

Fractales de Rauzy (*Ver V. Sirvent, p. 191*)

Boletín de la Asociación Matemática Venezolana

Volumen XI, Número 2, Año 2004

I.S.S.N. 1315-4125

Editor

Argimiro Arratia

Comité Editorial

Oswaldo Araujo Eduardo Lima de Sá

Alejandra Cabaña Gerardo Mendoza Joaquín Ortega

El Boletín de la Asociación Matemática Venezolana se publica dos veces al año en forma impresa y en formato electrónico. Sus objetivos, información para los autores y direcciones postal y electrónica se encuentran en el interior de la contraportada. Desde el Volumen VI, Año 1999, el Boletín aparece reseñado en *Mathematical Reviews*, *MathScinet* y *Zentralblatt für Mathematik*.

Asociación Matemática Venezolana

Presidente

Carlos A. Di Prisco

Capítulos Regionales

CAPITAL

Carlos A. Di Prisco, Matemáticas, IVIC
cdiprisc@ivic.ve

LOS ANDES

Oswaldo Araujo, Matemáticas, ULA
araujo@ciens.ula.ve

ZULIA-FALCON

Fernando Sánchez, Matemáticas, LUZ
fsanchez@luz.ve

CENTRO-OCCIDENTAL

Neptalí Romero
nromero@uicm.ucla.edu.ve

Matemáticas, UCLA

ORIENTE

Jacques Laforgue
laforgue@sucre.udo.edu.ve
Matemáticas, UDO

La Asociación Matemática Venezolana fue legalmente fundada en 1990 como una organización civil cuya finalidad es trabajar por el desarrollo de la matemática en Venezuela. Para más información ver su portal de internet o escribir a su dirección postal.

Asociación Matemática Venezolana

Apartado 47.898, Caracas 1041-A, Venezuela

amv@usb.ve <http://amv.ivic.ve/>

El Boletín de la Asociación Matemática Venezolana está dirigido a un público matemático general que incluye investigadores, profesores y estudiantes de todos los niveles de la enseñanza, además de profesionales de la matemática en cualquier espacio del mundo laboral. Son bienvenidos artículos originales de investigación en cualquier área de la matemática; artículos de revisión sobre alguna especialidad de la matemática, su historia o filosofía, o sobre educación matemática. El idioma oficial es el español, pero también se aceptan contribuciones en inglés, francés o portugués.

Todas las contribuciones serán cuidadosamente arbitradas.

El Boletín publica información sobre los eventos matemáticos más relevantes a nivel nacional e internacional, además de artículos de revisión y crítica de libros de matemática. Se agradece el envío de esta información con suficiente antelación.

Todo el material a ser publicado es revisado cuidadosamente por los editores. Sin embargo, el contenido de toda colaboración firmada es responsabilidad exclusiva del autor.

Cualquier colaboración debe ser enviada al Editor, preferiblemente por correo electrónico (via bol-amv@ma.usb.ve) como archivo postscript, pdf, o un dialecto estándar de TeX. Las colaboraciones en forma impresa deben enviarse por triplicado con figuras y símbolos cuidadosamente dibujados a la Dirección Postal. Para la preparación del manuscrito final recomendamos y agradecemos usar los archivos de estilo LaTeX del Boletín que se encuentran en su página web.

El precio actual de un ejemplar es de Bs. 10.000 (US\$ 10).

The Boletín de la Asociación Matemática Venezolana (Bulletin of the Venezuelan Mathematical Association) is address to a broad mathematical audience that includes researchers, teachers and students at all collegiate levels, and also to any mathematics professional wherever in the labour world. We welcome papers containing original research in any area of mathematics; expository papers on any topic of mathematics, its history or philosophy, or education. The official language is Spanish, but contributions in English, French or Portuguese are also acceptable.

All contributions will be carefully refereed.

The Boletín publishes information on any relevant mathematical event, national or international, and also book reviews. We appreciate receiving this type of information with plenty of time in advance.

All material to be published is carefully revised by the editors. Nonetheless, the content of all signed contributions is of the exclusive responsibility of the author.

All contributions should be sent to the Editor, preferably by email (via bol-amv@ma.usb.ve) in postscript, pdf, or any standard self-contained TeX file. Submissions in printed form should be sent in triplicate with figures and symbols carefully drawn to our Postal Address. For the preparation of the final manuscript we recommend and appreciate the use of the appropriate LaTeX style file of the Boletín, which can be downloaded from its web page.

The current price for one issue is Bs. 10.000 (US\$ 10).

Boletín de la Asociación Matemática Venezolana
Apartado 47.898, Caracas 1041-A, Venezuela
Tel.: +58-212-5041412. Fax: +58-212-5041416
email: bol-amv@ma.usb.ve
URL: <http://boletinamv.ma.usb.ve/>

Impreso en Venezuela por Editorial Texto, C.A.

Telfs.: 632.97.17 – 632.74.86

Boletín de la Asociación Matemática Venezolana
Volumen XI, Número 2, Año 2004

PRESENTACIÓN	131
ARTÍCULOS	
Classical analogues of quantum paradoxes Henryk Gzyl	133
Conservación de propiedades del anillo de polinomios y del anillo de series en la clase de los anillos de conductor finito Nelsy González, Omaid Sepúlveda & Oswaldo Lezama	151
Simetría y nuevas soluciones de la ecuación de Black Scholes Nikolay Sukhomlin	175
Sistemas de numeración, sistemas dinámicos substitutivos y fractales de Rauzy Víctor F. Sirvent	191
ENSAYO	
Determinismo, indeterminismo y la flecha del tiempo en la ciencia contemporánea Daniel A. Morales	213
MATEMÁTICAS RECREATIVAS	
La importancia de cada nodo en una estructura de enlaces: Google-PageRankTM Roberto Markarian & Nelson Möller	233
INFORMACIÓN INTERNACIONAL	
La Esquina Olímpica Rafael Sánchez Lamonedá	253
LIBROS	
<i>Los Crímenes de Oxford</i>, por Guillermo Martínez Reseñado por Carlos Augusto Di Prisco	257
EDITORIAL	
Quince años de la AMV Carlos Augusto Di Prisco	261
AGRADECIMIENTO	263

Asociación Matemática Venezolana
Apartado 47.898, Caracas 1041 – A, Venezuela

**Boletín
de la
Asociación
Matemática
Venezolana**

Vol. XI • No. 2 • Año 2004

Presentación

En este Boletín ofrecemos al lector los artículos de investigación de Henryk Gzyl, en mecánica cuántica; Nelsy González, Omaid Sepúlveda y Oswaldo Lezama, en álgebra; Nikolay Sukhomlin, sobre la ecuación de Black Scholes; Víctor F. Sirvent, sobre sistemas dinámicos y fractales; un ensayo filosófico de Daniel A. Morales sobre el determinismo e indeterminismo en las ciencias naturales y en la matemática; por último, la explicación de la matemática que fundamenta el popular buscador de internet *Google*, por Roberto Markarian y Nelson Möller. Todos artículos de gran calidad, aceptados por nuestro comité editorial luego de riguroso arbitraje por expertos en cada uno de los temas de estas investigaciones.

Las secciones habituales de nuestro Boletín traen información que no pueden dejar de leer: los nuevos logros de nuestros campeones de olimpiadas matemáticas reportado por Rafael Sánchez, y reseña de lo más reciente en la literatura matemática por Carlos Di Prisco. Además, Di Prisco, quien se estrena como nuevo presidente de la Asociación Matemática Venezolana, cierra este Boletín con un breve homenaje a los quince años bien cumplidos de la AMV.

A. A.

Classical analogues of quantum paradoxes

Henryk Gzyl

Abstract

In this note we discuss a few simple classical (as opposed to quantum) prediction problems. The thrust of this work is to examine the predictive role of probability theory and we shall see that some situations are not really paradoxical. In particular we want to separate the role of probability as a predictive tool from one of the basic sources of strangeness in quantum mechanics, namely the principle of superposition of states. For that we shall re-examine some of the quantum paradoxes, recast as a problem about making a prediction about the result of a random experiment in classical physics when some information is known.

1 Preliminaries

There is a vast, academic and non-academic, which does not mean non-serious, popular literature, about the strangeness of quantum phenomena and how this strangeness forces us to re-examine our image or our paradigms about the material world. A very small sample is [A],[C],[d'E], [F],[K],[KN],[M],[T] and the fabulous collection [WZ]. But no item in this collection separates the two issues mentioned in the abstract, namely, that the strangeness in quantum theory does not come from its being a curious mathematical model about phenomena that seem to be intrinsically random, but from the inner structure of the mathematical model itself. In quantum mechanics the strangeness does not come from its being a probabilistic theory, but from a combination of two facts: on one hand the superposition principle (implicit in the fact that states are modeled by vectors over the field of complex numbers), and on the other hand, from the fact that probabilities are obtained from the absolute values of the components of the vector describing the states.

What we are going to do in this note is to examine some problems in which a prediction has to be made about a classical experimental setup that imitates and parallels the setup associated to some of the quantum paradoxes. Thus in spirit we are close to Mermin's [Me]. Our emphasis is on examining in **what does the process of prediction consist of**, and **what is the influence of the available information** on it.

In a lengthy (but necessary for making this note self contained) appendix we recall the very basic ingredients of mathematical probability theory. The readers familiar with that material, may skip it, but those that have never heard of it, must read it. It is basically a collection of standard definitions accompanied by an explanation for their need. The important part of that section is the description of the predictive tool in probability theory, namely the concept of conditional expectation, and the formal use of the concept of σ -algebras as carriers of information. The content of each section is described below, but the message is that a predictor is described by a conditional expectation, which is a random variable whose value depends on information provided by a measurement, and once the measurement is made, the predictor provides us with a prediction.

In section 2 we consider an alternative, simple version of the two slit experiment. Instead of two slits, we have two point light sources. Here the randomness will come from the fact that we do not know which source is on, and we have to predict the observed signal.

In section 3 we shall consider the classical version of Schroedinger's cat and in section 4 the classical version of the EPR paradox (after Einstein, Rosen and Podolsky). Both in the first and last example the role of prior information role will be important for completing the prediction of the outcome. In section 5 we present a proof of the proof of Bell's inequalities, just to emphasize that there is nothing in them having to do with quantum phenomena. See [K] for a glimpse at the vast literature in this area.

We close this section with a few words about **random variables**. The formal definition is given in section 6, but an understanding of the concept is essential for following section (2)-(5). A random variable is the mathematical object (a function, to be specific) designed to model the result of a measurement in an experiment, and an experiment can be thought of as a sequence (finite, infinite, discrete or continuous, of measurements). The values taken by the random variable, are identified with the results of possible measurements of the variable, but the actual mathematical framework chosen does not have anything to do with the actual measurement process. A couple of example will clarify the issue. For example, to describe the results of the toss of a coin we need functions taking values in the set $\{H, T\}$, or to describe life-times, we shall use variables taking values in $[0, \infty)$, regardless of how toss the coin or how we measure the life-times.

But more important, a random variable is a function which stands for a collection of values. This issue is related to the question: In what state is the coin? Or, has the particle decayed or not?. The state is known after the coin or the particle are observed. This is the exact counterpart of the fact that a function is a collection of values taken at points and should not be confused with the values of the function at specific points. This is essential when interpreting

conditional expectations, which are random variables, whose values are the best predictors of some variable when another variable is observed.

2 The two sources experiment

To avoid dealing with waves behind diffracting holes, we shall consider an experimental setup in which we have two sources emitting continuously monochromatic light of frequency ν , located at $\mathbf{x}_{\pm} = (0, 0, \pm h)$. An observer at \mathbf{x} would see a signal $u_{\pm}(\mathbf{x}) = \frac{e^{i\nu\|\mathbf{x}-\mathbf{x}_{\pm}\|}}{2\pi\|\mathbf{x}-\mathbf{x}_{\pm}\|}$ coming from each of them. But regretfully Mr. Prankster is in charge of the lights and he tosses one or two coins to decide which source he will let shine or not. Thus the observer does not know in advance what Mr. Prankster will do but he has to predict what will be observed on the basis of any observation that he may gather.

Assume first that the decision will be made on the basis of a coin toss. In order to model this set up he chooses $\Omega = \{H, T\}$ and then the signal at any arbitrary (but fixed \mathbf{x} , to avoid dealing with function valued random variables) all he knows is that the signal has to be modeled by a random variable $U = u_+(\mathbf{x})I_H + u_-(\mathbf{x})I_T$, and if does not have any other information, then the best prediction he can make is that on the average he will observe $E[U] = p_H u_+(\mathbf{x}) + p_T u_-(\mathbf{x})$.

This is a good point to insist on the difference between a prediction like $E[U]$ and an observed value: When one observes a random variable, U in this case, each time one of its possible values is seen, even if the initial set up is the same. Since these differ from observation to observation, one performs a statistical analysis on the data, and an expected (value to be seen upon average) is provided. But assume that he is given information consisting of the specification of which bulb is on. This is modeled by a σ -algebra $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$, i.e., the full information about the setup. He knows that his best predictor must be computed as $E[U | \mathcal{F}] = U$, and when he is told that the result of the toss is an H , he then predicts that he will observe $u_+(\mathbf{x})$.

Assume now that the observer knows that Mr. Prankster is using two coins. Then to model the outcome of a single toss he considers $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$, and all the information about the experiment is contained in the σ -algebra $\mathcal{P}(\Omega)$. Assume that he knows the probabilities $\{p_1, p_2, p_3, p_4\}$ as listed above. The signal to be observed is modeled by the random variable (random field)

$$U = (u_+(\mathbf{x}) + u_-(\mathbf{x}))I_{\{(H,H)\}} + u_+(\mathbf{x})I_{\{(H,T)\}} + u_-(\mathbf{x})I_{\{(T,H)\}}.$$

Note again that the vales of U , that is of the observed field, depend on the result of the throw of the two coins. Now in the absence of any information, the best prediction that the observer can make is given by $E[U] = p_1(u_+(\mathbf{x}) + u_-(\mathbf{x})) +$

$p_2 u_+(\mathbf{x}) + p_3 u_-(\mathbf{x}) = (p_1 + p_2) u_+(\mathbf{x}) + (p_1 + p_3) u_-(\mathbf{x})$. Clearly $p_1 + p_2$ and $p_1 + p_3$ are, respectively the probabilities that the sources at \mathbf{x}_\pm are on. These were denoted by p_H and p_T a few lines above.

Assume now that all the observer is able to find out is that either the two sources may be in the same state. That is, assume that both of the sources are on or off, or that either of them may be on and the other off. The given information corresponds to the partition of $\{\{(H, H), (T, T)\}, \{(H, T), (T, H)\}\}$ of Ω . Denote by \mathcal{G} the σ -algebra that it generates. The best predictor given this information is the random variable $E[U | \mathcal{G}]$ taking values

$$E[U | \{(H, H), (T, T)\}] = \frac{p_1(u_+(\mathbf{x}) + u_-(\mathbf{x}))}{p_1 + p_4} = \frac{p_1}{p_1 + p_4}(u_+(\mathbf{x}) + u_-(\mathbf{x}))$$

$$E[U | \{(H, T), (T, H)\}] = \frac{p_2 u_+(\mathbf{x}) + p_3 u_-(\mathbf{x})}{p_2 + p_3} = \frac{p_2}{p_2 + p_3} u_+(\mathbf{x}) + \frac{p_3}{p_2 + p_3} u_-(\mathbf{x}).$$

respectively whenever the event $\{(H, H), (T, T)\}$ or $\{(H, T), (T, H)\}$ occurs. The probabilities of occurrence of each of the events are, respectively $p_1 + p_4$ and $p_2 + p_3$. To make the role of the prior information more apparent, assume that the observer knows that one of the events in the following partition of Ω occurs: $\{\{(H, H), (T, T)\}, \{(H, T)\}, \{(T, H)\}\}$. The difference with the previous case is that when the observer finds out that one of the sources is uncovered, he knows that the other is covered. If we denote by \mathcal{G}_1 the σ algebra generated by that partition, then the best predictor of the random field U is the random variable (random field)

$$E[U | \mathcal{G}_1] = u_+(\mathbf{x}) I_{\{(H, T)\}} + u_-(\mathbf{x}) I_{\{(T, H)\}} + \frac{p_1}{p_1 + p_4} (u_+(\mathbf{x}) + u_-(\mathbf{x})) I_{\{(H, H), (T, T)\}}.$$

Observe for example, that if the toss of the coins were (T, T) , then we would see no signal, but if it were (H, T) , we would see $u_+(\mathbf{x})$. And equally important, that the three possible results would occur with probabilities p_2 , p_3 and $p_1 + p_4$. The difference between this and the first case treated, is that now both sources may be on or off, whereas in the first case one was on while the other was off.

3 Is the bird in the cage or not?

Let us now examine the classical variant of the Schroedinger cat paradox. Our setup consists of a light bulb with an exponential lifetime, and a bird in a cage. The setup is such that if the light bulb burns out, the latch on the bird's cage is released and the bird flies away. The issue is whether, upon observation, will we find the bird in the cage or not? To describe the result of one measurement of a lifetime, the following sample space is adequate: $\Omega = [0, \infty)$, and all questions we can ask about the measurement are represented by $\mathcal{F} = \mathcal{B}([0, \infty))$. As

probability measure we choose $P(dt) = \frac{1}{\tau}e^{-t/\tau}$. The random variable of interest is the lifetime of the bulb, described by $T : \Omega \rightarrow [0, \infty)$ such that $T(u) = u$, that is we shall find convenient to think of the points $u \in \Omega$ as “life-histories” of the light bulb. The distribution of this variable is $P(\{T > t\}) = e^{-t/\tau}$, which is to be read as: the fraction of life-histories longer than t is $e^{-t/\tau}$.

Let the position of the latch be denoted by \top if it is open and by \perp if it is closed. The random state of the latch at time t is described by a random variable $X_t : \Omega \rightarrow \{\top, \perp\}$ given by

$$X_t(u) = \top \text{ whenever } T(u) \leq t; \text{ or } X_t(u) = \perp \text{ whenever } T(u) > t.$$

What happens here? Well, the observer decides when he is going to look at the bulb, that is the observer chooses t , whereas nature chooses the life history u of the bulb. If $u = T(u) \leq t$ then the bulb is still on at the time of observation, whereas if $u = T(u) > t$, the bulb has burnt out and the bird has flown away. Is there anything paradoxical here? It does not seem so.

If the caretaker of the bird is to receive a payment d if nothing has happened, and he has to pay a penalty δ if the bird escapes, what is our best prediction of his fortune at time t ? Clearly if no information is available, his fortune at time t is $d(1 - e^{-t/\tau}) - \delta e^{-t/\tau}$. We leave it up to the reader to see why, and direct her/him to Peres’ [P] for a discussion from the point of view of a physicist.

4 Split coins and recoiling particles

In this section we shall consider non quantum variants of the EPR paradox. Again the aim is to emphasize the role of prior knowledge. Suppose you have a coin-like object with two faces which can be either R(ed) or B(lue), but these properties for all you know, have been assigned at random. The assignment is by means of a random variable taking values in the set $\{(R, R), (R, B), (B, B)\}$. When performing an experiment consisting of one measurement, it suffices to take $\Omega = \{\{R, R\}, \{R, B\}, \{B, B\}\}$. Clearly we think of the color assignment as a multi set: sets in which an element may appear repeated a number of times. This description takes care of the intuitive symmetry of a coin respect to the labeling of the faces. An observation of a coloring of a coin is actually a pair of observations, which we list sequentially. Formally speaking $X : \Omega \rightarrow \{R, B\} \times \{R, B\}$ and an equivalence relation has been defined on the range, which identifies (R, B) and (B, R) . The values of X are obviously $X(\{B, B\}) = (X_1(\{B, B\}), X_2(\{B, B\})) = (B, B)$, etc. This may seem like overdoing it, but is really necessary if we want our model to reflect the symmetry of the coin with respect to “flips” (We may also think of colorings as mappings $\{1, 2\} \rightarrow \{B, R\}$ which are invariant under the “flip” $F : \{1, 2\} \rightarrow \{1, 2\}$, $F(1) = 2$, and $F(2) = 1$).

If we assign probabilities to the elementary events by Ω by $P(\{B, B\}) = p_1$, $P(\{R, B\}) = p_2$ and $P(\{R, R\}) = p_3$, then clearly $P(\{X_1 = B\}) = p_1 + \frac{p_2}{2}$ and $P(\{X_1 = R\}) = p_3 + \frac{p_2}{2}$.

Suppose now that Mr. Prankster splits the coin in two (each half comprising a face), and a half given to some team mate. You have to guess the color of that face. You will receive either of the following pieces of information about the half in Mr. Prankster's hand: (i) the color of that half, (ii) the color of that half and the coloring of the coin, or (iii) only the coloring of the coin.

Case 1 Assume you are told that $X_2 = R$. In this case all you can predict is that either $X_1 = B$ or that $X_1 = R$ with the two probabilities given above.

Case 2 Assume that you are told that $X_2 = B$ and that the coloring is $\{B, R\}$. Clearly in this case you predict that $P(X_1 = R | X = \{B, R\}, X_2 = B) = 1$. You do not have to invoke super luminal propagation of information or anything. Just apply basic probability (which in this case could not be more trivial).

Case 3 Here there are several possibilities depending on the explicit form of the information. For example, if told that both faces are equal but nothing else, you will compute basic conditional probabilities and conclude that the coloring is B with probability $\frac{p_1}{p_1+p_3}$ or R with the probability $\frac{p_3}{p_1+p_3}$. If you are told that the coloring is mixed, you will assert that one face is R and the other is B with probability 1.

A variation on this theme, corresponding to a simple version of the problem of the recoiling particles, is the following: Assume that the particles are marked (distinguishable) and an integer $1 \leq X_i \leq N, i = 1, 2$ is assigned to each. Assume that the numbers are chosen at random but not necessarily independently of each other. For example, the numbers may measure "quanta" of energy transferred to each particle during the splitting process. Assume that you may observe X_1 and you want to predict X_2 .

The underlying sample space for this situation is $\Omega = \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$, and the information about the measurement is described by a probability law \mathcal{P} on Ω given by $P(X_1 = i, X_2 = j) = p_{ij}$, and each observable is modeled by a coordinate map, for example $X_1 : \Omega \rightarrow \{1, \dots, N\}$, such that $X_1(a, b) = a$. When the X_i are independent $P(X_1 = i, X_2 = j) = P(X_1 = i)P(X_2 = j) = p_i p_j$, but this is not necessary for what comes below.

Any prediction about X_2 will depend about the information we are given about X_1 and about the pair. Let us consider three cases (i) we only know X_1 , (ii) we are given X_1 and we now that $a \leq X_1 + X_2 \leq b$ is constant, but we do not know which constant, and (iii) is as the second case, but the value of the constant $X_1 + X_2 = k$ is specified.

Case 1 We know from Example 1 in the appendix that $E[X_2 = j | X_1] = \frac{p_{X_1 j}}{p_{X_1}}$. That is the best predictor of the probability of the event $\{X_2 = j\}$ depends on the observation of X_1 . When we learn that $X_1 = i$, at that moment we know that we predict $X_2 = j$ with probability $\frac{p_{ij}}{p_i}$. But before the observation of X_1 , the best predictor is a random variable.

Case 2 This time the generic information is the trace of $\sigma(X_1)$ on the event $\{a \leq X_1 + X_2 \leq b\}$, that is the σ -algebra generated by the collection $\{a - i \leq X_2 \leq b - i; 1 \leq i \leq N\}$. Let us denote this by the obvious $\sigma(X_1) \cap \{a \leq X_1 + X_2 \leq b\}$. In this case the best predictor of X_2 (not of any particular value of the variable) is

$$E[X_2 | \sigma(X_1) \cap \{a \leq X_1 + X_2 \leq b\}]$$

and we leave it up to the interested reader to compute the possible values of this random variable according to Example 1 in the appendix.

Case 3 Let us examine the most standard example corresponding to the classical version of the EPR paradox. Assume that it is known that $X_1 + X_2 = k$. If nothing is known about X_1 , the best predictor of X_2 is

$$E[X_2 | X_1 + X_2 = k] = \sum_{i=1}^N \frac{(k-i)^+ p_i}{p_i} I_{\{X_1=i\}}$$

where the explanation of the term $(k-i)^+$ is clear: the value $k-i$ is an allowed value for X_2 while it is bigger or equal to 0.

Notice that when we are given the event $\{X_1 = i\} \cup \{X_1 + X_2 = k\}$ (which is not empty, that is it describes a possible event only when $k \geq i$, then

$$E[X_2 | \{X_1 = i\} \cup \{X_1 + X_2 = k\}] = k - i!$$

This is clear because $\{X_1 = i\} \cup \{X_1 + X_2 = k\} = \{X_2 = k - i\}$, that is specifying the conserved quantity and one of the variables and the value of one of them, automatically specifies the other, regardless whether we measure it or not. In this case knowledge does not have to do with transmission of information, it has to do with logic.

Let us now consider these three cases under the assumption that X_1 and X_2 are continuous, say, real valued random variables, having a joint distribution function $\rho(x_1, x_2)$. Let us now re-examine some possibilities.

Case 4 Assume that all knowledge that may be available to us consists of the result of a measurement of X_1 . What is our best predictor of X_2 ?. We already

know that it is $E[X_2 | X_1]$, and in example 2 of the appendix, we show that this is the random variable $\int x_2 \frac{\rho(x_2, X_1)}{\rho(X_1)} dx_2$, whose value becomes known once X_1 is measured. When $X_1 = x_1$ is observed, our best predictor of X_2 becomes

$$E[X_2 | X_1 = x_1] = \int x_2 \frac{\rho(x_2, x_1)}{\rho(x_1)} dx_2.$$

Case 5 Let us consider now the case where only $Z \equiv X_1 + X_2$ is available for observation. What is our best predictor of X_2 in this case? This is similar to the case considered above, except that now we must first find the joint distribution of (Z, X_2) . It is a simple computation to verify that it is $\hat{\rho}(z, x_2) = \rho(z - x_2, x_2)$ and that $\hat{\rho}(z) = \int \rho(z - x_2, x_2) dx_2$. Now the computation of $E[X_2 | Z]$ is a particular case of *Case 4*.

Case 6 Assume to finish that we may measure both Z and X_1 . In this case note that for bounded continuous functions h, f, g , a simple change of variables should convince anyone that

$$\begin{aligned} E[h(X_2)f(X_1 + X_2)g(X_1)] &= E[h(Z - X_1)f(Z)g(X_1)] \\ &= E[E[h(X_2) | Z, X_1]f(Z)g(X_1)] \end{aligned}$$

from which it is clear (see definition (6.4) that $E[h(X_2) | Z, X_1] = h(Z - X_1)$). This is clearly what is expected, and the predictive formalism reflects it.

5 Bell inequalities

There are some inequalities, called Bell inequalities, satisfied by any set of three random variables taking values in $[-1, 1]$. And it said that their violation is a manifestation of the fact that the system under study is quantum and not classical, or that it “obeys a logic” different that the logic of classical probability. Actually, the following is an exercise of chapter 1 of [B]:

Let X, Y, Z , be three random variables taking values in $[-1, 1]$. Show that $1 - E[XY] \geq |E[XZ] - E[YZ]|$.

One possible proof of the inequality runs as follows: Consider

$$(1 + X)(1 - Y)(1 + Z) = 1 + X - Y - Z - XY + XZ - YZ$$

$$(1 - X)(1 + Y)(1 - Z) = 1 - X + Y + Z - XY + XZ - YZ$$

adding these two and noticing that the left hand side is always positive, rearranging and taking expected values, we obtain

$$E[YZ] - E[XZ] \leq 1 - E[XY].$$

To obtain the other half of the inequality, consider

$$(1 - X)(1 + Y)(Z - 1) = -1 + X - Y + Z + XY - XZ + YZ$$

$$(1 + X)(Y - 1)(1 + Z) = -1 - X + Y - Z + XY - XZ + YZ$$

adding and noticing that the left hand side is always negative, we obtain after rearranging and taking expected values, that

$$1 - E[XY] \leq E[XZ] - E[YZ]$$

which finishes the proof. And for fun consider the particular case: Let ξ, η, ζ be random variables taking (any) two values each, such that $P(\xi = a) = 1/2$, $P(\eta = b) = 1/2$ and $P(\zeta = c) = 1/2$, and consider that random variables $X = I\{\xi = a\}$; $Y = I\{\eta = b\}$ and $Z = I\{\zeta = c\}$. A direct application of the inequalities plus multiplication by 2, lead us to the inequality

$$2 - P(\xi = a | \eta = b) \geq |P(\eta = b | \zeta = c) - P(\xi = a | \zeta = c)|.$$

Comment 5.1 *Note that in the assumption about X, Y and Z , nothing is said about whether these variables represent measurements at near or far away locations, nor about the times the measurements are to be taken, or about possible physical or other kind of interpretation of the random variables. The only assumption made is about their possible values, nothing is even assumed about their joint distribution. If the inequalities are violated, it must be that the underlying mathematical structure of the probabilistic model under examination does not coincide with the classical probabilistic model, not due to some non-explicit assumption about causality or whatever.*

Here is an example of a quantum system that does seem to satisfy the Bell inequalities: Consider three identical, distinguishable and spinless particles moving on the real line, under the action of a three particle force given by the potential $V(x_1, x_2, x_3) = 0$ or $= +\infty$ depending on whether $(x_1^2 + x_2^2 + x_3^2)$ is respectively ≤ 1 or ≥ 1 . Mathematically speaking, this system is equivalent to a particle confined to move freely inside a sphere of radius 1. Can there exist a state $\Psi(x_1, x_2, x_3)$ such that the probability density $|\Psi(x_1, x_2, x_3)|^2$ violates Bell inequalities? Offhand, it seems that the answer is: “No, there is not.” But life is full of surprises.

To finish, we direct the reader to the exposé [B] for a nice presentation of the “geometric” ideas behind the inequalities as well as further references.

6 Appendix: The very basics of probabilistic modeling

6.1 Ensembles a.k.a Sample Spaces

The basic construct in probability theory is that of a sample space. This corresponds to what is called (but never explicitly defined) an ensemble in most

books in statistical thermodynamics. The construct is specified by a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where the set Ω should be thought of as the “experiments” (i.e., sequences of measurements), or sometimes it may be convenient to *think* of it as “the set of states of nature”. The set \mathcal{F} is the mathematical construct describing the information about our system (i.e., it contains the objects that describe all questions **we assume valid** to ask about our system). Its elements are called events and formally it is a σ -algebra of subsets of Ω , that is a collection of sets satisfying:

- i) $\emptyset \in \mathcal{F}$
- ii) $A \in \mathcal{F} \Rightarrow A^c \equiv \Omega - A \in \mathcal{F}$
- iii) If A_k for $k = 1, 2, \dots$ is a countable collection of elements of \mathcal{F} , then $\bigcup_k A_k \in \mathcal{F}$.

Observe that the arithmetical structure of the set operations parallel that of the logical connectives in ordinary speech. That is why \mathcal{F} is to be thought of as the class of allowed questions about the collection of experiments described by Ω . And to finish with the list, we have

Definition 6.1 We say that \mathbb{P} is a probability measure or law on the sample space (Ω, \mathcal{F}) , that is a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfying the conditions

- 1) $\mathbb{P}(\Omega) = 1$
- 2) If A_k for $k = 1, 2, \dots$ is a countable collection of **disjoint** elements of \mathcal{F} , then $\mathbb{P}(\bigcup_k A_k) \equiv \sum_k \mathbb{P}(A_k)$.

Comment 6.1 A pair (S, \mathcal{S}) where S is any set and \mathcal{S} is a σ -algebra of subsets of S is called a **measurable space**, because an additive function (called measure or volume) can be consistently defined on it. The difference between probability and arbitrary measures is that the later do not have to be normalized to 1.

6.2 Observables a.k.a Random Variables

And important notion in both classical and quantum mechanics is the concept of random variable, it is central to mathematical probability theory, just because it is the mathematical object embodying the notion of measurement, **not the actual process or act of measurement**, but the object describing the result of the measurement. The technical definition is the following:

Definition 6.2 We assume we have already chosen a sample space (Ω, \mathcal{F}) and we also have a measurable space (S, \mathcal{S}) . An S -valued random variable is a function $X : \Omega \rightarrow S$ such that for any $A \in \mathcal{S}$ we have $\{X \in A\} \equiv X^{-1}(A) \in \mathcal{F}$.

Comment 6.2 We can interpret $X(\omega)$ as the value of the observable X in the experiment ω . It is here where the mathematical modeling of the randomness is

located. Notice that different values of the observable are achieved in different experiments. Thus, the interpretation of the ω and of $X(\omega)$ are tailored to complement each other. Also, if the sets in \mathcal{S} describe questions about the results of measurements, then the measurability condition just transfers questions about measurements onto questions about experiments taking given values. When $S = \mathbb{R}$ and $\mathcal{S} = \mathcal{B}(\mathbb{R})$, we just say that X is a random variable. For any topological space S , the class $\mathcal{B}(S)$ denotes the smallest σ -algebra containing the open sets of S and is called the Borel σ -algebra of S .

Also, now and then we shall make use of the notation I_A to denote the indicator function of the set A , defined by $I_A(x) = 1$ or 0 depending on whether $x \in A$ or not.

The important example for us corresponds to the case in which \mathcal{F} is generated by a partition $\pi = \{\Lambda_1, \dots, \Lambda_K\}$. In this case any random variable is described by

$$X = \sum_{j=1}^K x_j I_{\Lambda_j}$$

that is, X takes value x_j whenever the event Λ_j takes place.

Three simple but important **examples** of σ -algebras are worth recording.

- (i) First, consider a partition $\Pi = \{A_i, \dots, A_n, \dots\}$, that is, a finite or infinite disjoint collection of subsets of Ω whose union is Ω . The σ -algebra $\mathcal{F} = \sigma(\Pi)$ that the partition generates is the collection of all possible union of sets of Π . Any \mathcal{F} -measurable function (random variable) is of the form $X = \sum_n x_n I_{A_n}$.
- (ii) Conversely, any random variable X taking only a countable collection of values $\{x_1, x_2, \dots\}$ generates a σ -algebra, obviously denoted by $\sigma(X)$ generated by the partition $\{\{X = x_n\} \mid n = 1, 2, \dots\}$.
- (iii) To finish the list, there is the trivial σ -algebra $\mathcal{F}_0 \equiv \{\emptyset, \Omega\}$. The reader should verify that all \mathcal{F}_0 -measurable random variables are constant.

Actually, usually the construction of sample spaces and of the basic random variables is carried out simultaneously. A typical example of sample space is the following: Consider a finite set $R = \{r_1, \dots, r_k\}$ and form $\Omega \equiv R^N$ where N is either finite or ∞ . In any case think as follows $\Omega = \{\omega : \{1, 2, \dots, N\} \rightarrow R\}$ is the collection of sequences of length N (finite or infinite). Each sequence describes an experiment and to describe the measurements it is convenient to introduce the following R -valued random variables: $X_i : \Omega \rightarrow R$ defined by $X_i(\omega) \equiv \omega(i)$. Mathematically speaking this is just a coordinate map, but it is the mathematical object that models the result of the i -th measurement in the experiment Ω . The information about this setup is provided but the results of all possible experiments. This corresponds to the σ -algebra generated by the collection (of "cylinders") $\{X_{i_1} \in B_1, \dots, X_{i_m} \in B_m : 1 \leq i_1 \leq i_2 \leq \dots \leq i_m; B_1, \dots, B_m \subseteq R\}$

To construct a probability on (Ω, \mathcal{F}) there are many ways to proceed. If we want the X_i to be **statistically independent** we start from any assignment $\mathbb{Q}(B) = \sum_{q_i \in B} q_i$ where $\sum_{q_i \in R} q_i = 1$ and define $\mathbb{P}(\{X_{i_1} \in B_1, \dots, X_{i_m} \in B_m\}) = \prod \mathbb{Q}(B_i)$. That is we start from a probability \mathbb{Q} on R which we may have to find by doing statistical analysis, and we end up with a probability law on the class of all possible questions about the experiments.

But more interesting probability laws are conceivable. For example, when $N < \infty$, we may define a measure on Ω by putting

$$\mathbb{P}(\omega) \equiv \frac{e^{\beta \sum E(i,j)X_i(\omega), X_j(\omega)}}{Z(\beta)}$$

where $Z(\beta)$ is a normalization factor, which should remind us of the partition function. Anyway, this is as simple as you can get and say something interesting and familiar. But the list examples in the applied literature, ranging from signal processing to mathematical finance is enormous.

6.3 A simple ensemble

A standard construction in probability theory, which can obviously identified with the physicist's notion of ensemble (which by the way is never made explicit in any book on statistical physics), is useful for describing sequences of independent measurements. Denote by S the set in which the measurements take value. Let us assume that it is either a discrete set or a subset of some R^d . In the first case we consider $\mathcal{S} = \mathcal{P}(S)$ to be the σ -algebra on S . In the second case $\mathcal{S} = \mathcal{B}(S)$. We set $\Omega \equiv \prod_{j=1}^N S$, and we describe all the questions about the experiments by $\mathcal{F} \otimes_{j=1}^N \mathcal{S}$. We understand that N is either finite or $N = \infty$ depending on whether we need to describe finite or infinite sequences of measurements. The results of the measurements are described by the coordinate maps $X_n : \Omega \rightarrow S$ defined by $X_n(\omega) = s_n$ if $\omega = \{s_1, s_2, \dots\}$

Thus, note that in this set up, $\omega = \{s_1, s_2, \dots\}$ clearly denotes an experiment, that is, a sequence of measurements yielding values s_1, s_2, \dots

A probability assignment which makes the X_n statistically independent is the following: Let $\mu : \mathcal{S} \rightarrow [0, 1]$ be a given probability on (S, \mathcal{S}) , which we want to describe the distributions of particular measurements, and we put $P(\{X_{n_1} \in A_1, \dots, X_{n_k} \in A_k\}) \equiv \mu(A_1) \dots \mu(A_k)$, for any $n_1 < \dots < n_k$ and any A_1, \dots, A_k in \mathcal{S} . In particular $P(\{X_n \in A\}) = \mu(A)$, and thus the X_n are **by construction** independent and identically distributed.

For example in order to describe an experiment consisting on the measurement of the life-time of a light bulb, or the life-time of a decaying particle, or the penetration of a particle in an absorbing medium, we take $N = 1$, $S = [0, \infty)$, $\mathcal{S} = \mathcal{B}(S)$ and $\mu(A) = \frac{1}{\tau} \int_A e^{-t/\tau} dt$.

6.4 Basic predictions

In order to do predictions we need to introduce two notions, that of mathematical expectation and that of conditional expectation. Logically speaking, the former is a particular case of the later, but it is easier to begin with the simpler notion. The definition or construction of the mathematical expectation is done stepwise. We shall define a **simple function** or **simple random variable** by $X \equiv \sum x_i I_{A_i}$ where $\{A_1, A_2, \dots, A_m\}$ denotes a finite partition of Ω .

Now we **define** the expected value (or the mathematical expectation) of a simple random variable by

$$E[X] = \sum x_i \mathbb{P}(A_i)$$

This is the intuitive thing to do, and so it is to extend this to any variable: one can prove that any **positive** random variable X is an increasing limit of simple functions X_n , then we define $E[X] \equiv \lim_n E[X_n]$, and to finish one notices that any random variable can be written as $X = X^+ - X^-$, and when the expected of each of these is finite we say that X is integrable and write $E[X] = E[X^+] - E[X^-]$. Actually to express that summarily we write $E[|X|] < \infty$.

We are finally ready to introduce the most basic probabilistic notion: that of conditional expectation.

Definition 6.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space, and \mathcal{G} be a sub- σ -algebra of \mathcal{F} . Let Y be an integrable, \mathcal{F} -measurable, real valued random variable. The **conditional expectation of Y given \mathcal{G}** is a \mathcal{G} -measurable random variable denoted by $E[Y | \mathcal{G}]$ satisfying the following condition*

$$E[YH] = E[E[Y | \mathcal{G}]H] \text{ for any } \mathcal{G} - \text{measurable, bounded } H. \quad (1)$$

Let us list some of its properties and then show how it is computed starting from (1).

Theorem 6.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . Then the following hold:*

- (i) $E[1 | \mathcal{G}] = 1$.
- (ii) If $X_1 \geq X_2$ are random variables, then $E[X_1 | \mathcal{G}] \geq E[X_2 | \mathcal{G}]$.
- (iii) For $a, b \in \mathbb{R}$ $E[aX_1 + bX_2 | \mathcal{G}] = aE[X_1 | \mathcal{G}] + bE[X_2 | \mathcal{G}]$.
- (iv) (Filtering or tower property) If $\mathcal{H} \subset \mathcal{G}$ is another sub- σ -algebra, then for any integrable X ; $E[E[X | \mathcal{G}] | \mathcal{H}] = E[X | \mathcal{H}]$.
- (v) If Z is any bounded, \mathcal{G} -measurable random variable, then $E[ZX | \mathcal{G}] = ZE[X | \mathcal{G}]$.
- (vi) $E[X | \mathcal{F}_0] = E[X]$

Even though we do not state it explicitly, the mapping $X \rightarrow E[X | \mathcal{G}]$ behaves in all respect as an expected value, except that it is a random variable. When

$\mathcal{G} = \sigma(Y)$, where Y is a given random variable, we shall write $E[X | Y]$ instead of $E[X | \sigma(Y)]$. The result that explains why conditional expectations are (best) predictors is contained in the following

Theorem 6.2 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X be an integrable random variable and let Y be another random variable. Then the function $g(Y)$ that minimizes $E[(X - g(Y))^2]$ is $g(Y) = E[X | Y]$*

Comment 6.3 *Implicit in the statement of the theorem is an intuitive result that we need to complete the proof.*

Let us state a simple version of that result:

Theorem 6.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let Y_1, \dots, Y_N be a collection of random variables. If Z is a random variable, measurable with respect to $\sigma(Y_1, \dots, Y_N)$, then there exists a Borel measurable function $h : \mathbb{R}^N \rightarrow \mathbb{R}$ such that $Z = h(Y_1, \dots, Y_N)$.*

Proof of theorem (6.2). It is basically a computation. Consider

$$E[(X - g(Y))^2] = E[((X - E[X | Y]) + (E[X | Y] - g(Y)))^2] =$$

$$E[(X - E[X | Y])^2] + E[(E[X | Y] - g(Y))^2] + 2E[(X - E[X | Y])(E[X | Y] - g(Y))]$$

We shall now verify that the last term vanishes. Therefore the assertion is clear, since we can choose $G(Y) = E[X | Y]$. Observe that from the definition it follows that by (1) (i.e., the defining property of the conditional expectation)

$$E[(X - E[X | Y])(E[X | Y] - g(Y))] = E[E[(X - E[X | Y]) | Y](E[X | Y] - g(Y))]$$

and now by properties (i) and (v) in theorem (6.1), with X there replaced by 1 and Z by $E[X | Y]$, it follows that the conditional expectation under the integral sign vanishes.

To illustrate how the definition is applied, let us now work out two standard examples:

Example 1 Consider the a σ -algebra $\sigma(\pi)$, generated by the partition $\pi = \{A_1, \dots, A_n, \dots\}$ of Ω . Assume that $\mathbb{P}(A_j) > 0 \forall j \geq 1$. Then since $E[X | \sigma(\pi)]$ is $\sigma(\pi)$ -measurable, it can be written as $E[X | \sigma(\pi)] = \sum_j c_j I_{A_j}$. Clearly, to compute the c_j it suffices to apply the defining identity (1) to the binary random variables I_{A_k} . Thus

$$E[E[X | \sigma(\pi)]I_{A_k}] = E[XI_{A_k}] = E[(\sum_j c_j I_{A_j})I_{A_k}] = c_k \mathbb{P}(A_k),$$

and therefore

$$E[X | \sigma(\pi)] = \sum_j \frac{E[XI_{A_j}]}{\mathbb{P}(A_j)} I_{A_j}. \tag{2}$$

Example 2 Consider now two random variables X and Y , about which you know the joint distribution $\rho(x, y)$, i.e., for any bounded function $F(X, Y)$, $E[F(X, Y)] = \int F(x, y)\rho(x, y)dxdy$. Assume that X is integrable, which presently means that $E[|X|] = \int |x|\rho(x, y)dxdy < \infty$. To compute $E[H(X) | Y]$ for any bounded $H(X)$, note that for any bounded $g(Y)$,

$$E[H(X)g(Y)] = \int H(x)g(y)\rho(x, y)dxdy = \int g(y)\left(\int h(x)\frac{\rho(x, y)}{\hat{\rho}(y)}dx\right)\hat{\rho}(y)dy$$

where we set $\hat{\rho}(y) = \int \rho(x, y)dx$. Also, temporarily put $E[H(X) | Y] = h(y)$, then

$$E[E[H(X) | Y]g(Y)] = E[h(Y)g(Y)] = \int h(y)g(y)\hat{\rho}(y)dy.$$

Since this identity is true for any $g(y)$, then it must be true that

$$h(Y) = E[H(X) | Y] = \int h(x)\frac{\rho(x, Y)}{\hat{\rho}(Y)}dx.$$

Comment 6.4 *what is important is that we are displaying the conditional expectation as a random variable. That is, the best predictor of $H(X)$ given that Y is measured is a random variable, whose value is known once Y is observed, and not before.*

6.5 The notion of independence

The proper definition of independence is at the level of σ -algebras. We have, within the usual model (Ω, \mathcal{F}, P)

Definition 6.4 (a) *A collection $\{\mathcal{G}_i | i \in \mathcal{I}\}$ (where \mathcal{I} is any set of indices) of sub- σ -algebras of \mathcal{F} , is said to be independent whenever for any finite subset $J \subset \mathcal{I}$, and for any collection $\{A_i | i \in J\}$ we have*

$$P(\cap_{i \in J} A_i) = \prod_{\{i \in J\}} P(A_i).$$

(b) *A family $\{X_i | i \in \mathcal{I}\}$ is independent whenever $\sigma(X_i)$ are independent.*

Independence is a family property, and is a tricky property to verify. And it is sometimes confused with functional independence. Let us examine this source of confusion in a simple case. Consider $\Omega = \mathbb{R}^2$ and $\mathcal{F} = \mathcal{B}(\mathbb{R}^2)$. Consider the

two random variables $X, Y : \Omega \rightarrow \mathcal{R}$ $X(x, y) = x$, $Y(x, y) = y$. These are functionally independent at least in two senses. First they are linearly independent as vectors in the class of functions defined over Ω , and second they are functionally independent, that is $Y \notin \sigma(X)$ (nor viceversa), that is, Y cannot be written as function of X . But they may not be independent as random variables, as the following two examples show:

Example 3 Consider the uniform distribution on the unit disk on Ω , that is $P(dx, dy) = \frac{1}{\pi}$ whenever $(x, y) \in \{(\xi, \eta) \mid (\xi)^2 + (\eta)^2 \leq 1\}$, and equal to zero otherwise. We leave it up to reader to find two sets A, B in the disk and verify that $P(\{X \in A, Y \in B\}) \neq P(\{X \in A\})P(\{Y \in B\})$.

Example 4 Within the same setup of the previous example, Assume that X and Y have a joint Gaussian distribution with correlation $\rho \neq 0$. They again, they are not independent, no matter how functionally independent they are.

Let us now examine another source of confusion. Consider the following model for the coin tossing experiment. $\Omega = \{0, 1\}^{\mathbb{N}} = \{\omega : \mathbb{N} \rightarrow \{0, 1\}\}$. Now experiments are specified by the “observation” of the random variables $X_n : \Omega \rightarrow \{0, 1\}$ by $X_n(\omega) = \omega(n)$. Observe that specifying an experiment $\omega \in \Omega$ amounts to prescribing the result of the measurement of all X_n . The information available to the observer is modeled by the σ -algebra generated by the cylinder sets $\{\cap_{i \in J} \{X_i = \alpha_i\}; \alpha_i \in \{0, 1\}; J \text{ finite}\}$. One extends P from its values on cylinders, which to make life easy we take as $P(\{\cap_{i \in J} \{X_i = \alpha_i\}) = (\frac{1}{2})^{|J|}$, where we denote the cardinality of J by $|J|$. Let $\Theta : \Omega \rightarrow \Omega$ be the shift operator defined by $X_n \circ \Theta = X_{n+1}$, that is, the shift of a sequence by one unit to the left, and let Θ^n describe the n -th iterate of Θ . It is rather easy to see that the mappings Θ^n are not independent, whereas the X_n are independent by construction. Consider for example the events $\Lambda_1 = \{X_2 = 0\}$ and $\Lambda_2 = \{X_1 = 1, X_2 = 1\}$. Consider for example: $\{\Theta \in \Lambda_1\} \cap \{\Theta^2 \in \Lambda_2\} = \emptyset$, but $P(\{\Theta \in \Lambda_1\}) = 1/2$ and $P(\{\Theta^2 \in \Lambda_2\}) = 1/4$.

Notice that the flow Θ^n is deterministic, that is, given the initial point ω of the orbit, the values $\Theta^n(\omega)$ are completely determined. But to give the initial point an infinite sequence of independent random variables has to be observed.

References

- [A] Aczel, A. *“Entanglement”* Plume-Penguin, New York, 2003.
- [B] Borkar, V.A. *“Probability Theory”* Springer-Verlag, Berlin, 1995.
- [C] Cromer, A. *“Uncommon Sense”* Oxford Univ. Press, Oxford, 1993.
- [D] Deutsch, D. *“The fabric of reality”* Penguin Books, New York, 1997.
- [d’E] d’Espagnat, B. *“Reality and the Physicist”* Cambridge Univ. Press, Cambridge, 1990.
- [F] Fine, A. *“The Shaky Game: Einstein Realism and the Quantum theory”* Chicago Univ. Press, Chicago, 1990.
- [K] Kavafof, M. *“Bell’s Theorem, Quantum Theory and Conceptions of the Universe”* Kluwer Acad. Pubs., Dordrecht, 1989 .
- [KN] Kavafof, M. and Nadeau, R. *“The Conscious Universe”* Springer-Verlag, Berlin, 1990.
- [L] Lindley, D. *“Where did the weirdness go?”* Basic Books, New York, 1996.
- [M] Malament, D.B. *“Notes on the geometric interpretation of Bell type inequalities”* Downloaded from the authors web page.
- [Ma] Mayants, L. *“The Enigma of Probability in Physics”* Kluwer Acad. Pubs., Dordrecht, 1984.
- [Me] Mermin, D. *“Bringing home the atomic world: quantum mysteries for anyone”* Am. J.Phys. **49** (1981), pp. 940-943.
- [P] Peres, A. *“The classic paradoxes of quantum theory”* The Foundations of Physics, **14** (1984),pp. 1131-1145.
- [S] Silverman, M.P. *“More than a mystery”* Springer-Verlag, New York, 1995.
- [TM] Tarozzi, G. and v.d Merwe, A. *“The Nature of Quantum Paradoxes”*, Kluwer Acad. Press, Dordrecht, 1988.
- [WZ] Wheeler, J. and Zurek, H. *“Quantum Theory and Measurement”* Springer-Verlag, Berlin, 1990.

HENRYK GZYL

DEPTO. DE ESTADÍSTICA; UNIV. CARLOS III DE MADRID.

ESPAÑA

hgzyl@est-econ.uc3m.es; hgzyl@reacciun.ve

Conservación de propiedades del anillo de polinomios y del anillo de series en la clase de los anillos de conductor finito

Nelsy González, Omaidá Sepúlveda y Oswaldo Lezama

Resumen

En este artículo se analiza la conservación de propiedades del anillo de coeficientes R a los anillos de polinomios y series $R[x]$ y $R[[x]]$, respectivamente. Este estudio se realiza en la clase de los anillos de conductor finito, en particular se estudian propiedades como GCD , $G - GCD$, coherencia, cuasi coherencia y conductor finito.

ABSTRACT. In this paper is considered the preservation of properties of the ring of coefficients R to the rings $R[x]$ and $R[[x]]$, where $R[x]$ denotes the polynomials ring and $R[[x]]$ is the power series ring. This study is carry out in the context of finite conductor rings; in particular, properties like GCD , $G - GCD$, coherence, quasi coherence and finite conductor are analyzed.

1 INTRODUCCIÓN

Para un anillo conmutativo R se tiene que $R \subset R[x] \subset R[[x]]$, donde $R[x]$ es el anillo de polinomios y $R[[x]]$ el anillo de series. Hay propiedades de R que conserva tanto $R[x]$ como $R[[x]]$, pero algunas propiedades sólo las toma $R[x]$ ó $R[[x]]$: por ejemplo, es conocido que si R es un anillo local, $R[x]$ no es local y $R[[x]]$ es local. (Ver [19]). Así mismo, es conocido que si R es un anillo Noetheriano, entonces $R[x_1, \dots, x_n]$ es un anillo Noetheriano, el mismo resultado se tiene para el anillo $R[[x]]$, por otro lado, se sabe que los anillos de polinomios y de series son un dominio de Prüfer si y sólo si los coeficientes están en un cuerpo y $n = 1$ (ver [22]). También, si R es un anillo coherente, $R[x]$ no es necesariamente un anillo coherente (ver [12]). Resulta entonces interesante estudiar la preservación de propiedades del anillo de coeficientes, al anillo de polinomios $R[x_1, \dots, x_n]$ y al anillo de series $R[[x]]$. Dentro de las múltiples propiedades que pueden ser investigadas se encuentran aquellas que

definen subclases de anillos en la colección de los dominios de conductor finito, y que en la actualidad son materia de investigación entre los expertos del área. En este sentido se deben destacar los trabajos recientes de Sarah Glaz (Universidad de Connecticut, Storrs, Estados Unidos) y de Daniel Anderson (Universidad de Iowa, Iowa, Estados Unidos), los cuales constituyeron la motivación para la realización del presente trabajo.

2 PRELIMINARES

Un dominio se dice que es de **conductor finito** si la intersección de dos ideales principales es un ideal finitamente generado. Entre las subclases de los dominios de conductor finito se encuentran los **dominios cuasi-coherentes**, en los cuales la intersección de un número finito de ideales principales es nuevamente un ideal finitamente generado. Como una subclase de estos últimos se encuentran los **dominios coherentes**, cuya particularidad radica en que la intersección de dos ideales finitamente generados es un ideal finitamente generado. Otra subclase especial de los dominios de conductor finito la conforman los llamados dominios **G-GCD**, los cuales requieren que la intersección de dos ideales enteros invertibles sea un ideal entero invertible. Estos dominios tienen como subclase a los dominios **GCD**, que son aquellos en los cuales cualquier par de elementos posee máximo común divisor (*m.c.d.*). En general, propiedades como GCD, G-GCD, coherencia y conductor finito aparecen frecuentemente en la literatura, debido, a que como se mencionó antes, es de gran interés estudiar bajo qué condiciones estas propiedades son conservadas por los anillos $R[x]$ y $R[[x]]$. Por tal razón, uno de los propósitos de este trabajo es presentar el estudio de la preservación de estas propiedades.

Se adopta la notación y terminología que se encuentra en [8], [14] ó [17]. Los anillos considerados en este artículo son conmutativos con elemento unidad. Usualmente, o salvo que se especifique otra situación, R denota un dominio de integridad (DI) con cuerpo de fracciones K , $Spec(R)$ denota el espectro primo de R conformado por todos los ideales primos de R y $Max(R)$ el espectro maximal de R conformado por todos los ideales maximales de R . Además, si M es un R -módulo y $P \in Spec(R)$, entonces M_P denota la localización de M por el ideal primo P . R_P denota la localización de R por el ideal primo P , y $c(f)$ denota el ideal fraccionario generado por los coeficientes de f , con $f \in K[x]$.

Definición 2.1 Sean I, J ideales fraccionarios de R . Se define el ideal $[I :_R J] = \{r \in R \mid rJ \subset I\}$, a este ideal se le suele denominar el ideal **conductor** de J en I .

Definición 2.2 Sea I un ideal fraccionario de R . Se define el conjunto $\mathbf{I}^{-1} = [R :_K I] = \{x \in K \mid xI \subseteq R\}$.

En adelante se usará $[R : I]$ para denotar el cociente $[R :_K I]$.

Definición 2.3 Un dominio de integridad R es llamado un **dominio de Prüfer** si cada ideal finitamente generado de R es invertible.

En [22] se encuentra un estudio detallado de estos dominios, los cuales constituyen otra subclase importante de los dominios de conductor finito. Nótese que todo dominio de Prüfer es coherente. Una subclase importante de los dominios de Prüfer la conforman los dominios de Bézout. Hay que recordar que un dominio de integridad R es llamado un **dominio de Bézout** si cada ideal no nulo finitamente generado de R es principal.

Definición 2.4 Un ideal fraccionario I de un dominio de integridad R se dice **divisorial** si es una intersección de ideales fraccionarios invertibles.

El ideal fraccionario divisorial más pequeño que contiene a I se denota por I_v . Además, I_v es llamado la **clausura divisorial** de I .

Un ideal fraccionario I de R es un **v-ideal de tipo finito** si existe un ideal fraccionario finitamente generado J de R tal que $I = J_v$.

Proposición 2.5 [8] Sea R un DI e I un ideal fraccionario de R .

(i) El ideal $(I^{-1})^{-1}$ es la intersección de los ideales principales que contienen a I .

(ii) $I_v = (I^{-1})^{-1}$.

(iii) El ideal I es divisorial si y sólo si $I = I_v$.

Demostración: (i) Sea $\{\langle a_\lambda \rangle\}_{\lambda \in \Lambda}$ la familia de ideales fraccionarios principales de R que contienen a I , sea J la intersección de esta familia y sea I^{-1} el ideal $[R : I]$. Se debe concluir que $(I^{-1})^{-1} = J$.

En efecto, sea $x \in (I^{-1})^{-1}$ y $\lambda \in \Lambda$, entonces de $I \subseteq \langle a_\lambda \rangle$, se sigue que $a_\lambda^{-1}I \subseteq R$ y que $a_\lambda^{-1} \in I^{-1}$. Como $xI^{-1} \subseteq R$, entonces $xa_\lambda^{-1} \in R$ y así $x \in \langle a_\lambda \rangle$. En consecuencia, $(I^{-1})^{-1} \subseteq J$.

Recíprocamente, si $x \notin (I^{-1})^{-1}$, entonces $xI^{-1} \not\subseteq R$, esto es, existe $t \in I^{-1}$ tal que $xt \notin R$ y así $x \notin \langle t^{-1} \rangle$. Como $\langle t \rangle \subseteq I^{-1}$, entonces $I \subseteq \langle t^{-1} \rangle$. Por consiguiente, $x \notin J$.

(ii) Por la definición del ideal I_v se concluye que $I_v \subseteq (I^{-1})^{-1}$. Ahora, si $x \in (I^{-1})^{-1}$ y si J es un ideal fraccionario invertible tal que $I \subseteq J$, entonces $xI^{-1} \subseteq R$ y como $J^{-1} \subseteq I^{-1}$, se tiene que $xJ^{-1} \subseteq R$. Luego, $x \in J$ y así, $x \in I_v$.

(iii) Es consecuencia inmediata de la definición de ideal divisorial. ■

Definición 2.6 Sea D un dominio de integridad, K su cuerpo de fracciones y $t \in K$:

(i) $t \in K$ es **entero** sobre D si es raíz de un polinomio mónico con coeficientes en D .

(ii) $t \in K$ es **casi entero** sobre D si existe un elemento $d \in D$, $d \neq 0$, tal que $dt^n \in D$ para todo entero positivo n .

(iii) Un dominio de integridad D es **enteramente cerrado** si cada elemento de K entero sobre D pertenece a D .

(iv) Un dominio de integridad D es **completamente enteramente cerrado** si cada elemento de K casi entero sobre D pertenece a D .

Nótese que ser completamente enteramente cerrado implica ser enteramente cerrado (ver [22]). En el caso Noetheriano las dos condiciones coinciden.

A continuación se presentan algunas propiedades y caracterizaciones de los dominios GCD y $G - GCD$; las pruebas que incluimos son claras y sencillas y difieren o no son presentadas en la literatura disponible.

Proposición 2.7 Un dominio de integridad R es un GCD si y sólo si para todo $a, b \in R$, $\langle a \rangle \cap \langle b \rangle$ es principal.

Demostración: Para la implicación directa, si $a = 0$ ó $b = 0$, entonces $\langle a \rangle \cap \langle b \rangle$ es principal.

Sean $a, b \neq 0$ y sea $d = m.c.d(a, b)$, entonces $d \mid a$ y $d \mid b$, esto es, $a = dx$ y $b = dy$, luego $m.c.d(x, y) = 1$, nótese que $\langle x \rangle \cap \langle y \rangle = \langle xy \rangle$: como $xy \in \langle x \rangle \cap \langle y \rangle$ implica que $\langle xy \rangle \subseteq \langle x \rangle \cap \langle y \rangle$, sea ahora $z \in \langle x \rangle \cap \langle y \rangle$, entonces z es múltiplo de x y de y , por tanto z es múltiplo de $m.c.m(x, y) = m$, pero $m.c.d(x, y) = 1$, entonces $m = xy$, es decir $z \in \langle xy \rangle$ así que $\langle x \rangle \cap \langle y \rangle \subseteq \langle xy \rangle$.

Finalmente, hay que ver que $\langle a \rangle \cap \langle b \rangle = \langle dxy \rangle$: claramente $dxy \in \langle a \rangle \cap \langle b \rangle$, luego $\langle dxy \rangle \subseteq \langle a \rangle \cap \langle b \rangle$, ahora $\langle a \rangle \cap \langle b \rangle = \langle dx \rangle \cap \langle dy \rangle = \langle d \rangle \langle x \rangle \cap \langle d \rangle \langle y \rangle = \langle d \rangle (\langle x \rangle \cap \langle y \rangle) = \langle d \rangle \langle xy \rangle = \langle dxy \rangle$.

Recíprocamente, sean a, b no nulos y sea $\langle c \rangle = \langle a \rangle \cap \langle b \rangle$, $c \neq 0$ ya que R es

un dominio de integridad y $ab \in \langle c \rangle$. Note que $c = m.c.m(a, b) : c \in \langle a \rangle$ y $c \in \langle b \rangle$, entonces $a \mid c$ y $b \mid c$, sea ahora x tal que $a \mid x$ y $b \mid x$, esto indica que $x \in \langle a \rangle \cap \langle b \rangle$, entonces $x \in \langle c \rangle$ así $c \mid x$. Como $a \mid ab$ y $b \mid ab$, entonces $c \mid ab$, esto es, $ab = cd$. Se tiene que $d = m.c.d(a, b) : a \mid c$ implica que $c = aa'$, entonces $ab = aa'd$ así $b = a'd$, es decir $d \mid b$, similarmente $d \mid a$.

Sea $e \mid a$ y $e \mid b$, entonces $a = ex$ y $b = ey$, así $\langle e \rangle \langle xy \rangle \subseteq \langle e \rangle (\langle x \rangle \cap \langle y \rangle) = \langle e \rangle \langle x \rangle \cap \langle e \rangle \langle y \rangle = \langle ex \rangle \cap \langle ey \rangle = \langle a \rangle \cap \langle b \rangle = \langle c \rangle$, es decir, $\langle exy \rangle \subseteq \langle c \rangle$ luego $exy = \lambda c$.

Ahora, como $ab = cd$ se tiene que $exey = cd$, así que $e(xey) = cd$, esto es, $e(\lambda c) = cd$, entonces $\lambda e = d$, es decir, $e \mid d$. ■

Proposición 2.8 *Sea R un dominio de Prüfer, entonces R es un dominio GCD si y sólo si R es un dominio de Bézout.*

Demostración: Para la implicación directa basta probar que cada ideal generado por dos elementos no nulos a y b es principal:

Sea $d = m.c.d(a, b)$, la idea es mostrar que $\langle d \rangle = \langle a, b \rangle$. Como $d \mid a$ y $d \mid b$, entonces $a = d\alpha$ y $b = d\beta$, así cada combinación de a y b es múltiplo de d , esto implica que $\langle a, b \rangle \subseteq \langle d \rangle$. Nótese ahora que $\langle d \rangle \subseteq \langle a, b \rangle$: como R es dominio de Prüfer, entonces $I = \langle a, b \rangle$ es invertible, por tanto $1 = ax + by$ con $x, y \in I^{-1}$, sea $x = \frac{m}{n}$, $y = \frac{p}{q}$, entonces $d = da\frac{m}{n} + db\frac{p}{q}$, pero $ax, bx, by, ay \in R$, así que $ax = a\frac{m}{n} = d_1 \in R$, dualmente $bx = b\frac{m}{n} = d_2 \in R$, por tanto $am = nd_1$ y $bm = nd_2$, con lo cual $n \mid am$ y $n \mid bm$ esto implica que $n \mid m.c.d(am, bm) = md$, esto es $md = nd'$, $d' \in R$, similarmente $b\frac{p}{q} = e_1 \in R$, $a\frac{p}{q} = e_2 \in R$, entonces $q \mid bp$ y $q \mid ap$, así que $q \mid m.c.d(ap, bp) = pd$, es decir $pd = qd''$, $d'' \in R$, en consecuencia $d = \frac{and'}{n} + \frac{bqd''}{q} = ad' + bd''$. Por tanto $\langle d \rangle = \langle a, b \rangle$.

Recíprocamente, Sean $a, b \in R$ no nulos, entonces $\langle a, b \rangle = \langle d \rangle$, hay que ver que $d = m.c.d(a, b) : a \in \langle a, b \rangle = \langle d \rangle$, entonces $a = d\lambda$, de donde $d \mid a$. Análogamente, $d \mid b$. Sea $e \mid a$, i.e, $a = ea'$ y sea $e \mid b$, entonces como $d = \alpha a + \beta b$ se tiene que $d = \alpha ea' + \beta eb'$ por consiguiente $e \mid d$. ■

Proposición 2.9 *Sea R un dominio GCD, entonces R es enteramente cerrado.*

Demostración: La prueba se realiza en dos pasos:

Paso 1. Sea $t = \frac{r}{s} \in K - \{0\}$ raíz de un polinomio $a_0 + a_1x + \dots + a_nx^n \in R[x]$. Sea $d = m.c.d(r, s)$, así $r = dp$ y $s = dq$, es decir, $t = \frac{p}{q}$, además $m.c.d(p, q) = 1$, nótese que $p \mid a_0$ y $q \mid a_n : a_0 + a_1(\frac{p}{q}) + a_2(\frac{p^2}{q^2}) \dots + a_{n-1}(\frac{p^{n-1}}{q^{n-1}}) + a_n(\frac{p^n}{q^n}) = 0$,

entonces $a_0q^n + a_1pq^{n-1} + \dots + a_np^n = 0$, esto indica que $p \mid a_0q^n$, pero como $m.c.d(p, q) = 1$, se tiene que $p \mid a_0$. También, $q \mid a_np^n$ y $q \mid a_n$.

Paso 2. Sea $t = \frac{r}{s} \in K - \{0\}$ entero en R , es decir, satisface un polinomio mónico con coeficientes en R . Se puede tomar $t = \frac{p}{q}$ con $m.c.d(p, q) = 1$, luego por el Paso 1, $q \mid a_n = 1$, esto es $q \in R^*$, entonces $\frac{p}{q} \in R$. ■

Proposición 2.10 *R es un dominio $G - GCD$ si y sólo si la intersección de dos ideales fraccionarios invertibles de R es invertible.*

Demostración: Sean J_1, J_2 ideales fraccionarios invertibles de R , entonces existen r_1, r_2 no nulos en R , tales que $r_1J_1, r_2J_2 \subseteq R$. Además, r_1J_1, r_2J_2 son ideales enteros invertibles $((r_iJ_i)^{-1} = r_i^{-1}J_i^{-1}, i = 1, 2)$. También, $r_1r_2J_1, r_1r_2J_2$ son ideales enteros invertibles, por lo tanto $r_1r_2J_1 \cap r_1r_2J_2 = I \subseteq R$ es invertible, entonces $(r_1r_2)^{-1}(r_1r_2J_1 \cap r_1r_2J_2) = (r_1r_2)^{-1}I$, así $J_1 \cap J_2 = (r_1r_2)^{-1}I$ y $(r_1r_2)^{-1}I$ es invertible.

La inclusión recíproca se tiene del hecho que todo ideal entero invertible es un ideal fraccionario invertible. ■

Proposición 2.11 [7] *Sea R un dominio de integridad, entonces las siguientes afirmaciones son equivalentes:*

- (i) R es un dominio $G - GCD$.
- (ii) $\langle a \rangle \cap \langle b \rangle$ es invertible $\forall a, b \in R - \{0\}$.
- (iii) $[a :_R b]$ es invertible $\forall a, b \in R - \{0\}$.
- (iv) El grupo de ideales fraccionarios invertibles de R , denotado por $Inv(R)$, es un grupo reticular ordenado bajo el orden parcial $I \leq J$ si y sólo si $J \subseteq I$.
- (v) Cada $v -$ ideal de tipo finito es invertible.

Demostración: (i) \Rightarrow (ii) Es claro, ya que todo ideal principal es invertible.

(ii) \Rightarrow (i) Sean $I = \langle a_1, \dots, a_n \rangle$ y $J = \langle b_1, \dots, b_m \rangle$ ideales invertibles de R , es suficiente mostrar que $I \cap J$ es finitamente generado y localmente invertible, ver [22, Capítulo 1]. Sea R un anillo local e I un ideal invertible, se tiene entonces que $1 = x_1y_1 + \dots + x_ny_n$, donde $x_i \in I, y_i \in I^{-1}$ e $I = \langle x_1, x_2, \dots, x_n \rangle$, dado que $x_iy_i \in R, 1 \leq i \leq n$, se debe tener que $x_iy_i \in R^*$ para algún i , entonces existe $c \in R$ tal que $x_iy_ic = 1$, así $x_j = x_i(y_icx_j)$, como $y_i \in I^{-1}, y_ix_j \in R$, e igualmente $y_ix_jc \in R$, esto es $x_j \in \langle x_i \rangle$, entonces $I \subseteq \langle x_i \rangle \subseteq I$, así que I es principal y generado por algún x_i . Ahora, si $M \in Max(R)$ se tiene que

$I_M = \langle \frac{a_{i_1}}{1} \rangle$ y $J_M = \langle \frac{b_{j_1}}{1} \rangle$ para algún i_1, j_1 con $1 \leq i_1 \leq n$, $1 \leq j_1 \leq m$, así $I_M \cap J_M = (I \cap J)_M = \langle \frac{a_{i_1}}{1} \rangle \cap \langle \frac{b_{j_1}}{1} \rangle = \langle a_{i_1} \rangle_M \cap \langle b_{j_1} \rangle_M = (\langle a_{i_1} \rangle \cap \langle b_{j_1} \rangle)_M = \langle c \rangle_M$, es decir, $I \cap J$ es localmente invertible. Para ver que $I \cap J$ es finitamente generado nótese que $\sum_{i,j} \langle a_i \rangle \cap \langle b_j \rangle \subseteq I \cap J$, así que

$$\begin{aligned} (I \cap J)_M / (\sum_{i,j} \langle a_i \rangle \cap \langle b_j \rangle)_M &= (I_M \cap J_M) / \sum_{i,j} [\langle a_i \rangle_M \cap \langle b_j \rangle_M] \\ &= \langle a_{i_1} \rangle_M \cap \langle b_{j_1} \rangle_M / \sum_{i,j} [\langle a_i \rangle_M \cap \langle b_j \rangle_M] \\ &\subseteq \sum_{i,j} [\langle a_i \rangle_M \cap \langle b_j \rangle_M] / \sum_{i,j} [\langle a_i \rangle_M \cap \langle b_j \rangle_M] = 0, \end{aligned}$$

por consiguiente $I \cap J = \sum_{i,j} \langle a_i \rangle \cap \langle b_j \rangle$ es finitamente generado.

(i) \Rightarrow (v) Sea I un v -ideal de tipo finito, es decir $I = J_v$ para algún J fraccionario finitamente generado de R , $J = \langle u_1, \dots, u_n \rangle$ con $u_i \in K$.

Nótese que $J^{-1} = (u_1, \dots, u_n)^{-1} = \langle u_1 \rangle^{-1} \cap \dots \cap \langle u_n \rangle^{-1}$ es invertible y así $(J^{-1})^{-1}$ es invertible, es decir, $I = J_v$ es invertible.

(ii) \Leftrightarrow (iii) Es inmediata de la observación $\langle a \rangle \cap \langle b \rangle = [a :_R b] \langle b \rangle$.

(i) \Rightarrow (v) Sea $G = \text{Inv}(R)$; sean $I, J \in G$, entonces $I \cap J \subseteq I$ y $I \cap J \subseteq J$, así $I \leq I \cap J$ y $J \leq I \cap J$, además, por hipótesis $I \cap J \in G$, sea ahora K tal que $I \leq K$, $J \leq K$, entonces $K \subseteq I$ y $K \subseteq J$ con lo cual $K \subseteq I \cap J$ y por lo tanto $I \cap J \leq K$. así $I \cap J = \text{Sup}(I, J)$ indicando esto que $\text{Inv}(R)$ es un grupo reticular ordenado.

(iv) \Rightarrow (i) Sean $I, J \in G$ y sea $K = \text{Sup}(I, J)$ así $I \leq K$, $J \leq K$ lo cual indica que $K \subseteq I$ y $K \subseteq J$ y por tanto $K \subseteq I \cap J$. Sea ahora $x \in I \cap J$ entonces $x \in I$ y $x \in J$ así $\langle x \rangle \subseteq I$, $\langle x \rangle \subseteq J$, esto es, $I \leq \langle x \rangle$ y $J \leq \langle x \rangle$ y dado que $\langle x \rangle \in G$ se tiene que $K \leq \langle x \rangle$, i.e., $\langle x \rangle \subseteq K$ de donde $x \in K$; en conclusión $I \cap J = K$.

(v) \Rightarrow (iv) Si $I, J \in G$ entonces $(I + J)_v$ es invertible, ya que es un v -ideal de tipo finito. Nótese que $\text{Inf}(I, J) = (I + J)_v$: como $I \subseteq I + J \subseteq (I + J)_v$ y $J \subseteq I + J \subseteq (I + J)_v$ se tiene que $(I + J)_v \leq I$ y $(I + J)_v \leq J$. Sea ahora $K \in G$ tal que $K \leq I$ y $K \leq J$, vease entonces que $K \leq (I + J)_v$: $(I + J) \subseteq K$ implica que $(I + J)_v \subseteq K_v = K$ así $K \leq (I + J)_v$. ■

Proposición 2.12 Si R es un dominio GCD, entonces R es un dominio G -GCD.

Demostración: Sean I, J ideales enteros invertibles de R , entonces I, J son finitamente generados, denotados por $I = \langle a_1, a_2, \dots, a_t \rangle$, $J = \langle b_1, b_2, \dots, b_l \rangle$

hay que ver que $I \cap J$ es invertible, para esto se puede demostrar que $I \cap J$ es finitamente generado y que todas sus localizaciones maximales son invertibles, nótese que esto se tiene inmediatamente con un razonamiento análogo a la prueba de la proposición anterior para el caso $(ii) \Rightarrow (i)$. ■

Nótese que si R es un dominio de Prüfer, entonces R es un dominio G - GCD .

Corolario 2.13 [7] *Sea R un dominio G - GCD , S un sistema multiplicativo, entonces RS^{-1} es un dominio G - GCD , además R_P es un dominio GCD $\forall P \in \text{Spec}(R)$ y entonces R es enteramente cerrado.*

Demostración: Dados $\langle \frac{a}{r} \rangle, \langle \frac{b}{s} \rangle \in RS^{-1}$, se tiene que $\langle \frac{a}{r} \rangle \cap \langle \frac{b}{s} \rangle = \langle \frac{a}{1} \rangle \cap \langle \frac{b}{1} \rangle = \langle a \rangle S^{-1} \cap \langle b \rangle S^{-1} = (\langle a \rangle \cap \langle b \rangle) S^{-1}$. Si $\langle a \rangle \cap \langle b \rangle = I$ es invertible, entonces $II^{-1} = R$, así que $(II^{-1})S^{-1} = RS^{-1}$ y por tanto $IS^{-1}I^{-1}S^{-1} = RS^{-1}$ indicando esto que IS^{-1} es invertible, con lo cual $\langle \frac{a}{r} \rangle \cap \langle \frac{b}{s} \rangle$ es invertible, en consecuencia RS^{-1} es un dominio G - GCD . Véase ahora que R_P es un dominio GCD : dado que R_P es un anillo local y en anillos locales los ideales invertibles son principales se tiene que la intersección de ideales principales es un ideal principal.

Finalmente, ya que $R = \bigcap_{P \in \text{Max}(R)} R_P$ (ver [22, Capítulo 1]) y que cada R_P es enteramente cerrado, (por ser R_P un GCD) se puede concluir que R es enteramente cerrado. ■

3 ANILLO DE POLINOMIOS

A la fecha son muy pocos los resultados que se tienen en lo concerniente a la preservación de la propiedad de conductor finito de un anillo de polinomios, es más, este aún continua siendo un problema abierto. En esta sección se presentan los resultados concocidos en este contexto, mostrando el estado del arte del problema.

Para probar los cuatro teoremas centrales de esta sección son necesarias varias propiedades de tipo técnico relativas a ideales divisoriales y polinomios primitivos. Las demostraciones de varias de estas afirmaciones preliminares no se encuentran en la literatura y aquí son presentadas en forma clara.

Lema 3.1 [7] *Sea R un dominio de integridad con cuerpo de fracciones K , e I un ideal fraccionario no nulo de R , entonces:*

$$[R[x] : IR[x]] = [R : I][x], \text{ es decir,}$$

$$(IR[x])^{-1} = I^{-1}R[x] \text{ y así } (IR[x])_v = I_vR[x]$$

Demostración: Nótese que $[R[x] : IR[x]] \subseteq K[x]$: en efecto, existe $t \neq 0$ en R , tal que $tI \subseteq R$, sea $\frac{p(x)}{q(x)} \in [R[x] : IR[x]]$, entonces $\frac{p(x)}{q(x)}I[x] \subseteq R[x]$, sea $\alpha \neq 0$ en I , luego $\frac{p(x)}{q(x)}\alpha 1 = r(x)$, entonces $\frac{p(x)}{q(x)}\alpha t = tr(x)$, pero $\alpha t = s \in R$, por consiguiente $\frac{p(x)}{q(x)} = \frac{tr(x)}{s} \in K[x]$.

Ahora, sea $g \in (IR[x])^{-1}$, equivalentemente $g \in (R[x] : IR[x])$ sí y solo si $g \in [R[x] : I[x]]$ lo que equivale a $gI[x] \subseteq R[x]$ sí y solamente si $c(g)I \subseteq R$, lo cual conlleva a que $c(g) \subseteq [R : I]$, es decir $c(g) \subseteq I^{-1}$, así que $g \in I^{-1}R[x]$. Para la última parte si $f \in (IR[x])_v$, es decir $f \in ((IR[x])^{-1})^{-1}$ equivale a que $f \in (I^{-1}R[x])^{-1}$, por consiguiente $f \in (I^{-1})^{-1}R[x]$ así que $f \in I_vR[x]$. ■

Lema 3.2 [7] *Si I_1 es un ideal divisorial de $R[x]$ e $I = I_1 \cap K \neq 0$, entonces I es un ideal divisorial de R . Más precisamente,*

$$I = \cap \left\{ r(c(g))^{-1} \mid I_1 \subseteq \left\langle \frac{r}{g} \right\rangle R[x], r \in R, g \in R[x] \right\}.$$

Demostración: Observese que si I_1 es un ideal divisorial de $R[x]$, entonces $I_1 = \cap \left\{ \left\langle \frac{f}{h} \right\rangle R[x] \mid I_1 \subseteq \left\langle \frac{f}{h} \right\rangle R[x], h \neq 0 \right\}$; Si $I = I_1 \cap K \neq 0$, y además $I_1 \subseteq \left\langle \frac{f}{h} \right\rangle R[x]$, entonces $I_1 \cap K \subseteq \left\langle \frac{f}{h} \right\rangle \cap K$, por consiguiente $\left\langle \frac{f}{h} \right\rangle R[x] \cap K \neq 0$, y existen $r \in R, g \in R[x]$, tales que $\left\langle \frac{f}{h} \right\rangle R[x] = \left\langle \frac{r}{g} \right\rangle R[x]$: en efecto, sea $\alpha \in K \cap \left\langle \frac{f}{h} \right\rangle R[x]$, esto es, $\alpha = \frac{r}{t} = p\frac{f}{h}$, con $r, t \in R - \{0\}, p \in R[x] - \{0\}$, de tal forma que $\frac{f}{h} = \frac{r}{pt} = \frac{r}{g}, g \in R[x] - \{0\}$. Así $I_1 = \cap \left\{ \left\langle \frac{r}{g} \right\rangle R[x] \mid I_1 \subseteq \left\langle \frac{r}{g} \right\rangle R[x], r \in R, 0 \neq g \in R[x] \right\}$. Sea $J = \cap \left\{ r(c(g))^{-1} \mid I_1 \subseteq \left(\frac{r}{g} \right) R[x], 0 \neq g \in R[x], r \in R \right\}$, nótese que $I = J$:

⊆) Sea $u \in I$, así $u \in I_1$, por tanto $u \in \left\langle \frac{r}{g} \right\rangle$, es decir $u = h\frac{r}{g}$, de tal forma que $ug = hr \in \langle r \rangle$, por consiguiente $ug \in rR[x], r \in R, 0 \neq g \in R[x]$, $I_1 \subseteq \left\langle \frac{r}{g} \right\rangle R[x]$, entonces $uc(g) \subseteq rR$, luego $u \in J$.

⊇) Sea $u \in J, u \in rc(g)^{-1}$, para todo r, g tal que $I_1 \subseteq \left\langle \frac{r}{g} \right\rangle R[x], c(g)u \subseteq rR$, entonces $ug \in rR[x]$, luego $u \in \frac{r}{g}R[x]$, así que $u \in I_1 \cap K = I$.

La conclusión de que I es un ideal divisorial de R se debe a que I es intersección de ideales divisoriales de R . Nótese $r(c(g))^{-1}$ es divisorial; en efecto, $r[\langle g_0, g_1, \dots, g_n \rangle]^{-1} = r[\langle g_0 \rangle^{-1} \cap \dots \cap \langle g_n \rangle^{-1}]$, y $(r(\langle g_0 \rangle^{-1} \cap \dots \cap \langle g_n \rangle^{-1})^{-1})^{-1} = r[\langle g_0 \rangle^{-1} \cap \dots \cap \langle g_n \rangle^{-1}] = rc(g)^{-1}$. ■

Proposición 3.3 [7] *Sea R un dominio enteramente cerrado, I_1 un ideal entero divisorial de $R[x]$, entonces:*

(i) *Si $I = I_1 \cap R \neq 0$, entonces I es un ideal entero divisorial de R y además $I_1 = IR[x]$.*

(ii) *Si $I_1 \cap R = 0$, entonces existe $f \in R[x]$ y un ideal I divisorial de R tal que $I_1 = fIR[x]$.*

Demostración: (i) Por hipótesis $I_1 \subseteq R[x]$, además $I = I_1 \cap R = I_1 \cap K = \cap \left\{ r(c(g))^{-1} : I_1 \subseteq \left(\frac{r}{g}\right)R[x] \right\}$; nótese que por el lema anterior I es divisorial, véase entonces que $I_1 = IR[x]$: es claro que $IR[x] \subseteq I_1$, para la inclusión recíproca sea $f \in I_1$, $fg \in rR[x]$, $r \in R, 0 \neq g \in R[x]$, así $c(fg) \subseteq rR$ y como R es enteramente cerrado se tiene que $c(f)c(g) \subseteq (c(f)c(g))_v = c(fg)_v \subseteq rR$, ([7, Lema 6.1.2]) entonces $c(f) \subseteq [rR : c(g)] = r[R : c(g)]$ y $c(f) \subseteq I$, esto es, $f \in IR[x]$.

(ii) Si $I_1 \cap R = 0$, existe $f \in I_1, f \neq 0$ tal que $I_1K[x] = fK[x]$: en efecto, $fK[x] \subseteq I_1K[x]$, ahora $I_1K[x] = \langle h(x) \rangle_{K(x)}, h(x) = \frac{g(x)}{r} \in I_1$, entonces $h(x) = \left\langle \frac{g(x)}{r} \right\rangle = \left\langle \frac{1}{r}g(x) \right\rangle = \langle f(x) \rangle$. Luego $I_1 \subseteq I_1K[x] \cap R[x] = fK[x] \cap R[x] = f(c(f))^{-1}R[x]$, sea $0 \neq a \in R$ tal que $a(c(f))^{-1} \subseteq R$, como $I_1 \subseteq f(c(f))^{-1}R[x]$, $aI_1 \subseteq f(ac(f)^{-1}R[x])$, $aI_1 \subseteq fL$, donde $L = ac(f)^{-1}R[x]$, así que $\frac{a}{f}I_1 \subseteq L \subseteq R[x]$, observese que $J_1 = \frac{a}{f}I_1$ es un ideal entero divisorial de $R[x]$, así $J_1K[x] = \left(\frac{a}{f}I_1\right)K[x] = K[x]$, luego $J_1K[x] = K[x]$ y $J = J_1 \cap R \neq 0$, por tanto $a \in J$, por parte (i) $J_1 = JR[x]$, con lo cual $I_1 = fa^{-1}J_1$, $I_1 = fa^{-1}JR[x]$ donde $a^{-1}J$ es un ideal divisorial de R . ■

Proposición 3.4 [7] *Sea R un dominio de integridad con cuerpo de fracciones K , entonces R es enteramente cerrado si y sólo si cada ideal entero I de $R[x]$ con $I_v \cap R \neq 0$ satisface $I_v = (c(I)[x])_v = c(I)_v[x]$; donde $c(I)$ es el ideal de R generado por los coeficientes de elementos de I .*

Demostración: \Rightarrow) Dado que R es enteramente cerrado, se aplica la proposición anterior y se tiene que $I_v = (I_v \cap R)[x]$, donde $I_v \cap R$ es un ideal entero

divisorial de R , así $c(I_v) = I_v \cap R$, claramente $I \subseteq c(I)[x]$, por consiguiente $I_v \subseteq (c(I)[x])_v$, además por el Lema 14 se concluye que $(c(I)[x])_v = c(I)_v[x]$, por tanto $c(I_v) \subseteq c(I)_v$, además $c(I)_v \subseteq c(I_v)$; en efecto, dado que $I \subseteq I_v$, se tiene que $c(I) \subseteq c(I_v)$ y como $c(I_v) = I_v \cap R$ es un ideal divisorial de R , se deduce que $(c(I_v))_v = c(I_v)$, por tanto $c(I)_v \subseteq (c(I_v))_v = c(I_v)$. Así que $I_v = (I_v \cap R)[x] = c(I_v)[x] = c(I)_v[x] = (c(I)[x])_v$.

\Leftarrow) Para esta implicación sean $f, g \in R[x] - \{0\}$, y $r \in R - \{0\}$ tal que $c(fg) \subseteq rR$, de esta inclusión se tiene que $\langle rf, rg \rangle \subseteq rR[x] : rf \in rR[x], r \in \frac{r}{f}R[x], fg \in rR[x], g \in \frac{r}{f}R[x]$, así que $\langle r, g \rangle \subseteq \frac{r}{f}R[x]$ es divisorial por ser principal. Por hipótesis $(\langle r, g \rangle)_v = c(\langle r, g \rangle)_v[x]$, (note que $\langle r, g \rangle_v \cap R \neq 0; \langle r, g \rangle \subseteq \langle r, g \rangle_v, r \neq 0$) por tanto $c(g)[x] \subseteq c(\langle r, g \rangle)[x] \subseteq c(\langle r, g \rangle)_v[x] = \langle r, g \rangle_v \subseteq \frac{r}{f}R[x]$, de donde $fc(g)[x] \subseteq rR[x]$ y así $c(f)c(g) \subseteq rR$, lo cual implica que R es enteramente cerrado (ver [15, Lema 1.31 (iv)]). ■

Proposición 3.5 (Lema de Gauss) *Sea R un dominio GCD, si $f, g \in R[x]$ son polinomios primitivos, entonces fg es un polinomio primitivo.*

Demostración: Por efectos de espacio la prueba se ilustra para polinomios de primer y tercer grado; mostrando un modelo que puede seguirse para polinomios de cualquier orden finito.

Sean $f = f_0 + f_1x$ y $g = g_0 + g_1x$, entonces $h = fg = f_0g_0 + (f_0g_1 + f_1g_0)x + f_1g_1x^2$. Sean $d = m.c.d(f_0g_0, f_0g_1 + f_1g_0, f_1g_1)$, $u = m.c.d(f_0, d)$, así $u \mid f_0g_1, u \mid f_0g_1 + f_1g_0, u \mid f_1g_1$, luego $u \mid f_1g_0$, y $u \mid f_1g_1$, entonces $u \mid m.c.d(f_1g_0, f_1g_1)$, así que $u \mid f_1m.c.d(g_0, g_1)$, entonces $u \mid f_1$, pero como $u \mid f_0$, se tiene que $u \mid m.c.d(f_0, f_1)$, por consiguiente $u = 1$, es decir, $m.c.d(f_0, d) = 1$, pero dado que $d \mid f_0g_0$, se tiene que $d \mid g_0$. Análogamente, sea $v = m.c.d(f_1, d)$, así $v \mid f_0g_0, v \mid f_0g_1 + f_1g_0, v \mid f_1g_1$, luego $v \mid f_0g_1$, y $v \mid f_1g_1$, entonces $v \mid m.c.d(f_0g_1, f_1g_1)$, así que $v \mid g_1m.c.d(f_0, f_1)$, entonces $v \mid g_1$, pero como $v \mid g_0$, se tiene que $v = 1$, es decir, $m.c.d(f_1, d) = 1$, y $d \mid f_1g_1$, así que $d \mid g_1$, por tanto $d = 1$ y el caso del producto de polinomios primitivos lineales queda concluido.

Ahora, sean $f = f_0 + f_1x + f_2x^2 + f_3x^3$ y $g = g_0 + g_1x + g_2x^2 + g_3x^3$, entonces $h = fg = f_0g_0 + (f_0g_1 + f_1g_0)x + (f_0g_2 + f_1g_1 + f_2g_0)x^2 + (f_0g_3 + f_1g_2 + f_2g_1 + f_3g_0)x^3 + (f_1g_3 + f_2g_2 + f_3g_1)x^4 + (f_2g_3 + f_3g_2)x^5 + (f_3g_3)x^6$. Sea $d = m.c.d(f_0g_0, f_0g_1 + f_1g_0, f_0g_2 + f_1g_1 + f_2g_0, f_0g_3 + f_1g_2 + f_2g_1 + f_3g_0, f_1g_3 + f_2g_2 + f_3g_1, f_2g_3 + f_3g_2, f_3g_3)$. Existen una serie de máximos común divisores u, u', u'' , y para cada uno una serie de divisibilidades así: sean $u = m.c.d(f_0, d)$, $u' = m.c.d(u, f_1) = m.c.d(d, f_0, f_1)$ y $u'' = m.c.d(u', f_2) = m.c.d(d, f_0, f_1, f_2)$ entonces $u'' \mid f_3g_0, u'' \mid f_3g_1, u'' \mid f_3g_2$, y $u'' \mid f_3g_3$, por consiguiente $u'' \mid m.c.d(f_3g_0, f_3g_1, f_3g_2, f_3g_3)$, es decir $u'' \mid f_3m.c.d(g_0, g_1, g_2, g_3)$, así $u'' \mid f_3$. Como $u'' \mid f_0, u'' \mid f_1, u'' \mid f_2$ y $u'' \mid f_3$, se tiene que $u'' = 1$.

Ahora para u' se tiene que $u' \mid f_2g_0$, pero como $m.c.d(u', f_2) = u'' = 1$, se tiene que $u' \mid g_0$. También $u' \mid f_2g_1 + f_3g_0$, así que $u' \mid f_2g_1$, pero $m.c.d(u', f_2) = u'' = 1$, así que $u' \mid g_1$. De igual forma $u' \mid f_2g_2 + f_3g_1$, luego $u' \mid f_2g_2$, pero $m.c.d(u', f_2) = u'' = 1$, así que $u' \mid g_2$. Así mismo $u' \mid f_2g_3 + f_3g_2$, por tanto $u' \mid f_2g_3$, pero $m.c.d(u', f_2) = u'' = 1$, así que $u' \mid g_3$. Como $u' \mid g_0$, $u' \mid g_1$, $u' \mid g_2$, $u' \mid g_3$, se tiene que $u' = 1$.

Razonando para u se tiene que $u \mid f_1g_0$, pero $m.c.d(u, f_1) = u' = 1$, así que $u \mid g_0$. Además $u \mid f_1g_1 + f_2g_0$, por tanto $u \mid f_1g_1$, pero $m.c.d(u, f_1) = u' = 1$, así que $u \mid g_1$. Similarmente, $u \mid f_1g_3 + f_2g_1 + f_3g_0$, así $u \mid f_1g_3$, pero $m.c.d(u, f_1) = u' = 1$, así que $u \mid g_3$. Finalmente, $u \mid f_1g_3 + f_2g_1 + f_3g_0$, luego $u \mid f_1g_3$, pero $m.c.d(u, f_1) = u' = 1$, así que $u \mid g_3$. Recapitulando, $u \mid g_0$, $u \mid g_1$, $u \mid g_2$ y $u \mid g_3$, por consiguiente $u = 1$.

Con un procedimiento análogo al anterior se obtiene que $m.c.d(g_0, d) = 1$. Como $d \mid f_0g_0$ y además $u = m.c.d(f_0, d) = 1$ y $m.c.d(g_0, d) = 1$ se concluye que $d = 1$. ■

Proposición 3.6 *Sea R un dominio GCD y sea $f = hg$ un polinomio primitivo, entonces h y g son polinomios primitivos.*

Demostración: Sean $h = h_0 + h_1x + \dots + h_tx^t$ y $g = g_0 + g_1x + \dots + g_mx^m$, entonces $f = h_0g_0 + (h_0g_1 + h_1g_0)x + \dots + h_tg_mx^{t+m}$. Sea $d = m.c.d(h_0, h_1, \dots, h_t)$, entonces $d \mid h_i$ $0 \leq i \leq t$, luego d divide a todo coeficiente de f , es decir, d divide al máximo común divisor de los coeficientes de f , esto es, $d \mid 1$, luego $d = 1$ y h es primitivo. El mismo razonamiento se utiliza para g . ■

Proposición 3.7 *Sea R un dominio GCD, si g es un polinomio primitivo no constante en $R[x]$, entonces g es un producto finito de polinomios primos.*

Demostración: Sea g un polinomio primitivo no constante de $R[x]$, entonces $g = k_1k_2\dots k_t$ donde los k_i son irreducibles en $K[x]$, ahora $g = \frac{1}{r_1}h_1\dots\frac{1}{r_t}h_t$ con $h_i \in R[x]$, entonces $r_1\dots r_tg = h_1\dots h_t$, como R es un GCD se tiene que $r_1\dots r_tg = h_1\dots h_t = a_1g_1\dots a_tg_t$, con $g_i \in R[x]$ polinomios primitivos. Tomando máximo común divisor a los coeficientes de los polinomios obtenidos en la igualdad anterior se tiene que $r_1\dots r_t = a_1\dots a_t$, con lo cual $g = g_1\dots g_t$, $g_i \in R[x]$ primitivos. Nótese que cada g_i es irreducible en $R[x]$, (de lo contrario g_i sería reducible en $K[x]$ y entonces $k_i = \frac{a_i}{r_i}g_i$ sería reducible, lo cual es una contradicción).

Véase ahora que $\langle g_i \rangle$ es primo en $R[x]$: en efecto, dada la inclusión canónica $\iota : R[x] \hookrightarrow K[x]$, $\langle g_i \rangle$ en $K[x]$ es primo debido a que $K[x]$ es un PID y g_i irreducible, resta ver que $\langle g_i \rangle$ es primo en $R[x]$, i.e. $\iota^{-1}(\langle g_i \rangle) = \langle g_i \rangle_{R[x]}$: en efecto,

$\iota^{-1}(\langle g_i \rangle)$ es un ideal primo en $R[x]$, sea $hg_i \in \langle g_i \rangle_{R[x]}$, entonces $hg_i \in K[x]$, así que $hg_i \in \iota^{-1}(\langle g_i \rangle)$, luego $\langle g_i \rangle_{R[x]} \subseteq \iota^{-1}(\langle g_i \rangle)$.

Para la otra contención, sea t un polinomio de $R[x]$ tal que $t \in \iota^{-1}(\langle g_i \rangle)$, entonces $\iota(t) \in \langle g_i \rangle$, así $t = kg_i$, $k \in K[x]$, más precisamente, $t = \frac{a}{r}hg_i$, $h \in R[x]$ primitivo, por tanto, $rt = ahg_i$, entonces $r.b.t' = a.h.g_i$, con t' un polinomio primitivo, así tomando $m.c.d$ a ambos miembros de la igualdad anterior se sigue que $r.b = a$, y reemplazando se obtiene que $t = \frac{rb}{r}hg_i$, $t = bhg_i \in R[x]$, y $t \in \langle g_i \rangle_{R[x]}$. Por tanto, $g = g_1 \dots g_t$ es producto de polinomios primos. ■

Proposición 3.8 Sea R un dominio GCD, si f_1, g_1 son dos polinomios primitivos en $R[x]$, entonces f_1 y g_1 poseen máximo común divisor.

Demostración: Para la prueba se presentan dos casos:

Caso 1. Por proposición anterior $f_1 = P_1 \dots P_l$ y $g_1 = Q_1 \dots Q_s$, con P_i, Q_i primos y f_1, g_1 polinomios primitivos, si $\{P_1, \dots, P_l\} \cap \{Q_1, \dots, Q_s\} = \emptyset$ entonces $m.c.d(f_1, g_1) = 1$: sea h tal que $h \mid f_1$ y $h \mid g_1$, entonces $f_1 = hf'$ y $g_1 = hg'$, como f_1, g_1 son primitivos, h es un polinomio primitivo, (ver Proposición 19). Si h fuese una constante, sería invertible y así $m.c.d(f_1, g_1) = 1$, si no es constante, por ser h, f', g' polinomios primitivos, se descomponen en producto de primos, así $h = T_1 \dots T_k$, $f' = U_1 \dots U_w$, $g' = V_1 \dots V_v$, con T_i, U_i, V_i polinomios primos. Como $\langle f_1 \rangle$ es principal invertible y se descompone en un producto finito de ideales primos, esta descomposición es única, dualmente para g_1 , pero como no hay elementos comunes entre f_1 y g_1 , entonces h es constante invertible. Luego $m.c.d(f_1, g_1) = 1$.

Caso 2. $\{P_1, \dots, P_l\} \cap \{Q_1, \dots, Q_s\} \neq \emptyset$, se supone que P_1, \dots, P_t son los primos comunes tales que $f_1 = (P_1^{\theta_1} \dots P_t^{\theta_t}) \cdot (\text{otros primos})$ y $g_1 = (P_1^{\theta_1} \dots P_t^{\theta_t}) \cdot (\text{otros primos})$, hay que ver que $m.c.d(f_1, g_1) = (P_1^{\theta_1} \dots P_t^{\theta_t})$: en efecto, sea h tal que $h \mid f_1$ y $h \mid g_1$, entonces h es un polinomio primitivo, si h no es constante, entonces por ser h primitivo, es un producto finito de primos, es decir, $h = (Q_1^{\gamma_1} \dots Q_t^{\gamma_t})$, con Q_i primos, así mismo f' y g' son también un producto de polinomios primos, y la descomposición de f', g' es única, así que $\{Q_1, \dots, Q_t\} \subseteq \{P_1, \dots, P_t\}$, nótese que $\gamma_j \leq \theta_j$: en efecto, si para algún j , $\gamma_j > \theta_j$, entonces hay un primo común $P_j^{\gamma_j} \mid f_1$, $P_j^{\gamma_j} \mid g_1$ lo cual es una contradicción, luego $\gamma_j \leq \theta_j$. ■

Teorema 3.9 Sea R un dominio GCD con cuerpo de fracciones K , entonces las siguientes afirmaciones son equivalentes:

- (i) R es un dominio GCD.
- (ii) $R[x]$ es un dominio GCD.

Demostración: (i) \Rightarrow (ii) Sean f, g no unidades en $R[x] - \{0\}$, tales que $f = af_1$, $g = bg_1$, donde $a, b \in R$ y f_1, g_1 son polinomios primitivos en $R[x]$, por hipótesis sea $c = m.c.d(a, b)$, $c \in R$, como f_1, g_1 son polinomios primitivos, tienen máximo común divisor h en $R[x]$, (proposición anterior) y es tal que h es primitivo. De otro lado, ch es divisor común de f y g en $R[x]$, nótese que ch es el máximo común divisor: en efecto, supóngase que s es divisor común de f y g en $R[x]$, entonces se escribe $s = ds_1$, $d \in R$ y s_1 primitivo en $R[x]$, así existen $t, l \in R[x]$ tales que $f = st$ y $g = sl$, con $t = e_1t_1$ y $l = e_2l_1$, donde $e_1, e_2 \in R$ y t_1, l_1 son polinomios primitivos en $R[x]$, de esto se sigue que $f = ds_1e_1t_1 = de_1s_1t_1 = af_1$, con s_1t_1 primitivo, además como la presentación de f es única, se tiene que $d \mid a$ y $s_1 \mid f_1$ en $R[x]$, similarmente $g = ds_1e_2l_1 = de_2s_1l_1 = bg_1$, de donde $d \mid b$ y $s_1 \mid g_1$ en $R[x]$, como d divide simultáneamente a a y a b , se tiene que $d \mid c$ de igual forma $s_1 \mid h$ en $R[x]$, lo cual conduce a que $s \mid ch$. En consecuencia ch es el máximo común divisor de f y g .

(ii) \Rightarrow (i) Es inmediata, ya que $R \subseteq R[x]$, y para $f \in R[x]$, f tiene grado positivo, así por razones de grado los polinomios constantes tienen máximo común divisor en R . ■

Una prueba inductiva sobre el número de variables de $R[x_1, \dots, x_n]$, muestra que el teorema anterior es también válido para el caso de n variables.

En virtud del teorema anterior y del hecho que dentro de la gama de los dominios de Prüfer los únicos dominios que son GCD son los Bézout, se concluye que en las subclases de los dominios de Prüfer los únicos dominios que satisfacen que el anillo $R[x]$ sea GCD son los dominios de Bézout.

Para el siguiente teorema se considera la propiedad $G-GCD$. En particular, ahí se muestra que $R[x_1, \dots, x_n]$ es un dominio $G-GCD$ si y sólo si R es un dominio $G-GCD$. De igual forma, el resultado obtenido en este teorema permite concluir que si R es un dominio de Prüfer, entonces $R[x_1, \dots, x_n]$ es un dominio $G-GCD$.

Es conocido que el conjunto $S = \{f \in R[x] : c(f) = R\}$ es un sistema multiplicativo de $R[x]$.

Teorema 3.10 *Sea R un dominio de integridad con cuerpo de fracciones K , entonces las siguientes afirmaciones son equivalentes:*

- (i) R es un dominio $G-GCD$.
- (ii) $R[x]$ es un dominio $G-GCD$.
- (iii) $R(x)$ es un dominio GCD , donde $R(x) = R[x]S^{-1}$.

Demostración: (i) \Rightarrow (ii) Según la Proposición 11 se debe probar que cada v -ideal de tipo finito es invertible; sea I un v -ideal de tipo finito en $R[x]$, i.e., $I = \langle k_1 \dots k_n \rangle_v$ donde $k_i \in L$, el cuerpo de fracciones de $R[x]$, como I es un ideal fraccionario en $R[x]$ se puede escoger un $g \in R[x] - \{0\}$ tal que $gI \subseteq R[x]$. Note que gI es un ideal entero divisorial de tipo finito y que I es invertible si y sólo si gI es invertible. Así, se puede asumir que I es entero en $R[x]$, para probar que I es invertible se consideran dos casos:

(1) $I \cap R \neq 0$, aplicando la Proposición 17 y el Corolario 13 se obtiene que $I = \langle k_1 \dots k_n \rangle_v = c \langle k_1 \dots k_n \rangle_v[x]$. Dado que R es un $G - GCD$ se puede ver que $c \langle k_1 \dots k_n \rangle_v$ es un ideal invertible de R y así I es un ideal invertible de $R[x]$ (ver Lema 14).

(2) $I \cap R = 0$, argumentando como en la Proposición 16 (ii), existe un $0 \neq a \in R$ y $f \in I$, $f \neq 0$ tal que $J = \frac{a}{f}I$ es un ideal entero divisorial de $R[x]$ de tipo finito y $J \cap R \neq 0$, así por el caso (1), J es un ideal invertible de $R[x]$ y por tanto I es un ideal invertible de $R[x]$.

(ii) \Rightarrow (iii) Recuérdese que por Corolario 13, $R(x)$ es un dominio $G - GCD$, y que cada ideal invertible de $R(x)$ es principal ($PicR(x) = 0$), lo cual indica que $R(x)$ es un dominio GCD .

(iii) \Rightarrow (i) Sean I, J ideales invertibles de R , por Lema 14, $IR(x)$ y $JR(x)$ son ideales invertibles de $R(x)$ y por hipótesis son ideales principales de $R(x)$, por lo tanto $(I \cap J)R(x) = IR(x) \cap JR(x)$ es un ideal principal de $R(x)$ y así $I \cap J$ es un ideal invertible de R . ■

Nótese que la Proposición 16 fue establecida para dominios de integridad, desconocemos si se cumple para anillos con divisores de cero. De ser así, ayudaría a resolver el siguiente interrogante: ¿Si R es un anillo $G - GCD$, entonces $R[x]$ es un anillo $G - GCD$?

En lo referente a la preservación de la coherencia por el anillo de polinomios la situación es más complicada, ya que las técnicas clásicas de manejo de ideales no son suficientes. Para este caso se adoptan procedimientos propios de álgebra homológica, algunas condiciones de planitud y finitud.

Aquí vale la pena resaltar que existen dominios coherentes sobre los cuales el anillo de polinomios $R[x]$ no es coherente, esto lo muestra el ejemplo de Soublin en el cual $S_i = Q[[t, u]]$ denota el anillo de series en dos variables t, u sobre los racionales, y $S = \prod_{i=1}^{\infty} S_i$. Soublin mostró que S es un anillo coherente de $w.dim S = 2$ y que el anillo de polinomios $S[x]$ no es un anillo coherente (ver [14]).

La técnica usada en [9] para la prueba del siguiente teorema es el álgebra homológica y resulta demasiado extensa. A continuación se presenta una prueba corta a través de ideales divisoriales.

Teorema 3.11 *Sea R un dominio coherente enteramente cerrado, entonces $R[x]$ es un dominio cuasi-coherente.*

Demostración: Sean $f_1, f_2, \dots, f_n \in R[x]$, $I = \langle f_1 \rangle \cap \langle f_2 \rangle \cap \dots \cap \langle f_n \rangle$ y $J = I \cap R$. Para ver que I es finitamente generado se consideran dos casos:

(1) Si $J \neq 0$: nótese que I es un ideal divisorial de $R[x]$, entonces por la Proposición 16, $I = JR[x]$, con J un ideal entero divisorial de R . Por consideraciones de grado $f_1, f_2, \dots, f_n \in R$, y teniendo en cuenta que R es coherente, se sigue que J es finitamente generado. Luego I también es finitamente generado.

(2) Si $J = 0$: por la demostración de la parte (ii) de la Proposición 16 se tiene que existen un elemento no nulo a en R y un polinomio no nulo $f \in I$, tales que $J_1 = \frac{a}{f}I$ es un ideal entero divisorial de $R[x]$. Sea $J_2 = J_1 \cap R$, según la prueba mencionada, J_2 es un ideal entero divisorial no nulo de R y se puede aplicar el caso (1) para decir que J_1 es finitamente generado como ideal de $R[x]$. Entonces $J_1 = \langle j_1(x), \dots, j_t(x) \rangle$ y $\langle \frac{f}{a}j_1(x), \dots, \frac{f}{a}j_t(x) \rangle = I$, es decir, I es finitamente generado. ■

Del teorema anterior se puede deducir que si R es un dominio de Prüfer, entonces $R[x]$ es un dominio cuasi-coherente.

La misma técnica usada en el teorema anterior permite demostrar el siguiente teorema.

Teorema 3.12 *Sea R un dominio cuasi coherente enteramente cerrado, entonces $R[x]$ es un dominio de conductor finito.*

En la misma dirección de los resultados anteriores se plantea la siguiente pregunta, cuya respuesta desconocemos: ¿Si R es dominio de conductor finito, entonces $R[x]$ es $G - GCD$? (ver [11]).

4 ANILLO DE SERIES

En esta sección se consideran las propiedades GCD y $G-GCD$ para el caso del anillo de series, mostrando a través de dos ejemplos que dichas propiedades no se preservan. Es de notar que la propiedad de coherencia tampoco se prevea al anillo de series (ver [14]). Para comenzar veamos algunos preliminares.

Teorema 4.1 [17] *Sea K es un cuerpo, entonces $K[[x]]$ es dominio de factorización única (UFD) y por lo tanto, completamente enteramente cerrado.*

Demostración: Como K es cuerpo, $K[[x]]$ es un dominio de ideales principales, más exactamente $\{x^n\}_{n=1}^{\infty}$ es el conjunto de todos los ideales no nulos, con único ideal maximal $\langle x \rangle$. Así, como es un dominio de ideales principales, por el Teorema 1.6 de [24] es UFD y por el Teorema 1.7 de [24] si es UFD es completamente enteramente cerrado. ■

Teorema 4.2 [8] *Si D es un dominio de integridad, entonces $D[[x]]$ es completamente enteramente cerrado, si y sólo si, D es completamente enteramente cerrado.*

Demostración: Se tiene el siguiente diagrama conmutativo para D dominio de integridad, K su cuerpo de fracciones, L el cuerpo de fracciones de $D[[x]]$ y $K((x))$ el cuerpo de fracciones de $K[[x]]$:

$$\begin{array}{ccccc} K & \hookrightarrow & L & \hookrightarrow & K((X)) \\ & \uparrow & \uparrow & & \uparrow \\ D & \hookrightarrow & D[[X]] & \hookrightarrow & K[[X]] \end{array}$$

Para la implicación directa, sea K el cuerpo de fracciones de D , y se tiene que $D = D[[x]] \cap K$. Por hipótesis, $D[[x]]$ es completamente enteramente cerrado y como K es completamente enteramente cerrado y la intersección de dominios completamente enteramente cerrados es completamente enteramente cerrado (ver [8]), entonces D es completamente enteramente cerrado.

Recíprocamente, el cuerpo de fracciones L de $D[[x]]$ es subcuerpo de $K((x))$ y $K[[x]]$ (según el diagrama inicial de esta prueba); y $K[[x]]$ es un dominio de factorización única, por tanto por el Teorema 26 es completamente enteramente cerrado.

Sea $\alpha \in L$ casi entero sobre $D[[x]]$; existe $h \in D[[x]]$ tal que $h\alpha^n \in D[[x]]$ para todo $n \geq 1$. Pero como $h, h\alpha^n \in K[[x]]$ y este es completamente enteramente cerrado, entonces $\alpha \in K[[x]]$. Es decir, podemos suponer $\alpha = \sum_{i=0}^{\infty} a_i x^i \in K[[x]] \cap L$, y se prueba que $\alpha \in D[[x]]$.

Existe un elemento no nulo $g \in D[[x]]$ tal que $g\alpha^n \in D[[x]]$ para cada entero positivo n .

Se prueba por inducción sobre i que $a_i \in D$. Así, si se asume que $a_0, \dots, a_{j-1} \in D$,

con j entero no negativo, entonces sea $\alpha^* = a_0 + \dots + a_{j-1}x^{j-1} \in D[[x]]$, y $g(\alpha - \alpha^*)^n \in D[[x]]$, para cada entero positivo n . Si $a_j = 0$, entonces $a_j \in D$; si $a_j \neq 0$, sea $g = \sum_{i=0}^{\infty} b_i x^i$, con b_m el primer coeficiente no nulo de g de tal forma que $g = \sum_{i=m}^{\infty} b_i x^i$.

Entonces si n es un entero positivo,

$$\begin{aligned} g(\alpha - \alpha^*)^n &= \left(\sum_{i=m}^{\infty} b_i x^i \right) \left(\sum_{i=j}^{\infty} a_i x^i \right)^n \\ &= b_m a_j^n x^{m+nj} + (\text{otros términos de orden superior}) \in D[[x]]. \end{aligned}$$

Se sigue que $b_m a_j^n \in D$ para cada entero positivo n , así que a_j es un elemento de K casi entero sobre D , y por tanto, por hipótesis es un elemento de D , así $\alpha = \sum_{i=0}^{\infty} a_i x^i \in D[[x]]$. ■

Teorema 4.3 [8] *Sea D es un dominio de integridad con cuerpo de fracciones K . Si $D[[x]]$ es enteramente cerrado, entonces D es enteramente cerrado.*

Demostración: Sea $\alpha \in K$ tal que α satisface un polinomio mónico con coeficientes en D , entonces $\alpha \in L$ y satisface un polinomio mónico con coeficientes en $D[[x]]$. Por la hipótesis, $\alpha \in D[[x]]$, esto es $\alpha \in K \cap D[[x]] = D$. ■

El siguiente teorema establece que el recíproco del teorema anterior no es siempre cierto.

Teorema 4.4 [8] *Sea D un dominio enteramente cerrado que no es completamente enteramente cerrado. Entonces, $D[[x]]$ no es enteramente cerrado.*

Demostración: Se tiene aplicando el Teorema 1.11 y la Proposición 1.12 de [24]. ■

D. D. Anderson muestra en su último artículo “*GCD domains and power series rings*” (2002), que si V es un dominio de valuación y $V[[x]]$ un dominio *GCD*, entonces V debe ser de dimensión de Krull 1 con grupo de valuaciones \mathbb{Z} ó \mathbb{R} . Además, da un ejemplo de un dominio de valuación W de dimensión de Krull 1 con grupo de valuaciones \mathbb{R} , para el cual $W[[x]]$ no es dominio *GCD*.

El siguiente teorema es útil para ilustrar ejemplos que muestran que la propiedad *GCD* no se preserva al anillo de series.

Teorema 4.5 [8] *Sea V un dominio de valuación no trivial en el cuerpo K , con ideal máxima M . Entonces:*

- (i) V es enteramente cerrado.
- (ii) V es Noetheriano, si y sólo si, V es de dimensión de Krull 1 y discreto.
- (iii) V es completamente enteramente cerrado, si y sólo si, V es de dimensión de Krull 1.

Demostración: (i) Si $x \in K - V$, entonces $V \subset \langle x \rangle$, así que $V \subset \langle x \rangle \subset \langle x^2 \rangle \subset \dots \subset \langle x^n \rangle$ para cada $n \in \mathbb{Z}^+$. Así, si $\{d_i\}_{i=0}^{n-1}$ es un conjunto finito de elementos de V , entonces

$d_0 + d_1x + \dots + d_{n-1}x^{n-1} \in \langle x^{n-1} \rangle$, así que $x^n \neq d_0 + d_1x + \dots + d_{n-1}x^{n-1}$. Por tanto, x no es entero sobre V . Se sigue que V es enteramente cerrado.

(ii) Para la implicación inversa, se asume que V es 1-dimensional y discreto. Entonces $\{M^k\}_{k=1}$ es el conjunto de ideales propios de V . Por tanto, si $x \in M - M^2$, entonces $M^2 \subset \langle x \rangle \subseteq M$, y como consecuencia, $M = \langle x \rangle$. Se sigue que cada ideal de V es principal, y V es Noetheriano.

Recíprocamente, si V es Noetheriano, V es un dominio de ideales principales. En consecuencia, V es 1-dimensional y contiene ideales propios no idempotentes, así V es discreto.

(iii) Para la implicación directa, sea V completamente enteramente cerrado y sea P un ideal primo de V contenido propiamente en M . Si $x \in M - P$, entonces $P \subset \bigcap_{n=1}^{\infty} \langle x^n \rangle$. Pero el Teorema 1.11 de [24] muestra que $\bigcap_{n=1}^{\infty} \langle x^n \rangle = \langle 0 \rangle$. Así, $P = \langle 0 \rangle$ y V es 1-dimensional.

Recíprocamente, sea $x \in K - V$. Entonces $y = x^{-1} \in V$. Por el Teorema 1.11 de [24] $\bigcap_{n=1}^{\infty} \langle y^n \rangle$ es primo en V . Y como V es 1-dimensional, $\bigcap_{n=1}^{\infty} \langle y^n \rangle = \langle 0 \rangle$. Por tanto, si $d \neq 0$, $d \in V$, entonces $\langle y^n \rangle \subset \langle d \rangle$ para algún $n \in \mathbb{Z}^+$. Se sigue que $\frac{d}{y^n} = dx^n \notin V$, y como d es un elemento arbitrario de V , se tiene que x no es casi entero sobre V . Luego, V es completamente enteramente cerrado. ■

Corolario 4.6 *Sea V un dominio de valuación de dimensión > 1 , entonces V no es completamente enteramente cerrado, y por tanto $V[[x]]$ no es *G - GCD*.*

Demostración: Se tiene como consecuencia del teorema anterior y el Teorema 29. ■

A continuación se presenta un ejemplo de un dominio de valuación y por tanto dominio *GCD* (ver Proposición 8), cuyo anillo de series no es dominio *GCD*.

Ejemplo 4.7 Sea $A = K + xK[[x]]$ y $V = R + xK[[x]]$ un subanillo de A , con K el cuerpo de fracciones de R $xK[[x]] = M$ el ideal máximo de A y $R \cap M = 0$.

Se tiene entonces que:

(i) Si R es un dominio de valuación y en consecuencia GCD , entonces $V = R + xK[[x]]$ es dominio de valuación:

Se consideran en $R + M$ los ideales $\langle a + m \rangle$ y $\langle b + n \rangle$, con $a, b \in R$ y $m, n \in M$; se puede suponer en R que $\langle a \rangle \subseteq \langle b \rangle$, es decir, $a = rb$. Si $b = 0$, $a = 0$ y $\langle m \rangle, \langle n \rangle$ son comparables en A ; se supone ahora que $\langle m \rangle \supseteq \langle n \rangle$, entonces existe $v \in A$, $v = k + m'$ tal que $n = vm = (k + m')m$. Como $R \in DV$, $k \in R$ ó $k^{-1} \in R$. Si $k \in R$, entonces en $R + M$ se tiene que $\langle n \rangle \subseteq \langle m \rangle$. Si $k^{-1} \in R$, $\frac{1}{r} = k$, y $n = (\frac{1}{r} + m')m$, con lo cual $rn = (1 + rm')m$; además, $1 + rm' \in U(A)$ con lo que en A se tiene $\langle n \rangle = \langle rn \rangle = \langle m \rangle$. Entonces, existe $(k'' + m'') \in A$ tal que $m = (k'' + m'')(\frac{1}{r} + m')m$. Si $m = 0$, $n = vm = 0$ y entonces $\langle a + m \rangle \subseteq \langle b + n \rangle$. Si $m \neq 0$, $(k'' + m'')(\frac{1}{r} + m') = 1$ y se sigue que $k'' = r$ ya que $K \cap M = 0$, y entonces $m = (r + m'')n$ con lo que $\langle m \rangle \subseteq \langle n \rangle$.

Si $b \neq 0$, $b + n \neq 0$ porque $R \cap M = 0$ y $a + m = rb + m$. Se define en K , $z = \frac{m - rn}{b + n}$, entonces $z \in A$ ó $z^{-1} \in A$. Si $z^{-1} \in A$, como $m - rn \in M$, $z^{-1}(m - rn) \in M$ y por lo tanto $b + n \in M$ de lo que $b = 0$, una contradicción. Entonces, $z \in A$ y $z(b + n) = m - rn \in M$ y como $b + n \notin M$ entonces $z \in M$ y $a + m = (r + z)(b + n)$, es decir, $\langle a + m \rangle \subseteq \langle b + n \rangle$.

(ii) Sea $V = \mathbb{Z}_{(p)} + x\mathbb{Q}[[x]]$, entonces su dimensión de Krull es ≥ 2 , con p un número primo.

(iii) Según el Teorema 30, $V = \mathbb{Z}_{(p)} + x\mathbb{Q}[[x]]$ no es completamente enteramente cerrado y aplicando el Corolario 31, $V[[x]]$ no es $G - GCD$.

El siguiente ejemplo, similar al anterior, muestra que un anillo de series sobre un dominio Bézout no es necesariamente $G - GCD$.

Ejemplo 4.8 Sea R un dominio Prüfer, $K \neq R$ su cuerpo de fracciones. Sea $A = K + xK[[x]]$, $T = R + xK[[x]]$ un subanillo de A , $xK[[x]] = M$ el ideal máximo de A y $R \cap M = 0$.

Se tiene entonces que:

(i) Si R es un dominio Bézout y en consecuencia GCD , entonces $T = R + xK[[x]]$ es dominio Bézout:

Sea $J = \langle a_1 + m_1, a_2 + m_2 \rangle$ un ideal de $R + M$, así el ideal $I = \langle a_1, a_2 \rangle = \langle b \rangle$. Si $b = 0$, $J = \langle m_1, m_2 \rangle$ es un ideal del dominio de valuación $K[[x]]$

con lo cual J resulta principal generado por m_1 ó m_2 . Si $b \neq 0$, $a_i = c_i b$, $i = 1, 2$, luego $a_i + m_i = b(c_i + \frac{1}{b}m_i)$ con $\frac{1}{b}m_i \in M$, o sea en $R + M$ se tiene que $J \subseteq \langle b \rangle$; nótese que $b \in J$: existen $d_1, d_2 \in R$ tales que $b = d_1 a_1 + d_2 a_2$. Se supone, sin pérdida de generalidad que $a_1 \neq 0$, entonces $a_1 + m_1 \notin M$ y por tanto $a_1 + m_1 \in U(K[[x]])$, luego $(d_1 m_1 + d_2 m_2)(a_1 + m_1)^{-1} \in K + M$ y existe $k + n \in K + M$ tal que $(a_1 + m_1)(k + n) = (d_1 m_1 + d_2 m_2)$. Pero $K \cap M = 0$, con lo cual $k = 0$ y $(a_1 + m_1)(n) = d_1 m_1 + d_2 m_2$. Por lo tanto, $b = (d_1 - n)(a_1 + m_1) + d_2(a_2 + m_2)$, es decir, $b \in J$ de lo cual $J = \langle b \rangle$.

(ii) Sea $T = \mathbb{Z} + x\mathbb{Q}[[x]]$ un dominio Bézout de dimensión de Krull ≥ 2 .

(iii) Según el Teorema 30, $T = \mathbb{Z} + x\mathbb{Q}[[x]]$ no es completamente enteramente cerrado y aplicando el Corolario 31, $T[[x]]$ no es $G - GCD$.

Quedan planteados varios interrogantes como:

¿Bajo qué condiciones $R[[x]]$ es dominio GCD ?

En el caso del anillo de series cuando R es enteramente cerrado pero no completamente enteramente cerrado, el Corolario 31 permite concluir que su anillo de series no es dominio $G - GCD$. Así queda la siguiente conjetura: Si R es un dominio completamente enteramente cerrado y GCD , entonces $R[[x]]$ es un dominio GCD .

Referencias

- [1] D.D. Anderson, *GCD Domains, Gauss' Lemma, and contents of polynomials*, Kluwer Math And Appl. Vol 520, (2000), 1-31
- [2] D.D. Anderson And R.O. Quintero, *Some generalizations of GCD domains*, Lecture Notes in pure And Applied Math # 189, Marcel Dekker, (1997), 189-195.
- [3] D.D. Anderson And B.G. Kang, *Content formulas for polynomials and power series and completed integral closure*, J. Algebra 181 (1996), 82-94.
- [4] M.F. Atiyah, I.G. Macdonald, *Introducción al Álgebra Conmutativa*, Ed. Reverté S.A, Barcelona, 1978.
- [5] N. Bourbaki, *Commutative Algebra*, Addison Wesley, Massachusetts, 1972.
- [6] J. Brewer, *Power series over commutative rings*, Marcel Dekker, New York, 1981.

- [7] M. Fontana, J. A. Huckaba and I. Papick, *Prüfer Domains*, Marcel Dekker, New York, 1997.
- [8] R. Gilmer, *Multiplicative Ideal Theory*, Marcel Dekker, New York, 1972.
- [9] S. Glaz, *Finite conductor rings*, Proc. of AMS, to appear.
- [10] S. Glaz, *Finite conductor rings with zero-divisors*, Kluwer Math. And Appl. Vol 520 (2000), 251-271
- [11] S. Glaz, and S. Chapman, *One hundred problems in commutative rings theory*, Kluwer Math. and Appl. Vol 520 (2000), 459-477.
- [12] S. Glaz, *Commutative ring characterized by homological conditions*, Handbook to the Heart of Algebra, Kluwer Pub, to appear.
- [13] S. Glaz, *Finite conductor properties of $R(x)$ and $R\langle x \rangle$* , Dekker, Proc. of Huckaba Retirement Conference, Missouri (2000), to appear.
- [14] S. Glaz, *Commutative Coherent Rings*, Lecture Notes in Mathematics, No. 1371, Springer-Verlag, New York, 1989.
- [15] N. González, *Anillo de Polinomios Sobre Anillos de Conductor Finito*, Tesis de Maestría, Universidad Nacional de Colombia, Bogotá, 2002.
- [16] J. A. Huckaba, *Commutative Rings With Zero Divisors*, Marcel Dekker, New York, 1988.
- [17] I. Kaplansky, *Commutative Rings*, Allyn- Bacon, Boston, 1970.
- [18] E. Kunz, *Introduction to Commutative Algebra and Algebraic Geometry*, Birkhauser Boston, 1985.
- [19] O. Lezama y G. Rodriguez, *Anillos Modulos y Categorías*, Universidad Nacional de Colombia, Bogotá, 1994.
- [20] H. Matsumura, *Commutative Algebra*, W.A. Benjamin, Inc, New York, 1970.
- [21] G. Picavet, *About GCD domains*, to appear in Advances in Commutative Ring Theory, Marcel Dekker, (2001).
- [22] S. Rocuts, *Dominios de Prüfer: Caracterizaciones Subclases y Ejemplos*, Tesis de Grado, Universidad Nacional de Colombia, Bogotá, 1999.
- [23] J.J. Rotman, *An Introduction to Homological Algebra*, Academic Press, Orlando, 1979.

-
- [24] O. Sepúlveda, *Anillo de Series Formales Sobre Anillos Conductor Finito*, Tesis de Maestría, Universidad Nacional de Colombia, Bogotá, 2002.
- [25] K. Spindler, *Abstract Algebra with Applications*, Vol I and II, Marcel Dekker, New York, 1994.
- [26] W. Vasconcelos, *The Rings of Dimension Two*, Lecture Notes in Pure and Appl. Math. No. 22, Marcel Dekker, New York, 1976.
- [27] O. Zariski and P. Samuel, *Commutative Algebra*, Vol. I, D. Van Nostrand and Company, New Jersey, 1958.

NELSY GONZÁLEZ & OMAIDA SEPÚLVEDA
UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA
CAMPUS UNIVERSITARIO, DUITAMA - BOYACÁ, COLOMBIA
ngonzalez@duitama.uptc.edu.co omaidasd@latinmail.com

OSWALDO LEZAMA
UNIVERSIDAD NACIONAL DE COLOMBIA
CAMPUS UNIVERSITARIO, BOGOTÁ, COLOMBIA
olezama@matematicas.unal.edu.co
<http://www.virtual.unal.edu.co/cursos/ciencias/15930/olezama/>

Simetría y nuevas soluciones de la ecuación de Black Scholes

Nikolay Sukhomlin

Resumen

En el enfoque local estudiamos la ecuación de Black Scholes, construimos sus operadores de simetría, mostramos que su función de Green y otras soluciones usadas en la práctica financiera son las funciones propias de un operador de simetría (o de la combinación lineal de ellas). Este hecho sugiere la importancia del estudio de la simetría del “master equation” del modelo dinámico. Introducimos el concepto de leyes de conservación en el modelo de Black Scholes, las cuales en los casos sencillos se manifiestan por la existencia de relaciones estables durante la evolución del proceso entre el valor de una opción y la rapidez de su cambio. También encontramos el grupo de equivalencia de la ecuación de Black Scholes lo que permite clasificar los operadores de simetría diferenciales hasta tercer orden, realizar la separación de variables y encontrar varias clases de nuevas soluciones de esta ecuación.

Introducción

El modelo de Black Scholes es el más usado en la teoría de los mercados financieros. La ecuación de Black Scholes constituye la parte principal del modelo, véase por ejemplo [1]:

$$AV \equiv \left\{ \frac{\partial}{\partial t} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2}{\partial x^2} + r x \frac{\partial}{\partial x} - r \right\} V(t, x) = 0 \quad (1)$$

donde $V(t, x)$ es el valor de la opción de comprar o de vender un bien financiero con el precio x en el momento t . En la teoría de finanzas la constante real σ ($\sigma \neq 0$) se llama volatilidad; la constante r define el riesgo de la inversión. En el modelo de Black Scholes la ecuación (1) se deduce a partir de un proceso estocástico de tipo de difusión. Se conoce que la transformación

$$\tau = -\sigma^2 t + \text{const}, \quad \xi = \ln x, \quad (2 \text{ a})$$

$$V(t, x) = U(\tau, \xi) \exp \{ \beta \ln x - (\beta - 1)^2 t / 2 \} \quad (2 \text{ b})$$

$$\beta \equiv \frac{1}{2} - \frac{r}{\sigma^2} \quad (2 \text{ c})$$

reduce la ecuación (1) a la ecuación de difusión:

$$-\frac{\partial U}{\partial \tau} + \frac{1}{2} \frac{\partial^2 U}{\partial \xi^2} = 0 \quad (2 \text{ d})$$

En la teoría de procesos estocásticos la ecuación de Black Scholes (1) corresponde a la ecuación retrógrada de Kolmogorov adonde la difusión se realiza "hacia el pasado". Actualmente la ecuación de Black Scholes se estudia activamente en diferentes aspectos, véase por ejemplo [2-7].

La estructura del artículo es la siguiente: en la sección 1 establecemos los operadores de simetría de primer orden, notamos que ellos crean un álgebra de Lie, constatamos la estructura especial del operador A y encontramos las funciones propias de dos tipos de estos operadores. También introducimos el concepto de *ley de conservación en el modelo de Black Scholes*. En la sección 2 construimos el grupo de equivalencia de dicha ecuación como el producto cartesiano de dos subgrupos: uno es continuo y otro discreto. En la sección 3 clasificamos los operadores de simetría hasta el tercer orden relacionado al grupo de equivalencia. En la sección 4, al separar las variables, encontramos los representantes más sencillos de las soluciones correspondientes a todas las cinco clases de los operadores de simetría de segundo orden y construimos algunas otras nuevas soluciones que entran en estas clases.

1 Simetría de la ecuación de Black Scholes y Leyes de conservación

La idea básica de nuestro enfoque consiste en la búsqueda y en el estudio de ciertas relaciones entre el valor de una opción $V(t, x)$ y la rapidez de su cambio relacionado con el precio. El hecho de la existencia de relaciones que se conservan durante todo el proceso modelado, nos lleva a información práctica muy importante. Por ejemplo, este nos permite evaluar la volatilidad, o sugiere la formulación apropiada de los problemas de frontera.

En la teoría de las ecuaciones diferenciales parciales el primer paso del programa de resolución consiste en el estudio del álgebra de simetría de la ecuación. Este estudio puede tener varios objetivos: con la finalidad de la separación de variables [8, 9] o para la aplicación de la teoría de Lie [4, 7, 10-12]. Nosotros seguimos la primera orientación.

La simetría de la ecuación de difusión es bien conocida. Dado que la ecuación de Black Scholes se puede reducir a la ecuación de difusión, su simetría es similar. Sin hacer los cálculos, pasamos directamente a los resultados.

El conjunto de los operadores de simetría de la ecuación (1) diferenciales lineales de primer orden y que conmutan con el operador A constituye un espacio vectorial con la base:

$$B_1 = x \frac{\partial}{\partial x} - \beta, \quad (3 \text{ a})$$

$$B_2 = \sigma^2 t x \frac{\partial}{\partial x} - \ln x - \sigma^2 t \beta \quad (3 \text{ b})$$

y también un operador de simetría trivial del orden cero: $B_3 = 1$. Observamos que:

$$B_2 = \sigma^2 t B_1 - \ln x \quad (3 \text{ c})$$

y que se verifican las condiciones de conmutación:

$$[B_2, B_1] = B_3, \quad [B_2, B_3] = [B_1, B_3] = 0 \quad (3 \text{ d})$$

La simetría mencionada es intrínseca de la ecuación de Black Scholes (1). Es fácil verificar que el operador (3a) define totalmente el operador A :

$$A = \sigma^2 \left\{ \frac{1}{\sigma^2} \frac{\partial}{\partial t} + \frac{1}{2} B_1^2 - \frac{1}{2} (\beta - 1)^2 \right\}. \quad (3 \text{ e})$$

Visto que $[A, B_i] = 0$ ($i = 1, 2, 3$), es obvio que toda función de los operadores B_1, B_2, B_3 conmutará con A y en consecuencia, será un operador de simetría de la ecuación (1). Este hecho nos permitirá en las secciones 3 y 4 clasificar los operadores de simetría y construir las nuevas soluciones exactas de la ecuación de Black Scholes usando el metodo de separacion de variables en el enfoque de V. Shapovalov [8].

En las teorías económicas, sociales y financieras no se suele usar la noción de operador de simetría. Sin embargo la importancia de este concepto reside en el hecho de que cada operador de simetría está ligado con una ley de conservación, pero el concepto de ley de conservación tampoco está desarrollado en dichas teorías. Como en la mecánica clásica y la cuántica, aquí este concepto tiene un papel importante en la descripción del sistema dinámico. Vamos introducir el concepto de ley de conservación y lo ilustramos para el modelo de Black Scholes.

Hallamos las soluciones de la ecuación (1) que son las funciones propias del operador de simetría (3 a), es decir resolvemos el sistema:

$$A V(t, x) = 0, \quad (4 \text{ a})$$

$$B V(t, x) = \lambda V(t, x). \quad (4 \text{ b})$$

con $B = B_1$ en este caso. Después de cálculos sencillos encontramos una familia de soluciones parametrizadas por λ :

$$V_\lambda(t, x) = C x^{\lambda+\beta} \exp \left\{ \frac{\sigma^2}{2} [(\beta - 1)^2 - \lambda^2] t \right\} \quad (C = \text{const}) \quad (5)$$

Esta familia de soluciones corresponde a la ley de conservación que se expresa por la relación (4 b): elasticidad-precio del valor de la opción que se conserva en este caso:

$$E_p \equiv \frac{x}{V} \frac{\partial V}{\partial x} = (\lambda + \beta) = \text{const.} \quad (6)$$

Este hecho significa que al elegir la solución (5), elegimos la relación (6) entre el valor V y la rapidez de su variación. La existencia de esta ley de conservación permite experimentalmente evaluar la potencia del precio en la función (5): $\lambda + \beta$.

En el caso general las relaciones entre la función incógnita y sus derivadas pueden ser mucho menos evidentes: véase por ejemplo las formulas (7), (14) y (15). La importancia práctica de una ley de conservación se encuentra en el hecho de que matemáticamente es mucho más fácil encontrarla que resolver la ecuación (1). Además, la existencia de una ley de conservación no depende del sistema de coordenadas en las cuales está escrita la ecuación. Nuestro estudio permite explicitar la simetría que verifica la solución (5) y establecer el sentido de las constantes que la definen. Esta función es cóncava a la condición: $0 < \lambda + \beta < 1$.

El segundo operador de simetría de primer orden B_2 define otra ley de conservación: la relación (4 b) informa que la misma elasticidad ahora no se conserva, sino depende linealmente del logaritmo del precio y es inversamente proporcional al tiempo. En consecuencia se conserva la expresión:

$$\sigma^2 t \left[\frac{x}{V} \frac{\partial V}{\partial x} - \beta \right] - \ln x = \lambda = \text{const.} \quad (7)$$

Esta igualdad se construye usando la formula (3 c). También es posible interpretar esta ley como conservación del valor inicial del precio x_0 : para el tiempo $t \rightarrow 0$ el operador B_2 se reduce a la expresión: $-(\ln x_0) \Rightarrow x_0 = \exp\{-\lambda\}$. Ahora la relación (7) se puede escribir:

$$E_p \equiv \frac{x}{V} \frac{\partial V}{\partial x} = \frac{1}{\sigma^2 t} \ln \left(\frac{x}{x_0} \right) + \beta.$$

La resolución del sistema de tipo (4) con el operador B_2 (3 b) nos lleva al resultado siguiente:

$$V_\lambda(t, x) = \frac{C}{\sqrt{t}} \exp \left\{ \frac{[\ln x + \sigma^2 \beta t + \lambda]^2}{2\sigma^2 t} \right\} \quad (C = \text{const}). \quad (8)$$

Esta función representa la función de Green de la ecuación de Black Scholes (1) bien conocida en la literatura. Mencionamos que si $r = \sigma^2/2$ entonces la solución (8) queda:

$$V_\lambda(t, x) = \frac{C}{\sqrt{t}} \left(\frac{x}{x_0} \right)^{\frac{1}{2\sigma^2 t} \ln \left(\frac{x}{x_0} \right)}.$$

Podemos constatar que la función (8) es cóncava con la condición:

$$\ln(x/x_0) - rt^2 + 1 < \left(\frac{\sigma^2 t}{2} - 1 \right),$$

lo que se verifica a partir del momento del tiempo $t = 4/\sigma^2$.

2 Grupo de equivalencia de la ecuación de Black Scholes

Sea la ecuación de Black Scholes (1). Estudiamos las transformaciones de variables:

$$\tau = \tau(t), \quad \xi = \xi(t, x), \quad V(t, x) = a(t, x) Q(\tau, \xi) \quad (9)$$

tales que la ecuación (1) no cambia su estructura, salvo que aparece un factor multiplicativo a la izquierda:

$$f(t, x) \left\{ \frac{\partial Q}{\partial \tau} + \frac{1}{2} \sigma_0^2 \xi^2 \frac{\partial^2 Q}{\partial \xi^2} + r_0 \xi \frac{\partial Q}{\partial \xi} - r_0 Q \right\} = 0, \quad (10)$$

donde Q es la nueva incógnita. Así aparece una libertad complementaria: considerar las constantes σ_0, r_0 como distintas de las constantes σ, r o como iguales. El problema de encontrar todas las transformaciones con la propiedad mencionada se resuelve por el Teorema 1 y para prepararlo enunciamos dos proposiciones.

Proposición 1. *Sea la ecuación de Black Scholes (1). La transformación siguiente:*

$$\tau = \frac{\alpha^2 \sigma^2}{\sigma_0^2} (t - t_0), \quad \xi = \left(\frac{x}{x_0} \right), \quad (11 \text{ a})$$

$$V(t, x) = Q(\tau, \xi) \exp\{(\beta - \alpha \beta_0) \ln x + \frac{\sigma^2}{2} [(\beta - 1)^2 - \alpha^2 (\beta_0 - 1)^2] t\} \quad (11 \text{ b})$$

convierte la ecuación (1) en la ecuación de tipo (10) con el factor exterior:

$$f(t, x) = \frac{\alpha^2 \sigma^2}{\sigma_0^2} \exp\{(\beta - \alpha \beta_0) \ln x + \frac{\sigma^2}{2} [(\beta - 1)^2 - \alpha^2 (\beta_0 - 1)^2] t\}$$

para cualesquiera valores de las constantes: α, t_0, x_0 ($\alpha \neq 0, x_0 \neq 0$). Aquí la constante β está definida por (2 c) y $\beta_0 \equiv \frac{1}{2} - \frac{r_0}{\sigma_0^2}$.

Las transformaciones (11) constituyen un grupo continuo con tres parámetros: la constante α corresponde a un cambio en la escala del tiempo con la variación simultánea de la coordenada x ; la constante t_0 representa la libertad de elección del origen del tiempo y la constante x_0 está ligada con la libertad de elección del precio inicial (notamos la imposición: $x_0 \neq 0$).

Llamamos este grupo galileano G por analogía con el grupo correspondiente en la teoría de difusión. El grupo galileano G crea una relación de equivalencia sobre el conjunto de las soluciones de la ecuación (1). Además de este grupo, existe una transformación especial que tampoco cambia la estructura de la ecuación (1).

Proposición 2. *Sea la ecuación de Black Scholes (1). La transformación siguiente:*

$$\tau = -\frac{1}{\sigma_0^2 \sigma^2} \frac{1}{t}, \quad \xi = \exp \left\{ -\frac{\ln x}{\sigma^2 t} \right\}, \quad (12 \text{ a})$$

$$V(t, x) = \frac{Q(\tau, \xi)}{\sqrt{t}} \exp \left\{ \frac{(\ln x + \beta_0)^2}{2\sigma^2 t} + \beta \ln x + \frac{\sigma^2(\beta-1)^2}{2} t + \frac{r_0}{\sigma_0^2 \sigma^2} \frac{1}{t} \right\} \quad (12 \text{ b})$$

convierte la ecuación (1) en la ecuación de tipo (10) con el factor exterior:

$$f(t, x) = \frac{1}{\sigma_0^2 \sigma^2 t^2 \sqrt{t}} \exp \left\{ \frac{(\ln x + \beta_0)^2}{2\sigma^2 t} + \beta \ln x + \frac{\sigma^2(\beta-1)^2}{2} t + \frac{r_0}{\sigma_0^2 \sigma^2} \frac{1}{t} \right\}.$$

Ambas proposiciones se comprueban fácilmente por la sustitución. En realidad las constantes en la fórmula (12 a) se pueden eliminar por una transformación del grupo galileano G : por ejemplo, usando la transformación (11), la ecuación (1) se puede reducir a otra de la misma estructura con $\sigma_0 = 1$ y la constante r_0 puede ser igual a cero o a una constante cualquiera. En este caso la transformación (12 a) se presenta por:

$$x^I = \exp \left\{ -\frac{\ln x}{t} \right\}, \quad t^I = -\frac{1}{t}. \quad (12 \text{ c})$$

La denotamos ν . Estudiamos ahora las potencias de ν :

$$\nu^2 \text{ corresponde a } x^{II} = \exp \left\{ -\frac{\ln x^I}{t^I} \right\} = \frac{1}{x}, \quad t^{II} = -\frac{1}{t^I} = t;$$

$$\nu^3 \text{ corresponde a } x^{III} = \exp \left\{ \frac{\ln x}{t} \right\}, \quad t^{III} = -\frac{1}{t};$$

$$\nu^4 \text{ corresponde a } x^{IV} = x, \quad t^{IV} = t: \text{ transformación idéntica } i.$$

Concluimos que el conjunto $\{\nu, \nu^2, \nu^3, \nu^4 = i\}$ constituye un grupo discreto que denotamos N .

Teorema 1. *Sea la ecuación de Black Scholes (1) con las constantes σ ($\sigma \neq 0$), r dadas. El grupo de equivalencia más amplio que admite la ecuación (1) es $\Gamma = G \otimes N$ donde G es el grupo galileano y N es el grupo discreto definidos arriba por las fórmulas (11) y (12).*

El teorema se verifica sin dificultad por la sustitución de la transformación (9) en la ecuación (1) y la imposición de la estructura (10). Notamos que el grupo galileano G (y en consecuencia el grupo Γ) es parametrizado por tres parámetros α, t_0, x_0 ($\alpha \neq 0, x_0 \neq 0$). Así se crea una partición sobre el conjunto de todas las soluciones de la ecuación (1). Observamos que al mismo tiempo se crea una clasificación de las leyes de conservación.

Es fácil verificar que el operador (3 a) se transforma por el elemento ν del grupo N en el operador (3 b) y viceversa, lo que significa que las soluciones (5) y (8) entran en la misma clase de equivalencia. Más exactamente, la aplicación de la transformación (9) en la forma (12) con el factor definido por la relación (12 b) resulta:

$$a^{-1} B_1 a = B_2'; \quad a^{-1} B_2 a = -B_1'; \quad (13)$$

donde los operadores B'_1, B'_2 tienen las formas (3 a) y (3 b) escritas en las coordenadas τ, ξ (12 a) y con σ_0, β_0 . Recordamos que los operadores B_1 y B_2 no conmutan: véase la formula (3 d); entonces es normal que se pongan en relación de equivalencia sólo por los elementos del grupo discreto N y no se puedan poner en equivalencia por los elementos del grupo continuo G .

3 Clasificación de los operadores de simetría de la ecuación de Black Scholes hasta el tercer orden

En nuestro artículo [13] hicimos la clasificación de los operadores de simetría hasta tercer orden para la ecuación de Schrödinger libre. La misma clasificación relacionada con el grupo de equivalencia Γ del Teorema 1 se puede hacer para la ecuación de difusión (2 d) y en consecuencia para la ecuación de Black Scholes (1). Los resultados se presentan en el teorema siguiente.

Teorema 2. *Sea la ecuación de Black Scholes (1) con el operador A por ejemplo en la forma (3 e). Todos los operadores lineales diferenciales de hasta tercer orden que conmutan con A entran en una de las siguientes 6 clases de equivalencia y en 8 agrupaciones de clases de equivalencia:*

1) *Todos los operadores de primer orden constituyen una sola clase de equivalencia cuyo representante más sencillo es B_1 (ó B_2) definido por la fórmula (3 a).*

2) *Todos los operadores de segundo orden constituyen cinco clases de equivalencia cuyos representantes más sencillos son:*

$$\text{a) } B_2^2 + B_1^2, \quad (14 \text{ a})$$

$$\text{b) } B_2^2 - B_1^2, \quad (14 \text{ b})$$

$$\text{c) } B_1 B_2 + B_2 B_1, \quad (14 \text{ c})$$

$$\text{d) } B_1^2 + B_2, \quad (14 \text{ d})$$

$$\text{e) } B_1^2. \quad (14 \text{ e})$$

3) *Todos los operadores de tercer orden constituyen ocho agrupaciones de clases de equivalencia parametrizadas por tres o cuatro constantes arbitrarias ω, μ, ν, η y cuyos representantes más sencillos son:*

$$1. \quad B_2^3 + (B_2 B_1^2 + B_1^2 B_2) + \omega B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 \text{ a})$$

$$2. \quad B_2^3 - (B_2 B_1^2 + B_1^2 B_2) + \omega B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 \text{ b})$$

$$3. \quad B_2^3 + B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 \text{ c})$$

$$4. \quad B_2^3 - B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 \text{ d})$$

$$5. \quad B_2^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 \text{ e})$$

$$6. (B_2^2 B_1 + B_1 B_2^2) + B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 f)$$

$$7. (B_2^2 B_1 + B_1 B_2^2) - B_1^3 + \mu B_1^2 + \nu B_2 + \eta B_1, \quad (15 g)$$

$$8. (B_2^2 B_1 + B_1 B_2^2) + \mu B_1^2 + \nu B_2 + \eta B_1. \quad (15 h)$$

La demostración del teorema similar para la ecuación de Schrodinger libre está hecha en nuestro artículo [13]. El Teorema 2 se comprueba de la misma manera. La idea de la demostración es la siguiente: se construye la forma cúbica de los operadores B_1, B_2, B_3 de estructura más general. Luego se aplican las transformaciones del grupo Γ para simplificar el operador inicial. El hecho que cualquier operador de simetría diferencial lineal de tercer orden de la ecuación de Black Scholes (1) tiene la estructura de esta forma cúbica es también nuevo.

Notamos que el Teorema 2 también dice que existen 6 clases y 8 agrupaciones de clases de las leyes de conservación, cada una está definida por la relación (4 b) y las fórmulas (14) y (15).

Como mencionamos arriba, cada operador de simetría está ligado con una ley de conservación específica. Además, a cada operador corresponde un conjunto de funciones propias que son las soluciones de la ecuación (1) y que constituyen una base en un espacio funcional. También cada operador de simetría corresponde a un sistema privilegiado de coordenadas que permite separar las variables en la ecuación (1) y resolverla exactamente. Realizamos este programa en la sección siguiente.

4 Nuevas soluciones exactas de la ecuación de Black Scholes

El Teorema 2 de la sección anterior nos permite abordar el problema de la resolución sistemática de la ecuación de Black Scholes (1). Algunas soluciones son muy conocidas como aquellas que presentamos en la sección 1; pero la mayoría de las soluciones que enumeramos abajo son nuevas. Recordamos que las soluciones con la simetría dada las buscamos a partir del sistema de tipo (4).

La solución (5) representa la solución más sencilla de la clase de soluciones equivalentes que admiten un operador de simetría de primer orden. Cualquier solución de esta clase se puede construir al desarrollar la función (5) usando todos los elementos del grupo Γ del Teorema 1. Visto los resultados de la fórmula (13) podemos constatar que la función de Green (8) entra en la misma clase de equivalencia.

En nuestro artículo [14] encontramos las soluciones de la ecuación de Schrödinger libre correspondientes a las cinco clases de operadores de segundo orden correspondientes a las fórmulas (14) arriba. Usando las relaciones de similitud entre esta última ecuación, ecuación de difusión (2 d) y la ecuación de Black

Scholes (1) podemos construir por el mismo procedimiento las soluciones de cada una de las cinco clases de equivalencia de la ecuación de Black Scholes. Todos estos resultados son nuevos.

Conforme al Teorema 2 el conjunto de las soluciones de la ecuación (1) que verifican el sistema (4) con los operadores de simetría de segundo orden se separa en 5 clases de equivalencia. Los operadores más sencillos de cada clase de equivalencia son enumerados en las fórmulas de (14 a) hasta (14 e). Pasemos ahora a la resolución del sistema (4) para cada una de estas cinco clases.

a) Sea la clase de equivalencia definida por el operador (14 a):

$$B = B_2^2 + B_1^2 = (\sigma^4 t^2 + 1) B_1^2 - 2\sigma^2 t (\ln x) B_1 + \ln^2 x - \sigma^2 t.$$

Es fácil verificar que la función siguiente cumple el sistema (4) y en consecuencia es la solución de la ecuación de Black Scholes:

$$\begin{aligned} V_\lambda(t, x) &= \frac{\Phi(\xi)}{(\sigma^4 t^2 + 1)^{1/4}} \exp\left\{\frac{\sigma^2 t}{2(\sigma^4 t^2 + 1)} \ln^2 x + \right. \\ &+ \beta \ln x - \frac{\lambda}{2} \operatorname{arctg}(\sigma^2 t) + \frac{\sigma^2}{2} (\beta - 1)^2 t \left. \right\}, \quad (16) \\ \xi &\equiv (\sigma^4 t^2 + 1)^{-1/2} \ln x. \end{aligned}$$

La función $\Phi(\xi)$ verifica la ecuación diferencial ordinaria siguiente:

$$\Phi'' + \xi^2 \Phi = \lambda \Phi,$$

cuyas soluciones son las funciones del cilindro parabólico. Las soluciones correspondientes a la familia de soluciones (16) describen en la mecánica cuántica los estados llamados coherentes.

b) Sea la clase definida por el operador (14 b):

$$B = B_2^2 - B_1^2 = (\sigma^4 t^2 - 1) B_1^2 - 2\sigma^2 t (\ln x) B_1 + \ln^2 x - \sigma^2 t.$$

En este caso la solución de la ecuación de Black Scholes se presenta en la forma:

$$\begin{aligned} V_\lambda(t, x) &= \frac{\Phi(\xi)}{|\sigma^4 t^2 - 1|} \exp\left\{\frac{\sigma^2 t}{2(\sigma^4 t^2 - 1)} \ln^2 x + \right. \\ &+ \beta \ln x + \frac{\lambda}{2} \operatorname{arctg}(\sigma^2 t) + \frac{\sigma^2}{2} (\beta - 1)^2 t \left. \right\}, \quad (17) \\ \xi &\equiv |\sigma^4 t^2 - 1|^{-1/2} \ln x. \end{aligned}$$

La función $\Phi(\xi)$ verifica la ecuación diferencial ordinaria de tipo de oscilador armónico en la mecánica:

$$\Phi'' - \xi^2 \Phi = \lambda \Phi.$$

Las soluciones que se anulan al infinito se expresan por los polinomios de Tchebyshev - Hermite $H_n(\xi)$; el espectro de los valores propios del operador mencionado es discreto:

$$\Phi_n(\xi) = C_n H_n(\xi) \exp\{-\xi^2/2\}, \quad \lambda_n = -(2n+1), \quad n = 0, 1, 2, \dots$$

con las funciones $H_n(\xi)$ que verifican la ecuación de Hermite:

$$H_n'' - 2\xi H_n' + 2n H_n = 0.$$

Como se conoce las funciones $\Phi_n(\xi)$ representan una base ortonormal en L^2 . Las soluciones (17) se definen sobre un intervalo finito del tiempo y se pueden fácilmente adaptar al modelo de Black Scholes.

c) Sea la clase de equivalencia definida por el operador (14 c):

$$B = B_2 B_1 + B_2 B_1 = 2\sigma^2 t B_1^2 - 2(\ln x) B_1 - 1.$$

La solución de la ecuación de Black Scholes se presenta como:

$$\begin{aligned} V_\lambda(t, x) &= \Phi(\xi) \exp\left\{\frac{1}{4\sigma^2 t} \ln^2 x + \beta \ln x - \frac{\lambda+1}{4} \ln(\sigma^2 t) + \frac{\sigma^2}{2} (\beta-1)^2 t\right\}, \\ \xi &\equiv (\sigma^2 t)^{-1/2} \ln x, \end{aligned} \quad (18)$$

$$\Phi'' - (\xi^2/4)\Phi = (\lambda/2)\Phi.$$

Las funciones $\Phi(\xi)$ de la última ecuación como en el caso anterior se expresan por los polinomios de Tchebyshev - Hermite H_n ; el espectro de los valores propios del operador mencionado es también discreto:

$$\Phi_n(\xi) = C_n H_n(\xi/\sqrt{2}) \exp\{-\xi^2/4\}, \quad \lambda_n = -2n-1, \quad n = 0, 1, 2, \dots$$

Finalmente la solución (18) en este caso se presenta como:

$$V_\lambda(t, x) = C_n t^{\frac{n}{2}} H_n(\xi/\sqrt{2}) \exp\{\beta \ln x + \frac{\sigma^2}{2} (\beta-1)^2 t\} \quad (19)$$

Esta solución tiene una propiedad interesante: ella verifica una de las condiciones complementarias del modelo de Black Scholes (si $\beta > 0$):

$$\begin{aligned} x \rightarrow 0^+ &\Rightarrow V_n \sim x^\beta (\ln x)^n \Rightarrow \\ \lim_{x \rightarrow 0^+} (V_n) &= \lim_{x \rightarrow 0^+} (const x^\beta) = 0. \end{aligned}$$

Por otro lado constatamos que si $x \rightarrow \infty \Rightarrow V_n \sim x^\beta (\ln x)^n$. Esto permite aproximar bastante bien la otra condición de frontera si β es pequeña y construir una serie de tipo exponencial usando las soluciones (19). Más exactamente para $C_n = \sigma^n$:

$$V^\circ(t, x) \equiv \sum_{n=0}^{\infty} \frac{1}{n!} V_n(t, x) = x^{1+\beta} \exp\left\{\frac{\sigma^2}{2} (\beta-1)^2 t\right\} \quad (20)$$

Esta función es el caso particular de la solución (5) para $\lambda = 1$. En la aproximación $\beta \approx 0$, $\beta \neq 0$, la solución de la ecuación de Black Scholes que verifica las condiciones complementarias será la combinación lineal de (20) y de (19) para $n = 0$:

$$V(t, x) = x^\beta \exp \left\{ \frac{\sigma^2}{2} (\beta - 1)^2 (t - T) \right\} \left[x e^{-\frac{\sigma^2}{2} (t-T)} - K \right].$$

Precisamente la aproximación es en la presencia del factor $x^\beta \approx 1$ si $\beta \approx 0$. Por ejemplo, para $r = 0.05$; $\sigma = 0.317 \Rightarrow \beta = 0.002$.

d) Sea la clase de equivalencia definida por el operador (14 d): $B = B_1^2 + B_2$. La solución de la ecuación de Black Scholes en este caso tiene la forma siguiente:

$$V_\lambda(t, x) = \Phi(\xi) \exp \left\{ \left(\beta - \frac{\sigma^2 t}{2} \right) \ln x - \frac{\sigma^6 t^3}{12} + \frac{\sigma^2}{2} [(\beta - 1)^2 - \lambda] t \right\} \quad (21)$$

$$\xi \equiv \ln x + \sigma^4 t^2 / 4.$$

La función $\Phi(\xi)$ verifica la ecuación de Airy: $\Phi'' - \xi \Phi = \lambda \Phi$.

e) La última clase de equivalencia de los operadores de segundo orden tiene como operador más sencillo (14 e): $B = B_1^2$. Las soluciones de la ecuación de Black Scholes con $B = B_1^2$ para cualesquiera valores del parámetro λ son parecidas a las funciones (5) pero aquí dependen de dos constantes arbitrarias C , D (porque el operador de simetría en este caso es de segundo orden):

$$V_\lambda(t, x) = (C e^{\sqrt{\lambda} \ln x} + D e^{-\sqrt{\lambda} \ln x}) \exp \left\{ \beta \ln x + \frac{\sigma^2}{2} [(\beta - 1)^2 - \lambda] t \right\} \quad (22)$$

En conclusión notamos que se pueden construir otras soluciones de cada clase de equivalencia a partir de las soluciones mencionadas usando todos los elementos del grupo de equivalencia Γ del Teorema 1. Al aplicar, por ejemplo la transformación (12 c) del subgrupo discreto N a la última función (21), llegamos a nueva solución de la ecuación de Black Scholes (que queda sin embargo equivalente a la solución (22)):

$$V'_\lambda = \frac{C'}{\sqrt{t}} \exp \left\{ \frac{\lambda + 2\sqrt{\lambda} \ln x + \frac{(\ln x + \beta \sigma^2 t)^2 + 2r \sigma^2 t^2}{2\sigma^2 t}} \right\} +$$

$$+ \frac{D'}{\sqrt{t}} \exp \left\{ \frac{\lambda - 2\sqrt{\lambda} \ln x + \frac{(\ln x + \beta \sigma^2 t)^2 + 2r \sigma^2 t^2}{2\sigma^2 t}} \right\} \quad (23)$$

Aquí no utilizamos las primas arriba de las variables. Observamos que si por ejemplo $\lambda < 0$, las soluciones (22) y (23) contienen las funciones *senos* y *cosenos* del $\ln x$, lo que permite usar el desarrollo en la serie de Fourier para verificar las condiciones de frontera. Recordamos que nuestro enfoque es local y entonces el problema de condiciones de frontera necesita un paso más para su estudio.

También la nueva solución de la ecuación de Black Scholes que pertenece a la clase de equivalencia definida por la solución (19) es:

$$V'_n(t, x) = C'_n t^{-\frac{n+1}{2}} H_n(\rho) \exp\left\{\frac{\ln^2 x}{2\sigma^2 t} + \beta \ln x + \frac{\sigma^2}{2}(\beta - 1)^2 t\right\},$$

$$\rho \equiv \frac{\ln x}{\sqrt{-2\sigma^2 t}}.$$

La unidad imaginaria $\sqrt{-1}$ en la expresión ρ no influye en el resultado porque los polinomios de número par $H_{2k}(\rho)$ contienen únicamente los términos con potencias pares de ρ y por lo tanto son reales. En los polinomios $H_{2k+1}(\rho)$ todos los términos son de potencia impar y la unidad imaginaria se puede sacar como factor común, el que queda incluido en la constante C'_{2k+1} que será en este caso también imaginaria.

La solución "clásica" de la ecuación de Black Scholes es la función siguiente (véase por ejemplo [2], p. 17):

$$V(t, x) = x\Phi(\theta) - Ke^{-r(T-t)}\Phi(\omega) \quad (24)$$

con la función de distribución normal estandar:

$$\Phi(z) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du,$$

$$\theta \equiv \frac{\ln x - (1-\beta)\sigma^2 t + T\sigma^2(1-\beta) - \ln K}{\sigma\sqrt{T-t}},$$

$$\omega \equiv \frac{\ln x + \beta\sigma^2 t - (T\sigma^2\beta + \ln K)}{\sigma\sqrt{T-t}},$$

En realidad cada término de la fórmula (24) representa una solución de la ecuación de Black Scholes. Dichos términos tienen la simetría superior: el operador de simetría correspondiente es un operador integral. Observando que las funciones x , y $e^{-r(T-t)}$ son las soluciones particulares de la ecuación de Black Scholes (1), constatamos que cada uno de los términos de la solución (24) represente un producto de dicha solución particular por la función $\Phi(z)$ que en consecuencia verifica la ecuación diferencial ordinaria: $\Phi'' + z\Phi' = 0$.

La nueva solución de la ecuación de Black Scholes que es equivalente a (24) relacionada con el grupo discreto N es la función siguiente:

$$V'(t, x) = V_1 - V_2 \quad (25)$$

donde cada término es también una solución de la ecuación de Black Scholes:

$$V_1 \equiv \frac{\Phi(\nu(t, x))}{\sqrt{\sigma^2 t}} \exp\left\{\frac{\ln^2 x}{2\sigma^2 t} + \left(\beta + \frac{1-b}{\sigma^2 t}\right) \ln x + \frac{(b-1)^2}{2\sigma^2 t} + \left(r + \frac{\sigma^2 \beta^2}{2}\right) t\right\},$$

$$\nu(t, x) \equiv \frac{\ln x + (1-b) + [T\sigma^2(1-b) - \ln K]\sigma^2 t}{\sqrt{\sigma^2 t(T\sigma^2 t + 1)}},$$

$$V_2 \equiv \frac{Ke^{-\rho T} \exp((2b-1)/2\sigma^2 t)}{\sqrt{\sigma^2 t}} \Phi(\mu(t, x)) \exp\left\{\frac{\ln^2 x}{2\sigma^2 t} + \right.$$

$$+ \left(\beta - \frac{b}{\sigma^2 t} \right) \ln x + \frac{(b-1)^2}{2\sigma^2 t} + \left(r + \frac{\sigma^2 \beta^2}{2} \right) t \},$$

$$\mu(t, x) \equiv \frac{\ln x - b - [T\sigma^2 b + \ln K]\sigma^2 t}{\sqrt{\sigma^2 t(Ts^2\sigma^2 t + 1)}},$$

Concluimos que la función (25) $V'(t, x) = V_1 - V_2$ es una solución de la ecuación de Black Scholes (1) para cualesquiera valores de las constantes K, T, s, ρ ($K > 0, T > 0, s > 0, b \equiv \frac{1}{2} - \frac{\rho}{s^2}$) y la función $\Phi(z)$ es la función de distribución normal.

Encontrar las soluciones exactas a partir de los operadores de tercer orden y de orden superior es más difícil porque sus órdenes son superiores al orden de la ecuación (1). Por ejemplo en nuestro artículo [15] encontramos tal solución de la ecuación de Schrödinger libre correspondiente a una subclase de la clase 5 de los operadores de tercer orden correspondiente a la fórmula (15 e). Esta solución se presenta como una serie de potencias con coeficientes dependientes del tiempo. Usando este resultado es fácil construir la nueva solución de ecuación de Black Scholes.

Finalizamos al mencionar que el modelo estudiado se puede extrapolar en diferentes direcciones. Por ejemplo, es posible considerar la ecuación de Black Scholes como la "proyección" de una ecuación de este tipo relacionada con varias variables $\{x_1, x_2, \dots, x_n\}$. También es posible considerar que las fluctuaciones aleatorias no son válidas para todas las variables x_i : así llegamos a una ecuación ultraparabólica.

Otro modelo puede estudiar la influencia de la variación de los valores de unas opciones sobre otras, lo que nos lleva al sistema de ecuaciones ligadas que generalizan el modelo de Black Scholes.

También es posible usar el paralelismo entre la mecánica cuántica y la mecánica clásica que pone en relación la ecuación de Schrödinger y la ecuación correspondiente de Hamilton Jacobi. En este sentido es posible también desarrollar el paralelismo entre la ecuación de Black Scholes y una ecuación diferencial no lineal de primer orden de tipo de Hamilton Jacobi. Este permitirá usar los métodos poderosos de la mecánica analítica en la teoría de Black Scholes.

El autor agradece a los árbitros anónimos por las sugerencias interesantes y a Rodolfo Echarri y Alexander Shapovalov por las discusiones fructíferas sobre el tema y al Departamento de Física de la Universidad Autónoma de Santo Domingo por el apoyo.

Bibliografía

1. J. Stampfi, V. Goodman (2002), *Matemáticas para las finanzas*, Thomson, Australia.

2. H. Fontanals, R. Lacayo, J. Vives (2002), *Alternative solutions of the Black Scholes equation*, (<http://www.ub.es/div2/recerca/documents/resum/e58.htm>).
3. P. Poncet, R. Abgrall (2002), *A few Multiresolution Schemes for the Black Scholes equation*, (<http://perso.wanadoo.fr/poncet/research/C-2000-Cemracs.pdf>).
4. A. Charlemagne (2003), *Classification of invariant solutions of the Black-Scholes equation*, (<http://duck.cs.und.ac.za/~banasiak/poe.pdf>).
5. D. Silverman (1999), *Solution of the Black Scholes equation using the Green's function of the Diffusion equation*, (<http://www.physics.uci.edu/~silverma/bseqn/bs.pdf>).
6. R. Almgren (2001), *Solving the Black-Scholes Equation*, (<http://www.math.toronto.edu/almgren/finmath/pde-01/notes1.pdf>).
7. C Wafo Soh, N H Ibragimov (1999), "Solution of the Cauchy Problem for the Black-Scholes Equation using its symmetries", Proceedings of the International Conference at the Sophus Lie Conference Centre, Nordfjordeid, 30 June-5 July (Norway: Mars Publishers). (<http://www.cam.wits.ac.za/staff/wafo.html>).
8. V. Shapovalov (1980), *Differential Equations*, 16, pp. 1864 - 1874.
9. W. Miller, Jr. (1977), *Symmetry and Separation of Variables*, Addison-Wesley Publishing Company, London.
10. L. V. Ovsiannikov (1982), *Group analysis of differential equations*, Boston, Academic Press.
11. P. Oliver (1996), *Equivalence, Invariants and Symmetry*, Cambridge.
12. M. Cohen de Lara (1995), "Geometric and symmetry properties of non-degenerate diffusion processes", *The Annals of Probability*, Vol 25, pp. 1557 - 1604.
13. N. Sukhomlin, M. Arias (2004), "Estudio de simetría y de posibilidades de la resolución exacta de las ecuaciones de Schrödinger y Hamilton - Jacobi para un sistema aislado, I. Clasificación de los operadores de simetría de tercer orden" (en publicación).
14. N. Sukhomlin (2004), "Estudio de simetría y de posibilidades de la resolución exacta de las ecuaciones de Schrödinger y Hamilton - Jacobi para un sistema aislado, II. Resolución exacta de las ecuaciones de Schrödinger y de Hamilton - Jacobi" (en publicación).

15. N. Sukhomlin, J. Álvarez, D. Pérez (2003), "Función de Green para una partícula cuasi relativista I. Partícula libre" (en publicación).

NIKOLAY SUKHOMLIN
DEPARTAMENTO DE FÍSICA,
UNIVERSIDAD AUTÓNOMA DE SANTO DOMINGO,
REPÚBLICA DOMINICANA

Sistemas de numeración, sistemas dinámicos substitutivos y fractales de Rauzy

Víctor F. Sirvent

1 Introducción

En este trabajo se presenta una introducción a la representación geométrica de sistemas dinámicos substitutivos. Estas representaciones geométricas se basan en sistemas de numeración *ad hoc* asociados a la substitución. Y se obtienen unos conjuntos fractales, conocidos como fractales de Rauzy. El objetivo de este trabajo es mostrar como estos tres conceptos están relacionados. Se comienza con dos ejemplos representativos, el caso Fibonacci y el caso Tribonacci, donde se exponen las ideas principales que luego son usadas en el caso general, el cual se presenta en la última sección.

Los sistemas dinámicos substitutivos o substituciones están muy ligados con la teoría de *tilings*, la cual no se expone aquí. Para ello ver [71, 67, 66]. Tampoco se presentan aquí los aspectos relacionados con la teoría ergódica asociada a este tipo de sistemas dinámicos.

2 El caso Fibonacci

Considere los números de Fibonacci $f_{k+2} = f_{k+1} + f_k$ con $f_0 = 1, f_1 = 1$. A cada número natural n se le puede representar de manera única como suma de números de Fibonacci, siempre que no haya dos números de Fibonacci consecutivos. Es decir:

Proposición 2.1 ([76]) *Dado n existe n_{i_0}, \dots, n_{i_k} tal que $n = f_{n_{i_0}} + \dots + f_{n_{i_k}}$ con $n_{i_j} > n_{i_{j-1}} + n_{i_{j-2}}$, con $2 \leq j \leq k$.*

A esta representación de los números naturales se le llama *representación de Zeckendorff*.

Esto nos permite, escribir $n = \sum_{i \geq 0} \epsilon_i(n) f_i$ donde $\epsilon(n) = \epsilon_0(n) \epsilon_1(n) \dots$ es una sucesión infinita en $\{0, 1\}$ tal que no contiene dos “1” consecutivos y su cola consiste de “0”. Consideremos la clausura de estas sucesiones en la topología producto de $\{0, 1\}^{\mathbb{N}}$ esto nos da el espacio:

$$\mathcal{F} = \{\underline{a} = a_0 a_1 \dots \mid a_i \cdot a_{i+1} = 0 \forall i\}.$$

Aquí \cdot es el producto usual de números enteros.

La relación de recurrencia de los números de Fibonacci está asociada a la ecuación polinómica $x^2 = x + 1$. La cual tiene como raíces $\lambda = (1 + \sqrt{5})/2$, la razón dorada, y $\alpha = -1/\lambda$.

Proposición 2.2 Sea $J = [-1, -1/\alpha]$. Entonces

$$\left\{ \sum_{i \geq 0} a_i \alpha^i \mid a_0 a_1 \dots \in \mathcal{F} \right\} = J$$

Demostración: Como α es negativo, se tiene que $\sum_0^\infty \alpha^{2i} = -1/\alpha$ es la cota superior y $\sum_0^\infty \alpha^{2i+1} = -1$ la cota inferior, por lo que

$$\left\{ \sum_{i \geq 0} a_i \alpha^i \mid a_0 a_1 \dots \in \mathcal{F} \right\} \subset [-1, -1/\alpha].$$

Por otro lado, considere las transformaciones $h_1, h_2 : J \rightarrow J$,

$$h_1(x) = \alpha x \quad h_2(x) = \alpha^2 x + 1$$

Debido a que la unión de $h_1(J)$, $h_2(J)$ es J y sus interiores son disjuntos, a cada elemento x de J se le puede escribir de la forma:

$$x = \bigcap_{n \geq 1} h_{x_1} \cdots h_{x_n}(J)$$

para una cierta sucesión $x_1 x_2 \dots$, donde $x_i = 1$ ó 2 . Debido a las expresiones de h_1, h_2 , el elemento de $\bigcap_{n \geq 0} h_{x_1} \cdots h_{x_n}(J)$ es de la forma $\sum_{i \geq 0} a_i \alpha^i$, para un cierto $\underline{a} \in \mathcal{F}$.

Más aún, si se definen $f_0, f_1 : J \rightarrow J$,

$$f_0(x) = \alpha x, \quad f_1(x) = \alpha x + 1$$

al elemento $x = \sum a_i \alpha^i$, se le expresa de la forma

$$x = \bigcap_{n \geq 0} f_{a_0} \cdots f_{a_n}(J).$$

Como $h_1 = f_0$ y $h_2 = f_1 \circ f_0$, este hecho nos da la relación entre \underline{a} y la sucesión $x_1 x_2 \dots$. ■

A continuación se hará un paréntesis, en esta discusión, ya que se requiere introducir unos conceptos básicos de fractales.

Definición 2.3 Un sistema iterativo de funciones o IFS (por sus siglas en inglés: *iterated function system*) consiste en un espacio métrico completo (X, d) junto con un conjunto finito de contracciones $f_i : X \rightarrow X$ con constante de contracción $0 \leq c_i < 1$, para $i = 1, \dots, n$. Se dice que el factor de contracción del IFS es $c = \max\{c_i \mid i = 1, \dots, n\}$.

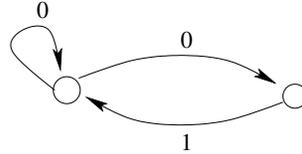


Figura 1: Grafo Fibonacci

Denotaremos por $\mathcal{H}(X)$ el conjunto formado por todos los subconjuntos compactos no vacíos de X . Este espacio tiene una distancia natural, la métrica de Hausdorff:

$$d_H(A, B) = \inf\{\delta \mid A \subset B_\delta \text{ y } B \subset A_\delta\}$$

donde $A_\delta = \{x \in X \mid d(x, A) \leq \delta \text{ para algún } \delta\}$. Si (X, d) es completo entonces $(\mathcal{H}(X), d_H)$ también es completo.

Teorema 2.4 (Hutchinson) *Sea $\{f_1, \dots, f_n\}$ un IFS con factor de contracción c . Entonces la transformación $F : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$ definida por*

$$F(E) = \bigcup_{i=1}^n f_i(E)$$

para todo $E \in \mathcal{H}(X)$, es una contracción en $(\mathcal{H}(X), d_H)$ con factor c .

Para la demostración del teorema véase [11, 20].

Definición 2.5 *Al único punto fijo de la contracción del teorema 2.4 se le llama atractor del IFS.*

Este tipo de conjuntos fue introducido originalmente por Moran [35] y sin embargo su definición fue modificada posteriormente por Hutchinson [27].

En el caso que se presenta en la demostración de la Proposición 2.2, el par $\{h_1, h_2\}$ forma un IFS y el intervalo J es su atractor. La relación entre las funciones h_i y f_i , nos permite presentar el IFS $\{h_1, h_2\}$ de la siguiente manera: Considere el grafo dirigido, de la figura 1.

Note que todos los elementos de \mathcal{F} se obtienen como caminos infinitos en el grafo anterior. Como para cada elemento x de J existe un $\underline{a} \in \mathcal{F}$ tal que $x = \bigcap_{n \geq 0} f_{a_0} \cdots f_{a_n}(J)$.

El par $\{f_0, f_1\}$ junto con el grafo dirigido, es un ejemplo de un GIFS, teniendo a J como atractor.

Sea $E = \{1, \dots, k\}$ y G_E un grafo dirigido primitivo con etiquetas en E . Sea $\{h_1, \dots, h_k\}$ un IFS en un espacio métrico completo (X, d) y

$$\Sigma_{G_E} = \{a_0 a_1 \dots \in E^{\mathbb{N}} \mid a_0 a_1 \dots a_n \text{ es un camino en } E \forall n\}.$$

Se define la aplicación $\pi : \Sigma_{G_E} \rightarrow X$, donde $\pi(\underline{a})$ es el único elemento de $\bigcap_{i \geq 0} h_{a_i}(X)$. El IFS $\{h_1, \dots, h_k\}$ junto con Σ_{G_E} se le denomina *sistema iterativo de funciones controlado por el grafo dirigido* o simplemente GIFS, por sus siglas en inglés *graph-directed iterated function system*. Al conjunto formado por la imagen de Σ_{G_E} bajo π , se le denomina el atractor del GIFS.

En general, no todo atractor de un GIFS se puede presentar como el atractor de un IFS.

En \mathcal{F} se puede definir un sistema dinámico de la siguiente manera: cualquier elemento en \mathcal{F} , cuya cola consiste en “0”, corresponde a un número natural, por lo tanto se puede definir su consecutivo. Esto induce por continuidad una transformación continua T en \mathcal{F} . Al sistema dinámico (\mathcal{F}, T) se le denomina *sistema ádico de Fibonacci*. El cual es minimal, es decir toda órbita es densa en \mathcal{F} .

Sea la transformación

$$\begin{aligned} \xi : \mathcal{F} &\longrightarrow J \\ \underline{a} &\longmapsto \sum_{i \geq 0} a_i \alpha^i. \end{aligned}$$

La cual es sobreyectiva y continua. De la demostración anterior se tiene que $\xi(\underline{a}) = \bigcap_{n \geq 0} f_{a_0} \dots f_{a_n}(J)$. Más aún el siguiente diagrama conmuta

$$\begin{array}{ccc} \mathcal{F} & \xrightarrow{T} & \mathcal{F} \\ \xi \downarrow & & \downarrow \xi \\ J & \xrightarrow{f} & J \end{array} \quad (2-1)$$

donde $f(x) = x + 1 \pmod{J}$.

Al sistema simbólico (\mathcal{F}, T) se le asocia el sistema “geométrico” (J, f) , de manera tal que existe una estrecha relación entre la estructura de las órbitas de ambos sistemas. Como la transformación ξ es sobreyectiva, el sistema (\mathcal{F}, T) tiene una estructura de órbitas más rica que (J, f) . Se dice que el sistema dinámico (J, f) es una realización geométrica del sistema ádico (\mathcal{F}, T) .

Esta dinámica tiene una estructura autosimilar. Nótese que $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_{10}$ donde $\mathcal{F}_0 = \{\underline{a} \in \mathcal{F} : a_0 = 0\}$ y $\mathcal{F}_{10} = \{\underline{a} \in \mathcal{F} : a_0 = 1, a_1 = 0\}$. Lo cual se puede expresar de la forma: $\mathcal{F}_0 = \tau(\mathcal{F}_{10})$ y $\mathcal{F}_{10} = T\tau^2(\mathcal{F})$, donde $\tau : \mathcal{F} \rightarrow \mathcal{F}$, $\tau(a_0 a_1 \dots) = 0 a_0 a_1 \dots$. Es decir \mathcal{F} es el punto fijo del IFS: $\{\tau, T\tau^2\}$. Esta

estructura se refleja en la dinámica. Sea \tilde{T} la transformación inducida por T en \mathcal{F}_0 , es decir, $\tilde{T}(\underline{a}) = T^{N(\underline{a})}(\underline{a})$ donde $N(\underline{a}) = \min\{n : T^n(\underline{a}) \in \mathcal{F}_0\}$. Como el conjunto \mathcal{F} admite la partición $\mathcal{F}_{00} = \tau(\mathcal{F}_0)$, $\mathcal{F}_{010} = T\tau^2(\mathcal{F}_0)$. Se tiene que \tilde{T} es T en \mathcal{F}_{010} y sobre \mathcal{F}_{00} es igual a T^2 .

Esta estructura autosimilar se traduce en el intervalo J de la siguiente forma. Sea $h : J \rightarrow J$ definido por $h(x) = \alpha x$ y $J_0 = h(J)$, $J_{10} = fh^2(J)$. Note que $h = h_1$, $fh = h_2$, donde h_1, h_2 son las transformaciones definidas en la demostración de la proposición 2.2. Como se mencionó anteriormente el par $\{h, fh^2\}$ forma un IFS, cuyo atractor es el intervalo J .

Substitución de Fibonacci: Considere el alfabeto $\mathcal{A} = \{a, b\}$ y $\mathcal{A}^* = \cup_{i \geq 0} \mathcal{A}^i$ el conjunto de palabras finitas en dicho alfabeto. Sea la transformación $\zeta : \mathcal{A} \rightarrow \mathcal{A}^*$ dada por $a \rightarrow ab, b \rightarrow a$. La aplicación ζ induce una transformación en \mathcal{A}^* , por juxtaposición de palabras: $\zeta(UV) = \zeta(U)\zeta(V)$ y esta a su vez induce otra transformación en $\mathcal{A}^{\mathbb{N}}$. La cual se seguirá denotando por ζ . Esta transformación es la *substitución de Fibonacci*. Note que la longitud de la palabra $\zeta^i(a)$ es igual al número de Fibonacci f_i .

Esta transformación tiene un único punto fijo:

$$\mathbf{u} = \zeta^\infty(a) = abaababaaba \dots$$

Sea σ el “shift” en $\mathcal{A}^{\mathbb{N}}$: $\sigma(v_0v_1v_2\dots) = v_1v_2\dots$, donde $v_0v_1v_2\dots \in \mathcal{A}^{\mathbb{N}}$. Se define

$$\Omega = \overline{\{\sigma^N(\underline{u}) : N \in \mathbb{N}\}}.$$

El sistema (Ω, σ) se llama sistema dinámico asociado a la substitución ζ . Es fácil verificar que dicho sistema dinámico es minimal.

Los sistemas dinámicos (Ω, σ) y (\mathcal{F}, T) son isomorfos. El isomorfismo viene dado por el siguiente hecho. Sea $U_N = u_0u_1\dots u_{N-1}$, la palabra formada por los N primeros símbolos del punto fijo. Si la representación de Zeckendorff del número N es $\sum_{j=0}^L \epsilon_j f_j$, para un L . Entonces se tiene

$$U_N = \zeta^L(\delta(\epsilon_L)) \cdots \zeta(\delta(\epsilon_1))\delta(\epsilon_0)$$

donde $\delta(0) = \emptyset$, la palabra vacía y $\delta(1) = a$.

Por lo mencionado anteriormente se tiene que el sistema dinámico (J, f) es una realización geométrica del sistema simbólico (Ω, σ) .

Se dice que el sistema dinámico simbólico (Σ, σ) se *realiza geoméricamente* si existe un sistema dinámico (X, f) donde X es una estructura geométrica, por ejemplo una variedad diferenciable, y una aplicación $\psi : \Sigma \rightarrow X$ continua y sobreyectiva tal que el siguiente diagrama conmute:

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma} & \Sigma \\ \psi \downarrow & & \downarrow \psi \\ X & \xrightarrow{f} & X \end{array} \tag{2-2}$$

Al sistema (X, f) se le llama una *realización geométrica* de (Σ, σ) .

El tener una realización geométrica de un sistema simbólico nos permite codificar las órbitas del sistema geométrico, de manera tal que el itinerario de las órbitas viene dado por el sistema simbólico. Veamos como la codificación de las órbitas del sistema (J, f) viene dado por las sucesiones que se encuentran en Ω . Sea $\nu : J \rightarrow \{1, 2\}$, dada por $\nu(x) = i$ si $x \in J_i$, donde $J_i = h_i(J)$, para $i = 1, 2$. Aquí se considera a J, J_1, J_2 , semiabiertos, cerrados a izquierda, para que ν esté bien definida. Los intervalos J_i son los intervalos de continuidad de f . El itinerario o codificación de la órbita (futura), bajo f , del punto x de acuerdo con esta partición viene dado por la sucesión $\{\nu(f^n(x))\}_{n \geq 0}$, la cual denotaremos por $\theta(x)$.

Veamos que el itinerario de la órbita de $x = 0$, viene dado por \mathbf{u} el punto fijo de la substitución. Sea \tilde{f} la transformación inducida por f en J_1 . Como J_1 admite la partición dada por $h(J_1), h(J_2)$, el itinerario común de todos los puntos de $h(J_1)$, bajo f , es "12", mientras que bajo \tilde{f} es "1"; y para $h(J_2)$ es "1" y "2" para f y \tilde{f} respectivamente. Así se tiene que la relación entre la codificación de la órbita de f y \tilde{f} , en J_1 , viene dada precisamente por la substitución ζ . Si ahora se continúa este proceso de inducción en $h(J_1)$, luego en $h^2(J_1)$ y así sucesivamente, se tiene que en la etapa k -ésima el itinerario común de los puntos de $h^k(J_1)$, bajo f , es $\zeta^{k+1}(1)$. Mientras que bajo la transformación inducida en $h^k(J_2)$, es "1". En $h^k(J_2)$ es $\zeta^{k+1}(2)$ y "2" para f y la transformación inducida, en $h^k(J_2)$, respectivamente. Así la relación entre el itinerario de los puntos en $h^k(J)$, bajo f y la transformación inducida viene dada por ζ^k . Como $\bigcap_{k \geq 0} h^k(J) = \{0\}$, se tiene que $\theta(0) = \zeta^\infty(1) = \mathbf{u}$.

Esto nos permite tener el siguiente diagrama conmutativo:

$$\begin{array}{ccc}
 J & \xrightarrow{f} & J \\
 \theta \downarrow & & \downarrow \theta \\
 \Omega & \xrightarrow{\sigma} & \Omega
 \end{array} \tag{2-3}$$

Resulta ser que la transformación θ no es sobreyectiva, debido a la elección de intervalos semiabiertos en la partición de J . A cada elemento de J , su itinerario viene dado por un elemento de Ω .

3 El caso tribonacci y el fractal de Rauzy

El caso tribonacci sigue las mismas ideas expuestas en la sección anterior, sin embargo hay ciertas diferencias, que se expondrán. Los números de tribonacci se obtienen de la relación de recurrencia: $g_{k+3} = g_{k+2} + g_{k+1} + g_k$ con $g_0 = 1, g_1 = 2$ y $g_2 = 4$. De manera similar al caso Fibonacci, existe una representación de

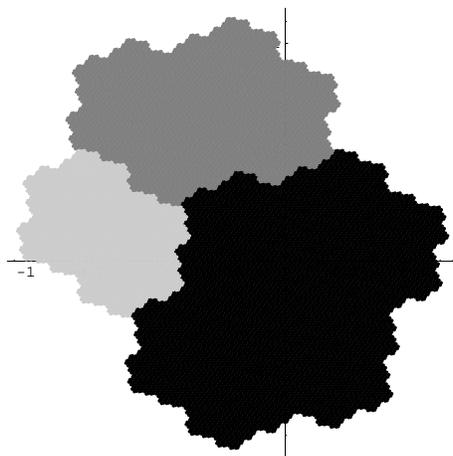


Figura 2: El fractal de Rauzy y su descomposición autosimilar.

Zeckendorff, asociada a esta familia de números. Lo cual nos permite representar $n = \sum_{i \geq 0} \epsilon_i(n) g_i$, aquí $\epsilon(n) = \epsilon_0(n) \epsilon_1(n) \dots$ es una sucesión infinita en $\{0, 1\}$ tal que no contiene tres “1” consecutivos y su cola está formada por “0”. Así obtenemos, la clausura de tales sucesiones:

$$\mathcal{T} = \{\underline{a} = a_0 a_1 \dots \mid a_i \cdot a_{i+1} \cdot a_{i+2} = 0 \forall i\}.$$

En el espacio \mathcal{T} se define una dinámica ádica T , siendo el sistema dinámico (\mathcal{T}, T) minimal.

La ecuación asociada a esta relación de recurrencia es $x^3 = x^2 + x + 1$, la cual tiene como raíz real λ y $\beta, \bar{\beta}$ como raíces complejas. Al conjunto

$$\mathcal{R} = \left\{ \sum_{i \geq 0} a_i \beta^i \mid a_0 a_1 \dots \in \mathcal{T} \right\}$$

se le llama el *Fractal de Rauzy*. Este objeto geométrico fue introducido originalmente por G. Rauzy en [49].

El conjunto \mathcal{R} es compacto, conexo, simplemente conexo [49]. Su frontera es fractal, con dimensión de Hausdorff no entera, la cual ha sido estudiada en [28, 57, 33].

De manera similar a la expuesto en la demostración de la proposición 2.2, el fractal de Rauzy es el atractor del IFS, formado por las transformaciones $\{h_1, h_2, h_3\}$, las cuales están definidas en el plano complejo de la siguiente manera:

$$h_1(z) = \beta z \quad h_2(z) = \beta^2 z + 1 \quad h_3(z) = \beta^3 z + \beta^2 + 1.$$

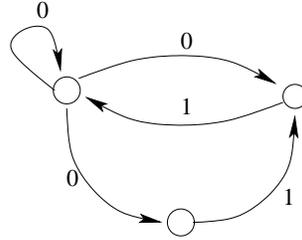


Figura 3: Grafo tribonacci

Esto nos da la siguiente estructura autosimilar en \mathcal{R} , sean $\mathcal{R}_1 = h_1(\mathcal{R})$, $\mathcal{R}_2 = h_2(\mathcal{R})$ y $\mathcal{R}_3 = h_3(\mathcal{R})$. Estos conjuntos tienen interior disjuntos y su unión nos da todo \mathcal{R} . Más adelante mostraremos la relevancia dinámica de esta descomposición autosimilar.

Este IFS, se puede presentar como un GIFS. Considere el grafo dirigido, de la figura 3

Todos los elementos de \mathcal{T} se obtienen como caminos infinitos en este grafo. Sean las transformaciones: $f_0(z) = \beta z$, $f_1(z) = \beta z + 1$. Como $h_1 = f_0$, $h_2 = f_1 f_0$, $h_3 = f_1^2 f_0$, se tiene que el IFS $\{h_1, h_2, h_3\}$ se obtiene del GIFS dado por el grafo anterior y las transformaciones $\{f_0, f_1\}$ y su atractor es precisamente el fractal de Rauzy \mathcal{R} .

La dinámica ádica se realiza geoméricamente de la siguiente manera. Sea

$$\begin{aligned} \xi_{\mathcal{R}} : \mathcal{T} &\longrightarrow \mathcal{R} \\ \underline{a} &\longmapsto \sum_{i \geq 0} a_i \beta^i. \end{aligned}$$

Como en el caso Fibonacci, se tiene $\xi_{\mathcal{R}}(\underline{a}) = \bigcap_{n \geq 0} f_{a_0} \cdots f_{a_n}(\mathcal{R})$. Esta aplicación es sobreyectiva y continua, y genera el siguiente diagrama conmutativo:

$$\begin{array}{ccc} \mathcal{T} & \xrightarrow{T} & \mathcal{T} \\ \xi_{\mathcal{R}} \downarrow & & \downarrow \xi_{\mathcal{R}} \\ \mathcal{R} & \xrightarrow{F} & \mathcal{R} \end{array} \tag{3-4}$$

donde la transformación F intercambia tres subregiones autosimilares de \mathcal{R} , descritas anteriormente [49]. Obviamente esta transformación F no es continua.

Sin embargo, resulta ser que \mathcal{R} es un dominio fundamental para el toro 2-dimensional, \mathbb{T}^2 , y la transformación F se convierte en una translación definida por un vector con componentes racionalmente independientes [49], en el toro. Obteniendo de esta manera el siguiente diagrama conmutativo:

$$\begin{array}{ccc}
 \mathcal{T} & \xrightarrow{T} & \mathcal{T} \\
 \widehat{\xi}_{\mathcal{R}} \downarrow & & \downarrow \widehat{\xi}_{\mathcal{R}} \\
 \mathbb{T}^2 & \xrightarrow{\widehat{F}} & \mathbb{T}^2
 \end{array} \tag{3-5}$$

Donde $\widehat{\xi}_{\mathcal{R}}$ y \widehat{F} , son las transformaciones que se obtienen de $\xi_{\mathcal{R}}$ y F , al tomar el cociente bajo el lattice que convierte a \mathcal{R} en un dominio fundamental de \mathbb{T}^2 .

Se pueden definir subconjuntos dinámicos de \mathcal{R} , prohibiendo ciertas subpalabras finitas en \mathcal{T} . En estos conjuntos se tiene una dinámica inducida. Las propiedades dinámicas y geométricas de estos sistemas han sido estudiadas en [54, 56, 61]

Veamos ahora como esta dinámica está relacionada con sustituciones.

Substitución Tribonacci: Considere el alfabeto $\mathcal{A} = \{a, b, c\}$ y $\mathcal{A}^* = \cup_{i \geq 0} \mathcal{A}^i$ el conjunto de palabras finitas en dicho alfabeto. Sea la transformación $\zeta : \mathcal{A} \rightarrow \mathcal{A}^*$ dada por $a \rightarrow ab, b \rightarrow ac, c \rightarrow a$. Como se explicó en el caso Fibonacci, esta transformación induce otra en $\mathcal{A}^{\mathbb{N}}$, la cual tiene un único punto fijo: $\mathbf{u} = \zeta^{\infty}(a)$. A él se le asocia el sistema dinámico minimal (Ω, σ) , donde Ω es la clausura bajo el “shift”, σ , del punto fijo \mathbf{u} . Este sistema dinámico se relaciona con el sistema ádico de manera similar al caso Fibonacci. Teniendo que (Ω, σ) y (\mathcal{T}, T) son isomorfos. Por lo que $(\mathbb{T}^2, \widehat{F})$ es una realización geométrica de (Ω, σ) .

Por la construcción anterior si se codifica la órbita del origen bajo \widehat{F} , usando las regiones $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$, se obtiene \mathbf{u} , el punto fijo de la substitución. Más aún el itinerario de la órbita de cada punto bajo \widehat{F} , usando la partición anterior nos da un elemento de Ω .

k-bonacci substituciones: Las construcciones expuestas anteriormente para los caso Fibonacci y tribonacci se generalizan de manera inmediata para el caso k-bonacci. Los números k-bonacci son aquellos que se obtienen de la relación de recurrencia de grado k: $g_{n+k} = g_{n+k-1} + \dots + g_{n+1} + g_n$, para $n \geq 0$, con $g_j = 2^j$ para $j = 0, \dots, k - 1$. A ellos se le asocia el sistema ádico $(\mathcal{T}(k), T)$, donde $\mathcal{T}(k)$ es, como antes, la clausura de las sucesiones de “0” y “1” asociada a la representación de Zeckendorf:

$$\mathcal{T}(k) = \{ \underline{a} = a_0 a_1 \dots \mid \prod_{j=0}^{k-1} a_{i+j} = 0 \forall i \}.$$

Y también se les asocia las siguientes substitutiones:

$$\zeta_k : \begin{cases} 1 & \rightarrow 12 \\ 2 & \rightarrow 13 \\ & \vdots \\ k-1 & \rightarrow 1k \\ k & \rightarrow 1. \end{cases}$$

El polinomio asociado a la relación de recurrencia de k -bonacci es $x^k - x^{k-1} - \dots - x - 1$. El cual tiene una única raíz real mayor que 1, en valor absoluto, λ . En el caso k impar las otras raíces son $\beta_1, \overline{\beta_1}, \dots, \beta_l, \overline{\beta_l}$, en el caso k impar las raíces son $\beta_1, \overline{\beta_1}, \dots, \beta_l, \overline{\beta_l}, \omega$ donde las β_j son complejas no reales y w real. Todos ellas tienen norma menor que 1.

El fractal de Rauzy asociado a la substitution ζ_k es el objeto en $\mathbb{C}^l \approx \mathbb{R}^{2l} = \mathbb{R}^{k-1}$, si k es impar, o en $\mathbb{C}^l \times \mathbb{R} \approx \mathbb{R}^{2l+1} = \mathbb{R}^{k-1}$, si k es par; definido por:

$$\mathcal{R}(k) = \left\{ \left(\sum_{i \geq 0} a_i \beta_1^i, \dots, \sum_{i \geq 0} a_i \beta_l^i, \sum_{i \geq 0} a_i \omega^i \right) \mid \underline{a} \in \mathcal{T}(k) \right\}.$$

Este conjunto es compacto, conexo y es igual a la clausura de su interior [67]. También tiene una estructura autosimilar, que nos da una partición natural en k sub-regiones. El IFS que nos da esta estructura autosimilar, es una generalización inmediata del IFS, definido en el caso tribonacci.

En $\mathcal{R}(k)$ se define una transformación que intercambia los k subregiones autosimilares. Más aún $\mathcal{R}(k)$ es un dominio fundamental para el toro $(k-1)$ -dimensional, y la transformación de intercambio de subregiones se convierte en una traslación dada por un vector con componentes racionalmente independientes.

Estas k regiones nos dan que la codificación de las órbitas de la traslación vienen dadas por elementos del sistema dinámico asociado a ζ_k .

4 Caso general

Sea \mathcal{A} un alfabeto finito y $\zeta : \mathcal{A} \rightarrow \mathcal{A}^*$ una substitution. La cual induce una transformación en el espacio de sucesiones infinitas (hacia un lado) sobre \mathcal{A} , tal como se explicó en los casos particulares estudiados anteriormente. Se está interesado en los puntos fijos de ζ o de ζ^k para algún k . Estos siempre existen, cuando la longitud de la palabra $\zeta^n(a)$ va a infinito, para todo a del alfabeto \mathcal{A} , cuando n crece. Si se remueven del alfabeto los símbolos que no aparecen en el punto fijo, se tiene que existe un símbolo que se denotará por 1, tal que $\zeta(1)$ comience por 1. Bajo estas condiciones se tiene \mathbf{u} un punto fijo de la substitution ζ , tal que $\mathbf{u} = \zeta^\infty(1)$. A esta sucesión se le asocia el sistema

dinámico (Ω, σ) tal como se describió en los ejemplos anteriores. Este sistema dinámico es minimal si y sólo si para todo $a \in \mathcal{A}$ existe un k tal que $\zeta^k(a)$ contiene el símbolo 1.

Un ejemplo de importancia histórica es la substitución de Thue-Morse:

$$\zeta : \begin{cases} 0 \rightarrow 01 \\ 1 \rightarrow 10. \end{cases}$$

Esta substitución tiene dos puntos fijos: $\zeta^\infty(0) = 01101001\dots$ y $\zeta^\infty(1) = 10010110\dots$. Nótese que se obtiene una sucesión de la otra intercambiando 0 por 1 y viceversa. Esta sucesión ha aparecido en varias áreas de la matemática. Originalmente surgió en el año 1851 en [46], estudiando potencias de números, luego fue introducida por Thue [73] al estudiar sucesiones no periódicas y posteriormente por Morse [36] al estudiar flujos geodésicos sobre superficies de curvatura negativa. Existe una amplia bibliografía sobre esta substitución. A diferencia de los ejemplos estudiados en las secciones anteriormente. Esta substitución es de longitud constante, es decir, la longitud de la palabra imagen de cada símbolo del alfabeto es la misma.

A la substitución ζ se le asocia su *matriz de incidencia* $M = M(\zeta)$, donde sus elementos M_{ij} es el número de veces que aparece el símbolo i en $\zeta(j)$.

Se dice que una matriz M es *positiva* si sus elementos son enteros no negativos. M es *primitiva* si existe un número positivo k tal que M^k es positiva.

Teorema 4.1 (Perron-Frobenius) *Sea M una matriz positiva primitiva. Entonces M admite un autovalor estrictamente positivo λ , tal que $\lambda > |\beta|$ para cualquier otro autovalor β de M . El autovalor λ es simple. Más aún, existe un autovector estrictamente positivo que corresponde con λ .*

Al autovector λ y al correspondiente autovector positivo del teorema anterior se le conoce como autovalor y autovector de Perron-Frobenius, respectivamente.

Se dice que la substitución ζ es primitiva si su matriz $M = M(\zeta)$ es primitiva. La substitución ζ es *Pisot unimodular* si $|\det(M)| = 1$, el polinomio característico de M es irreducible en \mathbb{Q} y este tiene una raíz real mayor que 1 y sus otras raíces tienen norma estrictamente menor que 1. La raíz mayor es el autovalor de Perron-Frobenius de la matriz M . Toda substitución de Pisot unimodular es primitiva [14].

La frecuencia con que aparecen los símbolos en la sucesión viene dada por el autovector de Perron-Frobenius de la matriz de incidencia. Muchas otras propiedades métricas del sistema dinámico substitutivo, vienen del autovalor de Perron-Frobenius y su correspondiente autovector.

Una propiedad importante de los sistemas dinámicos substitutivos es que

son auto-inducidos, es decir el siguiente diagrama conmuta,

$$\begin{array}{ccc}
 \Omega & \xrightarrow{\sigma} & \Omega \\
 \zeta \downarrow & & \downarrow \zeta \\
 \zeta(\Omega) & \xrightarrow{\tilde{\sigma}} & \zeta(\Omega)
 \end{array} \tag{4-6}$$

donde (Ω, σ) es el sistema que proviene de la substitución ζ , y $\tilde{\sigma}$ es la transformación inducida de σ en $\zeta(\Omega)$, o sea $\tilde{\sigma}(\mathbf{v}) = \sigma^{n(\mathbf{v})}(\mathbf{v})$, donde $n(\mathbf{v})$ es el tiempo de retorno de \mathbf{v} en $\zeta(\Omega)$. La transformación $\zeta : \Omega \rightarrow \zeta(\Omega)$ es continua y sobreyectiva. Esta propiedad de las substituciones induce una estructura autosimilar en el espacio Ω .

4.1 Sistemas de numeración y autómatas de Prefijos asociados a substituciones

Sea ζ una substitución primitiva con un único punto fijo $\mathbf{u} = \zeta^\infty(1)$ tal que $u_0 = 1$. A estas substituciones se le puede asociar un sistema de numeración basado en el autómata de prefijos.

Definición 4.2 *Un autómata es una quintupla (S, L, t, i, f) donde S es el conjunto de estados, L las etiquetas, $t : S \times S \rightarrow L$ la función de transición, i los estados iniciales y f los estados finales del autómata.*

Al autómata se representará por un grafo dirigido con etiquetas, donde los vértices están dados por S , las etiquetas por L , las transiciones por la función t . Se considerarán los caminos sobre el autómata que comiencen en los estados iniciales y terminen en los estados finales. En [18] se encuentra descrita la teoría general de autómatas.

El autómata de prefijos de ζ se define en [50] de la siguiente manera:

- El conjunto de los estados es el alfabeto \mathcal{A} .
- El conjunto de etiquetas está formado por los prefijos propios de las palabras obtenidas al aplicar ζ a cada elemento del alfabeto. Se admite la palabra vacía \emptyset como prefijo propio. Este conjunto se denotará por Pref.
- Transiciones: Si p, q son estados del autómata y W una palabra de Pref entonces existe una transición de p a q etiquetada por W , si y sólo si la palabra Wq es prefijo de $\zeta(p)$.
- El estado inicial es “1”.
- Todos los estados son estados de salida o finales.

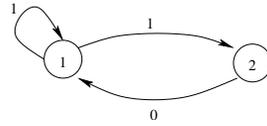


Figura 4: Autómata de prefijos asociado a la substitución de Fibonacci

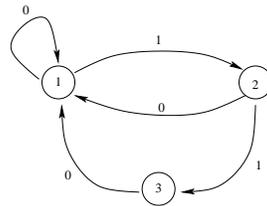


Figura 5: Autómata de prefijos asociado a la substitución de Tribonacci

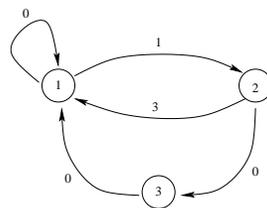


Figura 6: Autómata de prefijos asociado a la substitución $1 \rightarrow 12$, $2 \rightarrow 13$, $3 \rightarrow 1$

Estos autómatas tienen la propiedad:

Teorema 4.3 ([50]) *Sea U un prefijo no vacío de \mathbf{u} , el punto fijo de la sustitución ζ . Entonces existe un único camino en el autómata de prefijos, comenzando por el estado 1 y etiquetado por $(W(n), W(n-1), \dots, W(0))$ tal que $W(n) \neq \emptyset$, la palabra vacía, y $U = \zeta^n(W(n))\zeta^{n-1}(W(n-1)) \dots W(0)$.*

Recíprocamente, a cualquiera de estos caminos le corresponde un prefijo de \mathbf{u} , dado por la fórmula anterior.

De acuerdo con este teorema, se puede escribir $u_0 \dots u_{n-1}$ los primeros n símbolos de \mathbf{u} así: $u_0 \dots u_{n-1} = \zeta^k(W(k))\zeta^{k-1}(W(k-1)) \dots W(0)$. Por lo tanto $n = |\zeta^k(W(k))| + |\zeta^{k-1}(W(k-1))| + \dots + |W(0)|$ donde $|W|$ denota la longitud de la palabra W . De esta forma se tiene un sistema de numeración basado en el autómata de prefijos. Como se vió en las secciones anteriores, si se toma la sustitución de Fibonacci, el sistema de numeración que se obtiene es el sistema de Zeckendroff, similarmente en el caso k -bonacci.

De manera análoga se puede definir el autómata de sufijos y se tiene el mismo sistema de numeración [50, 51].

En [17] se encuentra una versión general del Teorema 4.3 y en [22] un estudio general de sistemas de numeración basados en relaciones de recurrencia y en autómatas. Con los sistemas de numeración basados en sustituciones se pueden definir operaciones binarias en los naturales que dan origen a semigrupos aritméticos que son representaciones de la estructura autosimilar de los sistemas dinámicos asociados a las sustituciones, para detalles véase [55, 58]. Otra operación binaria en los naturales asociada a sustituciones es la multiplicación de Fibonacci, la cual ha sido estudiada en [30, 7, 34]. En [21, 44] se pueden ver otros semigrupos aritméticos.

4.2 Sistemas ádicos

El sistema dinámico (Ω, σ) asociado a la sustitución ζ , admite una representación equivalente llamada sistema ádico. Sea

$$\mathcal{N} = \{ \underline{a} = a_0 a_1 \dots \in \text{Pref}^{\mathbb{Z}^+} \mid a_n \dots a_0 \text{ es un camino en Aut para todo } n \in \mathbb{Z}^+ \}$$

donde Aut es el autómata de prefijos de ζ .

Como todos los caminos finitos en Aut corresponden a números naturales, se puede definir una transformación continua T en \mathcal{N} , que es la extensión de la adición por 1 en los enteros. De esta forma se tiene el sistema dinámico (\mathcal{N}, T) , el cual es llamado sistema ádico de ζ . Este sistema también es llamado odómetro o máquina de sumar. Para más detalles sobre sistemas ádicos véase [74, 24].

Toda sustitución minimal es isomorfa a su transformación ádica ([26]).

4.3 Representación geométrica de (Ω, σ)

Sea \mathcal{A}^* el monoide generado sobre el alfabeto \mathcal{A} donde está definida la substitución ζ , la cual satisface las propiedades requeridas en la sección anterior, y $(G, +)$ un grupo topológico dotado de una distancia. Sea $\Delta : \mathcal{A}^* \rightarrow G$ una transformación tal que $\Delta(UV) = \Delta(U) + \Delta(V)$ y que $\Delta(\zeta^n(U))$ converge exponencialmente a cero. Sea $a_n \dots a_0$ un camino finito en el autómata de prefijos de ζ . Por el teorema 4.3 existe $W \in \mathcal{A}^*$ tal que $W = \zeta^n(a_n) \dots a_0$ por lo tanto $\Delta(W) = \Delta(\zeta^n(a_n)) + \dots + \Delta(a_0)$. Esto nos permite introducir la transformación: $\Phi : \mathcal{N} \rightarrow G$, donde \mathcal{N} es el conjunto de donde esta definido el sistema ádico. La cual está definida así: $\Phi(\underline{a}) = \sum_{i=0}^{\infty} \Delta(\zeta^i(a_i))$ se observa que no tiene problemas de definición por la propiedad de convergencia. Sea $\mathbf{u} = \zeta(\mathbf{u}) = U_n \sigma^n(\mathbf{u})$ se define la transformación $\delta : \Omega \rightarrow G$ de la siguiente manera $\delta(\mathbf{u}) = \Delta(U_n) + \delta(\sigma^n(\mathbf{u}))$. Como $\mathbf{u} = \zeta^\infty(1)$ se define $\delta(\mathbf{u}) = 0$, por lo tanto $\delta(\sigma^n(\mathbf{u})) = -\Delta(U_n)$. Sea \mathbf{v} un elemento de Ω , se puede escribir $\mathbf{v} = v_0 \sigma(\mathbf{v})$ por lo que $\delta(\sigma(\mathbf{v})) = \delta(\mathbf{v}) - \Delta(v_0)$. Por lo que podemos resumir que al *shift* σ en Ω le corresponde una translación en un número finito de pedazos, en G . La naturaleza del grupo $(G, +)$ y de la aplicación Δ depende del tipo de substitución, en particular de $M = M(\zeta)$ la matriz asociada a la substitución.

Sea ζ una substitución de longitud constante, se define $\Delta(U) = |U|$ la longitud de la palabra de U , es decir la distancia p -ádica, la cual tiene la propiedad $\Delta(\zeta(U)) = p\Delta(U)$, donde p es la longitud de la substitución. Por lo tanto

$$\Phi(\underline{a}) = |a_0| + p|a_1| + \dots + p^n|a_n| + \dots$$

La dinámica del *shift* se realiza como una translación en -1 sobre los números p -ádicos. En el caso de la substitución de Thue-Morse, ésta se realiza como una translación en -1 sobre los números 2-ádicos [23].

En caso que ζ sea Pisot unitaria en un alfabeto de k elementos se toma $G = \mathbb{R}^{k-1}$ y

$$\Delta(U) = |U| \begin{pmatrix} v_1 \\ \vdots \\ v_{k-1} \end{pmatrix} - \begin{pmatrix} r_1(U) \\ \vdots \\ r_{k-1}(U) \end{pmatrix}, \tag{4-7}$$

donde (v_1, \dots, v_k) es el autovector positivo de Perron-Frobenius de la matriz M y $r_i(U)$ es el número de veces que se encuentra el símbolo i en la palabra U . Esta aplicación tiene la propiedad $\Delta(UV) = \Delta(U) + \Delta(V)$ y $\Delta(\zeta(U)) = B\Delta(U)$ donde B es la restricción de la matriz B al autoespacio contractivo de M [49]. Por lo que $\Phi(\underline{a}) = \sum_{i \geq 0} B^i \Delta(a_i)$. Esta suma converge ya que los autovalores de B están todos en el interior del círculo unitario y los a_i son finitos ya que son prefijos propios de la substitución. De esta manera

$$\delta(\Omega) = \left\{ \sum_{i=0}^{\infty} B^i \Delta(a_i) \mid \underline{a} \in \mathcal{N} \right\}$$

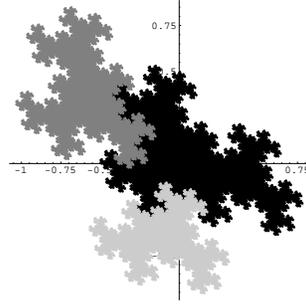


Figura 7: El fractal de Rauzy de la sustitución $1 \rightarrow 12, 2 \rightarrow 3, 3 \rightarrow 1$.

A este conjunto se le conoce como el *Fractal de Rauzy* asociado a la sustitución ζ .

Como se vió antes

$$\begin{aligned} \delta(\sigma(\mathbf{v})) &= \delta(\mathbf{v}) - \Delta(v_0) \\ &= \delta(\mathbf{v}) - \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix} + e_{v_0} \end{aligned}$$

donde $e_j = (0, \dots, 0)$ si $j = k$ y en el caso $j \neq k$ es el vector canónico en la dirección j -ésima. Por lo que *shift* se puede realizar como una translación en k direcciones en \mathbb{R}^{k-1} , lo cual es un intercambio de k pedazos en \mathbb{R}^{k-1} . Más aún el conjunto $\delta(\Omega)$ es un dominio fundamental del toro $(k-1)$ -dimensional y el shift se realiza como una translación sobre el toro, \mathbb{T}^{k-1} . [49, 70, 15]. La translación viene dada por el vector (v_1, \dots, v_k) , cuyas componentes son racionalmente independientes.

Se puede cambiar la definición de Δ , cambiando coordenadas. Si $\Delta(U) = |U|(w_1 + \dots + w_{k-1})$ donde w_1, \dots, w_{k-1} es una base del espacio contractivo de M , se tiene que $\Delta(UV) = \Delta(U) + \Delta(V)$ y $\Delta(\zeta(U)) = A\Delta(U)$, donde A es una matriz equivalente a B . Por lo tanto se puede escribir a la transformación Φ de la siguiente manera:

$$\Phi(a) = \begin{pmatrix} \sum_{i \geq 0} \beta_1^i \Delta(a_i) \\ \vdots \\ \sum_{i \geq 0} \beta_l^i \Delta(a_i) \end{pmatrix}$$

donde β_1, \dots, β_l son los autovalores de norma menores que uno de la matriz M .

En el ejemplo de la sustitución tribonacci, esta es la construcción presentada en la sección anterior.

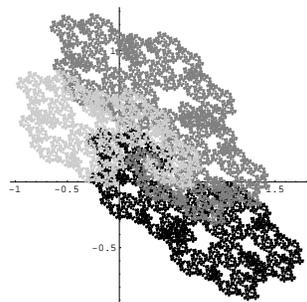


Figura 8: El fractal de Rauzy de la substitución $1 \rightarrow 32, 2 \rightarrow 1, 3 \rightarrow 2$.

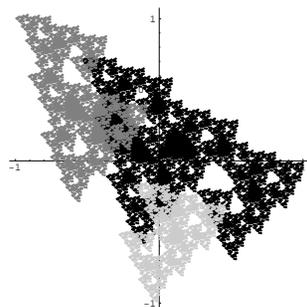


Figura 9: El fractal de Rauzy de la substitución $1 \rightarrow 12, 2 \rightarrow 31, 3 \rightarrow 1$.

Los fractales de Rauzy son conjuntos cerrados, de interior no vacío y coinciden con la clausura de su interior [67].

Por otro lado los fractales de Rauzy se pueden obtener como atractores de GIFS, donde el grafo asociado es el grafo del automata de prefijos donde los lados que unen vértices, se le ha dado la orientación opuesta. Para más detalles ver [26, 67].

Referencias

- [1] R. Adler, Symbolic dynamics and Markov Partition, *Bull. Amer. Math. Soc.*, **35**, (1998), 1-56.
- [2] V. Afraimovich, Pesin's dimension for Poincaré recurrences, *Chaos*, **7** (1997), 12-20.
- [3] V. Afraimovich, Poincaré recurrences of couple subsystems in synchronized regimes, *Taiwanese J. of Math.* **3** (1999), 139-161.
- [4] V. Afraimovich, A. Maass and J. Urias, Symbolic dynamics for sticky sets in Hamiltonian systems, *Nonlinearity*, **13** (2000), 617-637.
- [5] V. Afraimovich, J. Schmeling, E. Ugalde and J. Urias, Spectra of dimensions for Poincaré recurrences, *Discrete and Continuous Dynamical Systems*, **6**, (2000), 901-914.
- [6] V. Afraimovich and G.M. Zaslavsky. Fractal and multifractal properties of exit times and Poincaré recurrences, *Physical Review E* **55** (1997), 5418-5426.
- [7] P. Arnoux, Some remarks about Fibonacci multiplication, *Appl. Math. Lett.* **2** (1989), 319-320.
- [8] P. Arnoux, *Représentation géométrique des systèmes dynamiques symboliques*, Habilitation, Université D'Aix-Marseille II, 1991.
- [9] P. Arnoux and S. Ito, Pisot substitutions and Rauzy fractals, *Bull. Belg. Math. Soc. Simon Stevin* **8** (2001), 181-207.
- [10] P. Arnoux and G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, *Bull. Soc. Math. France*, **119**, (1991), 199-215.
- [11] M.F. Barnsley, *Fractals everywhere*, Second Edition, Academic Press Professional, London, 1993.
- [12] P. Billingsley, *Ergodic theory and information*, Wiley, 1965, New York.
- [13] H. Bruin, Dimensions of recurrence times and minimal subshifts. *Dynamical systems (Luminy-Marseille, 1998)*, 117-124, World Sci. Publishing, River Edge, NJ, 2000.

-
- [14] V. Canterini and A. Siegel, Automate des préfixes-suffixes associé à une substitution primitive, *J. Théor. Nombres Bordeaux* **13** (2001), 353–369.
- [15] V. Canterini and A. Siegel, Geometric Representation of primitive substitution of Pisot type, *Trans. Amer. Math. Soc.*, **353** (2001), 5121–5144
- [16] F.M. Dekking, Recurrent Sets, *Adv. Math.* **44** (1982), 78–104.
- [17] J.M. Dumont et A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoretical Computer Science*, **65** (1989), 153–169.
- [18] S. Eilenberg, *Automata, languages and machines*, Academic Press, New York, 1974.
- [19] K. Falconer, *The geometry of fractal sets*, Cambridge University Press, 1985, Cambridge.
- [20] K. Falconer, *Fractal Geometry - Mathematical foundations and applications*, John Wiley & Sons, 1990, Chichester.
- [21] A.S. Fraenkel, H. Porta and K.B. Stolarsky, Some arithmetic semigroups, in *Analytic Number Theory: Proceedings of a Conference in Honor of Paul T. Bateman*, edited by B.C. Berndt et al., Birkhauser: Boston 1990, 255–264.
- [22] C. Frougny, Representations of numbers and finite automata, *Math. Systems Theory*, **25** (1992), 37–60.
- [23] W.H. Gottschalk, Substitution minimal sets, *Trans. Amer. Math. Soc.*, **109**, (1963), 467–491
- [24] P.J. Grabner, P. Liardet and R. Tichy, Odometers and systems of numeration, *Acta Arithmetica*, **70** (1995), 103–123.
- [25] C. Holton and L. Zamboni, Geometric realization of substitutions, *Bull. Soc. Math. France*, **126** (1998), 149–179.
- [26] C. Holton and L. Zamboni, Directed graphs and substitutions. *Theory Comput. Sys.* **34** (2001), 545–564.
- [27] J.E. Hutchinson, Fractals and self-similarity, *Indiana U. Math. J.* **30** (1981), 713–747.
- [28] S. Ito and M. Kimura, On the Rauzy fractal, *Japan J. Indust. Appl. Math.* **8** (1991), 461–486.
- [29] D.E. Knuth, *The art of computer programming*, Vol 2, Seminumerical Algorithms, Addison–Wesley, Reading MA, 1981.
- [30] D.E. Knuth, Fibonacci multiplication, *Appl. Math. Lett.*, **1** (1988), 57–60.
- [31] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, UK, 1995.

- [32] R.D. Mauldin and S.C. Williams, Hausdorff dimension in graph directed constructions, *Trans. of the Amer. Math. Soc.* **309** (1988), 811-829.
- [33] A. Messaoudi, Frontière du fractal de Rauzy et système de numération complexe, *Acta Arith.* **95** (2000), 195-224.
- [34] A. Messaoudi, Tribonacci multiplication. *Appl. Math. Lett.* **15** (2002), no. 8, 981-985.
- [35] P.A.P. Moran, Additive functions of intervals and Hausdorff measure. *Proc. Cambridge Philos. Soc.*, **42**, (1946), 15-23.
- [36] M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.*, **22**, (1921), 84-110.
- [37] M. Morse, *Symbolic Dynamics, Lectures of 1937-38*, Notes by Rufus Oldenburger, The Institute for Advanced Study, Princeton, New Jersey, 1966.
- [38] M. Morse and G. Hedlund, Symbolic dynamics, *Amer. J. of Math.*, **60**, (1938), 286-303.
- [39] M. Morse and G. Hedlund, Symbolic dynamics II: sturmian trajectories, *Amer. J. of Math.*, **62**, (1940), 1-42
- [40] S. Ngai, V.F. Sirvent, J.J.P. Veerman and Y. Wang, 2-Reptiles in the Plane, *Geometriae Dedicata*, **82** (2000), 325-344.
- [41] V. Penné, B. Saussol and S. Vaienti, Dimensions for recurrence times: topological and dynamical properties, *Discrete and Continuous Dynamical Systems*, **5** (1999), 783-798.
- [42] C. Petronio, Thurston's solitaire tilings of the plane. *Rend. Istit. Mat. Univ. Trieste* **26** (1994) 261-295.
- [43] Y.B. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*. Chicago Lectures in Mathematics, The University Chicago Press (1997).
- [44] H. Porta and K.B. Stolarsky, Wythoff pairs as semigroup invariants, *Advances in Math.*, **85** (1991), 69-82.
- [45] B. Praggastis, Numeration systems and Markov partitions from self-similar tilings, *Trans. Amer. Math. Soc.* **351** (1999), 3315-3349.
- [46] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *C.R. Acad. Sci. Paris*, **33**, (1851), 31.
- [47] Pytheas Fogg *Substitutions in Dynamics, Arithmetics and Combinatorics* Editors: V. Berthé, S. Ferenczi, C. Mauduit, et al., Lecture Notes in Mathematics, Vol. 1794, Springer Verlag 2002

-
- [48] M. Queffélec, *Substitution Dynamical Systems -Spectral Analysis*, Lecture Notes in Mathematics, Vol 1294, Springer-Verlag, 1987, Berlin.
- [49] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* **110** (1982), 147-178.
- [50] G. Rauzy, Sequences Defined by Iterated Morphisms, in *Sequences* (Napoles/Positano, 1988), Edited by R.M. Capocelli, Springer, New York, 1990, 275-286.
- [51] G. Rauzy, Rotations sur les groupes, nombres algébriques et substitutions, *Séminaire de théorie des nombres de Bordeaux* **21** (1988), 1-12.
- [52] G. Rauzy, Numbers and automata. *Formal properties of finite automata and applications* (Ramatuelle, 1988), 176-185, Lecture Notes in Comput. Sci., 386, Springer, Berlin, 1989.
- [53] V.F. Sirvent, Properties of Geometrical Realizations of Substitutions associated to a Family of Pisot Numbers, Ph.D. thesis, University of Warwick, 1993.
- [54] V.F. Sirvent, Relationships between the dynamical systems associated to the Rauzy substitutions, *Theoret. Comput. Sci.* **164** (1996), 41-57.
- [55] V.F. Sirvent, A semigroup associated with k-bonacci numbers with dynamic interpretation, *The Fibonacci Quarterly*, **35** (1997), 335-340.
- [56] V.F. Sirvent, On some dynamical subsets of the Rauzy Fractal, *Theoret. Comput. Sci.* **180** (1997), 363-370.
- [57] V.F. Sirvent, Identifications and dimension of the Rauzy Fractal, *Fractals* **5** (1997), 281-294.
- [58] V.F. Sirvent, Semigroups and the self-similar structure of the flipped tribonacci substitution, *Applied Mathematics Letters*, **12** (1999), 25-29.
- [59] V.F. Sirvent, *Modelos geométricos asociados a substituciones*, Trabajo de ascenso, Universidad Simón Bolívar, 1998.
- [60] V.F. Sirvent, Geodesic laminations as geometric realizations of Pisot substitutions, *Ergodic Theory Dynam. Systems* (2000), **20**, 1253-1266.
- [61] V.F. Sirvent, The common dynamics of the tribonacci substitutions, *Bull. Belgian Math. Soc.* **7**, (2000), 571-582.
- [62] V.F. Sirvent, The Arnoux semi-conjugacy is Hölder continuous, *Journal of Mathematical Analysis and Applications* (2001), **259**, 357-367.
- [63] V.F. Sirvent, Spectra of recurrence dimension for adic systems. *Stochastics and Dynamics*, **2** (2002), 599-608.

- [64] V.F. Sirvent, Geodesic laminations as geometric realizations of Arnoux-Rauzy sequences, *Bull. Belg. Math. Soc. Simon Stevin*, **10** (2003), 221-229.
- [65] V.F. Sirvent, Hilbert's Space Filling Curves and Geodesic Laminations, *Mathematical Physics Electronic Journal*, **9** (2003)
- [66] V.F. Sirvent and B. Solomyak, Pure Discrete Spectrum for One-dimensional Substitution Systems of Pisot Type . *Canad. Math. Bull.*, **45** (2002), 697-710.
- [67] V.F. Sirvent and Y. Wang, Self-Affine Tiling Via Substitution Dynamical Systems and Rauzy Fractals. *Pacific Journal of Mathematics*, **206** (2002), 465-485.
- [68] B. Solomyak, A dynamical system with a discrete spectrum (Russian), *Uspekhi Mat. Nauk* **41** (1986), 209-210. English translation: *Russian Math. Surveys* **41** (1986), 219-220.
- [69] B. Solomyak, Substitutions, adic transformations, and beta-expansions, *Symbolic dynamics and its applications*(New Haven, CT, 1991), 361-372, *Contemp. Math.*, **135**, AMS, Providence, 1992.
- [70] B. Solomyak, *On the spectral theory of adic transformations*, *Advances in Soviet Mathematics*, **9** (1992), 217-230.
- [71] B. Solomyak, Dynamics of self-similar tilings, *Ergodic Theory Dynam. Systems*, **17**, (1997), no. 3, 695-738.
- [72] W. Thurston, Groups, tilings, and finite state automata, *AMS Colloquium Lecture Notes*, unpublished, 1889.
- [73] A. Thue, Uber unendliche Zeichenreihen (1906), *Selected mathematical papers of Axel Thue*, Universitetsforlaget, 1977.
- [74] A.M. Vershik, Uniform algebraic approximation of shift and multiplication operators, *Soviet. Math. Dokl.*, **24** (1981), 97-100.
- [75] A.M. Vershik, Arithmetic Isomorphism of Hyperbolic Toral Automorphisms and Sofic Shifts, UDC 519.56, 517.919.
- [76] E. Zeckendorf, Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas, *Bull. Soc. Roy. Sci. Liège* **3-4**, 1972, 179-182.

VÍCTOR SIRVENT
DEPARTAMENTO DE MATEMÁTICAS,
UNIVERSIDAD SIMÓN BOLÍVAR,
APARTADO 89000, CARACAS 1086-A, VENEZUELA.
sirvent@ma.usb.ve

ENSAYO

Determinismo, indeterminismo y la flecha del tiempo en la ciencia contemporánea

Daniel A. Morales

1 Introducción

La ciencia ha sido muy efectiva en su afán de entender el universo y la posición del hombre en éste. Lo increíble es que podamos lograr esto sin movernos prácticamente de nuestro propio planeta. Nuestro principal instrumento para lograr esta proeza es la mente humana. El entender cómo la mente es capaz de establecer ese vínculo entre el comportamiento del mundo exterior y nuestros conceptos es uno de los grandes retos futuros de la ciencia. Desde el punto de vista filosófico este problema está relacionado con las ideas de realismo e idealismo que discutiremos más adelante. Estas ideas filosóficas también se hacen evidentes en las teorías actuales de la ciencia.

Comenzaremos discutiendo algunas nociones fundamentales para el resto del ensayo. En primer lugar, ¿qué significan las ideas de complejidad y de simplicidad? Estas son nociones fundamentales en la ciencia. Podemos preguntarnos ¿qué es más complejo, un insecto o un ser humano?, o ¿qué es más complejo, un ser humano o el universo como un todo? Los logros antiguos y recientes de la ciencia demuestran que mientras que podemos entender el universo como un todo, se nos dificulta entender los aspectos más básicos del comportamiento humano. De manera que lo más importante para considerar un sistema como simple o complejo no puede ser solamente el número de partículas que lo componen. Esta diferencia ha sido expuesta elocuentemente por uno de los pensadores actuales más importantes como Murray Gell-Mann en su libro *el Quark y el Jaguar* [15]: ¿qué es más simple un quark, la partícula última constituyente de la materia, o un jaguar? ¿Podemos cuantificar las nociones de simplicidad y de complejidad, de tal manera que podamos afirmar inequívocamente que un jaguar es más complejo que un quark, o a la inversa?

La física hasta ahora ha tratado esencialmente con sistemas simples, la razón es que los sistemas simples pueden ser estudiados más fácilmente usando la matemática. C.N. Yang, uno de los físicos teóricos más famosos del siglo XX apunta que [38]:

Mientras más simple es el problema, es más seguro que el análisis esté más cercano a alguna estructura matemática básica.

Además, a la física le interesa los aspectos esenciales del universo y no los detalles. Esta manera de atacar los problemas ha sido muy fructífera y la física actual puede mostrar muy orgullosamente un conjunto de leyes de la naturaleza. En última instancia estas leyes se representan matemáticamente y no verbalmente. Aunque, al final cuando comunicamos estas leyes a los miembros de “la otra cultura” (las humanidades) como la llamó C.P. Snow (1905-1980) [29], debemos hacerlo verbalmente o por escrito.

Muchos aceptan que una ciencia es más fundamental en la medida en que sea más matematizable, afirmación ésta que podríamos designar como el sueño pitagórico o la “falacia jonia” como lo llama Isaiah Berlin (1909-1997), asociando esta idea con los filósofos presocráticos y la afirmación de que el mundo es ordenado y que puede ser explicado por leyes matemáticas que rigen el comportamiento de simples entidades [19]. De todas las ciencias naturales es la física la que más se acerca a esta premisa. El lenguaje matemático es el instrumento más poderoso de que dispone el hombre para entender el universo en el que vive. Albert Einstein (1879-1955) en su libro *Mi Visión del Mundo* escribe [13]:

Según nuestra experiencia estamos autorizados a pensar que la Naturaleza es la realización de lo matemáticamente más simple. Creo que a través de una construcción matemática pura es posible hallar los conceptos y las relaciones que iluminen una comprensión de la Naturaleza. Los conceptos usables matemáticamente pueden estar próximos a la experiencia, pero en ningún caso pueden deducirse de ella. Está claro que la experiencia es el único criterio que tiene la Física para determinar la utilidad de una construcción matemática. Pero el principio creativo se encuentra en realidad en la matemática. De algún modo creo que es cierto que a través del pensamiento puede comprenderse la realidad, tal como lo soñaron los antiguos.

El éxito que ha tenido la física en entender el universo usando la matemática es una evidencia de lo que Eugene Wigner (1902-1995) designó como la irracional efectividad de la matemática en las ciencias naturales [36]:

El milagro de lo apropiado del lenguaje matemático para la formulación de las leyes de la física es un regalo maravilloso que no entendemos ni merecemos.

La complejidad de la existencia y del conocimiento humano ha traído como consecuencia la creación de parcelas del conocimiento. Esta parcelación ha provocado el aislamiento entre las diferentes ramas del saber. Por esta separación y jerarquización del conocimiento ya no es posible para una sola persona poseer y estar al tanto de todo lo que ocurre en las diversas ramas del conocimiento. Pero puede ser que esta jerarquización sea simplemente ilusoria: a la naturaleza no le importa como el hombre ha clasificado el conocimiento.

Para tratar de entender la jerarquía del conocimiento han surgido dos corrientes filosóficas: el reduccionismo y la emergencia. ¿Puede la sociología o la psicología ser reducida a las ciencias físicas? En general, ¿puede una rama del conocimiento considerada como poco matematizada ser reducida a las leyes de otra más matematizada?, o, por el contrario, existen leyes emergentes, es decir leyes presentes en la rama inferior de la jerarquía que nunca van a ser obtenidas como consecuencia de las leyes presentes en las ramas superiores de la jerarquía? ¿Son estas leyes emergentes consecuencia de la complejidad de los sistemas con los que trabaja la rama inferior?

Otro aspecto relacionado con la discusión anterior lo constituye la noción de evolución. El pasaje de un organismo simple a uno complejo es una transformación en el tiempo que es irreversible, es decir, no observamos el caso contrario. Por otra parte, nuestra vida transcurre del nacimiento a la muerte de manera irremediable y nunca observamos una reversión de este hecho. Si tiramos un vaso de vidrio al piso éste se romperá en muchos pedazos y nunca observamos que el vaso de manera espontánea se reconstruya a sí mismo. Si nosotros miramos una película proyectada al revés observamos claramente que el transcurrir de las secuencias no es natural. En otras palabras, los fenómenos macroscópicos que observamos revelan claramente una flecha del tiempo, dirigida del pasado hacia el futuro [27,28]. Sin embargo, las leyes de la física son reversibles en el tiempo. Las ecuaciones matemáticas que expresan las leyes físicas no cambian cuando el parámetro tiempo t presente en estas ecuaciones es sustituido por el parámetro menos tiempo $-t$. Esto significa que las ecuaciones fundamentales de la física no distinguen entre el pasado y el futuro. Nos encontramos entonces ante una paradoja. Si consideramos a las leyes físicas como una síntesis del comportamiento del universo, entonces porqué estas leyes son indiferentes al sentido del tiempo, mientras que nuestra apreciación de los fenómenos naturales diariamente nos dice que existe claramente una flecha del tiempo. Los fenómenos observados en nuestro universo, sea en química, geología, biología, psicología o sociología muestran una dirección preferencial en su dinámica, una flecha del tiempo.

Ahora, ¿cómo puede surgir una flecha del tiempo de ecuaciones completamente simétricas? Este problema es conocido como la *paradoja temporal*. Fue identificado en el siglo XIX por el físico vienés Ludwig Boltzmann (1844-1906) al tratar de formular para la física una teoría de la evolución de la misma manera como Charles Darwin (1809-1882) lo había hecho para la biología. Sin embargo los intentos de Boltzmann tuvieron que luchar contra las leyes de Newton, el paradigma de determinismo y objetividad en la física. Posteriormente, la flecha del tiempo fue relegada al terreno de la fenomenología: se acepta que la flecha del tiempo es consecuencia de la incapacidad de la mente humana para percibir la totalidad de una sola vez, de manera que la flecha del tiempo aparece como consecuencia de las aproximaciones que nuestra mente introduce en la descripción

de la naturaleza. Sin embargo, hay varios intentos recientes para formular leyes del movimiento que describan la dinámica de procesos irreversibles en el tiempo [26].

Conectado al problema de la existencia o no de una flecha del tiempo real está la pregunta de si nuestro futuro está determinado por nuestro momento presente o, si por el contrario, no hay forma de predecir el futuro a partir del presente. El sentido común nos indica que hay algo de verdad en esta premisa: si cometemos un acto antisocial ahora, es bastante probable que en el futuro paremos en la cárcel. Pero, si esto fuera así en general, ¿podríamos predecir el resultado al girar una ruleta, simplemente mirando como el crupier tira la pelotita?

A comienzos del siglo XX el gran filósofo de la ciencia Karl Popper (1902-1994) escribió [25]:

El sentido común nos inclina, por una parte, a asegurar que cada evento es causado por algunos eventos precedentes, de manera que cada evento puede ser explicado o predicho. . . . Por otra parte. . . el sentido común atribuye a las personas maduras y sanas. . . la habilidad para escoger libremente entre posibilidades alternativas de actuación.

Los aspectos delineados anteriormente son sólo algunos de los problemas fundamentales a los que se enfrenta la ciencia actual. He querido englobar todos los problemas anteriores bajo el tema del determinismo e indeterminismo y la flecha del tiempo en la ciencia contemporánea. Estos aspectos han estado muy interrelacionados en las culturas humanas: ¿poseo libre albedrío para seleccionar mi futuro o por el contrario cualquier cosa que haga no será capaz de cambiar mi futuro que ya ha sido preestablecido desde mi nacimiento? Esta discusión también está conectada con la llamada ciencia del caos, muy de moda en la actualidad. Queremos ahora discutir estos aspectos con mayor profundidad.

2 Los griegos y el tiempo

Los problemas discutidos anteriormente no son nuevos, en el sentido de que preguntas similares ya se la hacían filósofos de la antigüedad. En particular quiero empezar discutiendo estas nociones en los griegos. El gran problema de la filosofía griega era el problema del ser y el devenir: de qué estaban formadas las diversas cosas que observamos a nuestro alrededor y si nuestra noción del pasaje del tiempo era real o tan sólo una ilusión o, en otras palabras, el problema fundamental era la naturaleza de la materia y del tiempo. Diversas teorías exponían que las cosas estaban formadas por uno o todos de cuatro elementos: agua, aire, fuego o tierra, y que fue la unión de estos cuatro elementos en diferentes proporciones y en diferentes organizaciones lo que dio origen a nuestro

universo. Asociado a la composición de las cosas estaba la noción de devenir, el decaimiento y muerte de las cosas. Filósofos como Heráclito (540-475 a.C) [18] sostenían que lo real era lo que se percibía por medio de los sentidos y estos nos indicaban que las cosas cambiaban, se transformaban:

A los que se bañan en los mismos ríos, aguas distintas cada vez fluyen, y las almas son exhaladas de la humedad.

Mientras que filósofos como Parménides (515-450 a.C) [18] sostenían que nuestros sentidos nos engañan y nos dan una idea ilusoria del devenir. Lo que es real es lo que es permanente, inmanente y que no cambia; el cambio es ilusorio:

Y del camino ya sólo queda un solo relato: que es. Y para éste hay muchísimas señales de que lo que es no engendrado e imperecedero. Pues es integro, inmóvil e infinito ni fue ni será, pues es ahora todo al mismo tiempo, uno solo, continuo.

Albert Einstein sostenía que el paso del tiempo era tan sólo una ilusión, razón por la cual Popper solía llamarlo “Parménides”. Entre los documentos más importantes que tenemos sobre las ideas filosóficas sobre el tiempo y la irreversibilidad son las cartas que Einstein intercambió con su amigo de toda la vida Michele Besso. Besso siempre insistía en interrogar a Einstein sobre el problema del tiempo, la irreversibilidad y su conexión con las leyes de la física. Einstein le respondía una y otra vez que la irreversibilidad era una ilusión, una impresión suscitada por unas condiciones iniciales improbables. No hay irreversibilidad en las leyes fundamentales de la física. Debes aceptar la idea de que el tiempo es subjetivo, con su insistencia sobre el “ahora” no debe tener ninguna significación objetiva. Cuando Besso muere, Einstein le escribe a la hermana e hijos de éste [12]:

Michele se me ha anticipado en dejar este mundo extraño. Esto nada significa. Para nosotros, físicos creyentes, la distinción entre pasado, presente y futuro es solo una ilusión, por persistente que ésta sea.

En esta búsqueda por la composición fundamental de la materia, la idea más ingeniosa y que parcialmente encuentra un asidero experimental en nuestra ciencia actual, es aquella expuesta por Demócrito (460-371 a.C) y Leucipo (430 a.C), de que todo lo existente (el alma incluida) puede concebirse como formada por entidades microscópicas, indivisibles, llamadas átomos y el vacío. Estas ideas son expuestas posteriormente de manera elocuente e ilustrativa en el gran poema de la antigüedad “De rerum natura” (De la naturaleza de las cosas) del gran poeta latino Tito Lucrecio Caro (95-55 a.C) [37]. Otro filósofo de la escuela atomística cuyas ideas quiero discutir en relación con la flecha del

tiempo es Epicuro (341-270 a.C). Éste, como representante de la escuela atomista, aceptaba la existencia de los átomos y el vacío. Pero esta escuela también asumía que los átomos se movían en trayectorias rectas y paralelas, a lo cual Epicuro se oponía. ¿Cómo se podía explicar la existencia de la diversidad en nuestro mundo si no se producían choques entre los átomos? ¿Cómo se puede explicar el libre albedrío de los seres humanos en un mundo completamente determinístico y sin azar? Para resolver este dilema Epicuro introduce la idea del *clinamen* (declinación): las trayectorias de los átomos no son siempre paralelas sino que algunas veces se desvían de estas trayectorias y chocan produciendo la gran diversidad de cosas que observamos en nuestro mundo. Como lo expone Lucrecio en su *De rerum natura* [37]:

En momentos inciertos y en lugares inciertos, ellos (los átomos) se desvían ligeramente de sus cursos, justamente lo suficiente como para que consideremos que han cambiado su dirección.

Sin embargo, Epicuro no proporciona ningún mecanismo por medio del cual el *clinamen* actúa.

Así, arribamos a los dos más grandes pensadores de la antigüedad, Platón (427-347 a.C) [23] y Aristóteles (384-322 a.C) [1] cuyas visiones filosóficas sobre el universo siguen vigentes y son punto de discusión en las teorías más actuales de la física. Platón sostenía que el mundo que observamos no representa sino una imagen imperfecta de un mundo más perfecto que es el mundo de las ideas inmanentes. En ese mundo se encuentran presentes en estado latente todas las realidades posibles que se pueden presentar en el universo, los arquetipos de las cosas visibles. Aristóteles, por el contrario, sostenía que la realidad es el mundo que observamos, lo que es real es el devenir, la transformación de las cosas, todo lo que observamos con nuestros sentidos. No podemos pensar como real otra cosa que no sea lo que observamos a nuestro alrededor. Un concepto importante en Aristóteles es el de potencia y acto. Potencia es la posibilidad latente que tiene un cuerpo para transformarse en otro, en acto, en realidad.

Estas ideas filosóficas de Platón y Aristóteles han sido asociadas a los puntos de vista del pájaro y la rana. El pájaro por su capacidad de volar tiene una visión más amplia de la realidad y puede tener acceso a realidades que la rana no tiene. La rana, por el contrario, al estar condicionada a moverse y tener acceso a una realidad más limitada que el pájaro no admite como real nada a lo que no pueda tener acceso. En la actualidad ambos puntos de vista han sido asociados con las visiones de muchos teóricos la de los experimentales .

Los puntos de vista de Platón y Aristóteles nos han acompañado aún en la ciencia actual como veremos más adelante.

3 El tiempo y el determinismo

Se considera que la ciencia contemporánea, basada en el llamado método científico, comienza con las ideas de Galileo (1564-1642), Kepler (1571-1630) y Newton (1642-1727). Con Newton se origina la verdadera revolución científica. Newton establece los principios físico-matemáticos que nos permiten entender el movimiento de los cuerpos: sus tres leyes de la mecánica. Estas leyes constituyen el núcleo central de lo que se conoce como mecánica clásica, el cuerpo de teorías desarrolladas por los físicos desde mediados del siglo XVII hasta finales del siglo XIX. La premisa fundamental de la mecánica clásica es que la materia consiste de un conjunto de partículas con masa y carga y que las posiciones instantáneas y velocidades de esas partículas se pueden determinar experimentalmente con precisión arbitraria. En este sentido, uno asocia con los sistemas físicos un conjunto de variables dinámicas que poseen en cada instante valores bien definidos que determinan el estado del sistema. La evolución futura del sistema queda completamente definida dando los valores de estas variables en el estado presente. La evolución en el tiempo de las variables que caracterizan al estado del sistema se determina entonces resolviendo un conjunto de ecuaciones de movimiento.

La discusión anterior nos permite formular una característica fundamental de la mecánica clásica: dada la función de fuerza para un sistema físico, el estado mecánico del sistema en cualquier otro instante estará completa y unívocamente especificado o determinado por el estado mecánico a algún instante inicial arbitrario. Es esta característica de las ecuaciones de movimiento la que define a la mecánica clásica como una teoría determinista. Esta característica aparece reflejada en un pasaje de Pierre Simon Laplace (1749-1827) muy citado [20]:

Debemos considerar el estado presente del universo como una consecuencia de su estado previo y como la causa del estado que seguirá. Una inteligencia que conozca todas las fuerzas que actúan en la naturaleza en un instante dado así como también las posiciones momentáneas de todas las cosas en el universo, sería capaz de comprender en una sola fórmula los movimientos de los cuerpos más grandes como también los de los átomos más ligeros en el mundo, con tal de que su intelecto fuera lo suficientemente poderoso para someter todos los datos al análisis; para ella nada sería incierto, tanto el futuro como el pasado estarían presentes simultáneamente ante sus ojos. La perfección que la mente humana ha sido capaz de darle a la astronomía proporciona una indicación de tal inteligencia. Descubrimientos en mecánica y geometría, acoplados con aquellos en gravitación universal, han traído a la mente dentro del alcance de comprender en la misma fórmula analítica el estado pasado y futuro del sistema del mundo. Todos los esfuerzos de la mente en la

búsqueda de la verdad tienden a aproximarnos a la inteligencia que hemos imaginado, aunque ella (la mente) permanecerá infinitamente remota de tal inteligencia.

La opinión anterior está muy ligada al concepto de tiempo. ¿Está el futuro escrito o en continua construcción?, ¿están nuestros actos determinados desde el principio o poseemos libre albedrío para escoger nuestro futuro?

4 Indeterminismo, probabilidad y mecánica cuántica

Durante el siglo XX aparecen dos grandes teorías de la física: la teoría de la relatividad y la mecánica cuántica. La teoría de la relatividad trata con fenómenos que se dan a altas velocidades mientras que la mecánica cuántica trata con el comportamiento de los entes microscópicos que constituyen la materia. Aquí nos concentraremos en las ideas fundamentales de la mecánica cuántica.

Experimentos que se hicieron al final del siglo XIX y en el primer cuarto del siglo XX indicaban que la mecánica clásica no era aplicable a los átomos. El comportamiento del mundo microscópico era tal que ameritaba la deducción de nuevas leyes a ese nivel. El desarrollo fundamental lo realizaron, entre otros, N. Bohr (1885-1962), W. Heisenberg (1901-1976), L. de Broglie (1892-1987) y E. Schrödinger (1887-1961). Heisenberg desarrolló su mecánica de matrices dentro del marco filosófico del pensamiento positivista de E. Mach (1838-1916) que establecía que cualquier teoría explicativa de la naturaleza debía contener solamente elementos que tuvieran una correspondencia con la realidad, es decir cualquier elemento teórico que no fuera observable no podía incluirse en la formulación de la teoría. Estas son extensiones de las ideas idealistas de Berkeley (1685-1753) [2]. Otra manera alternativa fue desarrollada por Schrödinger, basado en una idea de L. de Broglie. Este último decía que así como la luz poseía un comportamiento dual de onda y partícula (fotón) así también las partículas microscópicas, como los electrones, poseían un comportamiento dual, de partícula y onda: allí donde van las ondas van las partículas.

De Broglie entonces postuló que, debido al comportamiento ondulatorio de los entes microscópicos, le podíamos asociar una longitud de onda que era inversamente proporcional a la masa de la partícula. Así, los electrones que tienen una masa muy pequeña poseen una longitud de onda apreciable, y, por el contrario, partículas macroscópicas, tales como una pelota de béisbol, que poseen masas grandes no tienen comportamiento ondulatorio sino de partícula. Schrödinger toma esta idea de De Broglie y la combina con la ecuación clásica de las ondas para encontrar una ecuación diferencial para las ondas de materia, $\hat{H}\Psi = E\Psi$, donde \hat{H} es un operador hermítico y E son sus autovalores. Esto último introduce en la teoría física un elemento no observable: la función de

onda Y , que básicamente representa la amplitud de la onda de materia, en general un número complejo. Note las diferencias filosóficas de las dos teorías. La de Heisenberg, idealista: la teoría sólo contiene los elementos que podemos observar, todo lo demás no tiene existencia real, y la de Schrödinger: la teoría contiene elementos que no podemos medir pero que son necesarios para explicar eventos que sí son medibles.

Ahora, ¿cuál era el significado de esta “función de onda”? Esta pregunta todavía en la actualidad constituye uno de los acertijos fundamentales de la mecánica cuántica.

M. Born (1882-1970) estableció que ya que las partículas tienen comportamiento de ondas no es posible determinar con certeza sus posiciones. Lo único que podemos calcular es el cuadrado de la función de onda, lo cual representa la probabilidad de encontrar la partícula en una cierta posición del espacio. Esta interpretación introduce un aspecto de aleatoriedad en la mecánica cuántica y sugiere que aspectos de indeterminismo están presentes en las leyes de la naturaleza. Esta situación no le gustaba nada a Einstein, un convencido determinista, quién entonces expresa que “no creía que Dios jugara a los dados con el universo”. Para entender esto debemos recordar que Einstein creía en el Dios de Spinoza (1632-1677), un dios identificado con la naturaleza, donde el azar no es posible, lo que nos parece contingente es solo una ilusión; solo nos parece contingente porque no conocemos su causa ni su mecanismo.

A pesar de ello, de acuerdo a la teoría cuántica, y recordando las palabras de P. Dirac (1902-1984), toda la química y gran parte de la física podía ser calculada a partir de primeros principios sin necesidad de hacer más experimentos. La estructura y propiedades de cualquier sustancia desconocida podían ser predichas con gran precisión, y la velocidad, mecanismo y productos de cualquier proceso físico o químico desconocidos podían ser determinados. Lo único que se requería era resolver las ecuaciones de la mecánica cuántica.

Las aplicaciones de la mecánica cuántica han transformado nuestras vidas. Se estima que el 30% del producto territorial bruto de los E.E.U.U. se basa en invenciones desarrolladas gracias a la mecánica cuántica, desde semiconductores en los chips de las computadoras hasta los rayos láser utilizados en los lectores de discos compactos y DVDs, aparatos de resonancia magnética nuclear en los hospitales y celulares, entre otros.

Ahora quiero discutir el aspecto de indeterminismo y probabilidad en la mecánica cuántica. Uno de los principios fundamentales de esta teoría lo constituye el principio de indeterminación de Heisenberg que éste formuló en 1927. De acuerdo con este principio las ecuaciones de la mecánica cuántica restringen la precisión con la cual un observador puede medir simultáneamente ciertas cantidades, por ejemplo la posición y la velocidad de un electrón. Heisenberg estableció que aunque las leyes de la mecánica cuántica contenían elementos estadísticos en su formulación, estos elementos no eran propiedad de la misma

naturaleza sino que aparecían como consecuencia de la perturbación causada cuando el observador quería observar la naturaleza. Por ejemplo, la determinación de la posición de un electrón de manera precisa requiere la utilización de radiación electromagnética de alta frecuencia, lo que implica alta energía. En el proceso de colisión de un fotón de esta radiación con un electrón producirá un cambio brusco en el momento del electrón. Este cambio será mayor en la medida en que la frecuencia de la luz se haga cada vez mayor. Por otra parte, la determinación de la posición del electrón con poca precisión requerirá luz de baja frecuencia, esto es, baja energía. En este último caso la colisión de un fotón con un electrón no provocará un cambio muy importante en el momento de este último, de manera que éste puede ser medido con gran precisión. Este ejemplo ilustra el hecho de que existe una relación dual o complementaria entre la posición con que se mide el momento de un electrón y la precisión con que se determina su posición. Si x es la precisión con que se mide la posición de una partícula de características microscópicas y p es la precisión en la medida de su momento, entonces el principio de indeterminación establece que el producto de estas dos cantidades no puede ser menor que $h/2p$, esto es $\Delta x \Delta p \geq h/2p$, donde h es la constante de Planck. El principio de indeterminación tiene importantes y profundas consecuencias sobre el principio de causalidad. Este principio, uno de los más queridos y protegidos por los físicos del siglo XIX, establece que los efectos siempre preceden a sus causas. Este principio está entonces en la base del determinismo de Newton y Laplace.

El principio de indeterminación niega el principio de causalidad. Como dice Heisenberg [6]:

En la formulación estricta de la ley causal -si conocemos el presente, podemos calcular el futuro- no es la conclusión la que está errada sino la premisa.

Los valores iniciales de la posición y de la velocidad no pueden ser medidos simultáneamente con absoluta precisión, de manera que uno no puede calcular un futuro único para la partícula sino un conjunto de posibles futuros. Sólo la medición determinará que futuro se observará en la realidad. De esta manera la conexión causal entre presente y futuro se pierde y es sustituida sólo por probabilidades de los eventos posibles.

Estas ideas también se pueden discutir desde el punto de vista de la ecuación de Schrödinger, que es una ecuación simétrica en el tiempo y además determinista con respecto a su elemento principal, la función de onda. La función de onda contiene en potencia, en el sentido de Aristóteles, todas las posibilidades futuras del sistema bajo estudio. Aunque la evolución en el tiempo de la función de onda es perfectamente determinista: dada la función de onda en un instante dado se puede calcular exactamente la evolución en el tiempo de la función de onda, la ecuación de Schrödinger no nos puede decir con certeza cuál de

todas las posibilidades se convertirá en acto, en el sentido de Aristóteles, en el futuro. De esta manera, la irreversibilidad temporal en la mecánica cuántica entra como consecuencia del observador. Es la persona que realiza la medición la que determina cuál de entre todas las posibilidades se convertirá en acto.

Sin embargo, a pesar de que la mecánica cuántica nos ha permitido entender muchos de los aspectos del mundo microscópico, todavía existen aspectos que no entendemos. Uno de los problemas de la mecánica cuántica es que es una teoría que posee una ecuación determinista, la ecuación de Schrödinger, y un conjunto de reglas ad hoc que nos dicen como calcular las probabilidades de las diferentes posibilidades asociadas con un determinado evento. Sin embargo, a pesar de todo el tiempo transcurrido desde la creación de la mecánica cuántica hasta la actualidad, todavía nadie ha deducido una ecuación que nos diga con certeza cuál de las diferentes posibilidades se dará en la realidad.

Ya que de acuerdo a la mecánica cuántica, los electrones se comportan como ondas, ellos presentan las características que asociamos con las ondas. En particular, las funciones de onda describen combinaciones de diferentes estados, las llamadas superposiciones. Un electrón, por ejemplo, podría estar en una superposición de estados correspondientes a diferentes posiciones. Schrödinger afirmaba que este principio de superposición debía aplicarse también a los objetos macroscópicos ya que estos estaban formados por átomos. Un experimento mental muy conocido que ilustra este hecho es el experimento del gato encerrado en una caja. Supongamos que colocamos un gato en una caja herméticamente cerrada, con oxígeno para que el gato pueda respirar, y una muestra de un elemento radiactivo cuya emisión desencadena que un gas venenoso encerrado en un recipiente se libere y mate al gato. Como la emisión o el decaimiento del elemento radiactivo es aleatorio, la función de onda del átomo radiactivo es una superposición del estado que decae y el estado que no decae, y por lo tanto, produce un gato que está al mismo tiempo vivo y muerto. Sólo cuando una persona abre la caja la superposición se borra quedando un solo estado: el gato muerto o el gato vivo.

5 El reduccionismo matemático

Otro aspecto importante está relacionado con la aplicación de la mecánica cuántica a energías y distancias muy pequeñas. En física se trabaja en construir una teoría general de donde todo surja como consecuencia de los principios establecidos por esta teoría. Esta gran teoría, llamada teoría de unificación y en la cual el mismo Einstein estuvo trabajando gran parte de su vida sin lograr obtenerla, representa la gran ambición en el siglo XXI. Para lograr esto debemos construir una gran teoría que combine la teoría de la relatividad general de Einstein con la mecánica cuántica. A partir de esta teoría general se debería poder

construir todo el árbol jerárquico del conocimiento humano. En cada rama de ese árbol aparecen conceptos necesarios para entender aspectos relacionados con esa rama particular, tales como protones, electrones, átomos, moléculas, células, organismos, sociedades, culturas. La última gran teoría de la física no debe contener conceptos, ya que ello implicaría que deberíamos construir otra teoría más fundamental para derivar esos conceptos. La última teoría debería ser matemática, sin palabras ni conceptos y sin el parámetro tiempo, es decir matemática y atemporal a partir de la cual cualquier científico pueda derivar toda la existencia del universo [32,33].

Muchos no creen que lo anterior sea posible. Estos creen que en cada rama de la jerarquía del conocimiento aparecen leyes y propiedades que dependen de la complejidad de los sistemas estudiados en esa rama y que no son derivables de los conceptos de la rama inmediatamente sobre esta. Por ejemplo, un zoólogo tan eminente como Ernst Mayr arguye en su libro *Thisis Biology*[21] que las ciencias físicas no pueden explicar muchos aspectos relacionados con la vida y que muchos aspectos estudiados por los físicos no son relevantes para el mundo viviente, o para cualquier otra ciencia fuera de la física. Los organismos vivos deben ser entendidos a cada nivel de organización; no pueden ser reducidos a las leyes de la física y la química. Más adelante escribe [21]:

La unidad de la ciencia no podrá alcanzarse hasta que se acepte que la ciencia contiene un conjunto de provincias separadas, una de las cuales es la física, otra de las cuales es la biología. Sería fútil tratar de “reducir” la biología, una ciencia provincial, a la física, otra ciencia provincial, o viceversa.

Al comentar sobre las ideas de C.P. Snow expuestas en su libro *The Two Cultures* sobre la separación entre los científicos y los humanistas apunta que [29]:

La brecha entre la física y las humanidades, como él (Snow) correctamente apunta, es en verdad virtualmente infranqueable. Simplemente no existe un camino de la física a la ética, la cultura, la mente, el libre albedrío y otras preocupaciones humanísticas. La ausencia en la física de estos importantes tópicos contribuyó a la alienación de los científicos y los humanistas que Snow reclama. Pero, todas estas preocupaciones tienen relaciones substanciales con las ciencias de la vida.

Otro opositor a la teoría reduccionista es Roald Hoffmann, ganador del Premio Nobel de Química. En este sentido él comenta [11]:

... Un gran defecto de los físicos teóricos es su tendencia al reduccionismo: creen que todo en la naturaleza puede reducirse a unos

cuantos principios y partículas simples. “Ni la naturaleza ni la vida funcionan de ese modo en la realidad”. “La complejidad, y no la simplicidad, es la esencia de la vida”.

Vemos, de las dos posiciones anteriores, que el problema de la reducción y la emergencia en la ciencia no está resuelto. Estas posiciones apuntan al hecho de que cuando los sistemas adquieren cierta complejidad otras leyes aparecen. Y es precisamente en este terreno que otro tipo de indeterminación, ahora al nivel clásico, emerge: el caos.

6 El caos

De acuerdo al diccionario Larousse la palabra caos viene del griego khaos que significa abismo. En filosofía expresa la confusión general de los elementos y de la materia, antes de la creación del mundo. Este estado de confusión original ha sido plasmado en diversas obras de la antigüedad. A continuación transcribo un párrafo muy hermoso de Las Metamorfosis de Ovidio (43 a.C-17 d.C) [22], obra que indudablemente presenta elementos en su composición que reflejan las teorías que sobre el universo tenían los antiguos griegos:

Antes de existir el mar, la tierra y el cielo, continentes de todo, existía el Caos. El Sol no alumbraba aún el mundo. La Luna todavía no estaba sujeta a sus vicisitudes. La Tierra no se hallaba todavía suspensa en el vacío, o tal vez quieta por su propio peso. No se conocían las márgenes de los mares. El aire y el agua se confundían con la tierra, que todavía no se había solidificado. Toda era informe. Al frío se oponía el calor. Lo seco a lo húmedo. El cuerpo duro se hundía en el blando. Lo pesado era ligero a la vez. Dios, o la Naturaleza, puso fin a estos despropósitos, y separó al cielo de la tierra, a ésta de las aguas y al aire pesado del cielo purísimo. Y así, el Caos dejó de ser. Dios puso a cada cuerpo en el lugar que le correspondía y estableció las leyes que habían de regirlos. El fuego, que es el más ligero de los elementos, se ubicó en la región más elevada. Más abajo, el aire. La Tierra, encontrada en su equilibrio, la más profunda . . .

Algunas de las ideas sobre el determinismo y el caos ya habían sido anticipadas en 1903 por el gran matemático francés Henri Poincaré en su libro *The value of Science* [24]:

Si conociéramos con precisión infinita las leyes de la naturaleza y la situación inicial del universo, podríamos predecir exactamente la

situación de este mismo universo en un momento posterior. Pero incluso aunque las leyes naturales no tuvieran ningún secreto para nosotros, sólo podríamos conocer la situación inicial de modo aproximado. Todo lo que necesitamos para poder decir que un fenómeno ha sido predicho y que está regido por leyes es poder predecir la situación posterior con la misma aproximación que la inicial. Pero esto no siempre es posible; puede ocurrir que las pequeñas diferencias en las condiciones iniciales se hagan muy grandes en el resultado final. Un pequeño error al principio producirá un error enorme al final. La predicción se hace imposible y tenemos un fenómeno fortuito.

Murray Gell-Mann apunta [15]:

Dado que nada puede medirse con la precisión absoluta, el caos da origen a una indeterminación efectiva en el nivel clásico que se superpone a la indeterminación cuántica. La interacción entre estas dos clases de impredecibilidad es un aspecto fascinante y todavía poco estudiado de la física contemporánea.

Robert May, uno de los padres de la teoría del caos, define así esta ciencia [14]:

En la ciencia, el caos se refiere a la idea de que el comportamiento de algo puede ser a pesar de todos los intentos y propósitos impredecible aún cuando sea descrito por una “ecuación determinista” muy simple; por determinista queremos decir que las ecuaciones, y todos los parámetros en ella son completamente conocidos, sin elementos estadísticos o inciertos. Tal ecuación parece predecir con certeza el futuro de algo, dado su estado a algún momento inicial.

Un aspecto esencial del sistema caótico es que las ecuaciones que rigen su comportamiento son no lineales, de manera que un pequeño cambio en las condiciones iniciales de un experimento, o una situación en el mundo real, puede provocar un gran cambio en el comportamiento futuro del sistema [30]. Esto significa que si conducimos el experimento dos veces a partir de condiciones iniciales ligeramente diferentes, a medida que el tiempo transcurre las diferencias entre el comportamiento del sistema en las dos situaciones se vuelve tan diferente que es imposible hacer ninguna predicción. Esta situación es conocida como “el efecto mariposa”; el aleteo de una mariposa en Brasil podría provocar al cabo de unos meses un tornado en Texas [16]. La única manera entonces de hacer predicciones es conociendo las condiciones iniciales del sistema con absoluta certeza, algo imposible. En el caso de sistemas lineales, dos sistemas que

partan de condiciones iniciales ligeramente diferentes no difieren mucho en su comportamiento a medida que el tiempo transcurre.

Los efectos no lineales que aparecen en sistemas caóticos son típicos de sistemas complejos: por ejemplo el comportamiento de un cardumen en el tiempo, el comportamiento de poblaciones y culturas mundiales, el comportamiento de la economía mundial, la actividad eléctrica en nuestros cerebros y corazón, el congestionamiento en las autopistas, todos estos son ejemplos de sistemas reales cuyo comportamiento temporal puede ser estudiado con la teoría del caos.

A diferencia de los físicos fundamentales que quieren descubrir las leyes fundamentales de la naturaleza y sus restricciones, los científicos que trabajan en el campo de la teoría del caos y la complejidad utilizan ecuaciones que son sólo modelos aproximados de la realidad. El caotista quiere capturar los aspectos esenciales de su sistema bajo estudio, no quiere describir todos los aspectos del sistema. Para ello utiliza ecuaciones muy simples, que muestran universalidad, en el sentido de que la misma ecuación puede describir variados sistemas aparentemente diferentes: una población de pájaros y el comportamiento de la bolsa de Nueva York, por ejemplo [16].

La ecuación más sencilla que muestra caos y que ha sido utilizada ampliamente para describir muchos sistemas es la ecuación logística: $y = ax(1 - x)$, donde a es una constante.

Describiré la ecuación logística con un párrafo de la obra *Arcadia, Acto 1, Escena 4* [31]:

Lo que ella hace, cada vez que obtiene un valor para y , es usar ese como su próximo valor para x . Y así sucesivamente. Como una retroalimentación. Ella reintroduce la solución en la ecuación, y entonces la resuelve de nuevo. Iteración, tú sabes.

Lo que resulta del procedimiento anterior es que para ciertos valores del parámetro de la ecuación, el futuro no puede predecirse a partir del estado presente. La distribución final de valores de x luce aleatorio, destruyendo cualquier predicción, aún cuando la ecuación de donde partimos es completamente determinista. Este resultado representa caos, indeterminismo al nivel clásico. En otras palabras, caos a partir del orden: Lo impredecitivo y lo predeterminado se solapan de tal manera para hacer todo de la manera que es. Es como la naturaleza se crea a sí misma, en cada escala, el cristal de nieve y la tormenta.[31].

La moderna teoría del caos comenzó en el año 1966, cuando Edward Lorenz, estudiando un modelo matemático del comportamiento atmosférico, descubrió que sus ecuaciones mostraban caos. La resolución de esas ecuaciones con condiciones iniciales muy parecidas mostraba que al cabo de unos pocos segundos las soluciones se diferenciaban mucho unas de las otras. Con este resultado, Lorenz explicó porqué no se puede hacer predicciones del clima con más de unos pocos días de anticipación. Cualquier error en la medición de las condi-

ciones atmosféricas en un momento determinado se amplificará exponencialmente al cabo de poco tiempo invalidando cualquier predicción del clima. A este efecto Lorenz lo llamó “efecto mariposa”, recordando el cuento *El Sonido de un Trueno* del año 1952 del gran escritor de ciencia ficción Ray Bradbury [5]. En ese cuento la muerte de una mariposa prehistórica y su consecuente incapacidad para reproducirse, desencadena un final diferente en una elección presidencial.

Hasta ahora hemos discutido sólo la teoría del caos en la ciencia clásica. Pero la teoría del caos también hace su aparición en la mecánica cuántica. La ecuación de Schrödinger es una ecuación lineal de manera que el caos, en el mismo sentido que aparece en las leyes clásicas, no puede hacer su aparición en la mecánica cuántica. Sin embargo, el caos cuántico está ligado no a la aleatoriedad de los niveles energéticos o de la función de onda, sino a la aleatoriedad en la distribución del espaciado entre los niveles de energía. En síntesis, el caos cuántico se refiere a la aleatoriedad del espaciado de los niveles de energía a nivel cuántico de sistemas que presentan caos a nivel clásico [3, 7, 8, 17, 34]. El espaciado de los niveles energéticos de sistemas cuánticos no caóticos sigue una distribución de Poisson y la contraparte clásica de estos sistemas presenta una dinámica regular o completamente integrable, mientras que aquel de sistemas cuánticos caóticos sigue distribuciones correspondientes a ciertos tipos de matrices llamadas aleatorias y su contraparte clásica presenta una dinámica irregular, sin constantes de movimiento. Lo increíble es que este problema físico presente conexiones con uno de los problemas no resuelto más importante de la matemática actual, como lo es la hipótesis de Riemann. La hipótesis de Riemann establece que los ceros complejos de la función zeta de Riemann, $\zeta(z)$, tienen todos parte real igual a $1/2$, de manera que las cantidades $\{\Gamma_j\}$ definidas por $\zeta(\frac{1}{2} - i\Gamma_j) = 0$ son todas reales. Esta conjetura está, además, relacionada con la distribución de los números primos. La conexión se puede apreciar en la relación

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z} = \prod_{p=\text{primo}} \left(1 - \frac{1}{p^z}\right)^{-1}$$

con $Re(z) > 1$.

La conexión entre el caos cuántico y la hipótesis de Riemann tiene una historia muy interesante. En los años 50s, Eugene Wigner hizo la importante observación que la estadística de las matrices que describían el comportamiento de núcleos pesados no era muy diferente de la estadística correspondiente a matrices hermiticas aleatorias. Esta observación fue clave en la interpretación de los espectros de energía de núcleos pesados, ya que los elementos matriciales correspondientes a estos sistemas no eran conocidos. La observación de Wigner predice entonces que la distribución (la distribución, no los valores individuales) de los niveles de energía de un núcleo pesado sigue la misma ley que

la de matrices hermíticas aleatorias. Estas ideas fueron elaboradas posteriormente por Freeman Dyson, entre otros, y la proposición teórica fue confirmada hace 30 años en la distribución de los niveles nucleares del elemento erbio-166. La función de correlación de dos puntos de la distribución de los ceros de la función zeta de Riemann $1 - \left(\frac{\sin \pi x}{\pi x}\right)^2$ es la misma que la función de correlación de dos puntos de matrices hermíticas aleatorias. Esto sugiere que los ceros de la función zeta realmente poseen características espectrales. Si esto es así, los ceros de la función zeta deberían estar asociados a los autovalores de algún operador hermítico desconocido. Esta idea fue originalmente propuesta, de manera independiente, por David Hilbert y George Pólya. Si este operador existe, él determinará tanto la posición exacta de cada cero de la función zeta como la distribución exacta de los números primos. Por otra parte, debido a la conexión con el problema físico del caos cuántico, los físicos están también buscando un operador hamiltoniano correspondiente a un sistema físico y cuyos autovalores sean las cantidades reales $\{\Gamma_j\}$ correspondientes a la parte imaginaria de los ceros de la función zeta de Riemann. De acuerdo a Michael Berry, en el futuro estaremos leyendo los ceros de la función zeta de Riemann directamente de un espectro de un sistema físico, tomado en el laboratorio.

7 Conclusión

Hemos discutido el determinismo e indeterminismo en la ciencia natural, donde la matemática es un ingrediente fundamental en estos estudios. Sin embargo, existe también el problema de determinismo e indeterminismo en la matemática misma, problema éste relacionado con la lógica. Estas ideas son discutidas bajo el nombre de teoría de la información algorítmica [9,10]. Lo increíble es que estas ideas puramente matemáticas, sin relación aparente con el mundo natural, también están al final relacionadas con el mismo problema en las ciencias naturales. Para los platónicos, esto era de esperar: al final todo el universo y sus leyes no son más que representaciones de arquetipos que existen en el mundo perfecto de las ideas, el mundo de las matemáticas.

Concluiremos este artículo transcribiendo la parte final del ensayo de Jorge Luis Borges (*Nueva Refutación del Tiempo*). Después de describir sus ideas que presentan el tiempo como una ilusión, Borges comenta [4]:

And yet, and yet. . . Negar la sucesión temporal, negar el yo, negar el universo astronómico, son desesperaciones aparentes y consuelos secretos. Nuestro destino (a diferencia del infierno de Swedenborg y del infierno de la mitología tibetana) no es espantoso por irreal; es espantoso porque es irreversible y de hierro. El tiempo es la sustancia de que estoy hecho. El tiempo es un río que me arrebató, pero yo soy el río; es un tigre que me destroza, pero yo soy el tigre; es un fuego

que me consume, pero yo soy el fuego. El mundo, desgraciadamente, es real; yo, desgraciadamente, soy Borges.

Referencias

- [1] Aristotle, *Physics*, Oxford University Press, Oxford, 1999.
- [2] G. Berkeley, *Principles of Human Knowledge and Three Dialogues*, Oxford University Press, Oxford, 1996.
- [3] M. V. Berry y J. P. Keating, "The Riemann zeros and eigenvalue asymptotics", *SIAM Review*, 41 (1999), 236-266.
- [4] J. L. Borges, *Otras Inquisiciones*, Alianza Editorial, Madrid, 1997.
- [5] R. Bradbury, *El Sonido de un Trueno en Las Doradas Manzanas del Sol*, Minotauro, Barcelona, 2002.
- [6] D. C. Cassidy, "Heisenberg, uncertainty and the quantum revolution", *Sci. Amer.* May, (1992) 110-111.
- [7] B. Cipra, *What's Happening in the Mathematical Sciences 1998-1999*, vol 4, AMS, Providence, RI, 1999.
- [8] J. B. Conrey, "The Riemann hypothesis", *Notices of the American Mathematical Society*, 50 (2003), 341-353.
- [9] G. Chaitin, "Computers, Paradoxes and the Foundations of Mathematics", *American Scientist*, March-April (2002), 164-171.
- [10] G. Chaitin, "Información y Azar", *Boletín de la Asociación Matemática Venezolana*, IX (1) (2002), 55-81.
- [11] L. Chang, *The New York Times Grandes Logros de los Genios de Hoy*, McGraw-Hill, México, 2000.
- [12] A. Einstein, *Correspondencia con Michele Besso (1903-1955)*, Edición, prólogo y notas de P. Speziali, Traducción de M. Puigcerver, Tusquets Editores, Barcelona, 1994.
- [13] A. Einstein, *Mi visión del mundo*, Fábula Tusquets Editores, Barcelona, 1997.
- [14] G. Farmelo (editor), *It Must be Beautiful: Great Equations of Modern Science*, Granta Books, London, 2002.

-
- [15] M. Gell-Mann, *El Quark y el Jaguar*, Tusquets Editores, S.A., Barcelona, 1995.
- [16] J. Gleick, *Chaos*, Heinemann, 1988.
- [17] B. Hayes, "The spectrum of Riemannium", *American Scientist* 91 (2003), 296-300.
- [18] *Heráclito, Parménides y Empédocles, Textos Presocráticos*, traducción de M. Del Pino, Edicomunicación, Barcelona, 1995.
- [19] G. Holton, *Einstein, historia y otras pasiones*, Taurus, Madrid, 1998.
- [20] P. S. Laplace, *Philosophical Essays on Probabilities*, traducido por F. W. Truscott y F. L. Emory, Dover, New York, 1951.
- [21] E. Mayr, *This is Biology*, The Belknap Press, Cambridge, Mass., 1997.
- [22] Ovidio, *La Metamorfosis*, Edicomunicación, S.A., España, 1995.
- [23] Platón, *Diálogos*, Editorial Porrúa, México, 1996.
- [24] H. Poincaré, *The Value of Science*, The Modern Library, New York, 2001.
- [25] K. Popper, *El universo abierto*, Tecnos, Madrid, 1984.
- [26] H. Price, *Time's Arrow and Archimedes Point*, Oxford University Press, New York, 1996.
- [27] I. Prigogine, *The End of Certainty*, The Free Press, New York, 1997.
- [28] *Scientific American*, September, 2002. Número especial sobre el tiempo.
- [29] C.P. Snow, *The Two Cultures*, Cambridge University Press, Cambridge, 1998.
- [30] I. Stewart, *Does God Play Dice? The Mathematics of Chaos*, Basil Blackwell, New York, 1989.
- [31] T. Stoppard, *Arcadia*, Faber and Faber Limited, London, 1993.
- [32] M. Tegmark, "Parallel Universes", *Scientific American*, May 2003, 30-41.
- [33] M. Tegmark y J.A. Wheeler, "100 Years of Quantum Mysteries", *Scientific American*, February (2001), 54-61.
- [34] A. Terras, "Finite quantum chaos", *Amer. Math. Month.* 109 (2002), 121-139.

- [35] Tito Lucrecio Caro, *De la Naturaleza de las Cosas*, traducción de L. Alvarado, Equinoccio Editorial de la Universidad Simón Bolívar, Baruta, 1982.
- [36] E. Wigner, “The unreasonable effectiveness of mathematics in the natural sciences”, *Commun. Pure Appl. Math*, 13 (1960), 1-14.
- [37] S. Wolfram, *A New Kind of Science*, Wolfram Media, 2002.
- [38] D. Z. Zhang, “C. N. Yang and Contemporary Mathematics”, *Math. Intelligencer* 15 (1993), 13-21.

DANIEL A. MORALES
FACULTAD DE CIENCIAS
UNIVERSIDAD DE LOS ANDES
VENEZUELA

MATEMÁTICAS RECREATIVAS

La importancia de cada nodo en una estructura de enlaces: Google-PageRank™

Roberto Markarian & Nelson Möller

Resumen

Al buscar información en Internet, es importante cómo vemos ordenado el resultado de nuestra búsqueda. Este resultado es una numerosa cantidad de páginas con algo común a los temas o nombres consultados. En este trabajo, explicamos un procedimiento que asocia a cada página de la Red un número que cuantifica su “relevancia” y permite ordenar los resultados de la búsqueda. Este método fue popularizado por, y constituyó la base del buscador Google. Como se enlazan las páginas de la Red, determina una matriz cuyos vectores propios tienen propiedades que permiten utilizar sus componentes como una medida de “relevancia”. Nuestro objetivo es mostrar la utilización de nociones básicas del Álgebra Lineal en estos cálculos.

Introducción

La siguiente cita, extraída del periódico Le Monde [La], ilustra en términos generales las ideas que profundizaremos:

“A mediados de los '90, frente al creciente flujo de información, dos estudiantes de computación de la Universidad estadounidense de Stanford, Sergey Brin y Larry Page, intuyeron algo: un motor de búsqueda que se basara en el estudio matemático de las relaciones entre los diferentes sitios daría mucho mejor resultado que las técnicas rudimentarias que se empleaban entonces.

Convencidos de que las páginas más ‘pertinentes’ son las más frecuentemente citadas (las que los otros sitios ponen como referencia en su lista de enlaces en hipertexto) deciden hacer del tema su proyecto de estudios, estableciendo así las bases de un motor más ‘matemático’, al que bautizaron Google en el momento de crear su empresa, en setiembre de 1998.”

El mismo artículo continúa:

Para evaluar la ‘pertinencia’ de las páginas existentes en internet, Brin y Page inventaron el ‘Page Rank’, una escala de valores propia de Google. En la misma, la importancia de las páginas web es reevaluada permanentemente en función de la cantidad de menciones de que son objeto en los diferentes sitios. Por lo tanto, los sitios aislados, que no figuran en las listas de enlaces hipertextuales, resultan poco visibles, sin ‘legitimidad’. En cambio los sitios muy citados se convierten para Google en sitios de referencia. Ese original algoritmo ofrece resultados impresionantes.

Buscar material en Internet plantea simultáneamente dos problemas:

- ¿Qué páginas tienen elementos relacionados con lo que buscamos?
- ¿Cómo se presenta (ordena) todo lo hallado?

En este trabajo, miramos un aspecto matemático relacionado al segundo punto: analizaremos uno de los elementos, utilizado por el buscador Google para ordenar los resultados de una búsqueda. El primer problema, que tiene también gran relación con elementos del Algebra Lineal, no será tratado aquí; una referencia es [BDJ].

Los resultados iniciales de nuestra búsqueda suelen ser muchas páginas de direcciones relacionadas con el tema, pero pocas veces miramos más allá de las primeras. Por ello es muy útil un procedimiento que ordene los resultados de acuerdo a la “relevancia” que tienen las páginas.

Es allí donde interviene uno de los principales elementos introducidos por el Google en 1998, el PageRank [BP]: “Para medir la importancia relativa de las páginas web nosotros proponemos PageRank, un método para calcular un ordenamiento (*ranking* en inglés) para toda página basado en el gráfico de la Red.”

Expresándolo de manera un tanto simplificado, lo que buscamos es que *la importancia de cada página sea proporcional a la suma de las importancias de todos los sitios que enlazan con ella*. Matemáticamente, si llamamos K a la constante de proporcionalidad y x_i a la importancia, tenemos un sistema de ecuaciones del tipo

$$\begin{aligned} x_1 &= K(x_{14} + x_{97} + x_{1002}) \\ x_2 &= K(x_{11104} + x_{20006}) \\ &\vdots \\ &\vdots \end{aligned}$$

donde en cada igualdad el lado derecho es la suma de la importancia de todos los sitios que enlazan a la página correspondiente.

Sea A la matriz cuya entrada a_{ij} es 1 si el sitio j tiene un enlace con la página i y 0 caso contrario. Esta matriz permite reescribir las ecuaciones anteriores en

la forma

$$x_i = K \sum_{j=1}^n a_{ij} x_j.$$

Entonces, el problema de hallar vectores $x = (x_1, x_2, \dots, x_n)$ que satisfacen esa igualdad se transforma en encontrar x tal que $Ax = \frac{1}{K}x$, que es un problema de valores y vectores propios de una matriz que toma en cuenta la estructura de vínculos (*links* en inglés) de la Red.

El teorema de Perron - Frobenius sobre los valores propios de matrices con entradas reales no negativas es una pieza clave para mostrar que el método usado por PageRank funciona. En la versión original de Perron (1907) el teorema expresa que el valor propio de mayor valor absoluto de una matriz (con entradas) positiva(s) es positivo y su espacio propio es generado por un vector propio con coordenadas del mismo signo. Frobenius (1908, 1912) extendió estos resultados a matrices no negativas. Este resultado es central a la hora de implementar computacionalmente el cálculo.

En los años 50, ya se había observado el papel de este vector propio asociado al mayor valor propio de una matriz positiva, para obtener un ordenamiento [Ke]. Su vigencia actual se debe a su aplicación a Internet y a la posibilidad de implementar el cálculo para matrices muy grandes

Este trabajo, por su destino original, de divulgación, contiene detalles conocidos por toda aquella persona con una formación profesional en el cual el teorema de Perron-Frobenius sea utilizado frecuentemente. Por lo tanto puede ser leído de varias maneras; quienes estén interesados en la descripción general del PageRank y sus implicaciones pueden leer las Secciones 1, 2, 4 y 5. Quienes estén interesados en esos asuntos y en el planteamiento y solución de los principales problemas matemáticos deben leer además la Sección 3 y principalmente el Apéndice.

1. Algo de Historia

La Red ha crecido en una forma vertiginosa. Hagamos un poco de historia para situar el contexto de invención del procedimiento (algoritmo) utilizado por Google.

En 1996-98 ya comenzaba a notarse la dificultad de hallar material en internet debido a su rápido crecimiento. En ese momento “buscadores” también llamados “motores de búsqueda”, como Altavista, Lycos, Yahoo, ya tenían gran relevancia.

En principio [Pe], todo motor de búsqueda está compuesto de por lo menos tres elementos principales: un robot de indexación (también conocido como

araña, *spider o web crawler*), una base de datos y una interface de consulta de la base de datos. Normalmente los usuarios interactúan con la interface de consulta, y a través de ella consultan la base de datos. El robot de indexación es el encargado de “navegar” la Web colectando toda la información que este pueda procesar, almacenándola en la base de datos para su posterior consulta.

Muchos motores desarrollaban tecnologías que permitían restringir la búsqueda. Estas restricciones empleaban argumentos lógicos que no eran de manejo sencillo. Yahoo hacía “manualmente” el trabajo de ordenar de acuerdo a ciertos criterios “objetivos” las bases de datos disponibles. Dichas bases de datos tenían un tamaño considerable, por lo que ya estaba muy popularizado el uso de buscadores, y los que funcionaban bien eran un gran negocio: Yahoo se vendió en una abultada cifra. Los algoritmos de búsqueda recibían un gran impulso y a pesar de ello no se simplificaba el hallar lo deseado.

En ese contexto, y en pleno boom de las compañías puntocom, fue que comenzó en la Universidad de Stanford la historia de Google. Sergey Brin y Lawrence Page presentaron un trabajo de posgrado donde se definía la “importancia” de una página web tomando en cuenta los enlaces que recibe. Su buscador hace una lista de respuestas a nuestra búsqueda en un “orden de relevancia” decreciente, esta fue la mejora en su interface de consulta que popularizó su uso. Hemos puesto el comillado porque se señalan deficiencias y críticas al modo cómo se hace la cuantificación (de “relevancia”). Algunas de éstas serán comentadas más adelante.

2. Como ordenar las páginas de la Red.

Estando en una página web A determinada tenemos dos números importantes:

- cantidad de vínculos entrantes = cantidad de páginas que tienen un vínculo hacia la página A ;
- cantidad de vínculos salientes.

Las páginas web varían mucho en el número de vínculos entrantes que poseen. Generalmente las páginas que tienen muchos vínculos entrantes son más importantes que las que sólo tienen unos pocos.

Sin embargo, hay muchos casos en los cuales sólo el contar el número de vínculos entrantes no se corresponde con el sentido usual de la importancia de una página web.

Como escribían Brin y Page [BP]: “Por ejemplo, si una página tiene un vínculo de la página principal de Yahoo, éste puede ser un solo vínculo pero uno muy importante. Dicha página debería estar mejor clasificada que otras páginas con muchos vínculos pero de lugares desconocidos”.

Por tanto, una página tiene una clasificación alta si la suma de las clasificaciones de sus vínculos entrantes es alto. Esto cubre ambos casos: muchos vínculos entrantes o pocos con alta clasificación.

El algoritmo original del PageRank fue descrito en varios trabajos por Brin y Page [BP]. Posteriormente presentaron una versión mejorada, que es la que expondremos. El propósito es cuantificar la probabilidad de que un usuario (aleatorio) llegue a la página A utilizando la Red. Se define el PageRank por:

$$PR(A) = \frac{(1-d)}{N} + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

donde:

- N es el número total de páginas web desde las que salen vínculos.
- n es el número total de páginas web desde las que salen vínculos a la página A .
- $PR(A)$ es el PageRank de la página A .
- $PR(T_i)$ es el PageRank de las páginas T_i que tienen un vínculo hacia la página A .
- $C(T_i)$ es el número de vínculos salientes de la página T_i .
- d es un factor de amortiguación que puede ser tomado entre 0 y 1.

Como la suma de esos números sobre todas las páginas web, da uno es una distribución de probabilidad (indexada por el parámetro d). Esta “normalización” (suma=1) facilita la utilización de resultados generales que no dependen del tamaño del sistema (el número total de páginas).

Analizando con cuidado dicha fórmula se observarán las siguientes características del PageRank:

- está definido para cada página y es determinado por los PageRanks de las páginas que tienen un vínculo dirigido hacia ella;
- los sitios que enlazan a la página A no influyen uniformemente; depende del número de vínculos salientes que ellas posean: a más vínculos salientes de una página menos beneficiará el PageRank de las páginas a las que se una;
- un nuevo vínculo a una página siempre aumenta su valor;
- la definición es recursiva: la clasificación de una página depende de todas las otras que tienen vínculos hacia ella, por ello la clasificación de cada página depende de **todos** los sitios de la Red.

En sus explicaciones Brin y Page dan una justificación sencilla para el algoritmo. El PageRank modela el comportamiento de un usuario que estando en una página puede:

- elegir al azar entre los vínculos contenidos en la página actual, o
 - saltar al azar a cualquier página de la Red ingresando la dirección;
- todo ello sin tener en cuenta el contenido de los mismos (esto ha suscitado comentarios y modelos alternativos ver [DR]). Cuantificando esos comportamientos posibles, se supone que seguirá un enlace de la página en que está con probabilidad d , o que salta a cualquier página con probabilidad $1 - d$.

La definición del PageRank establece **un** procedimiento para determinar una probabilidad de que un usuario aleatorio llegue a la página web A . El navegante aleatorio visita una página web con una probabilidad proporcional al PageRank de la página. La probabilidad de elegir un vínculo depende de los vínculos que puede elegir en la página en que está.

El seguimiento de los vínculos está indexado probabilísticamente por el factor de amortiguamiento d . Parece razonable suponer que $d > 1/2$, o sea, estando en una página, se tiende a usar más los vínculos que allí están, que hacer una nueva elección al azar. En la Sección 3 profundizaremos en el significado y el uso de d .

La única excepción son las páginas hacia las que no va ningún vínculo, a las cuales en este modelo, por estar aisladas, sólo se llega al azar. No caben dudas que a ellas se puede llegar buscándolas explícitamente, pero para usar este procedimiento -que es el mejor procedimiento de búsqueda!- no se necesitan ‘buscadores’. El PageRank de estas páginas es $\frac{1-d}{N}$.

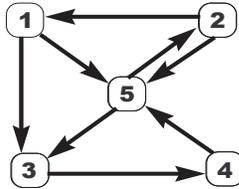
Vamos a ver que, por la naturaleza de la definición, es posible utilizar un algoritmo iterativo que aproxima los valores de PageRank. O sea, a cada página se le asigna un valor inicial y se realizan iteraciones que modifican sucesivamente estos valores iniciales. Esto es, a partir de distribuciones iniciales prefijadas, se repite un mismo procedimiento para obtener nuevos valores para cada página, y así sucesivamente. Este es un punto importante a la hora de implementar el mecanismo, pues en términos computacionales es más sencillo calcular iterativamente el valor y el vector propio que mediante otros procedimientos.

Algunas preguntas surgen naturalmente. ¿Por qué este procedimiento funciona? ¿Será que este procedimiento asigna a cada página un valor único, su PageRank? Explicaremos en detalle como se realiza este cálculo en el ejemplo de la siguiente Sección.

Las respuestas afirmativas, en general, incluyen el uso de una versión del teorema de Perron-Frobenius que se dará en el Apéndice.

3. Un ejemplo

Veamos ahora como es el procedimiento recursivo en un ejemplo dado por el siguiente diagrama.



Tenemos 5 páginas web e indicamos con una flecha los vínculos. Por ejemplo, de la página 1 salen dos vínculos a las 3 y 5, y entra un vínculo de la página 2.

Veamos las fórmulas de PageRank de una manera más compacta, intentando utilizar la nomenclatura probabilística relacionada con la distribución estacionaria de una cadena de Markov ¹. Llamamos $\pi_i = PR(i)$ al PageRank de la página i :

$$\begin{aligned} \pi_1 &= \frac{1-d}{5} + d \left(\frac{\pi_2}{2} \right), \\ \pi_2 &= \frac{1-d}{5} + d \left(\frac{\pi_5}{2} \right), \\ \pi_3 &= \frac{1-d}{5} + d \left(\frac{\pi_1}{2} + \frac{\pi_5}{2} \right), \\ \pi_4 &= \frac{1-d}{5} + d (\pi_3), \\ \pi_5 &= \frac{1-d}{5} + d \left(\frac{\pi_1}{2} + \frac{\pi_2}{2} + \pi_4 \right). \end{aligned}$$

Si definimos la matriz:

$$P = \frac{1-d}{5} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} + d \begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1/2 & 1 & 0 & 1 & 0 \end{pmatrix}$$

y $\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix}$, utilizando que $\sum_{i=1}^5 \pi_i = 1$ podemos resumir las 5 ecuaciones en²

$$\boxed{\pi = P\pi.}$$

¹Ver el capítulo 5 de [Ha].

²En lenguaje probabilístico suele ser más común llamar P a la transpuesta de nuestra matriz y llegar a $\pi = \pi P$.

Modelo de navegación³

Obtendremos la matriz P recurriendo a las explicaciones dadas en la sección anterior para justificar la definición. Resumamos la estructura de vínculos en la **matriz de conectividad** A , definida por

$$a_{ij} = \begin{cases} 1 & \text{si hay un vínculo de la página } j \text{ a la } i, \\ 0 & \text{si no hay un vínculo de la página } j \text{ a la } i. \end{cases}$$

En el caso del ejemplo

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

Supongamos que el usuario tiene los siguientes dos modos de navegación:

1. Elige una página al azar.
2. Sigue los vínculo de la página en la que está.

Elegimos un número d , $0 < d < 1$; la probabilidad del modo 2. Queremos saber lo siguiente: estando en una página determinada ¿cuál es la probabilidad de que en el próximo paso esté en otra determinada página?

Para esto introducimos otra matriz P , llamada **matriz de transición** cuya entrada p_{ij} es la probabilidad de que estando en la página j pase al sitio i . Tenemos que:

$$p_{ij} = \begin{cases} \frac{d}{\text{vínculos salen de } j} + \frac{1-d}{\text{total páginas}} & \text{si hay un vínculo de } j \text{ a } i \\ \frac{1-d}{\text{total páginas}} & \text{si no hay vínculo de } j \text{ a } i \end{cases}$$

Observe que la matriz P se puede obtener a partir de la matriz de conectividad A de la siguiente manera:

- $C(T_j) = \text{vínculos que salen de la página } j = \sum_{i=1}^5 a_{ij}$.
- Dividimos la columna j de A por $C(T_j)$.
- Formamos una nueva matriz C con la columnas del paso anterior.
- Entonces: $P = \frac{(1-d)}{5} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix} + dC$.

³Si ha estudiado cadenas de Markov es posible que las siguientes explicaciones le resulten elementales y rudimentarias.

Proceso de Iteración

Un vector de probabilidades es

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_5 \end{pmatrix}$$

donde $0 \leq p_j \leq 1$ y $\sum_{j=1}^5 p_j = 1$. El número p_j es la probabilidad de estar en la página j .

Si $p^{(k)}$ es el vector de probabilidades en el k -ésimo paso de la navegación, tenemos que

$$p^{(k+1)} = Pp^{(k)}.$$

Por ejemplo, si comenzamos en la página 1:

$$p^{(0)} = (1, 0, 0, 0, 0)$$

después del primer paso $p^{(1)} = Pp^{(0)}$ es el vector

$$\begin{pmatrix} \frac{(1-d)}{5} \\ \frac{(1-d)}{5} \\ d/2 + \frac{(1-d)}{5} \\ \frac{(1-d)}{5} \\ d/2 + \frac{(1-d)}{5} \end{pmatrix}.$$

De la misma manera, al segundo paso $p^{(2)} = Pp^{(1)} = P^2p^{(0)}$ y podemos continuar para obtener que $p^{(k)} = Pp^{(k-1)} = P^k p^{(0)}$. Observe que todos los vectores $p^{(k)}$ son de probabilidad.

Lo que nos interesa son las probabilidades a largo plazo; o sea, nos preguntamos si los vectores de probabilidad $p^{(k)} = P^k p^{(0)}$, $k = 1, 2, \dots$ convergen a algún vector de equilibrio π **independientemente** del vector de probabilidades inicial $p^{(0)}$. Si eso sucede, entonces en particular

$$\pi = \lim_{k \rightarrow \infty} P^{k+1} \pi = \lim_{k \rightarrow \infty} P P^k \pi = P \left(\lim_{k \rightarrow \infty} P^k \pi \right) = P \pi.$$

Por lo tanto π es un vector propio asociado al valor propio 1. Además por la independencia del vector inicial, si consideramos los vectores e_j de la base canónica tenemos que las columnas de P^k son

$$P^k e_j \rightarrow \pi,$$

lo que implica que $P^k \rightarrow P_\infty$ donde cada columna de P_∞ es el vector π .

Convergencia en norma uno⁴. Es importante tener en cuenta que los límites que aparecen en este trabajo se refieren a que la norma uno del vector diferencia

⁴En lenguaje probabilístico convergencia en variación total.

tiende a cero. Esto significa que si $v_j^{(t)}, v_j$ son las coordenadas j -ésimas de $v^{(t)}$ y v (vectores de \mathbb{R}^n), respectivamente, entonces $\lim_t v^{(t)} = v$ significa que

$$\|v^{(t)} - v\|_1 = \sum_{j=1}^n |v_j^{(t)} - v_j| \rightarrow 0 \quad \text{cuando } t \rightarrow \infty.$$

Veamos cómo hacemos el cálculo **recursivamente** en nuestro ejemplo⁵ utilizando $d = 0,85$

1. Comenzamos con un vector de probabilidad inicial

$$p^{(0)} = \begin{pmatrix} 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \end{pmatrix},$$

2. Calculamos

$$\begin{aligned} p^{(1)} = Pp^{(0)} &= \begin{pmatrix} ,03 & ,455 & ,030 & ,030 & ,030 \\ ,03 & ,03 & ,03 & ,03 & ,455 \\ ,455 & ,03 & ,03 & ,03 & ,455 \\ ,03 & ,03 & ,88 & ,03 & ,03 \\ ,455 & ,455 & ,03 & ,88 & ,03 \end{pmatrix} \begin{pmatrix} 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \\ 0,2 \end{pmatrix} \\ &= \begin{pmatrix} ,115 \\ ,115 \\ ,2 \\ ,2 \\ ,37 \end{pmatrix}. \end{aligned}$$

Miramos cuán próximos están:

$$\begin{aligned} \delta_1 &= \|p^{(1)} - p^{(0)}\|_1 \\ &= |,115 - 0,2| + |,115 - 0,2| \\ &\quad + |,2 - 0,2| + |,2 - 0,2| + |,37 - 0,2| \\ &= ,34. \end{aligned}$$

3. Calculamos $p^{(k)} = Pp^{(k-1)}$, hasta que estén suficientemente próximos, o sea $\delta_k < \varepsilon$.⁶
4. Si ε es muy pequeño, la componente i del vector $p^{(k)}$ será una buena aproximación al PageRank de la página i .

En el ejemplo si tomamos $k=11$, llegamos a:

$$p^{(10)} = \begin{pmatrix} 0,09934354879645 \\ 0,16700649449556 \\ 0,20994655573428 \\ 0,20521883387311 \\ 0,31848456710061 \end{pmatrix} \quad p^{(11)} = \begin{pmatrix} 0,10097776016061 \\ 0,16535594101776 \\ 0,20757694925625 \\ 0,20845457237414 \\ 0,31763477719124 \end{pmatrix}$$

$$\Rightarrow \delta_{11} = \|p^{(11)} - p^{(10)}\|_1 = 0,00973989973037.$$

⁵Sólo aparecen los resultados de los cálculos. Realizados en MATLAB, formato largo.

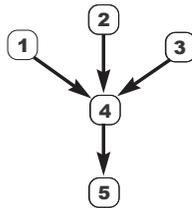
⁶El problema de estimar k en función de ε tiene importancia a la hora de implementar el cálculo; esta estimación involucra a otro valor propio de P [HK].

Si calculamos directamente⁷ el vector propio, se obtiene un resultado muy cercano al anterior

$$\begin{pmatrix} 0,10035700400292 \\ 0,16554589177158 \\ 0,20819761847282 \\ 0,20696797570190 \\ 0,31893151005078 \end{pmatrix}.$$

Observe que la página 5 es la que tiene mejor clasificación.

Si se realiza el cálculo con un esquema del tipo que sigue, se verá que nuevamente la página 5 será la más relevante.



¿Qué sucede si la página 5 no enlaza consigo misma? (En ese caso la página 5 representa lo que se conoce como **enlace colgado**.) Si vuelve al ejemplo anterior verá que aparece una división por 0 al definir la matriz P . En ese caso se calcula el de las páginas 1, 2, 3, 4 y después con esos números el de la 5. Esto es un fenómeno presente muchas veces en el cálculo del PageRank real, por ejemplo debido a enlaces a páginas que no han sido todavía descargadas por las “arañas” del Google (éstas aparentan no poseer enlaces salientes).

4. Google en serio

Se considera el conjunto W de páginas que se pueden alcanzar a partir de una página en Google. Sea N el número de páginas en W , este número varía con el tiempo, (en mayo 2002 era alrededor de 2700 millones [Mo]). Consideramos la matriz $N \times N$ de conectividad de W . La matriz es enorme pero tiene una gran cantidad de ceros (en inglés, *sparse matrix*). Consideramos la matriz P construida de manera análoga a lo hecho en nuestro ejemplo. Esta matriz tiene todas sus entradas no negativas y la suma de los elementos de cada columna da uno; se dice que es una matriz de Markov. De acuerdo con lo expresado en la Sección anterior se trata de encontrar un vector π tal que $\pi = P\pi$. Se prueba que si la matriz es de Markov y $\sum_i \pi_i = 1$, entonces π es único. El elemento π_j de π es el PageRank de la página j (a menos de posibles cambios de escala).

Observemos que la forma recursiva de implementar el algoritmo al realizar el cálculo no es algo menor, estamos hablando de manejar una matriz que tiene

⁷Utilizando las funciones eig y norm del MATLAB.

un tamaño de varios millones. En el Apéndice se mostrará por qué funciona esta implementación que asigna una calificación no nula única a cada página.

En teoría sucede que toda página posee un PageRank positivo pero en el ordenamiento real se introducen como penalización una calificación nula llamada **PR 0**. Desde que se popularizó la utilización del Google los responsables (*webmaster*) de algunas sitios han intentado aumentar la calificación de sus páginas intentando manipular sus enlaces. A su vez, los administradores de Google quieren evitar trampas de este tipo, por lo que se intenta detectar y penalizar tales intentos. Públicamente se desconoce la forma en que se realiza, puesto que, diversos elementos que hacen funcionar su buscador son secretos comerciales. En la Red, existe material que especula acerca de la implementación de esta penalización [EF].

5. Consideraciones generales

En este momento, Google no sólo es el buscador más utilizado, sino que, le vende servicios a portales importantes: Yahoo, AOL, etc. Se estima que, por venta de servicios y licencias de su tecnología de búsqueda tiene ganancias por 150 millones de dólares [Ec]. Un elemento no menor luego de la caída de las puntocom de marzo 2000.

El 27 de junio de 2002, la Comisión Federal de Comercio de los Estados Unidos estableció ciertas reglas; recomendando que cualquier ordenamiento influido por criterios monetarios más que por criterios “imparciales” y “objetivos” debía ser claramente indicado para proteger los intereses de los consumidores.

Por ello, cualquier algoritmo como éste, que aparenta ser objetivo, continuará siendo un aspecto importante para las búsquedas en la Red.

Google es también el único motor de búsqueda que recorre la Red frecuentemente para mantener actualizada su base de datos (por lo menos así lo ha hecho en los últimos dos años). Lleva, aproximadamente, una semana cubrir la Red y otra para calcular el PageRank. El ciclo de puesta al día de Google es de aproximadamente 30 días. Se ha advertido, que el PageRank vigente influye el recorrido mensual realizado por Google, hace que páginas con PageRank más alto sean recorridas más rápidamente y en mayor profundidad que otras con menor clasificación.

Este último punto, hace que se vea como discriminatoria la naturaleza del PageRank [La], [Bra]. Se llega a afirmar que, los nuevos sitios lanzados en el 2002 tienen mayor dificultad en conseguir tráfico que antes del dominio de Google y que la estructura de enlaces de la Red ha cambiado significativamente a partir del predominio del Google.

Debido a la naturaleza del orden que establece el PageRank, una búsqueda no lleva hacia la referencia “principal” sobre el tema sino hacia la acepción más

ampliamente citada. Ya hemos observado que existen quienes intentan mejorar su calificación, y que, Google trata de controlar tales comportamientos. Se han realizado experiencias exitosas que muestran las posibilidades de utilizar “artificialmente” esta característica para subir el PageRank de una página. En los términos utilizados en [La]:

“En realidad, el poder de influencia de los diferentes actores depende sobre todo de su grado de apropiación de la Red: no alcanza con desarrollar un sitio, también hay que ser capaz de establecer vínculos con los otros sitios y obtener el reconocimiento de ‘los que cuentan’ en internet.”

El artículo enfatiza aun:

“Es sin duda en los temas políticos -sobre los cuales cohabitan en internet puntos de vista radicalmente diferentes- donde Google pone de manifiesto sus límites: sus criterios matemáticos pueden privilegiar *de facto* ciertas opiniones y brindar una pertinencia indebida a textos que sólo representan la opinión de unos pocos. La base y la sobrerrepresentación de que se benefician los ‘adelantados’ de internet, la densidad de lazos que mantienen (sobre todo a través del fenómeno esencialmente estadounidense de los *weblogs*), designan -matemáticamente- a los actuales ‘gurús’ de Google. Por cierto que el sistema pasó brillantemente las pruebas en cuestiones técnicas y prácticas. Pero existen terrenos en los que la pertinencia escapa a los algoritmos.”

Google empresa, finalmente lanzó su cotización en el mercado electrónico NASDAQ, el 18 de agosto de 2004. Su lanzamiento, por un monto superior a 20 mil millones de dólares, utilizó un mecanismo no habitual, conocido como Remate Holandés Modificado (Modified Dutch Auction). El Google IPO (siglas en Inglés de Oferta Pública Inicial) fue la manera como se recabaron las ofertas a través de ciertas agencias.

Este proceso sufrió varios retrasos, desde observaciones por parte de la comisión reguladora (SEC) originados por la existencia de acciones en poder de ex-funcionarios, colaboradores, etc; hasta observaciones por las declaraciones de Brin y Page a Playboy en el período en el que se debía mantener reservas. Han habido varias quejas y rumores sobre el acceso a este proceso, además de los retrasos mencionados debidos a las observaciones de la SEC. Para el que esté interesado, la sigla en NASDAQ de Google es GOOG.

Estas son noticias ajenas al fin principal del artículo, sobre las cuales se producen novedades continuamente.

Se dice que Microsoft también estaría por lanzar su propia tecnología de búsqueda[Ec].

Apéndice: Por qué funciona el algoritmo.

Importancia de las matrices no negativas.

En este Apéndice daremos una demostración algebraica de una versión probabilística del Teorema de Perron-Frobenius.

Distintas versiones de este teorema fueron probadas en contextos totalmente abstractos, pero la importancia de la teoría de matrices no negativas se ha extendido a campos muy amplios: las teorías de probabilidad y de sistemas dinámicos, el análisis numérico, la demografía, la economía matemática y la programación dinámica. Ver, por ejemplo [MC].

Esto se debe a que diversas variables que se miden en el mundo real, interactúan a través de relaciones positivas o nulas. A su vez, una cantidad de modelos que formalizan esas interacciones son procesos iterativos lineales en que se comienza con un estado v y se evoluciona por la aplicación reiterada de una matriz A , de modo que luego de n pasos se tiene el estado $v^{(n)} = A^n v$. Muchas veces es fundamental saber cuándo este proceso converge a un estado único, cualquiera sea el estado de comienzo v . La teoría de matrices positivas responde a ésta (y muchas otras) cuestión(es).

El enfoque que haremos aquí tiene por prerequisites algún manejo algebraico y cursos elementales de álgebra lineal. Este enfoque sencillo y directo puede hacer perder parte del “sabor probabilístico” que en profundidad tienen muchos de los contenidos. Pedimos disculpas por esta opción que no es involuntaria.

Se pueden encontrar otras pruebas y desarrollo de estos temas en [MC]; [Ha] Ch. 5; [Re] Ch. 2. Un tratado muy completo sobre matrices no negativas, que comienza con los resultados que nos interesan, es [Se].

Convergencia

Sea $B \in M_{n \times n}(\mathbb{C})$ una matriz con valor propio $\lambda_1 = 1$ de multiplicidad algebraica 1, y los demás valores propios satisfaciendo $1 > |\lambda_2| \geq \dots \geq |\lambda_n|$. Los $\lambda_i, i \geq 2$ se pueden repetir. El uno es lo que se llama un **valor propio dominante**.

Si existe una base de vectores propios $v_i, i = 1, \dots, n$, entonces $v = \sum_{i=1}^n \alpha_i v_i$ y

$$B^k(v) = \alpha_1 v_1 + \sum_{i=2}^n \alpha_i \lambda_i^k v_i,$$

donde sabemos que $\lim_k \lambda_i^k = 0$ para $i = 2, \dots, n$.

Si no hay una base de vectores propios, o sea si la multiplicidad algebraica de alguno de los $\lambda_i, i \geq 2$; es diferente de la multiplicidad geométrica, entonces habrán elementos w_1, w_2 de una base (de Jordan) que satisfacen $Bw_1 = \lambda_i w_1 + w_2, Bw_2 = \lambda_i w_2$. Un breve cálculo permite deducir que $B^k w_1 = \lambda_i^k w_1 + k \lambda_i^{k-1} w_2$. Este sencillo cálculo es generalizable a todas las situaciones que se pueden presentar al tomar una base de Jordan.

En la matriz de Jordan se tienen los llamados bloques de Jordan que son de la forma $\lambda I + N$ donde λ es un valor propio (real o complejo) I es una matriz identidad, digamos que s por s , y N una matriz “nilpotente” todas cuyas entradas son cero, excepto la línea subdiagonal $(a_{i,i-1})$ que está formada por unos que verifica $N^s = 0$. Entonces $(\lambda I + N)^k = \sum_{i=0}^{s-1} \lambda^{k-i} C_i^k N^i$, donde C_i^k son las combinaciones de k elementos tomados de a i .

Por tanto $B^k v$ será combinación lineal de vectores de la base, cuyos coeficientes –con excepción del correspondiente al vector propio asociado al valor propio λ – tienden a cero cuando $k \rightarrow \infty$ por tener cada valor propio –distinto del primero– módulo menor que uno.

Así, sea la matriz B diagonalizable o no, si el uno tiene multiplicidad algebraica uno y las demás raíces del polinomio característico tienen módulo menor que uno podemos garantizar que

$$\lim_k B^k v = \alpha_1 v_1.$$

La igualdad anterior es la que nos permite realizar el cálculo iterativo cuando $\alpha_1 \neq 0$ porque la aplicación sucesiva de la matriz B a cualquier vector con $\alpha_1 \neq 0$ converge a múltiplos de un mismo vector (un vector propio de valor propio λ). El teorema de Perron-Frobenius nos permitirá también elegir vectores con los que comenzar el proceso, con $\alpha_1 \neq 0$, para los que la convergencia no será al vector nulo.

Matrices de Markov. Resultados principales.

Decimos que una matriz B es **positiva** si todos sus elementos son números positivos. Lo notamos $B > 0$. Si $X \in M_{n \times 1}(\mathbb{R})$ (o $M_{1 \times n}(\mathbb{R})$) es positivo decimos que X es un **vector positivo**. Dadas dos matrices A, B del mismo tamaño decimos que $A > B$ si $A - B$ es una matriz positiva. Tenemos definiciones análogas para $A \geq 0$ (no negativo) si sustituimos positivos por elementos mayores o iguales a cero. Diremos que $p \in \mathbb{R}^n$ es un **vector de probabilidad** si es no negativo y la suma de sus componentes es uno.

Definición 1 (Matrices de Markov) Dada una matriz positiva M del espacio $M_{n \times n}(\mathbb{R})$, decimos que es de **Markov** si la suma de los elementos de cada columna es uno ($\sum_{i=1}^n m_{ij} = 1, \forall j = 1, \dots, n$.)

⁸Recuerde que $\lim_k \lambda_i^k k^s = 0$.

Se probará que las matrices de Markov satisfacen las propiedades de la matriz B anterior: tienen un valor propio uno con multiplicidad algebraica uno y todos los demás valores propios con módulo menor que uno. También se probará que si el vector v es de probabilidad el valor α_1 antes referido es distinto de cero. Gran parte de estas demostraciones se harán usando la matriz traspuesta M^T de la matriz de Markov, que satisface que los elementos de sus filas suman uno, y tiene los mismos valores propios que M (por tener el mismo polinomio característico).

En el transcurso de la demostración de los puntos principales se demostrarán otros resultados interesantes. Resumimos todos los resultados en el siguiente

Teorema. *Si M es una matriz de Markov, entonces todos sus valores propios tienen módulo menor o igual que uno y sólo uno de ellos, el uno, tiene módulo uno. La multiplicidad algebraica del uno es uno y es el único que un vector propio no nulo es positivo.*

Para cualquier vector de probabilidad p , $M^k p$ converge en la norma 1⁹ al único vector de probabilidad que es vector propio del valor propio uno.

Como ya se vió estas condiciones aseguran la convergencia de $M^k v$ (para cualquier v) a un múltiplo del vector propio v_1 del valor propio uno. Comencemos probando que si M es de Markov y p es de probabilidad también lo es $M^k p$. Alcanza con probar que Mp es de probabilidad. En efecto, la suma de las componentes de Mp es

$$\sum_{i=1}^n \sum_{j=1}^n m_{ij} p_j = \sum_{j=1}^n p_j \sum_{i=1}^n m_{ij} = \sum_{j=1}^n p_j 1 = 1.$$

La segunda de las igualdades es consecuencia de que M es de Markov (ver definición). Entonces las iteraciones de un vector de probabilidad convergerán a un vector de probabilidad, que será el **único** vector probabilidad que es vector propio del valor propio uno. Por tanto el vector límite no es nulo y se muestra, de pasada, que al escribir un vector de probabilidad v como combinación lineal de los vectores de la base de Jordan, resultará $\alpha_1 \neq 0$.

Convergencia de M^T .

Probaremos ahora que para cualquier $w \in \mathbb{R}^n$, $(M^T)^k w$ converge cuando $k \rightarrow \infty$. Esta prueba será independiente de la estructura de valores y vectores propios de M^T ; sólo utilizará el hecho de que M es de Markov. Necesitaremos el siguiente resultado que presentamos en forma de ejercicio con sugerencia.

Ejercicio: Sean $c = (c_1, \dots, c_n) \in \mathbb{R}^n$, $0 < \gamma < 1/2$ tales que: $0 < \gamma \leq c_j$, $c_1 + \dots + c_n = 1$, entonces el promedio ponderado de los números w_1, \dots, w_n se

⁹Ver definición en Sección 3.

define como $\bar{w} = c_1 w_1 + \dots + c_n w_n$. Sean w_{min} y w_{max} los valores mínimo y máximo de los w 's. Entonces el promedio \bar{w} satisface:

$$\gamma w_{max} + (1 - \gamma)w_{min} \leq \bar{w} \leq (1 - \gamma)w_{max} + \gamma w_{min}, \text{ y } w_{min} \leq \bar{w} \leq w_{max}.$$

Se sugiere, para probar las primeras desigualdades, hacerlo por inducción completa en n , suponiendo, por ejemplo, que al pasar de n a $n + 1$, se agrega el w_{max} .

Para aplicar este resultado a nuestro problema, consideramos $\gamma = \min\{m_{ij}\}$ (si $n > 2$, resulta $\gamma < 1/2$). Como las filas de M^T suman 1, los elementos de $z = M^T w$ son promedios ponderados de los elementos de w . El resultado del ejercicio nos da estimativas para las componentes máxima y mínima de z ,

$$\begin{aligned} z_{max} &\leq (1 - \gamma)w_{max} + \gamma w_{min}, \\ \gamma w_{max} + (1 - \gamma)w_{min} &\leq z_{min}. \end{aligned}$$

Esas desigualdades implican que

$$w_{min} \leq z_{min} \leq z_{max} \leq w_{max} ; \quad z_{max} - z_{min} \leq (1 - 2\gamma)(w_{max} - w_{min});$$

como $0 < (1 - 2\gamma) < 1$ la diferencia entre el valor máximo y mínimo de la iteración resulta una contracción. Por tanto los vectores resultantes de la aplicación sucesiva de M^T convergen en la norma 1 a un vector no nulo con todas sus componentes iguales y distintas de cero si $w_{max} > 0$. Si se comienza el proceso tomando como vector w , cualquier vector con todas sus componentes iguales se observa que este vector z_1 es un vector propio asociado al valor propio uno.

Este es el único vector propio asociado al valor propio uno porque si hubiera otro $z_2 \neq z_1$, resultaría $(M^T)^m z_2 = z_2$, y convergería a z_1 . Absurdo.

Valores propios

Veremos ahora que el uno tiene multiplicidad algebraica uno. Si asumimos que su multiplicidad es $k > 1$, como fue observado al principio de este Apéndice, resultará que la matriz de Jordan tendrá un bloque $I + N$ con $N^k = 0$. Por tanto $(I + N)^m v = \sum_{i=0}^k C_i^m N^i v$ que no puede converger a un vector de coordenadas acotadas puesto que las combinaciones $C_i^m, 0 \leq i \leq k$ van para infinito con m .

Las mismas expresiones usadas al principio de este Apéndice, al introducir la matriz de Jordan, muestran que no pueden haber valores propios con módulo mayor que uno, pues $\lambda^{m-i} \rightarrow \infty$ cuando $m \rightarrow \infty$.

Por último, debemos probar que no hay valores propios complejos de módulo 1. Si existiera un $\lambda = a + bi$ con $b \neq 0$ y argumento $\varphi \neq 0, |\lambda| = 1$, la matriz de Jordan real tendrá una submatriz de la forma $J = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Entonces $J^m =$

$|\lambda|^m \begin{pmatrix} \cos m\varphi & \sin m\varphi \\ -\sin m\varphi & \cos m\varphi \end{pmatrix}$. Por tanto $M^m v$ tendrá dos componentes que no convergerán, porque las que correspondan a esa submatriz estarán dependiendo del valor de $m\varphi$ (si $\varphi \neq 0$, los senos y cosenos correspondientes convergen a por lo menos dos valores distintos al crecer m . Obsérvese que si φ es irracional, los senos y cosenos convergen a todos los valores entre -1 y 1).

Los resultados necesarios para asegurar la convergencia de $M^m v$ cuando v es un vector de probabilidad ya han sido probados. Ahora daremos otras pruebas de los mismos resultados, y completaremos el resultado faltante.

Otra demostración

El teorema de Gershgorin nos permite, sin calcular explícitamente los valores propios, tener una idea de su valor: *Los valores propios de una matriz $\{a_{ij}\}$ se encuentran en los círculos del plano complejo de centro a_{ii} y radio $\sum_{i \neq j} a_{ij}$.* Como nuestra matriz es de Markov los centros son $m_{ii} < 1$ y los radios $1 - m_{ii}$, por lo que los círculos que contienen los valores propios están todos dentro del círculo de centro en el origen y radio 1. Todos esos círculos contienen el punto $(1,0)$, siendo el único común con el círculo unitario centrado en el origen; de lo que se deduce que el único valor propio de módulo 1 es el 1.

Observación: Podemos tener matrices con valor propio 1 pero que M^k no converja o que, el límite $M^k p$ dependa de p . Ejemplos de esto son (en dimensión 2) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ o la matriz identidad.

Sea M una matriz de Markov. Consideramos

$$r = \sup\{\lambda \geq 0 : Mx \geq \lambda x \text{ para algún } 0 \neq x \geq 0\}.$$

Como el uno es valor propio de M existe un vector propio $z = (z_1, \dots, z_n)$. Si llamamos $M^{(j)}$ a la columna j de M la igualdad $Mz = z$ la podemos escribir

$$\sum_{j=1}^n z_j M^{(j)} = z.$$

Si tomamos valor absoluto a ambos lados de la igualdad, utilizamos la desigualdad triangular y llamamos $|z|$ al vector $(|z_1|, \dots, |z_n|)$ obtenemos:

$$M|z| \geq |z|.$$

De donde concluimos que $r \geq 1$.

Mostraremos que $r = 1$. Para ello alcanza con probar que r es valor propio de M pues los valores propios de M tienen módulos menores o iguales a uno.

Afirmación: r es valor propio de M .

Sea $0 \neq \xi \geq 0$ tal que $M\xi \geq r\xi$. Si $M\xi \neq r\xi$ tenemos que el vector $0 \neq y = M\xi - r\xi > 0$ como M es positiva se cumple que $My > 0$ y por lo tanto existe $\varepsilon > 0$ tal que $My \geq \varepsilon M\xi$ o sea

$$M(M\xi - r\xi) \geq \varepsilon M\xi \Rightarrow M(M\xi) \geq (r + \varepsilon)M\xi.$$

Considerando el vector $M\xi$ la desigualdad anterior contradice la definición del r . Entonces $M\xi = r\xi$ como queríamos mostrar. Obsérvese que como $M > 0$ llegamos a que el vector $\xi = (\xi_1, \dots, \xi_n)$ es **positivo**.

Afirmación: El uno es el único valor propio asociado con un vector propio z con todas sus componentes ≥ 0 .

Este enunciado sólo tiene sentido para valores propios reales porque los valores propios no reales (complejos) deben tener vectores propios con algunas o todas sus coordenadas complejas. Sea $z = (z_1, \dots, z_n)$ un tal vector propio de M , con valor propio λ . Algún z_i deber ser mayor que cero (*¿Por qué?*). Sea

$$\alpha = \min \left\{ \frac{z_i}{\xi_i}, \text{ con } z_i \neq 0 \right\}.$$

De la definición de α vemos que existe algún p , $1 \leq p \leq n$ con $z_p = \alpha \xi_p > 0$. Como $\lambda z = Mz \geq \alpha M\xi = \alpha \xi$ si miramos la componente p -ésima $\lambda z_p \geq \alpha \xi_p = z_p \Rightarrow \lambda \geq 1$.

Referencias

- [BDJ] M. Berry, Z. Drmac & E. Jessup, *Matrices, Vector Spaces and Information Retrieval*, SIAM Review **41** (1999), 335-362.
- [BP] Sergey Brin & Lawrence Page, *The anatomy of a large scale hypertextual web search engine*. Computer Networks and ISDN Systems, **33** (1998), 107-117.
- [Bra] Daniel Brandt, *PageRank: Google's original sin*.
<http://www.google-watch.org/pagerank.html>
- [DR] Pedro Domingos & Matthew Richardson, *The intelligent surfer: probabilistic combination of link and content information in PageRank*. Advances in Neural Information Processing Systems **14** (2002).
- [Ec] *How good is google?* The economist, print edition, October 30th, 2003.

- [EF] *A Survey of Google's PageRank*.
<http://pr.efactory.de>
- [Gr] Juan-Miguel Gracia, *Algebra Lineal tras los buscadores de Internet*.
<http://www.vc.ehu.es/campus/centros/farmacia/deptos-f/depme/gracia1.htm>
- [Ha] O. Haggstrom, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.
- [HK] T. Haveliwala and S. Kamvar, *The Second Eigenvalue of the Google Matrix*. A Stanford University Technical Report
<http://dbpubs.stanford.edu:8090/pub/2003-20>
- [Ka] Jerry Kazdan, *Solving Equations, An elegant Legacy*. American Math. Monthly, **105** (1998), 1-21. Versión expandida en <http://www.math.upenn.edu/~kazdan>
- [Ke] M. Kendall, *Further contributions to the theory of paired comparisons*. Biometrics **11** (1955), 43-62.
- [La] Pierre Lazuly, El mundo según Google. Le Monde diplomatique/el Dipló/, edición cono sur, Octubre 2003, 36-37.
- [MC] C.R. MacCluer, *The many proofs and applications of Perron's Theorem*, SIAM Review **42** (2000), 487-498.
- [Mo] Cleve Moler, *The World's Largest Matrix Computation. Matlab News and notes, Cleve's corner*.
<http://www.mathworks.com/company/newsletter/clevescorner/>
- [Pe] *Motores de consulta*.
<http://librosdigitales.net/eureka/eureka0903/motores.htm>
- [Re] S. Resnick, *Adventures in Stochastic Processes*, Birkhauser, 1992.
- [Se] E. Seneta, *Non-negative Matrices and Markov Chains*. 2md. Edition. Springer, 1981.
- [We] T. Wei, *The algebraic foundations of ranking theory*. Cambridge University, England (1952). Tesis no publicada.
- [Wi] Herbert Wilf, *Searching the web with eigenvectors*.
<http://www.math.upenn.edu/~wilf/>

ROBERTO MARKARIAN & NELSON MÖLLER
IMERL - FACULTAD DE INGENIERIA
UNIVERSIDAD DE LA REPÚBLICA - URUGUAY

INFORMACIÓN INTERNACIONAL

La Esquina Olímpica

Rafael Sánchez Lamonedá

En esta oportunidad reseñaremos la actividad olímpica del segundo semestre de 2004. Como ya es tradición participamos en la 45 Olimpiada Internacional de Matemáticas, IMO, celebrada en Atenas, Grecia, del 6 al 18 de Julio y la XIX Olimpiada Iberoamericana de Matemáticas, OIM, celebrada en Castellón, España, del 18 al 26 de Septiembre. Además recibimos los resultados de la Olimpiada Bolivariana de Matemáticas.

En todas estas competencias ganamos premios, mostrando que vamos por buen camino y que nuestros estudiantes alcanzan niveles de competencia adecuados, cuando se trabaja fuerte con ellos. Antes de enumerar los premios y los ganadores, quiero aprovechar la oportunidad para reconocer el trabajo desinteresado que llevan adelante Adolfo Rodríguez, Héctor Chang, David Seguí y Tomás Kabbabe, jóvenes exolímpicos, que ahora estudian matemáticas o carreras afines y que comparten sus actividades con la labor de entrenamiento de los estudiantes que conforman nuestras delegaciones. A continuación se listan las delegaciones y premios obtenidos en la segunda mitad del año.

45^a IMO

Leonardo Urbina. Mención Honorífica.

Rodrigo Ipince

Tomás Kabbabe. Tutor de delegación.USB.

Rafael Sánchez. Jefe de delegación. UCV.

XIX OIM

Maximiliano Liprandi. Medalla de Bronce.

Rodrigo Ipince. Mención Honorífica.

Roland Hablutzel. Mención Honorífica.

Andrés Guzmán. Mención Honorífica.

David Seguí. Tutor de la delegación. USB.

Henry Martínez. Jefe de la delegación. UPEL-IPC.

Mención aparte haremos de la Olimpiada Bolivariana de Matemáticas. Este evento lo organiza la Universidad Antonio Nariño de Colombia y participan por correspondencia los países andinos y Panamá. Hay dos niveles de competencia, los cuales tienen que ver con la escolaridad de los concursantes, *nivel intermedio* y *nivel superior*. Los resultados obtenidos este año son bastante buenos. En el nivel intermedio el joven Andrés Guzmán ganó medalla de oro y en el nivel superior Leonardo Urbina obtuvo medalla de oro y Tamara Mendt y Carlos Molina ganaron medallas de bronce. Una información completa sobre este evento y los otros que hemos reseñado, la pueden encontrar visitando nuestra página web, <http://ares.unimet.edu.ve/matematica/acm>. Sin más preámbulos los dejo con los problemas de la 45ª IMO.

45ª Olimpiada Internacional de Matemáticas

Primer día

Lunes 12 de julio de 2004

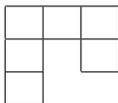
Problema 1. Sea ABC un triángulo acutángulo con $AB \neq AC$. La circunferencia de diámetro BC corta a los lados AB y AC en M y N , respectivamente. Sea O el punto medio del lado BC . Las bisectrices de los ángulos $\angle BAC$ y $\angle MON$ se cortan en R . Demostrar que las circunferencias circunscritas de los triángulos BMR y CNR tienen un punto en común que pertenece al lado BC .

Problema 2. Encontrar todos los polinomios $P(x)$ con coeficientes reales que satisfacen la igualdad

$$P(a - b) + P(b - c) + P(c - a) = 2P(a + b + c)$$

para todos los números reales a, b, c tales que $ab + bc + ca = 0$.

Problema 3. Un *gancho* es una figura formada por seis cuadrados unitarios como se muestra en el diagrama



o cualquiera de las figuras que se obtienen de ésta rotándola y/o reflejándola.

Determinar todos los rectángulos $m \times n$ que pueden cubrirse con ganchos de modo que

- el rectángulo se cubre sin huecos y sin superposiciones;

- ninguna parte de ningún gancho sobresale del rectángulo.

Duración del examen: 4 horas 30 minutos.

Cada problema vale 7 puntos.

45^a Olimpiada Internacional de Matemáticas

Segundo día

Martes 13 de julio de 2004

Problema 4. Sea $n \geq 3$ un entero. Sean t_1, t_2, \dots, t_n números reales positivos tales que

$$n^2 + 1 > (t_1 + t_2 + \dots + t_n) \left(\frac{1}{t_1} + \frac{1}{t_2} + \dots + \frac{1}{t_n} \right).$$

Demostrar que t_i, t_j, t_k son las medidas de los lados de un triángulo para todos los i, j, k con $1 \leq i < j < k \leq n$.

Problema 5. En un cuadrilátero convexo $ABCD$ la diagonal BD no es la bisectriz ni del ángulo $\angle ABC$ ni del ángulo $\angle CDA$. Un punto P en el interior de $ABCD$ verifica

$$\angle PBC = \angle DBA \quad \text{y} \quad \angle PDC = \angle BDA.$$

Demostrar que los vértices del cuadrilátero $ABCD$ pertenecen a una misma circunferencia si y sólo si $AP = CP$.

Problema 6. Un entero positivo es *alternante* si, en su representación decimal, en toda pareja de dígitos consecutivos uno es par y el otro es impar. Encontrar todos los enteros positivos n tales que n tiene un múltiplo que es alternante.

Duración del examen: 4 horas 30 minutos.

Cada problema vale 7 puntos.

LIBROS

Guillermo Martínez, *Los Crímenes de Oxford*
Editorial Destino, Barcelona, Colección Ancora y Delfín
Volumen 992, 2004

Reseñado por Carlos Augusto Di Prisco

El último teorema de Fermat y el anuncio de su demostración por Andrew Wiles, el teorema de incompletitud de Gödel y sus implicaciones filosóficas, la doctrina pitagórica y su representación mística de los números, todos estos hitos de la historia de las matemáticas, antigua y reciente, aparecen en esta obra de Guillermo Martínez ligados a una serie de misteriosos crímenes y a una relación amorosa para ofrecer al lector una novela enmarcada en la tradición de la novela negra inglesa. En este ámbito nuestro autor elabora un juego en el que se combinan las actividades casi rituales de la vida académica y la rica complejidad de las relaciones humanas.

En Guillermo Martínez se conjugan dos vidas, la del matemático y la del escritor. Nació en Bahía Blanca, Argentina en 1962. Doctor en matemáticas, docente e investigador en la Facultad de Ciencias exactas de la Universidad de Buenos Aires, se ha destacado como nueva figura literaria en Argentina y ahora a nivel mundial. *Los Crímenes de Oxford*, su tercera novela publicada, recibió el premio Planeta Argentina 2003. En 1989 publicó la colección de cuentos, *Infierno grande*, que contiene piezas de gran valor, algunas de las cuales dejan en el lector una satisfacción no desprovista de angustia por los desenlaces de las situaciones narradas; otras, deleitarán a todo aquel que se haya enfrentado alguna vez a ochenta estudiantes principiantes sin más instrumentos de defensa que una barra de tiza y una pizarra. Guillermo Martínez ha publicado también una serie de trabajos matemáticos en revistas de gran prestigio científico tales como *Algebra Universalis*, *Studia logica*, *Order*, y *Annals of Pure and Applied Logic*. No es la primera vez que se da una afortunada combinación de matemáticas y literatura en un autor. Bertrand Russell, cuyo trabajo matemático de principios del siglo veinte tuvo una gran importancia para el desarrollo de los fundamentos de las matemáticas, recibió el Premio Nobel de literatura; y el autor de *Alicia en el país de las maravillas*, Lewis Carroll, era también especialista en lógica matemática.

Un joven matemático argentino llega a Oxford con una beca para estudiar durante un año bajo la guía de una destacada profesora. A los pocos días de su llegada, la propietaria de la casa donde ha alquilado una habitación aparece asesinada, y así, sin quererlo, se encuentra en el centro de una investigación policial que debe ir descifrando las señales cargadas de contenido matemático

que a manera de reto intelectual deja el supuesto criminal, aparentemente para medir fuerzas con uno de los matemáticos más famosos de la época, Arthur Seldom, quien años antes había publicado un libro sobre series en el que incluye, justamente, una sección sobre crímenes en serie. Ya no de forma tan azarosa se envuelve también este joven en una relación amorosa que lo sumerge aún más en la intriga. El estudiante, quien es el narrador de la historia, observa cuidadosamente su alrededor y describe modales y características inglesas con fino humor. De los ingleses y su cuidadoso ahorro de gestos dice: “aún en esa risa espontánea había algo contenido, como si se tomaran una libertad infrecuente de la que no debían abusar...”. Martínez reconoce, en una entrevista reciente, una cierta coincidencia “entre su mirada y la del protagonista de su novela”; de hecho, algunos pasajes parecieran tener un revelador carácter autobiográfico:

“¿Porqué se hizo Ud. matemático?- me preguntó sorprendentemente. -No sé- quizás fue una equivocación, siempre creí que iba a seguir una carrera humanística. Supongo que lo que me atrajo de las matemáticas es la clase de verdad que encierran los teoremas: atemporal, inmortal, suficiente en si misma, y a la vez absolutamente democrática.”

Esta novela fue publicada en Argentina en 2003 con el título *Crímenes imperceptibles*, a mi manera de ver, más apropiado que el título adoptado para la edición española de 2004. Durante la búsqueda de pistas que orienten la investigación hacia el culpable, el joven argentino y el profesor Arthur Seldom, conocido por sus estudios sobre las implicaciones del teorema de Gödel, llegan a la conclusión de que el autor de los asesinatos desea precisamente que sus actos sean imperceptibles, crímenes perfectos que no dejen huellas, que no causen daño ni sufrimiento innecesario, buscando sus víctimas entre personas cuyas expectativa de vida ya ha sido sobrepasada. El crimen perfecto no es aquel que permanece sin solución, sino el que “se resuelve con un culpable equivocado”. En palabras del autor, se trata en este caso de crímenes casi abstractos.

Quizás el verdadero protagonista de la novela es el profesor Seldom quien se convierte en la figura clave para descifrar los mensajes que anuncian las muertes. Este personaje hace reflexiones profundas, y nos hace intuir una compleja personalidad de la que quedamos con ganas de saber mucho más.

“Hay una diferencia entre la verdad y la parte de la verdad que puede demostrarse, ése es en realidad un corolario de Tarski sobre el teorema de Gödel- dijo Seldom-. Por supuesto, los jueces, los forenses, los arqueólogos, sabían esto mucho antes que los matemáticos. Pensemos en cualquier crimen con sólo dos posibles sospechosos. Cualquiera de ellos sabe toda la verdad que interesa: *yo fui* o *yo no fui*. Pero la justicia no puede acceder directamente a esa verdad, y tiene que recorrer un penoso camino indirecto para reunir pruebas: interrogatorios, coartadas, huellas digitales... Demasiadas veces las evidencias que se encuentran no alcanzan para probar la culpabilidad de uno ni la inocencia del otro. En el fondo, lo que demostró Gödel en 1930 con su teorema de incompletitud

es que exactamente lo mismo ocurre en la matemática... ”

“...el método axiomático puede ser a veces tan insuficiente como los criterios precarios de aproximación de la justicia...”

Esta bien lograda novela corta, de poco más de doscientas páginas, capta desde la primera línea la atención del lector, y lo mantiene en suspenso hasta el inesperado desenlace final. Guillermo Martínez demuestra una vez más su calidad de escritor. Su primera novela, *Acerca de Roderer*, fue publicada en 1996, y recibió comentarios elogiosos de la crítica. Posteriormente publicó *La mujer del maestro* (1998), y un libro de ensayos titulado *Borges y la matemática* (2003).

CARLOS AUGUSTO DI PRISCO

INSTITUTO VENEZOLANO DE INVESTIGACIONES CIENTÍFICAS

EDITORIAL

Quince años de la Asociación Matemática Venezolana

En mayo de 1989 se realizó en la Universidad de Oriente, Cumaná, una reunión para promover la creación de la Asociación Matemática Venezolana. Unos sesenta matemáticos estuvieron presentes, y acordaron establecer una asociación sin fines de lucro y de carácter científico con el objetivo de contribuir al desarrollo de la matemática en Venezuela. El 10 de enero de 1990 se registró el acta constitutiva de la Asociación Matemática Venezolana para formalizar su existencia y para darle personalidad jurídica. La Asociación se inició con base en algunas actividades que la comunidad matemática venía desarrollando en Venezuela desde un poco antes: la Escuela Venezolana de Matemáticas y las Jornadas Venezolanas de Matemáticas. Desde entonces las actividades de la AMV se han consolidado y diversificado. Hoy en día, además de las Jornadas y la Escuela, la Asociación promueve la realización de los talleres de formación matemática TForMa, y publica un boletín con dos números por año.

Las Jornadas Venezolanas de Matemáticas se realizan anualmente como encuentro de los matemáticos venezolanos y son el foro natural para la presentación de resultados de investigación de los distintos grupos que mantienen su actividad en las universidades e institutos nacionales. Las jornadas se organizan en torno a sesiones temáticas de presentación de trabajos, e incluyen algunas conferencias plenarias. En marzo de 2005 se realizarán las décimo octavas jornadas en la Universidad Central de Venezuela.

La Escuela Venezolana de Matemáticas se realiza anualmente también desde hace dieciocho años. Su sede permanente es la Facultad de Ciencias de la Universidad de Los Andes. En cada oportunidad, la Escuela ofrece varios cursos intensivos, generalmente cuatro, de nivel de postgrado, con el objetivo principal de ofrecer a los estudiantes que se inician en el cuarto nivel de estudios de matemáticas una perspectiva amplia de los temas que se desarrollan en el país. Se pretende que en cada curso se llegue a plantear problemas o direcciones de investigación que puedan servir como punto de partida para la realización de tesis y trabajos de grado. Se pide al profesor de cada curso que escriba unas notas de una cien páginas que sirvan de bibliografía básica para el curso y de material de estudio que el interesado pueda utilizar posteriormente. Se cuenta ya con una colección de más de ochenta títulos en diversas áreas de las matemáticas escritos para estas escuelas, lo que constituye una fuente bibliográfica valiosa que está al alcance de cualquier estudioso que desee profundizar conocimientos.

Los Talleres de Formación Matemática TForMa tienen la finalidad de darle a los estudiantes de las licenciaturas de matemáticas una oportunidad de recibir información científica complementaria a los cursos que usualmente se dictan

en los programas de licenciatura y a la vez sirven de lugar de encuentro de estudiantes de distintas universidades. Estos talleres se iniciaron en 2000 y se han realizado en distintas instituciones de educación superior. Para los cursos de los talleres también se elaboran textos de apoyo.

El Boletín de la Asociación se publica de manera continua desde su primer número aparecido en 1994. Inicialmente su contenido fue más de carácter informativo y de divulgación científica especializada, pero recientemente se han comenzado a publicar en el Boletín artículos de investigación con resultados originales. Además, el Boletín incluye secciones de información matemática nacional e internacional, temas de matemáticas recreativas, y una sección llamada La Esquina Olímpica que contiene información sobre competencias matemáticas y los problemas que allí se plantean.

La Asociación ha dado su apoyo también al proyecto de olimpiadas matemáticas, tanto a nivel nacional como internacional, y en la actualidad se están estableciendo los mecanismos para que todos los capítulos de la Asociación participen activamente en este programa.

Alguna información adicional sobre la AMV y sus actividades se puede obtener en <http://amv.ivic.ve>

CARLOS A. DI PRISCO
PRESIDENTE DE LA AMV

AGRADECIMIENTO

Agradecemos la colaboración prestada por las siguientes personas en el trabajo editorial del volumen XI del Boletín de la AMV:

Isaías Alonso Mallo, Julio César Alvarez, Diómedes Bárcenas, Pedro Berrizbeitia, Antonio Campillo, Carlos Di Prisco, Vicenç Font, José Giménez, Luis Gómez Sánchez, Henryk Gzyl, John Neuberger, Neptalí Romero, Jesús Tapia y Juan Tena