(*Kurt Gödel 1906-1978*, ver p. 177)

Vol. XIII • No. 2 • Año 2006

Boletín de la Asociación Matemática Venezolana

Volumen XIII, Número 2, Año 2006

I.S.S.N. 1315–4125

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

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/

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

Boletín de la Asociación Matemática Venezolana

Vol. XIII • No. 2 • Año 2006

Boletín de la Asociación Matemática Venezolana
Volumen XIII, Número 2, Año 2006

ARTÍCULOS

DIVULGACIÓN MATEMÁTICA

INFORMACIÓN INTERNACIONAL

Ramsey theory for structures: Nešetřil's result on finite metric spaces

Carlos Augusto Di Prisco

Contents

1 Introduction

The main objective of these notes is to give a mostly self contained presentation of J. Nešetřil's recent proof of the Ramsey property for the class of ordered finite metric spaces [6]. This result was motivated by a question posed in [3], where the connections between Ramsey theory and the dynamics of groups of automorphisms are explored (see also [5]. These notes were written for a course on Ramsey Theory given by the author in Caracas during the first term of 2005.

A class of finite ordered structures is a Ramsey class if given structures A, B in the class, and a positive integer t, there is another structure C in the class such that for every partition of the set of substructures of C which are isomorphic to A into t pieces, there is a substructure of C isomorphic to B which is homogeneous, in the sense that all of its substructures isomorphic to A are in the same piece. Often, a partition into t pieces is seen as a t-coloring, and a homogeneous set for the partition is then said to be monochromatic.

Finite ordered metric spaces can be seen as labelled binary relational structures of a particular kind. For such a structure the triangular inequality can be obtained using the notion of l-metric system (see section 3); we will see that a finite binary relational structure of the appropriate kind is a metric space if it

is l-metric for a sufficiently large number l. The Ramsey property is proved by induction on l for the class of l- metric systems (Main Lemma). The first step of the induction (the case $l = 1$) follows from the fact that the class of finite ordered relational structures has the Ramsey property (Theorem 4 of section 2), which is a result of [8].

Two of the most emblematic results of Ramsey Theory are Ramsey's theorem about partitions, or colorings, of the k element subsets of a finite set, and the Hales -Jewett theorem about colorings of the n^{th} power of a finite set. Ramsey's theorem can be considered the starting point of the theory, and has been extended in many directions. The Hales-Jewett theorem is a powerful result which contains the combinatorial essence of the famous result of van der Waerden about arithmetic progressions. Both results will be used in the following sections.

We introduce some notation in order to state these two theorems. Every natural number n is identified with the set of its predecessors $\{0, 1, \ldots, n-1\}$. Given a set A and $k \in \mathbb{N}$, $A^{[k]}$ denotes the set $\{s \subseteq A : |s| = k\}$. Given positive integers n, m, k, t, the partition symbol

$$n \to (m)^k_t$$

is used to express that for every coloring $c : n^{[k]} \to t$ there is $H \subseteq n$ with $|H| = m$ such that c is constant on $H^{[k]}$.

Theorem 1 *(Ramsey's Theorem) Given positive integers k, r and m there is a positive integer n such that*

$$n \to (m)^r_k.$$

Before stating the Hales-Jewett theorem we need some definitions. Let $k \in \mathbb{N}$ and let $\Lambda_k = \{1, 2, \ldots, k\}$. Given $n \in \mathbb{N}$, Λ_k^n is the set of n-tuples of elements of Λ_k, or words of length n in the alphabet Λ_k.

Definition 2 *A combinatorial line in Λ_k^n is a set $\{x_1, x_2, \ldots, x_k\}$ of elements of Λ_k^n such that for each coordinate j, $1 \leq j \leq n$, either*

$$x_1(j) = x_2(j) = \cdots = x_k(j)$$

or

$$x_i(j) = i \text{ for every } i = 1, \ldots, k,$$

and the second possibility occurs at least once.

Another way to define combinatorial lines is by variable words. A variable word is a word in the alphabet $\{1, \ldots, k, x\}$ where x appears at least once. The symbol x acts as a variable. Given a variable word $w(x)$, we write $w(i)$ to denote the word (in the alphabet Λ_k) resulting from substituting i for x in $w(x)$. If $w(x)$ is a variable word, the combinatorial line associated to $w(x)$ is $\{w(1), w(2), \ldots, w(k)\}$.

Theorem 3 *(Hales-Jewett)*
Given positive integers $k, r \in \mathcal{N}$, there is a number $n = n(k, r)$ such that for every r-coloring Λ_k^n there is a monochromátic combinatorial line.

The proofs of these two theorems can be found in [2, 4].

2 Ramsey properties for relational structures.

We consider finite relational structures defined in the following way. A type is a sequence $\Delta = (\delta_i : i \in I)$ of natural numbers, where I is a finite set. Given a type Δ, a structure of type Δ is a pair $(X, \mathcal{M})$ such that

(i) X is a linearly ordered set, and

(ii) $\mathcal{M} = (\mathcal{M}_i : i \in I)$, and $\mathcal{M}_i \subseteq X^{[\delta_i]}$

The linear order of X is called the standard order.

$Rel(\Delta)$ denotes the class of finite structures of type Δ. Note that these relational structures are labelled hypergraphs.

Given structures $A = (X, \mathcal{M})$ and $B = (Y, \mathcal{N})$ of type Δ, a function $f : X \to Y$ is an embedding if

(i) f is one-one and monotone with respect to the standard linear orderings of X and Y, and

(ii) for every $i \in I$ and every subset M of $X^{[\delta_i]}$, $M \in \mathcal{M}_i$ if and only if $\{f(x) : x \in M\} \in \mathcal{N}_i$.

We write $A \leq B$ to express that there is an embedding from A to B, and $A \cong B$ when A and B are isomorphic.

Given structures A, B, $\binom{B}{A}$ denotes the set of all substructures of B which are isomorphic to A.

If $A \leq B \leq C$, the partition symbol

$$C \to (B)_t^A$$

expresses that for every coloring $c : \binom{C}{A} \to t$, there is a $B' \in \binom{C}{B}$ such that the collection $\binom{B'}{A}$ is monochromatic.

Theorem 4 *For any given type Δ, the class $Rel(\Delta)$ is a Ramsey class. In other words, given structures A, B in $Rel(\Delta)$ with $A \leq B$, and a positive integer t, there is a structure C in $Rel(\Delta)$ such that $B \leq C$ and $C \to (B)_t^A$.*

To prove this theorem, we define partite systems and use the amalgamation technique, following [4]. This is done in the next two sections.

2.1 Partite systems.

Definition 5 *Given a type $\Delta = (\delta_i : i \in I)$ and $a \in \mathbb{N}$, an a-partite system of type Δ is a pair $((X_j)_{j=1}^a, \mathcal{M})$ where*

(a) $X = \bigcup_{i=1}^a X_i$ is a linearly ordered set satisfying $X_1 < X_2 < \cdots < X_a$, i.e., for every $i, j \in \{1, \ldots, a\}$ with $i < j$, if $x \in X_i$ and $j \in X_j$, then $x < y$.

(b) $\mathcal{M} = (\mathcal{M}_i : i \in I)$, and $\mathcal{M}_i \subseteq X^{[\delta_i]}$

(c) $|M \cap X_j| \leq 1$ for every $M \in \mathcal{M}_i$, $j = 1, \ldots, a$, $i \in I$.

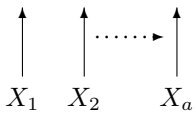


Fig. 1 Partite system

Given a subset $Y \subseteq X$, we denote by $tr(Y)$ the trace of Y, i.e. the set $\{j : X_j \cap Y \neq \emptyset\}$.

A system A is transversal if $|X_j| = 1$ for every $j = 1, \ldots, a$.

The system A is a subsystem of $B = ((Y_k)_{k=1}^b, \mathcal{N})$ if there exists a monotone injection $h : \{1, \ldots, a\} \to \{1, \ldots, b\}$ such that $X_j \subseteq Y_{h(j)}$ for every $j = 1, \ldots, a$ and $\mathcal{M}_i = \mathcal{N}_i \cap X^{[\delta_i]}$ for $i \in I$. An isomorphism is an order preserving isomorphism of structures which also preserves parts.

Lemma 6 *(The Partite Lemma) Let A and B be a-partite systems of type Δ, A transversal, and let t be a positive integer, then there exists an a-partite system C of type Δ such that*

$$C \to (B)_t^A.$$

Proof. Set $A = ((X_j)_{j=1}^a, \mathcal{M})$ and $B = ((Y_j)_{j=1}^a, \mathcal{N})$. Since A is transversal, we may assume without loss of generality that $\bigcup_{i \in I} \mathcal{M}_i$ is the set of all subsets of X. We also can assume that every vertex $y \in Y$ belongs to a copy of A. This is so because otherwise we can work with B^*, the subsystem of B induced by $\binom{B}{A}$, which satisfies this property, and if $C^*$is such that $C^* \to (B^*)_t^A$, then we can obtain C such that $C \to (B)_t^A$ enlarging each copy of B^* in C^* to a copy of B.

We fix a sufficiently large positive integer N, and define an a- partite system $C = ((Z_j)_{j=1}^a, \mathcal{O})$, $\mathcal{O} = (\mathcal{O}_i : i \in I)$ where $Z_j = Y_j \times \cdots \times Y_j$ (N times). Thus,

every element of Z_j has the form $(x_l : l = 1, \dots, N)$ with each $x_l \in Y_j$. We will say more about the number N later on.

Set $Z = \bigcup_{j=1}^{a} Z_j$. For each $l = 1, \dots, N$, the projection $\pi_l : Z \to Y$ is defined by $\pi_l(x_k : k = 1, \dots, N) = x_l$. For every l, π_l maps Z_l into Y_l.

We now define $\mathcal{O} = (\mathcal{O}_i : i \in I)$. Put first $\mathcal{N}_i = \mathcal{N}_i' \cup \mathcal{N}_i''$, where $\mathcal{N}_i'$ is the set of edges of $\mathcal{N}_i$ which belong to a copy of A in B, and $\mathcal{N}_i'' = \mathcal{N}_i \setminus \mathcal{N}_i'$.

We put

$$\{(x_1^k, \dots x_N^k) : k = 1, \dots, n_i\} \in \mathcal{O}_i$$

if $tr(\{x_j^k : k = 1, \dots, n_i\}) = tr(\{x_{j'}^k : k = 1, \dots, n_i\})$ for all $j, j' \leq N$, and one of the following possibilities occur:

1. $\{x_j^k : k = 1, \dots, n_i\} \in \mathcal{N}_i'$ for every $j = 1, \dots, N$,

2. there exists a non-empty set $\Gamma \subseteq \{1, \dots, N\}$ such that

$\{x_j^k : k = 1, \dots, n_i\} = \{x_{j'}^k : k = 1, \dots, n_i\} \in \mathcal{N}_i''$ for all $j, j' \in \Gamma$, and
$\{x_j^k : k = 1, \dots, n_i\} \in \mathcal{N}_m'$ for all $j \notin \Gamma$

In general, $m \neq i$, but m is uniquely determined by $tr(x_j^k : k = 1, \dots, n_i)$.

We now prove that $C \to (B)_t^A$ provided N is large enough. This will follow from the two facts stated below.

Fact 1. $A' \in \binom{C}{A}$ if and only if $\pi_l(A') \in \binom{B}{A}$ for every $l = 1, \dots, N$. This is an immediate consequence of the definition of $\mathcal{O}$. If $\pi_l(A') \in \binom{B}{A}$ for every $l = 1, \dots, N$, then clearly $A' \in \binom{C}{A}$. Conversely, let $A' = \{(x_1^k, \dots, x_N^k) : k = 1, \dots, a\}$ be a substructure of C which forms a copy of A, and suppose that $\{(x_1^k, \dots, x_N^k) : k = k_{m_1}, \dots, k_{m_{n_i}}\} \in \mathcal{O}_i$, then for every $j = 1, \dots, N$, the projection $\{x_j^k : k = k_{m_1}, \dots, k_{m_{n_i}}\} \in \mathcal{N}_i$. This is so because by the definition of $\mathcal{O}_i$, this projection is always in $\mathcal{N}_i$: if the second case of the definition occurs, then either $\{x_j^k : k = k_{m_1}, \dots, k_{m_{n_i}}\}$ belongs to $\mathcal{N}_i''$ and thus to $\mathcal{N}_i$, or to $\mathcal{N}_m'$ for some m, but since this edge forms part of a copy of A (because it is in $\mathcal{N}_m'$), it is also in $\mathcal{N}_i$. Note that an edge of A' in $\mathcal{O}_i$ must come then from the first clause of the definition of $\mathcal{O}_i$.

Let $\binom{B}{A} = \{A_1, \dots, A_r\}$, and put $R = \{1, \dots, r\}$. Given $\alpha = (\alpha_1, \dots, \alpha_N) \in R^N$, denote by $V(\alpha)$ the set of all the vertices $x \in Z$ which satisfy $\pi_j(x) \in A_{\alpha_j}$. If L is a combinatorial line in R^N, set $V(L) = \bigcup_{\alpha \in L} V(\alpha)$. By Fact 1, the set $\binom{C}{A}$ is in $1-1$ correspondence with R^N.

Fact 2. Let L be a combinatorial line of R^N. Then, $V(L)$ induces a copy of B in C.

Clear from the definition of C, since B is the union of the r copies of A it contains. Notice that the second option in the definition of $\mathcal{O}_i$ is important to obtain a copy of B in C from the union of all these copies of A; notice also that our assumption that every subset of X is an edge of A is used here.

Now, by the Hales-Jewett Theorem, if N was chosen large enough, for every partition of R^N into t classes, there is a combinatorial line contained in one of the classes. This implies $C \to (B)^A_t$. In fact, if $\binom{C}{A} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_t$ is a partition, it induces a partition $R^N = \mathcal{A}'_1 \cup \cdots \cup \mathcal{A}'_t$ by $\alpha \in \mathcal{A}'_i$ if $V(\alpha)$ induces a copy of A which is in $\mathcal{A}_i$. By the Hales-Jewett Theorem, there is a monochromatic line L which, by Fact 2, induces a $B'' \in \binom{C}{B}$, such that $\binom{B''}{A}$ is contained in a single class $\mathcal{A}_i$. □

2.2 The partite construction.

To prove Theorem 4, we use an amalgamation technique called the partite construction first used by Nešetřil and R ödl (see [8] , [4]).

Proof of Theorem 4. Let t, and A, B be given as in the statement of Theorem 4. We consider A as a transversal a-partite system and B as a transversal b-partite system. Put $B = ((y_1, \ldots, y_b), \mathcal{N})$ Let p be the minimal n such that $n \to (b)^a_t$, and let $q = \binom{p}{a}$, and put $\binom{\{1,\ldots,p\}}{\{1,\ldots,a\}} = \{M^1, \ldots, M^q\}$.

We will define a sequence $P^0, P^1, \ldots, P^q$ of "pictures", the last of which, P^q, will be the desired system C.

Let $P^0 = ((X^0_i)^p_{i=1}, \mathcal{O})$ be a p-partite system such that for each choice of b parts of P^0, $X^0_{i_1}, \ldots, X^0_{i_b}$, the subsystem of P^0 induced by them contains a copy of B. This can be obtained taking a disjoint union of copies of B.

If the picture $P^k = ((X^k_i)^p_{i=1}, \mathcal{O}^k)$ has been defined, consider M^{k+1} and the a-partite system D^{k+1} induced in P^k by the parts X^k_i for which i belongs to M^{k+1}. By the Partite Lemma 6, there is an a-partite system E^{k+1} such that

$$E^{k+1} \to (D^{k+1})^A_t.$$

Extend each copy of D^{k+1} in E^{k+1} to a copy of P^k in such a way that the distinct copies of P^k intersect only in vertices of E^{k+1}. The resulting a-partite system is P^{k+1}. Finally $C = P^q$. We claim that C has the desired properties.

By a backward induction we verify that

$$C \to (B)^A_t.$$

In the inductive step from $k+1$ to k, by the use of the partite lemma in the construction of P^{k+1}, we can find a copy of P^k in P^{k+1} in which all copies of A with trace M^k have the same color.

We end up with a copy P of P^0 such that the color of a copy of A in P depends only on its trace. This induces a t-coloring of $p^{[a]}$, the collection of a-element subsets of p: the color of s is defined as the color of any copy of A whose trace is s. Since $p \to (b)^a_t$, there is a monochromatic subset of p of size b.

By construction, the subsystem of P_0 induced by any b elements of p contains a copy of B, and therefore there is a monochromatic copy of B in P. □

Given a type Δ, if for every pair of structures A, B of type Δ such that B has substructures isomorphic to A, there is a structure C of type Δ such that $C \to (B)^A_2$, then for every positive integer r, and every pair A, B of structures with the same properties as above, there exists C such that $C \to (B)^A_r$.

3 Finite metric spaces.

In this section we present a proof due to J. Nešetřil of the Ramsey property for the class of finite ordered metric spaces. This result answers a question of [3], and gives information about the group of automorphisms of the Urysohn space.

A finite metric space can be viewed as a labelled complete finite graph: a pair of elements forms an edge labelled by the distance between them. These graphs are, in turn, special cases of relational structures.

We denote by *Rel* the class of all finite ordered relational structures of all possible finite types. Given $d, D \in \mathbb{R}$, with $d < D$, $Rel(d, D)$ is the subclass of *Rel* of all systems $A = (X, (R_i; i \in I))$ where I is a finite subset of the interval $[d.D]$, and for every $i \in I$, $R_i \subseteq X^{[2]}$.

Given structures $A = (X, (R_i; i \in I))$ and $B = (Y, (S_i; i \in J))$, a function $f : X \to Y$ is an embedding if

(i) f is one-one and monotone with respect to the standard linear orderings of X and Y, and

(ii) for every $i \in I$ and every pair $\{x, y\}$ of elements of X, $\{x, y\} \in R_i$ if and only if $\{f(x), f(y)\} \in S_i$ (thus, $I \subseteq J$).

If the embedding f is a bijection,, we say it is an isomorphism. Given structures A, B, $\binom{B}{A}$ denotes the set of all substructures of B which are isomorphic to A.

As a consequence of Theorem 4 we have the following theorem, which will be used in the proof of the result for finite ordered metric spaces.

Theorem 7 *(Nešetřil, [8]) For every pair of real numbers d, D, $0 < d < D$, the class $Rel(d, D)$ is Ramsey.*

Let $0 < d < D$ be real numbers, and let l be a positive integer. Consider a structure $A = (X, (R_i : i \in I))$ where I is a finite subset of the interval $[d, D]$ and each R_i is a symmetric binary relation. An edge of $\{x, y\} \in R_i$ of A is l- metric if for every path $x = x_0, x_1, \ldots, x_t = y$, with $t \leq l$ such that $\{x_{k-1}, x_k\} \in R_{i_k}$ (i.e. the distance between x_{k-1} and x_k is i_k) it holds that $i \leq i_1 + i_2 + \cdots + i_t$.

For every positive integer l, and every pair of real numbers $0 < d < D$, the class $Rel_l(d, D)$ is defined as follows. The class $Rel_l(d, D)$ is the subclass of $Rel(d, D)$ formed by the structures $A = (X, (R_i : i \in I))$ that satisfy:

(i) for every $i \in I$, $R_i \subseteq X^{[2]}$ for every $i \in I$, in particular every R_i is symmetric and anti-reflexive, as before, and the following additional properties

(ii) $R_i \cap R_j = \emptyset$ whenever $i \neq j$ for $i, j \in I$,

(iii) every edge of A is l-metric.

The objects of $Rel_l(d, D)$ are relational structures of type $\Delta = (\delta_i : i \in I)$, where for each $i \in I$, $\delta_i = 2$. For a pair $\{x, y\} \in R_i$, the index $i \in I$ is a real number which is called the length, or the weight, of the pair, and sometimes this is expressed writing $\rho(x, y) = i$.

Note that $Rel_1(d, D)$ is the sub-collection of $Rel(d, D)$ formed by the structures with pairwise disjoint binary relations.

If an edge (x, y) is l-metric for every l, then we say it is a metric edge. If for a system A every pair (x, y) of vertices is an edge and it is a metric edge then A is just a metric space (A, ρ).

Note that in case every pair of vertices of A is an edge, if every edge is 2-metric then every edge is metric.

The objects of $Rel_l(d, D)$ need not be metric spaces, but since an edge (x, y) cannot be shortened by paths of length $\leq l$, then the larger l is the better an approximation to a metric we have.

For $l = 1$, the notion of l-metric system coincides thus with the notion of relational structure with pairwise disjoint binary relations

The following lemma generalizes Theorem 4

Lemma 8 *(Main Lemma) For every positive integer l, and every pair of real numbers $0 < d < D$, if A is metric in $Rel(d, D)$, then the class $Rel_l(d, D)$ is A- Ramsey, i.e. for every $B \in Rel_l(d, D)$ such that $A \leq B$, there exists $C \in Rel_l(d, D)$ such that $B \leq C$ and in $Rel_l(d, D)$ the following partition relation holds,*

$$C \to (B)^A_2.$$

Before we give the proof we need to consider partite l-metric systems and their amalgamation. That will be done in the next section. Now we show that from lemma 8 we can derive that the class of ordered finite metric spaces is a Ramsey class.

Theorem 9 *The class of finite ordered metric spaces is a Ramsey class.*

Proof. Let (X, ρ) and (Y, σ) be finite ordered metric spaces, and assume that (Y, σ) contains an isometric copy of (X, ρ). Let $d = \min\{\sigma(x, y) : x, y \in Y\}$, and

$D = \max\{\sigma(x,y) : x,y \in Y\}$, and let $l \geq D/d$. Consider the binary relational systems $A = (X, (R_i : i \in I))$ and $B = (Y, (S_j : j \in J))$ corresponding to the metric spaces (X,ρ) and (Y,σ). Clearly, all edges in A, and B are metric.

By lemma 8 there is a binary relational system $C = (Z, (T_k : k \in K))$ such that $C \to (B)^A_2$ in the class $Rel_l(d,D)$.

Define a metric θ on Z by $\theta(x,y) = \min\{D, SP\}$, where $SP(x,y)$ (shortest path from x to y) is the minimum value of $i_1 + \cdots + i_t$ where $x = x_0, x_1, \ldots, x_t = y$ is a path such that for every $r \leq t$, $(x_{r-1}, x_r) \in T_{i_r}$. All the values taken by θ lie in the interval $[d,D]$, and since $ld \geq D$, for every edge (x,y) of C, $(x,y) \in T_i$ if and only if $\theta(x,y) = i$.

(Suppose $(x,y) \in T_i$, and $x = x_0, x_1, \ldots, x_t = y$ is a path from x to y. If $t \leq l$, then, since (x,y) is l-metric, $i \leq i_1 + i_2 + \cdots + i_t$. And if $t > l$, $i_1 + i_2 + \cdots + i_t > ld \geq D$. Therefore i is the length of the shortest path. Conversely, if $\theta(x,y) = i$, since (x,y) is an edge of C, $(x,y) \in T_j$ for some $j \in K$, and $i \leq j$. If $i < j$, it is because there is a path $x = x_0, x_1, \ldots, x_t = y$ from x to y of length i, but, as before, any path $x = x_0, x_1, \ldots, x_t = y$ must have length $\geq j$, and thus $j = i$.)

From this follows that any embedding from A into C (in $Rel_l(d,D)$) is an isometry (an isometric embedding) of (X,ρ) into (Z,θ), and similarly any embedding from B into C is an isometry from (Y,σ) into (Z,θ). From this we conclude that $Z \to (Y)^X_2$. □

3.1 Partite l-metric systems and their amalgamation

We define now the partite approximation classes $PartiRel_l(d,D)$. An object in $PartiRel_l(d,D)$ is a triple (B,A,ι) where A and B are ordered binary relational structures, $A \in Rel_{l-1}(d,D)$ and $B \in Rel_l(d,D)$. More explicitly, $A = (X, (R_i : i \in I))$ and $B = (Y, (S_j : j \in J))$, I, J finite sets of reals contained in the interval $[d,D]$, and $\iota : B \to A$ is a monotone homomorphism satisfying:

(i) If $(x,y) \in S_j$, then $(\iota(x), \iota(y)) \in R_j$ (thus, $J \subseteq I$),

(ii) for every $x \in A$, the set $\iota^{-1}(x)$ is an interval in the ordering of Y.

An embedding from (B,A,ι) into (B',A',ι') is a pair (f,α) such that

(i) $\alpha : A \to A'$ is an embedding in the class $Rel_{l-1}(d,D)$

(ii) $f : B \to B'$ is an embedding in the class $Rel_l(d,D)$

(iii) $\iota' \circ f = \alpha \circ \iota$

If for (B, A, ι), the ι is an injective mapping, we say that (B, A, ι) is a transversal system.

Any $B \in Rel_l(d, D)$ can be viewed as a transversal system(B, B, ι) in $PartiRel_l(d, D)$ where ι is the identity function.

Lemma 10 *(Amalgamation lemma)*

Let $C \in Rel_l(d, D)$, and A a metric subsystem of C (in $Rel_l(d, D)$), with $1 : A \to C$ the inclusion map. For $i = 1, 2$, let (B_i, C, ι_i) be systems in $PartiRel_{l+1}(d, D)$. Let (B_0, A, ι_0) be a system in $PartiRel_l(d, D)$, with embeddings $(f_i, 1) : (B_0, A, \iota_0) \to (B_i, A, \iota_i)$ in $PartiRel_l(d, D)$, for $i = 1, 2$.

Then, there exists $(B_3, C, \iota_3) \in PartiRel_{l+1}(d, D)$, and embeddings $(g_i, 1) : (B_i, c, \iota_i) \to (B_3, C, \iota_3)$ in $PartiRel_{l+1}(d, D)$ such that $g_1 \circ f_1 = g_2 \circ f_2$, and $\iota_3 \circ g_2 = \iota_2$ and $\iota_3 \circ g_1 = \iota_1$. In other words, (B_3, C, ι_3) is an amalgam of the systems (B_i, C, ι_i)

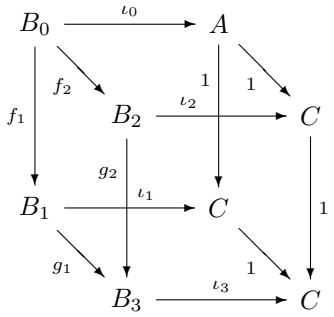


Fig.2 Amalgamation

Proof. We are given the systems (B_1, C, ι_1) and (B_2, C, ι_2), which can be represented as in Fig.3, where the two lines marked C should be identified; the partite subsystem (B_0, A, ι_0) is embedded in both (B_1, C, ι_1) and (B_2, C, ι_2).

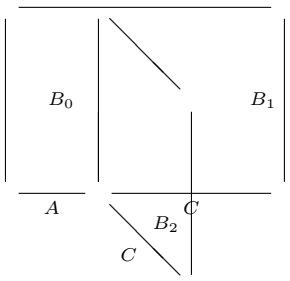


Fig. 3

Let (B_3, C, ι_3) be the free amalgamation of (B_1, C, ι_1) and (B_2, C, ι_2). We have to show that (B_3, C, ι_3) belongs to $PartiRel_{l+1}(d, D)$. Let $\{x, y\}$ be an edge in B_3, and $P = \{x_0 = x, x_1, \ldots, x_t = y\}$ be a path in B_3 from x to y of length $\leq l + 1$. We want to prove that the length $\rho(x, y)$ of the edge $\{x, y\}$ is at most $\rho(P) = \sum_{i=1}^{t} \rho(x_{i-1}, x_i)$.

Consider the projection of P, $i_3(P) = \{i_3(x_0), i_3(x_1), \ldots, i_3(x_t)\}$. For each $j = 1, \ldots, t$, $\rho(x_{j-1}, x_j) = \rho(i_3(x_{j-1}), i_3(x_j))$.

$\iota_3(P)$ is a sequence in C, in which some vertices and edges of P might be identified by ι_3. If this in fact occurs, then the length of $i_3(P)$, $\rho(\iota_3(P)) = \rho(P)$ is bounded by the length of a sub-path P' of $\iota_3(P)$ of length $\leq l$. and thus, since $C \in Rel_l(d, D)$, we have that $\rho(x, y) = \rho(i_3(x), i_3(y)) \leq \rho(P') \leq \rho(P)$.

We may thus assume that $\iota_3(P)$ is a path in C of length $l + 1$.

If $\iota_3(P)$ is a path in A, then $(\iota_3(x), \iota_3(y))$ is a metric edge, since A is metric, and then $\rho(x, y) = \rho(\iota_3(x), \iota_3(y)) \leq \rho(P)$.

If P is a subset of B_1 or B_2, then also $\rho(x, y) = \rho(\iota_3(x), \iota_3(y)) \leq \rho(P)$, since (B_1, C, ι_1) and (B_2, C, ι_2) are in $PartiRel_{l+1}(d, D)$.

So we have to examine the case in which there are $x_{j_1} \in B_1 \setminus A$ and $x_{j_2} \in B_2 \setminus A$. Since B_3 is a free amalgamation, there are no edges with one vertex in $B_1 \setminus A$ and the other in $B_2 \setminus A$, and so there are at least two vertices x_{k_1} and x_{k_2} for which $\iota_3(x_{k_1})$ and $i_3(x_{k_2})$ lie in A and $\{x_{k_1}, x_{k_2}\}$ are not consecutive in the path P.

Any path in A between $\iota_3(x_{k_1})$ and $\iota_3(x_{k_2})$ adds up to at least $\rho(\iota_3(x_{k_1}), \iota_3(x_{k_2}))$, since A is metric. Now, $\rho(P) \geq \rho(P')$ where P' is the path from $\iota_3(x)$ to $\iota_3(y)$ which goes through $\{x_{k_1}, x_{k_2}\}$, i.e.

$$P' = \{\iota_3(x) = \iota_3(x_0), \iota_3(x_1), \ldots, \iota_3(x_{k_1}), \iota_3(x_{k_2}), \ldots, \iota_3(x_t) = \iota_3(y)\},$$

and $\rho(P') \geq \rho(\iota_3(x), \iota_3(y)) = \rho(x, y)$, since P' is of length at most l and C is in $Rel_l(d, D)$. □

3.2 Proof of the main Lemma

Proof of lemma 8: the proof is by induction on l. For $l = 1$ the lemma follows from Theorem 7. Recall that $Rel_1(d, D)$ is the subclass of $Rel(d, D)$ of structures for which the binary relations R_i are pairwise disjoint. Theorem 7 gives us a structure C in $Rel(d, D)$, but from it we can extract one in $Rel_1(d, D)$ by taking the substructure induced by the copies of B in C. More precisely, we take only the vertices which belong to a copy of B in C, and the edges which lie within a copy of B. By the definition of the embeddings, a copy of B cannot have a pair belonging to two different relations (i.e. no pair has more than one label).

Assume the lemma holds for l, and let $B \in Rel_{(l+1)}(d, D)$. Consider A, B as transversal systems in $PartiRel_{(l+1)}(d, D)$, and let $R \in Rel_{(l)}(d, D)$ be a system

satisfying $R \to (B)^A_2$ in $Rel_{(l)}(d, D)$. Fix R, and consider it as a transversal system in $PartiRel_l(d, D)$. We construct now a sequence $P^0, P^1, \ldots, P^a$, of R-partite systems, where $a = |\binom{R}{A}|$. The system P^a will satisfy the required properties.

(P^0, R, ι_0) is the lifting of R obtained by separating all the copies of B contained in R. In other words, P^0 is the disjoint union of $\binom{R}{B} = \{B^1, B^2, \ldots, B^b\}$ with the natural projection to R. Notice that $P^0 \in PartiRel_{l+1}(d.D)$, since $B \in Rel_{l+1}(d, D)$.

Let $\{A^1, \ldots, A^a\}$ list the elements of $\binom{R}{A}$. For the inductive step from i to $i + 1$, let (P^i, R, ι^i) be an R-partite system in $PartiRel_{l+1}(d, D)$, and let (D^i, A, ι^i) be the subsystem of (P^i, R, ι^i) induced by the set $(\iota^i)^{-1}(A^i)$. Clearly $(D^i, A, \iota^i) \in PartiRel_{l+1}(d, D)$, and by the inductive hypothesis, there is a system (E^i, A, λ^i) such that

$$E^i \to (D^i)^A_2.$$

Let $(P^{i+1}, R, \iota^{i+1})$ be a free amalgamation of copies of (P^i, R, ι^i) such that every copy of (D^i, A, ι^i) in (E^i, A, λ^i) is extended to a unique copy of (P^i, R, ι^i). According to lemma 10, we know that $(P^{i+1}, R, \iota^{i+1}) \in PartiRel_{l+1}(d, D)$.

Put $(C, R, \iota) = (P^a, R, \iota^a) \in PartiRel_{l+1}(d, D)$. It remains to show that

$$C \to (B)^A_2.$$

This is done by reverse induction from a to 0 in the same fashion as in the end of the proof of Theorem 4 in 2.2 . □

References

[1] Fraïsse, R., Theory of relations. Springer Verlag, 2000.

[2] Graham, R. L., B. L. Rothschild and J. H. Spencer, Ramsey Theory. Wiley, 1990.

[3] Kechris, A., V. Pestov and S. Todorcevic, Fraïssé limits, Ramsey Theory, and Topological Dynamics of Automophism Groups. *Geometric and Functional Analysis*, 15 (2005) 106-189.

[4] Nešetřil, J., Ramsey Theory. In, Handbook of Combinatorics (R. Graham, M. Grötschel and L. Lovász, eds.) Elsevier Science BV and MIT Press, 1995.

[5] Nešetřil, J., Ramsey classes and homogeneous structures. *Combinatorics Probability and Computing*, 14 (2005) 171-189.

[6] Nešetřil, J., Metric spaces are Ramsey. *European Journal of Combinatorics*, 28 (2007) 457-468.

[7] J. Nešetril and V. Rödl, Mathematics of Ramsey Theory. Springer Verlag, 1990.

[8] Nešetřil, J. and V. Rödl, Partitions of finite relational and set systems. *Journal of Combinatorial Theory Ser. A*, 22 (1978) 289-312.

[9] Ramsey, F. P. , On a problem of formal logic. *Proceedings of the London Mathematical Society*, 30 (1930), 264-286.

Carlos A. Di Prisco
Instituto Venezolano de Investigaciones Científicas
Venezuela
cdiprisc@ivic.ve

From elementary martingale calculus to rigorous properties of mixtures of experts

Badih Ghattas & Gonzalo Perera

Abstract

Authors have performed learning algorithms, based on mixtures of experts, who achieve a good performance under severe time/cost restrictions, and that can be applied to non-stationary data. This is of particular interest for applications like quality of Service (QoS) prediction on IP data networks (see [12]). In this paper we show how can all the properties of this algorithms be proved in a strictly rigorous manner, with no other tools that elementary martingale theory at hand.

Mathematics Subject Classification (2000): 68T05,93E35, 62J20.
Key words: learning algorithm, prediction model, classification, regression, martingales, mixture of experts, limit theorems.

1 Introduction: motivation and basic ideas on Supervised Learning Algorithms.

1.1 On-line prediction for non-stationary engineering data.

In this paper we show how very simple probabilistic tools give rigorous proof for learning algorithms developed to solve real Engineering Applications. It is quite usual in real Engineering problems to use machine learning algorithms, or more precisely, supervised learning algorithms (see details in next subsection) to control a process. Of course, those algorithms must be as efficient and low-cost as possible. But many real engineering sets exhibit a clear non-stationary behavior, which, a priori, are out of the scope of the most usual learning algorithms.

For an example, on-line estimation of Data Networks performances is crucial to guarantee Quality of Service (QoS) for multi-purpose networks, whose traffic is usually non-stationary at all the possible time scales (see [14], [20]).

Algorithms for risk analysis of credit cards operations (see [5]) or surveillance of atmospheric pollution (see [3]) are other examples of Engineering problems where the same type of requirements (efficiency at the same time than low cost

and short computation time for non-stationary data) appears.

Coming back to the example of QoS prediction in Data Networks, large deviation principles (based in the notion of *effective bandwidth* as presented in [15]) have produced a series of results allowing to predict QoS at a given link, at the level of the backbone of the network (see for instance [1], [16]) and some partial results allowing to assure QoS from end-to-end (see for instance [4],[6]).

A relevant current of research in Data Networks has been active measurement of end-to-end QoS via probe packets. The basic idea is that if one sends some packets, the observed delay at their arrival will allow to estimate how heavy is the current traffic in the Network and what level of QoS can be assured. Even if the methodology of probe packets is not universally appliable (see [7]), for some particular Networks and parameters, this idea has shown to be successful. In [2] a functional regression method for non-stationary and dependent data was developed, providing a Learning Machine algorithm that, given the empirical distribution of the delay of the probe packets, predicts the QoS for a video or any other heavy network process that one wants to run. In [12], a much more general learning strategy for non-stationary data has been built up, based on mixtures of experts with different skills.

However, there was not a detailed, rigurous proof of the results provided in [12]. In this paper, we present a careful proof of the properties of this method. We will prove the results in the simplest possible context. In this way, an elementary knowledge of martingale calculus and its limit theorems will be largely enough to understand most of this paper. The extension from this context to the general setting of [12] is an exercise for the reader that knows Machine Learning Theory well, and is a merely technical effort for the general reader. We hope that our choice will make this work readable for a wide mathematical public that, hopefully, may feel interested by Machine Learning or Data Network Performance matters.

To get started in the following section we present the general framekork for *Supervised Learning Algorithms*.

1.2 Basic ideas of Supervised Learning.

Let P denote a probability on an underlying probability space $(\Omega, \mathcal{A})$, where the couple (X, Y) is defined, where X takes values in an arbitrary mesurable space S_X and Y takes values on a measurable space S_Y.

Some notation we will use:

- ρ is the joint distribution of the couple (X, Y), that is, for any measurable sets $A \subset S_X$ and $B \subset S_Y$, we have

$$\rho(A \times B) = P(X \in A, Y \in B).$$

- $p(./X)$ denotes the conditional probability distribution of Y given X, defined in a rigurous way as the almost surely (a.s., for short) unique measurable function of X satisfying:

$$E(1_{\{Y \in B\}} 1_{\{X \in A\}}) = E(p(B/X) 1_{\{X \in A\}})$$

  for any measurable sets A and B. We assume that this conditional distribution is regular, what is true if S_X, S_Y are standard spaces.

- π is the marginal law of X (i.e., $\pi(A) = P(X \in A)$); then we can write

$$\rho(A \times B) = \int_A p(B/x)\pi(dx) = \int_A \int_B p(dy/x)\pi(dx).$$

We assume that both ρ and p are unknown. In this paper we also assume that π is unknown but the key point of the prediction problem concerns the process of *learning* p (and hence, ρ).

In the prediction problem, we observe the value of $X(\omega)$ and we are requested to guess the value of $Y(\omega)$. We will often call X the *input* or the *pattern*, and Y *output* or *label*. In general, our prediction will be $f(X(\omega))$ where f is a measurable function from S_X on S_Y, that we call *predictor*. The main problem is to find a "good"predictor, what previously requires to set a criterion to determine whether a given predictor is "good"or not.

If we have a criterion to quantify how much we "loose"by predicting a value u for $X(\omega) = x$ where the true value was $Y(\omega) = y$ and this quantification is denoted by $L(x, u, y)$, we may introduce the *loss function*

$$L : S_X \times S_Y \times S_Y \Longrightarrow \mathbb{R}.$$

We will assume in the sequel that $L(x, u, y) \geq 0$ and that $L(x, u, y) = 0$ if and only if $u = y$.

From now on (X, Y) denotes a generic random vector distributed according to ρ. As said before, a *predictor* or *prediction rule* is in general a measurable function $f : S_X \Longrightarrow S_Y$ and its quality is measured by means of the expected loss:

$$\tau_L(f) = E\{L(X, f(X), Y)\} = \int_{S_X} \int_{S_Y} L(x, f(x), y)\rho(dx, dy) =$$

$$\int_{S_X} \{ \int_{S_Y} (L(x, f(x), y)) p(dy/x) \} \pi(dx).$$

Hence, we say that a predictor f is better than a predictor g if $\tau_L(f) \leq \tau_L(g)$. For instance, if we take $L(x, u, y) = 1_{\{u \neq y\}}$ then $\tau_L(f) = P(f(X) \neq Y)$ (the *overall error rate*).

As another classical example, if S_Y is a normed space, $\|.\|$ denotes its norm, we assume that $E\{\|Y\|^2\} < \infty$ and set $L(x, u, y) = \|u - y\|^2$, then $\tau_L(f) = E\{\|f(X) - Y\|^2\}$ (the *mean integrated squared error*, MISE, for short).

In general, if we assume that, for any x in a set of π-probability one, there exists an unique value $f^*(x)$ such that

$$\int_{S_Y} L(x, f^*(x), y) p(dy/x) \leq \int_{S_Y} L(x, u, y) p(dy/x) \text{ a.s. with respect to } u \in S_Y,$$

and if the function $f^* : S_X \Longrightarrow S_Y$ is measurable, then f^* is the optimal predictor, since a straightforward computation shows that $\tau_L(f^*) \leq \tau_L(f)$ for any predictor f.

In the case of the overall error rate and when $p(./x)$ is unimodal, $f^*(x)$ is called *conditional mode* or *Maximum A Posteriori* (MAP) given x, and corresponds to the value of u that maximizes $p(u/x)$. In the case of the MISE, $f^*(X) = E(Y/X)$. If, for instance, S_Y is a topological vector space and L satisfies some regularity and convexity conditions with respect to u, the existence of $f^*(x)$ can be shown.

The problem is that, in general, one is not able to look for a predictor on the whole set of functions from S_X to S_Y (that may be a very huge set) but only on a given class of functions $\mathcal{F}$ that corresponds to the kind of predictors that we may practically compute. In such a case, the optimal predictor f^* may be not included in $\mathcal{F}$ and , therefore, the best predictor that we will be able to find is f^{**}, such that

$$f^{**} = \text{argmin}_{f \in \mathcal{F}} \tau_L(f).$$

As we will see later, this predictor f^{**} is not available in practice, since the law ρ is unknown and should be estimated from data. Thus, one is not able to minimize τ_L but only an empirical estimation of it.

Remark 1.1: It is also a common practice (and in fact, this will be our case), to consider a *reward function* $R(x, u, y)$, giving the reward to be assigned to a prediction of u for the value x when the real value is y, instead of the loss function. Despite of the fact that the practical motivation may make one approach more appealing than the other, from the mathematical point of view, they are

completely equivalent, since if L is a loss function and C is a suitable constant, then $C - L$ is a reward function.

Remark 1.2: It is also common (and wise), in practice, to make use of the advice of experts, that may be human experts or previously tested algorithms. In our case we make use of experts advice, and we think an expert as a transition matrix $A(./.)$, that gives, for any input x, the probability $A(y/x)$ that this expert assigns to the output y. If an expert is used just by means of a MAP procedure, then, we consider that his answer for an input value x is the (unique) value $a(x) \in S_Y$ such that $A(a(x)/x) \geq A(y/x)$ for any y (if such an $a(x)$ is not unique, then some ordering or sampling procedure may be used to choose only one).

In *supervised learning*, the predictor $\hat{f}_n$ that we can use in practice, is based on a *training sample* $(X_1, Y_1), ..., (X_n, Y_n)$, often assumed to be independent and identically distributed (*iid*, for the sequel) according to the law ρ. If a class $\mathcal{F}$ of functions is used, then we take as our prediction rule $\hat{f}_n$, the element of $\mathcal{F}$ that minimizes

$$\tau_{L,n}(f) = \frac{1}{n} \sum_{i=1}^{n} L(X_i, f(X_i), Y_i),$$

i.e.

$$\hat{f}_n = \text{argmin}_{f \in \mathcal{F}} \tau_{L,n}(f).$$

As an estimation of the performance of this prediction rule, we might take $\tau_{L,n}(\hat{f}_n)$, but this is usually a biased estimation: it overestimates the performance of the predictor. This is also related to what is usually called "overfitting": if, for instance, $\mathcal{F}$ is as big as the whole family of functions from S_X to S_Y, and $X_1, ..., X_n$ contains n different values, it is clear that $\hat{f}_n(X_i) = Y_i$ for any i and that $\tau_{L,n}(\hat{f}_n) = 0$ (perfect fitting over the trainig sample), but when $\hat{f}_n$ is applied to new data the result may be catastrophic (the prediction follows so closely the particular features of the training sample, that it is statistically very poor). This type of problems is detected if the performance of the predictor rule is measured by means of a new sample, called the *evaluation sample*, which is another *iid* sample of the distribution ρ, $(X_1^v, Y_1^v), ..., (X_m^v, Y_m^v)$, independent with respect to the training sample, and the performance of our predictor is estimated by means of

$$\tau_{L,m}^V(f) = \frac{1}{m} \sum_{i=1}^{m} L(X_i^v, f(X_i^v), Y_i^v).$$

Another type of performance estimation, based on well-known procedures such as cross-validation, bootstrap and other resampling techniques, may be used

in practice to give unbiased and numerically efficient estimations of the performance, but we refer to [13] for an extensive account.

Finally, it should be noticed that when the best of all predictors is f^* and the best of possible predictors on our class is f^{**}, the real predictor we use in practice is $\hat{f}_n$. The loss of perfomance due to the difference between f^* and f^{**} is of modellistic nature, it depends on how clever is our choice of $\mathcal{F}$. If a bad choice of $\mathcal{F}$ is made, no further sampling allows to overcome this loss of performance. This is why the difference $f^* - f^{**}$ is often called *approximation error*. On the other hand, the second loss of performance, due to the difference between f^{**} and $\hat{f}_n$ is purely of statistical nature. If very large training samples were available (i.e., if n tends to infinity), under suitable hypothesis on the model (see for instance [9], [10], [18] for a general exposition), $\hat{f}_n$ tends to f^{**}. This explains why the difference $f^{**} - \hat{f}_n$ is often called *estimation error*.

At last, but not least, one must mention the fact that Machine Learning, and, in particular, Supervised Learning Techniques, are also used as means to gains insights on how does intelligence works. That is why the same subject is also presented under the more appealing title of "Artificial Intelligence". In fact, Steven Smale, when requested to list the 18 most relevant mathematical problems for the XXI century, included as Problem 18 the limits of the intelligence, and if it was possible to model and describe how does intelligence evolves (see [17]).

2 General description of the algorithm.

We will now describe the particular characteristics of our learning algorithm. As we said in the introduction, we will develop the whole method in the simplest case. Therefore, we will assume from now on that X takes values in a finite set $S_X = \{1, ..., I\}$ and that Y takes values in $S_Y = \{1, .., J\}$.

Despite the huge variety of procedures that have been proposed for supervised learning (linear methods, neural networks, CART, SVM, Boosting) most of them do work well under the condition that the size of the training sample (n), is assumed to be arbitrary large. This means that massive information is available, allowing to drastic reduction of the estimation error and, with suitable modeling, very efficient learning (see, for instance, [8],[10], [13]). This is clearly not possible for on-line applications or for applications that must exhibit a minimum delay to give a response. And that is the case of our motivating example of QoS prediction for Data Networks. We overcome this difficulty by means of an iterative procedure, that uses a limited ammount of information at each step, and that makes use of a mixture of experts.

Indeed, as predictors, we will have at hand k experts $A_1, ..., A_k$ and a class of

models $\mathcal{F}$. The experts will be fixed and will not change their behavior trough the whole process: given one expert A_i and an input x at any step of the algorithm, the expert always give the same advice and predict the same value of y. We call *advisors* both experts and the optimal predictor chosen from $\mathcal{F}$. Hence, we have $k+1$ advisors, where indexes $1, ..., k$ correspond to the experts and the index 0 to the model. It must be noticed that while $A_1, ..., A_k$ do not change over the whole execution of our algorithm, the specific function f_j selected in $\mathcal{F}$ at step j, changes from one step of the algorithm to the following. As explained in [12], the fact that the optimal predictor of the model is "fresh"(chosen again) at each step of the algorithm, helps to achieve better performances when dealing with non-stationary data and its a major difference with standard sequential procedures. We assume that a familly of $k+1$ reward functions is given. More precisely, denote $H = \{0, ..., k\}$; we consider a function

$$R : S_X \times S_Y \times S_Y \times H \Longrightarrow \mathbb{R}$$

such that for any $h = 0, ..., k$, $R(., ., ., h)$ is a reward function with $R(x, u, y, h)0$ if $u = y$ and $R(x, u, y, h) < 0$ if $u \neq y$. $R(x, u, y, h)$ represents the reward to be asigned to the advisor h if he assigns for $X = x$ the value u when the true value was $Y = y$.

A central role in our algorithm is played by the *credit matrix*.

$$(c_j(x, h))_{x \in S_X, h \in H}$$

which encodes for the step j our confidence in the advisor h to predict the output of x. More precisely denote:

$$h_j(x) = \text{argmax}_{h \in H} c_j(x, h),$$

the most credible advisor to predict x at step j (if there is more than one value of h where the maximum is reached, we may choose for instance the biggest of such values). Then :

- At step j, the prediction of the value to assign to x is done by the advisor $h_j(x)$ (recalling that the case $h_j(x) = 0$ corresponds to the model, i.e., $\hat{f}_j(x)$.

Once the training sample to be used at the step j is available, $(X_1^j, Y_1^j), \ldots, (X_T^j, Y_T^j)$ (which is assumed to be *iid*, following the law ρ), we choose $\hat{f}_j$ as the best candidate in $\mathcal{F}$ according to the following criteria:

$$\hat{f}_j = \text{argmax}_{f \in \mathcal{F}} \Gamma_T^j(f)$$

where,

$$\Gamma_T^j(f) = \frac{1}{T} \sum_{i=1}^{T} R(X_i^j, f(X_i^j), Y_i^j, 0)$$

is the empirical version of the *expected reward*

$$\Gamma(f) = E\{R(X, f(X), Y, 0)\} = \sum_{x \in S_X, y \in S_Y} R(x, f(x), y, 0)\rho(x, y)$$

(Γ is analogous to the the expected loss τ_L if we think in terms of a loss function L instead of a reward function).

Remark 2.1: Even if our method has been inspired from learning problems on network administration, where a merely objective learning seems to be adequate, it seems to be appealing to apply this system in a subjective context, for instance, for behavioral systems (see [11]). To allow the expression of subjective profiles, we need that different advisors gain different reward by a given decision (further, this difference on the credited rewards may be taken as patterns to identify such profiles). That is why we have included the "h "component on the function R and that is why we prefer to speak about "reward"instead of "loss "or "cost".

Once the model has been fitted, we proceed to the validation of the prediction rule and we update the credit matrix. Since each expert $A_1, ..., A_k$ uses a MAP criterion, we denote by $y_1(x), ..., y_k(x)$ the answer that each expert gives to the input x. In the cycle j of our algorithm, we use a validation sequence $(X_1^{v,j}Y_1^{v,j}), ..., (X_V^{v,j}Y_V^{v,j})$ (*iid* and distributed according to ρ), independent with respect to the training sequence of the same cycle j and independent with respect to both training and validation samples of previous cycles.

Then the credits are updated as following:

- For each $i = 1, ..., V$, compute $h_j(X_i^{v,j})$.
- Compute the prediction for each observation of the validation sequence by means of $y_h(X_i^{v,j})$ if $h_j(X_i^{v,j}) = h \geq 1$ or by means of $\hat{f}_j(X_i^{v,j})$ if $h_j(X_i^{v,j}) = 0$. In any case, let us denote by $f_j(X_i^{v,j})$ the predicted output.
- Update the credits as follows

$$c_{j+1}(x, h) = c_j(x, h) + \frac{1}{V} \sum_{i=1}^{V} R(X_i^{v,j}, f_j(X_i^{v,j}), Y_i^{v,j}, h) 1_{\{h_j(X_i^{v,j})=h, X_i^{v,j}=x\}}$$

Remark 2.2: Observe that to update the credit $c_j(x, h)$ we only use the observations of the validation sample where the input was x and the most credible expert was h. In particular, if a value of x does not appear in the validation sample, its credit is not changed, and if a given expert was less credible than

others for any input of the validation , its credit does not change.

Some final remarks on general notation: we use the symbol ":="for a definition that is set inside an equation. If $\mathcal{C}$ is a collection of random variables, $\sigma(\mathcal{C})$ denotes the σ-algebra generated by $\mathcal{C}$. If $\mathcal{F}, \mathcal{L}$ are σ-algebras on Ω, $\mathcal{F} \vee \mathcal{L} := \sigma(\mathcal{F} \cup \mathcal{L})$. As usual, convergence in law may be thought both at the level of random variables or at the level of probability distributions and notation may mix both levels. For instance, if $Z_1, ..., Z_n, ..$ is a sequence of random variables and we state

$$\lim_n Z_n = N(0,1) \text{ in law },$$

we are saying that, with respect to the topology of the weak convergence of probability measures, the sequence of distribution measures P^{Z_n} converges to a standard gaussian probability measure.

3 Theoretical results.

In this section we derive the asymptotic behavior of our Restricted Resources Learning Algorithm (RRLA, for short), when the number of iterations tends to infinity.

We will first obtain the limit of the credit matrix $(c_j(x,h))_{x \in S_X, h \in H}$ when j tends to infinity. Then we will compare the performance of the RRLA to that of an algorithm based on the whole set of training sequences (i.e., with No Restriction: we will call this algorithm NRLA, for short). In particular we will show that, under reasonable assumptions, RRLA behaves almost as well as NRLA, but with very lower cost and computation requirements, and therefore it can be seen as a performant alternative that respects restrictions.

Let us set some notation and assumptions.
First of all, to avoid trivialities, we assume that $\pi(x)0$ for any $x \in S_X$. For each one of the experts indexed by $h = 1, ..., k$ and any $x \in S_X, y \in S_Y$, we define

$$r_h(x,y) = R(x, y_h(x), y, h)$$

$$r_h(x) = E\{r_h(X,Y)/X = x\} = \sum_{y \in S_Y} r_h(x,y) p(y/x)$$

With the notation of the end of the previous section, denote

$$f^{**} = argmax_{f \in \mathcal{F}} \Gamma(f), \hat{f}_j = argmax_{f \in \mathcal{F}} \Gamma_T^j(f).$$

(We assume again that those maximum values are attained at a unique element of $\mathcal{F}$).
Define, for any $x \in S_X, y \in S_Y$:

$$r_0(x,y) = R(x, f^{**}(x), y, 0)$$

$$r_0(x) = E\{r_0(X,Y)/X = x\}$$
$$r_0^j(x,y) = R(x, \hat{f}_j(x), y, 0)$$
$$r_0^j(x) = \sum_{y \in S_Y} r_0^j(x,y) p(y/x)$$

Observe that $r_0^j(x,y)$ (resp. $r_0^j(x)$) is a random function of (x,y) (resp. x). Let us also call $\mathcal{S}$ the set of all the functions from S_X to S_Y. If (X,Y) is a random vector independent of the training sample and distributed according to ρ, we have that:

$$\begin{aligned} E\left(r_0^j(X)\right) &= E\left(E\{R(X, \hat{f}_j(X), Y, 0)/\hat{f}_j\}\right) \\ &= \sum_{f \in \mathcal{S}} E\{R(X, \hat{f}_j(X), Y, 0)/\hat{f}_j = f\} P(\hat{f}_j = f) \\ &= \sum_{f \in \mathcal{S}} E\{R(X, f(X), Y, 0)\} P(\hat{f}_j = f) \\ &= \sum_{f \in \mathcal{F}} \Gamma(f) P(\hat{f}_j = f) \qquad (1) \end{aligned}$$

In the last equality, we have used the fact that $\hat{f}_j \in \mathcal{F}$. In addition, we clearly have that:

$$E(r_0^j(x)) = \sum_{f \in \mathcal{F}} \sum_{y \in S_Y} R(x, f(x), y, 0) p(y/x) P(\hat{f}_j = f).$$

We assume from now on the following hypothesis:

- (H1) For any $x \in S_X$, there exists an unique $h(x) \in H$, such that

$$r_{h(x)}(x)0, r_h(x) < 0 \text{ if } h \neq h(x).$$

Remark 3.1: Observe that if for a given value x_0, $r_h(x_0)$ depends only on the values of $R(x_0, ., ., h)$. Hence, if $r_h(x_0) < 0$ for all the values of h but there is only one h corresponding to the maximum value $max_{h \in H} r_h(x_0)$, then, we can find a suitable constant C and modify R by means of $R_{mod}(x,u,y,h) = R(x,u,y,h) + C1_{\{x=x_0\}}$, for any x,u,y,h in such a way that for R_{mod} assumption (H1) holds true (and no change is introduced on the rewards for other values of x). Therefore, (H1) essentially means that for any x, there is only one value of h that maximizes $r_h(x)$. In practice, this is not a major restriction, since, again, this can be obtained by means of minor modifications of R and a fixed procedure to choose one h in case of "ties ".

Taking into account (H1), the following sets are well-defined

$$D_h = \{x \in S_X : r_h(x)0\},\ h = 0, ..., k.$$

and we have that

$$\bigcup_{h=0}^{k} D_h = S_X,\ D_h \cap D_l = \emptyset \text{ if } h \neq l.$$

The following lemma plays a key role in the rest of the paper. In two items of this lemma we will consider that T (size of the learning sequence) goes to infinity; since for the rest of the paper T will be fixed, we do not emphasize the dependence on T of $\hat{f}_j$, r_0, r_0^j.

We also denote

$$\gamma(x) := \sum_{f \in \mathcal{F}} \sum_{y \in S_Y} R(x, f(x), y, 0) p(y/x) P(\hat{f}_1 = f)$$

Lemma 3.1 *Let $\mathcal{F}$ be any class of functions from S_X on S_Y and let (X, Y) be a random vector, independent of the training sequence, distributed according to ρ. We have then that:*

- *i) For any j, $\lim_T \hat{f}_j(X) = f^{**}(X)$ in law*
- *ii) There exists a sequence of non-negative real numbers, $(a(T))_{T \in I\!N}$ such that $\lim_T a(T) = 0$ and, for any j,*

$$E\{\Big(r_0^j(X) - r_0(X)\Big)^2\} \leq a(T).$$

- *iii)Fix now T: for any x, $\lim_n \frac{1}{n} \sum_{j=1}^{n} r_0^j(x) = \gamma(x)$ a.s.*
- *iv) With the same notation as above,*

$$E\{(\gamma(X) - r_0(X))^2\} \leq a(T),$$

and for any $x \in S_X$,

$$|\gamma(x) - r_0(x)| \leq \left(\frac{a(T)}{\pi(x)}\right)^{\frac{1}{2}}.$$

Proof:

Fix j. Observe first that

$$\Gamma_T^j(f) - \Gamma(f) = \sum_{x \in S_X, y \in S_Y} R(x, f(x), y, 0)(\rho_T^j(x,y) - \rho(x,y)),$$

where

$$\rho_T^j(x,y) = \frac{1}{T} card\{i : 1 \leq i \leq T : X_i^j = x, Y_i^j = y\}.$$

Since $E\{\rho_T^j(x,y)\} = \rho(x,y)$ for any x, y, we deduce that

$$E\{\Gamma_T^j(f)\} = \Gamma(f). \tag{2}$$

In addition, by the Law of Large Numbers,

$$\lim_T |\rho_T^j(x,y) - \rho(x,y)| = 0 \ a.s., \text{ for any } j, x, y,$$

and, therefore

$$\lim_T max_{(x,y) \in S_X \times S_Y} |\rho_T^j(x,y) - \rho(x,y)| = 0 \ a.s.,$$

what implies in turn that

$$\lim_T max_{f \in \mathcal{F}} |\Gamma_T^j(f) - \Gamma(f)| = 0 \ a.s. \tag{3}$$

Let ω_0 be a point in the probability one set in which (3) holds. Define

$$C = \max_{f \in \mathcal{F}, f \neq f^{**}} \Gamma(f).$$

Let $\delta = \frac{1}{2}(\Gamma(f^{**} - C)$. From (3), there exists a natural number N (depending on ω_0) such that if $T \geq N$, then

$$\sup_{f \in \mathcal{F}} |\Gamma_T^j(f) - \Gamma(f)| < \delta$$

Therefore, if $T \geq N$ it must be $\hat{f}^j = f^{**}$ and (i) is proved.

From the previous argument, we also have that

$$\lim_T P(\hat{f}^j \neq f^{**}) = 0,$$

and, since R is bounded by (say) M,

$$E\{(r_0^j(X) - r_0(X))^2\} \leq (2M)^2 P(\hat{f}^j \neq f^{**})$$

which goes to zero with T and (ii) is proved.

For (iii), fix x. Then, $r_0^j(x)$, as said before, is a given function of the training sequence corresponding to the cycle j. Since training sequences of differents cycles are independent and follow the same distribution on the set of sequences of size T, we have that $r_0^1(x), ..., r_0^n(x), ...$ is an *iid* sequence, with mean

$$E\{r_0^j(x)\} = \sum_{f\in\mathcal{F}} \sum_{y\in S_Y} R(x, f(x), y, 0)p(y/x))P(\hat{f}_j = f)$$

Since the law of $\hat{f}_j$ does not depend on j we conclude that

$$E\{r_0^j(x)\} = \gamma(x)$$

and (iii) follows from the Law of Large Numbers.

Next, using (iii),

$$\begin{aligned} E\{(\gamma(X) - r_0(X))^2\} &= \lim_n E\{\left(\frac{1}{n}\sum_{j=1}^n (r_0^j(X) - r_0(X))\right)^2\} \\ &= \lim_n \|\frac{1}{n}\sum_{j=1}^n (r_0^j(X) - r_0(X))\|_{L^2}^2 \\ &\leq \lim_n \left(\frac{1}{n}\sum_{j=1}^n \|r_0^j(X) - r_0(X))\|_{L^2}\right)^2 \\ &\leq a(T) \text{ (by (ii))} \end{aligned}$$

Finally, pick any $x_0 \in S_X$ and write down

$$\begin{aligned} a(T) &\geq E\{(\gamma(X) - r_0(X))^2\} = \sum_{x\in S_X} (\gamma(x) - r_0(x))^2 \pi(x) \\ &\geq (\gamma(x_0) - r_0(x_0))^2 \pi(x_0) \end{aligned} \tag{4}$$

and we conclude that:

$$|\gamma(x_0) - r_0(x_0)| \leq \left(\frac{a(T)}{\pi(x_0)}\right)^{\frac{1}{2}} .\Diamond$$

We will also use in the sequel the following lemma, that is an easy consequence of the Law of Large Numbers for Martingales (see, for instance, [19]).

Lemma 3.2 *Assume that $(\mathcal{F}_j)_{j\in\mathbb{N}}$ is a filtration (i.e., each $\mathcal{F}_i$ is a sub-σ-algebra of the underlying σ-algebra $\mathcal{A}$ and, for any i, $\mathcal{F}_i \subset \mathcal{F}_{i+1}$) and that $\Delta_0, ..., \Delta_n, ...$ is a sequence of random variables such that:*

- *i)* Δ_i *is* $\mathcal{F}_{i+1}$*- measurable for any* i*.*
- *ii)*$E\{\Delta_i/\mathcal{F}_i\} = 0$ *for any* i*.*
- *iii) There exists* $K < \infty$ *such that* $sup_i|\Delta_i| \leq K$*, a.s.*

Then, if $M_n = \sum_{i=0}^{n-1} \Delta_i$*, we have that*

$$\lim_n \frac{M_n}{n} = 0 \ a.s.$$

Remark 3.3: The following fact also plays a key role in the proof of our main results. For any $x \in S_X$, define

$$\lambda_h(x) = r_h(x)1_{\{h>0\}} + \gamma(x)1_{\{h=0\}}.$$

Set

$$\Lambda_h = \{x \in S_X : \lambda_h(x)0\}.$$

It is clear that $\Lambda_h = D_h$ for $h > 0$. On the other hand, by Lemma 3.1 (iv), for T big enough, $\Lambda_0 = D_0$. More precisely, define

$$\eta = \min\{|r_0(x)| : x \in S_X\},\ T_0 = \inf\{T \in \mathbb{N} : \left(\frac{a(T)}{\pi(x)}\right)^{\frac{1}{2}} < \eta\ \forall x \in S_X\}.$$

Take $T \geq T_0$. If $x \in D_0$, then $r_0(x)\eta$ and by Lemma 3.1 (iv) and the definition of T_0, $\gamma(x)0$ and $x \in \Lambda_0$. If $x \notin D_0$, then $r_0(x) < -\eta$ and the same argument shows that $\gamma(x) < 0$, what implies that $x \notin \Lambda_0$. Therefore, if $T \geq T_0$,

$$\Lambda_h = D_h,\ h = 1, ..., 0.$$

We have then the first result, concerning the asymptotic behaviour of the credit matrix.

Theorem 3.1 *Let* T *be a fixed value,* $T \geq T_0$*, with* T_0 *as in Remark 3.3. As* n*, tends to infinity we have that*

$$\lim_n(\frac{1}{n}c_n(x,h))_{x\in S_X, h\in H} = (\lambda_h(x)\pi(x)1_{\{h=h(x)\}})_{x\in S_X, h\in H},\ a.s.$$

and

$$\lim_n h_n(x) = h(x)\ a.s.$$

(what implies that $h_n(x) = h(x)$ *for all* n *large enough, a.s.)*

Proof: Define

$$U_h^j = \{x \in S_X : h_j(x) = h\},$$

(set of points where the best advisor at cycle j is h)

$$\mathcal{T}_j = \sigma\left(\{(X_i^j, Y_i^j) : 1 \le i \le T\}\right)$$

(σ-algebra generated by the training sequence of cycle j),

$$\mathcal{V}_j = \sigma\left(\{(X_i^{v,j}, Y_i^{v,j}) : 1 \le i \le V\}\right)$$

(σ-algebra generated by the validation sequence of cycle j), and

$$\mathcal{F}_j = \bigvee_{i=1}^{j-1} (\mathcal{T}_i \vee \mathcal{V}_i)$$

(σ-algebra generated by training and validation sequences up to cycle $j-1$).

Set

$$\Delta_j = c_{j+1}(x,h) - c_j(x,h) - E\{c_{j+1}(x,h) - c_j(x,h)/\mathcal{F}_j\},$$

that clearly satisfies all the hypotheses of Lemma 3.2.
We have that:

$$\begin{aligned} c_{j+1}(x,h) - c_j(x,h) &= \frac{1}{V}\sum_{i=1}^{V} R(X_i^{v,j}, f_j(X_i^{v,j}), Y_i^{v,j}, h) 1_{\{X_i^{v,j}=x, h(X_i^{v,j})=h\}} \\ &= \frac{1}{V}\sum_{i=1}^{V} R(x, f_j(x), Y_i^{v,j}, h) 1_{\{X_i^{v,j}=x, x\in U_h^j\}}. \end{aligned}$$

We use in the following lines the fact that the validation sequence of cycle j is independent with respect to the training sample of cycle j and with respect to training and validation samples of previous cycles, and that U_h^j is $\mathcal{F}_j$-measurable. If $h > 0$ and $x \in U_h^j$, $f_j(x) = y_h(x)$ (deterministic) and

$$\begin{aligned} E\{c_{j+1}(x,h) - c_j(x,h)/\mathcal{F}_j\} &= E\{R(x, \hat{f}_j(x), Y_i^{v,j}, h) 1_{\{X_i^{v,j}=x\}} 1_{\{x\in U_h^j\}}\} \\ &= E\{R(x, y_h(x), Y_i^{v,j}, h) 1_{\{X_i^{v,j}=x\}}\} 1_{\{x\in U_h^j\}} \\ &= r_h(x)\pi(x) 1_{\{x\in U_h^j\}}. \end{aligned}$$

Therefore,

$$\Delta_j = c_{j+1}(x,h) - c_j(x,h) - r_h(x)\pi(x) 1_{\{x\in U_h^j\}} \text{ for } h \le k.$$

For $h = 0$, let us compute more carefully:

$$E\{c_{j+1}(x,0) - c_j(x,0)/\mathcal{F}_j\} = E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j\}1_{\{x\in U_0^j\}}.$$

But

$$\begin{aligned} E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j\} &= \\ E\{E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j \vee \mathcal{T}_j\}/\mathcal{F}_j\}. & \end{aligned}$$

Since $\hat{f}_j$ is $\mathcal{F}_j \vee \mathcal{T}_j$-measurable and $(X_i^{v,j}, Y_i^{v,j})$ is independent of $\mathcal{F}_j \vee \mathcal{T}_j$, we have that:

$$E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j \vee \mathcal{T}_j\} = \sum_{y\in S_Y} R(x,\hat{f}_j(x),y,0)p(y/x)\pi(x),$$

what implies in turn that:

$$E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j\} = \sum_{y\in S_Y} E\{R(x,\hat{f}_j(x),y,0)/\mathcal{F}_j\}p(y/x)\pi(x).$$

Observe now that $\hat{f}_j$ only depends on $\mathcal{T}_j$ and is independent of $\mathcal{F}_j$ (by its definition, $\hat{f}_j$ only depends on the performance of the elements of the model class $\mathcal{F}$ over the whole training sequence of cycle j), and thus,

$$\sum_{y\in S_Y} E\{R(x,\hat{f}_j(x),y,0)/\mathcal{F}_j\}p(y/x)\pi(x) =$$

$$\sum_{y\in S_Y}\sum_{f\in\mathcal{F}} R(x,f(x),y,0)P(\hat{f}_j = f)p(y/x)\pi(x).$$

Using now as in Lemma 3.1 the fact that the law of $\hat{f}_j$ does not depend on j, we conclude that

$$\begin{aligned} E\{R(x,\hat{f}_j(x),Y_i^{v,j},0)1_{\{X_i^{v,j}=x\}}/\mathcal{F}_j\} &= \\ \sum_{y\in S_Y}\sum_{f\in\mathcal{F}} R(x,f(x),y,0)P(\hat{f}_1 = f)p(y/x)\pi(x) &= \gamma(x). \end{aligned}$$

Therefore, for $h = 0$,

$$E\{c_{j+1}(x,0) - c_j(x,0)/\mathcal{F}_j\} = \gamma(x)1_{\{x\in U_0^j\}},$$

what shows that for any $h = 1, ..., k, 0$ we have

$$E\{c_{j+1}(x,h) - c_j(x,h)/\mathcal{F}_j\} = \lambda_h(x)1_{\{x\in U_h^j\}},$$

and that

$$\Delta_j = c_{j+1}(x,h) - c_j(x,h) - \lambda_h(x)\pi(x)1_{\{x\in U_h^j\}} \text{ for } h \leq 0.$$

Summing up both terms of this last equation with respect to j and dividing by n, we obtain as a consequence of Lemma 3.2 that:

$$\lim_n \left(\frac{c_n(x,h)}{n} - \lambda_h(x)\pi(x)\nu_n(h) \right) = 0, \ a.s.$$

where

$$\nu_n(h) = \frac{1}{n} card\{j : 0 \leq j \leq n-1, x \in U_h^j\}.$$

From now on, the rest of the proof is devoted to show the following two facts:

- a) $\lim_n \nu_n(h) = 1_{\{h=h(x)\}}$, for any x, h, and
- b) $\lim_n h_n(x) = h(x)$.

(Observe that b) it is not a direct consequence of a), since $\mu_n(h(x))$ gives only the asymptotic frequency of $h_n(x) = h(x)$).

To prove this, fix $x \in S_X$. Let ε be an arbitrary element of $(0,1)$. Set

$$a = inf_{h\in H}|\lambda_h(x)|\pi(x).$$

We already know that for almost any ω in our probability space, there exists $n_\varepsilon(\omega)$ such that

$$\max_{h\in H} |\frac{c_n(x,h)(\omega)}{n} - \lambda_h(x)\pi(x)\nu_n(h)| < \frac{1}{2}a\varepsilon \tag{5}$$

for any $n \geq n_\varepsilon(\omega)$.

Fix ω as before. Let us assume for a moment that:

$$\text{There exists } n_1 \geq n_\varepsilon(\omega) \text{ such that } \nu_{n_1}(h(x))(\omega) \geq \varepsilon. \tag{6}$$

By Remark 3.4, we know that $\Lambda_h = D_h$ for any h. Hence, $\lambda_h(x)0$ if and only if $h = h(x)$. We have then that:

$$\frac{c_{n_1}(x,h(x))(\omega)}{n_1} - \lambda_{h(x)}(x)\pi(x)\nu_{n_1}(h(x))(\omega) - \frac{a\epsilon}{2} \quad \geq$$

$$\lambda_h(x)\pi(x)\nu_{n_1}(h)(\omega) + \frac{a\epsilon}{2} > \frac{c_{n_1}(x,h)}{n_1}(\omega) \qquad \text{for any } h \neq h(x)$$

This implies that $h_{n_1}(x)(\omega) = h(x)$ and hence,

$$\nu_{n_1+1}(h(x))(\omega) = \frac{n_1\nu_{n_1}(h(x))(\omega) + 1}{n_1 + 1}\varepsilon.$$

Therefore, the same argument may be applied to $n_1 + 1$ instead of n_1 and we conclude that

$$h_n(x)(\omega) = h(x) \text{ for any } n \geq n_1$$

what clearly implies

$$\lim_n \nu_n(h)(\omega) = 1_{\{h=h(x)\}}(\omega), \text{ for any } h.$$

It is enough now to show that, on a set of probability one, there exists $\varepsilon \in (0, 1)$ such that (6) holds true.

Let us call A to subset of Ω where (6) does not hold for any $\varepsilon \in (0, 1)$. It is clear that

$$A = \{\omega \in \Omega : \lim_n \nu_n(h(x))(\omega) = 0\}.$$

We will prove that $P(A) = 0$. Observe that the reward function R is bounded (indeed, its domain is a finite set) and hence $\frac{c_n(x,h)}{n}$ is bounded, allowing to interchange limits and expectations in the following lines.
By the definition of $h_n(x)$ we have that, for any ω in Ω and h in H,

$$\frac{c_n(x, h_n(x))}{n}(\omega) \geq \frac{c_n(x, h)}{n}(\omega)$$

what implies that for any h,

$$E\{1_A \frac{c_n(x, h_n(x))}{n}\} \geq E\{1_A \frac{c_n(x, h)}{n}\} \tag{7}$$

Using that

$$\lim_n \max_{h\in H} |\frac{c_n(x, h)}{n} - \lambda_h(x)\pi(x)\nu_n(h)| = 0 \ a.s.$$

and taking limits in (7) we deduce that, for any h,

$$\limsup_n E\{1_A\lambda_{h_n(x)}(x)\pi(x)\nu_n(h_n(x))\} \geq \limsup_n E(1_A\lambda_h(x)\pi(x)\nu_n(h)\}.$$

Since the right-hand side of the last inequality is non-negative for $h = h(x)$, so is the left-hand side, hence

$$\limsup_n E\{1_A\lambda_{h_n(x)}(x)\pi(x)\nu_n(h_n(x))\} \geq 0.$$

But if we take

$$a(x) = \max_{h \neq h(x)} r_h(x),$$

which is negative, it is easy to check that the left-hand side of the last inequality is smaller than

$$\limsup_n E(1_A a(x)\pi(x)1_{\{h_n(x)\neq h(x)\}}) \tag{8}$$

and therefore (8) must be non-negative.

We will show that (8) is negative if $P(A)$ is greater than zero, leading to a contradiction. Taking into account that $a(x)$ is negative, using Fatou's lemma for negative functions and the fact that

$$\limsup_n 1_{\{h_n(x)\neq h(x)\}} = 1 \text{ over } A$$

we conclude that (8) is smaller than

$$E(1_A a(x)\pi(x)),$$

which is negative if $P(A) > 0$.$\Diamond$

Remark 3.4: It must be noticed that in the previous result we have assumed that T is big enough (i.e., $T \geq T_0$), but fixed.

Next result shows that a CLT holds for the convergence of Theorem 3.1. It is based on the following version of the CLT for martingales (see [19]).

Lemma 3.3 *Under the assumptions of Lemma 3.2, if in addition there exists a non-negative constant σ^2 such that*

$$\lim_n \frac{1}{n}\sum_{i=0}^{n-1} E\{\Delta_i^2/\mathcal{F}_i\} = \sigma^2 \text{ in probability,}$$

then

$$\lim_n \frac{1}{\sqrt{n}} M_n = N(0,\sigma^2) \text{ in law .}$$

Remark 3.5: A straightforward argument gives the multivariate version of Lemma 3.3., that may be stated as follows. Assume that $(\Delta_i(1),...,\Delta_i(d))_{i\in\mathbb{N}}$, is a d-dimensional sequence such that $(\Delta_i(s))_{i\in\mathbb{N}}$ satisfies the assumptions of Lemma 3.2 for each $s = 1,...,d$ with respect to the same filtration $(\mathcal{F}_i)_{i\in\mathbb{N}}$ and that there exists a covariance matrix M such that

$$\lim_n \frac{1}{n}\sum_{i=0}^{n-1} E\{\Delta_i(s)\Delta_i(t)/\mathcal{F}_i\} = M(s,t) \text{ in probability, for any } s,t = 1,...,d.$$

Then if $M_n = (M_n(1), ..., M_n(d))$ is defined by $M_n(s) = \frac{1}{n}\sum_{i=0}^{n-1} \Delta_i(s)$, we have that

$$\lim_n \frac{1}{\sqrt{n}} M_n = N(0, M) \text{ in law.}$$

where $N(0, M)$ denotes a d-dimensional centered gaussian random vector with covariance matrix M.

Theorem 3.2 *We have that*

$$\lim_n \sqrt{n} \left(\frac{1}{n} c_n(x,h) - \lambda_h(x)\pi(x) 1_{\{h=h(x)\}} \right)_{x \in S_X, h \in H} = N(0, M) \text{ in law.}$$

where, for any $x, x^* \in S_X,\ h, h^* \in H$,

$$M(x,h;x^*,h^*) = \frac{V-1}{V} \lambda_h(x)\lambda_{h^*}(x^*)\pi(x)\pi(x^*) 1_{\{h(x)=h, h(x^*)=h^*\}},$$

$$-\frac{1}{V}\theta_h(x)\pi(x) 1_{\{x=x^*, h=h^*=h(x)\}},$$

and

$\theta_h(x) := E\{R(x, y_h(x), Y, h)^2\}$ *for* $h = 1, ..., k;\ \theta_0(x) := E\{R(x, \hat{f}^1(x), Y, h)^2\}$.

Proof: Set

$$\Delta_j(x,h) = c_{j+1}(x,h) - c_j(x,h) - \lambda_h(x)\pi(x) 1_{\{x \in U_h^j\}} \text{ for } x \in S_X, h \in H.$$

After Remark 3.5, it is clear that is suffices to prove the following facts:

- a) $\lim_n \frac{1}{n}\sum_{j=1}^n E\{\Delta_j(x,h)\Delta_j(x^*,h^*)/\mathcal{F}_j\} = M(x,h;x^*,h^*)$ in probability, for any $x, x^* \in S_X,\ h, h^* \in H$.
- b)$\lim_n \sqrt{n}\left(\frac{1}{n}\sum_{j=0}^{n-1} 1_{\{x \in U_h^j\}} - 1_{\{h=h(x)\}}\right) = 0$ *a.s.* for any $x \in S_X$, $h \in H$.

By Theorem 3.1., for each $x \in S_X$, and for any ω on a set of total probability, there exists $n(\omega) \in \mathbb{N}$ such that for any $n \geq n(\omega)$ we have $h_n(x)(\omega) = h(x)$. It is then clear that, for $n \geq n(\omega)$, we have that

$$|\frac{1}{n}\sum_{j=0}^{n-1} 1_{\{x \in U_h^j\}}(\omega) - 1_{\{h=h(x)\}}| \leq \frac{n(\omega)}{n},$$

what clearly implies b).

We will then focus on a). It is easy to check that

$$E\{\Delta_j(x,h)\Delta_j(x^*,h^*)/\mathcal{F}_j\} = E\{(c_{j+1}(x,h)-c_j(x,h))(c_{j+1}(x^*,h^*)-c_j(x^*,h^*)\}$$

$$-\lambda_h(x)\lambda_{h^*}(x^*)\pi(x)\pi(x^*)P(h_j(x)=h, h_j(x^*)=h^*).$$

But

$$E\{(c_{j+1}(x,h)-c_j(x,h))(c_{j+1}(x^*,h^*)-c_j(x^*,h^*))\} =$$

$$\frac{1}{V^2}\sum_{s=1}^{V}\sum_{t=1}^{V} E\{R(x,f_j(x),Y_s^{v,j},h)R(x^*,f_j(x^*),Y_t^{v,j},h^*)1_{\{X_s,x,h\}}1_{\{X_s,x^*,h^*\}}\}$$

where $1_{\{X_s,x,h\}} := 1_{\{X_s^{v,j}=x,h_j(x)=h\}}$ and $1_{\{X_s,x^*,h^*\}} := 1_{\{X_t^{v,j}=x^*,h_j(x^*)=h^*\}}$.

If $s \neq t$, then

$$E\{R(x,f_j(x),Y_s^{v,j},h)R(x^*,f_j(x^*),Y_t^{v,j},h^*)1_{\{X_s,x,h\}}1_{\{X_s,x^*,h^*\}}\} =$$

$$\lambda_h(x)\lambda_{h^*}(x^*)\pi(x)\pi(x^*)P(h_j(x)=h, h_j(x^*)=h^*).$$

On the other hand, if $s = t$, then

$$E\{R(x,f_j(x),Y_s^{v,j},h)R(x^*,f_j(x^*),Y_t^{v,j},h^*)1_{\{X_s,x,h\}}1_{\{X_s,x^*,h^*\}}\} =$$

$$\theta_h(x)\pi(x)1_{\{x=x^*,h=h^*\}}P(h_j(x)=h).$$

Therefore, we have that

$$E\{(c_{j+1}(x,h)-c_j(x,h))(c_{j+1}(x^*,h^*)-c_j(x^*,h^*))\} =$$

$$\frac{(V-1)}{V}\lambda_h(x)\lambda_{h^*}(x^*)\pi(x)\pi(x^*)P(h_j(x)=h, h_j(x^*)=h^*)$$

$$-\frac{1}{V}\theta_h(x)\pi(x)1_{\{x=x^*,h=h^*\}}P(h_j(x)=h),$$

and applying Theorem 3.1, a) follows easily.◊

Remark 3.6: Observe that for each pair x, x^*, the limit covariance matrix is null except in the case $h(x)=h, h(x^*)=h^*$. Indeed, instead of the whole credit matrix, we may consider the reduced mean credit vector $(\frac{1}{n}c_n(x,h(x)))_{x\in S_X}$ since no other term is relevant for the asymptotic behaviour of the algorithm.

4 RRLA vs NRLA.

Next, we give the asymptotic behavior of the credit matrix when the NRLA algorithm is used.

Theorem 4.1 *For NRLA algorithm, we have that*

$$\lim_n(\frac{1}{n}c_n(x,h))_{x\in S_X,h\in H} = (\lambda_h(x)\pi(x)1_{\{h=h(x)\}})_{x\in S_X,h\in H},\ a.s.$$

Proof: For the sake of clearity, let us use a different notation for the principal ingredients of the algorithm in the NRLA case. Let us now denote $\hat{h}_j(x)$ the analogous of $h_j(x)$, g_j the analogous of f_j and $\hat{g}^j$ the analogous of $\hat{f}^j$. More precisely, we assume now that, at cycle j, the available training sample is

$$(X_i^s, Y_i^s)_{1\leq i\leq T, 1\leq s\leq j}.$$

Hence, from the model we choose $\hat{g}^j$ such that

$$\hat{g}^j = \text{argmax}_{f\in\mathcal{F}}\,\Gamma_{jT}(f)$$

where

$$\Gamma_{jT}(f) = \frac{1}{jT}\sum_{s=1}^{j}\sum_{i=1}^{T} R(X_i^s, f(X_i^s), Y_i^s, 0).$$

Finally, $\hat{h}_j(x)$ is now the most credible advisor among tke k experts and $\hat{g}^j$, the credit matrix is updated exactly as before (i.e., using only the validation sequence correspondig to each cycle) and g_j denotes the predictor.

If we now set

$$\hat{r}_0^j(x) = \sum_{y\in S_Y} R(x, \hat{g}^j(x), y, 0)p(y/x),$$

it is clear that

$$\left(\hat{r}_0^j(x)\right)_{x\in S_X}$$

has the same law as, in the RRLA,

$$\left(r_0^j(x)\right)_{x\in S_X}$$

when a training sample of size jT is used, and therefore, by Lemma 3.1 ii), it converges in L^2, as j goes to infinity, to $r_0(x)$.

From now on, the proof follows very closely the arguments used in Theorem 3.1 and may be easily reproduced by the reader. $\Diamond$

As a direct consequence of Theorem 3.1 and 3.2, we can finally compare the performance of RRLA and NRLA. We will compare performances by means of the following performance ratio:

$$\tau_j := \frac{E\{R(X, f_j(X), Y, h_j(X))\}}{E\{R(X, g_j(X), Y, \hat{h}_j(X))\}}.$$

We have then

Theorem 4.2 *If in RRLA we use $T \geq T_0$, then*

$$\lim_j \tau_j = \frac{E\{\lambda_{h(X)}\}}{E\{r_{h(X)}\}} = 1 - \frac{E\{(r_0(X) - \gamma(X))1_{\{X \in D_0\}}\}}{E\{r_{h(X)}\}}.$$

Proof: As seen before, we have that

$$E\{R(X, f_j(X), Y, h_j(X))\} =$$

$$\sum_{h=0}^{k} \sum_{x \in S_X} \sum_{y \in S_Y} R(x, y_h(x), y, h) P(h_j(x) = h) p(y/x) \pi(x) +$$

$$\sum_{x \in S_X} \sum_{f \in \mathcal{F}} \sum_{y \in S_Y} R(x, f(x), y, h) P(\hat{f}^j = f) P(h_j(x) = 0) p(y/x) \pi(x) =$$

$$\sum_{h=0}^{k} \sum_{x \in S_X} r_h(x) \pi(x) P(h_j(x) = x) + \sum_{x \in S_X} \gamma(x) \pi(x) P(h_j(x) = 0) =$$

$$\sum_{h=0}^{k} \sum_{x \in S_X} \lambda_h(x) \pi(x) P(h_j(x) = h).$$

By Theorem 3.1,

$$\lim_j P\{h_j(x) = h\} = 1_{\{h = h(x)\}},$$

and therefore,

$$\lim_j E\{R(X, f_j(X), Y, h_j(X))\} = \sum_{h=0}^{k} \sum_{x \in S_X} \lambda_h(x) \pi(x) 1_{\{h = h(x)\}} =$$

$$\sum_{x \in S_X} \lambda_{h(x)}(x) \pi(x) = E\{\lambda_{h(X)}(X)\}.$$

In a similar way, using Theorem 3.2, we deduce that

$$\lim_j E\{R(X, f_j(X), Y, h_j(X))\} = E\{r_{h(X)}(X)\}.$$

Finally, observe that

$$E\{\lambda_{h(X)}(X)\} = E\{\lambda_{h(X)}(X)1_{\{X\notin D_0\}}\} + E\{\lambda_0(X)1_{\{X\in D_0\}}\} = E\{r_{h(X)}(X)1_{\{X\notin D_0\}}\} + E\{\gamma(X)1_{\{X\in D_0\}}\},$$

and the result follows. $\Diamond$

The following corollary illustrates on the applications of Theorem 3.4.

Corollary 4.1

i) Under the assumptions of Theorem 3.4, we have

$$\lim_j \tau_j \geq 1 - \frac{(a(T)\pi(D_0))^{\frac{1}{2}}}{E\{r_{h(X)}\}}.$$

ii) Set $A = \min_{x\in D_0} r_0(x), B = \min_{x\notin D_0} r_{h(x)}(x)$.
Then

$$\lim_j \tau_j \geq 1 - \frac{(a(T)\pi(D_0))^{\frac{1}{2}}}{A\pi(D_0) + B(1-\pi(D_0))}.$$

Proof:

To prove i), observe that $E\{r_{h(X)}(X)\}$ is positive and apply Cauchy-Schwarz inequality and Lemma 3.1 iv) in Theorem 3.4.

To prove ii), observe that

$$E\{r_{h(X)}(X)\} \geq A\pi(D_0) + B\left(1-\pi(D_0)\right).\Diamond$$

Remark 4.7: Corollary 4.1 says that if T is large, both algorithms have very similar performances. On the other hand, if $T \geq T_0$ but T is not very large, the ratio of performances is close to one if $\pi(D_0)$ is small or if A or B are large. In simple words, this means that performance is almost the same if training samples are large, or experts are the most credible advisor for almost all inputs, or experts have a very good mean performance (even if they are not almost allways chosen as the best advisor) or the model has a very good mean performance. In other words if RRLA is not close in performance to NRLA, we have awfully chosen the ingredients of our learning machine! In [12] very impressive numerical examples of performances of both algorithms are given.

Acknowledgements: to an anonymous referee for his very precise and valuable suggestions.

References

[1] Aspirot L., Belzarena P., Bermolen P., Ferragut A., Perera G., Simón M. (2005). Quality of Service Parameters and Link Operating Point Estimation Based on Effective Bandwith. *Performance Evaluation* 59, 103-120.

[2] Aspirot L., Belzarena P., Perera G., Bazzano B. (2005) End-To-End Quality of Service Prediction Based On Functional Regression. *HET-NET 2005*, P46 (also available in http://iie.fing.edu.uy/investigacion/grupos/artes/).

[3] Bel L., Bellanger L., Bonneau V., Ciuperca G., Dacunha-Castelle D., Deniau C., Ghattas B., Misiti Y., Oppenheim G., Poggi J.M., Tomassone R.(1999). Elments de comparaison de prvisions statistiques des pics d'ozone. *Revue de Statistique Applique*, vol. XLVII (3), 7-25.

[4] Belzarena P., Bermolen P., Casas P., Simón M. (2005). Virtual Path Networks Fast Performance Analysis.(to apper in *Performance Evaluation*).

[5] Bolton R.J., Hand D.J.(2002). Statistical Fraude Detection: A Review (with discussants). *Statist. Sci.* Vol 17, No. 3, 235-255.

[6] Buschiazzo D., Ferragut A., Vázquez A., Belzarena P. (2005). Fast Overflow Probability Estimation Tool for MPLS Networks. *LANC 2005, Session 5 (MPLS)* (also available in

http://iie.fing.edu.uy/investigacion/grupos/artes/).

[7] Bacceli F., Bolot J., Machiraju S., Nucci A., Veitch D. (2005). Theory and Practice of Cross-Traffic Estimation. *SIGMETRICS* '05 June 6-10, 2005, Barnff,Alberta, Canada.

[8] Breiman L., Friedman J.H.,Olshen R.A., Stone C.J. (1999). *Classification And Regression Trees* CA.

[9] Cucker F., Smale S. (2002). *On the mathematical foundations of learning.* Bulletin of the American Mathematical Society, Vol. 39, N. 1, p. 1-49.

[10] Devroye L., Gyorfi L., Lugosi G. (1996) *A Probabilistic theory of Pattern Recognition.* Springer.

[11] Flash P.A. (2000). *On the state of the art in Machine Learning: a personal review.* Department of Computer Science, University of Bristol.

[12] Ghattas B., Perera G. (2005). A Resource-Restricted Learning Algorithm for on-line evaluation of non-stationary data. *Submitted.*

[13] Hastie T., Tibshirani R., Friedman J. (2001) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer .

[14] Karagiannis T., Molle M., Faloutsos M. and Broido A. (2004). A Nonstationary Poisson View of Internet Traffic. *IEEE INFOCOM 2004*.

[15] Kelly F. (1996). Notes on Effective Bandwith. *Stochastic Networks, Theory and Applications*, edited by Kelly, Zachiaris and Ziedis, Oxford University Press.

[16] Pechiar J., Perera G., Simón M. (2001) *Effective bandwith estimation and testing for Markov sources.*, Performance Evaluation 945,p. 1-19.

[17] Smale S. (2000) Mathematical problems for the next century. Arnold, V. (ed.) et al., *Mathematics: Frontiers and perspectives.* Providence, RI: American Mathematical Society (AMS). 271-294.

[18] Vapnik V. (1998) *Statistical Learning Theory.* Wiley.

[19] Williams D. (1991) *Probability with martingales.* Cambridge University Press.

[20] Zhang Y., Duffield N., Paxson V. and Shenker S. (2001). On The Constancy of Internet Path Properties. *ACM SIGCOMM Internet Measurement Workshop.*

Institut de Mathématiques de Luminy, CNRS, Marseille, France
Universidad de la República, Montevideo, Uruguay
ghattas@lumimath.univ-mrs.fr, gperera@fing.edu.uy

Composition Operators on the Dirichlet space and related problems

Gerardo A. Chacón, Gerardo R. Chacón & José Giménez

Abstract

In this paper we investigate the following question: when a bounded analytic function φ on the unit disk $\mathbb{D}$, fixing 0, is such that the family $\{\varphi^n : n = 0, 1, 2, \dots\}$ is orthogonal in the Dirichlet space $\mathcal{D}$?. We also consider the problem of characterizing the univalent, full self-maps φ of $\mathbb{D}$ in terms of the norm of the induced composition operator $C_\varphi : \mathcal{D} \to \mathcal{D}$. The first problem is analogous to a celebrated question asked by W. Rudin on the Hardy space setting that was answered recently ([3] and [14]). The second problem resembles a problem investigated by J. Shapiro in [13] about characterization of inner functions θ in the terms of $\|C_\theta\|_{H^2}$.

1 Introduction

Let $\mathbb{D}$ denote the unit disk in the complex plane. By a *self- map* of $\mathbb{D}$ we mean an analytic map $\varphi : \mathbb{D} \to \mathbb{C}$ such that $\varphi(\mathbb{D}) \subset \mathbb{D}$. The *composition operator* induced by φ is the linear transformation C_φ defined as $C_\varphi(f) := f \circ \varphi$ in the space of all analytic functions on $\mathbb{D}$.

The composition operators have been studied in many settings, and in particular in functional Banach spaces (cf. the books [4], [12], the survey of recent developments [7], and the references therein). The goal of this theory is to obtain characterizations of operator-theoretic properties of C_φ by function-theoretic properties of the symbol φ. Conversely, operator-theoretic properties of C_φ could suggest, or help to understand certain phenomena about function-theoretic properties of φ.

Recall that an analytic functional Banach space is a Banach space whose elements are analytic functions defined on a domain of $\mathbb{C}$ (or $\mathbb{C}^n$) such that the evaluation functionals are continuous.

Particular instances of functional Banach spaces are the Hardy space H^2, and the Bergman space A^2 of the unit disk. In these spaces, as a consequence of Littlewood's Subordination Principle [4], every self-map of $\mathbb{D}$ induces a bounded composition operator. Recently, there have been several articles that deal with the study of composition operators on the Dirichlet space: recall that if $dA(z) =$

$\frac{1}{\pi}dx\,dy = \frac{1}{\pi}r\,dr\,d\theta$, $(z = x + iy = re^{i\theta})$ denotes the normalized area Lebesgue measure on $\mathbb{D}$, the *Dirichlet space* $\mathcal{D}$ is the Hilbert space of analytic functions in $\mathbb{D}$ with a square integrable derivative and with norm given by

$$\|f\|_{\mathcal{D}} = \left(|f(0)|^2 + \int_{\mathbb{D}} |f'(z)|^2\, dA(z)\right)^{1/2}.$$

It is well known that $\mathcal{D}$ is a functional Hilbert space, and that for each $w \in \mathbb{D}$ the function

$$K_w(z) = 1 + \log\frac{1}{1-\overline{w}z},$$

is the *reproducing kernel* at w in the Dirichlet space, that is, for $f \in \mathcal{D}$ we have $f(w) = \langle f, K_w\rangle_{\mathcal{D}}$. It is easy to see that $\|K_w\|^2_{\mathcal{D}} = \log\frac{1}{1-|w|^2}$.

A self-map of $\mathbb{D}$ does not induces, necessarily, a bounded composition operator on $\mathcal{D}$. An obvious necessary condition would be that $\varphi \in \mathcal{D}$ which, of course, is not always the case. Actually this condition is not sufficient. A necessary and sufficient condition for φ to induce a bounded composition operator on $\mathcal{D}$ is given in terms of *counting functions* and *Carleson measures* (see [8]).

The counting function $n_\varphi(w)$, $w \in \mathbb{D}$, associated to φ is defined as the cardinality of the set $\{z \in \mathbb{D} : \varphi(z) = w\}$ when the latter is finite and as $+\infty$ otherwise, with the usual rules of arithmetics for $\mathbb{R} \cup \{\pm\infty\}$.

We will make use of a change of variable formula for non-univalent functions: Suppose $\varphi : \mathbb{D} \to \mathbb{D}$ is a non-constant analytic function with counting function $n_\varphi(w)$. If $f : \mathbb{D} \to [0, \infty)$ is any Borel function, then

$$\int_{\mathbb{D}} f(\varphi(z))|\varphi'(z)|^2\, dA(z) = \int_{\mathbb{D}} f(w)n_\varphi(w)\, dA(w).$$

In particular one obtains that $\int_{\mathbb{D}} |\varphi'(z)|^2\, dA(z) = \int_{\mathbb{D}} n_\varphi(w)\, dA(w)$. So, φ is in the Dirichlet space if and only if its counting function is an L^1 function.

In two recent papers, [9] and [10], M. Martín and D. Vukotić, studied composition operators on the Dirichlet space. In this article, based on the results in those works, we consider related questions. In Section 1, we investigate the analogous on the Dirichlet space to a problem proposed by W. Rudin in the context of Hardy spaces: When a bounded analytic function φ on the unit disk $\mathbb{D}$ fixing 0 is such that $\{\varphi^n : n = 0, 1, 2, \dots\}$ is orthogonal in $\mathcal{D}$? In Section 2 we consider the problem of characterizing the univalent, full self-maps φ of $\mathbb{D}$ in terms of the norm of $\|C_\varphi\|_{\mathcal{D}}$. This problem, is analogous to a question asked and answered by J. Shapiro in [13] about inner functions in the Hardy space setting.

We will denote as $\mathcal{D}_0$ the subspace of $\mathcal{D}$ of those function that vanish at 0, and we will use the notation $\|C_\varphi : \mathcal{H} \to \mathcal{H}\|$ to denote the norm of the induced composition operator on a Hilbert space $\mathcal{H}$.

2 Orthogonal functions in the Dirichlet space.

The problem of describing the isometric composition operators acting on Hilbert spaces of analytic functions has been studied in several contexts. Namely, it was proved by Nordgren in [11] that the composition operator C_φ induced on H^2 by an analytic map $\varphi : \mathbb{D} \to \mathbb{D}$ is an isometry on H^2 if and only if $\varphi(0) = 0$ and φ is an inner function (see also [4, p. 321]. In the Bergman space A^2 it is a straightforward consequence of Schwarz Lemma that φ induces an isometric composition operator if and only if φ is a rotation.

Recently, M. Martin and D. Vukotić showed in [10] that in $\mathcal{D}$, the isometric composition operators are those induced by univalent full self-maps of the disk that fix the origin (a self-map of $\mathbb{D}$ is said to be a *full map* if $A[\mathbb{D} \setminus \varphi(\mathbb{D})] = 0$).

W. Rudin in 1988 (MSRI conference) proposed the following problem: If φ is a bounded analytic on the unit disk $\mathbb{D}$ such that $\{\varphi^n : n = 0, 1, 2, \dots\}$ is orthogonal in H^2, does φ must be a constant multiple of an inner function? C. Sundberg [14] and C. Bishop [3] solved independently the problem. In fact, they showed that there exists a function φ which is not an inner function and the family $\{\varphi^n\}$ is orthogonal in H^2.

As asserted by M. Martín y D. Vukotić in [10], their characterization of the isometric composition operators acting on $\mathcal{D}$ can be interpreted as follows: the univalent full maps of the disk that fix the origin are the Dirichlet space counterpart of the inner functions that fix the origin for the composition operators on H^2. Now, we propose the following question: When a function $\varphi \in H^\infty(\mathbb{D}) \cap \mathcal{D}_0$ is such that the family $\{\varphi^n : n = 0, 1, 2, \dots\}$ is orthogonal in $\mathcal{D}$? (since $\mathcal{D} \cap H^\infty(\mathbb{D})$ is an algebra, $\{\varphi^n\}$ is in $\mathcal{D}$ for all n).

We will answer this question when n_φ is essentially bounded, that is, when there is a constant C so that $n_\varphi(w) \leq C$ for all w except those in a set of measure zero. Actually, we shall need the following possibly weaker hypothesis: that the functions $\int_0^{2\pi} e^{ik\theta} n_\varphi(re^{i\theta}) d\theta$ belong to $L^2[0, 1]$ for every positive entire k.

Our result is analogous to a characterization given by P. Bourdon in [2] in the context of H^2: the functions that satisfy the hypotheses of the Rudin's problem are characterized as those maps φ such that their Nevanlinna counting function N_φ is essentially radial. Our assumption that n_φ is essentially bounded is clearly stronger that assuming that $\varphi \in \mathcal{D}$ and it possibly can be weakened. The proof relies on the techniques of the proof given in [2].

Theorem 2.1. *Let φ be a self-map on $\mathbb{D}$ fixing 0 such that n_φ is essentially bounded. The family $\{\varphi^n : n = 0, 1, 2, \dots\}$ is orthogonal in $\mathcal{D}$ if and only if there is a function $g : [0, 1) \to [0, \infty)$ such that for almost every $r \in [0, 1)$, $n_\varphi(re^{i\theta}) = g(r)$ for almost every $\theta \in [0, 2\pi]$ (this is, n_φ is essentially radial).*

Proof. Suppose that n_φ is essentially radial. Let n, m be nonnegative integers

such that $n > m$. Then we have

$$\begin{aligned}\langle \varphi^n, \varphi^m \rangle_{\mathcal{D}} &= nm \int_{\mathbb{D}} \varphi(z)^{n-1} \overline{\varphi(z)^{m-1}} |\varphi'(z)|^2 \, dA(z) \\ &= nm \int_{\mathbb{D}} w^{n-1} \overline{w^{m-1}} n_\varphi(w) \, dA(w) \\ &= nm \int_0^1 r^{n+m-1} \left[\frac{1}{\pi} \int_0^{2\pi} e^{i(n-m)\theta} n_\varphi(re^{i\theta}) \, d\theta \right] dr \\ &= nm \int_0^1 r^{n+m-1} g(r) \left[\frac{1}{\pi} \int_0^{2\pi} e^{i(n-m)\theta} \, d\theta \right] dr \\ &= 0.\end{aligned}$$

Conversely, if the family $\{\varphi^n : n = 0, 1, 2 \dots\}$ is orthogonal in $\mathcal{D}$ and k is any positive integer, then for each integer $n > k$, we have

$$\begin{aligned}0 = \langle \varphi^n, \varphi^{n-k} \rangle_{\mathcal{D}} &= n(n-k) \int_{\mathbb{D}} \varphi(z)^{n-1} \overline{\varphi(z)^{n-k-1}} |\varphi'(z)|^2 \, dA(z) \\ &= n(n-k) \int_{\mathbb{D}} w^{n-1} \overline{w^{n-k-1}} n_\varphi(w) \, dA(w) \\ &= n(n-k) \int_0^1 r^{2n-k-1} \left[\frac{1}{\pi} \int_0^{2\pi} e^{ik\theta} n_\varphi(re^{i\theta}) \, d\theta \right] dr.\end{aligned}$$

Since n_φ is essentially bounded, the functions $f_k(r) := \int_0^{2\pi} e^{ik\theta} n_\varphi(re^{i\theta}) \, d\theta$ are in $L^2[0,1]$, and the preceding chain of equalities shows that they are orthogonal in $L^2[0,1]$ to the maps $\{r \mapsto r^{2n-k-1} : n > k\}$. The linear span of this latter set is dense in $L^2[0,1]$ (cf. [2]), and so $f_k(r) = 0$ for almost every $r \in [0,1]$. Taking complex conjugates, we see that $\int_0^{2\pi} e^{ij\theta} n_\varphi(re^{i\theta}) \, d\theta = 0$ for all $j \neq 0$, and almost every $r \in [0,1]$. Thus that $\theta \mapsto n_\varphi(re^{i\theta})$ is essentially constant for almost every r. □

Proposition 2.2. *Suppose that φ is a self-map with counting function essentially bounded, and essentially radial. Then φ is a constant multiple of a full self-map of $\mathbb{D}$.*

Proof. Suppose that φ is not constant. If the range of φ contains a point in the circle $S_r = \{re^{i\theta} : \theta \in [0, 2\pi]\}$, $\varphi(\mathbb{D})$ contains an arc because this is an open subset of $\mathbb{D}$. In this arc $n_\varphi \geq 1$, and so the range of φ may omit only a θ-zero-measure subset of S_r because n_φ is essentially constant on S_r.

Thus the range of φ contain almost every point in the disk $\{z : |z| < \|\varphi\|_\infty\}$. □

3 Composition Operators vs. Full Maps

In the Hardy space, J. Shapiro [13] characterized, in terms of their norms, those composition operators C_φ whose symbol is an *inner function.* In fact, J. Shapiro showed:

1. If $\varphi(0) = 0$ then φ is inner if and only if $\|C_\varphi : H^2_0 \to H^2_0\| = 1$, where H^2_0 is the subspace of functions in H^2 vanishing at 0, and

2. If $\varphi(0) \neq 0$ then φ is inner if and only if $\|C_\varphi : H^2 \to H^2\| = \sqrt{\frac{1+|\varphi(0)}{1-|\varphi(0)|}}$.

We are going to investigate the analogous questions on the Dirichlet space.

In [9] M. Martín and D. Vukotić calculated the norm of a composition operator $C_\varphi : \mathcal{D} \to \mathcal{D}$ induced by a univalent full map φ of $\mathbb{D}$. They obtain

$$\|C_\varphi : \mathcal{D} \to \mathcal{D}\| = \sqrt{\frac{L + 2 + \sqrt{L(4+L)}}{2}}, \tag{3.1}$$

where $L = \log \frac{1}{1-|\varphi(0)|^2}$, and show that it is an upper bound for the norms of composition operators, acting on the Dirichlet space, induced by univalent symbols.

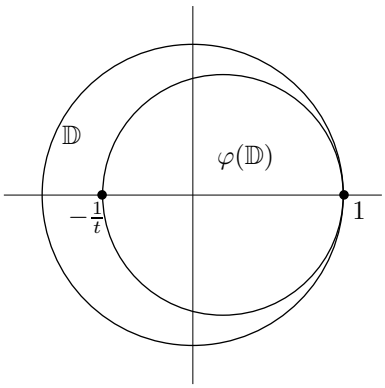
The fact that univalent full maps of the disk that fix 0 are the Dirichlet space counterpart of inner functions that fix the origin (for composition operators on H^2) [10], lead us to investigate if the equality in the equation (3.1) characterizes the univalent full maps of the disk among the univalent self-maps of $\mathbb{D}$.

In addition, the main result in [10] says that $\varphi(0) = 0$ and φ is a univalent full self-map of the disk if and only if C_φ is an isometry on $\mathcal{D}$, and hence on $\mathcal{D}_0$, so in particular its restriction to $\mathcal{D}_0$ has norm 1. Is the converse true?

It is easy to see that this is not true. In fact, let φ_t, $t \geq 1$, be the linear fractional transformation given by

$$\varphi_t(z) = \frac{2z}{(1-t)z + (1-z)}, \quad z \in \mathbb{D}.$$

One easily sees that $\varphi_t(\mathbb{D}) \subset \mathbb{D}$, $\varphi_t(0) = 0$, $\varphi_t(1) = 1$, and $\varphi_t(-1) = -1/t$ (see figure). If $t > 1$ clearly φ_t is not a full map, but a calculation in [1, Cor. 6.1] shows that $\|C_\varphi : \mathcal{D}_0 \to \mathcal{D}_0\| = 1$ when φ is a linear fractional self-map of $\mathbb{D}$ with a boundary fixed point.



Nevertheless, we have the following results, analogous to the results in [13].

Theorem 3.1. *Suppose that φ is a univalent, holomorphic self-map of $\mathbb{D}$, with n_φ essentially radial and $\varphi(0) = 0$. Then φ is a full map if and only if*

$$\|C_\varphi : \mathcal{D}_0 \to \mathcal{D}_0\| = 1.$$

Proof. One direction follows from Proposition 2.2. For the converse, suppose that φ is a univalent holomorphic self-map of $\mathbb{D}$, with n_φ essentially radial, $\varphi(0) = 0$, and that φ is not a full map.

We are going to show that the restriction of C_φ to $\mathcal{D}_0$ has norm < 1. We have that $\varphi(\mathbb{D})$ is contained in the disk $D(0,\rho) = \{z : |z| < \|\varphi\|_\infty = \rho\}$ and $A[D(0,\rho) \setminus \varphi(\mathbb{D})] = 0$ with $0 < \rho < 1$ (cf. proof of Proposition 2.2.)

Let $f \in \mathcal{D}_0$ and define

$$g(r) := \frac{1}{\pi} \int_0^{2\pi} |f'(re^{i\theta})|^2 \, d\theta.$$

Since $|f'|^2$ is subharmonic in $\mathbb{D}$ then g is monotone increasing for $0 \le r < 1$. The change of variable formula gives

$$\begin{aligned} \|C_\varphi f\|_{\mathcal{D}}^2 &= \int_{\mathbb{D}} |f'(\varphi(z))|^2 \, |\varphi'(z)|^2 \, dA(z) \\ &= \int_{\varphi(\mathbb{D})} |f'(w)|^2 \, dA(w) \\ &= \int_0^\rho g(r) \, r \, dr. \end{aligned}$$

and thus:

$$\begin{aligned}
\|f\|_{\mathcal{D}}^2 = \int_{\mathbb{D}} |f'(w)|^2 \, dA(w) &= \int_0^\rho g(r)\, r\, dr + \int_\rho^1 g(r)\, r\, dr \\
&\geq \int_0^\rho g(r)\, r\, dr + \frac{1-\rho^2}{2} g(\rho) \\
&= \int_0^\rho g(r)\, r\, dr + \frac{(1-\rho^2)/2}{\rho^2/2} (\rho^2/2) g(\rho) \\
&\geq \int_0^\rho g(r)\, r\, dr + \frac{(1-\rho^2)/2}{\rho^2/2} \int_0^\rho g(r)\, r\, dr \\
&= \left(1 + \frac{(1-\rho^2)/2}{\rho^2/2}\right) \int_0^\rho g(r)\, r\, dr \\
&= \left(1 + \frac{(1-\rho^2)/2}{\rho^2/2}\right) \|C_\varphi f\|_{\mathcal{D}}^2,
\end{aligned}$$

for each $f \in \mathcal{D}_0$. It yields the desired result: the restriction of C_φ to $\mathcal{D}_0$ has norm $\leq \nu = \left(1 + \frac{(1-\rho^2)/2}{\rho^2/2}\right)^{-1/2} < 1$. □

In the next theorem, we consider the case $\varphi(0) \neq 0$. The proof follows nearly the one in [13, Th. 5.2]).

Theorem 3.2. *Suppose that φ is a univalent, holomorphic self-map of $\mathbb{D}$ with n_φ essentially radial and $\varphi(0) \neq 0$. Then φ is a full map if and only if*

$$\|C_\varphi : \mathcal{D} \to \mathcal{D}\| = \sqrt{\frac{L + 2 + \sqrt{L(4+L)}}{2}},$$

where $L = -\log(1 - |\varphi(0)|^2)$.

Proof. The necessity follows as in [9, Th. 1]. For the converse, suppose that φ is a univalent, holomorphic self-map of $\mathbb{D}$ with n_φ essentially radial, such that $\varphi(0) = p \neq 0$, and that φ is not a full map. We want to show that the norm of C_φ is strictly less that $\sqrt{\frac{L+2+\sqrt{L(4+L)}}{2}}$, where $L = -\log(1 - |p|^2)$.

For this we consider α_p, the standard automorphism of $\mathbb{D}$ that interchanges p with the origin; i.e.,

$$\alpha_p := \frac{p - z}{1 - \overline{p} z}, \quad z \in \mathbb{D}.$$

Put $\varphi_p := \alpha_p \circ \varphi$. Then $\varphi_p(0) = 0$. Since this function is a univalent, self map of $\mathbb{D}$ with counting function essentially radial, but not a full map, Theorem 3.1 affirms that the restriction of the operator C_{φ_p} to $\mathcal{D}_0$ has norm $\nu < 1$.

Because α_p is self-inverse, $\varphi = \alpha_p \circ \varphi_p$, and so, for each $f \in \mathcal{D}$:

$$C_\varphi f = C_{\varphi_p}(f \circ \alpha_p) = C_{\varphi_p} g + f(p),$$

where $g = f \circ \alpha_p - f(p)$.

The function $C_{\varphi_p} g$ belong to $\mathcal{D}_0$ and thus:

$$\begin{aligned} \|C_\varphi f\|_{\mathcal{D}} &= \|C_{\varphi_p} g\|_{\mathcal{D}}^2 + |f(p)|^2 \\ &\leq \nu^2 \|g\|_{\mathcal{D}}^2 + |f(p)|^2 \\ &= \nu^2 \|(C_{\alpha_p} f) - f(p)\|_{\mathcal{D}}^2 + |f(p)|^2. \end{aligned} \tag{3.2}$$

Since $\langle h, 1\rangle_{\mathcal{D}} = h(0)$ for each $h \in \mathcal{D}$,

$$\langle C_{\alpha_p} f, f(p)\rangle_{\mathcal{D}} = \overline{f(p)} C_{\alpha_p} f(0) = |f(p)|^2,$$

and we obtain

$$\begin{aligned} \|(C_{\alpha_p} f) - f(p)\|_{\mathcal{D}}^2 &= \|C_{\alpha_p} f\|_{\mathcal{D}}^2 - 2\Re\langle C_{\alpha_p} f, f(p)\rangle_{\mathcal{D}} + |f(p)|^2 \\ &= \|C_{\alpha_p} f\|_{\mathcal{D}}^2 - 2|f(p)|^2 + |f(p)|^2 \\ &= \|C_{\alpha_p} f\|_{\mathcal{D}}^2 - |f(p)|^2. \end{aligned}$$

This identity and Equation (3.2) yield,

$$\|C_\alpha f\|_{\mathcal{D}}^2 \leq \nu^2 \|C_{\alpha_p} f\|_{\mathcal{D}}^2 + (1 - \nu^2)|f(p)|^2.$$

We know from [10, Th. 1] that $\|C_{\alpha_p} : \mathcal{D} \to \mathcal{D}\| = (L + 2 + \sqrt{L(4+L)})/2$, and we have the following estimate for $|f(p)|$:

$$|f(p)| \leq \|f\|_{\mathcal{D}} \|K_p\|_{\mathcal{D}} = \sqrt{1+L}\|f\|_{\mathcal{D}}.$$

Then

$$\|C_\alpha f\|_{\mathcal{D}}^2 \leq \left[\nu^2 \left(\frac{L + 2 + \sqrt{L(4+L)}}{2}\right) + (1-\nu^2)(1+L)\right] \|f\|_{\mathcal{D}}^2,$$

and $\delta = \left[\nu^2 \left(\frac{L+2+\sqrt{L(4+L)}}{2}\right) + (1-\nu^2)(1+L)\right] < 1$ because $p \neq 0$ and $L > 0$. □

3.1 The essential norm

Recall that the essential norm of an operator T in a Hilbert space $\mathcal{H}$ is defined as $\|T\|_e := \inf\{\|T - K\| : K \text{ is compact }\}$; that is, the essential norm of T is the norm of the equivalence class of T in the Calkin algebra. It is well known

[4] that in any Hilbert space of analytic functions containing every power of $f(z) \equiv z$, we have

$$\|C_\varphi\|_e = \lim_n \|C_\varphi R_n\|, \tag{3.3}$$

where R_n denotes the orthogonal projection of $\mathcal{H}$ onto $z^n\mathcal{H}$.

In [13], it is proved that a self-map $\varphi : \mathbb{D} \to \mathbb{D}$ is inner if and only if the essential norm of C_φ in the Hardy space is equal to $\sqrt{\frac{1+|\varphi(0)|}{1-|\varphi(0)|}}$. Because of the analogies discussed here between inner functions and univalent full-maps, one might ask: Are full-maps characterized by the fact that the essential norm of C_φ in the Dirichlet space is equal to $\sqrt{\frac{L+2+\sqrt{L(4+L)}}{2}}$? where $L = \log \frac{1}{1-|\varphi(0)|^2}$. The answer is not, in fact *every univalent full-map has essential norm equal to 1 in the Dirichlet space*:

Theorem 3.3. *Let $\varphi : \mathbb{D} \to \mathbb{D}$ a univalent full-map, then $\|C_\varphi\|_e = 1$ in the Dirichlet space.*

Proof. Suppose first that $\varphi(0) = 0$, then ([10]) C_φ is an isometry and equation 3.3 gives:

$$\|C_\varphi\|_e = \lim_n \{ \sup_{\|f\|=1} \|C_\varphi R_n f\| \} = \lim_n \|R_n\| = 1.$$

If $\varphi(0) = p \neq 0$, then the function $\varphi_p := \alpha_p \circ \varphi$ is a univalent full-map fixing the origin and then for every function $f \in \mathcal{D}$ with $\|f\| = 1$ we have that $\|C_\varphi R_n f\| = \|C_{\alpha_p} R_n f\|$. Thus, $\|C_\varphi\|_e = \|C_{\alpha_p}\|_e$.

But in [5, Cor. 5.9], it is proved that the essential norm of any composition operator induced by an automorphism of $\mathbb{D}$ is equal to 1 and the result follows. □

Acknowledgment

The authors would like to thank D. Vukotić for suggesting the study of composition operators on the Dirichlet space and for making available his works.

References

[1] E. Gallardo-Gutiérrez, and A. Montes-Rodríguez: *Adjoints of linear fractional composition operators on the Dirichlet space.* Math. Ann. 327, (2003) 117-234.

[2] P. Bourdon: *Rudin's orthogonality problem and the Nevanlinna counting function.* Proc. Amer. Math. Soc. 125 (1997), 1187-1192.

[3] C. Bishop: *Orthogonal functions in H^∞*. Preprint.

[4] C. Cowen and B. MacCluer: *Composition Operators on Spaces of Analytic Functions.* CRC Press, 1995.

[5] G.A. Chacón and G.R. Chacón: *Some Properties of Composition Operators on the Dirichlet Space.* Acta Math. Univ. Comenianae. 74, (2005) 259-272.

[6] C. Hammond: *The norm of a Composition Operator with Linear Fractional Symbol Acting on the Dirichlet Space* J. Math. Anal. Appl. 303 (2005), 499-508.

[7] F. Jafari et al., editors: *Studies on Composition Operators.* Comtemp. Math. Vol. 210 American Math. Soc., 1998.

[8] M. Jovović and B. MacCluer: *Composition operators on Dirichlet spaces.* Acta Sci. Math. (Szeged) 63 (1997), 229-247.

[9] M. Martín and D. Vukotić: *Norms and spectral radii of composition operators acting on the Dirichlet space* J. Math. Anal. Appl. 304 (2005), 22-32.

[10] M. Martín and D. Vukotić: *Isometries of the Dirichlet space among the Composition Operators.* Proc. Amer. Math. Soc. 134 (2006), 1701-1705.

[11] E. Nordgren: *Composition operators.* Canad. J. Math. 20 (1968) 442-449.

[12] J. Shapiro: *Composition Operators and Classical Function Theory.* Springer Verlag, 1993.

[13] J. Shapiro: *What do composition operators know about inner functions?* Monatshefte für Mathematik 130 (2000), 57–70.

[14] C. Sundberg: *Measures induced by analytic functions and a problem of Walter Rudin.* J. Amer. Math. Soc. 16 (2003) 69-90.

Gerardo A. Chacón
Nucleo Universitario del Táchira ULA.
gchacon@cantv.net

Gerardo R. Chacón
Dpto. De Medición y Evaluación ULA.
grchacon@ula.ve

José Giménez
Departamento de Matemáticas ULA.
jgimenez@ula.ve

El truco de m pilas de Gergonne y el sistema de numeración de base m

Roy Quintero*

Resumen

En este artículo, consideramos el truco de m pilas de Gergonne y su relación con el sistema de numeración de base m. El caso $m = 3$ produce uno de los más viejos trucos "mágicos" matemáticos que involucra el reordenamiento de 27 cartas. Joseph Diaz Gergonne [3], un matemático francés, fue el primero en analizarlo y generalizarlo en 1813. En [2, pág. 39], Gardner dice: *Mel Stover, of Winnipeg, Canada, calls my attention to the application of the ternary counting system to the Gergonne pile trick*. Inmediatamente, en [2, pág. 40], él también expresa: *Reflecting on the above matters led Mr. Stover to the invention of a truly stupendous breath-taking version of the trick. It makes use of the decimal system and a deck of 10 billion playing cards!*

Basados en estos casos ($m = 3$ y $m = 10$), demostramos matemáticamente la existencia de una relación formal entre la posición de la carta escogida después de aplicar el truco de Gergonne con una baraja de m^m cartas y el sistema de numeración de base m usando aritmética modular. También, damos pruebas matemáticas generales de algunas situaciones particulares como son: nombrar la posición de la carta, llevar la carta a una posición indicada y nombrar la carta.

Palabras y frases clave: A08 matemáticas recreativas, 11A07 congruencias; raíces primitivas; sistemas de residuos, 11B50 sucesiones (mod m).

The Gergonne m-pile trick and the base m counting system

Abstract

In this paper, we consider the Gergonne m-pile trick and its relation with the base m counting system. The case $m = 3$ produces one the oldest of mathematical "magic" tricks that involve the reordering of 27 cards. Joseph Diaz Gergonne [3], a French mathematician, was the first to analyze and generalize it in 1813. In [2, pag. 39], Gardner says: *Mel Stover, of Winnipeg, Canada, calls my attention to the application*

*El contenido de este trabajo es una versión ampliada de la ponencia, con igual título, presentada por el autor en el International Congress of Mathematicians celebrado en Madrid entre el 22 y el 30 de agosto de 2006.

of the ternary counting system to the Gergonne pile trick. Immediately, in [2, pag. 40], he also expresses: *Reflecting on the above matters led Mr. Stover to the invention of a truly stupendous breath-taking version of the trick. It makes use of the decimal system and a deck of 10 billion playing cards!*

Based on these cases ($m = 3$ and $m = 10$), we demonstrate mathematically the existence of a formal relation between the position of the selected card after applying the Gergonne trick with a deck of m^m cards and the base m counting system by using modular arithmetic. Also, we give general mathematical proofs of some particular situations as are: naming the position of the card, bringing the card to a named position and naming the card.

Key words and phrases: A08 recreational mathematics, 11A07 congruences; primitive roots; residue systems, 11B50 sequences (mod m).

1. Introducción

El truco de Gergonne, tal como lo conocemos hoy día, fue propuesto inicialmente por Joseph Diaz Gergonne en *Les Annales de Mathématiques Pures et Appliquées*, comúnmente llamados los Annales de Mathématiques de Gergonne, una especie de períodico matemático editado por Gergonne desde 1810 hasta 1832. En el cuarto tomo [3][1], Gergonne presenta la teoría general para un paquete de m^m cartas. En [5, pág. 328], Rouse Ball dice que el truco de tres pilas ($m = 3$) es mencionado por Bachet [1, prob. XVIII], pero que su análisis es insuficiente.

Existen algunas generalizaciones a su vez del truco de m pilas de Gergonne, una excelente referencia es el artículo de Harrison, Brennan y Gapinski [4].

En la Sección 2, una fórmula general (Teorema 2.1) será desarrollada para el caso general de m pilas con m^{m-1} cartas cada una, la cual permite expresar la posición de cualquier carta escogida después de la m-ésima recolección en términos del sistema de numeración de base m. Lográndose así formalizar matemáticamente y de manera general lo comentado por Gardner en [2], como fue mencionado en el resumen. Para éllo, empleamos aritmética modular con módulo m y tomamos, por razones de conveniencia técnica, como sistema

[1]Mi agradecimiento al Dr. Christian Gerini de la *Université du Sud Toulon Var* de Francia, por suministrarme una copia electrónica del artículo original de Gergonne.

de residuos el conjunto $I_m = \{1, 2, \ldots, m\}$. En el artículo original de Gergonne [3], el autor utiliza teoría de combinaciones en lugar de aritmética modular y la posición final de la carta escogida se expresa por medio de dos fórmulas, una para el caso par y otra para el impar.

En la Sección 3, damos pruebas matemáticas de algunas situaciones particulares, pero para cualquier m. Las situaciones consideradas son aquellas cubiertas heurísticamente en el capítulo 3 del excelente libro sobre el tema de Gardner [2] y que corresponden al caso $m = 3$. Específicamente, nombrar la posición de la carta, llevar la carta a una posición indicada y nombrar la carta.

2. El truco de m pilas de Gergonne y el sistema de numeración de base m

Sea $m \geq 3$ un entero fijo. Supongamos que tenemos una baraja de m^m cartas boca abajo y que n es la n-ésima carta contando desde el tope del paquete. El ejecutante realiza un paso del truco de Gergonne asumiendo que la carta escogida por el espectador es n; es decir, reparte las cartas boca arriba en m pilas o columnas con m^{m-1} cartas cada una, tal como se muestra en la Figura 1. Mientras tanto, el espectador observa con cuidado y al finalizar la repartición, indica la pila en la que ha caído la carta escogida. Inmediatamente, el ejecutante recolecta todas las pilas -con las cartas aún boca arriba- y coloca la pila indicada en la posición k-ésima contando desde 1 (fondo) hasta m (tope), como lo indica la Figura 2. Entonces, volteamos la baraja y la carta escogida pasa a ocupar la posición $P_k(n)$ dada por la fórmula siguiente:

$$P_k(n) = \frac{n-j}{m} + 1 + (k-1)m^{m-1},$$

si $n \equiv j \pmod{m}$ y j pertenece al sistema de residuos I_m. Esto se sigue claramente observando la Figura 1, ya que si $n = im + j$ $(0 \leq i \leq m^{m-1} - 1)$ entonces, $i + 1$ es la posición de la carta n dentro de la Pila j (contando desde el fondo) después de repartir las

Pila 1	Pila 2	...	Pila j	...	Pila m
$(m^{m-1}-1)m+1$	$(m^{m-1}-1)m+2$	...	$(m^{m-1}-1)m+j$	...	$m^{m-1}m$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$im+1$	$im+2$	...	$im+j$	...	$(i+1)m$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2m+1$	$2m+2$	...	$2m+j$	...	$3m$
$m+1$	$m+2$	...	$m+j$	...	$2m$
1	2	...	j	...	m

Figura 1: Repartición de las cartas

cartas. Realizando algunos cálculos simples obtenemos que,

$$\frac{n-j}{m}+1 = i+1 \in \{1,2,\ldots,m^{m-1}\}.$$

Por otra parte, observe que luego del ensamblaje de la pilas hay exactamente $k-1$ pilas debajo de la pila que contiene la carta escogida. Pero, cada pila tiene m^{m-1} cartas, así que hay exactamente $i+(k-1)m^{m-1}$ cartas debajo de la carta seleccionada. Por tanto, una vez que el paquete completo es volteado boca abajo para la próxima repartición, la carta escogida tendrá sobre sí misma $i+(k-1)m^{m-1}$ cartas y su nueva posición será $i+1+(k-1)m^{m-1} = P_k(n)$.

A continuación, presentamos una definición precisa del truco de Gergonne. Seguidamente, damos otra definición y una proposición referentes a una clase finita de conjuntos finitos, así como un lema sobre sucesiones módulo m, los cuales serán de suma utilidad para probar el principal resultado de esta sección (Teorema 2.1).

Definición 2.1 Sea $m \geq 3$ un entero fijo. Sean $n, k_1, \ldots, k_m$ enteros que satisfacen las condiciones:

1. $1 \leq n \leq m^m$.

2. $1 \leq k_i \leq m$.

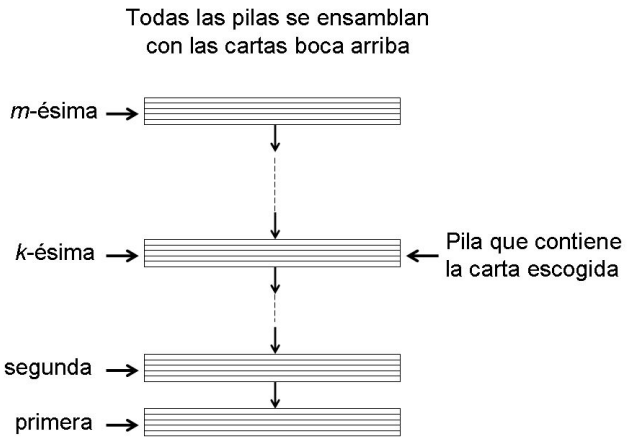


Figura 2: Recolección de las pilas

Diremos que el truco de m pilas de Gergonne se ha ejecutado sobre la carta n según el esquema $\{k_i\}_{i=1}^{m}$, si la función $G_{\{k_i\}_{i=1}^{m}} = P_{k_m} \circ \cdots \circ P_{k_1}$ es evaluada en n.

Observemos que ejecutar el truco es equivalente a hallar la posición final de la carta escogida. También es importante mencionar que existen m^m diferentes esquemas posibles para cada n y en total existen m^{2m} posibles variantes del truco de m pilas de Gergonne, sin considerar las diversas formas cómo pueden ser colocadas las restantes $m-1$ pilas que no contienen la carta escogida en cada recolección.

Definición 2.2 Sean $m \geq 3$ un entero fijo y $\{J_i\}_{i=0}^{m^{m-2}-1}$ la clase finita de conjuntos finitos de números naturales definida por:

$$J_i = \{lm^{m-1} + im + j : 0 \leq l \leq m-1 \text{ y } 1 \leq j \leq m\}.$$

Proposición 2.1 La clase finita $\{J_i\}_{i=0}^{m^{m-2}-1}$ satisface las siguientes propiedades:

(1) $J_{i_1} \cap J_{i_2} = \emptyset$, si $i_1 \neq i_2$.

(2) $\bigcup_{i=0}^{m^{m-2}-1} J_i = \{1, 2, \ldots, m^m\}$.

Demostración: (1) Supongamos que $J_{i_1} \cap J_{i_2} \neq \emptyset$. Sea $x \in J_{i_1} \cap J_{i_2}$. Entonces existen enteros $0 \leq l_1, l_2 \leq m-1$ y $1 \leq j_1, j_2 \leq m$ tales que

$$x = l_1 m^{m-1} + i_1 m + j_1 = l_2 m^{m-1} + i_2 m + j_2.$$

Sin pérdida de generalidad podemos asumir que $j_2 \geq j_1$. Primero consideremos el caso $j_2 > j_1$. Entonces, $j_2 - j_1 = ((l_1 - l_2)m^{m-2} + (i_1 - i_2))m$, luego necesariamente $i_2 - i_1 = m^{m-2}(l_1 - l_2)$ y como $i_1 \neq i_2$, tenemos que $l_1 = l_2$. Esto a su vez implica que $j_2 - j_1 = (i_1 - i_2)m$, lo cual es imposible porque $i_1 \neq i_2$ y $0 < j_2 - j_1 < m$. Así que el caso $j_2 > j_1$ es imposible, y en consecuencia $j_1 = j_2$. Pero en este caso tenemos nuevamente la ecuación $i_2 - i_1 = m^{m-2}(l_1 - l_2)$, lo cual es imposible porque $0 < |i_2 - i_1| < m^{m-2}$. Por tanto, nuestra suposición es falsa y los conjuntos J_{i_1} y J_{i_2} son disjuntos.

(2) Sea n, $1 \leq n \leq m^m$. Sea $j \in I_m$ (único) tal que $n \equiv j \pmod{m}$. Sea k el único entero $0 \leq k \leq m^{m-1} - 1$ tal que $n = km + j$. Por el algoritmo de la división, existen enteros l, i, $0 \leq l \leq m-1$ y $0 \leq i \leq m^{m-2} - 1$ tales que $k = lm^{m-2} + i$. Por tanto, $n \in J_i$, y tenemos que se cumple la inclusión $\bigcup_{i=0}^{m^{m-2}-1} J_i \supset \{1, 2, \ldots, m^m\}$. La otra inclusión es evidente porque cada J_i $(0 \leq i \leq m^{m-2} - 1)$ está contenido en $\{1, 2, \ldots, m^m\}$. ■

Lema 2.1 Sean $m \geq 3$ un entero fijo e i un entero $0 \leq i \leq m^{m-2} - 1$. Entonces existen sucesiones de enteros $\{i_s\}_{s=1}^{m-2}$ y $\{j_s\}_{s=2}^{m-2}$ que satisfacen las siguientes propiedades:

(1) $0 \leq i_s \leq m^{m-(s+1)} - 1$.

(2) $j_s \in I_m$.

(3) $i_s + 1 = i_{s+1} m + j_{s+1}$ $(1 \leq s \leq m-3)$.

Demostración: Sea $i_1 = i$. Claramente, i_1 satisface (1). Sea $j_2 \in I_m$ (j_2 satisface (2)) tal que $i_1 + 1 \equiv j_2 \pmod{m}$. Entonces, existe un entero i_2 tal que $i_1 + 1 = i_2 m + j_2$ (se cumple (3), para $s = 1$). Observemos que,

$$-1 + \frac{1}{m} \leq i_2 \leq m^{m-3} - \frac{1}{m}.$$

Así que $0 \leq i_2 \leq m^{m-3} - 1$. Repitiendo, el procedimiento $m-3$ veces, producimos las sucesiones requeridas. ■

Ahora presentamos el siguiente teorema, el cual expresa que al ejecutar el truco de Gergonne sobre una carta de una baraja de m^m cartas, la posición final de la carta escogida no depende de la carta en sí, sino de la forma como han sido ensambladas las pilas que la contenían en cada paso; es decir, que depende del esquema considerado. Ciertamente, esto es conocido (ver [3]), pero la fórmula (1), expresa adicionalmente que su valor posicional siempre se puede expresar en términos del sistema de numeración de base m, unificando a su vez las fórmulas dadas en [3] (casos par e impar). Este resultado, demuestra que los comentarios hechos por Gardner en [2, págs. 39 y 40] sobre los casos ternario ($m = 3$) y decimal ($m = 10$) son ciertos y además los generaliza.

Teorema 2.1 Sea $\{k_i\}_{i=1}^m$ un esquema cualquiera. Entonces,

$$G_{\{k_i\}_{i=1}^m}(n) = [(k_m - 1) \ldots (k_1 - 1)]_{base\ m} + 1, \tag{1}$$

para toda n $(1 \leq n \leq m^m)$.

Demostración: Dado n, existe (Proposición 2.1) un único entero i tal que $n \in J_i$. Sean l, j enteros tales $n = lm^{m-1} + im + j$ y sean $\{i_s\}_{s=1}^{m-2}$ y $\{j_s\}_{s=2}^{m-2}$ como en el Lema 2.1. Denotemos $c_s = (k_s - 1)m^s + \cdots + (k_1 - 1)m + l$ $(1 \leq s \leq m-2)$, entonces

$$\begin{aligned}
G_{\{k_i\}_{i=1}^m}(n) &= P_{k_m} \circ \cdots \circ P_{k_1}(lm^{m-1} + im + j) \\
&= P_{k_m} \circ \cdots \circ P_{k_2}(lm^{m-2} + i_1 + 1 + (k_1 - 1)m^{m-1}) \\
&= P_{k_m} \circ \cdots \circ P_{k_2}(c_1 m^{m-2} + i_2 m + j_2) \\
&= P_{k_m} \circ \cdots \circ P_{k_3}(c_1 m^{m-3} + i_2 + 1 + (k_2 - 1)m^{m-1}) \\
&= P_{k_m} \circ \cdots \circ P_{k_3}(c_2 m^{m-3} + i_3 m + j_3)
\end{aligned}$$

$$
\begin{aligned}
&\vdots \\
&= P_{k_m} \circ P_{k_{m-1}} \circ P_{k_{m-2}}(c_{m-4}m^2 + i_{m-3} + 1 + (k_{m-3} - 1)m^{m-1}) \\
&= P_{k_m} \circ P_{k_{m-1}} \circ P_{k_{m-2}}(c_{m-3}m^2 + i_{m-2}m + j_{m-2}) \\
&= P_{k_m} \circ P_{k_{m-1}}(c_{m-3}m + i_{m-2} + 1 + (k_{m-2} - 1)m^{m-1}) \\
&= P_{k_m} \circ P_{k_{m-1}}(c_{m-2}m + i_{m-2} + 1) \\
&= P_{k_m}(c_{m-2} + 1 + (k_{m-1} - 1)m^{m-1}) \\
&= P_{k_m}(((k_{m-1} - 1)m^{m-2} + \cdots + (k_1 - 1))m + (l + 1)) \\
&= (k_{m-1} - 1)m^{m-2} + \cdots + (k_1 - 1) + 1 + (k_m - 1)m^{m-1} \\
&= [(k_m - 1) \ldots (k_1 - 1)]_{base\ m} + 1
\end{aligned}
$$

■

Observación 2.1 Del Teorema 2.1 concluimos lo siguiente:

1. El truco no depende de la carta sobre la cual se ejecuta, lo cual lo hace infalible.
2. El truco depende del esquema seguido, lo cual permite al "mago" ejecutarlo a su conveniencia.
3. La posición final de la carta escogida se puede calcular a través de una única fórmula general, lo cual unifica y simplifica el resultado original de Gergonne dado en [3].
4. La posición final se puede calcular utilizando el sistema de numeración de base m, lo cual demuestra en general lo comentado por Gardner en [2], sobre los casos de 3 y 10 pilas.

3. Situaciones particulares

En [2, pág. 33], Gardner dice, una vez que se ha ejecutado el procedimiento de las tres pilas, lo siguiente:

> *(...) the magician is able to do one of three things:*
>
> 1. *Name the exact position of the chosen card from the top of the packet.*
> 2. *Find the chosen card at a position previously demanded by the spectator.*

3. Name the card.

Inmediatamente, procede a discutir de manera heurística cada cosa separadamente, pero no da pruebas matemáticas formales de las mismas. A continuación, damos una prueba formal de cada situación en general; es decir, para toda m.

3.1. Nombrar la posición de la carta

En esta versión, le es permitido al espectador ensamblar las pilas después de cada repartición, recogiéndolas en cualquier orden que él desee. Incluso, la repartición puede ser hecha por el espectador. En realidad, no es necesario que el ejecutante ("mago") toque las cartas. Simplemente, debe observar cuidadosamente la colocación de la pila que contiene la carta escogida después de cada recolección. Establecemos esto, matemática y formalmente, en el siguiente corolario.

Corolario 3.1 Si el truco de m pilas de Gergonne es ejecutado, entonces la posición de la carta escogida siempre puede ser descubierta.

Demostración: Supongamos que una vez ejecutado el truco, las m posiciones que tomaron las pilas que contenían la carta escogida son: primero k_1 (k_1-ésima), luego k_2 (k_2-ésima), ..., y finalmente k_m (k_m-ésima), entendiendo que estos valores son 1 para el fondo, 2 para la segunda posición, ..., y m para el tope. Entonces, de acuerdo con el Teorema 2.1, la carta escogida, digamos n, -sin conocer su identidad- toma la posición:

$$[(k_m-1)\dots(k_1-1)]_{base\ m}+1 \;=\; \sum_{i=m}^{1}(k_i-1)m^{i-1}+1$$

$$=\; k_1+(k_2-1)m+\cdots+(k_m-1)m^{m-1}.$$

Por tanto, la posición de la carta escogida siempre puede ser descubierta. ■

En la práctica, el mago sólo tiene que observar cuidadosamente y hacer mentalmente los cálculos simples y parciales siguientes: primero memorizar k_1, después de la primera recolección; luego sumarle $(k_2-1)m$, después de la segunda recolección, ..., y finalmente, sumar $(k_m-1)m^{m-1}$ después de la última recolección.

3.2. Llevar la carta a una posición indicada

En esta segunda versión, le es pedido al espectador que establezca, antes de la ejecución del truco, la posición en la cual desea que aparezca la carta seleccionada por él una vez finalizado el truco. En este caso, debe permitírsele al mago ensamblar las pilas después de cada repartición. Al final del truco, la carta escogida debe encontrarse en la posición requerida por el espectador. Establecemos esto, matemática y formalmente, en el siguiente corolario.

Corolario 3.2 Si la posición requerida es la p-ésima ($1 \leq p \leq m^m$), entonces la carta escogida siempre puede ser llevada a esa posición cada vez que se ejecuta el truco de m pilas de Gergonne.

Demostración: Supongamos que la carta escogida es n (nuevamente su identidad no es relevante). Expresemos el entero $p-1$ en base m, digamos que $p-1 = [k_m \dots k_1]_{base\ m}$. Entonces, por el Teorema 2.1, al ejecutar el truco colocando las pilas que contienen la carta escogida según el esquema $\{k_i+1\}_{i=1}^{m}$; es decir, k_1+1 para la primera recolección, k_2+1 para la segunda, ..., y k_m+1 para la última, tenemos que la carta escogida pasa a ocupar la posición:

$$[k_m \dots k_1]_{base\ m} + 1 = (p-1) + 1 = p.$$

Por tanto, la carta escogida siempre puede ser llevada a una posición cualquiera preestablecida. ■

En la práctica, el mago sólo tiene que restar 1 al número indicado por el espectador y expresar el resultado en base m, luego sumar 1 a cada cifra obtenida e invertir el orden, y entonces proceder a ejecutar el truco utilizando este esquema.

3.3. Nombrar la carta

Finalmente, en esta versión también el truco puede ser ejecutado por el mismo espectador si lo desea, pero sujeto a las condiciones adicionales siguientes:

1. m es impar.

2. En las m recolecciones se coloca la pila que contiene la carta escogida en el medio de las demás.

Esta variante es llamada '*truco de m pilas de Gergonne clásico*'. Al final, la carta escogida ocupará siempre la $\frac{m^m+1}{2}$ -ésima posición, lo cual significa que la carta escogida pasa a ocupar la posición media de la baraja. Establecemos esto, matemática y formalmente, en el siguiente corolario.

Corolario 3.3 Si el truco de m pilas de Gergonne clásico es ejecutado, entonces la carta escogida siempre ocupará la $\frac{m^m+1}{2}$ -ésima posición y además puede ser nombrada.

Demostración: Supongamos que la carta escogida es n (nuevamente su identidad no es relevante). Observemos que la condición 1 garantiza que $m+1$ es par. Por otra parte, la condición 2 es equivalente a seguir el esquema $\{\frac{m+1}{2}\}_{i=1}^{m}$. Entonces, al aplicar el Teorema 2.1, tenemos que la carta n pasa a ocupar la posición

$$\left[\left(\frac{m+1}{2}-1\right)\ldots\left(\frac{m+1}{2}-1\right)\right]_{base\ m}+1 \quad =$$

$$\begin{aligned}\left[\left(\frac{m-1}{2}\right)\ldots\left(\frac{m-1}{2}\right)\right]_{base\ m}+1 &= \sum_{i=m}^{1}\left(\frac{m-1}{2}\right)m^{i-1}+1\\ = \quad \left(\frac{m-1}{2}\right)\left(\frac{m^m-1}{m-1}\right)+1 &= \frac{m^m-1}{2}+1\\ &= \frac{m^m+1}{2}.\end{aligned}$$

Por tanto, cada vez que la variante clásica es ejecutada la carta escogida siempre pasa a ocupar la $\frac{m^m+1}{2}$ -ésima posición. Finalmente, después de la última repartición el ejecutante observa las cartas medias de cada pila y una vez que el espectador indica la pila que contiene la carta escogida por él, simplemente recuerda la carta correspondiente a esa pila y la nombra. ■

En la práctica, el mago sabe que después de la m-ésima repartición debe memorizar la carta media de cada pila; es decir, la carta que ocupa la $\frac{m^{m-1}+1}{2}$ -ésima posición de cada pila, de manera que cuando el espectador diga cuál de éllas contiene la carta escogida, inmediatamente él sabrá cuál es la carta y entonces podrá nombrarla ("adivinarla").

Referencias

[1] BACHET C., *Problèmes plaisants & délectables qui se font par les nombres*, Gauthier-Villars, Paris, 1884.

[2] GARDNER M., *Mathematics Magic and Mystery*, Dover Publications Inc., Mineola, N. Y., 1956.

[3] GERGONNE, J. D., Récréations Mathématiques: Recherches sur un tour de cartes, *Annales de Mathématiques Pures et Appliquées*, IV (1813-1814), 276-283.

[4] HARRISON, J., BRENNAN, T., GAPINSKI, S., The Gergonne p-pile problem and the dynamics of the function $x \mapsto \lfloor (x+r)/p \rfloor$, *Discrete Applied Mathematics*, 82 (1998), 103-113.

[5] ROUSE BALL, W. W., COXETER, H. S. M., *Mathematical Recreations and Essays*, Dover Publications Inc., Mineola, N. Y., 1987.

ROY QUINTERO
DEPARTAMENTO DE FÍSICA Y MATEMÁTICAS,
UNIVERSIDAD DE LOS ANDES, TRUJILLO, VENEZUELA
rquinter@ula.ve

DIVULGACIÓN MATEMÁTICA

Kurt Gödel 1906-1978, una vida dedicada a la reflexión

Carlos Augusto Di Prisco

Kurt Gödel hizo contribuciones fundamentales a la lógica matemática y llevó a cabo profundas indagaciones filosficas. A pesar de la importancia de su trabajo, es poco lo que se sabe de su vida y de su obra fuera de reducidos círculos los especialistas. El centenario de su nacimiento es una ocasión apropiada para rendirle homenaje.

Gödel nació en Brno (o Brünn) ciudad de la antigua provincia astro-húngara de Moravia, el 26 de abril de 1906. Fue un niño extraordinariamente inquisitivo, y mostró desde muy joven un gran interés por las matemáticas. Su formación universitaria la obtuvo en la Universidad de Viena, en contacto con personajes de la talla del matemático Hans Hahn y el filósofo Rudolf Carnap, y bajo la influencia del grupo filosófico que dirigía Moritz Schlick, conocido posteriormente como el Círculo de Viena. Se casó con Adele Porkert en 1938, una mujer divorciada, seis años mayor que él, que trabajaba como bailarina. El matrimonio no se celebró sino años después de haberla conocido por las objeciones que ponía la familia de Gödel. Adele, quien no tenía grandes pretensiones intelectuales, lo acompañó por el resto de su vida y lo sobrevivió por tres años. Según testimonios de conocidos se trataban con devoción.

Los trabajos que publicó durante la década 1929-1939 transformaron la lógica matemática de una manera extraordinaria, por lo que se le ha considerado como el lógico más importante del siglo XX. Para comprender el alcance de los principales resultados que Gödel obtuvo durante estos años conviene recordar, aún si muy esquemáticamente, lo que es un sistema axiomático y algunos otros conceptos básicos de la lógica.

Un sistema axiomático consta de un lenguaje formal, con un conjunto definido de símbolos: variables, conectivas, cuantificadores, y posiblemente símbolos no lógicos que se usan para denotar relaciones, funciones o constantes de una teoría dada. Las expresiones bien formadas, o fórmulas del lenguaje, se obtienen agrupando símbolos mediante un conjunto bien determinado de reglas. Consta también de axiomas, y de reglas de inferencia . Los axiomas son expresiones del lenguaje formal tomadas como punto de partida para deducir otras, los teoremas, utilizando las reglas de inferencia.

Una demostración en el sistema axiomático de una expresión φ es entonces una sucesión finita de fórmulas,

$$\varphi_0, \varphi_1, \ldots, \varphi_n,$$

cada una de las cuales o bien es un axioma o se obtiene de expresiones anteriores en la sucesión aplicando alguna regla de inferencia, y tal que $\varphi_n = \varphi$.

Diremos que una fórmula es un enunciado si todas sus variables aparecen bajo el alcance de un cuantificador.

Como ejemplo de un sistema axiomático presentamos un sistema para la aritmética, que llamaremos AP, abreviando Aritmética de Peano.

- Lenguaje
  - Símbolos lógicos: variables, $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\forall$, (,).
  - Símbolos no lógicos: $+$, $\times$, S, 0.
    Para cada n, el término
    $$S(\ldots \text{ n veces} \ldots (S(0)) \ldots)$$
    se abrevia $\underline{n}$.
- Axiomas:
  - Axiomas Lógicos
  - Axiomas de Peano
- Reglas de inferencia:
  - Modus Ponens: de φ y $\varphi \rightarrow \psi$ se infiere ψ.
  - Generalización: de $\varphi(x)$ se infiere $\forall x \varphi(x)$.

Los axiomas lógicos son los de Hilbert y Ackermann [9], o algún sistema equivalente. A continuación listamos los Axiomas de Peano.

1. $0 \neq S(x)$
2. $S(x) = S(y) \rightarrow x = y$
3. $x + 0 = x$
4. $x + S(y) = S(x + y)$
5. $x \times 0 = 0$
6. $x \times S(y) = (x \times y) + x$

7. Para cada fórmula $\varphi(x)$,
   $(\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(S(x)))) \rightarrow \forall x\varphi(x)$

Un mismo lenguaje formal puede interpretarse de diversas maneras, y una expresión del sistema puede ser verdadera según una interpretación y falsa según otra.

Por ejemplo, el lenguaje de la aritmética puede ser interpretado en la estructura $\langle \mathbb{N}, +, \times, S, 0\rangle$, donde $+$ y $\times$ son respectivamente la suma y el producto de números naturales, S es la función sucesor y 0 corresponde al número cero; o puede ser interpretado en la estructura correspondiente de los números enteros con sus operaciones de suma y producto, función sucesor y cero.

El enunciado $\forall x \exists y(x + y = 0)$ es falso en $\langle \mathbb{N}, +, \times, S, 0\rangle$, pero es cierto en $\langle \mathbb{Z}, +, \times, S, 0\rangle$.

Incluso se se podría pensar en una interpretación más caprichosa de este mismo lenguaje: dado un conjunto A, en $\mathcal{P}(A)$, el conjunto de subconjuntos de A, $+$ podría denotar unión, $\times$ intersección, S la función complemento, y 0 el conjunto vacío.

Un enunciado se dice lógicamente válido si es verdad en cualquier interpretación. Por ejemplo en un lenguaje con símbolos no lógicos H y M, la expresión silogística

Si todo H es M, y x es H, entonces x es M.

es lógicamente válida, ya que resulta una expresión verdadera independientemente de cómo se interpreten los predicados H y M.

En [9], se plantea por primera vez el problema de la completitud del cálculo de predicados. En ese texto, Hilbert y Ackermann presentan un sistema axiomático para el cálculo de predicados y preguntan si este sistema es suficiente para demostrar todos los enunciados lógicamente válidos.

En su tesis doctoral, titulada Uber die Vollständigkeit des Logikkalculus, de 1929 (y luego publicada como [3]), Gödel muestra que los enunciados demostrables a partir de los axiomas del cálculo de predicados son exactamente los lógicamente válidos, es decir, aquellos que son verdad bajo cualquier interpretación. De esta manera Gödel dió respuesta afirmativa al problema planteado por Hilbert y Ackermann.

Unos años después, Gödel sorprendió al mundo matemático con otro resultado de gran impacto, su teorema de incompletitud de la aritmética y sistemas relacionados, que ha trascendido el ámbito matemático y filosófico. Se ha especulado mucho sobre el significado y el alcance de este teorema, se ha dicho, por ejemplo, que establece límites para el pensamiento racional. Gödel prefería pensar más bien que su teorema reafirma la necesidad de la creatividad humana para el desarrollo de las matemáticas ya que como consecuencia del teorema se sabe que es imposible mecanizar completamente el quehacer matemático.

El teorema de incompletitud establece que todo sistema axiomático consistente, recursivo, y que contenga la aritmética, es incompleto en el sentido de que hay un enunciado φ no demostrable en el sistema y tal que su negación $\neg\varphi$ tampoco es demostrable.

Que el sistema sea consistente significa que no demuestra ninguna contradicción; y que sea recursivo quiere decir que hay un algoritmo (un procedimiento mecánico) que sirve para determinar, dado un enunciado, si es o no uno de los axiomas del sistema.

Para un sistema que cumpla con las condiciones del teorema siempre habrá, entonces, algún enunciado aritmético verdadero (en la estructura de los números naturales) que no es demostrable en el sistema.

La demostración del teorema tiene dos ingredientes importantes, a saber, la aritmetización del lenguaje y, en segundo lugar, la representatividad en el sistema de las relaciones y funciones aritméticas recursivas. Esto da lugar a la aparición del fenómeno de la autoreferencia, lo que a su vez permite la construcción del enunciado indecidible de Gödel.

La aritmetización del lenguaje se logra asignando números a los elementos sintácticos, es decir a los símbolos del lenguaje, a las fórmulas, y a las sucesiones finitas de fórmulas, en particular a las demostraciones. Esta asignación se hace en forma algorítmica, de modo que existe un procedimiento que permite, dado un símbolo, calcular el número que le corresponde, e igualmente, dado un enunciado o una sucesión finita de enunciados, se puede determinar el número que le corresponde. Más aún, hay un procedimiento algorítmico para determinar en un número finito de pasos, dado un número, si ese número corresponde a algún símbolo, término, fórmula, enunciado, o sucesión finita de fórmulas o enunciados; y en caso afirmativo, el mecanismo permite determinar el elemento sintáctico correspondiente.

De este modo, mediante expresiones puramente aritméticas en el lenguaje formal del sistema podemos referirnos a enunciados, axiomas, demostraciones , y otros elementos sintácticos del mismo sistema axiomático.

Para lograr su demostración, Gödel inventó el concepto de función recursiva (y de relación recursiva). Ya esto es un logro extraordinario, puesto que ese concepto ha resultado central en la teoría de la computabilidad. Hoy sabemos que una función numérica es recursiva si y solamente si existe un algoritmo para calcularla (por ejemplo, un programa en algún lenguaje de programación). Del mismo modo, una relación numérica $R(x_1, \ldots, x_n)$ (de n argumentos) es recursiva si hay un algoritmo que determina, para cada tupla $k_1, \ldots, k_n$ de números naturales, si está o no en la relación R. Gödel demostró que toda relación recursiva $R(x_1, \ldots, x_n)$ es representable en AP. Es decir, existe una fórmula $\varphi(x_1, \ldots, x_n)$ del lenguaje, con las variables libres $x_1, \ldots, x_n$ tal que si los números $k_1, \ldots, k_n$ están relacionados por R, entonces el enunciado $\varphi(\underline{k_1}, \ldots, \underline{k_n})$ es demostrable en el sistema AP (recordemos que $\underline{n}$ es una

abreviación del término $S\ldots(\text{n veces})\ldots S(0)$ que denota al número n). Y si esos números no están relacionados por R, entonces en AP podemos demostrar la negación $\neg\varphi(\underline{k_1},\ldots,\underline{k_n})$.

Por ejemplo, la relación (unaria) "x es un número par" se representa por la fórmula $\exists y(y+y=\underline{n})$.

Es interesante mencionar que Gödel utilizó el Teorema Chino del Resto para demostrar la represntabilidad de las relaciones recursivas en el sistema AP.

Con estas herramientas, la demostración del teorema se centra en encontrar un número k, tal que el enunciado G de la aritmética que expresa:

"el enunciado aritmético cuyo número de Gödel es $\underline{k}$ no es demostrable en el sistema AP"

tiene número de Gödel k. En otras palabras, el enunciado G expresa de si mismo que no es demostrable. Se demuestra que si AP no es contradictorio, ese enunciado no es demostrable en AP; y su negación tampoco. El enunciado G es una versión sofisticada de la paradoja del mentiroso.

Para construir ese enunciado, Gödel consideró la relación $W(u,y)$ dada por

"u es el número de una fórmula $\varphi(v)$ (con variable libre v) y y es el número de la demostración en AP de $\varphi(\underline{u})$",

que, bajo las condiciones del teorema, es una relación recursiva, y por lo tanto representada en AP por una fórmula $\mathcal{W}(x,y)$.

Consideremos ahora la fórmula

$$(*)\ \forall y\neg\mathcal{W}(x,y),$$

y sea k su número de Gödel. Entonces

$$(G)\ \forall y\neg\mathcal{W}(\underline{k},y)$$

expresa " yo no soy demostrable".

Se prueba que, si AP es consistente, entonces (G) no es demostrable en AP y $\neg(G)$ tampoco lo es.

Una consecuencia de el teorema de incompletitud (y de su demostración)es el llamado Segundo Teorema de Incompletitud, que afirma que si un sistema cumple las mismas condiciones del Teorema de Incompletitud, entonces en ese sistema no se puede demostrar que el sistema es consistente. En particular, la aritmética no puede demostar su propia consistencia, a menos que sea inconsistente.

Expliquemos esto de forma más precisa. Si n es el número de Gödel de una fórmula, denotemos por $\ulcorner n\urcorner$ a dicha fórmula.

La relación

> "x es el número de una fórmula, y es el número de la negación de $\lceil x \rceil$, u es el número de una demostración de $\lceil x \rceil$, y v es el número de una demostración de $\lceil y \rceil$".

es recursiva, y luego representable en AP mediante una fórmula $\mathcal{K}(x, y, u, v)$. Por lo tanto

$$\forall x, y, u, v(\neg \mathcal{K}(x, y, u, v))$$

expresa que AP es consistente.

El Segundo Teorema de Incompletitud afirma que si AP es consistente, entonces

$$\forall x, y, u, v(\neg \mathcal{K}(x, y, u, v))$$

no es demostrable en AP.

Poco después de haber presentado su resultado de incompletitud como Habilitationsschrift en la Universidad de Viena, Gödel fue invitado por von Neumann a visitar Princeton, y entre 1933 y 1939 visitó los EEUU varias veces, donde llevó a cabo una intensa actividad. Entre una y otra visita sufrió serias crisis depresivas que lo obligaron a recluirse varias veces en sanatorios.

El otro resultado de Gödel que comentaremos se refiere a la hipótesis del continuo. La demostración de la consistencia del axioma de elección y de la hipótesis del continuo, otra de sus grandes contribuciones, fue anunciada en [5], en 1938, y desde entonces ha tenido una gran importancia en el desarrollo de la teoría de conjuntos.

Cantor demostró que para cualquier conjunto A, no hay una biyección entre A y su conjunto de partes $\mathcal{P}(A)$. En ese sentido, $|A|$, la cardinalidad de A, es estrictamente menor que $|\mathcal{P}(A)|$, la cardinalidad del conjunto de partes de A. En particular, como el conjunto $\mathbb{R}$ de los números reales se puede poner en biyección con los subconjuntos de $\mathbb{N}$, el conjunto de los números naturales, la cardinalidad de $\mathbb{N}$, es estrictamente menor que la cardinalidad de $\mathbb{R}$.

¿Existe algún tamaño intermedio? La Hipótesis del Continuo (HC) afirma que no hay conjuntos de tamaño intermedio entre $|\mathbb{N}|$ y $|\mathbb{R}|$.

Cantor llamó $\aleph_0$ a la cardinalidad de $\mathbb{N}$, el primer cardinal infinito; $\aleph_1$ es el primer cardinal mayor que $\aleph_0$, y por lo tanto el primer cardinal no numerable; $\aleph_2$ es el primer cardinal mayor que $\aleph_1$, etc.

Los cardinales forman una colección bien ordenada

$$\aleph_0 < \aleph_1 < \aleph_2 < \cdots < \aleph_\omega < \aleph_{\omega+1} < \ldots$$

Como $2^{\aleph_0}$ es la cardinalidad de $\mathcal{P}(\mathbb{N})$, se tiene que $|\mathbb{R}| = 2^{\aleph_0}$, y por lo tanto la Hipótesis del Continuo se expresa por:

$$2^{\aleph_0} = \aleph_1.$$

Gödel halló un modelo, llamado L, o el universo de los conjuntos constructibles, donde valen todos los axiomas de la teoría de conjuntos y también la HC. De aquí sigue que no se puede demostrar la negación de la HC a partir de los axiomas de la teoría de conjuntos (a menos que la teoría de conjuntos sea contradictoria).

Gödel sospechaba que la HC tampoco es demostrable, lo quedó confirmado muchos años después por el trabajo de Paul Cohen. De modo que la HC es indecidible en el sistema usual de la teoría de conjuntos. Algunos autores afirmado que estos resultados indican que el problema del continuo no tiene sentido, y de esta manera ha sido definitivamente resuelto. La tendencia contemporánea se inclina más bien por considerar que la independencia de la HC indica la necesidad de reformular el sistema axiomático de la teoría de conjuntos para obtener un sistema más fuerte donde esta hipótesis se pueda decidir. En 1947 Gödel publicó un excelente ensayo ([6]) donde expone sus ideas sobre cual sería una vía para lograr esto. Las ideas contenidas allí ponen de manifiesto la profunda intuición de Gödel y de manera casi profética apuntan en direcciones que ha resultado sumamente enriquecedoras. (Véase [12] para una presentación de resultados recientes sobre la HC, o [2] para un resumen expositivo de algunos de estos resultados).

Gödel publicó algunos otros trabajos durante la década 1929-1938, pero los que hemos mencionado son ya suficientes para dar una idea de la importancia de su obra.

Gödel publicó poco, pero cada uno de sus artículos es una joya por su precisión y claridad, que pone de manifiesto la profundidad y el alcance del pensamiento de su autor.

En 1940 Gödel abandonó Europa, huyendo del militarismo totalitario impuesto por los nazis y de la guerra que se había iniciado. Viajó en pleno invierno, atravesando Siberia, para llegar a Vladivostok, luego Japón, y de allí a los Estados Unidos, para establecerse en Princeton. En el Instituto de Estudios Avanzados de Princeton entabló una duradera amistad con Albert Einstein, con quien hablaba frecuentemente sobre temas científicos y filosóficos. Ya en los ú ltimos años de su vida, Einstein llegó a decir que su trabajo ya no significaba mucho para él, y que iba al Instituto para gozar del privilegio de caminar de regreso a su casa conversando con Gödel.

A partir de 1943, Gödel se dedicó casi exclusivamente a los estudios filosóficos, primero a la filosofía de las matemáticas y luego a la filosofía general. Sin embargo, entre 1947 y 1959 trabajó en problemas de la teoría general de la relatividad, y produjo unos sorprendentes resultados sobre modelos cosmológicos donde, en principio, es posible viajar al pasado. Gödel fue el primero en demostar que las ecuaciones de Einstein admiten soluciones que describen universos rotantes. Según el mismo Gödel, estos estudios no surgieron de discusiones con Einstein, sino de su interés en la filosofía kantiana del espacio y el

tiempo.


Gödel (c. 1937), y con Einstein en Princeton (1950)

Debido a su personalidad complicada, llena de temores y fobias, Gödel fue aislándose de colegas y amigos, y llevó durante sus últimos años una vida de reclusión. Cuando la Universidad de Princeton decidió otorgarle un doctorado honoris causa, no quiso comprometerse a asistir, y el mismo día de la ceremonia decidió finalmente no presentarse, por lo que el título nunca fue conferido. Meses más tarde, recibió un reconocimiento aún más importante: la medalla nacional de ciencia, que por cierto se confería al mismo tiempo al Premio Nobel Linus Pauling. De igual forma, y a pesar de que le ofrecieron todo tipo de facilidades para trasladarse a Washington a recibir el galardón de manos del Presidente Ford, decidió no asistir, y la medalla fue recibida en su nombre por el Profesor Saunders Mac Lane, para entonces Presidente de la American Mathematical Society.

Gödel estaba convencido de que el mundo se rige por principios racionales susceptibles de ser comprendidos por la mente humana, y consideraba fundamental el papel de la reflexión introspectiva para alcanzar el conocimiento. Fue un pensador extraordinario, cuya vida estuvo signada por las tensiones entre su racionalidad científica y la inestabilidad emocional. Gödel trabajó en el Instituto de estudios Avanzados hasta que murió el 15 de enero de 1978, por desnutrición e inanición, como consecuencia de paranoias que lo llevaron a dejar de alimentarse por temor a envenenamientos. Su pensamiento, que ha orientado de manera significativa el desarrollo de la lógica matemática, merecería ser mejor conocido.

Después de su muerte han aparecido varios libros sobre Gödel, como las recomendables obras [1] y [11]. Parte del trabajo de Gödel ha sido presentado en el exitoso [10]. La obra [7] reúne todos los escritos de Gödel publicados hasta 1981, traducidos al castellano. Una edición en cinco volúmenes hecha por Oxford bajo los auspicios de la Association for Symbolic Logic ([8]) recoge toda la obra escrita por este extraordinario y singular personaje, incluyendo escritos no publicados anteriormente y toda su correspondencia. En esta extraordinaria recopilación, los artículos van precedidos de comentarios críticos escritos por renombrados especialistas.

Referencias

[1] Dawson, Jr., J. W., Logical Dilemmas. The life and work of Kurt Gödel. A K Peters, 1997.

[2] Di Prisco, C. A., Are we closer to a solution of the Continuum Problem? Manuscrito - Rev. Int. Fil., Campinas, v. 28 (2005) 331-350.

[3] Gödel, K., Die Vollständigkeit der Axiome des logischen Funktionenkalkulus, Monatshefte für Mathematik und Physik 37 (1930) 349-360.

[4] Gödel, K., Über formal unentscheidbare Sätze des Principia mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik 38 (1931) 173-198.

[5] Gödel, K., The consistency of the axiom of choice and the generalized continuum hypothesis. Proceedings of the National Academy of Sciences, U.S.A. 24 (1938) 556-557.

[6] Gödel, K., What is cantor's continuum problem?, American Mathematical Monthly 54 (1947) 515-525; errata 55, 151.

[7] Gödel, K., Obras completas. Alianza Editorial, 1981.

[8] Gödel, K., Collected Works. (S. Feferman et al. Eds.) Oxford University Press, Vol. 1, 1986; Vol. 2, 1990; Vol. 3, 1995; Vol. 4, 2003; Vol. 5, 2003.

[9] Hilbert, D. y W. Ackermann, Grundzuge der teoretischen logik. Springer, Berlin, 1928

[10] Hofstadter, D. R., Gödel, Escher, Bach: an eternal golden braid. Basic Books, Inc., 1979.

[11] Wang, H., Reflections on Kurt Gödel. MIT Press, 1988.

[12] Woodin, W. H., The Continuum Hypothesis, *Notices of the Amer.Math, Soc.*. 48 (2001). Part I 567-576, Part II 681-690.

Carlos A. Di Prisco
Instituto Venezolano de Investigaciones Científicas
Venezuela
cdiprisc@ivic.ve

DIVULGACIÓN MATEMÁTICA

Problemas con Subgrupos Discretos y Subgrupos Densos

José O. Araujo & Laura B. Fernández

Resumen

En este trabajo presentamos algunas aplicaciones de la geometría reticular y subgrupos densos en la recta y el plano real, especialmente del teorema de Minkowski en el plano.
Los problemas tratados son sobre polígonos regulares, aproximación y teoría elemental de números.
Palabras y frases claves: Minkowski, retículos, aproximación. [1]

Discrete and Dense Subgroup Problems

Abstract

In this work we present some aplications of the reticular geometry and dense subgroups of the real line and the real plane, especially of the Minkowski's theorem in the plane.
The problems we deal are over regular polygons, approximation and elemental theory of numbers.
Key words and phrases: Minkowski, lattices, approximation.

1 Introducción

En estas notas presentamos los conceptos de conjuntos densos y conjuntos discretos sobre la recta y el plano real. Se analiza particularmente, los subgrupos de la recta real con su estructura aditiva y, un poco más general, los subconjuntos aditivos de los números reales. Por otra parte, se presenta el teorema de Minkowski en el plano relativo a puntos reticulares en una figura convexa. Con el propósito de ilustrar sobre la utilidad de estos conceptos, las conclusiones obtenidas se aplican a una serie de problemas de aproximación de números por elementos de un subgrupo o de un conjunto aditivo. También se tratan aplicaciones del teorema de Minkowski relacionadas con el teorema de los cuatro

[1] 1991 Mathematics Subject classfication: Primary 52C05

cuadrados y con la ecuación de Pell. Finalmente se plantean problemas sobre los números complejos unitarios, teniendo en cuenta que la estructura multiplicativa de estos responde a la estructura aditiva de sus argumentos.

Los símbolos utilizados corresponden a la siguiente asignación:
$\mathbb{N}$: números naturales
$\mathbb{Z}$: números enteros
$\mathbb{R}$: números reales
$\mathbb{C}$: números complejos.

2 En la Recta Real

Un subconjunto $\mathcal{A}$ de los números reales $\mathbb{R}$ se dirá un *subconjunto denso* si para $r \in \mathbb{R}$ y $\varepsilon > 0$, existe $a \in A$ tal que $|r - a| < \varepsilon$.
Un subconjunto $\mathcal{A}$ de los números reales $\mathbb{R}$ se dirá un *subconjunto discreto* si para cada $a \in A$ existe $\varepsilon > 0$ tal que $(a - \varepsilon, a + \varepsilon) \cap \mathcal{A} = \{a\}$.
Por ejemplo, los siguientes subconjuntos son densos:
i) Los números racionales.
ii) Los números irracionales.
iii) $\{x \in \mathbb{R} / sen\,(x) \neq 0\}$.
iv) El complemento de un subconjunto discreto.
Los siguientes subconjuntos son discretos:
i) Cualquier conjunto finito.
ii) Los números naturales
iii) Los enteros múltiplos de 7.
iv) $\{x \in \mathbb{R} / \cos(x) = 0\}$.
Se propone como ejercicio comprobar las afirmaciones precedentes.

El concepto de densidad está estrechamente ligado al concepto de aproximación, por ejemplo al decir que los números racionales son densos, decimos que todo número real puede aproximarse arbitrariamente con números racionales. En general, que el conjunto $\mathcal{A}$ sea denso en $\mathbb{R}$, significa que cualquier número real puede ser aproximado arbitra-riamente con elementos de $\mathcal{A}$.
Es particularmente interesante el caso en que los subconjuntos considerados son subgrupos de $\mathbb{R}$ considerado con su estructura aditiva. Damos a continuación los conceptos de grupo abeliano y subgrupos.

Un conjunto $\mathcal{G}$ provisto de una operación binaria "+" se dice *grupo* si se verifican:
i) $(a + b) + c = a + (b + c) \quad \forall\, a, b, c \in \mathcal{G}$ *(Asociativa).*
ii) Existe $o \in \mathcal{G}$ tal que $a + o = o + a = a \quad \forall\, a \in \mathcal{G}$ *(con elemento neutro).*

iii) $\forall\, a \in \mathcal{G}$ existe $b \in \mathcal{G}$ tal que $a + b = b + a = o$ *(con inverso).*
Un grupo $\mathcal{G}$ se dice *abeliano* si además se verifica:
iv) $a + b = b + a \quad \forall a, b \in \mathcal{G}$.

El elemento b de la condición *iii)* resulta único, se llama el *inverso de* a y se notará con $-a$, y como es usual, se usará $a - b$ para indicar la suma $a + (-b)$.

Un subconjunto $\mathcal{H}$ de un grupo $\mathcal{G}$ se dice un *subgrupo de* $\mathcal{G}$ si se cumplen:
i) $o \in \mathcal{H}$.
ii) Si a y $b \in \mathcal{H}$ entonces $a - b \in \mathcal{H}$.

Como consecuencias de las condiciones *i)* y *ii)* precedentes, se tiene:
Si $\mathcal{H}$ es un subgrupo de un grupo $\mathcal{G}$, se verifican:
i) Si $a \in \mathcal{H}$ entonces $-a \in \mathcal{H}$.
ii) Si $a, b \in \mathcal{H}$ entonces $a + b \in \mathcal{H}$.

Como ejemplos de grupos abelianos tenemos:
i) Los números enteros $\mathbb{Z}$ con la suma usual.
ii) Los números reales con la suma usual.
iii) Los vectores en el plano con la suma usual de vectores.
iv) El conjunto $\mathbb{Z}_n = \{0, 1, ..., n-1\}$ con la suma módulo n.
v) $\mathbb{R} - \{0\}$ con el producto usual.
Como ejemplos de subgrupos:
i) Los números pares forman un subgrupo de los enteros con la suma.
ii) Los números enteros forman un subgrupo de los reales con la suma.
iii) $\mathbb{Z}\left[\sqrt{2}\right] = \left\{m + n\sqrt{2} : m, n \in \mathbb{Z}\right\}$ es un subgrupo de $\mathbb{R}$ con la suma.
iv) Los números reales positivos forman un subgrupo de $\mathbb{R} - \{0\}$ con el producto usual.
Se propone como ejercicio comprobar las afirmaciones precedentes.

Naturalmente que hay subconjuntos de $\mathbb{R}$ que no son discretos ni densos por ejemplo los reales positivos entre otros tantos, pero si consideramos como $\mathbb{R}$ el grupo abeliano con la operación suma, el teorema a conti-nuación, no deja otra alternativa para un subgrupo de $\mathbb{R}$ que la de ser un subconjunto discreto o un subconjunto denso.

Teorema 2.1. *Si* $\mathcal{H}$ *es un subgrupo de* $\mathbb{R}$*, entonces* $\mathcal{H}$ *es un subgrupo discreto o* $\mathcal{H}$ *es un subconjunto denso.*

Demostración: Comencemos observando que, si $n \in \mathbb{Z}$ y $h \in \mathcal{H}$, entonces $nh \in \mathcal{H}$. En efecto: $1h = h \in \mathcal{H}$, $2h = h + h \in \mathcal{H}$, $3h = 2h + h \in \mathcal{H}$, y en general se tiene

$$(n+1)\,h = nh + h$$

identidad que permite probar por inducción que $nh \in \mathcal{H}\,, \forall\, n \in \mathbb{N}$ y h arbitrario en $\mathcal{H}$. Ahora si n es un entero negativo, expresando $nh = (-n)\,(-h) \in \mathcal{H}$ y teniendo en cuenta que $-h \in \mathcal{H}$, del caso anterior se tiene $nh \in \mathcal{H}$.
Finalmente, $0h = 0$, lo que concluye con la prueba de nuestra afirmación.
Si $\mathcal{H} = \{0\}, \mathcal{H}$ es discreto. En caso contrario, $\mathcal{H}$ tiene un elemento $h \neq 0$. Dado que $-h \in \mathcal{H}$, resulta que $\mathcal{H}$ tiene un elemento positivo h_0 y podemos considerar

$$r = \inf\ \{h \in \mathcal{H} : h > 0\}$$

Si $r = 0$, sea x arbitrario en $\mathbb{R}$ y definimos los conjuntos

$$\mathcal{I} = \{h \in \mathcal{H} : h < x\} \quad \text{y} \quad \mathcal{J} = \{h \in \mathcal{H} : h > x\}$$

Acorde con lo observado al comienzo de la demostración, se tiene que

$$\mathbb{Z}h_0 = \{nh_0 : n \in \mathbb{Z}\} \subseteq \mathcal{H}$$

En consecuencia, $\mathcal{I}$ y $\mathcal{J}$ resultan conjuntos no vacíos. Es claro que

$$\sup \mathcal{I} \leq x \leq \inf \mathcal{J}$$

Como $r = 0$, para $\varepsilon > 0$ existe $h \in \mathcal{H}$ tal que $0 < h < \varepsilon$. En tal caso podemos elegir $k \in \mathbb{Z}$ de modo que

$$kh \leq x < (k+1)\,h$$

es decir $kh \in \mathcal{I}$ y $(k+1)\,h \in \mathcal{J}$, luego

$$\inf \mathcal{J} - \sup \mathcal{I} \leq (k+1)\,h - kh = h < \varepsilon \qquad \forall \varepsilon > 0$$

Debe ser $\sup \mathcal{H} = x = \inf \mathcal{J}$. Esto indica que un número real cualquiera puede ser aproximado arbitrariamente, tanto por la izquierda como por la derecha por elementos de $\mathcal{H}$, siendo $\mathcal{H}$ de este modo un subconjunto denso de $\mathbb{R}$.
En otro caso, $r > 0$ y $\mathcal{H}$ no puede tener más que un elemento en el intervalo

$$[r, 2r) = \{x \in \mathbb{R} : r \leq x < 2r\}$$

pues de haber dos elementos $h < h'$ de $\mathcal{H}$ en este intervalo, tendríamos que

$$0 < h - h' < r$$

pero esto contradice la condición de ínfimo del número r. En conclusión, $r \in \mathcal{H}$.
En este caso tenemos

$$\mathbb{Z}r = \{nr :\ n \in \mathbb{Z}\} \subseteq \mathcal{H}$$

Razonando como antes, encontraremos que

$$[nr, (n+1)\ r) \cap \mathcal{H} = \{nr\}$$

es decir

$$\mathcal{H} = \mathbb{Z}r$$

y resulta claro que $\mathbb{Z}\,r$ es un subconjunto discreto de $\mathbb{R}$. ■

Observemos que los subgrupos discretos de $\mathbb{R}$ quedan caracterizados por los subconjuntos de la forma $\mathbb{Z}\,r$, para algún $r \in \mathbb{R}$. Usaremos los términos *subgrupos discretos* o *subgrupos densos* para referirnos a subgrupos que son respectivamente conjuntos discretos o conjuntos densos.
Como aplicación del teorema 2.1, consideraremos los siguientes pro-blemas.

Problema 1. *Demostrar que todo número real puede aproximarse arbitrariamente por elementos del conjunto*

$$\mathbb{Z}\left[\sqrt{2}\right] = \left\{n + m\sqrt{2} : n, m \in \mathbb{Z}\right\}$$

El conjunto $\mathbb{Z}\left[\sqrt{2}\right]$ es un subgrupo de $\mathbb{R}$ con la suma, luego es denso o de la forma $\mathbb{Z}r$ para algún $r \in \mathbb{R}$. Si no fuese denso tendríamos enteros n y m tales que

$$1 = nr \quad \text{y} \quad \sqrt{2} = mr$$

Luego $\sqrt{2} = m/n$ resultaría un número racional.

Problema 2. *Sea α un número irracional, mostrar que todo número real y tal que $0 < y < 1$ puede aproximarse arbitrariamente por elementos del conjunto*

$$\{\{n\alpha\} : n \in \mathbb{Z}\}$$

donde $\{x\}$denota la mantisa del número real x, más precisamente $\{x\}$ es $x - [x]$ donde $[x]$ es la parte entera de x.

Como en el caso anterior, dado que α es irracional, puede mostrarse sin mayor dificultad que el conjunto

$$\{n + m\alpha : n, m \in \mathbb{Z}\}$$

es un subgrupo denso de $\mathbb{R}$. Escribimos

$$n + m\alpha = n + [m\alpha] + \{m\alpha\} = [n + m\alpha] + \{m\alpha\}$$

Si y es un número real tal que $0 < y < 1, y$ se aproxima arbitrariamente por la elementos de la forma $n + my$, dándose las mejores aproximaciones cuando $[n + my] = 0$, es decir por los elementos de la forma $\{my\}$.

Consideremos $\mathcal{C} = \{z \in \mathbb{C} : |z| = 1\}$, $\mathcal{C}$ es un grupo abeliano con el producto de números complejos. Los elementos en $\mathcal{C}$ pueden presentarse en su forma exponencial o polar como

$$z = \exp(2\pi i\theta) = cos(2\pi i\theta) + i\, sen(2\pi i\theta) \quad \text{con} \quad 0 \leq \theta < 1$$

es decir, $z \in \mathcal{C}$ depende sólo del parámetro real θ.

Problema 3. *Si $\mathcal{H}$ es un subgrupo de $\mathcal{C}$, mostrar que $\mathcal{H}$ es finito o bien todo elemento de $\mathcal{C}$ puede ser aproximado arbitrariamente por elementos de $\mathcal{H}$, esto último se expresa diciendo que $\mathcal{H}$ es denso en $\mathcal{C}$.*

Consideremos

$$\mathcal{H}' = \{x \in \mathbb{R} : \exp(2\pi ix) \in \mathcal{H}\}$$

Se tiene

i) $0 \in \mathcal{H}'$, pues $\exp(2\pi i0) = 1$.

ii) Si $x, y \in \mathcal{H}'$, entonces $x - y \in \mathcal{H}'$ pues

$$\exp(2\pi i(x-y)) = \exp(2\pi ix)\exp(2\pi iy)^{-1}$$

De aquí resulta que $\mathcal{H}'$ un subgrupo de $\mathbb{R}$ con la estructura aditiva, luego $\mathcal{H}'$ es denso o discreto. Si $\mathcal{H}'$es denso, todo número real θ en $[0,1)$ se aproxima arbitrariamente por elementos de $\mathcal{H}'$, luego todo elemento de $\mathcal{C}$ se aproxima arbitrariamente por elementos de la forma $\exp(2\pi ix)$ con $x \in \mathcal{H}'$, es decir, con elementos de $\mathcal{H}$.

Para precisar esta última afirmación, usaremos la desigualdad

$$|sen(t)| \leq |t| \qquad \forall t \in \mathbb{R}$$

En primer lugar notemos que

$$\begin{aligned} |\exp(2\pi ix) - \exp(2\pi i\theta)| &= |\exp(2\pi i\theta)|\,|\exp(2\pi i(x-\theta)) - 1| \\ &= |\exp(2\pi i(x-\theta)) - 1| \end{aligned}$$

Por otra parte, para $\alpha \in \mathbb{R}$

$$\begin{aligned} |\exp(i\alpha) - 1| &= \sqrt{(cos(\alpha) - 1)^2 + sen(\alpha)^2} \\ &= 2sen\left(\frac{\alpha}{2}\right) \end{aligned}$$

Se sigue que

$$\begin{aligned} |\exp(2\pi ix) - \exp(2\pi i\theta)| &= 2\,|sen(\pi(x-\theta))| \\ &\leq 2\pi\,|x-\theta| \end{aligned}$$

lo que muestra la densidad de $\mathcal{H}$ en $\mathcal{C}$.
Si $\mathcal{H}'$ es discreto, hay sólo un número finito de elementos de $\mathcal{H}'$ en el intervalo $[0,1)$, en consecuencia hay sólo un número finito de elementos en $\mathcal{H}$.

La afirmación en el teorema expuesto sigue siendo válida con bajo una hipótesis ligeramente más débil.
Diremos que $\mathcal{H}$ es un *subconjunto aditivo* de $\mathbb{R}$ si dados a y b en $\mathcal{H}$ entonces $a+b$ también pertenece a $\mathcal{H}$.
Observemos que si $\mathcal{H}$ es aditivo, usando inducción, puede mostrarse que si $a \in \mathcal{H}$ y $n \in \mathbb{N}$, entonces $na \in \mathcal{H}$.
Ejemplos de subconjuntos aditivos que no sean subgrupos podemos citar:
i) El conjunto $\mathbb{N}$ de los números naturales.
ii) Los número racionales menores que -1.
iii) $\left\{\ln\left(\frac{2^n}{3^m}\right) : n, m \in \mathbb{N}\right\}$
Tenemos entonces el siguiente teorema:

Teorema 2.2. *Si $\mathcal{H}$ es un subconjunto aditivo de $\mathbb{R}$ conteniendo elementos positivos y elementos negativos, entonces $\mathcal{H}$ es un subconjunto denso en $\mathbb{R}$ ó $\mathcal{H}$ es un subgrupo discreto de $\mathbb{R}$.*

Demostración: Notemos con I y S respectivamente

$$I = \inf\{h \in \mathcal{H} : h > 0\}$$

$$S = \sup\{h \in \mathcal{H} : h < 0\}$$

Se tiene $S \leq 0 \leq I$, de este modo es

$$S \leq S + I \leq I$$

En el intervalo $[S, I]$ pueden aproximarse arbitrariamente con elementos de $\mathcal{H}$ únicamente S, I y eventualmente el cero. Dado que $S+I$ puede ser aproximado arbitrariamente por elementos de $\mathcal{H}$, las posibilidades son

$$S + I = I \quad S + I = 0 \quad \text{o} \quad S + I = S$$

Si $S + I \neq 0$, entonces $S = 0$ ó $I = 0$. Supongamos que $S = 0$, debe ser $I > 0$. De aquí que podemos elegir $h \in \mathcal{H}$ tal que

$$-I < h < 0$$

pues $S = 0$ es el supremo de los elementos negativos de $\mathcal{H}$. Si $h' \in \mathcal{H}$ es positivo, $h' + nh \in \mathcal{H}$ para todo $n \in \mathbb{N}$, siendo $h' \geq I$, existe $k \in \mathbb{N}$ tal que

$$h' + kh \geq I > h' + (k+1)\,h$$

luego

$$I > h' + (k+1)\, h = h' + k\, h + h \geq I + h > 0$$

Encontramos una contradicción pues I es el ínfimo de los elementos positivos de $\mathcal{H}$.
En forma análoga, se trata el caso $I = 0$, y como conclusión se obtiene que $I + S = 0$.
Si $I = S = 0$, $\mathcal{H}$ posee elementos positivos y elementos negativos de módulos arbitrariamente pequeños, o sea cero puede ser aproximado, en forma arbitraria, por la izquierda y por la derecha con elementos de $\mathcal{H}$.
Sea $\varepsilon > 0$ y $r \in \mathbb{R}$ cualquier número positivo. Consideremos $h \in \mathcal{H}$ tal que

$$0 < h < \min\{r, \varepsilon\}$$

La sucesión nh con $n \in \mathbb{N}$, está formada por elementos de $\mathcal{H}$. Existe un número natural k tal que

$$kh \leq r < (k+1)\, h$$

Se tiene

$$r - \varepsilon < r - h < kh \leq r < (k+1)\, h = kh + h \leq r + h < r + \varepsilon$$

es decir el intervalo $(r - \varepsilon, r + \varepsilon)$ contiene dos elementos de $\mathcal{H}$, uno a la izquierda y otro a la derecha de r.
En forma similar se trata el caso en que r sea negativo, concluyendo que $\mathcal{H}$ es denso.
Sea ahora $I > 0$, $S = -I$. Supongamos que $\mathcal{H}$ posea un elemento h en el intervalo $(I, 2I)$. Consideremos h' en $\mathcal{H}$ tal que

$$-I - (h - I) < h' \leq -I$$

es decir

$$0 \leq h' + h$$

pero como

$$h < 2I \quad \text{y} \quad h' \leq -I$$

tendremos que $h' + h < I$ y esto no es posible pues I es el ínfimo de los elementos positivos de $\mathcal{H}$.
Resulta entonces que $I \in \mathcal{H}$. Al sumarle $2I$ a los elementos de $\mathcal{H}$ en el intervalo $(-2I, -I)$ obtenemos elementos de $\mathcal{H}$ en el intervalo $(0, I)$, por lo que no hay elementos de $\mathcal{H}$ en $(-2I, -I)$ y en consecuencia $-I \in \mathcal{H}$.
Finalmente, tenemos $\mathbb{Z}I \subseteq \mathcal{H}$ y en forma análoga a la demostración del teorema 2.1, obtenemos que $\mathcal{H} = \mathbb{Z}I$. ■

Es preciso mostrar que hay subconjuntos aditivos en las condiciones del teorema 2.2 que no son subgrupos, y en tal casos son subconjuntos densos.

Por ejemplo

$$\mathcal{H} = \left\{ n - m\sqrt{2} : n, m \in \mathbb{N} \right\}$$

es un subconjunto aditivo con elementos positivos y elementos negativos. Es simple mostrar que $0 \notin \mathcal{H}$, más aún, resulta claro que si $h \in \mathcal{H}$ entonces $-h \notin \mathcal{H}$. Como aplicación del teorema 2.2 tenemos:

Problema 4. *Sean p y q números naturales con $q > 1$, sea*

$$\mathcal{K} = \left\{ \frac{p^i}{q^j} : i, j \in \mathbb{N} \right\}$$

Entonces $\mathcal{K}$ es denso en los reales positivos ó p y q son ambos potencias de un mismo número natural h.

En efecto, sea

$$\mathcal{H} = \{\ln(k) : k \in \mathcal{K}\}$$

Como $\mathcal{K}$ es multiplicativo y contiene elementos mayores que 1 y elementos menores que 1, resulta $\mathcal{H}$ un subconjunto aditivo de $\mathbb{R}$ en las condiciones del teorema 2.2. Si $\mathcal{H}$ es discreto, entonces $\mathcal{H} = \mathbb{Z}\alpha$ para algún real α positivo. Definiendo

$$\beta = e^{\alpha} = \frac{p^r}{q^s}$$

los elementos de $\mathcal{K}$ son exactamente los números

$$\mathcal{K} = \{\beta^m : m \in \mathbb{Z}\}$$

Por otra parte como $0 \in \mathcal{H}$, $1 \in \mathcal{K}$, de modo que existen números naturales i y j tales

$$1 = \frac{p^i}{q^j}$$

luego, de las identidades

$$p = \frac{p^{i+1}}{q^j} \quad \text{y} \quad \frac{1}{q} = \frac{p^i}{q^{j+1}}$$

obtenemos que p y $1/q$ están en $\mathcal{K}$. Dado que $\beta > 1$, existen números naturales n y m tales que

$$p = \beta^n \quad \text{y} \quad q = \beta^m$$

de donde β, que en principio es racional, debe ser un número natural.
Por otra parte, si $\mathcal{H}$ es denso en $\mathbb{R}$, $\mathcal{K}$ es denso en los reales positivos por la continuidad de la función exponencial.

Problema 5. *Mostrar que dado un número natural k existen infinitas potencias de 2 cuyo desarrollo decimal comienza con k.*

Por ejemplo:

$$\begin{array}{cc} k=1 & 2^0=1 \\ k=3 & 2^5=32 \\ k=6 & 2^6=64 \\ k=10 & 2^{10}=1024 \\ k=13 & 2^{17}=131072 \end{array}$$

El problema se reduce a encontrar números naturales n y m tal que

$$10^m k \leq 2^n < 10^m\,(k+1)$$

En tal caso $2^n = 10^m k + h$ con $0 \leq h < 10^m$ lo que garantiza que los dígitos iniciales de 2^n son los dígitos de k.
Del problema anterior sabemos que

$$\{2^n/10^m : n, m \in \mathbb{N}\}$$

es un conjunto denso en los reales positivos, luego el intervalo $[k, k+1)$ contiene infinitos números de este conjunto.

Nota: Por el problema 4, en el problema precedente, puede reemplazarse 2 por cualquier número natural que no sea una potencia de 10. También podría cambiarse la base de numeración y enunciarse un problema análogo.

3 En el Plano

Si consideramos el plano real $\mathbb{R}^2$ como grupo abeliano con la suma de vectores, no es cierto que un subgrupo de $\mathbb{R}^2$ sea denso o discreto, entendiendo en este caso por *subconjuntos densos*, aquellos conjuntos cuyos elementos pueden aproximar arbitrariamente cualquier vector del plano, y por *subconjuntos discretos*, aquellos conjuntos en los que cada uno de sus puntos puede ubicarse en el centro de un círculo que deje en su exterior a los puntos restantes del conjunto. Una definición más formal de estos conceptos se dará más adelante.
Como ejemplos tenemos:

i) Una recta por el origen, en el plano, es un subgrupo de $\mathbb{R}^2$ que no es discreto ni es denso en $\mathbb{R}^2$.

ii) Los puntos de coordenadas enteras forman un subgrupo discreto del plano.

iii) Los puntos de coordenadas racionales forman un subgrupo denso del plano.

iv) Los puntos con primer coordenada entera y segunda coordenada racional, forman un subgrupo del que no es discreto ni denso.

En el plano, considerando la distancia y norma euclídeas dadas por

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad \|x\| = \sqrt{x_1^2 + x_2^2}$$

para cada $x, y \in \mathbb{R}^2$, $x = (x_1, x_2)$ e $y = (y_1, y_2)$, puede afirmarse lo siguiente:

Teorema 3.1. *Sea $\mathcal{H}$ un subgrupo de $\mathbb{R}^2$ y*

$$\delta = \inf\{\|h\| : h \in \mathcal{H} \text{ y } h \neq 0\}$$

Entonces, $\mathcal{H}$ es un subgrupo discreto si y sólo si $\delta > 0$.

Demostración: Supongamos que $\mathcal{H}$ es un subgrupo discreto. Dado que $\mathcal{H}$ es un subgrupo, $0 \in \mathcal{H}$, y por ser $\mathcal{H}$ un subconjunto discreto de $\mathbb{R}^2$, existe $\varepsilon > 0$ tal que

$$\|h - 0\| = \|h\| > \varepsilon \qquad \forall h \in \mathcal{H}, h \neq 0$$

Luego $\delta > 0$.
Recíprocamente, si $\delta > 0$, para $h \in \mathcal{H}$ consideremos el círculo dado por

$$\|x - h\| < \delta$$

es decir la totalidad de puntos del plano que distan de h en menos que δ.
Si h' es un elemento de $\mathcal{H}$ en dicho círculo, se tiene:

$$\|h' - h\| < \delta$$

por la definición de δ, debe ser $h' - h = 0$, o sea $h' = h$. Se sigue que $\mathcal{H}$ es un subconjunto discreto de $\mathbb{R}^2$. ■

Observación: Un subconjunto discreto en el plano, y a la vez de la recta, es el formado por los elementos de la sucesión $\left(\frac{1}{n}, 0\right)$. A medida que n aumenta, es necesario un círculo más pequeño para aislar a $\left(\frac{1}{n}, 0\right)$ del resto de los elementos de la sucesión. En cambio, en un subgrupo discreto, sea en la recta o el plano, es posible aislar todos sus elementos con círculos del mismo radio. En efecto, el caso de la recta es claro, a partir de la caracterización dada en el teorema 2.1. En el plano, sea $\varepsilon > 0$ de modo que un círculo con radio ε aisla un elemento $h \in \mathcal{H}$ del resto de los elementos de $\mathcal{H}$, es decir

$$\left\{x \in \mathbb{R}^2 : \|x - h\| < \varepsilon\right\} \cap \mathcal{H} = \{h\}$$

Si dos elementos $f, g \in \mathcal{H}$ se encontraran a menor distancia que ε, entonces la distancia entre $h + f - g \in \mathcal{H}$ y h es

$$\|h + f - g - h\| = \|f - g\| < \varepsilon$$

por lo que debe ser

$$h + f - g = h$$

o sea $f = g$.
Concluímos que todos los puntos de $\mathcal{H}$ pueden ser aislados usando círculos con mismo radio.
Como consecuencia de la observación precedente, tenemos:

Proposición 3.2. *Si $\mathcal{H}$ es un subgrupo discreto del plano, una región acotada $\mathcal{F}$ del plano sólo puede contener un número finito de elementos de $\mathcal{H}$.*

Demostración: Dado que $\mathcal{F}$ es acotada, podemos elegir un círculo $\mathcal{D}$ que contenga a $\mathcal{F}$. Supongamos que con círculos de radio δ se puede aislar los elementos de $\mathcal{H}$ entre sí. Si c es el centro de $\mathcal{D}$ y ρ su radio, el círculo $\mathcal{D}'$ con centro c y radio $\delta + \rho$, contiene todos los discos, disjuntos dos a dos, dados por

$$\left\{x \in \mathbb{R}^2 : \|x - h\| < \delta\right\} \qquad \forall h \in \mathcal{H} \cap \mathcal{D}$$

lo que resulta de

$$\|x - c\| = \|x - h + h - c\| \leq \|x - h\| + \|h - c\| < \delta + \rho$$

El área de la figura formada por estos discos es

$$|\mathcal{H} \cap \mathcal{D}| \times \pi\delta^2$$

y no puede exceder al área de $\mathcal{D}'$, de modo que $|\mathcal{H} \cap \mathcal{D}|$, el número de elementos de $\mathcal{H}$ en $\mathcal{D}$, debe ser finito, y en consecuencia, también resulta finito el número de elementos de $\mathcal{H}$ en $\mathcal{F}$. ■

Consideremos ahora $\mathcal{C}^2 = \mathcal{C} \times \mathcal{C}$, el producto cartesiano del conjunto de complejos unitarios $\mathcal{C}$ consigo mismo. $\mathcal{C}^2$ tiene estructura de grupo abeliano definiendo

$$(z, w) \cdot (z', w') = (zz', ww')$$

En $\mathcal{C}^2$ definimos la distancia entre dos de sus elementos como

$$d\left((z, w), (z', w')\right) = \sqrt{|z - z'|^2 + |w - w'|^2}$$

Conservando las notaciones precedentes, una aplicación del teorema 3.1 es la siguiente:

Problema 6. *Si $\mathcal{H}$ es un subgrupo de $\mathcal{C}^2$, entonces $(1,1)$ se aproxima arbitrariamente con elementos de $\mathcal{H}$, o $\mathcal{H}$ es finito.*

Poniendo

$$z = \exp(2\pi i\alpha), \quad w = \exp(2\pi i\beta) \quad \text{con} \quad 0 \leq \alpha, \beta < 1$$

$\mathcal{C}^2$ queda parametrizado por

$$[0,1) \times [0,1) \subset \mathbb{R}^2.$$

Si $\mathcal{H}'$ es el subconjunto de $\mathbb{R}^2$ dado por

$$\mathcal{H}' = \left\{(\alpha,\beta) \in \mathbb{R}^2 : (\exp(2\pi i\alpha), \exp(2\pi i\beta)) \in \mathcal{H}\right\}$$

comprobamos sin mayor dificultad que $\mathcal{H}'$ es un subgrupo de $\mathbb{R}^2$ con su estructura aditiva. Si $(1,1)$ no pudiera ser aproximado arbitrariamente con elementos de $\mathcal{H}$, entonces $(0,0)$ no podrá ser aproximado arbitrariamente por elementos de $\mathcal{H}'$, en este caso, del teorema 3.1 se sigue que $\mathcal{H}'$ es discreto y, según la proposición 3.2, sólo puede tener un número finito de puntos en la región acotada $\mathcal{F} = [0,1) \times [0,1)$, luego $\mathcal{H}$ sería finito.

Llamaremos *rango de un subgrupo de* $\mathbb{R}^2$, a la dimensión del subespacio generado por sus elementos. Los subgrupos de rango 1, pueden ser tratados en forma análoga a los de la recta real, es decir, son discretos o densos en la recta que los contiene. Es claro que un subgrupo discreto de rango 1 tendrá la forma

$$\mathbb{Z}v = \{nv : n \in \mathbb{Z}\}$$

para algún vector v no nulo y de longitud mínima entre los vectores del subgrupo.
Un *retículo* en el plano, es un conjunto de la forma

$$\mathbb{Z}v \oplus \mathbb{Z}w = \{nv + mw : n, m \in \mathbb{Z}\}$$

donde v, w son vectores linealmente independientes de $\mathbb{R}^2$.
A continuación daremos una caracterización de los subgrupos discretos de rango 2 en el plano desde un contexto algebraico.

Teorema 3.3. *$\mathcal{H}$ es un subgrupo discreto de $\mathbb{R}^2$ si, y sólo si $\mathcal{H}$ es un retículo.*

Demostración: Sea $\mathcal{H}$ un retículo de $\mathbb{R}^2$ que indicaremos con $\mathbb{Z}v \oplus \mathbb{Z}w$. Es claro que $\mathcal{H}$ es un subgrupo de $\mathbb{R}^2$. Por otra parte, si en el vector $u = nv + mw$ es $m \neq 0$, consideremos l la recta que une el origen con v. Denotando por d a la distancia, tenemos las siguientes desigualdades

$$\|u\| \geq d(u,l) = d(mw,l) = |m|\, d(w,l) \geq d(w,l)$$

Si θ es el ángulo que encierran v y w, resulta

$$d(w,l) = \|w\| \times sen(\theta)$$

Simétricamente, se trata el caso $n \neq 0$ y en conclusión se obtiene que para todo $u \in \mathcal{H}$, $u \neq 0$ tenemos que

$$\|u\| \geq sen(\theta) \times \min\{\|v\|, \|w\|\}$$

Recíprocamente, dado $u \in \mathcal{H}$, $u \neq 0$, se sigue de la proposición 3.2 que el círculo dado por

$$\|x\| \leq \|u\|$$

contiene un número finito de elementos de $\mathcal{H}$. Podemos encontrar entonces, entre los elementos no nulos de $\mathcal{H}$, un vector v cuya longitud sea mínima. Sea entonces $\delta > 0$ definido por

$$\delta = \inf\{\|h\| : h \in \mathcal{H} \text{ y } h \neq 0\} = \|v\|$$

De esto se desprende que la distancia entre dos elementos distintos en $\mathcal{H}$ es mayor o igual que $\delta = \|v\|$.
Sea l la recta que une el origen con v. Es claro que $\mathcal{H} \cap l$ es un subgrupo discreto de l, y por la elección de v, debe ser $\mathcal{H} \cap l = \mathbb{Z}v$.
Dado que $\mathcal{H}$ es de rango 2, existen elementos de $\mathcal{H}$ que no están en l. Fijado $u \in \mathcal{H} - l$, la recta $l + u$ es paralela l y se tiene que

$$\mathcal{H} \cap (l + u) = \mathbb{Z}v + u$$

puesto que no hay puntos distintos en $\mathcal{H}$ que disten en menos que $\delta = \|v\|$. Por la misma razón, resulta que cualquier segmento en $l+u$ cuya longitud sea mayor que δ, debe contener un elemento de $\mathcal{H}$ en su interior, y consecuentemente la recta $l + u$ cortaría al círculo

$$\|x\| \leq \delta$$

en un segmento de longitud menor o igual que δ.
Es decir, que las rectas l y $l+u$ tienen una distancia que, como mínimo, es igual a $\frac{\sqrt{3}}{2}\delta$.
Pongamos

$$\gamma = \inf\{d(l + u, l) : u \in \mathcal{H} - l\}$$

Para dos rectas distintas $l + p$ y $l + q$ tenemos

$$d(l + q, l + p) = d(l + p - q, l) \geq \frac{\sqrt{3}}{2}\delta$$

lo que indica que γ es en realidad un mínimo. Sea $w \in \mathcal{H}$ tal que

$$\gamma = d(l + w, l)$$

Naturalmente que

$$\mathbb{Z}v \oplus \mathbb{Z}w \subseteq \mathcal{H}$$

Dado $u \in \mathcal{H}$ podemos expresar

$$u = \alpha v + \beta w$$

descomponiendo β en su parte entera más su mantisa

$$\beta = [\beta] + \{\beta\}$$

tenemos que

$$\alpha v + \{\beta\} w \in \mathcal{H}$$

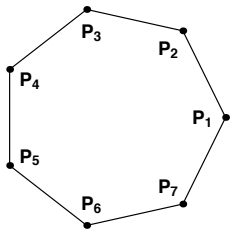
siendo

$$\begin{aligned} d(\alpha v + \{\beta\} w, l) &= d(\{\beta\} w, l) \\ &= \{\beta\} d(w, l) \\ &= \{\beta\} \gamma < \gamma \end{aligned}$$

por la minimalidad de γ, debe ser $\{\beta\} = 0$. Resulta entonces $\alpha v \in \mathcal{H} \cap l$, y con esto, $\alpha \in \mathbb{Z}$, es decir $\mathbb{Z}v \oplus \mathbb{Z}w = \mathcal{H}$. ■

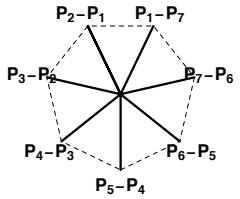
Problema 7. *Supongamos que un polígono regular de n lados tiene todos sus vértices en un retículo de* $\mathbb{R}^2$*. Entonces* $n = 3,\ 4$ *ó* 6*.*

Consideremos primero $n \geq 7$. Si $P_1, P_2, \ldots, P_n$ son los sucesivos vértices del polígono regular sobre un retículo los puntos $P_2 - P_1, P_3 - P_2, \ldots, P_n - P_{n-1}, P_1 -$

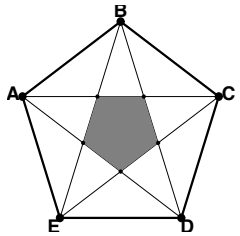


P_n serán también los vértices de un polígono regular sobre el mismo retículo, sólo que más pequeño, ya que el radio de la circunferencia que circunscribe a este último coincide con la longitud del lado del polígono original. Si R y r denotan los radios de las respectivas circunferencias circuns-criptas al primer y segundo polígono, tenemos

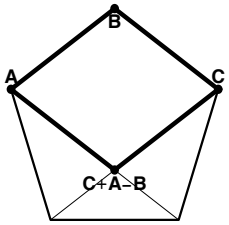
$$r = 2R\, sen\left(\frac{\pi}{n}\right)$$



Iterando este proceso, encontraríamos una sucesión de polígonos re-gulares sobre el retículo que converge a un punto, pero esto contradice el hecho que dos puntos en un retículo deben distar en más que un número $\delta > 0$.
Si $n = 5$ y A, B, C, D, E son los vértices de un pentágono regular sobre un retículo, los puntos $C + A - B$, $D + B - C$, $E + C - D$, $A + D - E$ y $B + E - A$ son los vértices de un pentágono regular sobre el mismo retículo, pero estos



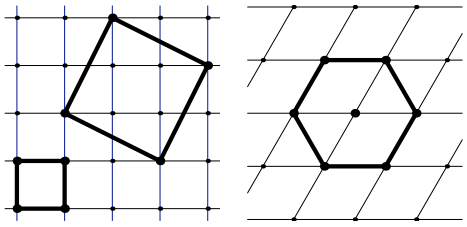
puntos son todos los que se obtienen al intersecar, dos a dos, las diagonales del pentágono A, B, C, D, E, y ahora utilizamos el mismo argumento que en el caso anterior.



Sobre el retículo $\mathbb{Z}^2$ se puede inscribir cuadrados y sobre el retículo

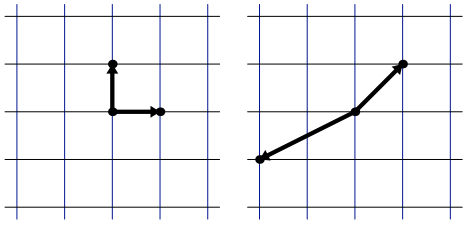
$$\mathbb{Z}\,(1,0) \oplus \mathbb{Z}\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

se puede inscribir hexágonos regulares, y en consecuencia también triángulos equiláteros.



A continuación presentamos una versión en el plano, a la manera de lo expuesto en [3], de dos hechos que pueden ser enunciados con mayor generalidad (ver por ejemplo [4], [5] ó [9]). Estos son el lema de Blichfëldt y el teorema de Minkowski. Particularmente, el teorema de Minkowski es central en el estudio de la geometría de números.

Fijemos un retículo $\mathbb{Z}v \oplus \mathbb{Z}w$ en el plano. En particular, si $v = (1, 0)$ y $w = (0, 1)$ el correspondiente retículo es $\mathbb{Z}^2$. En lo que sigue, nos referiremos a los puntos del retículo como *puntos reticulares*. La si-guiente figura ilustra como un mismo retículo puede ser generado por distintos pares de vectores



El área del paralelogramo con vértices $0, v, w, v + w$ se llama *discriminante del retículo*. Es posible ver que los paralelogramos determinados por un par de vectores que generen el retículo, tienen todos la misma área (ver ejercicio *iii)* al final de estas notas).

Lema 3.4. *(Blichfëldt) Sea $n \geq 0$ un número entero. Una figura $\mathcal{F}$ acotada de área $\delta > n$ puede ser ubicada en el plano de modo que cubra al menos $n + 1$ puntos reticulares en $\mathbb{Z}^2$.*

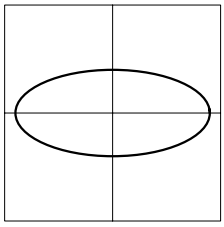
Demostración: Dado $(i,j) \in \mathbb{Z}^2$, consideremos los cuadrados dados por

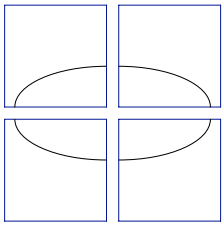
$$\mathcal{C}_{ij} = \{(x,y) : i < x < i+1, j < y < j+1\}$$

El área de $\mathcal{F}$ resulta igual a la suma de las áreas de las figuras dadas por

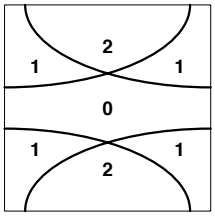
$$\mathcal{F}_{ij} = \begin{cases} (\mathcal{F} \cap \mathcal{C}_{ij}) - (i,j) & \text{si} \quad \mathcal{F} \cap \mathcal{C}_{ij} \neq \emptyset \\ \emptyset & \text{si} \quad \mathcal{F} \cap \mathcal{C}_{ij} = \emptyset \end{cases}$$

Todas las figuras $\mathcal{F}_{ij}$ tienen área menor o igual a 1. Podemos descomponer el





cuadrado $\mathcal{C}_{00}$ en regiones $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_m$ donde $\mathcal{R}_k$ es el conjunto de puntos cubiertos por exactamente k de las $\mathcal{F}_{ij}$. Ahora las regiones $\mathcal{R}_k$ son disjuntas



dos a dos y si $\delta_0, \delta_1, \ldots, \delta_m$ denotan sus respectivas áreas, tenemos

$$\begin{aligned} \delta &= 0 \times \delta_0 + 1 \times \delta_1 + \cdots + m \times \delta_m \\ &\leq m\,(\delta_0 + \delta_1 + \cdots + \delta_m) \leq m \end{aligned}$$

Como $\delta > n$, se sigue que existe un punto (a, b) que es cubierto por al menos $n + 1$ de las figuras $\mathcal{F}_{ij}$, para estos pares (i, j), se tiene que los puntos

$$(a, b) + (i, j)$$

son puntos en $\mathcal{F}$. Si trasladamos la figura para que uno de estos puntos quede sobre un punto reticular, entonces, todos los puntos $(a, b) + (i, j)$ indicados anteriormente serán puntos reticulares. ■

Nota: En el enunciado del lema, la hipótesis que la figura considerada sea acotada no es necesaria, se agrega para simplificar la demostración. Es posible mostrar que hay una parte acotada de la figura cuya área es mayor que n.

Una figura $\mathcal{F}$ es *convexa* si para cada par de puntos en $\mathcal{F}$ el segmento que los une está contenido en $\mathcal{F}$.
El segmento que une dos puntos p y q puede parametrizarse como:

$$[p, q] = \{\lambda p + (1 - \lambda)\, q : 0 \leq \lambda \leq 1\}$$

Una figura $\mathcal{F}$ es *simétrica* cuando verifica que; si $p \in \mathcal{F}$, entonces $-p \in \mathcal{F}$.

Teorema 3.5. *(Minkowski) Dado un retículo $\mathcal{R}$ con discriminante Δ, cualquier figura convexa y simétrica cuya área sea mayor que 4Δ, contiene al menos un punto reticular no nulo.*

Demostración: Sea

$$\mathcal{R} = \mathbb{Z}v \oplus \mathbb{Z}w$$

La transformación lineal $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ dada por

$$\varphi\,(\alpha, \beta) = \alpha v + \beta w$$

es un isomorfismo que aplica $\mathbb{Z}^2$ en $\mathcal{R}$ y su jacobiano es precisamente el discriminante de $\mathcal{R}$, es decir Δ. Sea $\mathcal{G} \subseteq \mathbb{R}^2$ la preimagen de $\mathcal{F}$ a través de φ, entonces

$$\Delta \times |\mathcal{G}| = |\mathcal{F}| > 4\Delta$$

donde con las barras indicamos el área de la figura. Luego

$$|\mathcal{G}| > 4$$

Si definimos

$$\frac{1}{2}\mathcal{G} = \left\{\frac{1}{2}u : u \in \mathcal{G}\right\}$$

resulta

$$\left|\frac{1}{2}\mathcal{G}\right| = \frac{1}{4}\left|\mathcal{G}\right| > 1$$

Por el lema de Blichfëldt, $\frac{1}{2}\mathcal{G}$ puede desplazarse en el plano de modo que cubra al menos dos puntos reticulares en $\mathbb{Z}^2$. En consecuencia, $\frac{1}{2}\mathcal{F}$, la imagen por φ de $\frac{1}{2}\mathcal{G}$, puede desplazarse en el plano de modo que cubra al menos dos puntos reticulares en $\mathcal{R}$. Notemos estos dos puntos como

$$p = \tfrac{1}{2}p_0 + r \quad \text{y} \quad q = \tfrac{1}{2}q_0 + r$$

donde $p, q \in \mathcal{R}$ y $p_0, q_0 \in \mathcal{F}$. Se sigue que

$$0 \neq p - q = \frac{1}{2}p_0 + \frac{1}{2}\left(-q_0\right) \in \mathcal{R}$$

Por ser $\mathcal{F}$ simétrica, $-q_0 \in \mathcal{F}$, y $p - q$ es el punto medio del segmento que une p_0 con $-q_0$, resulta $p - q \in \mathcal{R} \cap \mathcal{F}$ pues $\mathcal{F}$ es convexa. ■

Los problemas que siguen a continuación ilustran aplicaciones del teorema de Minkowski.

Problema 8. *Todo número primo de la forma $4k+1$ es suma de dos cuadrados.*

Si p es un primo de la forma $4k + 1$, es conocido que -1 es residuo cuadrático módulo p. Es decir, existe un entero a tal que $a^2 + 1$ es divisible por p. Una demostración de este hecho puede obtenerse usando el teorema de Wilson que establece que

$$(p-1)! \equiv -1 \quad (\mod p)$$

Por otra parte, si elegimos el sistema de restos $0, \pm 1, \pm 2, \cdots, \pm\frac{p-1}{2}$, tenemos

$$(p-1)! \equiv (-1)^{\frac{p-1}{2}} \left(\left(\frac{p-1}{2}\right)!\right)^2 \quad (\mod p)$$

siendo p de la forma $4k + 1$, resulta $(p-1)!$ un residuo cuadrático.
Consideremos el retículo $\mathcal{R}$ dado por

$$\mathbb{Z}\left(p, 0\right) \oplus \mathbb{Z}\left(a, 1\right)$$

donde $a \in \mathbb{Z}$ es tal que $a^2 + 1$ es divisible por p. El discriminante de este retículo es p. Si $\mathcal{C}$ es el círculo dado por

$$\mathcal{C} = \left\{(x, y) : x^2 + y^2 < 2p\right\}$$

el área de $\mathcal{C}$ es

$$2p\pi > 4p$$

Por el teorema de Minkowski, $\mathcal{C}$ contiene un punto reticular no nulo, es decir, existe $(n, m) \neq (0, 0)$ tal que

$$0 < (np + ma)^2 + m^2 < 2p$$

Pero

$$(np + ma)^2 + m^2 \equiv m^2 \left(a^2 + 1\right) \equiv 0 \quad (\mod p)$$

De aquí que

$$p = (np + ma)^2 + m^2$$

Observación: En forma similar al problema anterior, a partir de la versión general del teorema de Minkowski (ver [4], [5], [6] ó [9]), se puede probar el *teorema de Langrange de los cuatro cuadrados* que afirma que todo número natural es suma de cuatro cuadrados, por ejemplo

$$\begin{aligned} 1 &= 1^2 + 0^2 + 0^2 + 0^2 \\ 7 &= 2^2 + 1^2 + 1^2 + 1^2 \\ 30 &= 5^2 + 2^2 + 1^2 + 0^2 \end{aligned}$$

Este teorema, también conocido como la conjetura de Bachet, fue probado por Lagrange en 1770. Usando propiedades básicas de los números cuaterniónicos, el problema puede reducirse a ver que todo número primo positivo p es suma de cuatro cuadrados. A tal fin será necesario además establecer que -1 es suma de dos cuadrados, módulo p (ver el ejercicio *xv)* al final de estas notas).
Asociados con las descomposiciones de un número natural en suma de cuadrados podemos mencionar los siguientes teoremas debidos a Jacobi, (ver [1] ó [7]).

Sea n un número natural.
El número de pares enteros (p, q) tales que $p^2 + q^2 = n$ es igual a 4 veces la diferencia entre el número de divisores de n congruentes con 1 módulo 4 y el número de divisores de n congruentes con 3 módulo 4.
El número de cuaternas enteras (p, q, r, s) tales que $p^2 + q^2 + r^2 + s^2 = n$ es igual a 8 veces la suma de todos los divisores de n que no son congruentes con 0 módulo 4.

Volviendo al teorema de Minkowski, si la figura considerada en él es además compacta, o sea cerrada y acotada, la condición sobre el área puede ser debilitada como se muestra a continuación.

Proposición 3.6. *Dado un retículo $\mathcal{R}$ con discriminante Δ, cualquier figura compacta convexa y simétrica cuya área sea mayor o igual que 4Δ, contiene al menos un punto reticular no nulo.*

Demostración: Sea $\mathcal{F}$ la figura considerada. Para $\lambda > 1$ considere-mos la figura

$$\mathcal{F}_\lambda = \{\lambda v : v \in \mathcal{F}\}$$

Es claro que $\mathcal{F}_\lambda$ es convexa y simétrica, además

$$|\mathcal{F}_\lambda| = \lambda^2 \, |\mathcal{F}| > |\mathcal{F}| \geq 4\Delta$$

Por el teorema de Minkowski, $\mathcal{F}_\lambda$ contiene un punto reticular no nulo.
Consideremos la sucesión de figuras dadas por

$$\mathcal{F}_{1+\frac{1}{n}} \qquad n \geq 1$$

Supongamos que $0 = (0,0)$ sea el único punto reticular en $\mathcal{F}$. En la región dada por

$$\mathcal{F}_2 - \mathcal{F}$$

existe un conjunto finito $v_1, v_2, \ldots, v_k$ de puntos reticulares.
Dado que el área de $\mathcal{F}$ es mayor que cero, $\mathcal{F}$ no puede estar contenida en una recta, en consecuencia $\mathcal{F}$ contiene dos vectores u y v que son linealmente independientes, luego, por ser $\mathcal{F}$ convexa y simétrica, el paralelogramo con vértices $\pm u, \pm v$ está incluído en $\mathcal{F}$. De este hecho se sigue que si l es una recta que pasa por el origen, $l \cap \mathcal{F} \neq \{0\}$ y por ser esta intersección un subconjunto simétrico, convexo, cerrado y acotado de l, se tiene que existe un vector $w \neq 0$ en l tal que

$$l \cap \mathcal{F} = [-w, w]$$

En particular, para las rectas $l_i = \mathbb{R}v_i$, $(1 \leq i \leq k)$, encontraremos escalares λ_i con $0 < \lambda_i < 1$ y tales que

$$l_i \cap \mathcal{F} = [-\lambda_i v_i, \lambda_i v_i] \quad \text{y} \quad \lambda v_i \notin \mathcal{F} \quad \text{si} \quad \lambda > \lambda_i$$

Si elegimos $n \in \mathbb{N}$ tal que

$$1 + \frac{1}{n} < \frac{1}{\lambda_i}, \ \forall i$$

encontramos que

$$v_i \notin \mathcal{F}_{1+\frac{1}{n}}, \ \forall i$$

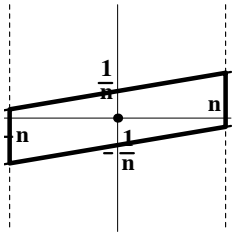
pues

$$\left\| \left(1 + \frac{1}{n}\right) \lambda_i v_i \right\| = \left(1 + \frac{1}{n}\right) \lambda_i \, \|v_i\| < \|v_i\|$$

Resulta entonces que 0 es el único punto reticular en $\mathcal{F}_{1+\frac{1}{n}}$, pero el área de $\mathcal{F}_{1+\frac{1}{n}}$ es mayor que 4Δ, y esto contradice el teorema de Minkowski. En consecuencia $\mathcal{F}$ contiene un punto reticular distinto de 0. ■

Problema 9. *Dados un número real α y un número entero n existen números enteros p y q tales que $0 < q \leq n$ y*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{qn}$$



Consideremos la figura $\mathcal{P}$ dada por el interior del paralelogramo cerrado limitado por las rectas

$$y - \alpha x = \tfrac{1}{n}, \quad y - \alpha x = -\tfrac{1}{n}, \quad x = n \quad \text{y} \quad x = -n$$

$\mathcal{P}$ es una figura convexa y simétrica y su área es 4. Por la proposición 3.6 existe un par $(q, p) \neq (0, 0)$ en el retículo $\mathbb{Z}^2$ tal que

$$p - \alpha q \geq \tfrac{1}{n}, \quad p - \alpha q \geq -\tfrac{1}{n}, \quad q \leq n \quad \text{y} \quad q \geq -n$$

o sea

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{n\,|q|}$$

Para concluir, observemos que, teniendo en cuenta la simetría de $\mathcal{F}$, q puede elegirse positivo.

Problema 10. *Mostrar que si $d \in \mathbb{N}$ y d no es un cuadrado perfecto, entonces la ecuación*

$$x^2 - dy^2 = 1$$

tiene infinitas soluciones enteras.

Supongamos que el par (m, n) sea una solución entera de la ecuación distinta de $(\pm 1, 0)$ y sin pérdida de generalidad asumamos que $m > 0$. Entonces descomponemos

$$m^2 - n^2 d = \left(m + n\sqrt{d}\right)\left(m - n\sqrt{d}\right) = 1$$

y si $k \in \mathbb{N}$, tenemos que

$$\left(m + n\sqrt{d}\right)^k \left(m - n\sqrt{d}\right)^k = 1$$

Pero entonces existen enteros m_k y n_k tales que

$$\left(m + n\sqrt{d}\right)^k = m_k + n_k\sqrt{d} \quad \text{y} \quad \left(m - n\sqrt{d}\right)^k = m_k - n_k\sqrt{d}$$

de donde

$$m_k^2 - n_k^2 d = 1$$

La sucesión (m_k, n_k) está dada por la ley recursiva

$$\begin{bmatrix} m_{k+1} & n_{k+1} \end{bmatrix} = \begin{bmatrix} m_k & n_k \end{bmatrix} \begin{bmatrix} m & n \\ dn & m \end{bmatrix}$$

Si notamos con A a la matriz

$$A = \begin{bmatrix} m & n \\ dn & m \end{bmatrix}$$

se tiene que A es inversible con determinante igual a 1 y la sucesión puede reescribirse como

$$\begin{bmatrix} m & n \end{bmatrix} A^k \quad \text{con} \quad k \geq 0$$

Ahora, si para valores dos distintos de k los correspondientes elementos de esta suceción coincidieran, podríamos simplificar la identidad a una expresión del tipo

$$\begin{bmatrix} m & n \end{bmatrix} A^j = \begin{bmatrix} m & n \end{bmatrix}$$

para algún entero $j > 0$. Esto significa que 1 debe ser valor propio de A^j, pero siendo los valores propios de A iguales a

$$m + \sqrt{m^2 - 1} \quad \text{y} \quad m - \sqrt{m^2 - 1}$$

es decir el primero mayor que 1 y el segundo menor que 1, los valores propios de A^j son precisamente

$$\left(m + \sqrt{m^2 - 1}\right)^j \quad \text{y} \quad \left(m - \sqrt{m^2 - 1}\right)^j$$

siendo el primero mayor que 1 y el segundo menor que 1, lo que contradice la condición que 1 sea valor propio de A^j, luego la sucesión no tiene términos repetidos.
Resta ver que hay al menos una solución distinta de $(\pm 1, 0)$.
Usando el resultado del problema 9, para $n = 1$, existe un par $(q_0, p_0) \in \mathbb{Z}^2$ tal que

$$\left| q_0\sqrt{d} - p_0 \right| < 1 \quad \text{y} \quad 0 < q_0 \leq 1$$

es claro que q_0 es igual a 1 y p_0 es la parte entera de $\sqrt{d}$. Elijamos ahora $n_1 \in \mathbb{N}$ tal que

$$\frac{1}{n_1} < \left| q_0\sqrt{d} - p_0 \right|$$

y nuevamente usando el problema 9, tomemos un par $(q_1, p_1) \in \mathbb{Z}^2$ tal que

$$\left| q_1\sqrt{d} - p_1 \right| < \tfrac{1}{n_1} \quad \text{y} \quad 0 < q_1 \leq n_1$$

Ahora fijamos $n_2 \in \mathbb{N}$ tal que

$$\frac{1}{n_2} < \left| q_1\sqrt{d} - p_1 \right|$$

y elegimos $(q_2, p_2) \in \mathbb{Z}^2$ tal que

$$\left| q_2\sqrt{d} - p_2 \right| < \tfrac{1}{n_2} \quad \text{y} \quad 0 < q_2 \leq n_2$$

continuando de esta manera obtenemos una sucesión $(q_i, p_i) \in \mathbb{Z}^2$ y una sucesión $n_i \in \mathbb{N}$ tales que

$$\tfrac{1}{n_{i+1}} < \left| q_i\sqrt{d} - p_i \right| \qquad \left| q_i\sqrt{d} - p_i \right| < \tfrac{1}{n_i} \qquad 0 < q_i \leq n_i$$

De estas desigualdades, encontramos que

$$|p_i| < q_i\sqrt{d} + \frac{1}{n_i} < n_i\sqrt{d} + 1$$

y luego

$$\left| q_i^2 d - p_i^2 \right| = \left| q_i\sqrt{d} - p_i \right| \left| q_i\sqrt{d} + p_i \right| < \frac{1}{n_i}\left(2n_i\sqrt{d} + 1\right) < 2\sqrt{d} + 1$$

Por otra parte, en la sucesión (q_i, p_i) todos los pares son distintos entre sí dado que los valores

$$\left| q_i\sqrt{d} - p_i \right|$$

forman una sucesión estrictamente decreciente. Como la sucesión de enteros $p_i^2 - q_i^2 d$ está acotada, debe haber una cantidad infinita de pares (q_i, p_i) tal que

$$p_i^2 - q_i^2 d = k$$

para algún número entero $k \neq 0$. Entonces podemos elegir dos soluciones distintas (α, β) y (γ, δ) de la ecuación anterior tales que

$$\alpha \equiv \gamma \pmod{k} \quad \text{y} \quad \beta \equiv \delta \pmod{k}$$

Si denotamos

$$\begin{aligned} \zeta &= \left(\alpha + \beta\sqrt{d}\right)\left(\gamma - \delta\sqrt{d}\right) \\ &= (\alpha\gamma - \beta\delta d) + (\beta\gamma - \alpha\delta)\sqrt{d} \\ &= p + q\sqrt{d} \end{aligned}$$

tenemos que

$$\begin{aligned} p &= \alpha\gamma - \beta\delta d \equiv \alpha^2 - \beta^2 d = 0 \pmod{k} \\ q &= \beta\gamma - \alpha\delta \equiv \beta\alpha - \alpha\beta = 0 \pmod{k} \end{aligned}$$

luego existen enteros m y n tales que

$$p = km \quad \text{y} \quad q = kn$$

y resulta

$$\begin{aligned} m^2 - n^2\sqrt{d} &= \left(m + n\sqrt{d}\right)\left(m - n\sqrt{d}\right) \\ &= \frac{1}{k^2}\left(p + q\sqrt{d}\right)\left(p - q\sqrt{d}\right) \\ &= \frac{1}{k^2}\left(\alpha + \beta\sqrt{d}\right)\left(\gamma - \delta\sqrt{d}\right)\left(\alpha - \beta\sqrt{d}\right)\left(\gamma + \delta\sqrt{d}\right) \\ &= \frac{1}{k^2}\left(\alpha^2 - \beta^2\sqrt{d}\right)\left(\gamma^2 - \delta^2\sqrt{d}\right) \\ &= 1 \end{aligned}$$

La ecuación $x^2 - dy^2 = k$ es conocida como *la ecuación de Pell* y fue tratada por Lagrange usando fracciones continuas (ver [8]).

Finalizamos estas notas incluyendo en ellas las definiciones formales, en el espacio $\mathbb{R}^n$, de algunos de los conceptos utilizados hasta aqui.

Consideremos $\mathbb{R}^n$ provisto con la métrica usual. Para $x, y \in \mathbb{R}^n$, con $d(x, y)$ denotaremos la distancia euclídea entre x e y dada por

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Dados $\mathcal{A}$ y $\mathcal{B}$, con $\mathcal{A} \subseteq \mathcal{B}$, dos subconjuntos de $\mathbb{R}^n$, decimos que $\mathcal{A}$ *es denso en* $\mathcal{B}$ si dados $b \in \mathcal{B}$ y $\varepsilon > 0$ y , existe $a \in \mathcal{A}$ tal que $d(a,b) < \varepsilon$.
Un subconjunto $\mathcal{A}$ de $\mathbb{R}^n$ se dice *discreto* si dado $a \in \mathcal{A}$, existe $\varepsilon > 0$ tal que

$$\mathcal{A} \cap \{x \in \mathbb{R}^n : d(x,a) < \varepsilon\} = \{a\}$$

$\mathbb{R}^n$ con la suma usual es un grupo abeliano. Un subgrupo G de $\mathbb{R}^n$ se dirá *subgrupo denso* o *subgrupo discreto* si G es un conjunto denso o si es un conjunto discreto. El *rango de un subgrupo* es la dimensión del subespacio generado por sus elementos.
Finalmente, *un retículo* es un subgrupo de rango n de la forma

$$\mathcal{R} = \mathbb{Z}v_1 \oplus \mathbb{Z}v_2 \oplus \cdots \oplus \mathbb{Z}v_n \qquad (v_i \in \mathbb{R}^n)$$

siendo su *discriminante* el valor absoluto del determinante de la matriz que tiene por filas a los vectores $v_1, v_2, \ldots, v_n$. Los resultados vistos en el plano se extienden en el ejercicio *iv)* y la consistencia de la definición del discriminante se plantea en el ejercicio *iii)*.

4 Ejercicios propuestos

i) Sea $\mathcal{H}$ el subgrupo de $\mathbb{R}$ dado por

$$\mathcal{H} = \left\{ \frac{2n}{3} + \frac{4m}{5} + \frac{6k}{7} : n, m, k \in \mathbb{Z} \right\}$$

Decidir si $\mathcal{H}$ es denso o discreto, de ser discreto expresarlo en la forma $\mathbb{Z}r$.
ii) Dados los números racionales $q_1, q_2, \ldots, q_m$ y el subgrupo de $\mathbb{R}$ definido por

$$\mathcal{H} = \{n_1 q_1 + n_2 q_2 + \cdots + n_m q_m : n_i \in \mathbb{Z}\}$$

Decidir si $\mathcal{H}$ es denso o discreto.
iii) Mostrar que el discriminante de un retículo no depende de la base que lo defina.
iv) Considerando $\mathbb{R}^n$ con la suma de vectores y la norma euclídea, generalizar la proposición 3.2 y el teorema 3.3.
v) Mostrar que la intersección de dos subgrupos es un grupo discreto si uno de ellos lo es.
vi) ¿Qué polígonos regulares pueden inscribirse en el retículo $\mathbb{Z}^2$?
vii) Si $\mathcal{H}$ es un subgrupo discreto del plano, mostrar que en un círculo dado contiene a lo sumo un número finito de puntos de $\mathcal{H}$. ¿Es cierta esta afirmación si $\mathcal{H}$ no es subgrupo?

viii) Si $\mathcal{H}$ es un subgrupo discreto del plano, mostrar que existen vectores u y v en $\mathbb{R}^2$, tales que $\mathcal{H} = \{nu + mv : n, m \in \mathbb{Z}\}$.
ix) Si $\mathcal{L}$ es una recta por el origen $\mathbb{R}^2$, un subgrupo de $\mathcal{L}$ es discreto o denso. Si se proyectan los puntos de coordenadas enteras sobre $\mathcal{L}$ se obtiene un subgrupo de $\mathcal{L}$, ¿en qué casos es este subgrupo discreto y en qué casos es denso?
x) Si las asíntotas de una hipérbola contienen cada una al menos dos puntos de coordenadas enteras, mostrar que en la hipérbola hay a lo sumo un número finito de puntos con coordenadas enteras.
xi) ¿Es cierto que si una de las asíntotas de una hipérbola pasa por dos puntos de coordenadas enteras, entonces ocurre lo mismo con la otra asíntota?.
xii) Mostrar que la ecuación

$$ax^2 + 2bxy + cy^2 = 1$$

tiene un número finito de soluciones enteras si a, b son enteros y existe un número natural n tal que $n^2 = b^2 - ac$.
xiii) Hallar las soluciones enteras de la ecuación

$$xy + 3x - 5y = 75$$

xiv) Sea

$$Q(x, y) = ax^2 + 2bxy + cy^2$$

una forma cuadrática positiva definida. Demostrar que

$$\min_{n,m \in \mathbb{Z}} \{Q(m, n) \neq 0\} \leq \frac{2}{\sqrt{3}} \sqrt{ac - b^2}$$

xv) Usando la versión general del teorema de Minkowski (ver [4], [5] ó [9]) es posible probar el teorema de los cuatro cuadrados:
a) Si dos números naturales son suma de cuatro cuadrados, probar que el producto de estos es suma de cuatro cuadrados.
b) Dado un número primo p, probar que la ecuación de congruencias

$$m^2 + n^2 \equiv -1 \quad (\mod p)$$

admite solución (ver por ejemplo [2] ó [5]).
c) Sea p un número primo y sean m y n soluciones de la ecuación en *b)*. Considerando el retículo de $\mathbb{R}^4$

$$\mathcal{R} = \mathbb{Z}(p, 0, 0, 0) \oplus \mathbb{Z}(0, p, 0, 0) \oplus \mathbb{Z}(m, n, 1, 0) \oplus \mathbb{Z}(m, -n, 0, 1)$$

1) Mostrar que el discriminante de $\mathcal{R}$ es p^2.
2) Probar que si $(x, y, z, t) \in \mathcal{R}$ entonces $x^2 + y^2 + z^2 + t^2 \equiv 0 \quad (\mod p)$.
3) Aplicar el teorema de Minkowski teniendo en cuenta la esfera centrada en el origen cuyo radio es $2\sqrt{p}$.

xvi) Sea z un complejo unitario, $\mathcal{H} = \{z^n : n \in \mathbb{N}\}$. Probar que $\mathcal{H}$ es denso en los complejos unitarios ó z es una raíz de la unidad .
xvii) ¿Es denso $\left\{\frac{n\,\pi}{m} : n, m \in \mathbb{N}\right\}$ en los reales positivos?
xviii) Sean $z_1, ..., z_n$ complejos unitarios. Mostrar que dado $\varepsilon > 0$, existe un número natural n tal que $|z_i^n - 1| < \varepsilon\ \forall i = 1, 2, ..., n$.
xix) El conjunto de todas las raíces de la unidad ¿es denso en los complejos unitarios?
xx) Sea $f(x)$ un polinomio mónico con coeficientes enteros tal que sus raíces son complejos unitarios. Probar que las raíces de $f(x)$ son raíces de la unidad. (Sugerencia: usar *xviii)*) y el siguiente hecho: si $z_1, ..., z_n$ son las raíces de $f(x)$ entonces para $m \in \mathbb{N}$ el polinomio

$$g\,(x) = \prod (x - z_i^m)$$

es mónico y con coeficientes enteros.

Referencias

[1] Andrews, G., Ekhad, S., Zeilberger, D., *A short proof of Jacobi's formula for the number of representations of an integer as the sum of four squares.* American Mathematical Monthly **100**, 1993, 274-276.
[2] Araujo, J.O., Fernández, L. B., *Contando con Sumas de Gauss.* Divulgaciones Matemáticas, vol. 12, N^{o}2, 2004, 171-180.
[3] De Guzmán, M., *Mirar y Ver.* Red Olímpica, 1993.
[4] Hardy, G., Wright, E., *An introduction to the theory of numbers.* 5^{o}edición, Oxford, 1979.
[5] Ivorra Castillo, C.,*Teoría de Números*, 2004. Url: www.uv.es/˜ivorra/ Libros/ Numeros.pdf
[6] Jacobi, C. G. J., *Note sur la décomposition d 'un nombre donné en quatre carrés.* J. Reine Angew. Math. **3** (1828)., 191. Werke, vol.I, 247.
[7] Lagrange, J. L., Nouveau Mém. Acad. Roy. Sci. Berlin (1772), 123-133; *Oevres*, vol. 3, 189-201.
[8] Le Veque, W.J., *Teoria Elemental de los Números.* Herreros Hnos. México. 1968.
[9] Narkiewicz, W., *Number Theory.* World Scientific Publishing Co. 1983.

Fac. de Ciencias Exactas, UNICEN,
Tandil, 7000 Buenos Aires, Argentina
araujo@exa.unicen.edu.ar, lfernan@exa.unicen.edu.ar

INFORMACIÓN INTERNACIONAL

La Esquina Olímpica

Rafael Sánchez Lamoneda

En esta oportunidad reseñaremos la actividad olímpica del segundo semestre del año 2006. Se destacan los siguientes eventos, la reunión anual del Canguro Matemático en Barcelona, España, a la cual asistimos los profesores Henry Martínez, José Nieto y Rafael Sánchez. La reunión se desarrolló entre los días 11 y 15 de Octubre en el Institut d'Estudis Catalans y contó con la presencia de representantes de 37 países. La Olimpiada Iberoamericana de Matemáticas, OIM que se realizó en Guayaquil, Ecuador, del 23 al 30 de Septiembre. La delegación venezolana estuvo integrada por:

Andrés Guzmán. Caracas. Medalla de Bronce.

Sofía Taylor. Caracas. Mención Honorífica.

Rafael Guédez. Maracaibo. Mención Honorífica.

Gilberto Urdaneta. Maracaibo. Mención Honorífica.

Silvina María de Jesús. Tutor de delegación.UPEL-IPC.

Henry Martínez. Jefe de delegación. UPEL-IPC.

Finalmente el último evento del año fue la Olimpiada Iberoamericana de Matemáticas Universitaria,OIMU, la cual se llevó a efecto el día 18 de Noviembre, a la misma fueron invitados a participar estudiantes de la UCV, USB, UPEL, LUZ, UC, ULA y UNIMET, aunque solo presentaron alumnos de las universidades UCV, USB y UC.

Como siempre finalizamos con algunos de los problemas de las competencias a las cuales asistimos. En esta oportunidad les mostramos los exámenes de la OIM 2006. Cada examen tiene una duración de cuatro horas y media y el valor de cada problema es de 7 puntos.

21^{a} Olimpiada Iberoamericana de Matemáticas

Primer día

Guayaquil, 26 de Septiembre de 2006

Problema 1. En el triángulo escaleno ABC, con $\angle BAC = 90^{\text{o}}$, se consideran las circunferencias inscrita y circunscrita. La recta tangente en A a la circunferencia circunscrita corta a la recta BC en M. Sean S y R los puntos de tangencia de la circunferencia inscrita con los catetos AC y AB, respectivamente. La recta

RS corta a la recta BC en N. Las rectas AM y SR se cortan en U. Demuestre que el triángulo UMN es isósceles.

Problema 2. Se consideran n números reales $a_1, a_2, \ldots, a_n$, no necesariamente distintos. Sea d igual a la diferencia entre el mayor y el menor de ellos y sea $s = \sum_{i<j} |a_i - a_j|$. Demuestre que

$$(n-1)d \leq s \leq \frac{n^2 d}{4}$$

y determine las condiciones que deben cumplir estos n números para que se verifique cada una de las igualdades.

Problema 3. Los números $1, 2, 3, \ldots, n^2$ se colocan en las casillas de una cuadrícula de $n \times n$, en algún orden, un número por casilla. Una ficha se encuentra inicialmente en la casilla con el número n^2. En cada paso, la ficha puede avanzar a cualquiera de las casillas que comparten un lado con la casilla donde se encuentra. Primero, la ficha viaja a la casilla con el número 1, y para ello toma uno de los caminos más cortos (con menos pasos) entre la casilla con el número n^2 y la casilla con el número 1. Desde la casilla con el número 1 viaja a la casilla con el número 2, desde allí a la casilla con el número 3, y así sucesivamente, hasta que regresa a la casilla inicial, tomando en cada uno de sus viajes el camino más corto. El recorrido completo le toma a la ficha N pasos. Determine el menor y el mayor valor posible de N.

Segundo día

Guayaquil, 27 de Septiembre de 2006

Problema 4. Determine todas las parejas (a, b) de enteros positivos tales que $2a + 1$ y $2b - 1$ sean primos relativos y $a + b$ divida a $4ab + 1$.

Problema 5. Dada una circunferencia Γ, considere un cuadrilátero $ABCD$ con sus cuatro lados tangentes a Γ, con AD tangente a Γ en P y CD tangente a Γ en Q. Sean X e Y los puntos donde BD corta a Γ, y M el punto medio de XY. Demuestre que $\angle AMP = \angle CMQ$.

Problema 6. Sea $n > 1$ un entero impar. Sean P_0 y P_1 dos vértices consecutivos de un polígono regular de n lados. Para cada $k \geq 2$, se define P_k como el vértice del polígono dado que se encuentra en la mediatriz de P_{k-1} y P_{k-2}. Determine para qué valores de n la sucesión $P_0, P_1, P_2, \ldots$ recorre todos los vértices del polígono.

ESCUELA DE MATEMÁTICAS
UNIVERSIDAD CENTRAL DE VENEZUELA, VENEZUELA
rsanchez@euler.ciens.ucv.ve

AGRADECIMIENTO

Agradecemos la colaboración prestada por las siguientes personas en el trabajo editorial del volúmen XIII del Boletín de la AMV:

Begoña Cano, Juan Cuesta, Félix Delgado, Paul Doukhan, Eva A. Gallardo, Sabrina Garbin, Michael Katz, Gabor Lugosi, Domingo Morales, Jaroslav Nešetřil, Ragnar Norberg, Henryk Gzyl, Philip Feinsilver, Ramón Puigjaner, Martin Leitz-Martini, Ragnar Norberg, Javier Peralta, Ramón Pino, Carlos Uzcátegui, José Vielma, Minaya Villasana

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 responsability 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