$

The Husimi transform is

$$\tilde{W}_u = W_u * \exp\Big(-\frac{x^2+p^2}{\hbar}\Big)(\pi\hbar)^{-n}$$
$$=[LPp.19]=2^{n/2}(2\pi\hbar)^{-n}\Big|\int u(z)e^{\frac{-(y-z)^2}{4\hbar}}(2\pi\hbar)^{-\frac{n}{4}}e^{-\frac{ipz}{2\hbar}}dz\Big|^2=$$
$$=(2)^{\frac{n}{2}}\Big|f_{uc}\Big(y,\frac{p}{2}\Big)\Big|^2$$

with $c(x)=(2\pi\hbar)^{-\frac{n}{4}}e^{-\frac{x^2}{4\hbar}}$. □

Remark that use of the Gaussian instead of dynamically evolving wave-packets like in (S′) sometimes makes formulas less adapted to the structure of the equations.

P5) Let $V \in C^1, u^0 \in L^2(\mathbb{R}^n)$. Let $u \in C(\mathbb{R}_+, H^1(\mathbb{R}^n))$ be the unique solution of the (SE) with initial value u^0. If $u^t \in H^2$, then in distribution sense

$$i\hbar\partial_t(\overline{u^t(x)u^t(x+\hbar X)}) = H(x, -i\hbar\nabla_x)u^t(x)\overline{u^t(x+\hbar X)} -$$
$$-u^t(x)\overline{H(x+\hbar X, -i\hbar\nabla_{x+\hbar X})u^t(x+\hbar X)}.$$

Proof This is formally obvious, for details via approximation with smooth functions see [2], [7].

P6) For $\hbar \approx 0$, $u \in H^2$ and $\mathcal{F}_p\varphi \in C_c^2(\mathbb{R}^{2n+1})$, by P2) and P5)

$$(2\pi)^{n/2}\int dxdp|f_{u^tc}|^2 i\partial_t\varphi \approx \int\int dxdX\, u^t(x)\overline{u^t(x+\hbar X)}i\partial_t\mathcal{F}_p\varphi(t,x,X)$$
$$= -\frac{1}{\hbar}\int\int dxdX[H(x,-i\hbar\nabla_x)u^t(x)\overline{u^t(x+\hbar X)} -$$
$$-u^t(x)\overline{H(x+\hbar X, -i\hbar\nabla_{x+\hbar X})u^t(x+\hbar X)} = \mathcal{F}_p\varphi(t,x,X).$$

P6a) The V-term of P6) satisfies (with the t variable suppressed in φ)

$$\int\int dxdX\, u^t(x)\overline{u^t(x+\hbar X)}\frac{1}{\hbar}(V(x) - V(x+\hbar X))\mathcal{F}_p\varphi(x,X) \approx$$
$$\approx \int\int dxdX\, u^t(x)\overline{u^t(x+\hbar X)}X\nabla_x V(x)\mathcal{F}_p\varphi(x,X) =$$
$$= -i\int\int dxdX\, u^t(x)\overline{u^t(x+\hbar X)}\nabla_x V(x)\mathcal{F}_p\nabla_p\varphi(x,X) \approx$$
$$\approx -i(2\pi)^{\frac{n}{2}}\int\int |f_{u^tc}|^2\nabla_x V(x)\nabla_p\varphi(x,p)dxdp,$$

the last step by P2). □

P6b) The Δ-term of P6) satisfies

$$-\frac{\hbar}{2m}\int\int dxdX(\Delta_x u^t(x)\overline{u^t(x+\hbar X)} - u^t(x)\overline{\Delta_x u^t(x+\hbar X)}\mathcal{F}_p\varphi(x,X))$$
$$= \frac{\hbar}{2m}\int\int dxdX\nabla_x\mathcal{F}_p\varphi(\nabla_x u^t(x)\overline{u^t(x+\hbar X)} - u^t(x)\overline{\nabla_x u^t(x+\hbar X)})$$
$$= \int\int dxdX\nabla_1\mathcal{F}_p\varphi(x - \frac{\hbar}{2}X, X)[-\nabla_X u^t(x - \frac{\hbar}{2}X)\overline{u^t(x + \frac{\hbar}{2}X)} -$$

$$-u^t(x-\frac{\hbar}{2}X)\overline{\nabla_X u^t(x+\frac{\hbar}{2}X)}] = \int\int dxdX D_X \nabla_1 \mathcal{F}_p \varphi(x-\frac{\hbar}{2}X, X)$$
$$\cdot u^t(x-\frac{\hbar}{2}X)\overline{u^t(x+\frac{\hbar}{2}X)} = I$$

Since

$$D_X\nabla_1\mathcal{F}_p\varphi(x-\frac{\hbar}{2}X,X) = -\frac{\hbar}{2}\nabla_1\nabla_1\mathcal{F}_p\varphi(x-\frac{\hbar}{2}X,X)+\nabla_2\nabla_1\mathcal{F}_p\varphi(x-\frac{\hbar}{2}X,X),$$

it follows that

$$I \approx \int\int dxdX\nabla_2\nabla_1\mathcal{F}_p\varphi(x-\frac{\hbar}{2}X,X)u^t(x-\frac{\hbar}{2}X)\overline{u^t(x+\frac{\hbar}{2}X)} =$$
$$= \int\int dxdX\nabla_1\mathcal{F}_p(ip\varphi)(x-\frac{\hbar}{2}X,X)u^t(x-\frac{\hbar}{2}X)\overline{u^t(x+\frac{\hbar}{2}X)} =$$
$$= \int\int dxdX\nabla_1\mathcal{F}_p(ip\varphi)(x,X)u^t(x)\overline{u^t(x+\hbar X)} \approx$$
$$\approx (2\pi)^{-n/2}\int\int dxdpip\nabla_x\varphi(x,p)|f_{u^tc}(x,p)|^2,$$

the last step by P2). □

We have thus proved that for $V \in C^1$ and $\hbar \approx 0$, in distribution sense μ satisfies the Liouville equation

$$\partial_t\mu = \nabla_x V(x)\nabla_p\mu - p\nabla_x\mu, \tag{LE}$$

where $(\mu^t,\varphi) = {}^0\!\int |f_{u^tc}|^2\, {}^*\varphi dydp$. This follows first for $u \in H^2$, then by density arguments and P1) for $u \in L^2$. By the remark of P2) μ^t only depends on the solution u^t of (SE) (for $\hbar \approx 0$) and not on c.

It also follows in the standard context that for each test function φ, the quantity $|f_\hbar|^2$ of the solution of (SE) for $\hbar > 0$ converges (after taking a subsequence) in the sense of measures to a solution of the Liouville equation. If $V \in C^{1,1}$, then the solution of Liouville equation is unique and the full sequence $|f_\hbar|^2$ converges to this solution.

This standard convergence is a consequence of the weak* compactness of the closure W of $\{|f_\hbar^t|^2, 0 < \hbar < 1\}$, which in turn follows from P1) together with the conservation of $\int |u^t|^2dx$ for solutions u of the Schrödinger equation. Let d be the metric for the weak*topology of measures on W. We recall the well known

Lemma 3.1 *If $\tilde{\mathcal{E}} \in {}^*W$, then there is $\mathcal{E} \in W$ so that ${}^*d(\mathcal{E},\tilde{\mathcal{E}}) \approx 0$, i.e., $\tilde{\mathcal{E}}$ is in the monad of $\mathcal{E}$.*

For a proof see any main text on NSA. Hence given a solution $u_\hbar$ of (SE) for $\hbar \approx 0$ there is $\mathcal{E} \in W$ such that ${}^*d(\mathcal{E}, |f_\hbar^t|^2) \approx 0$. And so there is a standard sequence $(\hbar_j), \hbar_j \searrow 0$, such that in weak* measure sense $|f_{u^t c}|^2$ converges to $\mathcal{E}$. Evidently $\mathcal{E} = \mu$.

Remark that for $V \in C^{1,1}$, (HE) and (LE) have unique solutions and μ is given by the Hamiltonian flow H_t,

$$\int \varphi d\mu(t) = \int \varphi \circ H_t d\mu_0.$$

For $V \in C^\beta$, β large enough, and for u^0 sufficiently smooth with compact support, it follows by (S') and without using the uniqueness of the Liouville equation, that μ is given by the Hamiltonian flow, and that the Schrödinger solutions in the limit $\hbar \to 0$ is built up from Hamiltonian path transport. The same then holds for initial L^2 data by a density argument. Similarly, by approximation with potentials in C^β, the case of $V \in C^{1,1}$ gives an analogous result.

We finally turn to the nonlinear case, starting from the Schrödinger Poisson system (SP) with for simplicity $x \in \mathbb{R}^3$,

$$\begin{aligned} i\hbar\partial_t\psi_m &= -\frac{\hbar^2}{2m}\Delta\psi_m + V\psi_m,\ m \in \mathbb{N}, \\ \psi_m(x,0) &= \varphi_m,\ x \in \mathbb{R}^3, \\ \Delta V &= -n,\ n = \sum \lambda_m|\psi_m(x,t)|^2. \end{aligned} \tag{SP}$$

Here $\lambda_m \geq 0, \sum \lambda_m = 1$, $\varphi_m \in L^2(\mathbb{R}^3)$, $m \in \mathbb{N}$, $\int \varphi_j\bar{\varphi}_k dx = \delta_{jk}$, $V = \mathcal{V} * n$, (cf. [5]).

By [3], [7] for $\mathcal{V} = \dfrac{1}{4\pi|x|}$ (repulsive case) there exists a unique solution to the problem (SP).

This is also so in the simpler case when the singularity of $\mathcal{V}$ at $x = 0$ is removed.

Lemma 3.2 $\int \psi_j(t)\overline{\psi_k(t)}dx = \delta_{jk}$, $t > 0$.

Proof From

$$\frac{d}{dt}||\psi(t)||^2 = 0,\ \frac{d}{dt}||\psi_j(t) + \psi_k(t)||^2 = 0$$

it follows that, for $j \neq k$,

$$\frac{d}{dt}\mathrm{Re}\int \psi_j(t)\overline{\psi_k(t)}dx = 0,\ \mathrm{Re}\int \psi_j(t)\overline{\psi_k(t)}dx = \mathrm{Re}\int \psi_j(0)\overline{\psi_k(0)}dx = 0.$$

Similarly,

$$\frac{d}{dt}||i\psi||^2 = 0, \quad \frac{d}{dt}||i\psi_j + \psi_k||^2 = 0$$

imply that

$$\frac{d}{dt} Im \int \psi_j(t)\overline{\psi_k(t)}dx = 0,\ Im \int \psi_j(t)\overline{\psi_k(t)}dx = 0,\ j \neq k.$$

Hence

$$\int \psi_j(t)\overline{\psi_k(t)}dx = 0,\ j \neq k.$$

We next observe that the earlier discussion holds for $\psi := \sum \sqrt{\lambda_m}\psi_m$ with $\psi_0 = \sum \sqrt{\lambda_m}\varphi_m$. In particular the $f_{\psi c}$ discussion holds if $V \in C^1$, i.e., in the case of $\mathcal{V}$ without singularity at $x = 0$. Then for $\hbar \approx 0$ the function μ defined by

$$(\mu^t, \varphi) = {}^0\!\!\int |f_{\psi^t c}|^2 \varphi dy dp$$

in distribution sense satisfies the Vlasov equation

$$\frac{\partial \mu}{\partial t} = \nabla_x V(x) \cdot \nabla_p \mu - p\nabla_x \mu.$$

Under slightly stronger assumptions on the initial values this analysis can also be carried out for $\mathcal{V} = \frac{1}{4\pi|x|}$ (cf. the Wigner transform approach in [8]), i.e., for the full Vlasov equation. □

References

1. Arkeryd, L. (1993), Some examples of NSA methods in kinetic theory, *Lecture Notes in Mathematics* **1551**, Springer, Berlin.
2. Brezzi, F and Markowich, P. (1991) The three-dimensional Wigner-Poisson problem: existence, uniqueness and approximation, *Math. Meth. Appl. Sci.* **14**, pp. 35-61.
3. Ginibre, J. and Velo, G. (1980) On a class of nonlinear Schrödinger equations with nonlocal interactions, *Math. Z.* **170**, pp. 109-136.
4. Harthong, J. (1984) Études sur la mechanique quantique, *Astérisque* **111**.
5. Illner, R. (1992) The Wigner-Poisson and Schrödinger-Poisson systems, *TTSP* **21**, pp. 753-767.
6. Illner R., Bohun, S. and Zweifel P.F. (1991), Some remarks on the Wigner transform and the Wigner-Poisson system, *Proc. VI Conf. Waves and Stability in Continuous Media.*
7. Illner, R., Lange, H. and Zweifel, P.F. (1994) Global existence, uniqueness, and asymptotic behaviour of solutions of the Wigner-Poisson and Schrödinger-Poisson systems, *Math. Meth. Appl. Sci.* **17**, pp. 349-376.
8. Lions, P.L and Paul, T. (1993) Sur les mesures de Wigner, *Revista Mat. Ib.*, pp. 553-618.

A NONSTANDARD APPROACH TO HYDROMECHANICS

Navier–Stokes Equations

M. CAPIŃSKI
Institute of Mathematics
Jagiellonian University
Reymonta 4
30–059 Kraków
Poland
email: capinski@im.uj.edu.pl

1. Introduction

The main object of our investigation is the system of stochastic Navier–Stokes equations

$$\begin{cases} \dfrac{\partial u_j}{\partial t} = \nu \displaystyle\sum_{i=1}^{n} \dfrac{\partial^2 u_j}{\partial x_i^2} - \displaystyle\sum_{i=1}^{n} u_i \dfrac{\partial u_j}{\partial x_i} + f_j(t,u) - \dfrac{\partial p}{\partial x_j} + g(t,u)\dfrac{\mathrm{d}w_t}{\mathrm{d}t} \\ \operatorname{div} u = \displaystyle\sum_{i=1}^{n} \dfrac{\partial u_i}{\partial x_i} = 0. \end{cases} \tag{1}$$

considered in a bounded domain D in $\mathbb{R}^n$ with boundary of class C^2. The vector $u = (u_1, \ldots, u_n)$ represents the velocity of a fluid, ν is the viscosity coefficient, (f_j) is the vector of external forces, p represents the pressure, $\frac{\mathrm{d}w_t}{\mathrm{d}t}$ is white noise. Functions with vanishing divergence are called solenoidal. This condition corresponds to the fact that the fluid is incompressible.

This system can be considerably simplified. First, we built the incompressibility condition into the functional-analytic setup by working in spaces of solenoidal functions. Then the gradient of the pressure can be discarded when we consider the equation in appropriate way since this term is orthogonal to solenoidal functions.

1.1. HISTORY

The deterministic system ($g = 0$) has a vast literature beginning with the pioneering papers of Leray in 1933–1934. Existence and uniqueness are

L.O. Arkeryd et al. (eds.), Nonstandard Analysis: Theory and Applications, 341–355.

proved for $n = 2$ and for the most interesting case $n = 3$ existence is proved but the uniqueness problem is open. The stochastic case with g depending on u has been investigated recently with first existence results obtained using nonstandard methods.

1.2. SPACES AND OPERATORS

We introduce an abstract framework suitable for the formulation of the equation.

Suppose that H is a real separable Hilbert space with an orthonormal basis $\{e_k\}$. Let λ_k be a sequence of positive numbers with $\lambda_k \nearrow \infty$. We define a scale of Hilbert spaces H^r by

$$H^r = \{u \in H : \sum_{k=1}^{\infty} \lambda_k^r (u, e_k)^2 < \infty\}$$

and writing $u_k = (u, e_k)$, $v_k = (v, e_k)$ we have

$$(u, v)_r = \sum_{k=1}^{\infty} \lambda_k^r u_k v_k, \quad |u|_r^2 = \sum_{k=1}^{\infty} \lambda_k^r u_k^2,$$

for $r > 0$, and we take H^{-r} to be the dual to H^r. The space H^{-r} can be regarded as the space of sequences (u_k) with finite $\sum_{k=1}^{\infty} \lambda_k^{-r} u_k$. For $u \in H^{-r}$ and $v \in H^r$ we denote the value of u on v by (u, v). This is justified since

$$(u, v) = \sum u_k v_k = \sum \lambda_k^{-r/2} u_k \lambda_k^{r/2} v_k \le |u|_{-r} |v|_r$$

thus the duality between H^{-r} and H^r extends that between H^0 and $(H^0)'$ (identified with H^0).

We write $V = H^1$, $V' = H^{-1}$, and we have $H = H^0$. Identifying H with its dual we have

$$V \subset H \subset V'.$$

The symbols $(\cdot, \cdot)$, $|\cdot|$ are used for the scalar product and norm in H and $((\cdot, \cdot))$, $\|\cdot\|$ for those in V.

Let

$$Au = \sum \lambda_k (u, e_k) e_k.$$

It is an unbounded operator in H with $\mathrm{dom} A = H^2$ and $|Au| = |u|_2$, whose eigenvectors are e_k: $Ae_k = \lambda_k e_k$. We can extend A to $A : V \to V'$ in a natural way: for $u \in V$, $v \in V$, we put

$$Au[v] := (Au, v) = \sum \lambda_k u_k v_k = ((u, v)).$$

Let $b : V \times V \times V \to \mathbb{R}$ be a trilinear form satisfying

$$b(u, v, z) = -b(u, z, v)$$

which implies

$$b(u, v, v) = 0.$$

We also assume that

$$b(u, v, z) \leq c|u|_\alpha |v|_\beta |z|_\gamma,$$

where $\alpha + \beta + \gamma > 1 + \frac{n}{2}$, $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$. In particular, for $n = 3$ we clearly have

$$b(u, v, z) \leq c\|u\| \, \|v\| \, \|z\|,$$

$$b(u, v, z) \leq c|u| \, \|v\| \, |Az|,$$

$$b(u, v, z) \leq c|Au| \, \|v\| \, |z|,$$

for suitable u, v, z.

For any $u, v \in V$, the mapping $z \to b(u, v, z)$ defines a continuous functional on V which gives rise to the bilinear operator $B : V \times V \to V'$ determined by

$$(B(u, v), z) = b(u, v, z).$$

The example. We introduce some concrete spaces and operators for studying the Navier–Stokes equations. Let $H = L^2_{\mathrm{div}}(D)$ be the space of vector functions $u = (u^1, \ldots, u^n)$, that is the completion of the set $C^\infty_{0,\mathrm{div}}(D)$ of smooth solenoidal functions in the L^2-norm. It is a separable Hilbert space with the scalar product

$$(u, v) = \sum_{i=1}^{n} \int_D u^i(x) v^i(x) \mathrm{d}x$$

and the norm $|u|^2 = (u, u)$. The Laplace operator $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ is densely defined in H on the set $C^\infty_{0,\mathrm{div}}(D)$ and we take A to be the self-adjoint extension of $-\Delta$ in H (the Stokes operator). It is a non-negative operator which gives an orthonormal basis of H consisting of eigenvectors $\{e_k\}$ of A with eigenvalues $0 < \lambda_k \nearrow \infty$.

Let V be the domain of $A^{1/2}$. We equip V with the scalar product

$$((u, v)) = \sum_{i=1}^{n} \left(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right).$$

We define a trilinear form b by

$$b(u,v,z)=\sum_{i,j=1}^{n}\int_D u^i(x)\frac{\partial v^j(x)}{\partial x_i}z^j(x)\mathrm{d}x.$$

It can be shown that it satisfies all the properties given before.

Here we demonstrate the crucial property $b(u,v,z) = -b(u,z,v)$. For $u, v, z \in C^{\infty}_{0,\mathrm{div}}(D)$ we integrate by parts

$$b(u,v,z) = -\sum_{i,j=1}^{n}\int_D \frac{\partial u^j(x)}{\partial x_i}v^j(x)z^j(x)\mathrm{d}x - \sum_{i,j=1}^{n}\int_D u^i(x)v^j(x)\frac{\partial z^j(x)}{\partial x_i}\mathrm{d}x$$

and the first term on the right vanishes since $\mathrm{div}u = 0$, and the second gives $-b(u,z,v)$. This can be extended by continuity to general u, v, z.

1.3. CONTENTS OF THE PAPER

In the next section we give a detailed proof of existence of solution for the deterministic Navier–Stokes equation based on the proof given in [1]. We also discuss the nonstandard approach to the uniqueness problem.

In Section 3 we introduce the notion of a statistical solution. To construct statistical solution we employ the representation of measures on space H by means of their nonstandard densities. Section 4 is devoted to stochastic equation. We state the result and give the idea of a proof. Section 5 contains some open problems.

Except for Theorem 2.2, the proofs are either sketched or omitted. Complete proofs can be found in [2].

2. Deterministic Navier–Stokes equations

We consider (1) with $g = 0$ in the abstract setting of the previous section.

2.1. DEFINITION OF SOLUTION

Let $u_0 \in H$ and $f : [0,T) \to V'$. Consider the following evolution equation in V':

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -Au(t) - B(u(t),u(t)) + f(t),$$

$$u(0) = u_0.$$

For the first two terms on the right-hand side of the equation to make sense we will seek the solution as a function $u : [0, T) \to V$. To avoid difficulties concerning the regularity of u in t we write the equation in the integral form at the same time evaluating the functionals in V' on test elements from V.

Definition 2.1 A weak solution of Navier–Stokes equations is a function $u \in L^2(0,T;V) \cap L^\infty(0,T;H) \cap C(0,T;H_{\text{weak}})$ satisfying

$$(u(t),v) - (u_0,v) = \int_0^t [-\nu((u(s),v)) - b(u(s),u(s),v) + (f(s),v)]\mathrm{d}s \quad (2)$$

for all $t \in [0,T)$, $v \in V$.

2.2. EXISTENCE OF SOLUTIONS

In this section we give a complete self-contained proof of the existence result.

Theorem 2.2 *For any $u_0 \in H$ and $f \in L^2(0,T;V')$ for all $T < \infty$, there exists a weak solution of the Navier–Stokes equations.*

Proof
Step 1 - nonstandard Galerkin approximation.

Take the nonstandard extension ${}^*H$ of H. Denote the *extension of $\{e_k\}_{k\in\mathbb{N}}$ by $\{E_k\}_{k\in{}^*\mathbb{N}}$. Fix $N \in {}^*\mathbb{N}\setminus\mathbb{N}$ and let H_N be the subspace of ${}^*H$ spanned by $\{E_1,\ldots,E_N\}$. Consider an H_N valued function

$$U(\tau) = \sum_{k=1}^{N} U_k(\tau)E_k$$

and the following nonstandard Galerkin equation in H_N:

$$\dot{U}_k(\tau) = -\nu\lambda_k U_k(\tau) - {}^*b(U(\tau),U(\tau),E_k) + F_k(\tau) \quad (3)$$

$k = 1,\ldots,N$, where $\tau \in {}^*(0,\infty)$, $F_k(\tau) = ({}^*f(\tau),E_k)$ and the initial condition is $U_0 = \sum_{k=1}^{N}({}^*u_0,E_k)E_k$. By transfer of the elementary theory there exists a unique local solution to this equation. We will prove that this solution satisfies the energy inequality

$$|U(\tau)|^2 + \nu\int_0^\tau \|U(\sigma)\|^2\mathrm{d}\sigma \le K = |U_0|^2 + \frac{1}{\nu}\int_0^\tau |F(\sigma)|^2_{V'}\mathrm{d}\sigma \quad (4)$$

for all $\tau \in {}^*[0,\infty)$. To this end compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\sigma}|U(\sigma)|^2 &= \sum_{k=1}^{N} \frac{\mathrm{d}}{\mathrm{d}\sigma} U_k^2(\sigma) \\
&= -2\nu \sum_{k=1}^{N} \lambda_k U_k^2(\sigma) - {}^*b(U(\sigma),U(\sigma),U(\sigma)) + (U(\sigma),F(\sigma)) \\
&= -2\nu \|U(\sigma)\|^2 + (U(\sigma),F(\sigma)) \\
&\le -2\nu \|U(\sigma)\|^2 + \nu\|U(\sigma)\|^2 + \frac{1}{\nu}|F(\sigma)|_{-1}^2
\end{aligned}$$

and integrating from 0 to τ we get (4).

Step 2 - construction of $u(t)$.

Since the right hand side of (4) is finite for finite τ, $|U(\tau)|$ is bounded by the finite constant K for finite τ, and so are $U_k(\tau)$. We shall show that for finite k the function $U_k(\tau)$ is S-continuous for finite τ. For, let $\tau \approx \sigma$ and integrate (3) to obtain

$$\begin{aligned}
U_k(\tau) - U_k(\sigma) &= \int_\sigma^\tau -\nu\lambda_k U_k(\rho)\mathrm{d}\rho + \int_\sigma^\tau -{}^*b(U(\rho),U(\rho),E_k)\mathrm{d}\rho \\
&\quad + \int_\sigma^\tau F_k(\rho)\mathrm{d}\rho.
\end{aligned}$$

The first integral is infinitesimal because λ_k is finite and $U_k(\tau)$ is bounded. The integrand in the third term is square integrable, so by Lindstrøm's Lemma it is S-integrable, and so the integral is infinitesimal. The same argument applies to the second term after noticing that

$${}^*b(U(\rho),U(\rho),E_k) \le |U(\rho)|\,\|U(\rho)\|\,|AE_k| \le c\lambda_k\|U(\rho)\|$$

and the bound is square integrable by (4).

Therefore, for finite k we can define a standard function

$$u_k(t) = {}^\circ U_k(t)$$

which is continuous. We write

$$u(t) = \sum_{k=1}^{\infty} u_k(t)e_k.$$

Step 3 - $u(t)$ is a solution.

We shall show that $u(t)$ is a weak solution to the Navier–Stokes equation satisfying all requirements of the theorem.

First we show that u is H-valued. For any n

$$\sum_{k=1}^{n} u_k^2(t) = {}^\circ \sum_{k=1}^{n} U_k^2(t) \leq {}^\circ|U(t)|^2 \leq K < \infty$$

and letting $n \to \infty$ we find that $|u(t)|$ is finite.

Second, we note that (4) and the construction of u imply that

$$u \in L^\infty(0,T;H) \cap L^2(0,T;V) \cap C(0,T;H_{\text{weak}}).$$

Third, we show that u satisfies (2) for $v = e_k$. This is sufficient since the general case can be obtained by exploiting the linearity and continuity of all the terms in (2) with respect to v.

To this end we show that the standard parts of the terms of (3) written in the integral form:

$$U_k(t) - (U_0)_k = \int_0^t [-\nu\lambda_k U_k(\tau) - {}^*b(U(\tau), U(\tau), E_k) + F_k(\tau)]\mathrm{d}\tau$$

give the corresponding terms of (2).

1. By construction ${}^\circ U_k(t) = u_k(t) = (u(t), e_k)$.
2. By the definition of U_0, ${}^\circ(U_0)_k = (u_0, e_k)$.
3. Since U_k is S-continuous, it is S-integrable and

$${}^\circ\!\!\int_0^t \lambda_k U_k(\tau)\mathrm{d}\tau = \int_0^t \lambda_k {}^\circ U_k(\tau)\mathrm{d}_{\mathrm{L}}\tau = \int_0^t \lambda_k(u(s), e_k)\mathrm{d}s = \int_0^t ((u(s), e_k))\mathrm{d}s.$$

4. Before we discuss the term involving the form b, we note that by S-integrability of F_k, which follows from square integrability of f, we have

$${}^\circ\!\!\int_0^t ({}^*f(\tau), {}^*e_k)\mathrm{d}\tau = \int_0^t (f(s), e_k)\mathrm{d}s.$$

5. Finally, ${}^*b(U(\tau), U(\tau), {}^*e_k)$ is S-integrable as noted earlier, hence

$${}^\circ\!\!\int_0^t {}^*b(U(\tau), U(\tau), {}^*e_k)\mathrm{d}\tau = \int_0^t {}^{\circ *}b(U(\tau), U(\tau), {}^*e_k)\mathrm{d}_{\mathrm{L}}\tau.$$

On the other hand,

$$\begin{aligned}\int_0^t b(u(s), u(s), e_k)\mathrm{d}s &= \int_0^t b(u({}^\circ\tau), u({}^\circ\tau), e_k)\mathrm{d}_{\mathrm{L}}\tau \\ &= \int_0^t b({}^\circ U(\tau), {}^\circ U(\tau), e_k)\mathrm{d}_{\mathrm{L}}\tau\end{aligned}$$

so it remains to show that

$$^*b(U(\tau), U(\tau), {}^*e_k) \approx b({}^\circ U(\tau), {}^\circ U(\tau), e_k) \quad \text{for a.a. } \tau,$$

where ${}^\circ U(\tau)$ denotes the H-valued function constructed coordinate wise as above.

Fix τ such that $\|U(\tau)\|$ is finite (this is true a.e.) and for simplicity of notation write $U = U(\tau)$, $u = {}^\circ U(\tau)$. Note that $\|u\|$ is also finite since

$$\sum_{k=1}^{n} \lambda_k u_k^2 = {}^\circ \sum_{k=1}^{n} \lambda_k U_k^2 \leq {}^\circ\|U\|^2$$

and letting $n \to \infty$ we get $\|u\| \leq {}^\circ\|U\| < \infty$.

We estimate the difference:

$$\begin{aligned} |^*b(U,U,E_k) - {}^*b({}^*u, {}^*u, E_k)| &\leq |^*b(U,U,E_k) - {}^*b({}^*u, U, E_k)| \\ &\quad + |^*b({}^*u, U, E_k) - {}^*b({}^*u, {}^*u, E_k)| \\ &= |^*b(U - {}^*u, U, E_k)| \\ &\quad + |^*b({}^*u, U - {}^*u, E_k)| \\ &\leq c(\|{}^*u\| + \|U\|)|U - {}^*u|\lambda_k, \end{aligned}$$

which follows from the estimates on b.

It remains to show that

$$|U - {}^*u| \approx 0.$$

Note that for each finite n,

$$\sum_{k=1}^{n} (U_k - ({}^*u)_k)^2 \approx 0$$

by the construction of u, hence

$$\begin{aligned} |U - {}^*u|^2 &= \sum_{k=1}^{n} (U_k - ({}^*u)_k)^2 + \sum_{k=n+1}^{N} (U_k - ({}^*u)_k)^2 \\ &\approx \sum_{k=n+1}^{N} (U_k - ({}^*u)_k)^2. \end{aligned}$$

By Robinson's Sequential Lemma there is an infinite $M \leq N$ such that

$$|U - {}^*u|^2 \approx \sum_{k=M}^{N} (U_k - ({}^*u)_k)^2.$$

The final estimation shows that the right hand side is infinitesimal:

$$\begin{aligned}\sum_{k=M}^{N}(U_k-({}^*u)_k)^2 &= \sum_{k=M}^{N}\frac{1}{\lambda_k}\lambda_k(U_k-({}^*u)_k)^2\\ &\le \frac{1}{\lambda_M}\sum_{k=M}^{N}\lambda_k(U_k-({}^*u)_k)^2\\ &\le \frac{1}{\lambda_M}\|U-{}^*u\|^2\end{aligned}$$

where we have employed the monotonicity of the sequence λ_k. For infinite M, λ_M is infinite and so the right hand side is infinitesimal. □

2.3. UNIQUENESS PROBLEM

We first show that any weak solution is the standard part of a solution of (3) with some perturbation of the initial data and the force f that is infinitesimal in a certain topology depending on the space dimension $n = 2, 3$.

The perturbed Galerkin equation that we consider is

$$\begin{cases} \dot{U}_k(\tau) = -\nu\lambda_k U_k(\tau) - {}^*b(U(\tau), U(\tau), E_k) + F_k(\tau) + G_k(\tau)\,, \\ U(0) = U_0 + Z, \end{cases} \tag{5}$$

where $Z \in H_N$ with $|Z| \approx 0$, and $G : {}^*[0,T] \longrightarrow H_N$ is such that $\int_0^T |G(\tau)|^2_{V'} \mathrm{d}\tau < \infty$, and

$$\int_0^T (G(\tau), E_k)\mathrm{d}\tau \approx 0 \quad \text{for finite } k.$$

It is easy to see that the standard part of the solution to (5) gives a weak solution to the Navier–Stokes equations. The proof is the same as in the case $Z = 0$, $G = 0$.

Two following standard function classes contain all weak solutions.

$$\begin{aligned} \mathcal{A}_2 = \Big\{u \;\; &: \;\; u = {}^\circ U,\ U \text{ solves (5) for some } Z,\ G \text{ with} \\ & |Z| \approx 0,\ \int_0^T |G(\tau)|^2_{V'}\mathrm{d}\tau \approx 0\Big\} \\ \mathcal{A}_3 = \Big\{u \;\; &: \;\; u = {}^\circ U,\ U \text{ solves (5) for some } Z,\ G \text{ with} \\ & |Z| \approx 0,\ \int_0^T |G(\tau)|^2_{-s}\mathrm{d}\tau \approx 0 \text{ with } s > \tfrac{3}{2}\Big\} \end{aligned}$$

where $°U$ is the vector with coordinates $\mathrm{st}U_k$ for finite k (as in the proof of Theorem 2.2).

Theorem 2.3 *The set of all weak solutions of the Navier–Stokes equations in dimension n is contained in $\mathcal{A}_n$, $n = 2, 3$, with $Z = 0$.*

Proof If u is a weak solution then $^*u$ solves the *Navier–Stokes equations. Put

$$U(\tau) = \mathrm{Pr}_{H_H}{}^*u(\tau).$$

It can be shown that U satisfies (5) with

$$G(\sigma)_k = -{}^*b({}^*u(\sigma), {}^*u(\sigma), E_k) + {}^*b(U(\sigma), U(\sigma), E_k).$$

Using some inequalities for b one can show that the integrals of G appearing in the definition of the sets $\mathcal{A}_i$ are infinitesimal.

Next, following the lines of the classical uniqueness proof we can show that for $n = 2$ the class $\mathcal{A}_2$ is a one-element set. For dimension $n = 3$ we cannot prove this for $\mathcal{A}_3$. However for any fixed T we can select a subclass of $\mathcal{A}_3$ with the uniqueness property: we write

$$\mathcal{B}_3^{\alpha T} = \Big\{u \quad : \quad u = °U,\ U \text{ solves (5) for some } Z,\, G \text{ with } |Z| < \alpha,\ \int_0^T |G(\tau)|_{V'}^2 \mathrm{d}\tau \approx 0\Big\}.$$

Then for each T there exists an α such that $\mathcal{B}_3^{\alpha T}$ is a one element set. This indicates some possible directions of further investigation. To prove uniqueness in dimension 3 one has to obtain some stability results for ODEs in H_N. On the other hand, since all weak solutions are contained in $\mathcal{A}_3$, seeking a counterexample would reduce to finding infinitely close data for the equation in H_N such that the corresponding solutions are not infinitely close.

3. Statistical solutions

We introduce the notion of statistical solution and then show how it can be constructed by nonstandard methods.

3.1. FOIAS EQUATION

We first explain the basic idea for an abstract evolution equation

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = F(u(t)), \quad t > 0. \tag{6}$$

Suppose that for each initial value v there is a unique solution $u(t)$ with $u(0) = v$, in a certain Hilbert space H. Denote this solution by $u(t) = S(t, v)$ to emphasize the dependence on the initial function.

Suppose now that the initial value is a random variable $v : \Omega \to H$. This random variable induces a probability measure μ_0 on H by

$$\mu_0(A) = \mathrm{Prob}(v \in A).$$

The function $S(t, v)$ is then a stochastic process with initial distribution μ_0. The probability distributions μ_t of the random variables $S(t, \cdot)$ are the measures on H given by

$$\mu_t(A) = \mathrm{Prob}(S(t, v) \in A) = \mu_0(\{v : S(t, v) \in A\}).$$

To describe a measure it is sufficient to characterize the integrals

$$\int_H \theta(u)\mathrm{d}\mu_t(u). \tag{7}$$

for a sufficiently broad class of functions θ.

Computing heuristically the time derivative of (7) we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int_H \theta(u)\mathrm{d}\mu_t(u) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_H \theta(S(t,v))\mathrm{d}\mu_0(v) \quad \text{definition of } \mu_t \\
&= \int_H (\theta'(S(t,v)), \frac{\mathrm{d}}{\mathrm{d}t}S(t,v))_H \mathrm{d}\mu_0(v) \quad \text{assuming } \theta' \in H \\
&= \int_H (\theta'(S(t,v)), F(S(t,v)))_H \mathrm{d}\mu_0(v) \quad \text{from (6)} \\
&= \int_H (\theta'(u), F(u))_H \mathrm{d}\mu_t(u) \quad \text{definition of } \mu_t \text{ again.}
\end{aligned}$$

After integrating from 0 to t we obtain the so-called Foias equation

$$\int_H \theta(u)\mathrm{d}\mu_t(u) - \int_H \theta(u)\mathrm{d}\mu_0(u) = \int_0^t \int_H (\theta'(u), F(u))_H \mathrm{d}\mu_s(u)\mathrm{d}s. \tag{8}$$

Any solution $(\mu_t)_{t\geq 0}$ to (8) is called a statistical solution to the equation (6).

The derivation above requires that (6) have the uniqueness property, and as we know this is not known for the Navier–Stokes equations in dimension 3. The crucial point, observed by Foias, is that S does not occur in the Foias equation so the final equation makes sense even when the underlying equation does not have a unique solution.

The Foias equation for the Navier–Stokes equations takes the form

$$\int_H \theta(u)\mathrm{d}\mu_t(u) = \int_H \theta(u)\mathrm{d}\mu(u)$$
$$+ \int_0^t \int_H [-\nu((u,\theta'(u))) - b(u,u,\theta'(u)) + (f(s),\theta'(u))]\mathrm{d}\mu_s(u)\mathrm{d}s \quad (9)$$

3.2. MEASURES BY NONSTANDARD DENSITIES

Measures on infinite dimensional spaces cannot be described by densities due to the lack of Lebesgue measure. However, since the hyperfinite space H_N carries the nonstandard Lebesgue measure as it can be identified with ${}^*\mathbb{R}^N$, densities can be introduced on it.

Definition 3.1 An internal function $\Phi : H_N \to {}^*\mathbb{R}$ is a *nonstandard density* of the probability measure μ on a Hilbert space H if Φ is non-negative, *integrable with respect to *Lebesgue measure on ${}^*\mathbb{R}^N$, with $\int \Phi(U)\mathrm{d}U = 1$, and

$$\mu(B) = M_{\mathrm{L}}(\mathrm{st}_H^{-1}(B) \cap H_N),$$

where M_{L} is the Loeb measure corresponding to the internal measure M on H_N given by

$$M(A) = \int_A \Phi(U)\mathrm{d}U.$$

Let $\mathcal{N}(X,C)$ denote the normal density on ${}^*\mathbb{R}^N$ with mean X and covariance C.

Proposition 3.2 *The function* Φ

$$\Phi(U) = \int_{{}^*H} \mathcal{N}(U - \mathrm{Pr}_N v, \varepsilon^2 \cdot I)\,\mathrm{d}\,{}^*\mu(v),$$

is a nonstandard density of μ*. Moreover*

$$\int |U|^2 \mathrm{d}M(U) \leq \int |u|^2\,\mathrm{d}\mu(u) + N\varepsilon^2.$$

3.3. CONSTRUCTION OF STATISTICAL SOLUTION

We first heuristically derive an equation for the densities of solutions to the Foias equation (9).

Suppose that $\Phi(t,U)$ is a density of μ_t so that for a test functional θ we have

$$\int \theta(u)\,\mathrm{d}\mu_t(u) \approx \int {}^*\theta(U)\,\Phi(t,U)\,\mathrm{d}U.$$

Then it is natural to rewrite the Foias equation replacing θ' by the vector $\left(\frac{\partial\theta}{\partial U_k}\right)$, $\mathrm{d}\mu_t$ by $\Phi\mathrm{d}U$ etc. Then after integration by parts and dropping ${}^*\theta$ from both sides we obtain

$$\frac{\partial}{\partial\tau}\Phi(\tau,U)+\sum_{k=1}^{N}\frac{\partial}{\partial U_k}\Big[\,(-\nu\lambda_k U_k - {}^*b(U,U,E_k)+F_k(\tau)\,)\cdot\Phi(\tau,U)\,\Big] = 0 \quad (10)$$

which is now a nonstandard equation with $\tau \in {}^*[0,T]$ and $U \in H_N$; we call it the *density equation.* It is in fact a hyperfinite version of the Liouville equation.

Theorem 3.3 *Let μ be a Borel probability measure on H satisfying*

$$\int |u|^2 \mathrm{d}\mu(u) < \infty.$$

Let Φ_0 be a nonstandard density of μ with

$$\int |U|^2 \Phi_0(U)\mathrm{d}U < \infty$$

Let $\Phi(\tau,U)$ be the solution to the density equation (10) *with initial function Φ_0. Then the internal measures M_τ determined by $\Phi(\tau)$ are nearstandardly concentrated and the standard family of measures given by*

$$\mu_t = (M_t)_{\mathrm{L}} \circ \mathrm{st}^{-1}, \quad t \in [0,T],$$

is a statistical solution of the Navier–Stokes equation with initial measure μ.

Proof The idea of the proof is as follows. Take a nonstandard density Φ_0 of the initial measure μ. Proposition 3.2 shows that if $\int |u|^2 \mathrm{d}\mu(u) < \infty$ we may take Φ_0 so that $|U|^2$ is S-integrable with respect to the measure

$$M_0(A) = \int_A \Phi_0(U)\mathrm{d}U.$$

Next, solve (10) internally using the method of characteristics with initial function $\Phi(0,U) = \Phi_0(U)$. Then show that the internal measures corresponding to the internal densities $\Phi(\tau,U)$ are nearstandardly concentrated, and finally show that the corresponding standard measures give a statistical solution.

4. Stochastic equations

In this section we give an existence theorem for a fairly general system of stochastic Navier–Stokes equations (1).

The set $K_m = \{v : \|v\| \leq m\} \subseteq V$ is considered with the strong topology of H. Note that continuity on each K_m is weaker than continuity on V in either the H-norm or the weak topology of V.

Theorem 4.1 *Suppose that* $u_0 \in H$ *and*

$$f : [0,\infty) \times V \to V', \qquad g : [0,\infty) \times V \to L(H,H)$$

are jointly measurable functions with the following properties

(i) $f(t,\cdot) \in C(K_m, V'_{\text{weak}})$ *for all* m,
(ii) $g(t,\cdot) \in C(K_m, L(H,H)_{\text{weak}})$ *for all* m,
(iii) $|f(t,u)|_{V'} + |g(t,u)|_{H,H} \leq a(t)(1+|u|)$ *where* $a \in L^2(0,T)$ *for all* T.

Then the equation

$$\begin{aligned}(u(t),v) - (u_0,v) \quad &= \quad \int_0^t [-\nu((u(s),v)) - b(u(s),u(s),v) + (f(s),v)]\mathrm{d}s \\ &\quad + \int_0^t (g(s,u(s)),v)\mathrm{d}w(s)\end{aligned}$$

has a solution u *on a filtered Loeb space, where the stochastic integral is understood in the sense of [3] for Wiener process with nuclear covariance.*

Proof The general idea of the proof is the same as in the proof of Theorem 2.2. The Galerkin approximation is a system of stochastic differential equations in the space H_N:

$$\begin{cases} \mathrm{d}U(\tau) = [-\nu^* AU(\tau) - \overline{B}(U(\tau),U(\tau)) + F(\tau,U(\tau))]\mathrm{d}\tau \\ \qquad\qquad +G(\tau,U(\tau))\mathrm{d}W(\tau), \\ U(0) = \mathrm{Pr}_N{}^* u_0. \end{cases}$$

This equation can be solved by transfer. The crucial property of the solution is the following stochastic version of the energy inequality (4)

$$E\left(\sup_{\sigma\leq\tau} |U(\sigma)|^2 + \int_0^\tau \|U(\sigma)\|^2 \mathrm{d}\sigma\right) < \infty$$

which is derived by applying Itô's lemma to $|U(\tau)|^2$. The standard process is defined by taking standard parts of the coordinates of $U(\tau)$ as in the deterministic case (see [2] for details).

5. Some open problems

The basic open problem is the uniqueness of solutions of Navier–Stokes equation for dimension 3. Since it has been open for a long time despite many investigations, it can be qualified as one of the *impossible* problems. The problems given below seem to be tractable.

1. Extend the nonstandard approach to the problem of uniqueness presented in Section 2.3 to the case of stochastic Navier–Stokes equations.
2. Can any statistical solution to Navier–Stokes be given in terms of nonstandard densities satisfying the density equation with some perturbation.
3. We say that intrinsic turbulence takes place if there is a statistical solution with Dirac initial measure such that for some positive time μ_t is not Dirac. Suppose that in the class of statistical solutions constructed by nonstandard densities we have intrinsic turbulence. Does this imply nonuniqueness of the underlying equation.
4. Prove uniqueness of statistical solutions for stochastic Navier–Stokes equations for dimension $n = 2$.

References

1. Capiński, M. and Cutland, N.J (1992) A simple proof of existence of weak and statistical solutions of Navier–Stokes equations, *Proceedings of the Royal Society, London, Ser.A*, **436**, pp. 1–11.
2. Capiński, M. and Cutland, N.J. (1995) *Nonstandard Methods for Stochastic Hydromechanics,* World Scientific, Singapore.
3. Ichikawa, A. (1982) Stability of semilinear stochastic evolution equations, *Journal of Mathematical Analysis and Applications* **90**, pp. 12–44.

INDEX