Qualitative Computing

F. Chaitin-Chatelin* E Traviesas†

CERFACS Technical Report TR/PA/02/58‡

1 Introduction

Computing is a human activity which is much more ancient than any historical record can tell, as testified by stone or bone tallies found in many prehistorical sites. It is one of the first skills to be taught to small children. However, everyone senses that there is more to computing than the multiplication table...

Why do we compute? What does it mean to compute? To what end? Our technological society wants to compute more and more efficiently. The question of "how to compute" overshadows that of "why compute" in high-technology circles.

And indeed the meaning of the act of computing seems to vanish, while we are gradually surrounded by a digital reality, which tends to shield us from the natural reality. It is therefore, more than ever, vital that we analyse computation through the two questions of **why**, as well as **how**. This can lead the way towards a better understanding of intensive computer simulations in particular, and of the dynamics of evolution in general.
These two questions about Computation have many different facets which are all interrelated. We refer to them under the global umbrella term of **Qualitative Computing**.

2 Numbers as building blocks for Computation

Natural integers such as $1, 2, 3, \ldots$ are recorded to have emerged in Sumer in the 3^{rd} Millennium BC, where scribes were skilled in basis 60. They used a positional representation and had a special mark for missing digits. However, neither them, nor the Greeks, which were the great computer scientists of the times, had the concept of zero, which is today so evident to us. Why?

2.1 Thinking the unthinkable

Because zero is not merely a number like any other. It is, before all, a formidable philosophical concept. It takes great courage to pose as an evident truth:"There exists the

*Université Toulouse 1 and CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique), 42 av. G. Coriolis, 31057 Toulouse Cedex, France, e-mail: chatelin@cerfacs.fr

†CERFACS, 42 av. G. Coriolis, 31057 Toulouse Cedex, France, e-mail: travies@cerfacs.fr

‡to appear as a Chapter in *Handbook for Numerical Computation,* Bo Einarsson editor, SIAM

non existence, and it is called Zero". Aristotle took the risk, then shied away from the logical impossibility. Finally the leap was taken gradually by various Indian thinkers, from Brahmagupta (600 A. D.) who conceived of zero to Bhaskara (1100 A. D.) who showed how to compute with it.

When unleashed and tamed by computation, zero allowed the art of Computing, initially developed in India, the Middle East and Central Asia, to blossom in the Western world. However, Algebra, the new Arabic art of Computing was at first met with strong opposition from the abacists which did not need zero to compute with their abacus. The new written arithmetic, done with pen and paper, eventually won because it was without rival to balance the checkbooks of the European merchants, while keeping track of all transactions. Once accepted, zero opened the door to the solution of equations, not only linear, but of degree 2, 3 and 4. Until a new challenge was met.

2.2 Breaking the rule

If the practical notion of debt and credit helped spreading the acceptance of *negative* numbers, it was not the case with the new concept $\sqrt{-1}$ created by Cardano to represent one of the two solutions of the equation $x^2+1=0$. Even with the computing rules enunciated by Bombelli, the strange $\sqrt{-1}$ was met by extremely strong opposition, as testified by the adjective *impossible* or *imaginary* which was invariably used to describe it. And the resistance was justified: $\sqrt{-1}$ is not a natural, or real, number, because its square is -1. It is not positive as it should to qualify: $\sqrt{-1}$ breaks the rule that for any real number $x \neq 0$, its square x^2 is positive. And it took another 300 years before complex numbers became fully accepted. This acceptance was brought about in two steps: first Euler invented the symbol i, then several scientists used the plane to give it a geometric interpretation. Because a complex number $a+ib$ is in effect *two* dimensional, it has 2 real components a and b on *two different* axes perpendicular to each other. The first, the real axis, represents the real numbers with positive square. And the second, the *imaginary* axis, represents "alien" numbers, the imaginary numbers with negative square. Therefore a complex number is a *new kind* of number, it is a mix of the two types of numbers, the real and the imaginary type, which are required for the complete solution of equations encountered in classical algebra.

The example of the introduction of 0 and $\sqrt{-1}$ shows clearly the long maturation process required, after their creation, for the use of new concepts which eventually turn out to be essential for our scientific representation of the world. But the story of Numbers does not stop with the realisation of the complex numbers, if one wants to go beyond classical algebra.

2.3 Hypercomputing: up the Dickson ladder

To make a long story short, the complex numbers open the door to more new numbers, the hypercomplex numbers which are vectors of dimension 2^k, $k \geq 1$, on which a *multiplication* can be defined, in conjunction to the usual vector addition. They form hypercomplex algebras in which the multiplication is defined recursively for $k \geq 1$, from that at the previous level, of dimension 2^{k-1}, by the Dickson doubling process. Two families of such algebras are important for applications:

i) the family of *real* algebras A_k, starting from $A_0 = \mathbb{R}$, the field of real numbers,

and

ii) the family of *binary* algebras B_k, starting from $B_0 = \{0,1\}$, the binary algebra $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ of computation mod 2.

The two families have, in a sense, complementary, properties. The real algebras express Euclidean geometry, their multiplication is not commutative for $k \geq 2$ and not associative for $k \geq 3$. Difficult and spectacular results in algebraic and differential topology, as well as in Lie groups, rely on the first four algebras A_0 to A_3, which are the four division algebras: the reals $\mathbb{R}$, the complex numbers $A_1 = \mathbb{C}$, the quaternions $A_2 = \mathbb{H}$ and the octonions $A_3 = \mathbb{G}$. Various models of the Universe in Theoretical Physics use also these algebras, as well as the hexadecanions A_4.
The binary algebras, on the other hand, do not yield a scalar product, their multiplication is associative and commutative for all $k \geq 0$. B_0 is the usual binary algebra (mod 2), and B_1 explains easily the $\sqrt{\text{not}}$ logical gate realised by quantum interference [4]. With an appropriate labeling of the sequences of 0 and 1 of dimension 1, 2, 4, the algebras B_0, B_1 and B_2 are closely related to computation mod 2, 4 and 8 respectively.

What makes the hypercomplex numbers so essential in Mathematics and Physics? They allow to multiply! And multiplication is at the heart of any serious computation, as we are taught by the Newcomb-Borel paradox.

2.4 The Newcomb-Borel paradox

In 1881, the astronomer Simon Newcomb reported the experimental fact that the significant digits of a number chosen at random from physical data or computation were not uniformly distributed. He stated that, instead, the logarithm of their mantissa is uniformly distributed.
In 1909, the mathematician Emile Borel proved that, if one chooses at random a real number between 0 and 1, its decimal digits are almost surely equally distributed.
These two truths, one experimental and the second mathematical, seem at odds [9]. Can they be reconciled?
Yes, if we realise that Borel and Newcomb do not see the same numbers...

Borel considers the static additive representation for $x \in \mathbb{R}^+$

$$x = [x] + \{x\},$$

where $[x]$ is the integer part of x, $0 \leq \{x\} < 1$.
Newcomb, on the contrary, looks at the dynamic floating point representation in base β for $x \in \mathbb{R}^+$:

$$x = \beta^{[\log_\beta x]+1} \beta^{\{\log_\beta x\}-1}$$

which is multiplicative. It is the representation actually used to compute in base β.
The law of Newcomb follows immediately from a theorem by P. Lévy (1939) [11] about the sum of random variables mod 1: $\{\log_\beta x\}$ is uniformly distributed on $[0,1]$.
The law of Newcomb is ubiquitous in intensive computing. It has found its way to Wall Street and the Internal Revenue Services (to detect frauds).

In any result of a serious computational process, the first decimal digit has 3 times more chances to be 1 than 9. The Newcomb-Borel paradox is a welcome reminder of

the fact that any actual computation creates meaning: the Newcomb law allows to discriminate between the leading digits which bear meaning, and the trailing digits which are uniformly distributed [9].
The Newcomb view (based on $\times$, and the floating point representation) is very different from Borel's (based on +). The paradox puts forward the natural hierarchy between operations on numbers which follow from Computation in a given basis: $\times$ is dominant over + and creates meaning in the chosen basis.

2.5 The discrete, the continuous and the connected

It was already clear to the Pythagorean School that Numbers have a dual personality : they can be perceived as discrete, as the integers 1,2,3,..., or continuous, as on the number line.
This dual character shows up, in the axiomatic presentation of the construction of Numbers based on addition, in the notion of limit of a Cauchy sequence. One creates the integers by repeatedly adding 1 to the preceding number, starting from 1 (or 0). Then solving $ax = b$ yields the set of rationals $\mathbb{Q}$. The real numbers (resp. complex numbers) are the *closure* of the rational (resp. algebraic) numbers.

Continuity is therefore a limit property, which might be viewed by some as superfluous. And, indeed, the finitist programme of Hilbert wanted to exclude any recourse to a limit process. The 1-dimensional version of this programme has been shattered by Gödel, Turing and, most of all, by Chaitin, in a way which allows Randomness to invade formal Mathematics, the domain of ultimate rigor [1]. Randomness (expressed as program size) exposes the limit of Turing computability with numbers which are 1-dimensional.

Does it mean that any finitist programme is doomed to failure, as many have argued in the aftermath of Gödel's incompleteness result for Arithmetic? To better understand the issues at stakes, let us look at another finitist programme, the Greeks programme, which was more flexible than Turing's in a subtle but essential way. They allowed all operations which could be realised with a ruler and a compass. This means, broadly speaking and in modern vocabulary, that they allowed quadratic numbers (i.e. numbers which are solutions of equations of degree 2 with integer coefficients). When they could not produce an exact rational solution, they worked with rational approximations of irrational numbers. They clearly realised that the world could not be captured with rational numbers only. The profoundity of their programme, the Greek miracle, was that they had a working compromise, by means of Geometry in 2 and 3 dimensions and of successive approximations, between what they conceived of computing and what they could actually compute. Time only adds to the shining perfection of Archimedes'method to approximate the transcendental number π.

It becomes clear that the main difference between Turing and the Greeks is in terms of the dimension of the numbers they allow in their computing game. In both cases, the procedure is algorithmic. But Turing numbers are 1-dimensional: they are discrete points on the number line. Whereas the Greeks numbers are 2-dimensional, their variety of quadratic numbers cannot be represented on a line, but in a plane. And we know that certain problems apparently stated on 1-dimensional numbers can be solved only by a call to 2-dimensional numbers. The most famous example is the d'Alembert-Gauss theorem of Algebra: any polynomial of odd degree with real coefficients has

at least one real solution. It cannot be proved without going complex. This is the first instance of the "topological thorn" planted in the "flesh" of algebra... The phenomenon is intimately connected with the notion of connectivity which is sensitive to the topological dimension.
Remarkably, at the time that the Austrian logician Gödel discovered the incompleteness of Arithmetic (1931), the Polish logician Tarski had already shown the completeness of elementary 2D-Geometry (1930) [12]. In other words, Randomness can exist with 1D numbers but not with 2D numbers! It is worth commenting that the *negative* result of Gödel attracted much more attention than the *positive* result of Tarski. The greater generality of Gödel's result overshadowed the constructive power of Tarski's own result. This is still true today, despite the fact that Tarski's result is at the foundation of Computer Algebra, one of the success stories of Computer Science!

In summary, the problem is not in the limit (potential versus actual infinity). It is in the **number of dimensions** of the building blocks used for Computation. And as Dickson showed, once you allow two dimensions instead of one, you might as well allow any number 2^k for dimension: you construct recursively all hypercomplex Numbers!

It is interesting to remark that Turing's dimensional limitation has been bypassed in the 1980s by the emergence of **Quantum Computing**, a theory of computability based on 2D-numbers to represent the probability amplitudes of Quantum Mechanics. Quantum Computing can be seen as the computer age version of the Greeks programme [4].

3 Exact versus Inexact Computing

3.1 What is Calculation?

Calculation is a transformation of information from the implicit to the explicit. For example, to solve an equation $f(x) = g$, where f and g are known, is to find the solution x. The implicit solution x becomes explicit in the form $x = f^{-1}(g)$. This can seldom be performed exactly. As an example, only polynomial equations in one variable of degree less than 5 are explicitly soluble with radicals. So one very quickly faces the need for approximation techniques, which is the domain of mathematical analysis.
But suppose we know how to compute $x = f^{-1}(g)$ exactly, a new and equally important question arises. Is the exact resolution always pertinent to understanding the visible world around us, a world of phenomena perceived with limited accuracy?
The fool says yes (exact is always better than approximate...), but the wise realises that in certain phenomena of very unstable evolution, the exact solution of the model can be so unstable that it is not realised in the physical world, because its window of stability is below the floor limit necessary for us to see it.

3.2 Exact and Inexact Computing

It is very important to keep in mind that the challenge of calculation is expressed by the two questions "how" and "why", as we recalled in the introduction.
The answer to the question of "how" appears easy: it suffices in principle to use an **algorithm**, a mechanical procedure, which after a finite number of steps, delivers a solution. However, it cannot be that simple in practice: everyone has experienced the

frustration created by a bad algorithm!

One of the difficulties of the "how to design a good algorithm" is that the designer should have a clear view of what is the type of understanding that is expected from running the algorithm. He should know "why compute". Because there are two very different types of world for which insight can be sought:

i) a type of world where accuracy is **unlimited**, as exemplified by Mathematics, or

ii) a type of world where accuracy on the available data is **intrinsically limited**, as in the phenomenological world of Natural Sciences (Physics, Biology,...).

To conveniently refer to the act of computing in one of these two types of world, we shall speak of *Exact* versus *Inexact Computing*.

3.3 The computer arithmetic

Implicitly, theoretical mathematical analysis applies to a world of type i) where abstract notions such as convergence to 0, exact arithmetic, equality, do have a meaning. This is why certain numerical methods which are proved convergent in mathematics fail miserably on a computer: they ignore the issue of algorithmic reliability required by the limited precision of the computer arithmetic.

The basic principles to address such a question, with a strong emphasis on "why", are given in **Lectures on Finite Precision Computation** [3], where the key notion of a **reliable** algorithm is developed.
In a nutshell, a reliable algorithm shields the user from the possibly negative consequences of the limited precision of the computer arithmetic.

> Expert numerical software developers make the necessarily finite precision of the computer arithmetic become transparent with respect to the effect of the limited accuracy available on the data.

More on reliable algorithms and the toolbox PRECISE to test reliability on a computer can be found in Chapter ??? of this book [5].

If a reliable algorithm is used, then the computer arithmetic is never to be blamed for an unexpected behaviour. On the contrary, it is an asset in the sense that it can reveal a computational difficulty, but it cannot create it.
The bad reputation of the computer arithmetic is therefore largely undeserved. It may not be the best of all possible arithmetics for Exact Computing (like for Mathematics), but it is, for essential reasons, without rival for Inexact Computing (that is for Experimental Sciences). It allows computer simulations to capture aspects of the phenomenological reality that exact computation would miss, specially in the area of chaotic evolution.

Despite its powerful practical implications for Science and Technology, this fact is far from being appreciated, even by software developers. So strong is the appeal of **exact** computing, even in a world of limits...

3.4 Singularities in Exact and Inexact Computing

Singularity and regularity are mathematical notions which give a technical content to the idea of an abrupt change: a property which was present disappears suddenly in certain conditions, or a new property emerges. Singularities are, in some sense, exceptional with respect to the background which is regular.
In many areas of classical mathematics, singularities can be forgotten because they are not generic (Sard, 1942 in [3]) : they form a set whose interior is empty. Does this extend to Inexact Computing?

Not at all. Let us look at the simple example of a matrix A. The singular points of the resolvent map $z \to (A - zI)^{-1}$ are the eigenvalues of A. They form the *spectrum* of A, a finite set of points in $\mathbb{C}$ of *zero measure*.
Now let us assume that we cannot distinguish between the given matrix A and any matrix of the form $A + \Delta A$, for ΔA such that $\| \Delta A \| \leq \alpha$, where α denotes the level of uncertainty on the data. Therefore any z in $\mathbb{C}$ which is an eigenvalue of $A + \Delta A$, but not of A, will nevertheless be interpreted as an eigenvalue of A. The spectrum of A in Exact Computing becomes, in Inexact Computing, the *pseudospectrum* of A (relative to the level α), that is

$$\{z \text{ eigenvalue of } A + \Delta A \text{ for any } \Delta A \text{ such that } \| \Delta A \| \leq \alpha\},$$

which is a closed set of **positive Lebesgue measure** in $\mathbb{C}$.

The classical theory of singularities does not apply in Inexact Computing, because singularities have a positive measure.
As a consequence, singularities cannot be ignored. Their influence can be enormous. Chapter 11 in [3] provides illuminating examples of this phenomenon.

3.5 Homotopic deviation

Considering a closeby matrix $A' = A + \Delta A$ to analyse the properties of A is one of the favourite devices in the Numerical Analyst's toolbox. And it can go a long way to give a useful description of the local neighborhood of A, for $\| \Delta A \|$ small enough [3, 7, 8, 10]. This amounts to explain a computation on A by the *local* topology in the variety of matrices around A.

If one is to take the consequences of limited accuracy seriously, one should not, however, rule out the possibility of non local effects in which ΔA is not small in a metric sense. In other words, how can we weaken the constraint of perturbation theory, which is expressed by $\| \Delta A \|$ small enough?

One possibility is by means of an **homotopic deviation**. In addition to A, let be given a deviation matrix E. We introduce a complex parameter t to define the homotopic family of matrices $A(t) = A + tE$, such that $A(0) = A$. The family realises an homotopic deviation of A, associated with E.
A linear problem such as solving a set of n linear equations in n unknowns is regular whenever the matrix of the system is regular, that is invertible. Associated with a matrix A, there are two dual classes of problems: the regular (resp. singular) problems, which are associated with the matrix $A - zI$, for $z \in \mathbb{C}$, whenever it is invertible (resp. singular).

Solving linear systems with the matrix $A - zI$ of full rank are all regular problems associated with A, while the eigenproblem for A is the associated singular problem. We address the general question: is it possible to relate the singular/regular problems posed on $A(t) = A + tE$ to the ones posed on A with no assumption on E? The answer turns out to be remarkably simple. This is possible by analysing the factorization,

$$A + tE - zI = (I + tE(A - zI)^{-1})(A - zI)$$

where $z \in \mathbb{C}$ is not an eigenvalue of A, that is $z \in \mathbb{C} \setminus \sigma(A)$. Let μ_z denote an eigenvalue of $F_z = -E(A - zI)^{-1}$, for $z \notin \sigma(A)$. The point $z \notin \sigma(A)$ is an eigenvalue of $A + tE$ iff there exists an eigenvalue $\mu_z \neq 0$ of F_z such that $t\mu_z = 1$.
Any z in $\mathbb{C}$ which is not an eigenvalue of A can be interpreted as an eigenvalue of at least one matrix $A + tE$, as long as $0 < \rho(F_z) < \infty$. This is not possible if $\rho(F_z) = 0$, because all $\mu_z = 0$, and t is not defined.
In parallel to this interpretation, any z such that $0 < \rho(F_z) < \infty$ can receive an interpretation as a regular point: $A + tE - zI$ is invertible for any t such that $t \neq \frac{1}{\mu_z}$.
The *two* interpretations (z eigenvalue of $A + tE$, versus $A + tE - zI$ invertible)*hold in parallel* for any z in $\mathbb{C} \setminus \sigma(A)$such that $\rho(F_z) > 0$. We introduce the

Definition 3.1 *A point $z \in \mathbb{C} \setminus \sigma(A)$ such that $\rho(E(A - zI)^{-1}) = 0$ is a critical point associated with (A, E). We denote by $K(A, E)$ the set of critical points for (A, E), and write $\Sigma(A, E) = \sigma(A) \bigcup K(A, E)$.*

Do critical points exist, that is can $K(A, E)$ be non empty?
We know that $\lim_{|z| \to \infty} \rho(F_z) = 0$. Do there exist points z at finite distance such that $\rho(F_z) = 0$?
If E is rank 1, such that $E^2 \neq 0$, the answer is yes in general. There exist at most $n - 1$ points in $\mathbb{C}$ such that $F_z^2 = 0$, hence $\rho(F_z) = 0$ [6, 2].

What is the meaning of critical points from the point of view of computation?
If $\zeta \in K(A, E)$, then ζ cannot be interpreted as an eigenvalue of $A + tE$. There is only *one* possible interpretation: $A + tE - \zeta I$ is a regular (invertible) matrix for which we know a $finite$ representation, valid for any t:

$$\begin{array}{rcl} (I - tF_\zeta)^{-1} & = & I + tF_\zeta \qquad \text{and} \\ (A + tE - \zeta I)^{-1} & = & (A - \zeta I)^{-1}(I + tF_\zeta) = R(t, \zeta). \end{array}$$

The analyticity domain (with respect to t) in $\mathbb{C}$ of $R(t, z)$, which is $\{t;\ |\ t\ |\ \rho(F_z) < 1\}$ for z in $\mathbb{C} \setminus \sigma(A)$ but not in $K(A, E)$, becomes $all\ of\ \mathbb{C}$ when z is a critical point. The convergence in t is not conditional to z : z and t are independent. This is a non local effect which depends on the algebraic structure of E (rank E =1) and not on a metric condition on $\|E\|$.

The above discussion can be summarized by looking at the properties of the resolvent matrix $R(t, z) = (A + tE - zI)^{-1}$ as a function of the two complex parameters t and z. Three properties are of interest: existence or non existence (singularity), and, if existence, analyticity in t for a given z.
The choice of one property induces a specific relationship between z and t which is listed in Table 1. In this Table, μ_z denotes any eigenvalue of $F_z = -E(A - zI)^{-1}$ which is not 0 and $\rho_z = \rho(F_z)$.

Singularity = Non Existence	Yes	i) $t = 0$ and $z \in \sigma(A)$ ii) $t = \frac{1}{\mu_z}$, $\mu_z \neq 0$ and $z \in \mathbb{C} \setminus \Sigma(A,E)$
	No	$z \in K(A,E)$

Existence	Yes	$t \neq \frac{1}{\mu_z}$, $\mu_z \neq 0$ and $z \in \mathbb{C} \setminus \sigma(A)$
	No	$t = 0$ and $z \in \sigma(A)$
Analyticity	Yes	i) asymptotic : $\mid t \mid < \frac{1}{\rho_z}$ and $z \in \mathbb{C} \setminus \Sigma(A,E)$ ii) polynomial : $t \in \mathbb{C}$ and $z \in$ and $z \in K(A,E)$
	No	$t = 0$ and $z \in \sigma(A)$

Table 1: Properties of $R(t,z)$ as a function of t and z

Table 1 shows clearly that the nature of the coupling between t and z changes with the point of view. The change can be smooth or abrupt, and this has computational consequences. The "why" rules the "how".

3.6 The map $\varphi : z \rightarrow \rho(F_z)$

The map $\varphi : z \rightarrow \rho(F_z)$ will provide a useful graphical tool to analyse homotopic deviation. The role of φ comes from the fundamental property that it is a **subharmonic** function of z. A subharmonic function is the analogue, for a function of two real variables, of a convex function in one real variable. The level curves $z \rightarrow \rho(F_z)$=constant are closed curves which enclose eigenvalues of A (maximal value $\rho = +\infty$) or local minima of ρ with values ≥ 0. There are two values for which the map φ has special properties: the value ∞ and the value 0.

a) the value $\rho = \infty$: there exist n points in $\mathbb{C}$, the n eigenvalues of A such that $(A - zI)^{-1}$ does not exist, and therefore $\rho(E(A - zI)^{-1}) = +\infty$ for any matrix E. The spectrum of A is the set of singular points of $z \rightarrow (A - zI)^{-1}$. It also belongs to the set of singular points of φ, because ρ is not defined at any eigenvalue. At such points in $\mathbb{C}$, the value of $\text{rank}(A - zI)$ jumps from n to a value $\leq n - 1$.

b) the value $\rho = 0$: we know that $\rho \rightarrow 0$ as $\mid z \mid \rightarrow \infty$. The critical points at finite distance which satisfy $\rho = 0$ belong also to the set of singular points of φ. At such points, there is a qualitative change in the behaviour of the Neumann series. It jumps from an asymptotic representation (infinity of non zero terms) to a finite

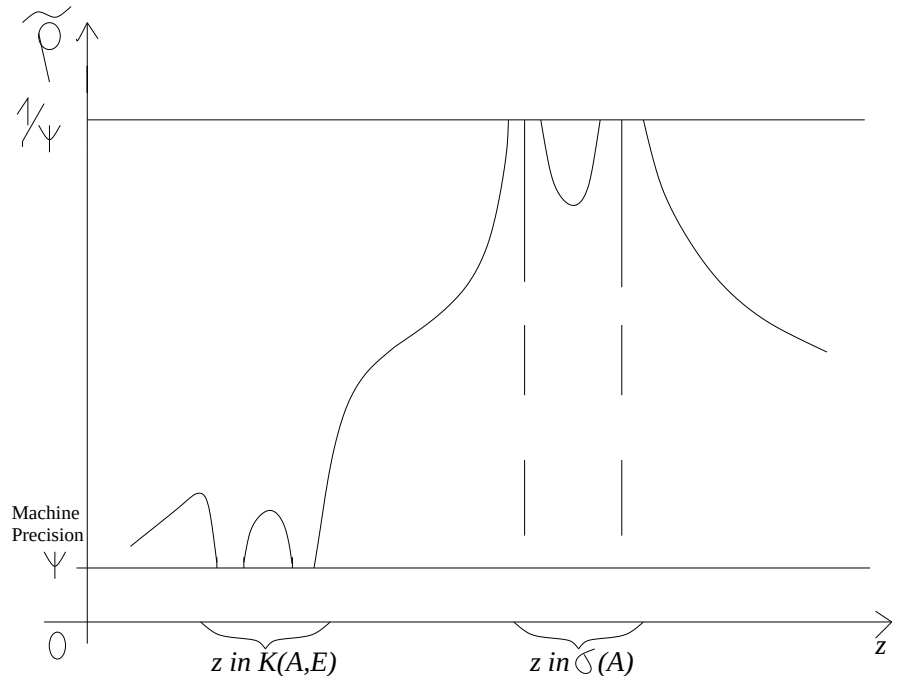

Figure 1: *Profile of the computed map* $\tilde{\varphi} : z \to \tilde{\rho}(E(A-zI)^{-1})$ *for* E *of rank 1*

one (for example, two non zero terms when rank $E = 1$).

This shows that globally the properties of $A+tE$ can be affected not only by the eigenvalues of A, but also by the critical points of (A,E) when they exist. The singular points of φ (where ρ takes the value 0 or $+\infty$) consist of both the eigenvalues of A and the critical points of (A,E).

In Exact Computing, these singular points have zero measure and their influence is almost surely nonexistent.
However this cannot be true any longer in Inexact Computing. And the phenomenon is clearly visible if one looks at the computed map $\tilde{\varphi} : z \to \tilde{\rho}(E(A-zI)^{-1})$. This amounts to take machine precision $\psi \sim 10^{-16}$ as the relative level of uncertainty on the data, see Figure 1. The maps φ and $\tilde{\varphi}$ do not describe the same reality. The map φ expresses the view of homotopic deviation in Exact Computing. Whereas the computed map $\tilde{\varphi}$ expresses the view in Inexact Computing. The two views differ in the region of $\sigma(A)$ in a way which is well understood, thanks to the notion of pseudospectrum. They also differ in the region of $K(A,E)$ in a way which has been uncovered only recently [6].

In both cases, the effect of Inexact Computing is to replace the set of singularities of measure 0 by a set of positive measure. In the case of $\sigma(A)$, this is usually viewed as negative because this decreases the domain where $A+tE-zI$ is invertible. But, in the case of $K(A,E)$, this has definitely a positive flavour: on a region of *positive* measure, the convergence is **better than predicted** because the computational process is essentially finite rather than asymptotic. The role of finite precision on the singular points of

φ is twofold:

i) in the pseudospectrum region around $\sigma(A)$, it makes the eigenvalues of A appear *closer* than they are in exact arithmetic (this is a *local* effect),

ii) in the critical region around $K(A,E)$, it makes the eigenvalues of A appear *further* than they are (this is a *non local* effect). Or said differently, it makes the solution of linear systems on the computer easier than predicted by mathematics: the computation is essentially a finite process.

For a general E, the set $K(A,E)$ can be empty. But we know that if E is of rank 1, $E^2 \neq 0$ then $K(A,E)$ consists of at most $n-1$ points in $\mathbb{C}$.

The question asked at the beginning of paragraph 3.5 has been answered in the following way by homotopic deviation: the introduction of the parameter t has allowed to replace the metric condition $\| E \|$ small enough by the algebraic structure condition E of rank 1. Meanwhile, this has complemented the set of n eigenvalues $\sigma(A)$ by the set of $n-1$ critical points $K(A,E)$. In the neighborhood of these points finite precision creates **computational opportunities** by providing better convergence than predicted: the computational process is finite rather than asymptotic.

In Figure 1, the two regions have been represented as neatly separated. This need not be the case. Let us look at an example where eigenvalues of A and critical points of (A,E) are interestingly intertwined.

3.7 Illustration

We illustrate the use of the map φ to analyse homotopic deviation on the following example. We call **One**(n) the matrix A of order n which is the companion matrix associated with the polynomial

$$p_n(x) = \sum_{i=0}^{n} x^i = x^n + p_{n-1}(x),\ n \geq 1.$$

The eigenvalues of A are the roots of $p_n(z) = 0$, that is the n roots of $z^{n+1} = 1$ distinct from 1.
We choose the deviation matrix E such that $A+E$ is the companion matrix associated with x^n. That is, we set $E = ee_n^T$, with $e = (1,...,1)^T$.
E is of rank 1. The critical points of (A,E) are the roots of $p_{n-1}(z) = 0$. Therefore the n eigenvalues and the $n-1$ critical points are the roots of 1 of order $n+1$ and n respectively, which are different from 1. They are intertwined on the unit circle $|z| = 1$.
We set $n = 8$. Figure 2 (resp. 3) displays the map $\varphi : z \to 1/\rho(E(A-zI)^{-1})$ (in logarithmic scale) in 2D (resp. in 3D). The black curve is the level curve $\rho = 1$. Figure 4 (resp. 5) displays in 2D (resp. 3D) a set of 11 level curves $\rho = constant$ corresponding to the values $\{0.001,\ 0.03,\ 0.05,\ 0.1,\ 0.4,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.8,\ 2\}$.
Figure 6 (resp. 7) displays the same level curves with the colour used to parameterize the variation of θ on $[0; 2\pi[$ in 2D (resp. 3D). The colour chart varies from blue to red for increasing values of the parameter r or θ.

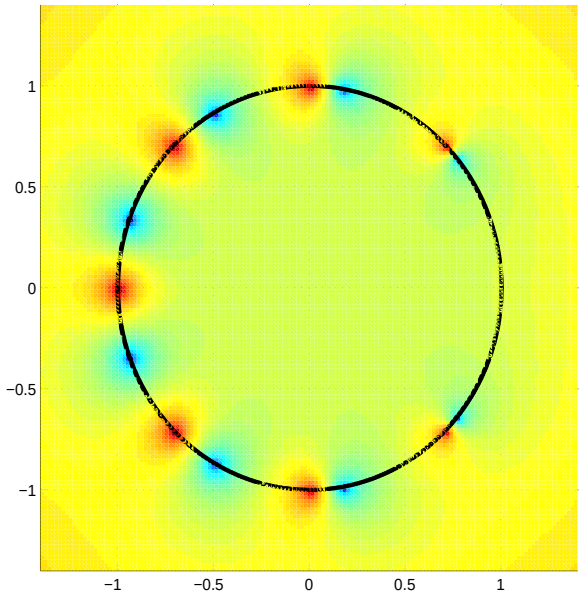

Figure 2: The map $\varphi : z \to 1/\rho(E(A-zI)^{-1})$ for the matrix One ($n = 8$) in 2D

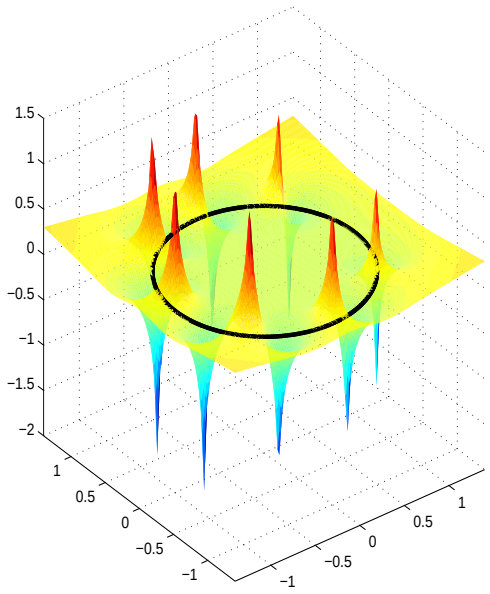

Figure 3: The map $\varphi : z \to 1/\rho(E(A-zI)^{-1})$ for the matrix One ($n = 8$) in 3D

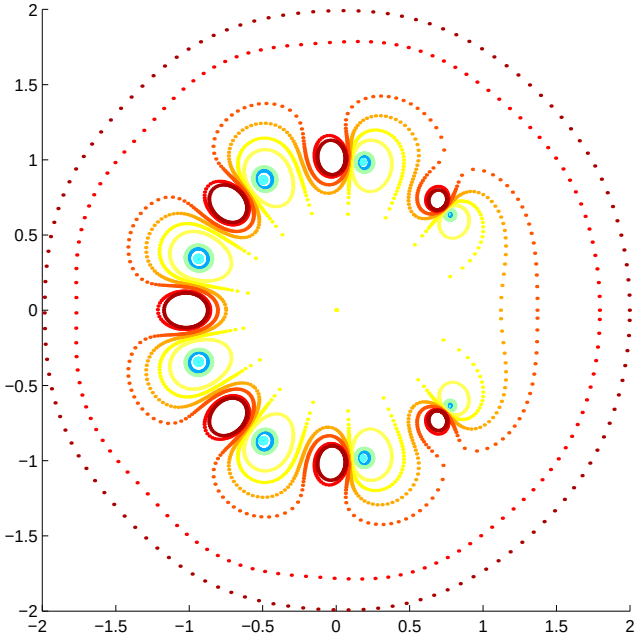

Figure 4: 11 level curves $\{0.001,\ 0.03,\ 0.05,\ 0.1,\ 0.4,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.8,\ 2\}$ for the matrix One ($n = 8$) in 2D

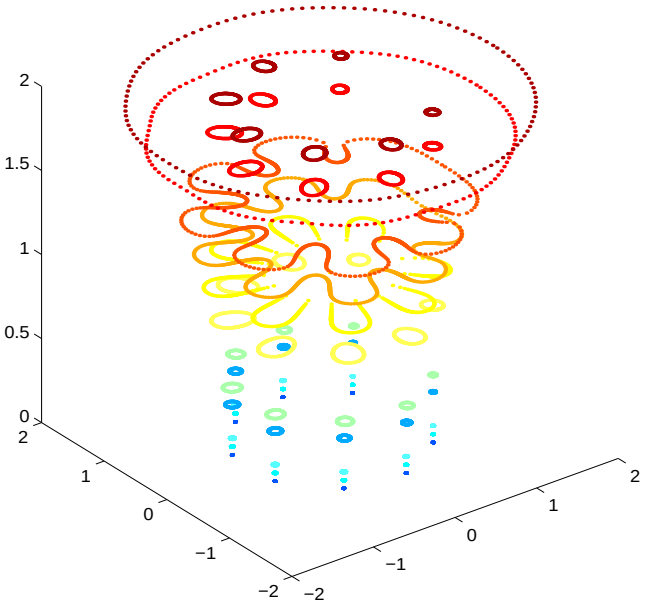

Figure 5: 11 level curves $\{0.001,\ 0.03,\ 0.05,\ 0.1,\ 0.4,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.8,\ 2\}$ for the matrix One ($n = 8$) in 3D

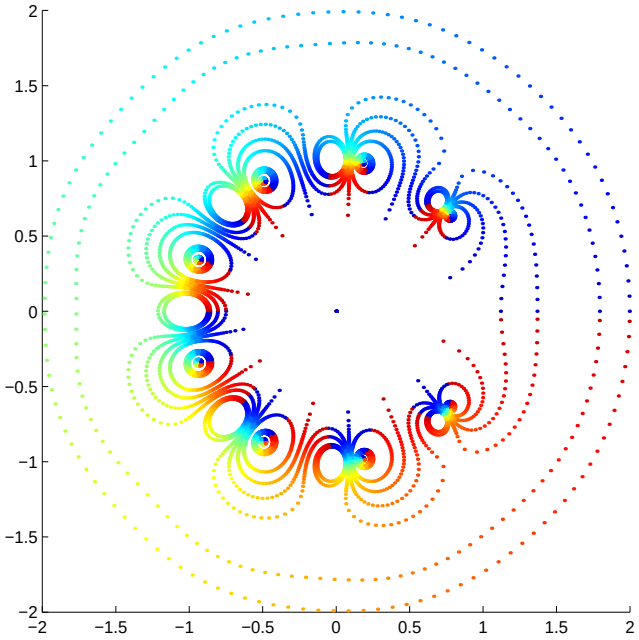

Figure 6: 11 level curves $\{0.001,\ 0.03,\ 0.05,\ 0.1,\ 0.4,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.8,\ 2\}$ for the matrix One ($n = 8$) in 2D parameterized by θ on $[0; 2\pi[$

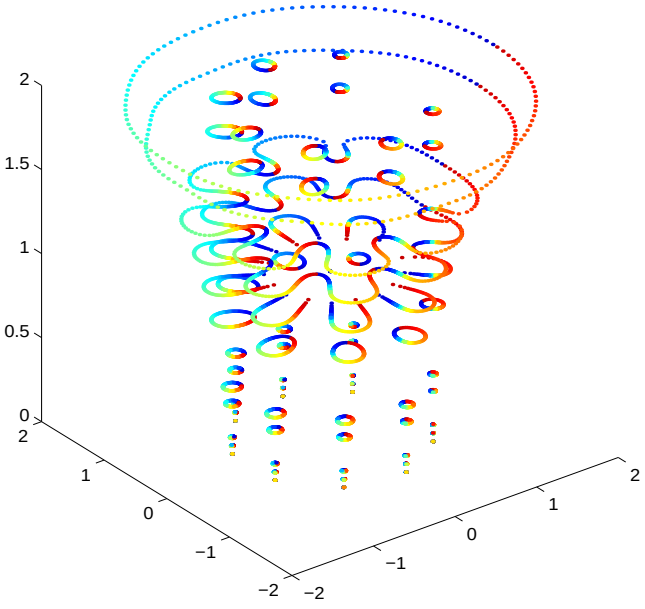

Figure 7: 11 level curves $\{0.001,\ 0.03,\ 0.05,\ 0.1,\ 0.4,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.8,\ 2\}$ for the matrix One ($n = 8$) in 3D parameterized by θ on $[0; 2\pi[$

4 Numerical Software

As we already pointed out in the previous section, the design and use of good numerical software destined to get insight on a particular question of Experimental Sciences, is best analysed with the conceptual tools of Inexact Computing. Most important amongst them is the method called **Backward error analysis** which emerged in the 1950s and 1960s, mainly under the influence of Wilkinson. The basic idea has the ideal beauty of simplicity: just consider any computed solution (any output of computer simulations) as the exact solution of a nearby problem. The curious reader should also look at Chapter ??? [5] which presents a complementary viewpoint on backward error analysis: it gives guidelines on how to realise such an analysis in practice.

4.1 Local error analysis

One of the most important features of backward error analysis is that it allows to factor out in the error bound the contribution of the algorithm (the backward error) from the contribution of the problem (the condition number).
Let us suppose that the problem to be solved is $Ax = b$. A computed solution $\tilde{x}$ is interpreted as the solution of a nearby problem of the same form, that is

$$(A+\Delta A)\tilde{x} = b+\Delta b \tag{1}$$

In general, ΔA and Δb are not unique, and are not known. However, in this simple case, there are formulae to compute $min(\|\Delta A\|, \|\Delta b\|)$ which is the backward error associated with $\tilde{x}$. Subtleties of perturbations ΔA, Δb are discussed at length in [3], see also Chapter ??? [5].
Therefore a local error analysis can be carried out by means of the 1st order bound

$$\textbf{Forward error} \leq \textbf{Condition Number} \times \textbf{Backward error}$$

whenever the problem to be solved is *regular*.

This approach can be easily related to the homotopic deviation presented in the previous section. If we choose to allow only perturbation on A, but not on b, then equation (1) becomes $(A+\Delta A)\tilde{x} = b$, where ΔA is a perturbation matrix which is *unknown*, we only have access to the metric quantity $min\,\|\Delta A\|$. Because there is a condition on $\|\Delta A\|$, we call ΔA a *perturbation* of A. And, as said previously, we reserve the term *deviation* for E without metric constraint.

4.2 Homotopic deviation versus normwise perturbation

It is interesting to compare the homotopic deviation theory of section 3 to the more traditional approach of *normwise perturbation theory* [3]. In particular, the normwise point of view leads to the notion of **normwise pseudospectrum** of a matrix A of level α ([3], chapter 11) which is defined for any $\alpha > 0$, as:

$$\begin{aligned}&\{z \in \mathbb{C}, \text{ z is an eigenvalue of } A+\Delta A, \text{ for all } \Delta A \text{ such that } \|\Delta A\| \leq \alpha\}\\ &= \{z \in \mathbb{C}; \left\|(A-zI)^{-1}\right\| \geq \tfrac{1}{\alpha}\}.\end{aligned}$$

The map $\psi : z \to \left\|(A-zI)^{-1}\right\|$ defined on $\mathbb{C} \setminus \sigma(A)$ is also subharmonic. However, the properties of a norm exclude the possibility of critical points.
Let us compare the maps φ and ψ on the following example.

Example 4.1:

We call **Venice** the companion matrix B associated with the polynomial

$$\begin{aligned} p(x) &= (x-1)^3(x-3)^4(x-7) \\ &= x^8 - 22x^7 + 198x^6 - 958x^5 + 2728x^4 - 4674x^3 + 4698x^2 - 2538x + 567. \end{aligned}$$

And we consider the rank one perturbation $E = ee_8^T$ with $e = (1,...,1)^T$ and $e_8 = (0,...,0,1)^T$. We set $A = B - E$: it is the companion matrix associated with the polynomial $q(x) = p(x) + r(x)$, with $r(x) = \sum_{i=0}^{7} x^i$. The 7 zeros of $r(z) = 0$ on the unit circle are the critical points of (A,E). Figures 8 and 10 display respectively the maps $\varphi : z \to \rho(E(A - zI)^{-1})$ and $\psi : z \to \left\|(A - zI)^{-1}\right\|$.
The critical zone for φ is clearly visible on Figures 8 and 9: there is a very sharp sink. Interestingly, the map ψ on Figure 10 displays also a local minimum in the same region of the complex plane. We know from theory that the exact value of this minimum is necessarily positive.

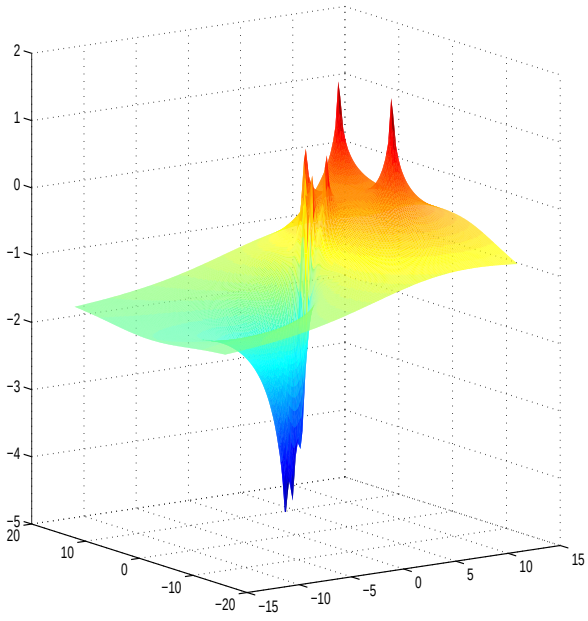

Figure 8: Map $\varphi : z \to \rho(E(A - zI)^{-1})$, $A = B - E$ with B=Venice and E is rank 1

4.3 Homotopic deviation with E of rank 2 in finite precision

When E is a matrix of rank 2, the eigenvalues of F_z which are not necessarily zero are that of a 2×2 matrix which we denote M_z.
F_z is nilpotent such that $F_z^3 = 0$ iff $M_z^2 = 0$. And M_z is nilpotent iff z satisfies the system

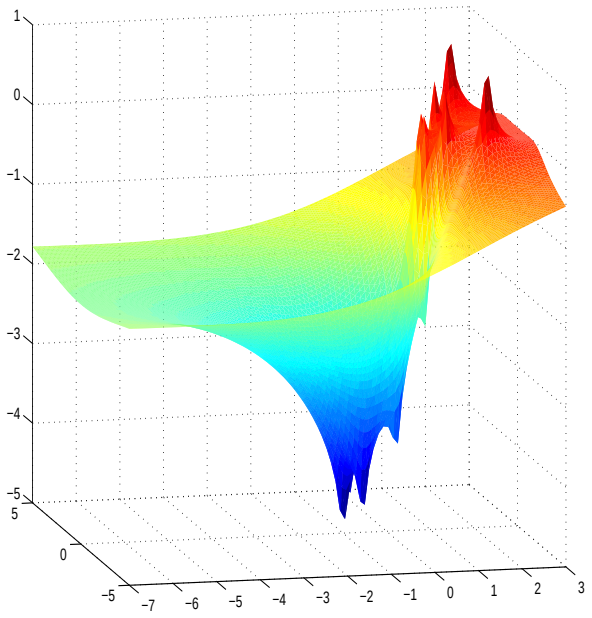

Figure 9: Map $\varphi : z \to \rho(E(A-zI)^{-1})$, $A = B - E$ with B=Venice and E is rank 1 - Zoom around the critical points

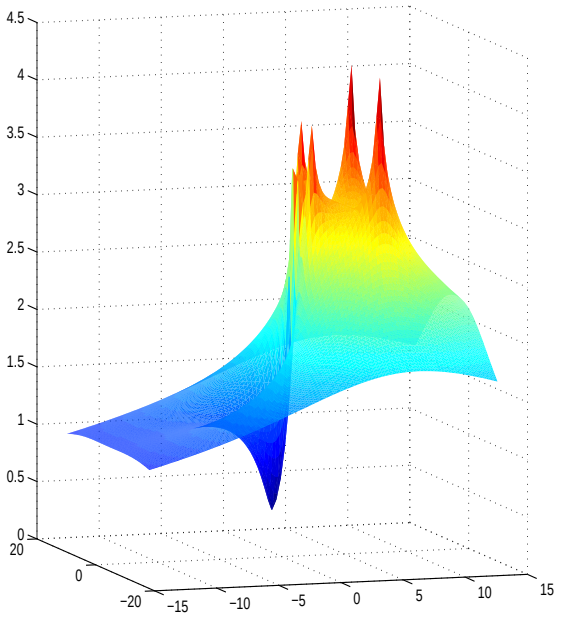

Figure 10: Map $\psi : z \to \left\|(A-zI)^{-1}\right\|$, $A = B - E$, with B =Venice

of 2 polynomial equations

$$\begin{cases} tr\ M_z = 0 & (\text{degree} \leq n-1) \\ det\ M_z = 0 & (\text{degree} \leq 2n-2) \end{cases}$$

Each equation has a set of roots which in general is disjoint one from the other. However in finite precision they might be close to being solutions of the two equations.

This fact is confirmed by the following example.

Example 4.2:

We consider again the matrix B called Venice of Example 4.1. And we consider the rank 2 deviation $H = ee_8^T + e_1 e_6^T$. We set $C = B - H$: C is not a companion matrix anymore. We give in Figure 11 the map $\varphi : z \to \rho(H(C-zI)^{-1})$. One sees distinctly acute sinks, which, however, are not as marked as those displayed on Figure 9. A detailed analysis of Venice can be found in [6].

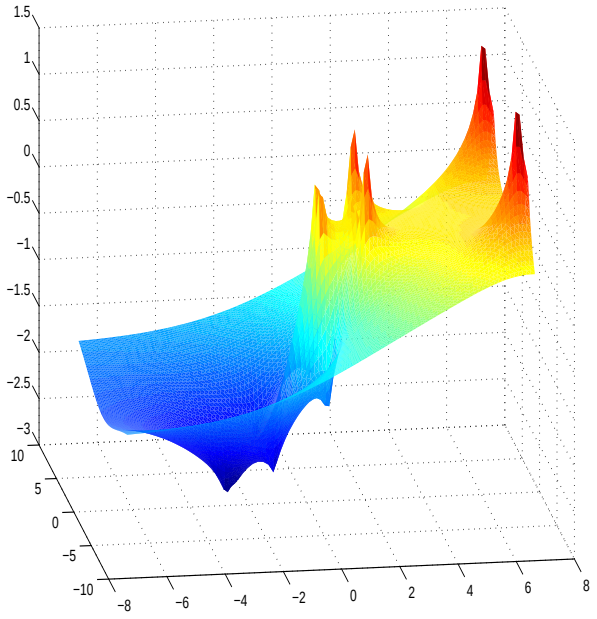

Figure 11: Map $\varphi : z \to \rho(H(C-zI)^{-1})$, $C = B - H$, with B=Venice and H of rank 2.

5 The Lévy law of large numbers for Computation

The law of large numbers which assesses the activity of measurements (data collection and interpretation) is given by the *Laplace-Gauss normal distribution*: the average (arithmetic mean) of a large number of random variables has a normal distribution.
Because this law is so ubiquitous in Experimental Sciences, it has been assumed that it also rules Computation. And indeed, this can be the case in very specific situations.

But **not in general**.

We already indicated that the Newcomb-Borel paradox means that the *Lévy uniform distribution* is at work in floating point computation (in a given base) and induces the dominant role of multiplication over addition. The Lévy law concerns the limit distribution of the sum mod 1 of N random variables as $N \to \infty$ [12].

In floating point computation, it applies to the logarithms of the mantissa. Another important domain of application of the Lévy law is clearly the multiplication of the complex numbers in the trigonometric form of Euler: the arguments add mod 2π, therefore they are uniformly distributed in $[0, 2\pi[$. This fact accounts for some of the differences, in Mathematics, between Analysis in real or complex variable.

Another way to reflect upon these phenomena is to realise that there are major differences, from the point of view of Computation, between working with numbers of dimension 1 (which are scalars) or with numbers of higher dimension (dim > 1).
We mentioned this fact when discussing the two finitist programmes of Hilbert and of the Greeks.
At that point, the *topological* notion of connectivity arose. Later, when presenting homotopic deviation, we encountered the *algebraic* notion of a nilpotent matrix (all the eigenvalues are zero but the matrix itself is nonzero), which enables a generically non polynomial computation to become polynomial at critical points.
These are two prominent aspects (Topology and Algebra) under which the dimensional quality of Numbers manifests itself most naturally in Computation.

References

[1] G. J. Chaitin. *Exploring randomness*. Springer Verlag, Singapore, 2001.

[2] F. Chaitin-Chatelin. About singularities in Inexact Computing, May 2002. Working Notes, CERFACS.

[3] F. Chaitin-Chatelin and V. Frayssé. *Lectures on Finite Precision Computations*. SIAM, Philadelphia, 1996.

[4] F. Chaitin-Chatelin, T. Meškauskas, and A. N. Zaoui. Hypercomplex division in the presence of zero divisors on $\mathbb{R}$ and $\mathbb{Z}_2$. Technical Report TR/PA/02/29, CERFACS, Toulouse, France, 2002.

[5] F. Chaitin-Chatelin and E. Traviesas. PRECISE and the reliability of numerical software. in *Handbook of Numerical Computation*, IFIP WG 2.5, SIAM, 2002. To appear.

[6] F. Chaitin-Chatelin and E. Traviesas. Homotopic perturbation - Unfolding the field of singularities of a matrix by a complex parameter: a global geometric approach. Technical Report TR/PA/01/84, CERFACS, 2001.

[7] F. Chatelin. *Spectral approximation of linear operators*. Academic Press, New York, 1983.

[8] F. Chatelin. *Valeurs propres de matrices*. Masson, Paris, 1988.

[9] F. Chatelin. Les ordinateurs calculent faux et c'est tant mieux, Juin 1991. Journées Observatoire de Paris.

[10] F. Chatelin. *Eigenvalues of matrices*. Wiley, Chichester, 1993. Enlarged Translation of the French Publication with Masson.

[11] P. Lévy. L'addition des variables aléatoires définies sur une circonférence. *Bull. Soc. Math. France*, 67:1–41, 1939.

[12] A. Tarski. La complétude de l'algèbre et la géométrie élémentaires. in *Logique, sémantique, métamathématique*, Mémoires écrits entre 1923 et 1944, Armand Colin, Paris, 2:203–242, 1974.