Nonstandard Analysis and representation of reality

Claude Lobry[*] Tewfik Sari[†]

March 2007

Abstract

The aim of this paper is to show that the representation with the help of Nonstandard Analysis of a real phenomenon, presenting different observation scales, allows an important simplification of language. Indeed, it is convenient to have available the concept of infinitely small and infinitely large quantities in dealing with the macroscopic effects of microscopic phenomena. This is illustrated on two examples : the representation of two time scales systems and the representation of noise.

Key words : Non standard analysis, stabilization, peaking, moiré, averaging, integration on finite sets, noise
AMS Classification : 03H10, 28E05, 34E18, 93D20, 94A12.

Contents

[*]Projet MERE, INRA-INRIA, claude.lobry@inria.fr
[†]Mulhouse University, tewfik.sari@uha.fr

Foreword by Claude Lobry

The present paper has been written by Tewfik Sari and myself on the basis of notes that I prepared for the conference in honor of Michel Fliess. Below is the introduction (in French) that I wrote in view of my oral presentation.

Il y a un peu plus de vingt ans Marc Diener et moi éditions un ouvrage collectif "Analyse non standard et représentation du réel" [16] dont j'ai repris le titre pour cette conférence car la question me semble toujours d'actualité. Dans les années mille neuf cent quatre-vingt il était difficile d'évoquer l'Analyse Non Standard sans déclencher des passions : Il y avait les "contre" et les "pour" dont je faisais partie. Maintenant que la passion est moins forte il semble possible d'aborder cette question de façon un peu plus scientifique. C'est ce que je vais tenter de faire en analysant sur deux exemples choisis en automatique la capacité de l'Analyse Non Standard à traduire formellement des discours mathématiques informels.

Le premier exemple que j'ai choisi est celui du phénomène de "peaking" et de son influence sur la stabilisation des systèmes ; il est tiré d'un travail commun avec T. Sari [25, 27, 28] et je me sens assez à l'aise pour en parler. En revanche, je me sens bien moins qualifié pour parler du second exemple puisqu'il touche au traitement du signal, discipline que je ne connais pas. J'ai cependant décidé d'en parler à cause de son actualité dans le cadre de ces journées en l'honneur de Michel Fliess.

En effet, peu de temps avant le colloque j'ai découvert, avec un immense plaisir, qu'il a l'intention d'aborder des questions de traitement du signal en *proposant une représentation du "bruit" par des fonctions "rapidement oscillantes", cette dernière expression étant définie de façon précise dans le cadre Non Standard [19]. Comme je l'ai déjà dit, je ne connais rien au traitement du signal, donc je ne peux pas avoir un avis scientifiquement fondé sur l'avenir de cette idée. Ce qui ne m'enpêche pourtant pas d'y croire et donc de vouloir lui faire de la publicité.*

Qu'est ce donc qui me donne cette foi du charbonnier ? C'est simplement le fait, maintes fois constaté au long de sa carrière remarquablement féconde, que chaque fois que Michel Fliess a eu une idée, cette dernière a fini par s'imposer. Je ne vois pas pourquoi il n'en serait pas de même encore une fois !

There is a long way from a text prepared for an oral presentation and a publication in a journal. A first version was prepared in French with my accomplice T. Sari and submitted by us. In view of the comments of the editor we tried to give the present English version (but the french original one is accessible at ....).

1 Introduction

Let us say some words about the title : Why "Nonstandard Analysis and Representation of Reality" ? Nonstandard Analysis (NSA) is a way to practice mathematical analysis in which it is perfectly legal to say that a real number is "infinitesimal and fixed for ever". By "legal" we mean that, considered from the point of view of the mathematical rigor, accordingly to the most recent canons of logic, there is nothing to say against such a sentence. By the way, any proof through NSA of a classical mathematical result can be transformed into a classical proof, which led some people to say that NSA does not produce anything new. This is not true. What NSA brings is a new way to formalize a mathematical discourse into a new formalism which is proved to be not more dangerous than the classical one. More precisely if there were a contradiction in mathematics formalized within NSA there would be one in mathematics formalized in the classical sense. Thus the question is not to discuss whether NSA is rigorous or not but rather to see if this new practice is more suitable than the ancient one.

The first criteria is "fertility" : How many new results where obtained thanks to NSA ? How many old conjectures have been demonstrated ? Very few, almost none. Does this disqualify definitively NSA ? Certainly not. Presently only one percent mathematicians use NSA in their research. Thus the disproportion is too large in order that global comparisons make sense. In some particular domains, like ordinary differential equations, the report is far from being equal to zero.

A second criteria, which is less decisive since it is more subjective, is the criteria of "elegance". Elegance of the style was always preoccupation for scientists, especially for mathematicians. We think that for certain questions the NSA style is more elegant than the classical one and it is on this ground that we shall put ourself in this paper. More precisely we shall try to have the point of view of a non mathematician user, more specifically an engineer in automatic control, for whom mathematics are a tool for representation and understanding of concrete phenomena.

In Automatic control science the matter is to observe in real time the system in order to get informations from which one deduces the way to achieve some goal. To achieve this purpose we built a "model", that is to say a "mathematical representation" of the real system which we are interested in. We think that when the real phenomenon presents different observation scales, a representation with the help of NSA allows an important simplification of language. But, what is the price to pay for this simplification ? Do the necessary efforts made to manipulate correctly the tools of NSA not annihilate the expected benefit ? The answer is no. With very little effort any specialist in automatic control can provide to himself, in addition to classical mathematical tools, a language that is very well fitted to the expression of certain ideas. This is, at least, what we shall try to demonstrate on two examples : The *representation of two time scale systems*, on one side, the *representation of noise*, on the other side. Our objective being purely pedagogical none of the presented results is really new except, perhaps, the idea of writing in "bold " the "external notions" with the hope to facilitate the understanding of this formalism. In any case we hope that our presentation will push some readers to see by themselves what is going on. In particular they will find in the collective works [14, 18] many other applications of the method we are advocating for.

The paper is divided in three parts. The first part presents, without any particular previous knowledge, the necessary tools of NSA ; the few examples given are introductory to further notions which will be exposed in parts two and three. The second part is devoted to the "peaking phenomenon" and the third one to the question of "noise".

2 Nonstandard Analysis

Nonstandard Analysis was invented by A. Robinson [39]. Our background is the theory developed by E. Nelson. This is not the only way to proceed but it is the one we know !

The system IST (for Internal Set Theory) of E. Nelson [34] is a formal language $\mathcal{L}$ which contains the language of set theory ZFC (standing for Zermelo, Fraenkel + axiom of Choice). It contains the non defined unitary predicate "st" which writes "st(x)" in the formulas and reads "x is standard". A set of three axioms : I for "Idealization", S for "Standardization" and T for "Transfer" define the rules of manipulation of the new predicate. E. Nelson proves that the theory IST is "conservative" and by the way is irreproachable from the point of view of rigor : if there would be a contradiction in IST the contradiction would be already present in ZFC. It is not the place in this paper to describe in details the three axioms and their immediate consequences but one must emphasize the "double language" that IST permits within the practice that was widely spread by G. Reeb and its school (see Sections 2.5 and 3.3).

In $\mathcal{L}$ one makes the distinction between the formulas which do not contain the predicate "st", which are called *internal* ("internal" to the language of set theory) from those which contain it, called *external*. Actually, when one writes mathematics, one does not write formulas of the formal language but writes in a technical language precise enough (at least when things are well done) to indicate a path which could go to the formal writing. By the way, since the first months of University, the young student knows that the sentence *when* t *tends to* 0 *the function* f *tends to* l is formalized into :

$$\forall\epsilon > 0\ \exists\eta > 0\ \ \forall t\ \ \{|t| \leq \eta \implies |f(t) - l| \leq \epsilon\}.$$

But all the art of formalism, when it is well understood, is to make allusion to formulas which could be written and write them as least as possible.

As a consequence, in E. Nelson's system, it is essential, when one uses expressions, to know whether they are internal or external and to take care that something which was in the classical system a loosely way of speaking could become in IST perfectly rigorous in the sense that the path toward a formal text is perfectly shown. Only a manual (for instance [12, 15, 30, 38]) and some experience are able to make you familiar with this practice. Nevertheless it is not too difficult to give an idea of it as we shall do now.

2.1 The "large" integers

From the Transfer axiom one derives that 0 is standard and that the successor of a standard integer is a standard integer. Thus 1 is standard, 2 is standard, and so on. From the axiom of Idealization one deduces[1] :

$$\exists\,\omega \in \mathbb{N}\ \ \forall^{\mathrm{st}} n \in \mathbb{N}\ \ \ n \leq \omega. \tag{1}$$

This integer ω which is greater than any standard integer deserves to be called "infinitely large" and also all those like it (for instance $\omega + 1$) which are greater than any standard integer. By the way, saying "infinitely large" is not a loosely way of speaking used to evoke some classical formulation but merely an abbreviation of an external formula (the formula (1) is external). In order to help the understanding we use bold characters formulations for sentences in the natural language which have an immediate external formal translation in the system IST. Thus, by the way the expresion :

ω is an **infinitely large** integer

is just a synonym for the external formula :

$$\forall^{\mathrm{st}} n \in \mathbb{N}\ \ \ n \leq \omega.$$

Of course, we give to this expression all the affective connotation contained in the idea of something which is "greater than everything". The choice of

[1] One uses the abbreviation : $\forall^{\mathrm{st}} x\ \ A(x)$ for : $\forall x\ \ \{\mathrm{st}(x) \Longrightarrow A(x)\}$

such a terminology can be criticized and was criticized. In fact, in the classical language, that we are anxious to preserve, a set is "finite" if there is a bijection between this set and the set $\{1, 2, \cdots, n\}$ where n is some natural number. Thus the following subset of $\mathbb{N}$:

$$\{0, 1, 2, \cdots, \omega\}$$

is finite, even in the case where ω is **infinitely large**. Thus in the system of E. Nelson the sentence :

Let ω be an **infinitely large** integer : the set $\{0, 1, 2, ..., \omega\}$ is finite,

is perfectly correct. This might be disrupting but not more than the discovery by a student that in a topological space a set can be both open and closed. It could have been wise to choose an other terminology but the tradition to which we belong decided that way and we conform to it. This tradition has the advantage to emphasize that what is metaphysical and disrupting is not the existence of **infinitely large** integers but rather the existence of *infinite sets* in ZFC. This relates to the discussions at the beginning of the twentieth century on the "actual infinite" and the possibility of "intuitionistic" foundations of NSA, which is an other subject[2].

Finally notice the important fact that external formulas may define sets in the *intuitive* or *naïve* sense which are not *formal* sets. Thus the external formula :

$$n \in \mathbb{N} \ \wedge \ \mathrm{st}(n)$$

defines no "set of standard numbers". This fact is at the origin of some powerful tools that we shall explain later : the so called "overspill" or "permanency" principles.

2.2 The infinitesimals

Let us forget our previous metaphysical considerations and let us come to the "infinitely small" which, according to our feeling, carries a less heavy emotional load than "infinitely large". A real number is said to be **infinitely large** if it is greater than some **infinitely large** integer. An **infinitesimal** (or an **infinitely small** real number) will be a real number either equal to zero or having an **infinitely large** inverse number. Two real numbers are **infinitely close** if their difference is **infinitesimal**. A real number which is not **infinitely large** is called **limited**. A real number which is **limited** and not **infinitesimal** is called **appreciable**. We also call "halo" of x the collection **hal**(x) of real numbers which are **infinitely close** to x : Note that it is not a formal set in IST.

[2] For more details the reader is refered to G. Reeb's paper *La mathématique non standard, vieille de soixante ans ?* This paper was first published in October 1979, in the *Series de Mathématiques Pures et Appliquées*, IRMA, Strasbourg. It appeared in 1999 as as an appendix of J-M. Salanskis's book *Le constructivisme non standard* [40], pages 277-289. A short version of this paper, including a discussion of the mathematical theory of the moiré phenomenon, was already published in 1981 [37].

Thus we can work with *actual* **infinitesimals**. Within the system IST it is perfectly legal to say :

Let ϵ be a strictly positive fixed **infinitesimal**.

The sum of two **infinitesimals** is as expected **infinitesimal**. The famous paradox of infinitesimals :

> If ϵ is infinitesimal, 2ϵ must also be infinitesimal an so on.... Let n_0 be the last integer such that $n_0\epsilon$ is infinitesimal. The sum of the two infinitesimals $n_0\epsilon + \epsilon$ is not an infinitesimal, which causes the trouble,

which was solved during the nineteenth century by the method of the "passage through the limit" is solved in IST by the interdiction which is made to consider the collection of all integers n such that $n\epsilon$ is **infinitesimal** as a true set and, as a consequence, to take its largest value n_0.

2.3 Continuous functions

Infinitesimal calculus was created for the purposes of analysis and, in particular, to define continuity. We introduce the notation $x \approx y$ to say that x is **infinitely close** to y. The following definition is an external definition.

Definition 2.1 (**S-continuity**) *A function f is* ***S-continuous*** *if x* ***infinitely close*** *to y implies $f(x)$* ***infinitely close*** *to $f(y)$, that is to say :*

$$\forall x \ \ \forall y \, (x \approx y \Longrightarrow f(x) \approx f(y)).$$

In order to understand what is the meaning of this definition let us take $a > 0$ and consider the piecewise constant function defined by :

$$x \in [ka, (k+1)a[\implies f(x) = ka. \tag{2}$$

This function is not continuous, as everybody knows. If a is standard, it is a standard function. It is easy to check that f is **S-continuous** if a is **infinitesimal**. In fact

$$|f(x) - f(y)| \leq |x - y| + a$$

and thus, if a is **infinitesimal** $f(x)$ is **infinitely close** to $f(y)$.

Let us prove that a standard **S-continuous** function is continuous.

- Let us prove the formula :

$$\forall^{\mathrm{st}}\epsilon \ \exists\eta \ \forall x \ \forall y \ \{|x - y| \leq \eta \implies |f(x) - f(y)| \leq \epsilon\}. \tag{3}$$

  It is enough to take an arbitrary **infinitesimal** η. If $|x - y| \leq \eta$, x and y are **infinitely close**, so are $f(x)$ and $f(y)$. Thus their difference, which is **infinitesimal**, is smaller than the standard number ϵ.

- Using the Transfer axiom the formula (3) is equivalent to the formula :

$$\forall\epsilon\ \exists\eta\ \forall x\ \forall y\ \{|x-y|\leq\eta \implies |f(x)-f(y)|\leq\epsilon\}$$

  which we recognize as the definition of uniform continuity. The axiom of Transfer is valid if the function f is standard.

We understand that the preceding proof is somewhat obscure since we never explained what the Transfer axiom is ! We have merely introduced it to give a general idea of the kind of proofs which appear. They have two parts : One is made of more or less easy majorations. The other one is just "formal nonsense"; there is nothing to understand, just apply the suitable axiom.

Conversely we could prove that if f is standard and uniformly continuous then it is **S-continuous**. Thus, for standard functions, uniform continuity is equivalent to **S-continuity**.

So we begin to understand how one plays the "nonstandard game". External definitions (here **S-continuous**) are equivalent to classical definitions (here uniform continuity) when applied to standard objects. They carry their own meaning when applied to nonstandard objects. The nonstandard piecewise constant function (2) with an infinitesimal step a (which is **S-continuous** but not continuous) "seems" to be continuous if its graph is observed from a sufficiently large distance. Hence, **S-continuous** can be read "seems continuous".

2.4 A Glance at some non standard functions

We are interested in some specific non standard functions that we shall use later. An easy way to "observe"[3] a non standard function is to consider a one parameter family of standard functions and to fix the value of the parameter to some non standard value, for instance **infinitely large**. A non standard function *must be a true classical function defined by an internal formula.* Be careful to not make a confusion between a non standard function and some external intuitive function. For instance, the "function" defined by the external relation :

$f(x) = 1$ if x is **infinitely small**, $f(x) = 0$ overwise,

is not a function, even in the sense of IST.

2.4.1 Fast oscillating trigonometric functions

Consider the function :

$$t \mapsto \sin(\omega t)$$

[3] Let us be clear about "observe". It was sometimes advocated against NSA that it is impossible to produce any **infinitely small** real number. This is perfectly true (but does not prove the uselessness of NSA because, if it where the case, the number π, which is also a complete abstraction would be useless). Thus our "observe" has to be understood in a very weak sense.

If ω is not **infinitely large**, this function is **S-continuous**. Conversely, if ω is **infinitely large** the number $\frac{\pi}{2\omega}$ is **infinitely close** to 0 and $\sin\left(\omega\frac{\pi}{2\omega}\right) = 1$ is not **infinitely close** to $\sin(0)$. Let us compute the integral :

$$I_\delta(x) = \frac{1}{\delta}\int_x^{x+\delta} \sin(\omega t)\, dt$$

which is equal to :

$$I_\delta(x) = \frac{\cos(\omega x) - \cos(\omega(x+\delta))}{\omega\delta}.$$

We see that if ω is **infinitely large** the integral $I_\delta(x)$ is **infinitely small** as soon as $\omega\delta$ is **infinitely large**, and thus, in particular, for every non **infinitely small** δ. In elementary lecture courses of physics one explains that the above moving average is a "low-pass" filter, that is to say, if one considers that an intensity of 0.01 is not appreciable (or equivalent to 0) all the frequencies greater than ω_0 with $\omega_0 = \frac{2}{0.01\delta}$ will be cut. Since the threshold 0.01 may vary, depending on circumstances, an ideal mathematical statement, somewhat universal, should be welcome. The following statement is possible within IST : A moving average of non **infinitesimal** length stops all **infinitely large** frequencies.

But we must acknowledge that this point is not much convincing. Without any call to NSA we can use the classical statement :

$$\omega \geq \frac{2}{\epsilon\delta} \implies I_\delta(x) \leq \epsilon$$

which is perfectly sound and much more precise. A a rule, like all classical asymptotic formulations, NSA is not useful each time we have an explicit inequality. But this is far to be the general situation.

2.4.2 Impulses

In this section we have a look on a class of function which is well known of physicists, the functions of Dirac. Let us recall that for a physicist, before the popularization of the theory of distributions, the "Dirac function" at 0 was a function equal to zero outside of zero, infinite at 0 and such that its integral over the real numbers is equal to 1. Within NSA one says :

Definition 2.2 (Dirac impulse) *One says that an integrable function f belongs to the class of Dirac functions at* 0 *if there exists an* ***infinitesimal*** $\delta > 0$ *such that :*

$$|x| \geq \delta \Rightarrow f(x) \approx 0, \qquad \int_{\mathbb{R}} f(x)dx \approx 1.$$

Notice that the "class" of Dirac functions is not a set. Thus we shall not say "equivalence class". One of the most famous Dirac function is :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

considered for an **infinitely small** value of the parameter σ. An other function that we shall use later is defined by :

$$f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ \omega^2 t e^{-\omega t} & \text{if} \quad x > 0 \end{cases}$$

which, for ω **infinitely large**, is also a Dirac impulse at 0. These non standard functions are external analogues of classical distributions ; indeed one checks very easily the :

Proposition 2.3 *If f belongs to the class of Dirac functions at* 0*, for every standard continuous function ψ with compact support one has :*

$$\int_{\mathbb{R}} f(x)\psi(x)dx \approx \psi(0).$$

As a matter of fact the class of Dirac functions contains all the objects as mad as an analyst might dream to.

This manner to look at things allows clarifications in certain subjects. For instance let us consider the problem of impulse controls for non linear systems. Consider the non linear system :

$$\frac{dx}{dt} = u_1 X^1(x) + u_2 X^2(x). \tag{4}$$

There is a problem to define the solutions when the inputs are distributions. For instance, take the two Dirac functions :

$$u_1(t) = \begin{cases} 0 & \text{if} \quad t \leq 0, \\ \frac{1}{\epsilon} & \text{if} \quad 0 < t < \epsilon, \\ 0 & \text{if} \quad \epsilon \leq t. \end{cases} \qquad u_2(t) = \begin{cases} 0 & \text{if} \quad t < \epsilon, \\ \frac{1}{\epsilon} & \text{if} \quad \epsilon \leq t < 2\epsilon, \\ 0 & \text{if} \quad 2\epsilon \leq t. \end{cases}$$

For ϵ **infinitesimal** these two functions are gentle functions, despite the fact they are non standard. There is no problem to integrate (4). The result is, using the usual notations for the one parameter group of diffeomorphisms generated by vector fields,

$$X_1^2 \circ X_1^1(x_0).$$

If the two vector fields X^1 and X^2 are not commuting the exchange of u_1 with u_2 will change the result of the integration. Thus it is not possible to define a single result for the integration of the whole class of Dirac functions at 0. However, it is not forbidden to use Dirac functions in non linear systems, provided one is careful !

2.5 Historical and bibliographical comments

A complete history of NSA is out of the scope of this paper. Let us just give some elements of recent history. The book of A. Robinson [39], considered as the creator of modern NSA, was published in 1966. An important school works within the formalism developed in this book. We do not report on this aspect here. For references and informations on the recent progress we recommend [6].

At the beginning of the seventies, the French mathematician G. Reeb, well known for his contributions to geometry (The *Reeb foliation*) discovered A. Robinson's book and was very anxious to share his enthusiasm. His very radical posture was contained in his favorite slogan :

Les entiers naïfs ne remplissent pas $\mathbb{N}$.

This did'nt make the communication very easy. Nevertheless he created a school in Strasbourg which decided to adopt the point of view of E. Nelson's paper on IST [34] published in 1977. A part of Reeb's school work is synthesized in the book by R. Lutz and M. Goze [30] published in 1982 and later in two collective works [14, 16] to which we refer much of the time. The interested reader will find other informations in these books. The method was presented to the French community in applied mathematics during the "Colloque d'Analyse Numérique" of 1981 [23] and the congress of Belle Ile en Mer in 1983 [22].

The existence of a French School in NSA was the occasion of a violent debate which was related (with a completely non objective point of view !) in the book of one of us [24]. More philosophical viewpoints, and by the way, in principle, more objective ones can be found in the books [1, 40, 41]. The reader interested by philosophical aspects related to NSA should look also at the home pages of J. Harthong [http://moire4.u-strasbg.fr/] and G. Wallet [http://perso.univ-lr.fr/gwallet/].

There exists numerous text books on NSA. For a short initiation we recommend [14, 15, 16, 38] and for more complete expositions see the bibliography of [14].

3 Stabilization of two times scales systems

Automatic control engineers like to stabilize plants as fast as possible. If the system is linear one tries to put the eigenvalues as far as possible in the left of the complex plane but it is well known that this may be dangerous in the presence of nonlinear, even small, perturbations. Far from being stabilized the system can explode to infinity. In 1988, M. Canalis and P. Yalo [10] remarked the paper [21] of P.V. Kokotovic and R. Marino which for the first time (to our knowledge) payed attention to this problem. They published a short paper which clarifies one of the thorough reasons of this phenomenon : Before it goes to zero a trajectory makes an excursion in the "neighborhood of infinity". Independently H.J. Sussmann and P.V. Kokotovic popularized this phenomenon under the name "peaking phenomenon" in a paper [43] which, moreover, gives some conditions to secure oneself against it. This is, of course, more difficult than to just mention it. Let us see how NSA allows to speak about this phenomenon.

3.1 The peaking phenomenon

Let us point out an old trap which it is advisable to avoid. Consider the system :

$$\frac{dx}{dt} = f(x, y), \qquad \frac{dy}{dt} = g(y) \tag{5}$$

where the systems :

$$\frac{dx}{dt} = f(x, 0) \tag{6}$$

and

$$\frac{dy}{dt} = g(y) \tag{7}$$

are Globally Asymptotically Stable (GAS) at 0. The reasoning below, a little to fast as we will see, leads to conclude to the global asymptotic stability of system (5).

> *Since equation (7) is GAS, $y(t)$ tends to 0. Thus the system :*
>
> $$\frac{dx}{dt} = f(x, y(t))$$
>
> *tends to the autonomous system (6) which is GAS at 0. Hence system (5) is GAS in 0.*

But this is not true as shown by the well known example :

$$\frac{dx}{dt} = -x(1 - xy), \qquad \frac{dy}{dt} = -y. \tag{8}$$

One sees immediately that $t \mapsto (x(t) = e^t, y(t) = 2e^{-t})$ is a solution which

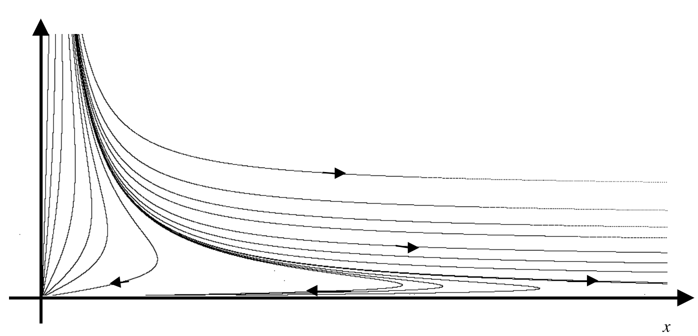

Figure 1: Phase portrait of (8).

separates the phase plane in two parts as indicated on Figure 1. One can think that the size of the basin of attraction depends on the speed with which y tends towards 0. More quickly one tends towards 0 the larger would be the basin of attraction. One can note it on the family of systems :

$$\frac{dx}{dt} = -x(1-xy), \qquad \frac{dy}{dt} = -\gamma y, \tag{9}$$

and make γ tend to infinity. For this system the separatrix is the trajectory defined by $t \mapsto (x(t) = e^{\gamma t}, y(t) = (\gamma+1)e^{-\gamma t})$. The basin of attraction of (9) grows indefinitely when γ tends towards infinity.

However this is not general because of the phenomenon of "peaking" which we expose now. Consider the system :

$$\left\{ \begin{array}{ccc} \frac{dx}{dt} & = & -\frac{x^3}{2}(1+y_2) \\ \frac{dy_1}{dt} & = & y_2 \\ \frac{dy_2}{dt} & = & -\gamma^2 y_1 - 2\gamma y_2 \end{array} \right. \tag{10}$$

The second and third equations in the variables y_1, y_2 form a linear system

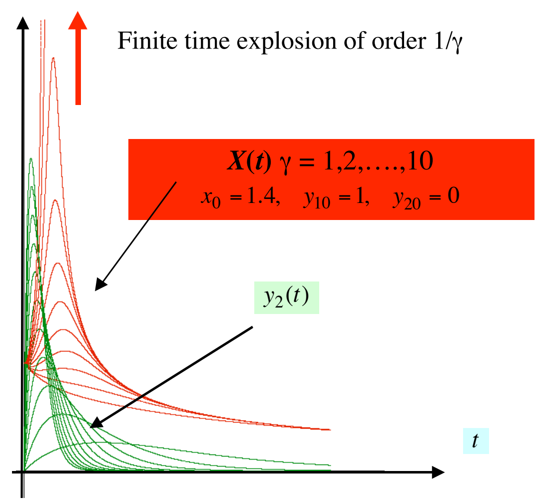

Figure 2: The peaking in equation (10).

which does not depend on x. It acts as a forcing term on the first equation in the variable x. The eigenvalues of the linear system are both equal to γ. The solutions of the linear system tend to 0 all the more quickly as γ is large. Integrating system (10) one obtains :

$$x(t,\gamma) = \frac{x_0}{\sqrt{1 + x_0^2[t - y_{10} + (y_{10}(\gamma t+1) + y_{20}t)e^{-\gamma t}]}}.$$

For the initial conditions $y_{10} = 1$ and $y_{20} = 0$, we get

$$x(t,\gamma) = \frac{x_0}{\sqrt{1 + x_0^2[t - 1 + (\gamma t + 1)e^{-\gamma t}]}}.$$

When γ is large, the quantity $t - 1 + (\gamma T + 1)e^{-\gamma t}$ is very close to -1 for small $t > 0$. Thus for $x_0^2 \geq 1$ the quantity under the radical is negative which says that the solution $x(t)$ has "exploded at infinity". On Figure 2 one can see the solutions with initial condition $x_0 = 1.4$, $y_{10} = 1$ and $y_{20} = 0$ for increasingly large values of γ. To understand what occurred it is enough to consider $y_1(t)$ and $y_2(t)$ for the preceding initial conditions, which writes :

$$t \mapsto \begin{cases} y_1(t) &= \quad (\gamma t + 1)e^{-\gamma t} \\ y_2(t) &= \quad -\gamma^2 t e^{-\gamma t} \end{cases}$$

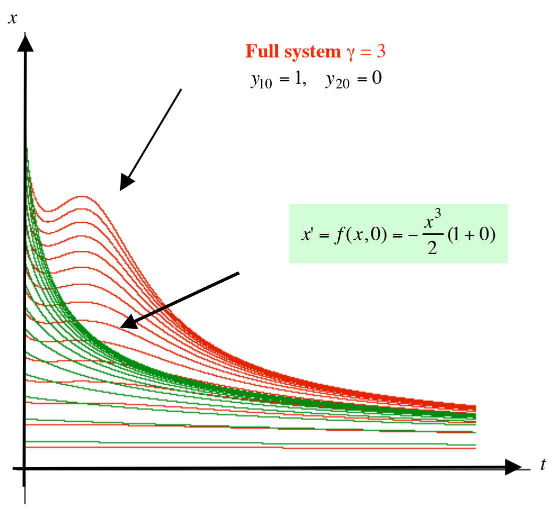

Figure 3: The peaking in equation (10)

The function $y_2(t)$ which appears in the first equation

$$\frac{dx}{dt} = -\frac{x^3}{2}(1 + y_2(t)) \tag{11}$$

of system (10) is a function equal to zero for $t = 0$ and negative for all $t > 0$. Its minimum is achieved for $t = \frac{1}{\gamma}$ and is equal to $-\frac{\gamma}{e}$. For γ large enough this minimum is smaller than -2 and, thus, for a certain amount of time one has :

$$\frac{dx}{dt} > \frac{x^3}{2}$$

which can "explode at infinity". The instability comes from the fact that $y_2(t)$ is an input of (11) more and more close to an impulse, together with the fact that the vector fields (11) is not complete. Consider the second and third equation

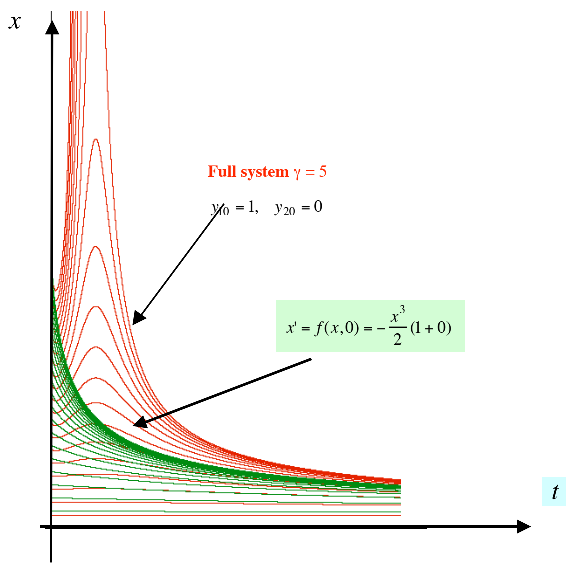

Figure 4: The peaking in equation (10).

of system (10) :

$$\left\{\begin{array}{lcr} \frac{dy_1}{dt} & = & y_2 \\ \frac{dy_2}{dt} & = & -\gamma^2 y_1 - 2\gamma y_2 \end{array}\right.$$

The peaking of this system is the fact that starting from the initial condition $y_1 = 0$ with a velocity equal to zero, to arrive near the point $(0, 0)$ during an elapsed time of the order of $\frac{1}{\gamma}$ it is definitely necessary to have, at some instant, a negative velocity with an absolute value of the order of γ.

On Figures 3 and 4 we represent, for two values of γ and several initial conditions x_0, the solutions of the limiting system, and the solutions of the complete system.

All this is not very mysterious but rather complicated to formalize in the general case. One could, for example, state that the family of linear systems :

$$\frac{dY}{dt} = A(\gamma)Y \tag{12}$$

"makes peaking" when γ tends to infinity if the two following conditions are satisfied. The first condition is that real parts of the eigenvalues of $A(\gamma)$ tend towards $-\infty$ when γ tends to infinity (for saying that the solutions tend more and more quickly to 0). The second condition is that it exists for all γ at least an initial condition of norm equal to 1 such that the maximum of the corresponding trajectory is larger than a function of γ which tends to infinity with γ. For a complete study the reader is refered to [43]. Let us note however that the system

(12) is linear. How to do when it is not the case ? Let us see how the peaking can be expressed within NSA.

3.2 An external view of the peaking phenomenon

3.2.1 Global Asymptotic Stability

Definition 3.1 (**S-GAS**) *The differential system :*

$$\frac{dx}{dt} = f(x)$$

*is **S-Globally Asymptotically Stable at 0 (S-GAS)** if for any **limited** initial condition x_0 and all **infinitely large** t, $x(t, x_0)$ is **infinitely small** (where $x(t, x_0)$ indicates the trajectory of initial condition x_0).*

This external definition of the asymptotic stability, indicated by **S-GAS** to distinguish it from usual global asymptotic stability, is equivalent for the standard systems to the classical definition because one can easily prove the :

Theorem 3.2 *If the differential system*

$$\frac{dx}{dt} = f(x)$$

*is standard it is **S-GAS** at 0 if and only if the two following properties are satisfied*

- *The system is stable at 0, which means : for every neighborhood $\mathcal{V}$ of 0 there exists a neighborhood $\mathcal{W}$ of 0 such that for every initial condition x_0 belonging to $\mathcal{W}$ the positive half trajectory issued from x_0 belong to $\mathcal{V}$ or, if one prefers quantifiers :*

$$\forall\, \mathcal{V}\, \exists\, \mathcal{W}\, \forall x_0 \in \mathcal{W}\, \forall t > 0\; x(t, x_0) \in \mathcal{V}$$

- *The point 0 is attractive, which means that for every initial condition, the corresponding trajectory tends to 0.*

In other words : A standard system is Globally Asymptotically Stable if and only if it is **S-GAS**. Therefore, thanks to this theorem, the situation of **S-Global Asymptotically Stability** in the same as the situation of **S-continuity** : It is an external definition which coincides, when the differential system is standard, with the concept of global asymptotic stability. It is also noticed that the definition of **S-GAS** is spectacularly more compact than that of global asymptotic stability.

On the following example we see the meaning of **S-GAS** in the case of a system which is not standard. Let us consider the system defined in polar coordinates by the equations :

$$\frac{d\theta}{dt} = 1, \qquad \frac{d\rho}{dt} = \rho(\epsilon - \rho).$$

It is a system which, for ϵ **infinitely small**, has a limit cycle of **infinitely small** radius, which is globally asymptotically stable. Therefore, for t **infinitely large**, the trajectory is **infinitely close to** 0, but outside a circle of radius ϵ. Thus the system is not stable *strictly speaking*. We will not seek with demonstration of Theorem 3.2 and we refer to [25] for its proof which is a routine exercise.

3.2.2 Instantaneous stability

We now consider the question of the measurement of the velocity with which a dynamical system GAS at 0 tends to 0. It is not, a priori, very simple to formulate since, in theory, the trajectory never reaches 0 : time to "reach" 0 is always infinite. But this is an "ideal" mathematical vision, in practice, any real system ends up at 0, with respect to the "precision of measurements". The only case where there is a simple answer is the linear case where, due to invariance, one can define the "characteristic time" as the time spent to get the norm of the initial condition is divided by two. Nothing like this is possible in the nonlinear case. Within NSA we propose the following definition :

Definition 3.3 (**S-IGAS**) *The system*

$$\frac{dx}{dt} = f(x)$$

*is **S-Instantaneously Globally Asymptotically Stable at** 0 (**S-IGAS at** 0) if for every **limited** initial condition x_0 and every non **infinitely small** $t > 0$, $x(t, x_0)$ is **infinitely small**.*

The meaning of this definition is clear : any "reasonable" initial condition is transferred instantaneously almost into 0. To see what is the classical equivalent of this external definition, one starts with a family of systems :

$$\frac{dx}{dt} = f(x, \gamma) \tag{13}$$

where γ is a positive real number.

Definition 3.4 (Seems IGAS) *One says that the family of systems (13) "seems Instantaneously Globally Asymptotically Stable (seems IGAS) at 0 when γ tends to infinity" if :*

$$\forall R > 0\, \forall r > 0\, \forall t_0 > 0\;\; \exists \gamma_0 > 0\, \forall \gamma > \gamma_0\, \forall x_0 \leq R\;\; \forall t > t_0\;\; x(t, x_0) < r$$

Or, in a little bit more sound words :

> For any limited set of initial conditions and any ball B centered at 0 of arbitrarily small radius, when γ tends to infinity, the trajectories penetrate more and more quickly in the ball B.

We have the following result

Theorem 3.5 *Assume that the family (13) is standard. Then the family of systems (13) seems IGAS at* 0 *when* γ *tends to infinity if and only if for each* ***infinitely large*** γ*, the system (13) is* ***S-IGAS****.*

In conclusion, thanks to such a theorem (and others of the same type [25]), everything we could prove on non standard systems about instantaneous stability will have a classical counterpart in terms of the asymptotic properties of *families of systems*, when the parameter tends to infinity.

3.2.3 Stability of cascade systems

In Section 3.1 we saw on the example (10) that a cause of nonstability was the fact that solutions "exploded to infinity" very quickly when γ increases. It is enough to forbid this phenomenon to obtain stability. Consider the (not necessarily standard) system

$$\frac{dx}{dt} = f(x,y), \qquad \frac{dy}{dt} = g(y). \tag{14}$$

Definition 3.6 (**UIB**) *The system (14) is* ***Uniformly Infinitesimally Bounded (UIB)*** *if for every* ***limited*** *initial condition* (x_0, y_0) *and every* ***infinitesimal*** *time* t *the* x *component of the corresponding solution is* ***limited****.*

We can now state the :

Theorem 3.7 *Assume that the system (14) is* ***UIB****, that the subsystem :*

$$\frac{dx}{dt} = f(x,0)$$

is ***S-GAS*** *at* 0 *and the subsystem :*

$$\frac{dx}{dt} = g(y)$$

is ***S-IGAS*** *at* 0*, then (14) is* ***S-GAS*** *at* 0*.*

The proof is very simple. We begin by showing that for any **limited** initial condition (x_0, y_0) there exists an **appreciable** real t_1 such as $x(t_1, x_0, y_0)$ is **limited**. To show this result we use a technique known under the name "permanence principle" which goes back to A. Robinson.

> **Lemma 3.8 (Robinson)** *Let* u_n *be a sequence of real number. Assume that for* n *standard* u_n *is* ***infinitesimal****. Then it exists an* ω ***infinitely large*** *such that* u_ω *is* ***infinitesimal****.*
>
> **Proof** One considers the set :
>
> $$A = \{n \in N \ \ : nu_n \leq 1\}.$$

Remark that A is a true set in the formal sense. The following derivation is obvious:

$$\forall n \in \mathbb{N}\ \{st(n) \Longrightarrow n \in A\}$$

since the product of a **standard** by an **infinitesimal** is **infinitesimal** and as such smaller than 1. On the other side a set which contains all the standard elements of $\mathbb{N}$ contains necessarily at least an other elements, which by this way is **infinitely large**. Let ω be such an element. One has $\omega u_\omega \leq 1$ which implies that u_ω is **infinitely small** an proves the lemma.

There are various alternatives of this lemma, for which one can find a synthesis in [5].

Let us come back to our proof with the same idea. Let :

$$A = \{t \in \mathbb{R} : t \geq 0 \quad t\, x(t, x_0) \leq 1\}.$$

The set A is a true set (in the formal sense) which contains all positive **infinitely small** real numbers, by definition of **UIB** (The product of an **infinitesimal** by a **limited** is **infinitesimal** and, by the way, smaller than 1). The set A which contains all **infinitesimals** contains at least a non **infinitesimal** that is to say an **appreciable** real t_1. The inequality :

$$t_1\, x(t_1, x_0) \leq 1$$

proves that $x(t_1, x_0)$ is **infinitesimal** which we wanted to prove. Since the subsystem :

$$\frac{dx}{dt} = g(y)$$

is **S-IGAS**, for $t \geq t_1$ we know that $y(t, y_0)$ is **infinitesimal**. From the time t_1 we integrate the system

$$\frac{dx}{dt} = f(x, y(t, y_0))$$

which is a regular perturbation of

$$\frac{dx}{dt} = f(x, 0)$$

Denote by $x^0(t, t_1, x(t_1, x_0, y_0))$ the solution of this last system issued from the point $x(t_1, x_0, y_0)$ at the instant t_1. From the continuous dependence of solutions of a differential equation with respect to parameters it follows immediately that for **limited** $t \geq t_1$:

$$||x^0(t, t_1, x(t_1, x_0, y_0)) - x(t, x_0, y_0)||$$

is **infinitely small**. The Lemma of Robinson once more enables us to thus go a little further and to say that there exists some **infinitely large** ω such than on the interval $[t_1, \omega]$:

$$||x^0(t, t_1, x(t_1, x_0, y_0)) - x(t, x_0, y_0)||$$

is **infinitesimal** and thus, since the system:

$$\frac{dx}{dt} = f(x, y(t, y_0))$$

is **S-GAS** at 0, for t **infinitely large** smaller than ω we know that $x(t, x_0, y_0)$ is **infinitesimal**. It remains to show that $x(t, x_0, y_0)$ is **infinitesimal** for any **infinitely large** t. Let us suppose that it is not true. Then there exists $\omega_1 > \omega$ such as $x(\omega_1, x_0, y_0)$ is not **infinitely small**. The difference $\omega_1 - \omega$ is necessarily **infinitely large**. The set:

$$B = \{t : t > \omega\ (t - \omega)x(t, x_0, y_0) > 1\}$$

is not empty (it contains ω_1) and its infimum l is such that $l - \omega$ is **limited**, which is a contradiction and achieves the proof. This proof is illustrated by Figure 5.

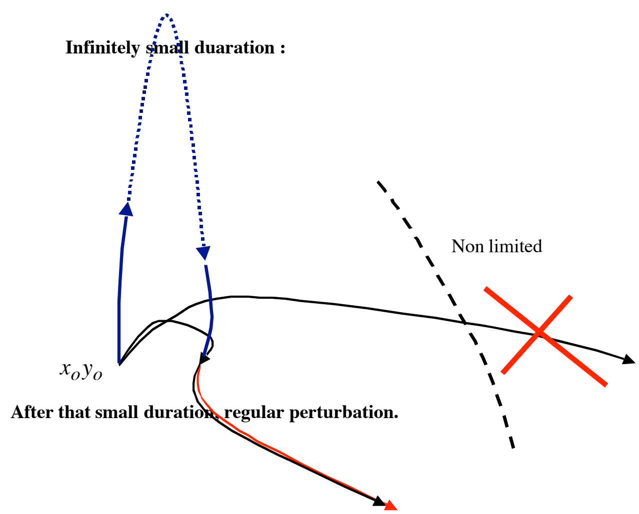

Figure 5: Illustration of the proof of Theorem 3.7.

A small remark to finish : Theorem 3.7 should not make illusion. The sufficient condition **S-UIB** does not relate directly to the data. It is difficult to give sufficient conditions relating to the second members so that a system is **S-UIB**. It is what is made in [43] in the case where the subsystem in the variable y is linear. We did not seek non standard equivalents of their conditions.

3.3 NSA and differential equations

G. Reeb was certainly the first to see all the benefit which NSA could bring to the *drafting* in the field of ordinary differential equations where the geometrical arguments are not allways simple to formalize. He pushed at the end of the seventies some young researchers of Strasbourg to be interested in the equation of Van der Pol via NSA. In the France of mathematics which hardly began to conceive that there were other pure mathematics than Bourbaki's ones and other applied mathematics than that of the digital simulation of partial differential equations, to push mathematicians to be interested in a very small equation that only the electronics specialists of the schools of engineers taught, was resolutely provocative. G. Reeb did not doubt that on this old subject a new glance would not fail to be fertile. It is what occurred with the discovery of the phenomenon called *canard* (or *duck*) phenomenon, i.e. of the importance of certain special solutions in the description of phase portrait of certain one parameter families of differential equations [4]. Our treatment of the peaking phenomenon very clearly claims the philosophy inaugurated in this article. NSA made many other intrusions in the theory of differential equations like, for example, the *stroboscopic method*, which is the external vision of the classical averaging methods [9, 42], the theory of differential equations with discontinuous right hand sides [26], the theory of *rivers* [17], which does not have yet a classical equivalent, and the consideration of the complex slow-fast differential equations [8]. We return to [14] and its bibliography for a rather complete vision of the subject.

4 Nonstandard Theory of noise by M. Fliess

In a recent note [19] M. Fliess uses a result of P. Cartier and Y. Perrin [11] to propose a purely deterministic approach to the question of the treatment of the noise in signal theory. We propose in this part a small history of the ideas which led to the result of [11] used in [19]. It consists essentially of some results from the articles [11, 20, 35, 36][4]. If we do not respect the letter of these authors (in particular certain definitions and/or results that we attribute to them do not necessarily appear in the original papers in the form that we give them) we hope, on the other hand, to respect the spirit of it.

4.1 The Moire's theory of J. Harthong

To jump from a microscopic scale to a macroscopic one by the mean of some averaging is certainly the very basic job of the physicist. The Reeb's school, mainly through the work of J. Harthong [20], gave a contribution to that question in NSA.

The "moiré phenomenon", or simply "moiré" is the following phenomenon. On Figure 6 are drawn two networks of fine black and white lines alternate. If one superimposes two such networks as on Figure 7 one sees appearing "dark stripes" separated by clearer ones. It is also seen that these stripes are all the more broad as the angle between the two networks is weaker. This optical

[4]Since it will be question, in particular, of theories of measurement note the existence of very many developments in this field known under the name of *Loeb measure* [6]. We will not evoke them because their background is the formalism of A. Robinson.

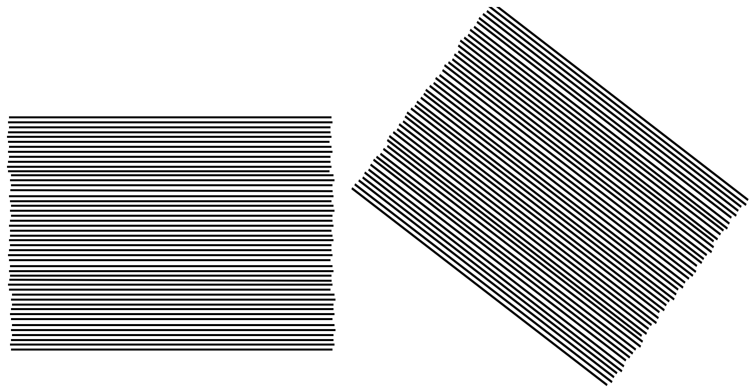

Figure 6: Two "microscopic" networks of straight lines.

phenomenon is very widespread. It is visible in particular when one looks at curtains by transparency, with television when the "pixelized" nature of the screen interferes with certain figures or image etc... J. Harthong immediately saw how to benefit from NSA to theorize this phenomenon. The matter is to model the fact that in certain areas of space, locally, the superposition of the black lines is done on white ones, resulting in a complete black, whereas in other areas of space a black line is superimposed with another black line, leaving clear the neighboring line.

J. Harthong models a "network of step h" in the following way (see Figure 8). It is decided that the network of fine straight lines is at the infinitely small scale. One thus gives oneself a strictly positive **infinitely small** real number h which will be the *mesh* of the network and a periodic function ψ, of period 1, which takes alternatively the value 0 or 1 on intervals of length $\frac{1}{2}$. The value 0 codes for "black", the value 1 for "white". The function ψ measures a "transmittance": 0 if the light does not pass, 1 if all the light passes. The function :

$$(x, y) \mapsto \psi\left(\frac{\lambda x - y}{h}\right)$$

takes the value 1 if $h(k - 0.5) \leq \lambda x - y < hk$, which means that on lines of **infinitesimal** wideness h and slope λ the function

$$(x, y) \mapsto \psi\left(\frac{-y}{h}\right)\psi\left(\frac{\lambda x - y}{h}\right)$$

takes the value 0 or 1 and codes for the superposition of an horizontal network and a network of slope λ. In order to represent the "moiré phenomenon" J. Harthong proposes to consider the function :

$$\mathcal{M}(x, y) = \frac{1}{\mu(\mathbf{hal}(x, y))} \int\int_{\mathbf{hal}(x,y)} \psi\left(\frac{-v}{h}\right)\psi\left(\frac{\lambda u - v}{h}\right) du dv \qquad (15)$$

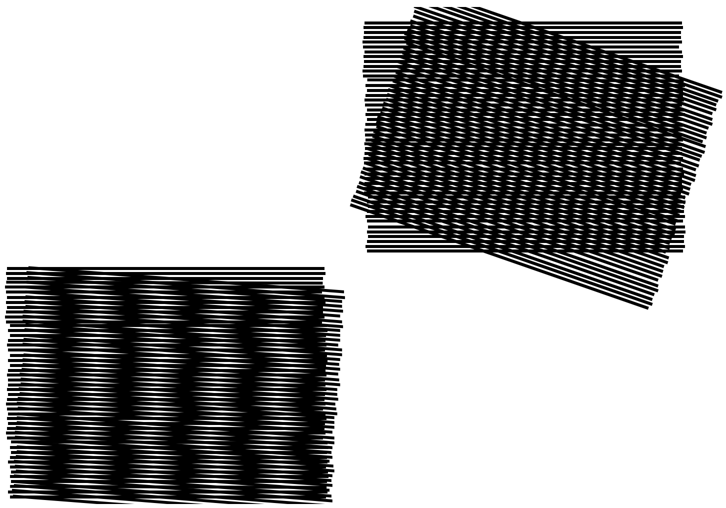

Figure 7: The "stripes" of moiré.

where **hal**(x, y) is the **halo** of the point (x, y) and μ indicates the Lebesgue measure of $\mathbb{R}^2$. The **halo** of (x, y) being made up of "all" points of the plane which are "very close" to (x, y) one can say that it is, at the same time, a small set "in the absolute", but "large" at the scale of h. Indeed the points distant of $\sqrt{h}$ of (x, y) are at a distance **infinitely large** compared to h of (x, y). The average "on the **halo**" of (x, y) of the product of the transmitances thus represents well locally (at the point (x, y)), on a macroscopic scale, the effect of the product of transmittances of the two microscopic networks.

Since the figures showed that the stripes appear for the small values of the slope, we will take for λ a number of the same order of magnitude as h, i.e. $\lambda = kh$ with k **limited**. Now let us replace λ by this value not in $\mathcal{M}(x, y)$ but in :

$$M_\epsilon(x, y) = \frac{1}{(2\epsilon)^2} \int_{v=y-\epsilon}^{v=y+\epsilon} \int_{u=x-\epsilon}^{u=x+\epsilon} \psi\left(\frac{-v}{h}\right) \psi\left(ku + \frac{-v}{h}\right) du dv,$$

where ϵ is **infinitesimal** and **infinitely large** compared to h. The change of variable $\frac{-v}{h} = w$ gives :

$$\frac{1}{2\epsilon} \int_{v=y-\epsilon}^{v=y+\epsilon} \psi\left(\frac{-v}{h}\right) \psi\left(ku + \frac{-v}{h}\right) dv = \frac{h}{2\epsilon} \int_{w=-y/h-\epsilon/h}^{w=-y/h+\epsilon/h} \psi(w)\psi(ku + w) dw.$$

When ϵ/h is **infinitely large** this last integral is the average on a very great number of periods of the periodic function $w \mapsto \psi(w)\psi(ku + w)$, and is thus **infinitely close** to the average to this function over one period, i.e. :

$$\int_0^1 \psi(w)\psi(ku + w) dw.$$

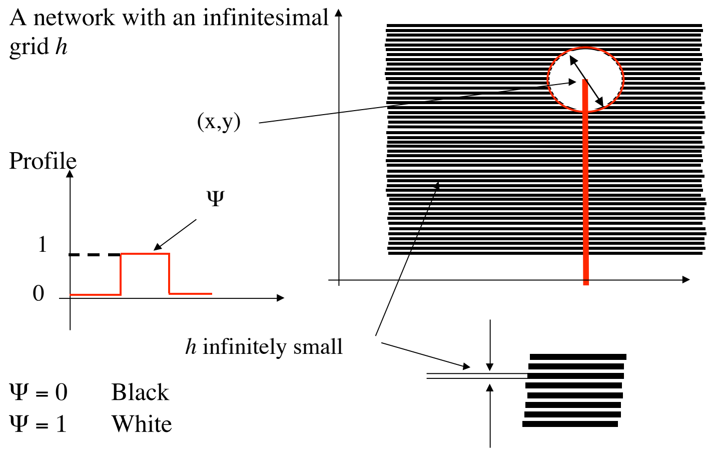

Figure 8: Modelisation

Moreover the function

$$u \mapsto \int_0^1 \psi(w)\psi(ku+w)dw$$

is **S-continuous**. So, we can write :

$$M_\epsilon(x,y) \approx \frac{1}{2\epsilon}\int_{u=x-\epsilon}^{u=x+\epsilon}\int_0^1 \psi(w)\psi(ku+w)dwdu.$$

Thus, for ϵ **infinitely small** this quantity is **infinitely close to** :

$$\int_0^1 \psi(w)\psi(kx+w)dw$$

and finally :

$$M_\epsilon(x,y) \approx \int_0^1 \psi(w)\psi\left(\frac{\lambda}{h}x+w\right)dw.$$

One sees that $M_\epsilon(x,y)$ is a quantity independent of y. This explains that stripes are vertical, and that the period is all the more large as λ is small. This explains also why the width of the stripes is all the more large as the slope is weak. We thus have a theory[5] which suitably describes the optical phenomenon presented at the beginning of this paragraph. The principle is simple : make an average on a set which is at the same time rather large on the scale of the mesh of the networks considered, but small enough to reflect a local property. One can imagine that this theory can extend to nonlinear networks. It is what is done in [20].

[5] Here we take "theory" within the meaning of "physical theory", i.e. of a convincing system of representation of reality, not within the meaning of "mathematical theory", i.e. an irreproachable text from the point of view of the accepted mathematical rigor of the moment.

4.2 The averaging theory of C. Reder

It undoubtedly does not have missed to the reader that, from the mathematical point of view, the formula (15) of definition of the average "on the **halo**" poses a problem, even within the framework of NSA[6]. Indeed "$\mathbf{hal}(x, y)$"is not a set. We thus do not have a theory of integration valid for such objects. In the calculation of the preceding paragraph we solved the problem by integrating not on all the **halo** of (x, y) but simply on a square of size 2ϵ. But this misses a little "canonicity". The clarification of these questions is not a purely formal problem and was undertaken by C. Reder in [36] from where we extract the matter for this paragraph. Whereas the question of the "moiré" effect is primarily a problem with two dimensions, C. Reder considered, to start, the unidimensional case which is simpler.

4.2.1 The apparent value at a point

One places oneself on $\mathbb{R}$ provided with the Lebesgue measure. One considers a (not necessarily standard) Lebesgue integrable function f. To simplify, we suppose that $|f|$ is bounded by a **limited** constant.

Definition 4.1 (Observability at a point) *Let x be a **limited** real number. It is said that the function f is **observable at x** and that a is an apparent value of f at x if there is an **infinitely small** real number $h_0 > 0$ such as:*

$$\forall h_1\ \forall h_2 \quad (h_1 \approx 0\ h_2 \approx 0\ h_1 \geq h_0\ h_2 \geq h_0) \Longrightarrow \frac{1}{h_1 + h_2} \int_{x-h_1}^{x+h_2} f(s)ds \approx a.$$

It is immediate to note that two apparent values at the same point are infinitely close. By abuse of language we call "apparent value" (when it exists) and note $\mathcal{F}(x)$ the collection of the apparent values of f at point x. If this collection were a formal set we could call upon the axiom of the choice "to choose" a value. In the system IST this possibility "of choosing" a particular value in a collection of **infinitely close** real numbers is offered by the axiom of "standardization". We could use it (it is besides what is made in [36]) but that would lead us formal developments which do not seem useful to us in this article. To paraphrase what has been just made we will say that the "apparent value ", when it exists, is the average "on a sufficiently large **infinitely small** interval".

In Definition 4.1 one makes the average on an interval which is contained in $\mathbf{hal}(x)$ and which contains x in its interior. The following definition does not make any more this restriction.

Definition 4.2 (Strong observability at a point) *Let x be a **limited** real number. It is said that the function f is **strongly observable at x** and that a is an apparent value, if there is a real **infinitely small** $h_0 > 0$ such that :*

$$\forall h\ \forall y \quad (h \approx 0\ h \geq h_0\ |y - x| \approx 0) \Longrightarrow \frac{1}{h} \int_{y}^{y+h} f(s)ds \approx a.$$

[6] Notice that J. Harthong took care well in his article not to write such a formula !

One notes easily that an **S-continuous function** f at x is observable at x and has as apparent value $f(x)$. One has the following properties :

- A periodic function f of **infinitesimal** period T is strongly observable at x and has as the apparent value a for any a such that:

$$a \approx \frac{1}{T}\int_0^T f(t)dt.$$

  Indeed, let $h_0 = \sqrt{T}$ which is **infinitely large** compared to T. Let h be an **infinitesimal** larger than h_0. One can write : $h = nT + r$ with n **infinitely large** and $r < T$. The average on $[y, y+h]$ breaks up into :

$$\frac{1}{h}\int_y^{y+h} f(t)dt = \frac{1}{nT+r}\left[n\int_y^{y+T} f(t)dt + \int_y^{y+r} f(t)dt\right]$$

  from which we deduce immediately what was claimed.

- The function which is equal to 0 for x negative and 1 for x positive and unspecified at 0 does not have an apparent value at 0. Indeed, the average of this function on an interval which contains 0 can vary between 0 and 1 according to whether 0 is at one edge or the other of the interval.

- One shows in [36] that the function f which is null for $x \leq 0$ and which is equal to $\sin\left(\frac{1}{x}\right)$ for x positive is observable at 0 but is not strongly observable there.

One can wonder under which conditions a function will often have an apparent value. The answer is that if f is integrable it has an apparent value in almost any point (see [36]).

4.2.2 Moving average

The purely external definition of the apparent value does not make it possible to speak about the function which associates to x an apparent value of f at point x when it exists. The result of this paragraph will give us the means of solving this problem. Let us consider the function (defined for $h > 0$) :

$$x \mapsto M_h(x) = \frac{1}{h}\int_x^{x+h} f(t)dt.$$

It is the "moving average" of f on a window of width h to the most elementary meaning of the term. Since unmemorable times, to take the moving average is a process used to regularize the functions. Is there a link between the moving average and the apparent value ? If we return to the definition of the apparent value at x we let find that the average on an **infinitesimal** interval h "large enough" should not depend any more on h, but, a priori, the rather large "h" depends on the point x. In fact it is not the case, as the result [36] shows :

Theorem 4.3 *There exists an* ***infinitely small*** *real number* $h_0 > 0$ *such that for all* ***infinitely small*** *real number* h *greater than* h_0*, at any point* x *where* f *is observable, the average* $M_h(x)$ *is an apparent value of* f*. Moreover, at this point, the function* $x \mapsto M_h(x)$ *is observable and an apparent value of* M_h *is an apparent value of* f*.*

This theorem is not trivial and we return to [36] for its proof. On the other hand it results from it some easy conclusions.

- Let f be integrable. There exists an **infinitesimal** h_0 such that for h_1 and h_2 **infinitesimals** greater than h_0 one has $M_{h_1}(x) \approx M_{h_2}(x)$ at each point where f is observable.
- Let h be such that In Theorem 4.3. The function $x \mapsto M_h(x)$ is continuous but is not **S-continuous**.
- Let I an interval such that at each point f is observable . Then $x \mapsto M_h(x)$ is **S-continuous** on I.

We thus can, from now, associate to a function f its "regularized" that we note $\mathcal{M}_f$. We understand there any unspecified functions M_h for h sufficiently large, defined by Theorem 4.3. This process of regularization through moving average is called regularization by convolution in [36]. If we start from a signal corrupted by noise, and we draw the graph of M_h for increasing values of h from 0 we observe that the graph becomes independent of h after a certain value of h, as it is illustrated on Figure 9. In this experiment we started from of a disturbed signal. The width of the figure represents on the whole seven units ; the graphs of the function M_h are shifted upwards for each new value of h, for values of h varying of 0.02, by step of 0.02, to 0.3. 0n can consider that after $h = 0.1$ function M_h does not depend any more on h.

4.3 Radically Elementary Probability Theory

In [35][7] E. Nelson proposes a Nonstandard version of the theory of the Brownian motion, which cannot leave indifferent if one knows that he is also the author of "Dynamical theories of Brownian motion" [33][8], a very great classic. The principle is as follows. Everyone knows the "walk of the drunkard". It is the stochastic process defined by :

$$x_{t+dt} = x_t + Z_t\sqrt{dt}, \tag{16}$$

where t takes discrete values $0, dt, ..., kdt, ...$ and where Z_t is a sequence of random variables taking values $+1$ or -1 with the probability $\frac{1}{2}$. We symbolize this by $Z_t = \pm$ which leads us to rewrite (16) as :

$$x_{t+dt} = x_t \pm \sqrt{dt}. \tag{17}$$

[7]Downloadable Book in English, French or Russian
[8]Downloadable Book

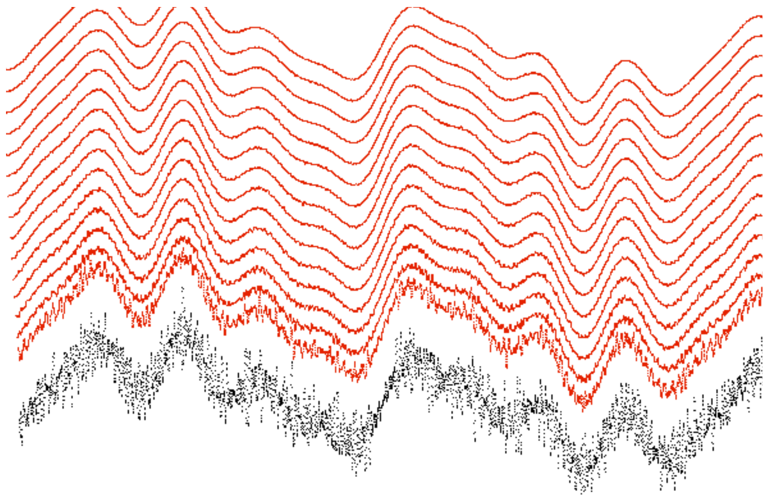

Figure 9: Evolution of M_h.

The length $\sqrt{dt}$ of the step is chosen in order to normalize the variance of x_1 at the value 1. The physical Brownian motion is the process in which a particle of size of the order of some micrometers undergoes each second an incredibly large number of shocks on behalf of the molecules of the fluid in which it is plunged. The mathematical Brownian motion, or Wiener process wants to be the idealization of this situation. In mathematics one idealizes "small" by the "continuous limit". This leads to the invention of the mathematical Brownian motion which is a *continuous time process* in which at *each moment* one choose with head or tail the direction of the motion. There is a very strong discrepancy between the idea of "continuous mathematics", represented by the real line, and a "succession of moment". This is why the process of Wiener is an abstract object so difficult to define and manipulate. E. Nelson proposes to idealize the physical Brownian motion, by considering simply the process (17) with dt **infinitesimal**. The unspecified character of dt gives all its canonicity to the process at least in all the assertions where dt does not appear explicitly. Such an assertion is, for example :

Theorem 4.4 (Nelson) *Almost surely the trajectories of the Brownian motion are continuous.*

One certainly understood that in this theorem "continuous" must be taken within the meaning of **S-continuous**. The "almost surely" deserves an explanation. Let us suppose that the interval of time $[0, 1]$ is discretized as :

$$0, dt, 2dt, \cdots, kdt, \cdots, Ndt = 1,$$

provided with the measure of probability product of the uniform measure on the set $\{-1, +1\}$. Each element of this set has a probability of $\frac{1}{2^n}$ and there is not set of null measure. In addition, for a trajectory, being **S-continuous** is an external property, which thus does not define a set. This is why one says of a property P (possibly external) that it is **rare** if, for any standard ϵ there is a set A of measure lower than ϵ such as if $P(z)$ is true then z belongs to A. An event is **almost sure** if its complement is a **rare** event.

The theory of E. Nelson is *radically elementary* in the sense that it uses only a weak version of the theory IST more intuitive and simpler than the complete theory, and a trivial theory of integration on finite sets. Most astonishing is that, in an appendix which is not elementary (one uses all the force of IST), E. Nelson shows that the elementary theory potentially contains any result which the continuous theory could. He do not hesitate to present its appendix by these words :

> *The purpose of this appendix is to demonstrate that theorems of the conventional theory of stochastic processes can be derived from their elementary analogues by arguments of the type usually described as generalized nonsense ; there is no probabilistic reasoning in this appendix. This shows that the elementary nonstandard theory of stochastic processes can be used to derive conventional results ; on the other hand, it shows that neither the elaborate machinery of the conventional theory nor the devices from the full theory of nonstandard analysis, needed to prove the equivalence of the elementary results with their conventional forms, add anything of significance : the elementary theory has the same scientific content as the conventional theory. This is intended to be a self-destructing appendix.*

This book of E. Nelson, if it does not have a direct relationship with averaging, the subject which interests us here, is certainly a source of inspiration for the authors who wish to consider the question of measure theory from an elementary point of view with the means of IST.

4.4 The measure theory of Cartier-Perrin

It is thus trying to forget the probabilities and to follow the path opened by E. Nelson to build an elementary measure theory. Let us take on the interval $[0, 1]$ the sequence of points $\{dt, 2dt...., kdt..., Ndt = 1\}$, and assign to each point the "mass" $\frac{1}{n}$. Let us call "measure of enumeration" of a set A of $[0, 1]$ the number :

$$\delta(A) = \frac{1}{N} \sum_{k=1}^{N} \chi_A(kdt),$$

where χ_A is the characteristic function of the set A. The quantity $\delta(A)$ has good properties like additivity, invariance by translation but has the unacceptable defect to give the mass 1 to the whole set of rational of $[0, 1]$. It is thus difficult to consider a theory of measurement on such a naïve basis. In [11], P. Cartier and Y. Perrin propose an "integration theory on finite sets" which tries to be enough rich to meet the needs for the current analysis. They consider a finite set (of **infinitely large** cardinal) $X = \{a_1, a_2..., a_i..., a_N\}$; to each element a_i is associated a positive or null number, its "mass" m_i. The measure $m(A)$ of a set A is the sum of the masses of the points which belong to him. Being given a function f defined on X, with value in $\mathbb{R}$ one can always consider the sum :

$$\sum_{i=1}^{N} m_i f(a_i)$$

and denote it by :

$$\sum_{i=1}^{N} m_i f(a_i) = \int_X f dm.$$

That class of all functions being too large we shall restrict to :

Definition 4.5 (**S-integrable** functions) *An application f from X to $\mathbb{R}$ is called **S-integrable** if and only if :*

$$\int_x |f| dm$$

*is **limited** and*

$$\int_A f dm \approx 0$$

*for each **rare** set A.*

A set A is **rare**[9] if, as in the theory of E. Nelson, it is a (possibly external) set such that for any standard $\alpha > 0$ there exists an internal set B such as A is contained in B and $m(B) \leq \alpha$. One provides then the set X with a distance d. One supposes that X is precompact, which means that the diameter of X is **limited** and that for any non **infinitely small** real number $r > 0$ there exists a covering of the space by a **limited** number of balls of radius smaller than r. The definition of **S-continuous** is the same one as that which we already defined and :

Definition 4.6 (**Almost S-continuous** functions) *A function f from X to $\mathbb{R}$ is **almost S-continuous** if an only if it is **S-continuous** on the complement of a **rare** set.*

This makes it possible to define the concept of "Lebesgue" integrable function.

Definition 4.7 (**L-integrable** functions) *A function is **L-integrable** if and only if it is **S-integrable** and **almost S-continuous**.*

Let us add the definition of "fast oscillating" function :

[9] In [2] which is the extension of the *Radically elementary probability theory* of E. Nelson to diffusion processes, E Benoit establishes nontrivial properties of certain **rares** sets

Definition 4.8 (**Fast oscillating** functions) *A function f from X to $\mathbb{R}$ is* ***fast oscillating*** *if and only if it is* ***S-integrable*** *and if for each* ***quarrable*** *subset of X one has :*

$$\int_A hdm \approx 0.$$

Definition 4.9 *A set A is* ***quarrable*** *if its boundary is a* ***rare*** *set.*

We can now state the :

Theorem 4.10 (Decomposition theorem of Cartier-Perrin) *Let f be an* ***S-integrable*** *function. Then one has :*

$$f = g + h$$

where g is ***L-integrable*** *and h is* ***fast oscillating****. The decomposition is unique up to an* ***infinitesimal***

To establish the link with the averaging theory of J. Harthong and C. Reder let us give an idea of the proof of this result. The idea is to define a sequence $\mathcal{P}_n$ of partitions of X by sets of diameters smaller and smaller and to calculate the average of f on each atom of the partition $\mathcal{P}_n$ which gives a sequence of function f_n ; for suitable values of the indices the function f_n will be **almost S-continuous** thus **Lebesgue integrable**. Since all the definitions involved are external this construction must be carried out carefully.

4.5 The Nonstandard definition of noise by M. Fliess

In a recent communication, M. Fliess [19] proposes to consider a signal (disturbed), not like a continuous function, but for what it is really, a sampled signal, i.e. a function f (the notations are not those of [19]) defined on a finite set, but idealized in a "almost interval". An **almost-interval** (to follow the terminology of E. Nelson taken again by Cartier-Perrin) is a finite set of points:

$$\{t_0, t_1 ..., t_i ..., t_N\}$$

of the interval $[a, b]$ such as $a = t_0$, $t_{i-1} \approx t_i$, $t_N = b$ affected of the masses

$$m_0 = 0, \text{ and } \quad m_i = t_i - t_{i-1}, \quad i = 1, 2, \cdots, ..N.$$

and provided with the distance induced by the natural one of $\mathbb{R}$. It is thus a finite set, precompact with the meaning of the preceding paragraph. One admits then that the disturbed signal is an **S-integrable** function and, according to Theorem 4.10, one has :

$$f = g + h$$

with g **L-integrable** and h **fast oscillating**. M. Fliess proposes to define the "noise" as the **fast oscillating** part h of signal f.

This double decision to define the noise within a *discrete* and *purely deterministic* framework is a strong decision which goes against the current mathematical tradition which thus treats the noise as a "white noise" like, to some extent, a continuous realization of the "pure chance". We will not discuss the relevance of this choice since we are not specialists in signal processing. We simply make some remarks of a mathematical nature.

- M. Fliess also defines in the same way the noise for signals in several dimensions.

- One could as well have chosen to treat continuous signals within the more classical (but still nonstandard) framework of C. Reder and the results of [36].

- As we did in the case of the peaking, we propose a classical equivalent of "fast oscillating". One gives oneself a family $f_\gamma(t)$ of applications of $\mathbb{R}$ into $\mathbb{R}$. It will be said that f_γ is asymptotically (when $\gamma \to +\infty$) fast oscillating if and only if:

  $$\forall \epsilon > 0\ \forall M > 0\ \exists \gamma_0\ \forall \gamma\ \forall a\ \forall B\ \{\gamma > \gamma_0 \& |b-a| < M \Rightarrow \left| \int_a^b f_\gamma(t)dt \right| < \epsilon\},$$

  that will be compared to : "On a **limited** interval the integral of f is **infinitely small**".

- Let us consider functions of period $[0, 2\pi]$. That one adopts the relatively classical point of view of the functions (possibly not standard) integrable or that more radical where $[0, 2\pi]$ is replaced by one **almost-interval** (as in [13]) one can develop the function into Fourier series. Thus let :

  $$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{int}.$$

  The theorem of decomposition of f in a regular part (**L-integrable function** or "observation function") and an **oscillating** part suggests the possibility of a decomposition of the kind :

  $$\sum_{n=-\infty}^{+\infty} c_n e^{int} = \sum_{n \in \{\textbf{limited}\}} c_n e^{int} + \sum_{n \in \{\textbf{infinitely large}\}} c_n e^{int}$$

  where the first sum would be the regular part and the second the **fast oscillating** part. For that it would be necessary to determine the conditions under which it is possible to give a sense to the two sums (on external sets of indices) which for the moment do not have any.

- In a very recent work [3], E. Benoit approaches this frequential point of view in the nonperiodic case via the Laplace transform. He shows that if f is **fast oscillating**, its Laplace transform $F(s)$ is **infinitely small** if $\mathcal{R}e(s) > 0$ and $\frac{\mathcal{I}m(s)}{\mathcal{R}e(s)}$ are **limited**.

5 Last remarks

The reader a little familiar with IST [34] will have noticed that we did not use the axiom of Standardization. What is it about ? It is known that in formal set theory (in ZFC) there is an axiom which is stated as follows : If P is an (internal) formula then

$$\forall x\ \exists y\ \forall z\{z \in x \wedge P(z) \Longleftrightarrow z \in y\},$$

which says that the (intuitive) set of all those $z \in x$ which have the property P is a true set y, a set within the formal meaning of ZFC. We sufficiently insisted on the fact that if P is not an internal formula such a formal set does not necessarily exist and that was there one of the key of the effectiveness of the nonstandard language.

The axiom of Standardization in IST is a provider of standard sets issued from external formulas. It is stated as follows : Let P be a formula (not necessarily internal)

$$\forall^{st} x\ \exists^{st} y\ \forall^{st} z\{z \in x \wedge P(z) \Longleftrightarrow z \in y\}$$

In other words, for any standard set x and any property there exists a standard set y whose standard elements are the standard elements of x which satisfy the property and only them. It is, in particular, the use of this axiom which makes it possible E. Nelson to show in the appendix of [35] that its radically elementary theory of the probabilities is equivalent to the classical theory. But as E. Nelson says *This appendix is to some extent self-destructing.* This is why a more "radical" nonstandard point of view is to recommend the use of weak nonstandard theories like E. Nelson in [35], J-L. Callot in [7] and R. Lutz in [29] (see also [31, 32]). In such theories one does not seek to distinguish the non standard objects from the standard objects, in all generality. One is satisfied to do it on integers which is sufficient to analysis. But it is not possible any more to compare with the classical results. Thus "a standard function is uniformly continuous if and only if it is **S-continuous**" does not have meaning since the notion of "standard function" is not defined. By keeping the concept of standard object, without using the axiom which accompanies it, we chose to allow such comparisons. But people convinced by NSA think that they are *self-destructing*.

References

[1] H. Barreau, J. Harthong (editors). *La mathématique non standard.* Editions du CNRS, Paris, 1989.

[2] E. Benoit. Diffusions discrètes et mécanique stochastique. Technical report, Université de Nice, Laboratoire J. A. Dieudonné, 1989.

[3] E. Benoit. Nonstandard averaging and signal processing. 2006. Slides at EMS 2006 congress ; http://www.dm.unipi.it/~nsm2006/slides/benoit-s.pdf.

[4] E. Benoit, J-L. Callot, F. Diener, M. Diener. Chasse au canard. *Collectanea Mathematica*, 31(1-3):37–119, 1981.

[5] I. van den Berg. *Nonstandard asymptotic analysis.* Lecture Notes in Mathematics, 1249. Springer-Verlag, Berlin, 1987.

[6] I. van den Berg, V. Neves (editors). *The Strenght of Nonstandard Analysis.* Springer-Verlag, Wien, 2007.

[7] J-L. Callot. Trois leçons d'analyse infinitésimale. In [41], pages 369–381.

[8] J-L. Callot. Champs lent-rapides complexes à une dimension lente. *Ann. Scient. Ec. Norm. Sup*, 4(26):149–176, 1993.

[9] J-L. Callot, T. Sari. Stroboscopie infinitésimale et moyennisation dans les systèmes d'équations différentielles à solution rapidement oscillantes. In [22], pages 345–353.

[10] M. Canalis-Durand P. Yalo. Robustesse des systèmes linéaires bouclés aux perturbations non linéaires. In A. Bensoussan and J.L. Lions, editors, *Analysis and Optimization of Systems*, volume 111 of *Lecture Notes in Control and Information Sciences*, pages 282–293. Springer, 1988.

[11] P. Cartier, Y. Perrin. Integration over finite sets. In [14], pages 185–204.

[12] A. Deledicq, M. Diener. *Leçons de calcul infinitésimal.* Armand Colin, Paris, 1989.

[13] P. Delfini, C. Lobry. The vibrating string. In [14], pages 51–70.

[14] F. Diener, M. Diener (editors). *Nonstandard Analysis in Practice.* Springer-Verlag, Berlin, 1995.

[15] F. Diener, G. Reeb. *Analyse non standard*, Collection Enseignement des Sciences, 40. Hermann, Paris, 1989.

[16] M. Diener, C. Lobry (editors). *Analyse Non Standard et représentation du réel.* OPU, Alger and CNRS, Paris, 1985.

[17] M. Diener, G. Reeb. Champs polynômiaux: nouvelles trajectoires remarquables. *Bull. Soc.Math. Belgique*, 38:131–150, 1987.

[18] M. Diener, G. Wallet (editors). *Mathématiques finitaires et analyse non standard. Tomes 1 and 2.* Publications Mathématiques de l'Université Paris VII, 31. Université de Paris VII, U.E.R. de Mathématiques, Paris, 1989.

[19] M. Fliess. Analyse non standard du bruit. *C. R. Acad. Sci. Paris, Ser I*, 342:797–802, 2006.

[20] J. Harthong. Le moiré. *Adv. in Appl. Math.*, 2(1):24–75, 1981.

[21] P.V. Kokotovic, R. Marino. On vanishing stability regions in nonlinear systems with high gain feedback. *IEEE. Trans. Autom. Control.*, AC-31(10), 1986.

[22] I. D. Landau (editor). *Outils et Modèles Mathématiques pour l'Automatique, l'Analyse de Systèmes et le Traitement du Signal*, Volume 3. Edition du CNRS, Paris, 1983.

[23] C. Lobry. Mathématiques non standard, mathématiques appliquées ? Technical report, Université de Bordeaux I, Mathématiques, 22p., 1981. Conférence pleinière, Colloque national d'analyse numérique, Aussois.

[24] C. Lobry. *Et pourtant ils ne remplissent pas* $\mathbb{N}$ *!* Editeur ALEAS, 15 quai Lassagne, Lyon, 1989.

[25] C. Lobry, T. Sari. The peaking pheneomenon and singular perturbations, an extension of Tikhonov's theorem. Technical report, INRIA, 2001. http://hal.inria.fr/inria-00072586

[26] C. Lobry, T. Sari. *Equations différentielles à second membres discontinus*, in Contrôle non linéaire et applications, pages 237–266. Hermann, Travaux en Cours 64, 2005.

[27] C. Lobry, T. Sari. *Singular perturbations in control theory*, in Contrôle non linéaire et applications, pages 151–177. Hermann, Travaux en Cours 64, 2005.

[28] C. Lobry, T. Sari, S. Touhami. On Tykhonov's theorem for convergence of solutions of slow and fast systems. *Electronic Journal of Differential Equations*, 1998 (19) : 1–22, 1998.

[29] R. Lutz. Rêveries infinitésimales. *Gazette des Mathématiciens*, 34 :79–87, 1987.

[30] R. Lutz, M. Goze. *Nonstandard analysis : a practical guide with applications.* Lecture notes in Mathematics, 881, Springer-Verlag, Berlin, 1982.

[31] R. Lutz, L. Gonzaga. Modern Infinitesimals as a Tool to Mach Intuitive an Formal Reasoning in Analysis. *Synthese*, 134 (1-2):325–351, 2003.

[32] R. Lutz, M. Musio. An alternative mathematical foundation for statistics. *Acta Appl. Math.* 89 (1-3):217–249, 2005.

[33] E. Nelson. *Dynamical Theories of Brownian Motion.* Princeton University Press, 1967. http://www.math.prince-ton.edu/˜nelson/books/

[34] E. Nelson. Internal set theory: a new approach to nonstandard analysis. *Bull. Amer. Math. Soc.*, 83 (6) : 1165–1198, 1977.

[35] E. Nelson. *Radically Elementary Probability Theory.* Princeton University Press, 1987. http://www.math.princeton.edu/~nelson/books/

[36] C. Reder. Observation macroscopique de phénomènes microscopiques. In [16], pages 195–244.

[37] G. Reeb. La mathématique non standard vieille de soixante ans ? Troisième Colloque sur les Catégories, dédié à C. Ehresmann, Amiens, 1980. *Cahiers Topologie Géom. Différentielle* 22(2):149–154, 1981.

[38] A. Robert. *Analyse non standard*, Presses polytechniques romandes, EPLF, Lausanne, 1985. English translation *Nonstandard Analysis*, Wiley, New York, 1988.

[39] A. Robinson. *Nonstandard analysis.* Noth Holland, Amsterdam, 1966.

[40] J-M. Salanskis. *Le constructivisme non standard.* Histoire des Sciences. Presses Universitaires du Septentrion, Villeneuve d'Ascq, 1999.

[41] J-M. Salanskis, H. Sinaceur (editors). *Le Labyrinthe du Continu.* Springer-Verlag France, Paris, 1992.

[42] T. Sari. Averaging for ordinary differential equation and functional differential equations. In [6], pages 286–305.

[43] H.J. Sussmann, P.V. Kokotovic. The peaking phenomenon and the global stabilization of nonlinear systems. *IEEE. Trans. Autom. Control.*, 36(4):224–440, 1991.