F.C. Claassens

Non-standard Analysis

Master Thesis

Supervisors: Dr. E. Coplakova & Dr. M.F.E. de Jeu

Date Master Exam: 14 January 2016

Mathematisch Instituut, Universiteit Leiden

Contents

1 Introduction

The great Gauss warned mathematicians to stay away from infinity. For years most mathematicians heeded his warning until Georg Cantor, at the end of the 19th century, invented set theory. Creating more than one kind of infinity with the cardinal and ordinal numbers. This paved the way for more study of infinity. In the field of analysis mathematicians were less interested in Cantors rather abstract cardinal numbers. Their interest in infinity mainly was focused around infinitesimals, the infinitely small numbers. The concept of infinitesimals had already been introduced in mathematics and used by mathematicians like Newton and Leibniz, however they never were able to develop a rigorous foundation, so they were discarded in favour of the more well-founded limit approach. That was, until the late 1960s when Abraham Robinson developed non-standard analysis.

Non-standard analysis is a powerful tool. Though it is not necessary to use it, it can bring great elegance to a proof. In this thesis we will look at three non-standard proofs of big theorems in Functional Analysis. The Theorem of Hahn-Banach, the Theorem of Hille-Yosida and the Theorem of Bernstein-Robinson, a case of the invariant subspace problem, which was initially proved using non-standard methods.

We will start in Section 2 with a short introduction to logic. We will assume that the reader has know knowledge of the subject so those who have followed a basic course on the subject should be already familiar with the material. We will finish by defining our non-standard model, using filters, and then expand our model so we can 'quantify sets'. In the remaining three sections we focus on our main results. In Section 4 we will explore non-standard Functional analysis and end by proving the Theorem of Bernstein-Robinson. In section 5 we look at a more abstract proof technique which we will use to proof the Theorem of Hahn-Banach. We will end in Section 6 by taking a look at semigroups and giving a non-standard proof for the Theorem of Hille-Yosida.

2 Logic

2.1 An introduction to logic

Non-standard analysis is a product of logic, hence we will start giving a short introduction to logic. We will start with some very basic definitions. One of the main goals of logic is giving an unambiguous language in which we can talk about mathematics. This so-called logical language will always contain the following logical symbols:

$\neg$	not
$\wedge$	and
$\vee$	or
$\rightarrow$	implies
$\leftrightarrow$	is equivalent to
$\forall$	universal quantifier
$\exists$	existential quantifier
$x_1, x_2, ...$	variables
(	left bracket
)	right bracket
,	comma

Besides the logical symbols there are constants, function symbols and predicates which can be more freely interpreted to pose theorems. Here constants and functions fulfill the role you would expect and predicates are statements.

Example 2.1. The statement 'Every mathematician thinks logically' can be formulated logically by $(\forall x)(A(x) \rightarrow B(x))$ where $A(x)$ means 'x is a mathematician.' and $B(x)$ means 'x thinks logically.'. In this case A and B are both predicates.
The statement 'Function f is continuous at c.' can be expressed in the logical language by:

$$(\forall \epsilon)(A(0,\epsilon) \rightarrow (\exists \delta)(A(0,\delta) \wedge (\forall x)(A(f_1(f_2(c,x)),\delta) \rightarrow A(f_1(f_2(f(c),f(x))),\epsilon))))$$

where $A(x_1, x_2)$ means '$x_1 < x_2$', c is a constant and the function symbols are interpreted as $f_1(x) = |x|$ and $f_2(x_1, x_2) = x_1 - x_2$. As you might expect we will usually denote functions in a more recognisable way to avoid horrible looking formulas as above.

Definition 2.2. A *language* $\mathcal{L}$ is a set of symbols containing:

(i) The logical symbols.

(ii) A set of constants $c_1, c_2, ...$.

(iii) A set of functions $f_1^{n_1}, f_2^{n_2}, ...$ where n_i is a positive integer describing the number of variables of the function.

(iv) A non-empty set of predicates $P_1^{n_1}, P_2^{n_2}, ...$ where n_i is a positive integer describing the number of variables of the predicate.

Since the logical symbols are always part of a language we will omit them when describing a language. So a language containing just the logical symbols we will denote as $\mathcal{L} = \emptyset$.

With a language we can now write down *formulas* which are a combination of symbols from the language in a correct order. We will not define what a correct order of symbols is, but assume that the reader intuitively knows what this means. You can find out more about this in [6].

We now have a language to talk about mathematics. But as in a normal language the words are useless without an interpretation, in a logical language the symbols are senseless without an interpretation as well. An interpretation of the language is given in so called structures.

Definition 2.3. Let $\mathcal{L}$ be a language. An $\mathcal{L}$-*structure* $\mathcal{M}$ is a non-empty set M called the *domain* and an interpretation of all symbols from $\mathcal{L}$. That is:

(i) To every predicate $P_i^{n_i}$ we associate a set $M_i \subset M^{n_i}$. (We will interpret $P_i^{n_i}(v_1, ..., v_{n_i})$ as 'true' if $(v_1, ..., v_{n_i}) \in M_i$.)

(ii) To every function symbol $f_i^{n_i}$ we associate a map from M^{n_i} to M.

(iii) To every constant c_i we associate an element $v_i \in M$.

Example 2.4. Consider the language $\mathcal{L} = \{c_1, c_2, f_1^2, f_2^2, P_1^2\}$. Then $M = \mathbb{N}$ with the interpretation $c_1 = 0$, $c_2 = 1$, $f_1^2(x_1, x_2) = x_1 + x_2$,$f_2^2(x_1, x_2) = x_1 \cdot x_2$ and predicate $P_1^2(x_1, x_2)$ means $x_1 < x_2$, i.e. $M_1 = \{(x_1, x_2) \in \mathbb{N}^2 : x_1 < x_2\}$ is an $\mathcal{L}$-structure

If we want to use a specific language we usually use the standard mathematical symbols. In this case we would write: $\mathcal{L} = \{0, 1, +, \cdot, <\}$.

If a formula B is true in $\mathcal{M}$ we say that $\mathcal{M}$ *satisfies* B. For short we write $\mathcal{M} \models B$.

We can now talk about mathematical structures, but they can be quite wild. If you want to, you can use '+' as a function symbol and interpret it as multiplication. To gain more control over our structures we want to make rules for how the function symbols, constants and predicates should behave. These rules are the axioms and a set of axioms will be called a theory.

Definition 2.5. Let $\mathcal{L}$ be a language. A non-empty set of formulas from $\mathcal{L}$ is called a *theory.*

Definition 2.6. Let T be a theory. An $\mathcal{L}$-structure $\mathcal{M}$ is a *model* of T if $\mathcal{M} \models \mathfrak{B}$ holds for each formula $\mathfrak{B} \in T$. We write $\mathcal{M} \models T$.

Example 2.7. We will now give a short example to get a better idea of theories, models and $\mathcal{L}$-structures and the differences between them.

Consider the language $\mathcal{L} = \{0, 1, +, \cdot, =\}$. We will take as theory T the group axioms:

$$(\forall x)(0 + x = x \wedge x + 0 = x)$$
$$(\forall x)(\forall y)(\forall z)(x + (y + z) = (x + y) + z)$$
$$(\forall x)(\exists y)(x + y = 0 \wedge y + x = 0).$$

An $\mathcal{L}$-structure in this case is for example $\mathbb{N}$ with the usual interpretation. This does not satisfy T due to the lack of inverses.
$\mathbb{Z}$ or $\mathbb{Q}$ with the usual interpretation however are models of T.
Of course we know that $\mathbb{Q}$ is a field as well, (hence the inclusion of '$\cdot$' and '1'

in our language), so sometimes a model satisfies formulas, like the existence of multiplicative inverses, which are not consequence of the theory. On the other hand there are formulas, like the uniqueness of 0, which are logical consequences of a theory, but which are not part of the theory. That is why it is important to make a clear difference between what is part of the theory, what are consequences of the theory and what is true in a specific model of the theory.

We have now covered the basic logic that we need. This allows us to focus on some more specific concepts of logic which we will need for non-standard analysis. In particular closed formulas which will become very important. Before we can define them however, we will first take a look at free variables.

Definition 2.8. Let $\mathfrak{B}$ be a formula. In $(\forall x)(\mathfrak{B})$ and $(\exists x)(\mathfrak{B})$, we call $\mathfrak{B}$ the *range of the quantifier.*

Definition 2.9. An occurrence of a variable x is called *bound* if it lies inside the range of a universal or existential quantifier. If x is not bound it is called *free.*

To really understand this definition it is important to understand that, while it is advisable to write down formulas unambiguously, it is not necessary. The formula $(\forall x)(x = y) \wedge (\exists y)((y + x < z) \rightarrow (\exists x)(z = y))$ is legal since the range of the universal quantifier is restricted to $(x = y)$ and thus does not create a conflict with the existential quantifier later. In this example x is bound in its first two occurrences and its last occurrence, but free in its third and y is free the first time and bound in the rest. Here z is free in all occurrences.

Definition 2.10. If in a formula all variables are bound we call it *closed* or we call it a *sentence.*

Example 2.11. The following formulas are closed:

$$(\forall x)(\forall y)(x < y \vee y < x \vee x = y).$$

For c a constant:

$$(\forall \epsilon)(\epsilon > 0 \rightarrow (\exists \delta)(\delta > 0 \wedge (|c - x| < \delta \rightarrow |f(c) - f(x)| < \epsilon))).$$

For f a function:

$$(\exists c)(\forall \epsilon)(\epsilon > 0 \rightarrow (\exists R)(R > 0 \wedge (\forall r)(R < r \rightarrow |f(r) - f(c)| < \epsilon))).$$

To really understand closed formulas it might be useful to look at some formulas with free variables. A well known and relevant example is the theory

$$T = T_{\mathbb{R}} \cup \{v > n : n \in \mathbb{N}\}$$

where $T_{\mathbb{R}}$ is the theory of the real numbers [6]. Since in every formula we used the same free variable v, any model $\mathcal{M}$ of T must contain an $x \in M$ such that $m > n$ for all $n \in \mathbb{N}$. Hence a model of T would be suitable for non-standard analysis. That such a model exists is a consequence of the Compactness Theorem [6], a well known theorem in logics.

2.2 Filters

Though we could use the Compactness Theorem to prove there exists models with infinitely large numbers and we could just pick one and work with that, it is a rather abstract approach. There is another more constructive way to construct infinitely large numbers using filters which is the way we will explore.
For this section κ is a cardinal number and J is a set of cardinality κ.

Definition 2.12. We call $\mathcal{F} \subset \mathcal{P}(J) = \{X : X \subset J\}$ a *filter* on J if the following hold:

(i) If $A \in \mathcal{F}$ and $A \subset B \subset J$, then $B \in \mathcal{F}$.

(ii) If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

(iii) $\emptyset \notin \mathcal{F}$ and $J \in \mathcal{F}$.

Furthermore we call $\mathcal{F}$ *principal* if there exists a set $A \subset J$ such that $\mathcal{F} = \{X \subset J : A \subset X\}$, an *ultrafilter* if for all $A \subset J$ either $A \in \mathcal{F}$ or $A^c \in \mathcal{F}$ and *maximal* if there exists no filter $\mathcal{F}_0 \supset \mathcal{F}$.

We will first look at some examples of filters and equivalent definition.

Example 2.13. **(i)** $\mathcal{F} = \{J\}$ is called the trivial filter.

(ii) $\mathcal{F}_A = \{X \subset J : A \subset X\}$, $A \subset J$.

(iii) $\mathcal{F} = \{X \subset \mathbb{N} : X^c \text{ is finite }\}$ is called the Fréchet filter.

(iv) $\mathcal{F} = \{X \subset J : |X^c| < \kappa\}$ which we will call the generalised Fréchet filter.

Note that the generalised Fréchet filter, and in particular the Fréchet filter both are non-principal.

Proposition 2.14. *Let $\mathcal{F}$ be a filter of J. Equivalent are:*

(i) *$\mathcal{F}$ is an ultrafilter.*

(ii) *$\mathcal{F}$ is maximal.*

Proposition 2.15. *Every filter can be extended to an ultrafilter.*

Proposition 2.16. *Every principal ultrafilter is of the form $\mathcal{F} = \{X \subset J : j \in X\}$ for some $j \in J$.*

Corollary 2.17. *An ultrafilter containing the generalised Fréchet filter is non-principal.*

We will leave these results as exercises for the reader where you can prove the second proposition by using Zorn's lemma.

Using filters we will now define a structure, which will be seen as non-standard. This we can do by defining an equivalence relation on M^J.

Definition 2.18. Let $\mathcal{M}$ be an $\mathcal{L}$-structure with domain M and let $\mathcal{F}$ be a filter on J. We define on M^J the relation:

$$a \sim b \Leftrightarrow \{j \in J : a(j) = b(j)\} \in \mathcal{F}.$$

Lemma 2.19. *$\sim$ is an equivalence relation.*

Proof. Since $J \in \mathcal{F}$ we find that $\sim$ is reflexive. Since equality is symmetric $\sim$ clearly is symmetric. To see that $\sim$ is transitive consider $a, b, c \in M^J$ and note that:

$$\{j \in J : a(j) = b(j)\} \cap \{j \in J : b(j) = c(j)\} \subset \{j \in J : a(j) = c(j)\}.$$

Hence if $a \sim b$ and $b \sim c$ then by definition of filters

$$\{j \in J : a(j) = b(j)\} \cap \{j \in J : b(j) = c(j)\} \in \mathcal{F}$$

and then again by definition of filters $\{j \in J : a(j) = c(j)\} \in \mathcal{F}$, thus $a \sim c$. $\square$

We now have a new set namely $M^J/\sim$ which we will denote by M^*. The equivalence classes we will denote by $[(a_j)_{j\in J}]$. Of course we want this to be the domain of some $\mathcal{L}$-structure $\mathcal{M}^*$. For this we must give an interpretation to all symbols in our language.

Definition 2.20. Let $\mathcal{M}$ an $\mathcal{L}$-structure with domain M and let $\mathcal{F}$ a filter on some set J. We call $\mathcal{M}^*$ the *ultrapower* of $\mathcal{M}$ if the domain of $\mathcal{M}^*$ is M^* and we interpret the symbols of $\mathcal{L}$ as follows:

(i) If $c \in \mathcal{L}$ is a constant, then
$$c^{\mathcal{M}^*} := [(c^{\mathcal{M}})_{j\in J}].$$

(ii) If $f(v_1, ..., v_n) \in \mathcal{L}$ is a function symbol, then for all $[(x_{1,j})_{j\in J}], ..., [(x_{n,j})_{j\in J}] \in M^*$ we define

$$f^{\mathcal{M}^*}([(x_{1,j})_{j\in J}], ..., [(x_{n,j})_{j\in J}]) := [(f^{\mathcal{M}}(x_{1,j}, ..., x_{n,j}))_{j\in J}].$$

(iii) If $P(v_1, ..., v_n) \in \mathcal{L}$ is a predicate, then for all $[(x_{1,j})_{j\in J}], ..., [(x_{n,j})_{j\in J}] \in M^*$ we define

$$\mathcal{M}^* \models P([(x_{1,j})_{j\in J}], ..., [(x_{n,j})_{j\in J}]) \Leftrightarrow \{j \in J : \mathcal{M} \models P(x_{1,j}, ..., x_{n,j})\} \in \mathcal{F}.$$

Proposition 2.21. *The interpretations in Definition 2.20 are well-defined.*

Proof. Obviously the constants are well-defined, leaving us with the functions and the predicates.

Let $f(v_1, ..., v_m) \in \mathcal{L}$ a function and let $a_1, ..., a_m, b_1, ..., b_m \in M^J$ with $a_i \sim b_i$ for all $1 \leq i \leq m$.
We want:
$$(f(a_{1,j}, ..., a_{m,j}))^{\mathcal{M}})_{j\in J} \sim (f(b_{1,j}, ..., b_{m,j}))^{\mathcal{M}})_{j\in J}.$$

To see this, consider

$$\{j \in J : (f(a_{1,j}, ..., a_{m,j}))^{\mathcal{M}} = (f(b_{1,j}, ..., b_{m,j}))^{\mathcal{M}}\} \supset \bigcap_{i=1}^{m} \{j \in J : a_{i,j} = b_{i,j}\}.$$

Since by assumption $\{j \in J : a_{i,j} = b_{i,j}\} \in F$, we have that the finite intersection $\bigcap_{i=1}^{m} \{j \in J : a_{i,j} = b_{i,j}\} \in F$. Because filters are closed by taking supersets we have

$$(f(a_{1,j}, ..., a_{m,j}))^{\mathcal{M}})_{j \in J} \sim (f(b_{1,j}, ..., b_{m,j}))^{\mathcal{M}})_{j \in J}.$$

We can use similar arguments to prove the predicates are well-defined, so we leave that part of the proof for the reader □

We have now constructed an $\mathcal{L}$-structure, though without some extra restrictions it can be either a bit to wild or to boring. First observe that if $\mathcal{F}$ is principal that $\mathcal{M}$ and $\mathcal{M}^*$ are basically the same. So we want that $\mathcal{F}$ is non-principal, we gain a 'new' $\mathcal{L}$-structure. Secondly, suppose our language contains '$=$' and '$\neq$' as predicates interpreted in the usual way in some $\mathcal{L}$-structure $\mathcal{M}$, then in the ultrapower it could happen that for some $x, y \in M^*$ we have that $M^* \not\models x = y$ and $M^* \not\models x \neq y$. To avoid this we will require $\mathcal{F}$ to be an ultrafilter.

When we require $\mathcal{F}$ to be a non-principal ultrafilter, we come to the point where the closed formulas become very important. Because then every closed formula true in $\mathcal{M}$ is also true in $\mathcal{M}^*$ as well and vice versa. This is expressed in the main theorem of non-standard analysis, the Theorem of Łoś.

Theorem 2.22 (Łoś)**.** *Let $\mathfrak{B}$ be a closed formula, let $\mathcal{M}$ be a $\mathcal{L}$-structure, and let $\mathcal{M}^*$ be its ultrapower. Then*

$$\mathcal{M} \models \mathfrak{B} \Leftrightarrow \mathcal{M}^* \models \mathfrak{B}.$$

Officially this is not the theorem of Łośbut a direct consequence of it. The proof is more deeply rooted in logic, and hence will be omitted. More on this subject can be found in [6].

Since closed formulas are so important we want as many of them as possible. This can be done by expanding our language by adding every element of the original model as a constant. Then the statement "Function f is continuous at c" becomes a closed formula which will prove to be very useful.

2.3 Ultrapowers

We will apply non-standard analysis on superstructures which we will define in the next section. Before we do this we first will get familiar with ultrapowers by looking at a relatively simple structure: $\mathbb{C}$. Of course, before we can talk about $\mathbb{C}$ we need a language. We will use a language $\mathcal{L}$ containing

1. A constant a_c for every $c \in \mathbb{C}$.

2. A function f_ϕ for every n-ary operation ϕ on $\mathbb{C}$.

3. A predicate A_φ for every n-ary relation on $\mathbb{C}$.

One of the great advantages of non-standard analysis versus standard analysis are the infinitesimals. Infinitely small numbers are wonderful for everyone who is not a fan of epsilon/delta definitions. Using non-standard analysis we can give an equivalent easy definition of continuity. Before this is possible we first will check whether our ultrapower actually contains the promised infinitely large and small numbers.

Definition 2.23. Let $\mathcal{M}$ be an $\mathcal{L}$-structure, let $\mathcal{F}$ be a filter on a set J and let $\mathcal{M}^*$ be the ultrapower. We call

$$* : M \to M^*, x \mapsto x^* := [(x)_{j \in J}]$$

a *non-standard map*. For every $X \subset M$, we define the copy of X in M^* as $X^\# := \{x^* : x \in X\}$.

From now on, for convenience sake, we will assume that our filter $\mathcal{F}$ is an ultrafilter on $\mathbb{N}$ containing the Frèchet filter. This means that any subset of $\mathbb{N}$ with a finite complement is an element of $\mathcal{F}$ which makes it easier to understand the proofs. Note that it does not really matter what filter we use since if a closed formula is true in one ultrapower, by Łoś' Theorem it is true in the original model and thus, again by Łoś' Theorem true in every ultrapower.

We call the elements of $M^\#$ the standard elements of M^*. Please note that there is no predicate in $\mathcal{L}$ describing $M^\#$, which means that the definition of standard elements is not a definition in the logical language, but in a meta language. We can now say an element ω of $\mathbb{C}^*$ is infinitely large if $|\omega| > |c|$ for all $c \in \mathbb{C}^\#$. An example of an infinitely large number is $\omega = [(n)_{n \in \mathbb{N}}]$. Evidently for all $c \in \mathbb{C}^\#$ the complement of $\{n \in \mathbb{N} : n > |c|\}$ is finite, hence $\{n \in \mathbb{N} : n > |c|\} \in \mathcal{F}$, so ω is larger than c. Since $\mathcal{C}^*$ is also a field multiplicative inverses exists, which implies that there exist infinitely small numbers.

Definition 2.24. Let $x \in \mathbb{C}^*$, then we call x *infinitesimal* if for all $r \in \mathbb{R}^\#_{>0}$ we have $|x| < r$. We denote the set of infinitesimal complex numbers by $\mathbb{C}_0$ and the set of infinitesimal real numbers by $\mathbb{R}_0$.
We call x *finite* if there is a $c \in \mathbb{C}^\#$ and an infinitesimal $y \in \mathbb{C}_0$ such that $x = c + y$. We call $x^\circ := c$ the *standard part* of x.
Finally if $x, y \in \mathbb{C}^*$ and there is an infinitesimal $z \in \mathbb{C}_0$ such that $x = y + z$ then we write $x \approx y$.

Proposition 2.25. *Let y_1 and y_2 be infinitesimal and let c be finite. Then $y_1 + y_2$, $-y_1$, $|y_1|$ and $y_1 c$ are infinitesimal.*

Proof. Let $r \in \mathbb{R}^{\#}_{>0}$ and let $z \in \mathbb{C}_0$ be such that $c = c^\circ + z$. Then

$$\begin{aligned}
&|y_1 + y_2| \leq |y_1| + |y_2| < \frac{1}{2}r + \frac{1}{2}r = r, \\
&|-y_1| = |y_1| < r, \\
&|y_1 c| \leq |y_1 c^\circ| + |y_1||z| < \frac{r}{2|c^\circ|}|c^\circ| + \sqrt{\frac{r}{2}} \cdot \sqrt{\frac{r}{2}} = r.
\end{aligned}$$

□

Proposition 2.26. *The standard part is unique and every finite $c \in \mathbb{C}^*$ has standard part.*

Proof. Let c be such that there exist $c_1, c_2 \in \mathbb{C}^{\#}$ and $y_1, y_2 \in \mathbb{C}_0$ such that $c = c_1 + y_1 = c_2 + y_2$. Then

$$0 = |c_1 - c_2 - y_2 + y_1| \geq |c_1 - c_2| - |y_2 - y_1| \approx |c_1 - c_2| \geq 0.$$

Hence $c_1 - c_2 = 0$ proving that the standard part is unique.
Now let $x \in \mathbb{C}^*$ and $r \in \mathbb{R}^{\#}$ such that $|x| < r$. Consider $A_r = \{x \in \mathbb{R}^{\#} : x > \mathrm{Re}(c)\}$, $B_r = \{x \in \mathbb{R}^{\#} : x < \mathrm{Re}(c)\}$, $A_i = \{x \in \mathbb{R}^{\#} : x > \mathrm{Im}(c)\}$ and $B_i = \{x \in \mathbb{R}^{\#} : x < \mathrm{Im}(c)\}$.
Since $\mathbb{R}$ is linearly ordered we find $c_r := \inf A_r = \sup B_r$ and $c_i := \inf A_i = \sup B_i$.
Now suppose that there is a $r_1 \in \mathbb{R}^{\#}$ $|\mathrm{Re}(c) - c_r| > r_1 > 0$.
Then either $\mathrm{Re}(c) - c_r > r_1 > 0 \Leftrightarrow \mathrm{Re}(c) > r_1 + c_r > c_r$, thus $r_1 + c_r \in B_r$ and $r_1 + c_r > \sup B_r$, which contradicts the definition of B_r, or $\mathrm{Re}(c) - c_r < -r_1 < 0 \Leftrightarrow \mathrm{Re}(c) > c_r - r_1 > c_r$, thus $c_r - r_1 \in A_r$ and $c_r - r_1 > \inf A_r$, which contradicts the definition of A_r.
So $|\mathrm{Re}(c) - c_r|$ is infinitesimal. In the same way we find that $|i\mathrm{Im}(c) - ic_r|$ is infinitesimal. Thus $|c - c_r - ic_i|$ is infinitesimal and c has standard part. □

The infinitesimal numbers are a very useful tool to simplify a lot of definitions. There are a lot of interesting results for $\mathbb{C}^*$ but since our goal is Functional analysis we will just look at one example.

Proposition 2.27. *Let $f : \mathbb{C} \to \mathbb{C}$ be a function and let $c \in \mathbb{C}$. Then f is continuous at c if and only if for all $x \in \mathbb{C}^*$ with $x \approx c$ we have $f(x) \approx f(c)$.*

Proof. Suppose f is continuous at c, then for all $\epsilon \in \mathbb{R}_{>0}$ there exists a $\delta \in \mathbb{R}_{>0}$ such that

$$\mathcal{C} \models (\forall x)(|c - x| < \delta \to |f(x) - f(c)| < \epsilon).$$

Since ϵ, δ and c are constants, this formula is closed and so we may apply the theorem of Łośwhich gives us

$$\mathcal{C}^* \models (\forall x)(|c - x| < \delta \to |f(x) - f(c)| < \epsilon).$$

Since for all $\delta \in \mathbb{R}^{\#}_{>0}$ we have that $|x - c| < \delta$ for all $x \in \mathbb{C}^*$ with $x \approx c$, we have that for all $x \in \mathbb{C}^*$ with $x \approx c$ and all $\epsilon \in \mathbb{R}^{\#}_{>0}$ it holds that $|f(x) - f(c)| < \epsilon$. Thus

$f(x) \approx f(c)$.
Now suppose that for $x \in \mathbb{C}^*$ with $x \approx c$ we have that $f(x) \approx f(c)$. Let $\epsilon \in \mathbb{R}^{\#}_{>0}$ and let $\delta \in \mathbb{R}_0$ be positive. Then for all $|x - c| < \delta$ implies that $x \approx c$ and so $f(x) \approx f(c)$. This gives $|f(x) - f(c)| < \epsilon$, so

$$\mathcal{C}^* \models (\exists\delta)(\forall x)(|x - c| < \delta \to |f(x) - f(c)| < \epsilon).$$

Hence also

$$\mathcal{C} \models (\exists\delta)(\forall x)(|x - c| < \delta \to |f(x) - f(c)| < \epsilon).$$

Since this holds for all $\epsilon \in \mathbb{R}_{>0}$ we find that f is continuous at c. $\square$

Although looking at $\mathbb{C}^*$ can be very useful it does not cover everything we want. If we want to make a statement about sets such as 'Every subset of $\mathbb{N}$ has a smallest element.' we run into trouble since subsets of $\mathbb{C}^*$ are not elements of $\mathbb{C}^*$, so we can't quantify over them. To solve this problem we will expand our original $\mathcal{L}$-structure.

2.4 Quantifying sets

Consider a model $\mathcal{M}$ with domain M. If we want to quantify over sets of M, we can simply add them to our domain, that is: $\mathcal{P}(M) \cup M$, this might however not be enough. An ordered pair $(x, y) \in M^2$ is defined as $(x, y) = \{\{x\}, \{x, y\}\}$ which is not an element of $\mathcal{P}(M) \cup M$, so if we want to talk about ordered pairs we need more. Functions are sets of ordered pairs, and we might want to talk about functions of functions. To accommodate such a big range of possibilities we take the following construction.

Definition 2.28. Let M be a set. Let $M_0 := M$. We define inductively $M_n := \mathcal{P}(M_{n-1}) \cup M_0$ for all $n \in \mathbb{N}_{>0}$. We call

$$\hat{M} := \bigcup_{n \in \mathbb{N}} M_n$$

the *superstructure* of M.

Having created such a big set, we now need a language to talk about it. This poses certain problems since we can not include a function $+$ on $\mathbb{C}$ like we did when we only had to deal with $\mathbb{C}$ since the domain of an n-ary function in this case should be $\hat{\mathbb{C}}^n$. To solve this problem we use the following language:

$$\mathcal{L} = \{=, \in\} \cup \hat{\mathbb{C}}.$$

You might wonder why we include so few predicates and functions. This is because we can 'construct' predicates and functions by using the constants we added. If for example we want to use the inequality in $\mathbb{R}$ we can use the set

$$X_< := \{(x, y) \in \mathbb{R}^2 : x < y\} \in \hat{\mathbb{C}}.$$

This set, which we described using metalanguage, is a constant in our language, hence we can express "$0 < 1$" by the formula

$$(0,1) \in X_<.$$

"$0 < 1$" is not a very hard statement. Now we will give a formula for the more complicated statement "for all $x \in \mathbb{R}$ there exists a $y \in \mathbb{R}$ such that $x + y = 0$". For this we consider the set

$$X_+ := \{(x,y,z) \in \mathbb{R}^3 : x + y = z\} \in \hat{\mathbb{C}}$$

which again is a constant in our language. Now we can express the statement as follows

$$(\forall x)(x \in \mathbb{R} \to (\exists y)(y \in \mathbb{R} \to (\exists a \in X_+)(\exists b \in a)(0 \in b \wedge (\exists c \in b)(\exists d \in c)(x \in d \wedge y \in d))))$$

where we use that

$$(x,y,0) := ((x,y),0) := \{\{\{\{x\},\{x,y\}\}\},\{\{\{x\},\{x,y\}\},0\}\}.$$

Obviously the more complicated theorems will be very hard to understand using this kind of formulas. That is why we will be using abbreviations, to reduce the formula above to just

$$(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})(x + y = 0).$$

With this language, alongside its natural interpretation, we now can talk about an $\mathcal{L}$-structure which we will denote by $\hat{\mathcal{C}}$. Using filters we can now look at the non-standard superstructure $\hat{\mathcal{C}}^*$.

2.5 Ultrapowers of superstructures

Let $\mathcal{F}$ be a non-principal ultrafilter of $\mathbb{N}$. Let $\hat{\mathbb{C}}^*$ be the ultrapower of $\hat{\mathcal{C}}$ with respect to $\mathcal{F}$ and let $* : \hat{\mathbb{C}} \to \hat{\mathbb{C}}^*, x \mapsto [(x)_{n\in\mathbb{N}}] = x^*$ be its non-standard map. When dealing with elements of $\mathbb{C}$ this was an easy map, however with sets it requires a bit more attention. As we already saw in Section 2.3, $\mathbb{C}^*$ is not the set $\{x^* : x \in \mathbb{C}\}$. As a result some of the operators we use on sets do not commute with the non-standard map. In this section we will find out more about the non-standardization of sets and what sets are part of our new structure.

Lemma 2.29. *1. $\emptyset^* = \emptyset$.*

2. If $a, b \in \hat{\mathbb{C}}$, then $a \subset b$ if and only if $a^* \subset b^*$.

3. If $a, b \in \hat{\mathbb{C}}$, then $a \in b$ is and only if $a^* \in b^*$.

4. For all $a \in \hat{\mathbb{C}}$ we have $\{a\}^* = \{a^*\}$.

5. If $a_1, ..., a_n \in \hat{C}$ then $(\bigcup_{i=1}^n a_i)^* = (\bigcup_{i=1}^n a_i^*)$, $(\bigcap_{i=1}^n a_i)^* = (\bigcap_{i=1}^n a_i^*)$, $\{a_1, ..., a_n\}^* = \{a_1^*, ..., a_n^*\}$ and $(a_1, a_2)^* = (a_1^*, a_2^*)$.

6. *For all $a, b \in \hat{\mathbb{C}}$ we have $(a \setminus b)^* = (a^* \setminus b^*)$.*

7. *If b is a binary relation, then $(dom\ b)^* = dom(b^*)$, $(ran\ b)^* = ran(b^*)$ and for all $a \in dom\ b$ we have $(b(a))^* = b^*(a^*)$.*

8. *If $\phi(v)$ is a formula with one free variable and $a, b \in \hat{\mathbb{C}}$ then $a := \{x \in b : \hat{\mathcal{C}} \models \phi(x)\}$ if and only if $a^* = \{x \in b^* : \hat{\mathcal{C}}^* \models \phi(x)\}$.*

Proof. 1. Note that for all $x = [(x(n))] \in \hat{\mathbb{C}}^*$ it holds that $x(n) \notin \emptyset$ for all $n \in \mathbb{N}$. Since $\mathbb{N} \in \mathcal{F}$ we have $x \notin \emptyset^*$ hence $\emptyset^* = \emptyset$.

2. Let $a, b \in \hat{\mathbb{C}}$. Note that $a \subset b$ if and only if

$$\hat{\mathcal{C}} \models (\forall x)(x \in a \to x \in b)$$

which by Łośis equivalent to

$$\hat{\mathcal{C}}^* \models (\forall x)(x \in a^* \to x \in b^*)$$

which is equivalent to $a^* \subset b^*$.

3. Let $a, b \in \hat{\mathbb{C}}$. Then $(a \in b)$ is a closed formula thus by Łoś$a^* \in b^*$ if and only if $a \in b$.

4. Let $a \in \hat{\mathbb{C}}$. We have $\{a\}^* = [(\{a\})]$ hence $x = [(x(n))] \in \{a\}^*$ if and only if

$$\{n : x(n) \in \{a\}\} \in \mathcal{F} \Leftrightarrow \{n : x(n) = a\} \in \mathcal{F} \Leftrightarrow x = a^*.$$

5. Let $a_1, ..., a_n \in \hat{C}$. Observe that

$$\begin{aligned}
&\hat{\mathcal{C}} \models (\forall x)(x \in (\cup_{i=1}^n a_i) \leftrightarrow x \in a_1 \vee ... \vee x \in a_n) \\
&\hat{\mathcal{C}} \models (\forall x)(x \in (\cap_{i=1}^n a_i) \leftrightarrow x \in a_1 \wedge ... \wedge x \in a_n) \\
&\hat{\mathcal{C}} \models (\forall x)(x \in \{a_1, ..., a_n\} \leftrightarrow x = a_1 \vee ... \vee x = a_n)
\end{aligned}$$

all are closed formulas, hence by Łoś' Theorem we find that

$$\begin{aligned}
&\hat{\mathcal{C}}^* \models (\forall x)(x \in (\cup_{i=1}^n a_i)^* \leftrightarrow x \in a_1^* \vee ... \vee x \in a_n^*) \\
&\hat{\mathcal{C}}^* \models (\forall x)(x \in (\cap_{i=1}^n a_i)^* \leftrightarrow x \in a_1^* \wedge ... \wedge x \in a_n^*) \\
&\hat{\mathcal{C}}^* \models (\forall x)(x \in \{a_1, ..., a_n\}^* \leftrightarrow x = a_1^* \vee ... \vee x = a_n^*).
\end{aligned}$$

This is what we wanted to prove. In particular we find

$$(a_1, a_2)^* = \{\{a_1\}, \{a_1, a_2\}\}^* = \{\{a_1\}^*, \{a_1, a_2\}^*\} = \{\{a_1^*\}, \{a_1^*, a_2^*\}\} = (a_1^*, a_2^*).$$

6. Let $a, b \in \hat{\mathbb{C}}$ and note that

$$\hat{\mathcal{C}} \models (\forall x)(x \in a \setminus b \leftrightarrow x \in a \wedge x \notin b)$$

is a closed formula and thus by Łoś' Theorem we find

$$\hat{\mathcal{C}}^* \models (\forall x)(x \in (a \setminus b)^* \leftrightarrow x \in a^* \wedge x \notin b^*)$$

and we already knew that

$$\hat{\mathcal{C}}^* \models (\forall x)(x \in a^* \setminus b^* \leftrightarrow x \in a^* \wedge x \notin b^*).$$

7. First note that if b is a binary relation with domain $\mathrm{dom}(b)$ and range $\mathrm{ran}(a)$ then
$$\hat{\mathcal{C}} \models (\forall x)(x \in b \to (\exists y \in \mathrm{dom}(b))(\exists z \in \mathrm{ran}(b))(x = (y, z))).$$
Since this is a closed formula, by Łoś' Theorem this is equivalent to
$$\hat{\mathcal{C}}^* \models (\forall x)(x \in b^* \to (\exists y \in \mathrm{dom}(b)^*)(\exists z \in \mathrm{ran}(b)^*)(x = (y, z))).$$
Thus we find that b^* is a binary relation with $\mathrm{dom}(b)^* = \mathrm{dom}(b^*)$, $\mathrm{ran}(b)^* = \mathrm{ran}(b^*)$.

Now let $a \in \mathrm{dom}(b)$ and $b(a) \in \mathrm{ran}(b)$. Then
$$\hat{\mathcal{C}} \models ((a, b(a)) \in b).$$
is closed, hence by Łoś' Theorem is equivalent to
$$\hat{\mathcal{C}}^* \models ((a, b(a))^* \in b^*).$$
By 2.29.5 we have that $(a, b(a))^* = (a^*, (b(a))^*) \in b^*$. Hence $b^*(a^*) = (b(a))^*$.

8. Let $\phi(v)$ a formula with one free variable and let $a, b \in \hat{\mathbb{C}}$. Note that by Łoś' Theorem
$$\hat{\mathcal{C}} \models (\forall x \in b)(x \in a \leftrightarrow \phi(x))$$
if and only if
$$\hat{\mathcal{C}}^* \models (\forall x \in b^*)(x \in a^* \leftrightarrow \phi(x))$$
and by 2.29.2 $a \subset b$ if and only if $a^* \subset b^*$. Hence $a := \{x \in b : \hat{\mathcal{C}} \models \phi(x)\}$ if and only if $a^* = \{x \in b^* : \hat{\mathcal{C}}^* \models \phi(x)\}$.

□

Although most operations on sets are the same for non-standard analysis, there are a few that are different. This has important consequences. To see this we again consider the statement that every set of $\mathbb{N}$ has a smallest element. However if we consider $\mathbb{N}^* \setminus \mathbb{N}^\#$, the set of infinite positive integers, we find that it does not have a smallest element. But the Theorem of Łośgives us that every set of $(\mathcal{P}(\mathbb{N}))^*$ has a smallest element. From this we must conclude that $\mathbb{N}^* \setminus \mathbb{N}^\# \notin (\mathcal{P}(\mathbb{N}))^*$. To distinguish the different sets and elements we introduce the following definition.

Definition 2.30. Consider $\hat{\mathbb{C}^*}$, the superstructure of $\mathbb{C}^*$. We call $x \in \hat{\mathbb{C}}^*$ *internal* if $x \in (\mathbb{C}_n)^*$ for some $n \in \mathbb{N}$. If there is an $a \in \hat{\mathbb{C}}$ such that $x = a^*$, then we call x *standard*. All entities which are not internal, we call it *external*. Right now we used brackets to differentiate $(\mathbb{C}_n)^*$ and $(\mathbb{C}^*)_n$ and make clear they are significantly different sets. In the future we will just use $\mathbb{C}_n^*$ to denote $(\mathbb{C}_n)^*$.

In Definition 2.30 we defined external entities instead of external elements. We did this since some of the entities we want to think about are not an element of $(\hat{\mathbb{C}})^*$. This leads us to two distinct classes of external entities, those that are an element of $\hat{\mathbb{C}}^*$ like $x = [(\{0\}, \{\{0\}\}, \{\{\{0\}\}\}, ...)]$ which for every n is not in $(\mathbb{C}_n)^*$, and those that are not like $\mathbb{N}^* \setminus \mathbb{N}^\#$.

Proposition 2.31. $\mathbb{N}^* \setminus \mathbb{N}^\# \notin \hat{\mathbb{C}}^*$.

Proof. Suppose $\mathbb{N}^* \setminus \mathbb{N}^\# = [(x_n)_{n\in\mathbb{N}}] \in \hat{\mathbb{C}}^*$. Note that $\mathbb{N}^* \setminus \mathbb{N}^\# \subset \mathbb{N}^*$. However we know that

$$\hat{\mathcal{C}} \models (\forall x)(x \subset \mathbb{N} \leftrightarrow x \in \mathcal{P}(\mathbb{N})).$$

which is a closed formula, thus by the Łoś' Theorem we have

$$\hat{\mathcal{C}}^* \models (\forall x)(x \subset \mathbb{N}^* \leftrightarrow x \in (\mathcal{P}(\mathbb{N}))^*).$$

With our earlier observation this means that $\mathbb{N}^* \setminus \mathbb{N}^\# \in (\mathcal{P}(\mathbb{N}))^*$ which we already found to be impossible. $\square$

Proposition 2.32. *There are internal sets which are not standard.*

Proof. Consider $\{[(n)_{n\in\mathbb{N}}]\} = [(\{n\})_{n\in\mathbb{N}}] \in \mathbb{C}_1^*$. This set is not standard, but is is internal. $\square$

Theorem 2.33. *Elements of internal set are internal.*

Proof. Let $x = [(x_n)_{n\in\mathbb{N}}] \in \mathbb{C}_n^*$ be an internal set, that means that $\{n \in \mathbb{N} : x_n \in \mathbb{C}_n\} \in \mathcal{F}$. Now let $y = [(y_n)_{n\in\mathbb{N}}] \in x$. Then $\{n \in \mathbb{N} : y_n \in x_n\} \in \mathcal{F}$. Since $\mathbb{C}_n = \mathcal{P}(\mathbb{C}_{n-1}) \cup \mathbb{C}$ by the definitions of filters we find

$$\{n \in \mathbb{N} : x_n \in \mathbb{C}_n\} \cap \{n \in \mathbb{N} : y_n \in x_n\} \subset \{n \in \mathbb{N} : y_n \in \mathbb{C}_{n-1}\} \in \mathcal{F}.$$

Hence $y \in \mathbb{C}_{n-1}^*$ and thus is internal. $\square$

So the statement in standard analysis that every subset of $\mathbb{N}$ has a smallest element, in non-standard analysis is replaced by the statement that every internal subset of $\mathbb{N}^*$ has a smallest element.
This means that if we are studying specific sets in $\hat{\mathbb{C}}^*$ we want to know whether they are internal or not. For this we look at the following technical but useful results.

Proposition 2.34. *Let $a_1, ..., a_k \in \hat{\mathbb{C}}^*$ be internal, $n_1, ..., n_k \in \mathbb{N}$ such that $a_j \in \mathbb{C}_{n_j}^*$ for all $j = 1, ..., k$ and let $\phi(v, w_1, ..., w_k)$ a formula with $k+1$ free variables v and $w_1, ..., w_k$. Let $m \in \mathbb{N}$ and $b := \{x \in \mathbb{C}_m^* : \hat{\mathcal{C}}^* \models \phi(x, a_1, ..., a_k)\}$, then b is internal.*

Proof. Consider the formula

$$\mathfrak{B} : (\forall a_1 \in \mathbb{C}_{n_1})...(\forall a_k \in \mathbb{C}_{n_k})(\exists b \in \mathbb{C}_{m+1})(\forall x)(x \in b \leftrightarrow (x \in \mathbb{C}_m \wedge \phi(x, a_1, ..., a_k))).$$

Note that $\mathfrak{B}$ is closed and $\hat{\mathcal{C}} \models \mathfrak{B}$ by the axioms of set theory. Hence Łoś' Theorem gives

$$\hat{\mathcal{C}}^* \models (\forall a_1 \in \mathbb{C}_{n_1}^*)...(\forall a_k \in \mathbb{C}_{n_k}^*)(\exists b \in \mathbb{C}_{m+1}^*)(x \in b \leftrightarrow (x \in \mathbb{C}_m^* \wedge \phi(x, a_1, ..., a_k))).$$

Hence b is internal. $\square$

Corollary 2.35. *Let $b \in \hat{\mathbb{C}}^*$ be an internal binary relation. Then* $\mathrm{dom}(b)$ *and* $\mathrm{ran}(b)$ *are internal.*

Proof. Let $n \in \mathbb{N}$ be such that $b \in \mathbb{C}_n^*$. Consider

$$\phi_1(v,w) : (\exists z)((v,z) \in w) \text{ and } \phi_2(v,w) : (\exists z)((z,v) \in w).$$

Note that $\mathrm{dom}(b) = \{x \in \mathbb{C}_n^* : \phi_1(x,b)\}$ and $\mathrm{ran}(b) = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}} \models \phi_2(x,b)\}$. Hence by Proposition 2.34 $\mathrm{dom}(b)$ and $\mathrm{ran}(b)$ are internal. $\square$

Corollary 2.36. *Let $b \in \hat{\mathbb{C}}^*$ be an internal binary relation and let $a \subset \mathrm{dom}(b)$ be an internal set. Then $b|_a$, i.e. b restricted to a and $b(a) = \mathrm{ran}(b|_a)$ are internal.*

Proof. Consider the formula

$$\phi(v, w_1, w_2) : (v \in w_2 \wedge (\exists y \in w_1)(\exists z)(y,z) = v)$$

and note that $b|_a = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}} \models \phi(x,a,b)\}$. Hence by Proposition 2.34 $b|_a$ is internal and thus by Corollary 2.35 $\mathrm{ran}(b_a) = b(a)$ is internal. $\square$

Of course the most important binary relations are maps for which we can prove a few extra results.

Proposition 2.37. *Let a, b be internal maps with* $\mathrm{ran}(a) \subset \mathrm{dom}(b)$. *Then $b \circ a$ is an internal map.*

Proof. Let $n \in \mathbb{N}$ be such that $a, b \in \mathbb{C}_n^*$. Evidently $b \circ a$ is again a map. To see it is internal consider:

$$\phi(x, w_1, w_2) : (\exists w)(\exists y)(\exists z)((y,w) \in w_1 \wedge (w,z) \in w_2 \wedge x = (y,z))$$

and note that $b \circ a = \{x \in \mathbb{C}_{2n}^* : \hat{\mathcal{C}}^* \models \phi(x,a,b)\}$. By Proposition 2.34 we see that $b \circ a$ is internal. $\square$

Again to make our formula less complicated, for a given internal function a and a $x \in \mathrm{dom}(a)$, instead of

$$(\exists y)((x,y) \in a)$$

and work with y we will just write $a(x)$.

Finally we will prove that the composition and evaluation maps are internal.

Proposition 2.38. *Let $X, Y, U \in \hat{\mathbb{C}}$ internal, then:*

(a) *For all $f \in X^U$ and all $g \in U^Y$ the function $M_f : Y^X \to Y^U,\ h \mapsto h \circ f$ and $M_g : Y^X \to U^X,\ h \mapsto g \circ h$ are internal.*

(b) *For all $a \in X$ the function $V_a : Y^X \to Y,\ h \mapsto h(a)$ is internal.*

Proof. Let $n \in \mathbb{N}$ be such that $Y \times X \times U \times Y, Y \times X \times U \times X \in \mathbb{C}_n^*$. Consider

$$\phi_f(x, w_1, w_2, w_3) : (\exists y \in w_1)(\exists z \in w_2)(x = (y, z) \wedge z = y \circ w_3),$$

$$\phi_g(x, w_1, w_2, w_3) : (\exists y \in w_1)(\exists z \in w_2)(x = (y, z) \wedge z = w_3 \circ y)$$

and

$$\phi_a(x, w_1, w_2, w_3) : (\exists y \in w_1)(\exists z \in w_2)(x = (y, z) \wedge z = y(w_3)).$$

Note that $M_f = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}}^* \models \phi_f(x, Y^X, X^U, f)\}$,
$M_g = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}}^* \models \phi_g(x, Y^X, U^Y, g)\}$ and $V_a = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}}^* \models \phi_a(x, Y^X, Y, a)\}$.
By Proposition 2.34 we conclude that M_f, M_g and V_x are internal. □

3 The invariant subspace problem

3.1 Dimension of a vector space

Our goal is to use non-standard analysis on vector spaces, in particular on function and sequence spaces. However some of the mathematical concepts, which are easily defined in mathematical language, are quite hard to describe in logical formulas. One of these concepts is the dimension of a vector space which is tied to a basis of the vector space. We will only describe a basis for normed vector spaces and for that we will have to define the span of a set of vectors, and the closure of a set in a vector space. Before we can define the span of a set, we first need the concept of finitude.

To define finitude we will use cardinality. The finite sets are luckily easily represented by the sets $K_n := \{1, ..., n\}$ with $n \in \mathbb{N}$. Note that the set

$$K := \{K_n : n \in \mathbb{N}\} \in \hat{\mathbb{C}}.$$

From there we can define the function: $f_K : \mathbb{N} \to K,\ n \mapsto K_n$. We can now express that a set A has finite cardinality by

$$\text{fin}(A) : \ (\exists n \in \mathbb{N})(\exists f \in K_n^A)(\text{ran}(f) = f_K(n) \wedge \text{inj}(f))$$

and we can express that A has cardinality n for some $n \in \mathbb{N}$ by:

$$|A| = n : \ (\exists f \in K_n^A)(\text{ran}(f) = f_K(n) \wedge \text{inj}(f))$$

and that A has cardinality $\aleph_0$ by

$$|A| = \aleph_0 : \ (\exists f \in \mathbb{N}^A)(\text{ran}(f) = \mathbb{N} \wedge \text{inj}(f)).$$

Here inj expresses that f is an injection. We leave that formula as an exercise for the reader. Note that we have the surjectivity by the fact that $\text{ran}(f) = f_K(n)$ or $\text{ran}(f) = \mathbb{N}$.

With this knowledge we can now define finite and infinite sums in a normed vector space. We will start with the finite sum which we can define by induction. For this let $X \in \hat{\mathbb{C}}$ be a vector space and we define for every $n \in \mathbb{N}$ the set $X_n = \{A \in \mathcal{P}(X) : |A| \leq n\}$. Consider the formula:

$$(\exists \Sigma \in X^{X_n})(\Sigma(\emptyset) = 0 \wedge ((\forall A \in X_n)(\forall x \in A)(\Sigma(A) = \Sigma(A \setminus \{x\}) + x))).$$

Note that this evidently holds for $n = 0$ and that if it holds for an arbitrary $N \in \mathbb{N}$ then it also holds for $N + 1$, hence we have finite sums.

For infinite sums consider that for every vector space X and every $n \in \mathbb{N}$ we can logically define the function

$$s_n : X^{\mathbb{N}} \to X, (x_n)_{n \in \mathbb{N}} \mapsto \sum_{i=1}^{n} x_n.$$

Then we can define the infinite sum as the limit of $s_n((x_n))$. Of course to describe that a sequence (v_n) converges to some v we should use the formula

$$(\forall \epsilon)(\exists N \in \mathbb{N})(\forall n \in \mathbb{N}_{>N})(\|v_n - v\| < \epsilon).$$

To simplify our formulas we will write $\lim_{n\to\infty} v_n = v$ instead.
Note that we can find the set $X_l = \{x = (x_n)_{n \in \mathbb{N}} \in X^{\mathbb{N}} : (\exists c \in X)(\lim_{n\to\infty} s_n(x) = c\}$ by

$$(\exists X_l)(x \in X_l \leftrightarrow (\exists c \in X)(\lim_{n\to\infty} s_n(x) = c)).$$

From this we can define the infinite sums by

$$(\exists \Sigma \in X^{X_l})((\forall c \in X)(\forall x \in X_l)\Sigma(x) = c \leftrightarrow \lim_{n\to\infty} s_n(x) = c)).$$

As a last step we will define the support of a function.

Definition 3.1. Let $T : A \to X$ be a function where X is a vector space. Then the *support* of T is the set $\text{supp}(T) := \{x \in A : T(x) \neq 0\}$. Note that we can express this in logic formulas.

Note that $\mathbb{C}$ is a vector space and therefore we can give the following definition.

Definition 3.2. Let X be a normed vector space and let $A \subset X$. Then the *span of* A is defined as

$$\text{span}(A) = \{\sum_{x \in A} c_x x | c_x \in \mathbb{C}, c_x \neq 0 \text{ for finitely many } x\}.$$

Logically we can express this as

$$(\exists B)(x \in B \leftrightarrow (\exists g \in \mathbb{C}^A)(\text{fin}(\text{supp}(g)) \wedge \sum_{y \in \text{supp}(g)} g(y)y = x)).$$

Also the notion of linear independence is now easily obtained for countable sets.

Definition 3.3. Let X be a vector space and let $A \subset X$. Then A is called *linearly independent* if

$$(\forall g \in \mathbb{C}^A)((\text{fin}(\text{supp}(g)) \wedge \sum_{x \in \text{supp}(g)} g(x)x = 0) \rightarrow (\forall y \in A)(g(y) = 0)).$$

We now have enough to define a basis of finite dimensional vector spaces. For countably infinite dimensional vector spaces it requires one more step.

Definition 3.4. Let X be a normed vector space and $A \subset X$. Then we call

$$\overline{A} := \{x \in X : (\exists (x_n)_{n \in \mathbb{N}} \in A^{\mathbb{N}})(\lim_{n \to \infty} x_n = x)\}$$

the closure of A in X.

We now call $A \subset X$ a basis of a normed vector space X if A is linearly independent and $\overline{\text{span}(A)} = X$. All the work we did in this section ensures that we can express this in our logical language. A consequence of this is that results about finite dimensional vector spaces in the standard analysis can now be carried over to the non-standard analysis.

3.2 Non-standard normed spaces

We now have enough knowledge to start working on the proof of our first major result, solving a case of the invariant subspace problem. The proof however does require some preliminary results. In particular results concerning $l_2(\mathbb{C})^*$ since $l_2(\mathbb{C})$ is up to isomorphism the only infinite dimensional separable Hilbert space. This will play an important role in the proof. We can view a sequence as a binary relation with domain $\mathbb{N}$ and co-domain $\mathbb{C}$. In non-standard analysis this gives us domain $\mathbb{N}^*$ and $\mathbb{C}^*$. We will first take a closer look at non-standard infinite sequences.

Theorem 3.5. *Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{C}$. Then (a_n) converges to a in $\mathbb{C}$ if and only if $|a - a_n|$ is infinitesimal for all $n \in \mathbb{N} \setminus \mathbb{N}^{\#}$.*

Proof. Suppose $\lim_{n \to \infty} a_n = a$. We find for all $\epsilon \in \mathbb{R}_{>0}$ an $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}_{>N}$ we have $|a - a_n| < \epsilon$. Thus

$$\hat{\mathcal{C}} \models (\forall n \in \mathbb{N}_{>N})(|a - a_n| < \epsilon).$$

Taking ϵ and N constant, this is a closed formula. Hence Łoś' Theorem gives

$$\hat{\mathcal{C}}^* \models (\forall n \in \mathbb{N}^*_{>N})(|a - a_n| < \epsilon).$$

Hence for all $\epsilon \in \mathbb{R}_{>0}$ and $n \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ we have $|a - a_n| < \epsilon$. Thus $|a - a_n|$ is infinitesimal.

Suppose $|a - a_n|$ is infinitesimal for all $n \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$. Then for all $\epsilon \in \mathbb{R}^{\#}_{>0}$ we find

$$\hat{\mathcal{C}}^* \models (\exists N \in \mathbb{N}^*)(\forall n \in \mathbb{N}^*_{>N})(|a - a_n| < \epsilon)$$

since we can take $N \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$. By Łoś' Theorem we find

$$\hat{\mathcal{C}} \models (\exists N \in \mathbb{N})(\forall n \in \mathbb{N}_{>N})(|a - a_n| < \epsilon).$$

Since this holds for all $\epsilon \in \mathbb{R}_{>0}$ we have that $\lim_{n\to\infty} a_n = a$. □

Theorem 3.6. *Let a be an internal sequence in $\mathbb{C}^*$ such that $a(n)$ is infinitesimal for all finite positive integers n. Then there exists an infinite positive integer ω such that $a(n)$ is infinitesimal for all $n < \omega$.*

Proof. Let $a \in (\mathbb{C}^{\mathbb{N}})^*$ satisfy the conditions of the theorem. Consider the formula

$$\phi(v, w_1, w_2) : (v \in w_1 \wedge ((\exists y)((v, y) \in w_2 \wedge v|y| \geq 1))).$$

Note that

$$\{n \in \mathbb{N}^* : n|a(n)| \geq 1\} = \{n \in \mathbb{C}^*_0 : \hat{\mathcal{C}}^* \models \phi(n, \mathbb{N}^*, a)\}.$$

By Proposition 2.34 $\{n \in \mathbb{N}^* : n|a(n)| \geq 1\}$ is internal.

If $\{n \in \mathbb{N}^* : n|a(n)| \geq 1\} = \emptyset$ we are done, otherwise $\{n \in \mathbb{N}^* : n|a(n)| \geq 1\}$ has a smallest element $\omega \in \mathbb{N}^*$. This cannot be finite since for all $n \in \mathbb{N}^{\#}$ we have that $a(n)$ infinitesimal, hence $n|a(n)|$ is infinitesimal.
For all infinite $n < \omega$ we find that $|a(n)| < \frac{1}{n}$, since n is infinite, $\frac{1}{n}$ is infinitesimal and thus $a(n)$ is infinitesimal. □

Lemma 3.7.

$$l_2(\mathbb{C})^* = \{s \in (\mathbb{C}^{\mathbb{N}})^* : \|s\|_2^2 = \sum_{n=1}^{\infty} |s_n|^2 \text{ exists }\}.$$

Proof. We know that by definition

$$l_2(\mathbb{C}) := \{s \in \mathbb{C}^{\mathbb{N}} : \|s\|_2^2 = \sum_{n=1}^{\infty} |s_n|^2 \text{ exists}\}.$$

Now consider the formula

$$\phi(v) : (\exists y \in \mathbb{R})(\lim_{n\to\infty} \sum_{i=1}^{n} |v_i|^2 = y)$$

and note that $l_2(\mathbb{C}) := \{s \in \mathbb{C}^{\mathbb{N}} : \hat{\mathcal{C}} \models \phi(s)\}$ By Proposition 2.29.8 we now find that

$$l_2(\mathbb{C})^* = \{s \in (\mathbb{C}^{\mathbb{N}})^* : \|s\|_2^2 = \sum_{n=1}^{\infty} |s_n|^2 \text{ exists}\}.$$

□

Definition 3.8. Let X be a normed vector space. We call $\sigma \in X^*$ *norm finite* if $\|\sigma\|$ is finite.
We call σ *near-standard* if there is a $\sigma^\circ \in X^\#$ such that $\|\sigma - \sigma^\circ\|$ is infinitesimal; we call σ° the *standard part* of σ. If $\sigma, \tau \in X^*$ are infinitely close, i.e. $\|\sigma - \tau\|$ is infinitesimal, then we write $\sigma \approx \tau$.

Proposition 3.9. *Let X be a normed vector space. If $\sigma \in X^*$ is near-standard, then its standard part $\sigma^\circ \in X^\#$ is unique.*

Proof. Let $\sigma_1, \sigma_2 \in X^\#$ be such that $\|\sigma - \sigma_1\|$ and $\|\sigma - \sigma_2\|$ are infinitesimal. Then

$$\|\sigma_2 - \sigma_1\| = \|\sigma_2 - \sigma + \sigma - \sigma_1\| \leq \|\sigma - \sigma_1\| + \|\sigma - \sigma_2\|.$$

Hence $\|\sigma_2 - \sigma_1\| \in \mathbb{R}^\#$ is infinitesimal, thus 0. So $\sigma_1 = \sigma_2$. □

Using Theorem 3.5 on the sequence of partial sums of a $\sigma = (s_n) \in l_2(\mathbb{C})$ we find:

Corollary 3.10. *For all $\sigma = (s_n) \in l_2(\mathbb{C})^\#$ and all $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ we have that $\sum_{n=\omega}^{\infty} |s_n|^2$ is infinitesimal.*

From this we can derive the following.

Theorem 3.11. *Let $\sigma = (s_n) \in l_2(\mathbb{C})^*$. Then σ is near-standard if and only if σ is norm finite and $\sum_{n=\omega}^{\infty} |s_n|^2$ is infinitesimal for all $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$.*

Proof. Suppose that σ is near-standard. Then

$$\|\sigma\|_2 = \|\sigma - \sigma^\circ + \sigma^\circ\|_2 \leq \|\sigma - \sigma^\circ\|_2 + \|\sigma^\circ\|_2 < 1 + \|\sigma^\circ\|_2.$$

So σ is norm finite. The other property follows directly from Corollary 3.10.

Suppose that σ is norm finite and $\sum_{n=\omega}^{\infty} |s_n|^2$ is infinitesimal for all $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$. Then for all $n \in \mathbb{N}^*$ $|s_n|$ is finite since σ is norm finite. Hence we may consider $\hat{\sigma} = (s_n^\circ)_{n \in \mathbb{N}}$, the sequence of standard parts of s_n in $\mathbb{C}^{\mathbb{N}}$. Note that for all $k \in \mathbb{N}^\#$ we have that

$$\left|\sum_{n=1}^{k} |s_n^\circ|^2 - \sum_{n=1}^{k} |s_n|^2\right| \leq \sum_{n=1}^{k} |(s_n^\circ)^2 - s_n^2|$$

is infinitesimal. Hence

$$\sum_{i=1}^{k} |s_n^\circ|^2 \leq \frac{1}{2} + \sum_{i=1}^{k} |s_n|^2 \leq 1 + \|\sigma\|_2^2.$$

Hence $\sum_{n=1}^{k} |s_n^\circ|^2$ is a monotonically increasing bounded sequence, thus converges. So $\hat{\sigma} \in l_2(\mathbb{C})$.
Consider $\sigma' = \hat{\sigma}^* \in l_2(\mathbb{C})^\#$. Note that for all $k \in \mathbb{N}^\#$ we have $\sum_{n=1}^{k} |s_n - s_n'|^2$ is infinitesimal, hence from Theorem 3.6 we have that there is an $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ such that $\sum_{i=1}^{\omega-1} |s_n - s_n'|^2$ is infinitesimal.

By assumption we know that $\sum_{n=\omega}^{\infty} |s_n|^2$ is infinitesimal and by Theorem 3.10 $\sum_{n=\omega}^{\infty} |s'_n|^2$ is infinitesimal. Hence we find that

$$\|\sigma - \sigma'\|_2^2 = \sum_{n=1}^{\omega-1} |s_n - s'_n|^2 + \sum_{n=\omega}^{\infty} |s_n - s'_n|^2 \quad \leq \sum_{n=1}^{\omega-1} |s_n - s'_n|^2 + ((\sum_{n=\omega}^{\infty} |s_n|^2)^{\frac{1}{2}} + (\sum_{n=\omega}^{\infty} |s'_n|^2)^{\frac{1}{2}})$$

which is infinitesimal. Hence σ is near-standard. □

Theorem 3.12. *Let $A \subset l_2(\mathbb{C})$ be a compact set. Then all $\sigma \in A^*$ are near-standard.*

Proof. Suppose there is a $\sigma = (s_n) \in A^*$ not near-standard.
Claim: There is an $r > 0$ such that $\|\sigma - \tau\|_2 > r$ for all $\tau \in l_2(\mathbb{C})$.
If σ is not norm finite it is clear, so assume that σ is norm finite. Then by Theorem 3.11 there is an $\omega \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ such that $\sum_{n=\omega}^{\infty} |s_n|^2 > 2r^2$ for some $r \in \mathbb{R}^{\#}_{>0}$. For all $\tau = (t_n) \in l_2(\mathbb{C})^{\#}$, $\sum_{n=\omega}^{\infty} |t_n|^2$ is infinitesimal. Hence

$$\begin{aligned}\|\sigma - \tau\|_2 &\geq (\sum_{n=\omega}^{\infty} |s_n - t_n|^2)^{\frac{1}{2}} \\ &\geq (\sum_{n=\omega}^{\infty} |s_n|^2)^{\frac{1}{2}} - (\sum_{n=\omega}^{\infty} |t_n|^2)^{\frac{1}{2}} > r\end{aligned}$$

which proves our claim.

Since A is compact we however have $\tau_1, ..., \tau_n \in A$ such that

$$\hat{\mathcal{C}} \models (\forall \xi \in A)(\|\xi - \tau_1\|_2 < r \vee ... \vee \|\xi - \tau_n\|_2 < r)$$

hence by Łoś' Theorem we find

$$\hat{\mathcal{C}} \models (\forall \xi \in A^*)(\|\xi - \tau_1^*\|_2 < r \vee ... \vee \|\xi - \tau_n^*\|_2 < r).$$

So in particular we find that there is a $j \in \{1, ..., n\}$ such that $\|\sigma - \tau_j\|_2 < r$. Contradiction. □

3.3 Operators in non-standard separable Hilbert spaces

We will now look at linear operators in vector spaces. For now we will focus on linear operators on $l_2(\mathbb{C})$, but we will start a bit more general. Let $H, K \in \hat{\mathbb{C}}$ be vector spaces over $\mathbb{C}$ and let $T : H^* \to K^*$ be an internal function. Note that

$$(\forall x, y \in H)(\forall r, s \in \mathbb{C})(T(rx + sy) = rT(x) + sT(y))$$

is a closed formula for a given $T \in K^H$. Let $L(K, H)$ be the set of all linear operators from H to K, then by Łoś' Theorem we find that $L(H, K)^*$ is the set of all internal non-standard linear operators from H^* to K^*. We also have $\mathcal{L}(H, K)$, the set of continuous linear operators from H to K. We can prove that continuity and boundedness are properties that can be caught in a closed formula, and hence are also equivalent properties in non-standard analysis.

Proposition 3.13. *Let $T : K \to H$ be an internal linear operator. T is bounded if and only if T is continuous. Furthermore $\mathcal{L}(H,K)^*$ is the set of all internal bounded linear operators from H^* to K^*.*

Proof. Consider the formula

$$\mathfrak{B}(T) : (\forall (x_n) \in H^{\mathbb{N}})(\exists x \in H)((\lim_{n\to\infty} x_n = x) \to (\lim_{n\to\infty} T(x_n) = T(x)))$$

i.e. T is continuous and

$$\mathfrak{C}(T) : (\exists C \in \mathbb{R}_{>0})(\forall x \in H)(\|x\| \leq C\|T(x)\|)$$

i.e. T is bounded.

We know that T is bounded if and only if T is continuous, hence

$$\hat{\mathcal{C}} \models (\forall T \in L(H,K))(\mathfrak{B}(T) \leftrightarrow \mathfrak{C}(T)).$$

Hence by Łoś' Theorem we have

$$\hat{\mathcal{C}} \models (\forall T \in L(H,K)^*)(\mathfrak{B}(T) \leftrightarrow \mathfrak{C}(T)).$$

□

As one might expect, the sums and scalar products of internal linear operators are again internal linear operators.

Proposition 3.14. *Let $H, K \in \hat{\mathbb{C}}^*$ be internal vector spaces, let $S, T \in L(H,K)^*$ and let $\lambda \in \mathbb{C}^*$. Then $S + T$ and λT are internal.*

Proof. Let $n \in \mathbb{N}$ be such that $H, K \in \mathbb{C}_n^*$. Consider:

$$\phi_+(x, w_1, w_2, w_3, w_4) : (\exists y \in w_3)(\exists z \in w_4)((y,z) = x \wedge z = w_1(y) + w_2(y))$$

and

$$\phi_\cdot(x, w_1, w_2, w_3, w_4) : (\exists y \in w_3)(\exists z \in w_4)((y,z) = x \wedge z = w_1 \cdot w_2(y))$$

and note that $S + T = \{x \in \mathbb{C}_{n+2}^* : \hat{\mathcal{C}} \models \phi_+(x, S, T, H^*, K^*)\}$ and
$\lambda T = \{x \in \mathbb{C}_{n+2}^* : \hat{\mathcal{C}} \models \phi_+(x, \lambda, T, H^*, K^*)\}$. Thus by Proposition 2.34 we are done. □

We will also consider some of the maps on the vector space of operators between two vector spaces.

Proposition 3.15. *Let $V, W \in \hat{\mathbb{C}}$ be internal vector spaces and let $R \in L(V,W)^*$. Then $f : L(V,W)^* \to L(V,W)^*$, $S \mapsto S + R$ is an internal function.*

Proof. Let $n \in \mathbb{N}$ be such that $L(V,W)^*, L(V,W)^* \in \mathbb{C}_n^*$ and consider

$$\phi(x, w_1, w_2) : (\exists y \in w_1)(\exists z \in w_1)((y,z) = x \wedge z = y + w_2).$$

Finally note that $f = \{x \in \mathbb{C}_{n+2}^* : \hat{\mathcal{C}} \models \phi(x, L(V,W)^*, R)\}$. Thus by Proposition 2.34 we are done. $\square$

We will now take a closer look at operators in $l_2(\mathbb{C})^*$. For an operator on T on $l_2(\mathbb{C})$ we have a matrix representation $T = (a_{jk})_{j,k\in\mathbb{N}}$ which satisfies

$$\sum_{k=1}^{\infty} |a_{jk}|^2 < \infty$$

and

$$\sum_{j=1}^{\infty} |a_{jk}|^2 < \infty.$$

For T^* we can extend this matrix representation and from Theorem 3.10 it follows that, for all finite j and $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$, $\sum_{k=\omega}^{\infty} |a_{jk}|^2$ is infinitesimal.

Theorem 3.16. *Let T be a compact linear operator on $l_2(\mathbb{C})$. Then T^* maps norm finite points to near-standard points.*

Proof. Let σ a norm finite point and let $r \in \mathbb{R}_{>0}$ be such that $\|\sigma\|_2 < r$. Then $\sigma \in B_r[0]^*$, where $B_r[0]^*$ is the closed ball of radius r in $l_2(\mathbb{C})^*$. Since T is compact, $A := \overline{T(B_r[0])}$ is compact. Hence $T(\sigma) \in T(B_r[0])^* \subset \overline{T(B_r[0])}^*$. By Theorem 3.12 we find that $T(\sigma)$ is near-standard. $\square$

Theorem 3.17. *Let $T = (a_{jk})$ be a compact linear operator on $l_2(\mathbb{C})$, then for all $j \in \mathbb{N}^*$ we have that a_{jk} is infinitesimal for all $k \in \mathbb{N}^* \setminus \mathbb{N}^\#$.*

Proof. For finite j it this follows directly from the fact that $\sum_{j=k}^{\infty} |a_{jk}|^2$ is infinitesimal. For infinite j consider $e_k = (s_n)$ defined by $s_n = 0$ for all $n \neq k$ and $s_k = 1$, i.e. the kth unit vector. Note that $\|\sigma\|_2 = 1$, hence $T\sigma = (t_j)$ is near-standard by Theorem 3.16. Theorem 3.11 gives us that for all $j \in \mathbb{N}^* \setminus \mathbb{N}^\#$, t_j is infinitesimal. Since $t_j = \sum_{n=1}^{\infty} a_{jn} s_n = a_{jk}$ we find that a_{jk} is infinitesimal. $\square$

Definition 3.18. We call $T = (a_{jk})$ an operator on $l_2(\mathbb{C})$ *almost superdiagonal* if $a_{jk} = 0$ for all $j > k+1$.

It is important to note that being almost superdiagonal depends on the basis.

Theorem 3.19. *Let T be a bounded linear operator on $l_2(\mathbb{C})$ which is almost superdiagonal. Let $m \geq 1$ and let*

$$p(z) = c_0 + c_1 z + ... + c_m z^m$$

be a polynomial of degree m with $c_1, ..., c_m \in \mathbb{C}$ such that $p(T)$ is compact. Then there is an $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ such that $a_{\omega+1,\omega}$ is infinitesimal.

Proof. Let $n \in \mathbb{N}^\#$ and let $T^n = (d_{jk})$.
Claim: For all $h \in \mathbb{N}^*$ we find that $d_{h+n,h} = a_{h+1,h}a_{h+2,h+1}...a_{h+n,h+n-1}$ and for all $l > n$ we have $d_{h+l,h} = 0$.
We will prove this by induction. Note that since T is almost superdiagonal it holds by definition for $n = 1$. Let $N \in \mathbb{N}^\#_{\geq 2}$ and suppose the claim holds for all $n < N$. Let $T^{N-1} = (f_{jk})$ and $T^N = (g_{jk})$. Since $T^N = T^{N-1}T$ for all $h \in \mathbb{N}^*$, we can compute

$$g_{h+N,h} = \sum_{i=1}^{\infty} f_{h+N,i}a_{i,h} = f_{h+N,h+1}a_{h+1,h} = a_{h+1,h}a_{h+2,h+1}...a_{h+N,h+N-1}$$

and

$$g_{h+l,h} = \sum_{i=1}^{\infty} f_{h+l,i}a_{i,h} = 0 \quad (\forall l > N)$$

since $a_{i,h} = 0$ for all $i > h + 1$, $f_{h+N,i} = 0$ for all $i < h + 1$ and $f_{h+l,i} = 0$ for all $i < h + l - N + 1$ by assumption. This proves our claim.

Now let $p(T) = (b_{jk})$. Then we find that $b_{h+m,h} = c_m a_{h+1,h}a_{h+2,h+1}...a_{h+m,h+m-1}$ for all $h \in \mathbb{N}^*$. In particular this holds for $h \in \mathbb{N}^* \setminus \mathbb{N}^\#$. Since $p(T)$ is compact we then find that $b_{h+m,h}$ is infinitesimal by Theorem 3.17. Since c_m is standard and p is of degree m we find that, for some $j \in \{1, ..., m\}$, $a_{h+j+1,h+j}$ is infinitesimal. □

A well known class of operators is the class of projection operators. We will investigate projection operators in $l_2(\mathbb{C})^*$ for which we recall the following definition.

Definition 3.20. Let $E \subset l_2(\mathbb{C})^*$ be an internal closed linear subspace. Then $P : l_2(\mathbb{C})^* \to l_2(\mathbb{C})^*$ is called an *orthogonal projection* on E if $P|_E = id_E$, $P^2 = P$ and $\text{ran}(P)^\perp = \ker(P)$.

Note that every projection operator has norm 1 and thus is continuous.
Since $l_2(\mathbb{C})$ is a Hilbert space we know that we can write $l_2(\mathbb{C})$ as the direct sum of a a linear closed subspace E and its orthogonal complement $E^\perp$. The same holds for $l_2(\mathbb{C})^*$.

Proposition 3.21. *Let $E \subset l_2(\mathbb{C})^*$ be an internal closed linear subspace of $l_2(\mathbb{C})^*$. Then there exists a unique closed linear internal subspace $E^\perp$ such that $l_2(\mathbb{C})^* = E \oplus E^\perp$.*

Proof. First we define the following formula:

$$\text{closed}(E) : (\forall (x_n) \in E^{\mathbb{N}})(\exists x \in l_2(\mathbb{C}))(\lim_{n\to\infty} x_n = x \to x \in E)$$

i.e. E is closed and

$$\text{linear}(E) : (\forall x, y \in E)(\forall \lambda \in \mathbb{C})(x + y \in E \wedge \lambda x \in E)$$

i.e. E is closed under sums and scalar multiplication. Then from standard analysis we know that:

$$\hat{\mathcal{C}} \models (\forall E \subset l_2(\mathbb{C}))(\text{linear}(E) \wedge \text{closed}(E) \to (\exists! F)((\forall x \in E)(\forall y \in F)$$
$$(\langle x, y\rangle = 0) \wedge \text{linear}(F) \wedge \text{closed}(F) \wedge (\forall z \in l_2(\mathbb{C}))(\exists! x \in E)(\exists! y \in F)(x + y = z))).$$

By Łoś' Theorem we are done. □

Note that since $E^\perp$ is unique we can define a function from the closed linear internal subspaces of $l_2(\mathbb{C})^*$ to their orthogonal complement, and thus we may use this in formulas. This also gives, as in standard analysis, that the orthogonal projection on a closed linear subspace is unique. Hence we find an internal function from closed internal linear subspaces of $l_2(\mathbb{C})^*$ to the orthogonal projections. From this we now obtain the following result.

Proposition 3.22. *Let $E \subset l_2(\mathbb{C})^*$ be an internal closed linear subspace and let P be the orthogonal projection on E.Then P is an internal function.*

Proof. Let $n \in \mathbb{N}$ be such that $(l_2(\mathbb{C}) \times l_2(\mathbb{C}))^* \in \mathbb{C}_n^*$ and consider

$$\phi(x, w_1, w_2) : (\exists y \in w_1)(\exists z \in w_2)(\exists w \in w_2^\perp)(w + z = y).$$

Note that $P = \{x \in \mathbb{C}_n^* : \hat{\mathcal{C}}^* \models \phi(x, l_2(\mathbb{C})^*, E)\}$. By Proposition 2.34 P is internal. □

As we could look at the standard part of numbers and sequences, we can also define a standard part of sets of normed vector spaces.

Definition 3.23. Let $H \in \hat{\mathbb{C}}$ be a normed vector space and let $E \subset H^*$ be an internal subset. We call $E^\circ := \{\sigma \in H : \exists \sigma' \in E$ such that $\|\sigma - \sigma'\|$ is infinitesimal$\}$ the *standard part* of E.

Clearly the standard part is unique. Now let E be an internal closed linear subspace of H^*, where H is a normed space in $\hat{\mathbb{C}}$. Then for all $\tau \in E$ we have that $\|\sigma - \tau\| \geq \|\sigma - P_E\sigma\|$ since P_E is an orthogonal projection. From this it follows that $\sigma \in E^\circ$ if and only if $\|\sigma - P_E\sigma\|$ is infinitesimal. Clearly if $\sigma \in E$ is near-standard, then $\sigma^\circ \in E^\circ$.

Theorem 3.24. *If E is a closed linear internal subspace of $l_2(\mathbb{C})^*$, then E° is a closed linear subspace of $l_2(\mathbb{C})$.*

Proof. Let $\sigma_1, \sigma_2 \in E^\circ$ and $\tau_1, \tau_2 \in E$ be such that $\|\sigma_1 - \tau_1\|$ and $\|\sigma_2 - \tau_2\|$ are infinitesimal. Then $\tau_1 + \tau_2$ and

$$\|(\sigma_1 + \sigma_2) - (\tau_1 + \tau_2)\| \leq \|\sigma_1 - \tau_1\| + \|\sigma_2 - \tau_2\|$$

gives that $\|(\sigma_1 + \sigma_2) - (\tau_1 + \tau_2)\|$ is infinitesimal. So $\sigma_1 + \sigma_2 \in E^\circ$. Also for standard $\lambda \in \mathbb{R}^\#$ we have $\|\lambda\sigma_1 - \lambda\tau_1\| = \lambda\|\sigma_1 - \tau_1\|$. Hence $\|\lambda\sigma_1 - \lambda\tau_1\|$ is infinitesimal, and since $\lambda\tau_1 \in E$ it follows that $\lambda\sigma_1 \in E^\circ$. Thus E° is a linear subspace.

To show that $E^\circ$is closed, let $(\sigma_n)_{n\in\mathbb{N}}$ be a sequence in E° converging to some $\sigma \in l_2(\mathbb{C})$. Consider $(\sigma_n)^*$ the non-standard extension in $l_2(\mathbb{C})^*$ and note that $\|\sigma_n - P_E\sigma_n\|$ is infinitesimal for all finite n. Then Theorem 3.6 gives that there is an $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ such that $\|\sigma_n - P_E\sigma_n\|$ is infinitesimal for all $n < \omega$. Moreover, since $\lim_{n\to\infty} \sigma_n = \sigma$, Theorem 3.5 gives that $\|\sigma_n - \sigma\|$ is infinitesimal for all $n \in \mathbb{N}^* \setminus \mathbb{N}^\#$. Therefore, for all $n \in \mathbb{N}^* \setminus \mathbb{N}^\#$ with $n < \omega$, we find

$$\|\sigma - P_e\sigma_n\| \leq \|\sigma - \sigma_n\| + \|\sigma_n - P_E\sigma_n\|,$$

which gives us that $\|\sigma - P_e\sigma_n\|$ is infinitesimal for such n. Since $P_E\sigma_n \in E$ we conclude that $\sigma \in E^\circ$, thus E° is closed. $\square$

In $l_2(\mathbb{C})^*$ we can find some natural subspaces for $n \in \mathbb{N}^*$, namely:

$$H_n := \{(s_n) \in l_2(\mathbb{C}) : s_n = 0 \text{ for all } n > \omega\}.$$

We will denote its corresponding orthogonal projection by P_n, or, if there is no confusion, simply by P.
Note that for $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ we have that for all $\sigma \in l_2(\mathbb{C})$ we have
$\|\sigma - P_\omega\sigma\| = (\sum_{n=\omega+1}^{\infty} |s_n|^2)^{\frac{1}{2}}$, which is infinitesimal by Theorem 3.10.
For a bounded operator T on $l_2(\mathbb{C})$, we denote $T' = P_\omega T P_\omega$ and $T_\omega = T'|_{H_\omega}$. Note that $\|T'\| = \|P_\omega\|^2\|T\| \leq \|T\|$ and hence $\|T_\omega\| \leq \|T\|$.

Theorem 3.25. *Let $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ and let E be an internal closed linear subspace of H_ω that is invariant for T_ω. Then E° is invariant for T.*

Proof. Let $\sigma \in E^\circ$ and let $\tau \in E$ be such that $\tau^\circ = \sigma$. Then $PT\tau = T_\omega\tau \in E$, hence $T\sigma \in E^\circ$ if $\|T\sigma - PT\tau\|$ is infinitesimal. Consider:

$$\begin{aligned}\|T\sigma - PT\tau\| &= \|T\sigma - PT\sigma + PT(\sigma - \tau)\| \\ &\leq \|T\sigma - PT\sigma\| + \|P\|\|T\|\|\sigma - \tau\|\end{aligned}$$

Since $\|T\|$ and $\|P\|$ are finite and $\|\sigma-\tau\|$ is infinitesimal we have that $\|P\|\|T\|\|\sigma-\tau\|$ is infinitesimal. Also since $T\sigma \in l_2(\mathbb{C})$ we find that $\|T\sigma - PT\sigma\|$ is infinitesimal. Hence $\|T\sigma - PT\tau\|$ is infinitesimal. $\square$

We will now take a closer look at the dimension of subspaces of H_ω, as described in Section 3.1.

Theorem 3.26. *Let $\omega \in \mathbb{N}^*\setminus\mathbb{N}^\#$ and E_1 and E_2 be admissible closed linear subspaces of H_ω such that $E_1 \subset E_2$ and $\dim(E_2) = \dim(E_1 + 1)$. Then $E_1^\circ \subset E_2^\circ$ and any two points of E_2° are linearly dependent modulo E_1°.*

Note that in the non-standard case we can talk about sums of dimension since H_ω has finite dimension, and hence E_1 and E_2 have finite dimension. In the standard case we have to be more careful since $H_\omega^\circ = H$, and thus E_1° and E_2° might have infinite dimension.

Proof. Note that since $E_1 \subset E_2$ evidently $E_1^\circ \subset E_2^\circ$. For the other statement suppose $\sigma_1, \sigma_2 \in E_2^\circ$ are linearly independent modulo E_1°. Let $\tau_1, \tau_2 \in E_2$ points infinitely close to σ_1, σ_2 respectively. Note that $\tau_1, \tau_2 \notin E_1$, since otherwise $\sigma_1 \equiv 0 \bmod E_1^\circ$ or $\sigma_2 \equiv 0 \bmod E_1^\circ$. Since $\dim(E_2) = \dim(E_1) + 1$ we now find $\lambda \in \mathbb{C}^*$ and $\tau \in E_1^*$ such that

$$\tau_2 = \lambda\tau_1 + \tau.$$

Note that λ is not infinitesimal, since then $\sigma_2 = \tau_2^\circ = \tau^\circ \in E_1^\circ$. Suppose λ were infinitely large. Then $\frac{1}{\lambda}$ would be infinitesimal, hence

$$\sigma_1 = \tau_1^\circ = (\frac{1}{\lambda}\tau_2 + \frac{1}{\lambda}\tau)^\circ = (\frac{1}{\lambda}\tau)^\circ \in E_1^\circ.$$

Thus we find that λ has non-zero standard part.
Claim: $\sigma_2 - \lambda^\circ \sigma_1 \in E_1$.
Let $\sigma = \sigma_2 - \lambda^\circ \sigma_1$. Consider

$$\begin{aligned} \|\tau - \sigma\| &= \|\tau_2 - \lambda\tau_1 - \sigma_2 + \lambda^\circ \sigma_1\| \\ &\leq \|\tau_2 - \sigma_2\| + |\lambda| \|\tau_1 - \sigma_1\| + |\lambda - \lambda^\circ| \|\sigma_1\|. \end{aligned}$$

Hence $\|\tau - \sigma\|$ is infinitesimal. So τ has a standard part which naturally belongs to E_1°. □

3.4 The Theorem of Bernstein-Robinson

The invariant subspace problem is a partially open problem in Functional Analysis. The question is whether for a bounded operator $T \in \mathcal{H}$, where H is a Banach space of dimension larger than one, there exists a non-trivial closed linear T-invariant subspace of H. Though P.H. Enflo gave a counter-example for H a Banach space in 1987, the problem remains open for Hilbert spaces. The problem originated from a problem raised by P.R. Halmos in 1963 in [3], in which it was credited to K.T. Smith. A generalization of the problem was solved three years later by A.R. Bernstein and A. Robinson in [1] using non-standard analysis. In the same journal, directly after the proof of Bernstein and Robinson, Halmos gave a proof for his own problem using only standard methods [4], though clearly inspired by the proof of Bernstein and Robinson which he reviewed for the journal. We have done most of the work already in the previous sections, but here we will give the main proof of the theorem.

Theorem 3.27 (Bernstein-Robinson). *Let T be a bounded linear operator on an infinite-dimensional Hilbert space H over the complex numbers and let $p(z) \neq 0$ be a polynomial with complex coefficients such that $p(T)$ is a compact operator. Then there exists a closed subspace of H, not equal to H or $\{0\}$, which is invariant under T.*

Proof. The proof relies on the fact that if H is finite dimensional, we can find a chain of closed T-invariant subspace of H

$$\{0\} = E_0 \subset E_1 \subset ... \subset E_n = H$$

where E_i is i-dimensional. This can be achieved by looking at the eigenspaces. Our finite-dimensional space will be $(l_2(\mathbb{C})^*)_\omega$ where $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$. First, however, we must prove that we only have to consider the case $H = l_2(\mathbb{C})$.

Claim: If there exists a non-zero $\sigma \in H$ such that $A_\sigma := \{\sigma, T\sigma, T^2\sigma, ...\}$ does not span H, then $\overline{\text{span}(A_\sigma)}$ is a non-trivial T-invariant closed subspace.
Clearly $\overline{\text{span}}(A)$ is a non-trivial closed subspace. So we have to prove that it is T-invariant. For this consider $x \in \overline{\text{span}}(A)$. We can find $\lambda_i \in \mathbb{C}$ for $i \in \mathbb{N}$ such that

$$x = \sum_{i=0}^{\infty} \lambda T^i \sigma.$$

Hence, by linearity and continuity of T,

$$Tx = \sum_{i=0}^{\infty} \lambda T^{i+1}\sigma \in \overline{\text{span}}(A).$$

Hence $\overline{\text{span}}(A)$ is T-invariant.

Now we can use that if H is not separable it does not have a countable basis, hence for all $\sigma \in H$ we find that A_σ does not span H since A_σ is countable. Therefore we only have to consider infinite-dimensional separable Hilbert spaces which we know to be isomorphic to $l_2(\mathbb{C})$.

With our claim it now is easy to see that if there is a σ such that A_σ is linearly dependent we can find a non-trivial closed T-invariant subspace. Therefore, from now on we will assume that $H = l_2(\mathbb{C})$, $\|\sigma\| = 1$ and A_σ spans H and linearly independent. We can apply the Gram-Schmidt method to A_σ to replace it with an orthonormal set $B = \{\sigma = \eta_0, \eta_1, \eta_2, ...\}$, and since A_σ is linearly independent B is an orthonormal basis of H. Also note that $\text{span}\{\sigma, T\sigma, ..., T^n\sigma\} = \text{span}\{\eta_0, ..., \eta_n\}$, therefore we find that for all $n \in \mathbb{N}$ we have that

$$T\eta_n \in \text{span}\{\sigma, T\sigma, ..., T^n\sigma\} = \text{span}\{\eta_0, ..., \eta_n\},$$

thus T is super-diagonal with respect to basis B. We can represent T in matrix form by taking $T = (a_{jk})$. Now passing to the realm of non-standard analysis by Theorem 3.19 we find that there exist a $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$ such that $a_{\omega+1,\omega}$ is infinitesimal. Keep this ω fixed and consider H_ω, P its orthogonal projection and $T' = PTP$ as seen before in section 3.3.

Let $\xi = (x_i) \in H^*$ be norm-finite and consider:

$$(TP - T')\xi = (I - P)TP\xi =: \zeta = (z_n)$$

By the super-diagonality of T we find that $z_n = 0$ for all $n \neq \omega + 1$ and $z_{\omega+1} = a_{\omega+1,\omega}x_\omega$. Hence $\|\zeta\| \leq |a_{\omega+1,\omega}|\|\xi\|$, i.e. ζ is infinitesimal. Thus we have that $TP\xi \approx T'\xi$ where by continuity of T and P both $TP\xi$ and $T'\xi$ are norm-finite.

Claim: $T^rP\xi \approx (T')^r\xi$ for al norm-finite ξ and all $r \in \mathbb{N}$ and both are norm-finite. We just proved the case $r = 1$. Let $R \in \mathbb{N}_{>1}$ be arbitrary and suppose the claim holds for all $r < R$ and all norm-finite ξ. Then

$$T^RP\xi \approx T(T')^{R-1}\xi = TP(T')^{R-1}\xi \approx T'(T')^{R-1}\xi = (T')^R\xi,$$

and again by the continuity of T and P both are norm finite. Thus by induction the claim holds.

Now applying our claim multiple times we see that, for all norm-finite $\xi \in H_\omega$ and polynomial q, we have

$$q(T)P\xi \approx q(T').$$

Let $T_\omega = T'|_{H_\omega}$. Since H_ω is 'finite dimensional' we can find a chain of closed T_ω-invariant internal subspaces

$$\{0\} = E_0 \subset E_1 \subset \ldots \subset E_\omega = H_\omega$$

where E_i is i-dimensional for all $0 \leq i \leq \omega$. Note that E_i is also a closed subspace of H^*. Let P_i be the orthogonal projection from H^* onto E_i. Let $p(z)$ be as in Theorem 3.26 and let $\xi \in H$, $\xi \neq 0$. Then $p(\xi) \neq 0$ since A_ξ is finitely linear independent by assumption. We choose $\xi \in H$ with $\|\xi\| = 1$. As we saw earlier $P\xi \approx \xi$, so $p(T)\xi \approx p(T)P\xi \approx p(T')\xi$, hence $p(T')\xi$ is not infinitesimal, i.e. there is some $0 < r \in \mathbb{R}^\#$ such that $\|p(T')\xi\| > r$. For all $0 \leq i \leq \omega$ we define

$$r_i = \|p(T')\xi - p(T')P_i\xi\|.$$

Note that $r_i \leq \|p(T')\|\|\xi - P_i\xi\|$. Also we have $r_0 = \|p(T')\xi\|$ hence $r_0 > r$ and $\|\xi - P_\omega\xi\|$ is infinitesimal. Thus $r_\omega < \frac{r}{2}$.

Claim: There exists a $\lambda \in \mathbb{N}^*_{\leq\omega}$ such that $\frac{r}{2} \leq r_{\lambda-1}$ and $r_\lambda < \frac{r}{2}$.
Clearly $(r_i)_{0\leq i\leq\omega}$ is a decreasing sequence so the obvious candidate is the smallest i such that $r_i \geq \frac{r}{2}$. Such an i exists since $\mathbb{N}^*$ is well-ordered if $\{i \in \mathbb{N}^* : r_i \geq \frac{r}{2}\}$ is an internal set.
For this note that we can view the chain $E_0 \subset \ldots \subset E_\omega$ as an internal function f with domain $\mathrm{dom}(f) = \{0, \ldots, \omega\}$ and $f(i) = E_i$. We know there is an internal function from the closed linear subspaces to the corresponding orthogonal projections, hence by Proposition 2.37 we find an internal function $\hat{f}$ given by $\hat{f}(i) = P_i$ for all $0 \leq i \leq \omega$.
Now by Proposition 2.38 and 3.15 we find that $\tilde{f}$ given by $\tilde{f}(i) = \|p(T')\xi - p(T')P_i\xi\| = r_i$ for all $0 \leq i \leq \omega$ is an internal function. Now by an easy application of Proposition 2.34 we find that $\{i \in \mathbb{N}^* : r_i \geq \frac{r}{2}\}$ is indeed internal, thus proving our claim.

Now let $\lambda \in \mathbb{N}^*_{\leq\omega}$ be such that $\frac{r}{2} \leq r_{\lambda-1}$ and $r_\lambda < \frac{r}{2}$. We will prove that either $E^\circ_{\lambda-1}$ or E°_λ is a T-invariant non-trivial closed linear subspace of H. Note that, by Theorem 3.24 and 3.25, $E^\circ_{\lambda-1}$ and E°_λ are T-invariant closed linear subspaces of H. So left to show is that one of them is non-trivial.
For this note that $E^\circ_{\lambda-1} \neq H$. If not, then $\xi \in E^\circ_{\lambda-1}$, so $\|\xi - P_{\lambda-1}\xi\|$ would be infinitesimal. Thus $\frac{r}{2} \leq r_{\lambda-1} \leq \|p(T')\|\|\xi - P_{\lambda-1}\xi\|$ is infinitesimal which is a contradiction.
Also note that $E^\circ_\lambda \neq \{0\}$. To see this consider $\eta = p(T')P_\lambda\xi$. Note that $\eta \in E_\lambda$ since $P_\lambda\xi \in E_\lambda$ and E_λ is $p(T_\omega)$-invariant, thus $p(T')$-invariant. Also since $P_\lambda\xi \in H_\omega$ we find

$$\eta = p(T')P_\lambda\xi \approx p(T)P_\lambda\xi.$$

Here $p(T)P_\lambda\xi$ is near-standard by Theorem 3.16, since $P_\lambda\xi$ is norm finite and $p(T)$ is compact. Hence η has a standard part η°. Suppose $\eta^\circ = 0$, then η would be infinitesimal. And thus

$$r_\lambda \geq \|P(T')\xi\| - \|p(T')P_\lambda\xi\| > r - \|\eta\|,$$

which contradicts that $\frac{r}{2} > r_\lambda$. Hence indeed $E_\lambda^\circ \neq \{0\}$. So we find the only way for both of them to be trivial is if $E_{\lambda-1} = \{0\}$ and $E_\lambda = H$. However, by Theorem 3.26 their dimension cannot differ more than one, which proves the theorem. $\square$

Seven years after this proof, in 1973, Lomonosov proved the even stronger Theorem which said that for every non-trivial compact operator T there exist a non-trivial closed linear subspace which is S-invariant for every operator S which commutes with T. [7, p. 269] This is a generalization of the Theorem of Bernstein and Robinson since every operator commutes with a polynomial of itself.

4 The Hahn-Banach Theorem

4.1 Non-standard maps

Non-standard analysis has another very strong tool at its disposal. For this we will take a deeper dive into logics and take a closer look at non-standard maps. From now on we will consider more abstract models and filters again.
Let $\hat{M}$ be a superstructure of some model M and let $* : \hat{M} \to \hat{M}^*, x \mapsto [(x)]$ be a non-standard map.

Definition 4.1. We call $*$ an κ-*enlargement* for some cardinal number κ if, for all non-empty systems $\mathcal{A}$ consisting of sets $A \in \hat{M}$ with the finite intersection property (i.e. all finite intersections are non-empty) and cardinality at most κ, we have that

$$\bigcap \mathcal{A}^\# = \bigcap \{A^* : A \in \mathcal{A}\} \neq \emptyset.$$

We call $*$ an *enlargement* if $*$ is a κ-enlargement for every κ.

Note that $\mathcal{A}$ is not per definition an element of $\hat{M}$. For example

$$\mathcal{A} := \{C_n : n \in \mathbb{N}\} \notin \hat{\mathbb{C}}.$$

Luckily we may ignore such systems.

Lemma 4.2. *To prove that $*$ is a κ-enlargement one only has to consider systems $\mathcal{A} \in \hat{M}$.*

Proof. To see this, take $A_0 \in \mathcal{A}$ and consider

$$\mathcal{A}_0 := \{A \cap A_0 : A \in \mathcal{A}\}.$$

Evidently $\mathcal{A}_0$ inherits the finite intersection property, has cardinality less than the cardinality of $\mathcal{A}$ and is non-empty. Since

$$\emptyset \neq \bigcap \mathcal{A}_0^\# = \bigcap \mathcal{A}^\# \cap A_0 = \bigcap \mathcal{A}^\#$$

we find that $*$ is a κ-enlargement if $\bigcap \mathcal{A}^\# \neq \emptyset$ for all $\mathcal{A} \in \hat{M}$ of cardinality less than κ with the finite intersection property. $\square$

Enlargement are very useful due to their connection to binary relations.

Definition 4.3. We call a binary relation ϕ *satisfied by* $b \in \text{ran}(\phi)$ *on* $A \subset \text{dom}(\phi)$ if for all $a \in A$ we have that $(a, b) \in \phi$.
We call ϕ *concurrent on* $A \subset \text{dom}(\phi)$ if for all finite $A_0 \subset A$ there is some $b \in \text{ran}(\phi)$ that satisfies ϕ on A_0. If ϕ is concurrent on $\text{dom}(\phi)$ we just call ϕ *concurrent.*

Note the similarity between a concurrent binary relation and the finite intersection property. This is something we can utilize.

Theorem 4.4. *The following are equivalent:*

(i) $*$ *is a* κ*-enlargement.*

(ii) *For any concurrent binary relation* $\phi \in \hat{M}$ *with* $|\text{dom}(\phi)| \leq |\kappa|$ *we have that* $\text{dom}(\phi)^{\#}$ *is satisfied by some* $b \in \text{ran}(\phi)^*$.

(iii) *For any* $A \in \hat{M}$ *of cardinality at most* κ *there is a* $*$*-finite* $B \in \hat{M}^*$ *(i.e.* B *has cardinality* $\{1, ..., \omega\}$ *for some* $\omega \in \mathbb{N}^*$*) such that* $A^{\#} \subset B \subset A^*$.

Proof. We start with proving (i)⇒(ii). So assume (i) and let $\phi \in \hat{M}$ be a concurrent binary relation with domain of cardinality at most κ. Let, for all $d \in \text{dom}(\phi)$,

$$A_d := \{y \in \text{ran}(\phi) : (d, y) \in \phi\},$$

and let $\mathcal{A} = \{A_d : d \in \text{dom}(\phi)\}$. Clearly $\mathcal{A}$ has cardinality at most κ and it has the finite intersection property since ϕ is concurrent. Hence $\bigcap \mathcal{A}^{\#} \neq \emptyset$, so there is a $b \in \bigcap \mathcal{A}^{\#} \subset \text{ran}(\phi)^*$ such that $(d, b) \in \phi$ for all $d \in \text{dom}(\phi)$, i.e. b satisfies $\text{dom}(\phi)^{\#}$.
We will now prove (ii)⇒(iii), so assume (ii) and let $A \in \hat{M}$ be of cardinality at most κ. We define the binary relation $\phi \subset A \times \mathcal{P}(A)$ by

$$\phi := \{(a, b) : a \in b \wedge b \text{ is finite}\}.$$

Clearly ϕ is concurrent, hence there is a $b \in \mathcal{P}(A)^*$ satisfying ϕ^* on $A^{\#}$.
Finally we prove (iii)⇒(i). Thus assume (iii) and let $\mathcal{A}$ be a non-empty system of entities $A \in \hat{M}$ of cardinality at most κ and with the finite intersection property. We may assume by Lemma 4.2 that $\mathcal{A} \in \hat{M}$. Thus there exists a $*$-finite $\mathcal{B}$ such that $\mathcal{A}^{\#} \subset \mathcal{B} \subset \mathcal{A}^*$. Since $\mathcal{A}^*$ has the $*$-finite intersection property we find that $\bigcap \mathcal{A}^{\#} \supset \bigcap \mathcal{B} \neq \emptyset$. □

We will see an application of this theorem in a proof of the Hahn-Banach Theorem. Before we do this we first have to find out whether these enlargements actually exist.

Definition 4.5. Let λ be a cardinal number. We call a filter $\mathcal{F}$ of a set J λ*-adequate* if for each non-empty family $\mathcal{A}$ of subsets of λ with the finite intersection property there exists a map $f : J \to \lambda$ such that for each $A \in \mathcal{A}$ there exists an $F \in \mathcal{F}$ with $f(F) \subset A$.

Theorem 4.6. *For each cardinal number* λ *there exists a* λ*-adequate ultrafilter* $\mathcal{F}$ *for an appropriate set* J.

Proof. Let J be the set of all finite collections of subsets of λ. For each $A \subset \lambda$ we define

$$F_A = \{j \in J : A \in j\}.$$

Then $\mathcal{F}_0 := \{F_A : A \subset \lambda\}$ has the finite intersection property. Indeed if we consider $A_1, ..., A_n \subset \lambda$ then $j = \{A_1, ..., A_n\}$ is in the intersection $\bigcap_{i=1}^n F_{A_i}$. We then find that

$$\tilde{F_0} := \{X : \text{there exists a finite subset } \mathcal{B} \subset \mathcal{F}_0 \text{ such that } \bigcap \mathcal{B} \subset X)\}$$

is a filter which we can extend to an ultrafilter $\mathcal{F}$. We will now prove that $\mathcal{F}$ is λ-adequate. For this let $\mathcal{A}$ be a non-empty family of subsets of λ with the finite intersection property. We define

$$J_{\mathcal{A}} := \{j \in J : B_{j,\mathcal{A}} := \bigcap_{A \in j \cap \mathcal{A}} A \neq \emptyset\}$$

and we define, using the axiom of choice, $f : J \to \lambda$ by $f(j) \in B_{j,\mathcal{A}}$ if $j \in J_{\mathcal{A}}$ and $f(j)$ arbitrarily otherwise. This function works. To see this let $A \in \mathcal{A}$. Then $f(F_A) \subset A$ since, for all $j \in F_A$ we have that $A \in j$, thus $B_{j,\mathcal{A}} \neq \emptyset$ since $\mathcal{A}$ has the finite intersection property. Hence $f(j) \in B_j \subset A$. Thus $\mathcal{F}$ is λ-adequate. $\square$

Theorem 4.7. *If $\mathcal{F}$ is an $\hat{M}$-adequate filter, then the corresponding map $*$ is an enlargement.*

Proof. Let $\mathcal{A}$ be a system of sets $A \in \hat{M}$ with the finite intersection property. Note that $\mathcal{A}$ has at most cardinality $\hat{M}$. We have to prove that there exists some $b \in \hat{S}^*$ such that $b \in A^*$ for all $A \in \mathcal{A}$. Recall, if we write $b := [(b_j)_{j \in J}]$, that $b \in A^*$ if and only if

$$\{j \in J : b_j \in A\} \in \mathcal{F}.$$

Since $\mathcal{F}$ is $\hat{M}$-adequate, there exists a map $f : J \to \hat{M}$ such that for each $A \in \mathcal{A}$ there exists an $F \in \mathcal{F}$ such that $f(F) \subset A$. So if we take $b = [(f(j))_{j \in J}]$ we find that that for all $A \in \mathcal{A}$ there exists an $F \in \mathcal{F}$ such that $f(F) \subset A$. Hence

$$F \subset \{j \in J : f(j) \in A\}.$$

By definition of filters this means that $\{j \in J : f(j) \in A\} \in \mathcal{F}$ and hence $b \in A^*$. $\square$

4.2 The Hahn-Banach Theorem

We will now go ahead and prove the Hahn-Banach theorem using the theory of the previous sections.

Theorem 4.8 (Hahn-Banach)**.** *Let X be a normed vector space over $\mathbb{K}$, where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, let $Y \subset X$ be a linear subspace and let $T : Y \to \mathbb{K}$ be a bounded linear functional. Then there exists a linear functional $S : X \to \mathbb{K}$ such that $S|_Y = T$ and $\|S\| = \|T\|$.*

Before proving the theorem we will prove that we can extend T in a finite way.

Lemma 4.9. *Let X be a normed vector space over $\mathbb{R}$, $Y \subset X$ a linear subspace, let $T : Y \to \mathbb{R}$ be a bounded linear functional and let $x_1, ..., x_n$ be elements of $X \setminus Y$. Then there exists a linear functional $S : \operatorname{span}(x_1, ..., x_n, Y) \to \mathbb{C}$ such that $\|S\| = \|T\|$ and $S_Y = T$.*

Proof. Note that, since we can extend T one element at a time, it is enough to prove this for $n = 1$. So let $x \in X \setminus Y$. Note that x is linear independent of Y, thus we can write each $u \in \operatorname{span}(x, Y)$ uniquely as $u = y + \lambda x$ where $\lambda \in \mathbb{R}$ and $y \in Y$. Since the only way of extending T linearly is by putting $S(u) = T(y) + \lambda r$ we only have to find an appropriate $r \in \mathbb{R}$, i.e. an $r \in \mathbb{R}$ such that $\|S\| = \|T\|$. So we have to prove there exists an r such that $|T(y) + \lambda r| \leq \|T\|\|y + \lambda x\|$ for all $y \in Y$ and $\lambda \in \mathbb{R}$. Since $\lambda = 0$ is trivial we can divide by λ or just assume that $\lambda = 1$. Hence we have to prove there exists a r such that

$$-\|T\|\|y + x\| - T(y) \leq r \leq \|T\|\|y + x\| - T(y)$$

for all $y \in Y$. To see that such an r exists, consider that, for all $y, z \in Y$, we have

$$\begin{aligned} T(z) - T(y) = T(z - y) &\leq \|T\|\|z - y\| = \|T\|\|z + x - (x + y)\| \\ &\leq \|T\|(\|z + x\| + \|x + y\|), \end{aligned}$$

from which we find that

$$\sup_{y \in Y} -\|T\|\|y + x\| - T(y) \leq \inf_{z \in Y} \|T\|\|z + x\| - T(z).$$

Hence we can find an r between the two values, proving our lemma. □

We will also use a result on suprema.

Proposition 4.10. *Let $r \in \mathbb{R}^{\#}$ and $A \subset \mathbb{R}^*$ such that for all $x \in A$ we have $x < r$ or $x \approx r$. Then $\sup_{x \in A} x < r$ or $\sup_{x \in A} x \approx r$.*

Proof. Suppose $\sup_{x \in A} x > r$ and $\sup_{x \in A} x \not\approx r$. Then there exists $y \in (r + \epsilon, \sup_{x \in A} x) \cap A$ for some $\epsilon \in \mathbb{R}^{\#}_{>0}$ which contradicts with the assumptions. □

Proof of the theorem of Hahn-Banach. We will now prove the Hahn-Banach Theorem in two stages. First we will give a non-standard proof for the case $\mathbb{K} = \mathbb{R}$ and with that result we will prove Hahn-Banach for the case $\mathbb{K} = \mathbb{C}$.
So let $T : Y \to \mathbb{R}$ a linear functional on $Y \subset X$. We define

$$\mathfrak{F}(X, \mathbb{R}) := \{S \subset X \times \mathbb{R} : S \text{ is a map}\},$$

and consider the binary relation

$$\begin{aligned} \phi = \{(x, y) \in X \times \mathfrak{F}(X, \mathbb{R}) : &\operatorname{dom}(y) \text{ is a linear subspace} \wedge \\ &x \in \operatorname{dom}(y) \wedge y \text{ is a linear functional} \wedge T \subset y \wedge \|y\| \leq \|T\|\}. \end{aligned}$$

Note that this relation is concurrent by Lemma 4.9, hence we can apply Theorem 4.4 to find a non-standard linear functional $S \in \mathfrak{F}(X, \mathbb{R})^*$ such that S satisfies ϕ^* on $X^\#$, i.e. $X^\# \subset \mathrm{dom}(S), T^* \subset S$ and $\|S\| \leq \|T\|$. From this we find that for all $x \in X^\#$ we have

$$|S(x)| \leq \|S\|\|x\| \leq \|T\|\|x\|$$

which is finite, so $S(x)^\circ$ exists for all $x \in X^\#$.
Claim: $U : X \to \mathbb{R}, x \mapsto S(x)^\circ$ is a linear functional, $\|U\| = \|T\|$ and $T \subset U$.
Since $S(x) = T(x)$ for all $x \in Y$ we find that $U(x) = S(x)^\circ = T(x)$ for all $x \in Y$ so $T \subset U$.
Also note that for all $x, y \in X$ and $\lambda \in \mathbb{R}$ we have that

$$U(x + y) = S(x + y)^\circ = (S(x) + S(y))^\circ = S(x)^\circ + S(y)^\circ = U(x) + U(y)$$

and

$$U(\lambda x) = S(\lambda x)^\circ = \lambda^\circ S(x)^\circ = \lambda U(x,)$$

so U is linear. Finally note that $|U(x)| \approx |S(x)|$. Then by Proposition 4.10 it holds that

$$\sup_{x \in X, \|x\| \leq 1} |U(x)| \leq \sup_{x \in X^*, \|x\| \leq 1} |S(x)| = \|S\|.$$

Hence $\|U\| = \|T\|$, proving the Hahn-Banach Theorem for $\mathbb{K} = \mathbb{R}$.
Now suppose $\mathbb{K} = \mathbb{C}$. By considering only multiplication by real scalars we can view X as a real vector space. We define $T_\mathbb{R} : Y \to \mathbb{R}, x \mapsto \mathrm{Re}(T(x))$. Then we can extend $T_\mathbb{R}$ on the entire space to a real functional S. Now define $U(x) = S(x) - iS(ix)$. Since S is linear and $U(ix) = S(ix) - iS(-x) = S(ix) + iS(x) = iU(x)$ we find that U is linear. Finally we look at the norm of U an we find that

$$\begin{aligned}\|U\| = \sup_{x \in X, \|x\| \leq 1} |U(x)| = \sup_{x \in X, \|x\| = 1} \sup_{|\lambda| = 1} \mathrm{Re}(\lambda U(x)) = \\ \sup_{x \in X, \|x\| = 1} \sup_{|\lambda| = 1} \mathrm{Re}(U(\lambda x)) \leq \sup_{x \in X, \|x\| = 1} |S_\mathbb{R}(x)| = \|S\| = \|T_\mathbb{R}\| = \|T\|.\end{aligned}$$

□

5 The Theorem of Hille-Yosida

5.1 Non-standard topology and limits

We will now move on to study non-standard analysis and semigroups. For this we will again assume that $\mathcal{F}$ is a non-principal ultrafilter over $\mathbb{N}$. Before we can start on semigroups, we will need some results on non-standard topology and non-standard limits. For convenience we will again assume that $\mathcal{F}$ is an ultafilter over $\mathbb{N}$ containing the Fréchet filter.

Proposition 5.1. *Let X be a normed vector space over $\mathbb{K}$, where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$. Then $C \subset X$ is closed if and only if $C = \{x \in X^\# : (\exists y \in C^*)(x \approx y)\}$.*

Proof. Suppose C is closed. Let $x \in X^{\#}$ and suppose there exists a $y \in C^*$ such that $x \approx y$. Let y be represented by $[(y_n)]$. Since for all $\epsilon \in \mathbb{R}^{\#}$ we have $\|x - y\| < \epsilon$ and $y \in C$ we find that

$$F_\epsilon = \{n \in \mathbb{N} : y_n \in C \wedge \|x - y_n\| < \epsilon\} \in \mathcal{F}.$$

Hence we can construct a sequence (z_n) in C converging to x by picking $z_n = y_m$ where $m = \min(F_{\frac{1}{n}})$. Thus, since C is closed, $x \in C$.

Now suppose $C = \{x \in X^{\#} : (\exists y \in C^*)(x \approx y)\}$. Let (y_n) be a sequence in C converging to some $x \in X$.
Claim: $y := [(y_n)] \in C^*$ is infinitely close to x.
To see this let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{N}$ be such that for all $n > N$ we have $\|x - y_n\| < \epsilon$. Hence

$$F_\epsilon = \{n \in \mathbb{N} : \|x - y_n\| < \epsilon\} \in \mathcal{F}$$

since it has a finite complement. So $\|x - y\| < \epsilon$. Because this holds for all $\epsilon \in \mathbb{R}^{\#}_{>0}$ we find that $y \approx x$ and clearly $y \in C^*$. Hence $x \in C$ and so C is closed. □

Proposition 5.2. *Let X be a normed space and let $A \subset X$. Then $x \in \overline{A}$ if and only if there exists a $y \in A^*$ such that $y \approx x$.*

Proof. From Proposition 5.1 we know that

$$\overline{A} = \{x \in X^{\#} : (\exists y \in \overline{A}^*)(x \approx y)\}.$$

By Łośwe have

$$(y \in \overline{A}^*) \leftrightarrow ((\exists z = (z_n) \in (A^{\mathbb{N}})^*)(\lim_{n \to \infty} z_n = y)).$$

Thus there exists some $\omega \in \mathbb{N}^*$ such that $y \approx z_\omega$ is infinitesimal. Since $z_\omega \in A^*$ we find

$$\overline{A} = \{x \in X^{\#} : (\exists y \in A^*)(x \approx y)\}.$$

□

Corollary 5.3. *Let X be a normed space and let $A \subset B \subset X$, with B closed. Then A is dense in B if and only if for all $x \in B$ there exists a $y \in A^*$such that $y \approx x$.*

This is a direct consequence of Proposition 5.2.

Proposition 5.4. *Let X be a normed vector space, let $R \subset \mathbb{R}$ and let $T : R \to X$ be continuous. Then, for all $y \in \overline{R}$ for which there is a $\delta > 0$ such that $(y, y + \delta) \subset R$, we have that $\lim_{x \downarrow y} T(x) = c \in X$ if and only if, for all $x \in R^*$ such that $x \approx y$ and $x > y$, it holds that $T(x) \approx c$.*

Proof. Let $y \in \overline{R}$ for which there is a $\delta > 0$ such that $(y, y + \delta) \subset R$.
Suppose $\lim_{x \downarrow y} T(x) = c \in X$, then for all $\epsilon \in \mathbb{R}_{>0}$ there is a $\delta \in \mathbb{R}_{>0}$ such that for all $x \in (y, y + \delta)$ we have $\|T(x) - c\| < \epsilon$. In particular we find for all $x \in R^*$ such that $y \approx x$ and $x > y$, that $x \in (y, y + \delta)^*$ and thus for all $\epsilon \in \mathbb{R}^{\#}_{>0}$ by Łoś' Theorem

we have $\|T(x) - c\| < \epsilon$ and so $T(x) \approx c$.

Now suppose that for all $x \in R^*$, $x \approx y$ and $x > y$ it holds that $T(x) \approx c$. Then we have that for all $\epsilon \in \mathbb{R}^{\#}_{>0}$

$$\mathcal{M}^* \models (\exists \delta \in \mathbb{R}^*_{>})((y, y+\delta) \subset R^* \wedge (\forall x \in (y, y+\delta))(\|T(x) - c\| < \epsilon))$$

Hence by Łoś' Theorem we find that $\lim_{x \downarrow y} T(x) = c$. □

From this we can also do the other limits.

Corollary 5.5. *Let X be a normed vector space, let $R \subset \mathbb{R}$ and let $T : R \to X$ be continuous. Then for all $y \in \overline{R}$ for which there is a $\delta > 0$ such that $(y - \delta, y) \subset R$, we have that $\lim_{x \uparrow y} T(x) = c \in X$ if and only if, for all $x \in R^*$ such that $x \approx y$ and $x < y$, it holds that $T(x) \approx c$.*

Corollary 5.6. *Let X be a normed vector space, let $R \subset \mathbb{R}$ and let $T : R \to X$ be continuous. Then, for all $y \in \overline{R}$ for which there is a $\delta > 0$ such that $(y-\delta, y+\delta) \subset R$, we have that $\lim_{x \downarrow y} T(x) = c \in X$ if and only if, for all $x \in R^*$ such that $x \approx y$, it holds that $T(x) \approx c$.*

Proposition 5.7. *Let X and Y be normed vector spaces over $\mathbb{K}$ and $T \in \mathcal{L}(X, Y)$. Then the following hold.*

(i) *If $y \in X$ is near-standard, then $T(y^\circ) = (Ty)^\circ$.*

(ii) *If $(x_n)_{n \in \mathbb{N}^*}$ is a near-standard sequence in X^*. i.e. if x_n is near-standard for all $n \in \mathbb{N}^*$ and $\lim_{n \to \infty} x_n = x$, then x is near-standard and $x^\circ = \lim_{n \to \infty} x_n^\circ$.*

Proof. We will prove (i) first. Note that if y is near-standard there exists a $z \in X^*$ such that $\|z\| \in \mathbb{R}_0$ and $y = y^\circ + z$. Now consider $Ty = T(y^\circ + z) = Ty^\circ + Tz$. Note that $\|Tz\| \leq \|T\|\|z\| \in \mathbb{R}_0$ since $\|z\|$ is infinitesimal and $\|T\|$ is finite. So Ty is near-standard. Furthermore $(Ty)^\circ = (Ty^\circ + Tz)^\circ = (Ty^\circ)^\circ + (Tz)^\circ = Ty^\circ$.

To prove (ii) note that, since $\lim_{n \to \infty} x_n = x$, there is an $N \in \mathbb{N}^*$ such that $\|x - x_n\|$ is infinitesimal for all $n \geq N$. Then the left-hand side of

$$\|x - x_N^\circ\| \leq \|x - x_N\| + \|x_N - x_N^\circ\|$$

is infinitesimal since the right-hand side is. Hence x is near-standard. Also note that the sequence $(x_n^\circ)_{n > N}$ is constant, hence $x^\circ = \lim_{n \to \infty} x_n^\circ$. □

Proposition 5.8. *Let X and Y be Banach spaces over $\mathbb{K}$. Let (B_n) be a bounded sequence in $\mathcal{L}(X, Y)$. Let $D \subset X$ be a dense subset of X, and assume that for all $x \in D$ the sequence $(B_n x)$ is convergent. Then $Bx := \lim_{n \to \infty} B_n x$ exists for all $x \in X$, and $B : X \to Y$, thus defined, is an operator $B \in \mathcal{L}(X, Y)$.*

Proof. We will first prove that Bx exists for all $x \in X$. For this let $x \in X$. Since $X = \overline{D}$ by Proposition 5.2 there is a $y \in D^*$ such that $x \approx y$. Note that for large enough $m, n \in \mathbb{N}^*$ we have

$$\begin{aligned} \|B_n x - B_m x\| &\leq \|B_n x - B_n y\| + \|B_n y - B_m y\| + \|B_m y - B_m x\| \\ &\leq (\|B_n\| - \|B_m\|)\|x - y\| + \|B_n y - B_m y\| \approx 0 \end{aligned}$$

since $\lim_{n\to\infty} B_n y = By$. Hence $(B_n x)$ is a Cauchy sequence and, since Y is a Banach space, converges. So we can define $Bx = \lim_{n\to\infty} B_n x$. Evidently B is also a bounded operator since the sequence (B_n) is bounded. The linearity of B follows from the linearity of the limit. $\square$

5.2 Semigroups

Semigroups are designed to give solutions to the well-known differential equation $\frac{d}{dt} f(t) = \lambda f(t)$. However instead considering the equation over $\mathbb{R}$ or $\mathbb{X}$ we consider $\frac{d}{dt} T(t) = AT(t)$ where $T : [0, \infty) \to \mathcal{L}(X)$ where X is a Banach space and A is a linear operator in X, i.e. $\text{dom}(A) \subset X$ and $\text{ran}(A) \subset X$. If $A \in \mathcal{L}(X)$ it is not hard to find a solution for this equation. Namely

$$T(t) = e^{tA} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}.$$

However, if $\text{dom}(A) \neq X$ or if A is unbounded it is unclear if there even is a solution, let alone what kind of solution. We will give an introduction to semigroups to prove the Theorem of Hille-Yosida, which will give us sufficient conditions to find a solution.

Definition 5.9. Let X be a Banach space, a *one-parameter semigroup* on X is a function $T : [0, \infty) \to \mathcal{L}(X)$ satisfying

(a) $T(t + s) = T(t)T(s)$, for all $t, s \geq 0$.

If additionally

(b) $\lim_{t\downarrow 0} T(t)x = x$, for all $x \in X$,

we call T a C_0*-semigroup* or a *strongly continuous semigroup*.

For the rest of this article we will simply write 'semigroup' instead of 'one-parameter semigroup'. Note that by this definition $T(0)$ is either 0, which implies that $T(t) = 0$ for all $t > 0$ since $T(t) = T(t + 0) = T(t)T(0)$, or $T(0) = I$.

As already noted we consider linear operators in X which do not necessarily have X as domain. For those who never encountered such operators we will look into them in detail.

Definition 5.10. Let X be a Banach space. We call A a *linear operator* in X if A is linear, $\operatorname{dom}(A) \subset X$ and $\operatorname{ran}(A) \subset X$. Instead of 'linear operator', we will usually just write 'operator'. We call A *bijective* or *invertible* if A is injective and $\operatorname{ran}(A) = X$. We define $A^{-1} = \{(x, y) \in X \times \operatorname{dom}(A) : Ay = x\}$. Note that $\operatorname{ran}(A^{-1}) = \operatorname{dom}(A)$. We call

$$\|A\| := \sup_{x \in \operatorname{dom}(A),\ \|x\| \leq 1} \|Ax\|$$

the *operator norm.* If A and B are both operators in X we can define $A+B$ by taking $\operatorname{dom}(A+B) = \operatorname{dom}(A) \cap \operatorname{dom}(B)$ and $(A+B)x = Ax + Bx$ for all $x \in \operatorname{dom}(A+B)$. We also define a norm on $X \times X$ by

$$\|(x, y)\|_{X \times X} = \|x\|_X + \|y\|_X.$$

A is called *closed* if it is a closed subset of $X \times X$ under the norm $\| \cdot \|_{X \times X}$.

Proposition 5.11. *Let X be a Banach space and let A be a standard bounded operator in X and let B and C be internal operators in X such that $B \approx C$. Then, if $\operatorname{ran}(B) \subset \operatorname{dom}(A)$ and $\operatorname{ran}(C) \subset \operatorname{dom}(A)$, $AB \approx AC$ and, if $\operatorname{ran}(A) \subset \operatorname{dom}(B)$ and $\operatorname{ran}(A) \subset \operatorname{dom}(C)$, $BA \approx CA$.*

Proof. Suppose $\operatorname{ran}(B) \subset \operatorname{dom}(A)$ and $\operatorname{ran}(C) \subset \operatorname{dom}(A)$. Let $x \in \operatorname{dom}(AB - AC)$ with $\|x\| \leq 1$ and consider

$$0 \leq \|ABx - ACx\| \leq \|A\| \|B - C\| \|x\| \in \mathbb{R}_0.$$

Since $\|A\|$ and $\|x\|$ are finite and $\|B - C\|$ is infinitesimal.
Hence by Proposition 4.10

$$\sup_{x \in \operatorname{dom}(AB - AC), \|x\| \leq 1} \|ABx - ACx\| \approx 0$$

and thus $AB \approx AC$.
For the other statement suppose $\operatorname{ran}(A) \subset \operatorname{dom}(B)$ and $\operatorname{ran}(A) \subset \operatorname{dom}(C)$ and note that

$$0 \leq \|BAx - CAx\| \leq \|A\| \|B - C\| \|x\| \in \mathbb{R}_0$$

and hence by the same arguments $BA \approx CA$. □

Proposition 5.12. *Let T be a C_0-semigroup on X and let $\tau \in \mathbb{R}^*_{\geq 0}$ be finite and let $t = \tau^\circ \in \mathbb{R}^\#$. Then $T(t) \approx T(\tau)$.*

Proof. It is enough to prove that $T(\delta) \approx I$ for all $0 < \delta \in \mathbb{R}_0$ since then for all finite $\tau \in \mathbb{R}^*$ we have that, if $\tau^\circ \leq \tau$,

$$T(\tau) = T(\tau^\circ + (\tau - \tau^\circ)) = T(\tau^\circ) T(\tau - \tau^\circ) \overset{5.11}{\approx} T(\tau^\circ),$$

and, if $\tau^\circ > \tau$,

$$T(\tau^\circ) = T(\tau + (\tau^\circ - \tau)) = T(\tau) T(\tau^\circ - \tau) \overset{5.11}{\approx} T(\tau^\circ).$$

That $T(\delta) \approx I$ is a direct consequence of Proposition 5.4. □

Lemma 5.13. *Let T be a semigroup on X such that there exists $\delta > 0$ such that $M := \sup_{0 \leq t < \delta} \|T(t)\| < \infty$. Then there exists $\omega \in \mathbb{R}$ such that*

$$\|T(t)\| \leq Me^{\omega t} \text{ for all } t \geq 0.$$

For the proof of this lemma see [2]

Proposition 5.14. *Let T be a C_0-semigroup on X. Then the following hold:*

(a) *There exists an $M \geq 0$ and $\omega \in \mathbb{R}$ such that*

$$\|T(t)\| \leq Me^{\omega t} \text{ for all } t \geq 0$$

(b) *For all $x \in X$ the function $[0, \infty) \to X, t \mapsto T(t)x$ is continuous.*

Proof. We start by proving a. By Proposition 5.13 we only have to prove that there exists $0 < \delta \in \mathbb{R}^{\#}$ such that $\sup_{0 \leq t < \delta} \|T(t)\| < \infty$. Suppose such δ does not exists. Then we can find a sequence $(t_n)_{n \in \mathbb{N}}$ in $\mathbb{R}$ such that $\lim_{n \to \infty} \|T(t_n)\| = \infty$ and $\lim_{n \to \infty} t_n = 0$.
Then $\tau = [(t_n)] \in \mathbb{R}^*$ is infinitesimal, hence $T(\tau) \approx I$. In particular $\|T(\tau)\| \approx 1$ which is a contradiction with $\lim_{n \to \infty} \|T(t_n)\| = \infty$ which implies that $\|T(\tau)\|$ should be infinitely large.

To prove b, by Corollary 5.6 we only have to prove that for all $x \in X$, for all $0 < \delta \in \mathbb{R}_0$ and for all $0 < t \in \mathbb{R}^{\#}$, we have $T(t+\delta)x \approx T(t)x \approx T(t-\delta)x$.
Thus let $x \in X$, $t > 0$ and $\delta \in \mathbb{R}_0$. Then $T(t+\delta)x = T(t)T(\delta)x \approx T(t)x$ and $T(t-\delta)x \approx T(t-\delta)T(\delta)x = T(t)x$. □

Though semigroups are an interesting field of study on their own, most people will study them to solve differential equations. For this the C_0-semigroups are very important since in a way we can differentiate them.

Definition 5.15. Let T be a C_0-semigroup on X. We call

$$A := \{(x, y) \in X \times X : y = \lim_{h \downarrow 0} h^{-1}(T(h)x - x) \text{ exists}\}.$$

the *generator* of $T(t)$

Note that A is an operator in X due to the linearity of limits. As promised we can now differentiate T.

Proposition 5.16. *Let T be a C_0-semigroup and A its generator. Then:*

(i) *For all $x \in \text{dom}(A)$ and for all $t \in (0, \infty)$ we have that $T(t)x \in \text{dom}(A)$, $T(t)x$ is differentiable and*

$$\frac{d}{dt} T(t)x = AT(t)x = T(t)Ax.$$

(ii) $\mathrm{dom}(A)$ *is dense and A is a closed operator.*

Proof. We will first prove (i). Let $0 < \delta \in \mathbb{R}_0$, and let $x \in \mathrm{dom}(A)$. We know that $Ax = \lim_{h\downarrow 0} h^{-1}(T(h)x - x)$ exists, so Proposition 5.4 then gives us $\delta^{-1}(T(\delta)x - x) \approx Ax$. Since $T(t)$ is bounded, by Proposition 5.14 we may use Proposition 5.11 and 5.12 to find

$$\delta^{-1}(T(t+\delta)x - T(t)x) = T(t)\delta^{-1}(T(\delta)x - x) \approx T(t)Ax$$

and

$$-\delta^{-1}(T(t-\delta)x - T(t)x) = T(t-\delta)\delta^{-1}(T(\delta)x - x) \approx T(t)Ax.$$

Thus by Corollary 5.6 we have $\frac{d}{dt}T(t)x = T(t)Ax$. Finally note that

$$T(t)h^{-1}(T(h)x - x) = h^{-1}(T(h) - I)T(t)x.$$

Since we know that this converges for $h \downarrow 0$ we find that $T(t)x \in \mathrm{dom}(A)$ and that $T(t)Ax = AT(t)x$.

Proving (ii) will take more theory than we have right now. People interested can find more information in [2]. $\square$

Finally we will take a look at an often used trick in semigroups, which is shifting semigroups.

Proposition 5.17. *Let T be a semigroup on X. Then $T_\lambda(t) = T(t)e^{\lambda t}$ is a semigroup on x. Furthermore if $\|T(t)\| \leq Me^{\omega t}$ with M and ω as in Lemma 5.13, then $\|T(t)\| \leq Me^{(\omega+\lambda)t}$. Additionally, if T is strongly continuous, with generator A, then T_λ is strongly continuous with generator $A + \lambda I$.*

The proof of this is elementary, hence we leave it to the reader.

Definition 5.18. Let T be a semigroup on X. We call T contractive if $\|T(t)\| \leq 1$ for all $t \in [0, \infty)$.

Contractive semigroups are one of the main reasons we want to be able to shift semigroups since they are far easier to work with. In several proofs you first shift a semigroup to a contractive one, work with that, and then shift it back. We will see an example of this in the proof of the Hille-Yosida Theorem.

5.3 The Hille-Yosida Theorem

Before we can finally prove the Theorem of Hille-Yosida, we take a glimpse into the world of resolvents. Resolvents and the spectrum of an operator play a major role in Functional Analysis, with fundamental results such as the Spectral Mapping Theorem [7, p. 263]. We will only take a look at a small part of the relation they have with semigroups since they are very important in the proof of the Hille-Yosida Theorem.
Let X be a Banach space over $\mathbb{K}$ and let A be an operator in X.

Definition 5.19. We call

$$\rho(A) := \{\lambda \in \mathbb{K} : \lambda - A \text{ is bijective and } (\lambda - A)^{-1} \in \mathcal{L}(X)\}$$

the *resolvent set of* A, $R(\lambda, A) := (\lambda - A)^{-1}$ the *resolvent of* A *at* λ, and

$$R(\cdot, A) : \rho(A) \to \mathcal{L}(X)$$

the *resolvent of* A. The set $\sigma(A) = \mathbb{K} \setminus \rho(A)$ is called the *spectrum of* A.

Definition 5.20. Let $T(t)$ and $S(t)$ be internal semigroups. We call $T(t)$ and $S(t)$ *infinitely close* if $T(t) \approx S(t)$ for all finite $t \in \mathbb{R}^*$. We write $S \approx T$. We call S *near-standard* if T is a standard semigroup, and we denote the standard part of $S(t)$ by $S(t)^\circ = T(t)$.

Proposition 5.21. *Let $S(t)$ be an internal semigroup such that for all $t \in [0, \infty)^\#$ there is a $T(t) \in \mathcal{L}(X)^\#$ such that $S(t) \approx T(t)$. Then $T(t)$ is a semigroup and, additionally, if $S(t)$ is a contractive semigroup, then so is $T(t)$.*

Proof. First note that $T(t) : [0, \infty) \to X$ and that for all $s, t \in \mathbb{R}$ we have

$$T(s + t) \approx S(s + t) = S(s)S(t) \overset{5.11}{\approx} T(s)T(t).$$

Thus T is a semigroup in $\mathcal{M}$ and therefore also in $\mathcal{M}^*$.
Suppose additionally that $S(t)$ is contractive. Then $\|S(t)\| \leq 1$ for all $t \in [0, \infty)^*$. In particular for all $t \in [0, \infty)^\#$ we find that $\|T(t)\| \approx \|S(t)\|$ and thus $\|T(t)\| \leq 1$ or $\|T(t)\| \approx 1$. Since T and t are standard we find that $\|T(t)\| \leq 1$ for all $t \in [0, \infty)$. $\square$

Theorem 5.22. *Let T be a C_0-semigroup on X, let A be its generator and let $M \geq 1$ and $\omega \in \mathbb{R}$ be such that $\|T(t)\| \leq Me^{\omega t}$. Then $\{\lambda \in \mathbb{K} : Re(\lambda) > \omega\} \subset \rho(A)$.*

We will not prove this theorem since the proof is not non-standard; those interested can find it in [2, p. 42]. Armed with this knowledge we can now prove the Theorem of Hille-Yosida.

Theorem 5.23 (Hille-Yosida)**.** *Let A be a closed densely defined operator in X. Assume that there exists $\omega \in \mathbb{R}$ such that $(\omega, \infty) \subset \rho(A)$ and that*

$$\|R(\lambda, A)\| \leq \frac{1}{\lambda - \omega} \quad (\lambda \in (\omega, \infty)).$$

Then A is the generator of a C_0-semigroup T satisfying the estimate

$$\|T(t)\| \leq e^{\omega t} \quad (t \geq 0).$$

Proof. Note that due to rescaling we may assume that $\omega = 0$.
Consider the Yosida approximations:

$$A_n = A(I - \frac{1}{n}A)^{-1} = nAR(n, A) = n^2 R(n, a) - nI \in \mathcal{L}(X).$$

Since $R(n,a)$ is bounded we easily find that A_n is bounded, hence e^{tA_n} is a C_0-semigroup. Also, using the made assumptions, we can obtain the estimate

$$\|e^{tA_n}\| = \|e^{tn^2R(n,a)-nI)}\| = e^{-tn}\|\sum_{k=0}^{\infty}\frac{(tn^2R(n,a))^k}{k!}\| \leq e^{-tn}\sum_{k=0}^{\infty}\frac{(tn)^k}{k!} = 1$$

Thus e^{tA_n} is a contractive C_0-semigroup.
Using Proposition 5.21 we only have to prove that e^{tA_ω} is near-standard for some $\omega \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ and that A is the generator of $(e^{tA_\omega})^\circ$.
So let $\omega \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ and let $t \in [0,\infty)^{\#}$. We find that, for all $m,n \in \mathbb{N}^*$and all $x \in \mathrm{dom}(A)^*$, we have

$$\begin{aligned} e^{tA_m}x - e^{tA_n}x &= \int_0^t \frac{d}{ds}(e^{(t-s)A_n}e^{sA_m}x)ds \\ &= \int_0^t e^{(t-s)A_n}(A_m - A_n)e^{sA_m}xds \\ &= \int_0^t e^{(t-s)A_n}e^{sA_m}(A_m - A_n)xds. \end{aligned}$$

Since e^{tA_n} is contractive we find that

$$\|e^{tA_m}x - e^{tA_n}x\| \leq t\|(A_m - A_n)x\|.$$

For $n \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ we have that $I - \frac{1}{n}A \approx I$, thus by Proposition 5.11 we find

$$A(I-\frac{1}{n}A)^{-1} = A(I-\frac{1}{n}A)^{-1}I \approx A(I-\frac{1}{n}A)^{-1}(I-\frac{1}{n}A) = A.$$

Hence for $m,n \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$ we find

$$\begin{aligned} \|(A_m - A_n)x\| &= \|(A(I-\frac{1}{m}A)^{-1} - A + A - A(I-\frac{1}{n}A)^{-1})x\| \\ &\leq \|(A(I-\frac{1}{m}A)^{-1} - A)x\| + \|(A - A(I-\frac{1}{n}A)^{-1})x\| \approx 0 \end{aligned}$$

Since t is finite, $e^{tA_n}x$ is a Cauchy sequence, and since X is a Banach space, this sequence converges for all $x \in \mathrm{dom}(A)$. Since A is densely defined we may now apply Proposition 5.8 to find that $\lim_{n\to\infty} e^{tA_n} \in \mathcal{L}(X)$. This means that, for all $t \in [0,\infty)^{\#}$ and for all $\omega \in \mathbb{N}^* \setminus \mathbb{N}^{\#}$, e^{tA_ω} is near-standard and thus, by Proposition 5.21, $T(t) = (e^{tA_\omega})^\circ$ is a contractive semigroup.
To see that $T(t)$ is strongly continuous we will focus on a small interval of the semigroup. Consider $C([0,1];X)$ the Banach space of continuous function from $[0,1]$ to X equipped with the supremum norm. Since e^{tA_n} is strongly continuous we find

$$\mathcal{T}_n : X \to C([0,1];X), x \mapsto [t \mapsto e^{tA_n}x].$$

We now can use our inequality from above to find

$$\|\mathcal{T}_m x - \mathcal{T}_n x\| \leq \|(A_m - A_n)x\| \quad (x \in \mathrm{dom}(A)).$$

Again by taking $m, n \in \mathbb{N}^* \setminus \mathbb{N}^\#$ we find that $\mathcal{T}_n x$ is a Cauchy sequence and thus by Theorem 5.8 $\mathcal{T}_n$ converges to some standard $\mathcal{T} : X \to C([0,1]; X)$. Note that for standard $t \in [0,1]^\#$ we have that $(\mathcal{T}_n x)(t) = e^{tA_n}x$, which converges to $T(t)x$, hence for the standard limit we find

$$\lim_{t \downarrow 0} T(t)x = \lim_{t \downarrow 0} \mathcal{T}x(t) = \mathcal{T}x(0) = x.$$

Thus we find that $T(t)$ is strongly continuous.
Finally we have to prove that the generator of $T(t)$ is A. For this let B be the generator of $T(t)$, let $0 < h \in \mathbb{R}_0$ and let $x \in \mathrm{dom}(A)$ and consider

$$\|\frac{T(h)x - x}{h} - Ax\|.$$

We want to prove that this is infinitesimal. Firstly recall that $\|A_\omega x - Ax\|$ is infinitesimal for all $\omega \in \mathbb{N}^* \setminus \mathbb{N}^\#$. Also recall that $(\mathcal{T}_\omega)_{\omega \in \mathbb{N}^*}$ converges to $T(t)|_{[0,1]}$, hence for large enough $\omega \in \mathbb{N}^*$ we find that $\left\|\frac{T(h)x - e^{tA_\omega}x}{h}\right\|$ is infinitesimal. Then we find

$$\begin{aligned}
\left\|\frac{T(h)x - x}{h} - Ax\right\| &= \left\|\frac{T(h)x - e^{tA_\omega}x}{h} + \frac{e^{tA_\omega}x - x}{h} - A_\omega x + A_{\omega x} - Ax\right\| \\
&\leq \left\|\frac{T(h)x - e^{tA_\omega}x}{h}\right\| + \left\|\frac{e^{tA_\omega}x - x}{h} - A_\omega x\right\| + \|A_{\omega x} - Ax\| \\
&\approx \left\|\frac{(\sum_{k=0}^\infty (hA_\omega)^k k!)x - x}{h} - A_\omega x\right\| \\
&= \left\|\frac{(\sum_{k=2}^\infty (hA_\omega)^k k!)x}{h}\right\| \\
&\leq \sum_{k=2}^\infty \frac{(h\|A_\omega\|)^k}{k!h}\|x\| \\
&= h\sum_{k=2}^\infty \frac{h^{k-2}\|A_\omega\|^k}{k!}\|x\| \\
&\leq he^{\|A_\omega\|} \approx 0.
\end{aligned}$$

Hence $\|\frac{T(h)x - x}{h} - Ax\|$ is infinitesimal. Since this holds for all $0 < h \in \mathbb{R}_0$ we find that $\lim_{h \downarrow 0} \frac{T(h)x - x}{h} = Ax$. Hence $A \subset B$.
Of course we want equality. For this note that by assumption $(0, \infty) \subset \rho(A)$ and by Theorem 5.22 we have that $(0, \infty) \subset \rho(B)$. Thus, since $I - A \subset I - B$, $I - B$ is injective and $\mathrm{ran}(I - A) = X$, we find that $I - A = I - B$ and thus $A = B$. $\square$

References

[1] Bernstein, A.R. & Robinson, A. 1966, 'Solution of an invariant subspace problem of K.T. Smith and P.R. Halmos', *Pacific journal of Mathematics*, Vol. 16, No. 3, pp. 421-431

[2] Engel, K.J. & Nagel, R. 2006, *A Short Course on Operator Semigroups*, Springer Science & Business Media, New York

[3] Halmos, P.R. 1963, 'A glimpse into Hilbert space' *Lectures on Modern Mathematics*, Vol. 1, pp. 1-22

[4] Halmos, P.R. 1966, 'Invariant subspaces of polynomially compact operators', *Pacific journal of Mathematics*, Vol. 16, No. 3, pp. 433-437

[5] Luxemburg, W.A. 1973, 'What is Nonstandard Analysis?', *The American Mathematical Monthly*, Vol. 80, No. 6, part 2, pp.38-67

[6] Mendelson, E. 2009, *Introduction to Mathematical Logic, Fifth Edition*, Chapman & Hall/CRC, Boca Raton

[7] Rudin, W. 2007, *Functional Analysis, Second Edition*, Tata McGraw-Hill, Hightstown

[8] Väth, M. 2007, *Nonstandard Analysis*, Birkhäuser Verlag, Basel