REVISTA
DE LA

ACADEMIA DE CIENCIAS

Exactas
Físicas
Químicas y
Naturales
DE
ZARAGOZA

Serie 2.ª
Volumen 56

2001

ÍNDICE DE MATERIAS

Rev. Academia de Ciencias. Zaragoza. **56** *(20001)*

Cianotoxinas y el metabolismo del hierro *

M. Teresa Bes, José A. Hernández y María F. Fillat

Departamento de Bioquímica y Biología Molecular y Celular

Facultad de Ciencias, Universidad de Zaragoza. 50009-Zaragoza, Spain

* Premio de la Academia a la investigación (2000–01)

Resumen

El hierro es un elemento esencial para todos los seres vivos. Aunque es muy abundante, paradójicamente es limitante en muchos casos, dada su escasa solubilidad en medios acuáticos. Además, los organismos deben controlar muy finamente su incorporación ya que puede generar radicales libres. La proteína Fur (ferric uptake regulation) es un represor que en presencia de hierro regula un gran número de genes en bacterias y cianobacterias. Algunas estirpes de cianobacterias, en condiciones no bien determinadas todavía, producen sustancias tóxicas de naturaleza neurotóxica o hepatotóxica, llamadas cianotoxinas. Las proliferaciones incontroladas de algas y la producción de cianotoxinas en agua potable constituyen un grave problema sanitario y medioambiental. Al igual que ocurre en bacterias patógenas, en que la expresión de algunos factores de virulencia y toxinas está mediada por *Fur*, se ha propuesto que esto podría ocurrir en el caso de expresión de cianotoxinas. En septiembre del año 2000, se dio en La Estanca de Alcañiz (Teruel, España) una proliferación incontrolada de *Microcystis* con producción de microcistina. En este trabajo se estudian muestras recogidas en las circunstancias citadas y se demuestra la presencia de una proteína con reactividad cruzada con anticuerpos anti Fur de otra cianobacteria, *Anabaena* PCC 7119.

Abstract

Iron availability limits cell division rates, abundance and production of phytoplancton. Iron nutritional status of the phytoplankton can be related with blooms and toxin production. Fur (ferric uptake protein) is a bacterial regulator which in presence of iron represses not only the expression of genes related to iron acquisition but also to other metabolic pathways. Fur has been described in cyanobacteria, but the role of Fur in photosynthetic organisms has not been characterised yet, even thought *fur* regulon can be one of the main regulation systems. Toxin production is

a mechanism developed by cells to face up different stress situations. In pathogenic bacteria, Fur is a key regulatory element in the expression of toxins in response to iron stress. Cyanobacteria produce toxins, called cyanotoxins. They constitute an increasing problem due to the eutrophization of the continental waters. The factors leading cyanobacteria to cyanotoxin production are not yet well established, but it has been suggested linkage between iron metabolism and blooms and cyanotoxin production. In this paper we study a *Microcystis* bloom with microcystin production in Alcañiz (Teruel, Spain), in September 2000. Samples from the lagoon "La Estanca", showed the presence of bands cross-reactive with the Fur antibodies (from *Anabaena* PCC 7119 protein).

1. Introducción

El hierro es un elemento muy abundante en la naturaleza, pero su disponibilidad para microorganismos, algas eucariotas y plantas superiores es muy baja, ya que en su mayor parte se encuentra en forma de hidróxido férrico que es muy insoluble. El hierro es un componente esencial de grupos redox en los seres vivos, y en el caso de los microorganismos, la carencia de hierro activa una serie de genes, cuyos productos permiten hacer frente de una forma eficaz a esta deficiencia, dando lugar a proteínas implicadas en la incorporación del hierro, y otros muchos polipéptidos de diversa naturaleza. Muchos de estos genes, como los que dan lugar a toxinas, tienen un papel importante como mecanismo de defensa a la deficiencia de hierro, ya que matando células u organismos cercanos expuestos a estas toxinas, provocan la liberación de hierro intracelular que pasa a estar disponible, a la vez que eliminan competidores. Las cianobacterias producen en determinadas condiciones toxinas que provocan grandes problemas sanitarios y económicos, tanto en mares y océanos dando lugar a las llamadas mareas tóxicas, como en aguas dulces, en las que pueden originarse graves problemas cuando estas proliferaciones y liberaciones de toxinas se producen en aguas para uso urbano o ganadero. En las aguas dulces, la presencia de cianotoxinas se está convirtiendo en un problema grave, debido a la creciente eutrofización de nuestros acuíferos. Este problema causa una gran alarma social (Heraldo de Aragón, 23 y 26 de septiembre, 2000) Las toxinas producidas por las cianobacterias son fundamentalmente hepatotóxicas y neurotóxicas, y no solo producen daños irreversibles en los organismos afectados, sino que bajas dosis de algunas de ellas en aguas de boca se han relacionado recientemente con alta incidencia de algunos tumores. *Microcystis* es una de las cianobacterias con alta capacidad de producir toxinas (fundamentalmente microcistinas), y que más problemas está causando (Carmichael 1997). Las microcistinas son unas hepatotoxinas constituidas por un heptapéptido cíclico, con 5 aminoácidos no proteicos, y dos de los que forman parte de proteínas. Hasta el momento se han identifica-

do del orden de unas 60 microcistinas diferentes, que son inhibidoras de protein-fosfatasas. La síntesis de microcistinas no ocurre constitutivamente y los factores que la inducen son objeto de numerosos estudios. En algún caso se ha propuesto que la deficiencia de hierro podría ser un factor desencadenante de la expresión de microcistinas (Lick et al., 1996).

El elemento clave de la regulación de la expresión de toxinas en bacterias heterotrofas es una proteína de unión a DNA, de unos 17 kDa, que se ha denominado proteína Fur (ferric uptake regulation). Cuando la disponibilidad de hierro es elevada, Fur reprime la expresión de los genes implicados en la incorporación de este elemento. Asímismo, la actividad de Fur está relacionada con procesos tan importantes como la defensa frente al estrés oxidativo, la virulencia de ciertos patógenos como el vibrión del cólera, algunas cepas de *Legionella*, o las bacterias causantes de la meningitis y la peste.

La regulación mediada por Fur tiene lugar a nivel transcripcional. Esta proteína se une a una secuencia consenso en los promotores de los genes diana; los pocos estudios estructurales llevados a cabo indican que Fur carece del típico motivo hélice-giro-hélice, característico de la mayor parte de las proteínas de unión al DNA. Hasta el momento se ha clonado el gen *fur* y se ha purificado la proteína de diversos microorganismos, aunque hasta la fecha el sistema de *E. coli* es el mejor caracterizado.

Fur es el único represor conocido que requiere hierro para activarse, y aunque se han llevado a cabo numerosos estudios sobre la interacción Fur-DNA, hasta el momento no se conoce cómo ocurre, puesto que no se ha determinado la estructura tridimensional de Fur. La resolución de dicha estructura es indispensable para conocer las bases moleculares del mecanismo de regulación de Fur.

Trabajo previo en nuestro laboratorio ha permitido clonar y sobreexpresar las proteínas Fur de *Anabaena* y *Synechococcus*, y se ha purificado a homogeneidad la proteína recombinante de *Anabaena*. A pesar de las analogías que podrían encontrarse entre la expresión de toxinas en bacterias heterótrofas reguladas por Fur, y la presumible implicación de Fur en la expresión de cianotoxinas, la búsqueda bibliográfica ha revelado que no hay información disponible.

En este trabajo se ha identificado una proteína con reacción cruzada con Fur de *Anabaena* en muestras procedentes de "La Estanca" de Alcañiz (Teruel) que contenían *Microcystis*, productoras de microcistina.

2. Materiales y Métodos

Las muestras (5 litros) se tomaron el 17-10-2000 en tres puntos (Figura 1), filtrando agua a través de una capa de gasa de 4 mm de poro. La obtención de células se llevó a cabo en rotor continuo (centrífuga Beckmann J2-21 con un rotor JCF-Z) a 7.000 $\times g$ y

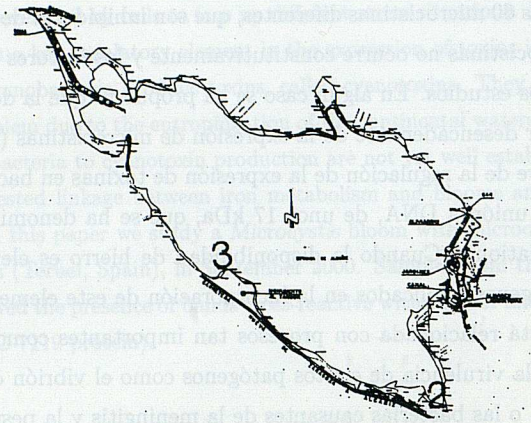

Figura 1.—Mapa de "La Estanca" de Alcañiz, mostrando los puntos de toma de las muestras.

posteriormente una segunda centrifugación en rotor J-14 a 15.000 $\times g$ durante 10 minutos. El pellet se conservó a -20°C hasta su posterior utilización. Una pequeña alicuota fue conservada a $-80°$ C en criotubos con 8% DMSO.

El pH se determinó con un pH-metro Crison micropH2002 y papel indicador (Panreac).

Las muestras se rompieron por sonicación utilizando como tampón de ruptura, 50 mM Tris-acetato pH 8, con 5 mM $MgCl_2$, 1 mM β-mercaptoethanol, 1 mM EDTA y 10 mM PMSF (phenylmethylsulfonyl fluouride). Tras retirar las células no rotas, las muestras fueron preparadas para llevar a cabo SDS-PAGE al 15% según el procedimiento descrito por Laemnli (1970). Las transferencias de Western se llevaron a cabo utilizando filtros de Inmobilon-P, y se trataron según se describe en Pueyo & Gómez-Moreno (1993).

Se clonó y se sobreexpresó el gen *fur* de *Anabaena* PCC 7119 en *E.coli* tal como se describe en Bes *et al.* (2001). La proteína Fur recombinante se ha purificado de acuerdo con el procedimiento descrito por Hernández et al. (2001). Se obtuvieron anticuerpos policlonales según el procedimiento descrito en Hernández *et al.* (2001).

Las proteínas totales fueron determinadas por el método de Lowry (1951). El contenido en clorofila *a* se determinó espectrofotométricamente según el procedimiento descrito por Mackinney (1941), utilizando acetona al 80% y la ecuación 12.7 (A_{663nm})-2.58 (A_{645nm}) = μg cl *a* / ml.

3. Resultados y Discusión

En la figura 1 se detallan los puntos de muestreo, que presentaban el típico aspecto tras una proliferación incontrolada de algas o cianobacterias: masas de células formando remansos en las orillas y en la superficie (Figura 2). Aunque las causas no están bien determinadas,

Figura 2.—Vista de las masas de fitoplancton que cubrían la superficie de "La Estanca", características de una proliferación incontrolada.

estas situaciones suelen darse en aguas quietas, ricas en nutrientes y normalmente en los meses más cálidos del año, en los que las temperaturas y la iluminación son óptimas. La inexistencia de movimiento de capas de agua, da lugar a acumulación de las capas de agua más cálidas en la superficie y la proliferación de estos organismos. La fecha en la que se tomaron las muestras está comprendida en el periodo de máxima proliferación y producción de toxinas (Quesada et *al.* 2000) descrito para otras zonas de España.

El pH de las aguas en el momento de la recolección de muestras fue de 8.2 en la estación de nuestreo 1, de 7 en la estación 2, y de 6.2 en la estación 3. Es de destacar la diferencia de dos unidades de pH entre la estación 1 y la estación 3, aunque podría estar justificado por la proximidad de la estación 1 a zonas con mayor impacto humano. En todos los casos, el pH está dentro del rango (5.5 a 9) permitido para las aguas superficiales de calidad tipo A2 destinadas a la producción de agua potable (WWW.chebro.es). La solubilidad del hierro esta directamente relacionada con el pH, y cuanto más alto es el pH, más baja su solubilidad.

El espectro UV-visible de las células sonicadas (Figura 3) mostró máximos de absorbancia a 667, 622, 437, 386 y 320 nm, no apreciándose presencia de ficobiliproteinas. Uno de los efectos de la deficiencia de hierro en numerosas cianobacterias es el reemplazamiento de la ferredoxina I por flavodoxina, cuyo gen esta regulado por Fur. Un método

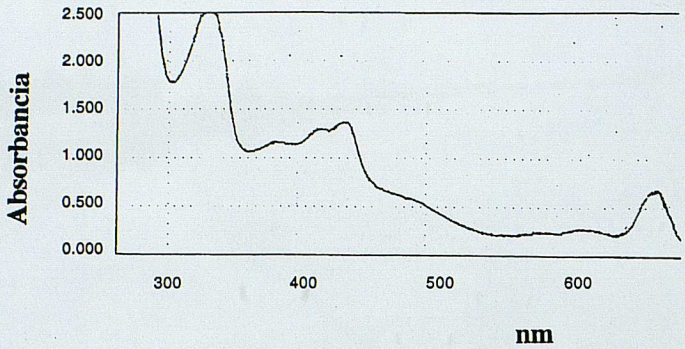

Figura 3.—Espectro UV-visible de extractos crudos de muestras recogidas en "La Estanca" de Alcañiz.

de detección cualitativo de flavodoxina en extractos crudos consiste en la obtención de un espectro diferencial usando como reductor ditionito sódico (M. Fillat, Tesis doctoral, 1988). Un espectro diferencial (no mostrado) del extracto crudo de *Microcystis*, en el que se aprecian dos máximos de absorbancia a 480 y 446 nm, parece indicar la presencia de flavodoxina. Esta observación es coherente con el hecho de que el hierro sería un factor limitante en la situación final de la proliferación, que es la que se realizó el muestreo. Por otra parte, Fur se inactivaría como represor, induciendo la síntesis de sideróforos, y posiblemente de microcistinas. Estas cianotoxinas son péptidos cíclicos que se sintetizan fuera de los ribosomas mediante péptido sintetasas. Se ha propuesto que estos péptidos cíclicos, excretados al exterior de las cianobacterias pudieran actuar como quelantes de hierro pero, hasta el momento, no se ha logrado demostrar.

Para confirmar la presencia en *Microcystis* de una proteína homóloga a Fur, se realizó una electrofóresis en geles de poliacrilamida del 15 % de los extractos crudos, y se transfirió el gel a un filtro de Inmobilon-P, revelado con anticuerpos anti-Fur de otra cianobacteria, *Anabaena* PCC 7119. En la figura 4 se muestra el patrón electroforético obtenido, y la presencia de dos bandas con reactividad cruzada con el anticuerpo de la proteína de *Anabaena*. El peso molecular de la proteína de *Anabaena* es de 17 kDa (carril 4), y las bandas observadas en el caso de las muestras de "La Estanca" es aproximadamente de 28 kDa y 37 kDa. (carril 1) Los pesos estimados son un poco superiores a los descritos para las proteínas Fur conocidas, que oscilan entre 14 y 20 kDa, pero no puede descartarse que otras proteínas Fur pudieran tener mayor peso de los descritos hasta el momento. La presencia de dos bandas pudiera achacarse a que si bien *Microcystis* era la población predominante, se observó visualmente la presencia de otras cianobacterias minoritarias. En la figura se muestra la presencia de Fur en *Anabaena* PCC 7119, crecidas en deficiencia

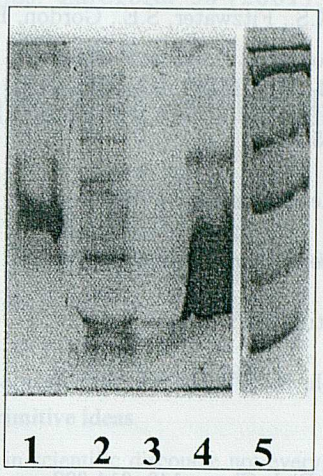

Figura 4.—Transferencia de Western de extractos de muestras recogidas en "La Estanca" (carril 1), *Anabaena* PCC7119 crecidas en deficiencia de hierro (carril 2) y en hierro completo (Carril 3), Fur recombinante pura (carril 4) y marcadores de peso molecular (carril 5).

de hierro (carril 2) y en condiciones óptimas en cuanto a disponibilidad de hierro (carril 3).

Hace unos años se demostró de una forma concluyente, que los niveles de hierro controlan la productividad de los océanos, mediante un experimento de fertilización "in situ" llamado IronEx II. Tras fertilizar con hierro aguas superficiales de una región ecuatorial del Océano Pacífico, se detectó un crecimiento masivo del fitoplancton, demostrando inequívocamente que la productividad estaba limitada por la disponibilidad de hierro (Coale *et al.* 1996). Esta situación se ha propuesto que podría ser extrapolable a aguas continentales, y el regulador Fur podría jugar un papel muy importante en la proliferación de fitoplancton descontrolada. Utkilen y Gjolme (1995) encontraron que la disponibilidad de hierro estimulaba la producción de toxinas en *Microcystis aeruginosa*, por lo que podría pensarse que Fur podría estar regulando la expresión de toxicidad. La proteína Fur, en función de la disponibilidad de hierro en el medio de cultivo, regularía la expresión de dichos genes.

Referencias

[1] Bes, M.T., Hernández, J.A., Peleato, M.L. and Fillat, M.F. FEMS Microbiology Letters. 194: 567-574. 2001

[2] Carmichael, W.W. (1997). Advances in Botanical Research, 27: 211-256.

[3] Coale, K. H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S. Chávez, F.P. Ferioli, L. Sakamoto, C. Rogers, P., Millero, F. Steinberg, P. Nightingale, P., Cooper, D. Cochlan, WP., Landry, M.R, Constaninou, J., Rollwagen, G., Trasvina, A. and Kudela, R. (1996). Nature 383, 495-501.

[4] Heraldo de Aragón, 23 de septiembre de 2000.

[5] Heraldo de Aragón, 26 de septiembre de 2000.

[6] Laemnli U.K. (1970) Nature 227:680-685.

[7] Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.

[8] MacKinney G. (1941) J. Biol. Chem. 140, 314-322.

[9] Pueyo, J.J. y Gómez-Moreno. (1993). Photosynthesis research, 38:35-39.

[10] Quesada, A. Sanchis, D., Carrasco, F., Leganes, E, Fernández-Valiente, E. y Fernández-del-Campo., F. (2000) Actas de "International Conference on Toxic Cyanobacterial blooms". Rabat (Marruecos).

[11] Lynk, S., Gjolme, N., Utkilen, H. (1996). Phycologia 35: 120-124.

Rev. Academia de Ciencias. Zaragoza. **56** *(2001)*

THE CREATION OF PRIMITIVE IDEAS: ITS ROLE IN MATHEMATICS AND THINKING IN GENERAL

F. G. Asenjo
Department of Mathematics
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

I
BACKGROUND

§1. No thought without primitive ideas

In common as well as in scientific discourse not every concept can be defined. These undefined concepts are also called primitive ideas and primitive categories. I shall use these three expressions interchangeably. In mathematics, primitive ideas have their meaning indirectly determined by axioms: points, lines, natural numbers, sets, etc. are not defined, they are made sense of as they occur in successive axioms, which implies that their connotations may never be completed. Axioms are independent primitive statements that give progressive meaning to such undefined concepts. These statements do not have to be proved; each is to be taken as true presumably for a substantial length of time, although, as we now well know, no axiom is eternal.

Primitive ideas are not innate: we are not born with the ideas of one and many, say. Further, their connotation does not stay the same after we acquire them. The meaning of every category changes with the ages, even within the limits of one's life span, indeed even in a day! As axioms can be substituted at will, undefined ideas can incorporate special new connotations. These ideas undergo either subtle or profound transformations. Some may be entirely created anew, suddenly or gradually, as I shall show in the present work. In fact, we are all engaged in this creative effort to one degree or another. There would be no intellectual progress without such effort; thought would be a mere unfolding of fixed, immutable contexts. There is, in effect, a continuous creation of categories – the inevitable outcome of the evolution of the mind and the increased sophistication of scientific experience.

§2. Categoreal systems

Aristotle proposed more than one list of basic categories, two with ten items, then a third with only eight, which seems to indicate that he himself – the first proponent of such lists – did not consider any of them as definitive. Kant's list of categories, which differs substantially from Aristotle's, started a very different trend. More recently, one of the last categoreal systems that have been offered is the one that Alfred North Whitehead presents in his *Process and Reality*.[1] It

contains forty-seven items, but it has no pretension of being complete. In fact, as I just implied, the number of categories that can be created is potentially infinite.

Although influenced by Bradley, Bergson, and Samuel Alexander, Whitehead's system has his own imprint. It is organized into four groups: The Category of the Ultimate, Categories of Existence, Categories of Explanation, and Categoreal Obligations, each with three, eight, twenty-seven, and nine items, respectively. The first group is composed of creativity, one, and many; the second includes actual entities, concrete facts of relatedness, private matters of fact, pure disjunctions of diverse entities, contrasts, etc.; the last two groups consist not of single, compactly worded concepts like the ones mentioned already, but of sentences and paragraphs, the categories themselves being not the sentences or paragraphs but the relations they describe among some of the categories previously listed. I abbreviate three examples from the third group in order to give a taste of this special kind of presentation.

(ix) The actual world is a process, and the process is the becoming of actual entities. Thus actual entities are creatures.

(x) All entities have the potential for being an element in any real concrescence of many entities into one actuality.

(xi) How an actual entity becomes constitutes what the actual entity is. Its being is constituted by its becoming.

These examples sound like axioms, but they are not. There is no other way of describing relations between given primitive ideas than through sentences. These relations, not the sentences themselves, are the categories.

The primitive ideas we think with give us the picture of the world we believe in. Kant's categories undermined realism and led to a century of transcendental idealism in Germany. Whitehead's system, on the other hand, is basically realistic; in addition, he warns against (i) "the trust in language as an adequate expression of propositions," (ii) "the subject-predicate form of expression," (iii) "the Kantian doctrine of the objective world as a theoretical construct from purely subjective experience," (iv) "arbitrary deductions in ex-absurdo arguments," and (v) the centuries-old "belief that logical inconsistencies can indicate anything else than some antecedent errors."[2]

As for (iv) and (v) in particular, Aristotle – and many others after him until today – kept drawing existential conclusions about the real world out of the belief in the sacrosanctity of the logical law of no contradiction,[3] including conclusions that are now negated by physics. In mathematics, intuitionism and constructivism wholeheartedly agree that from a mere contradiction one should not derive the existence of any mathematical object. For very different

reasons, inconsistent mathematics embraces contradictions,[4] and with this approach opens the gate to a new world of useful and fascinating structures.

In connection with his system as a whole, Whitehead also states that "the fundamental ideas in terms of which the scheme is developed presuppose each other so that in isolation they are meaningless. This requirement does not mean that they are definable in terms of each other; it means that what is indefinable in one such notion cannot be abstracted from its relevance to the other notions."[5]

But if all primitive ideas are so intimately interconnected to the point that they not only communicate with one another but transform and conceptually enrich one another, then absolute classification is impossible; every taxonomy becomes a mere approximation, a relative and imperfect division in which sorted-out ideas are really only separated by porous mental membranes through which meaning travels freely.

II

PLAYING WITH PRIMITIVE IDEAS

§3. The transformation of categories

Important as categoreal systems have been in the past, we may have seen the end of them. Nowadays, for example, given the fast and long-range spread of computer science and informatics, the proliferation of primitive ideas is not only unlimited but also constantly changing: categories are introduced only to be disposed of after having fulfilled their purpose. A new program often begins with its own undefined concepts meant to function in a new way in accordance with the program's specific objective. And the less the emphasis on proof, the more do primitive ideas emerge to fill the vacuum: the retreat from deduction means the necessity to be a constant beginner. Not that the past and present role of axiomatics is to be obliterated: merely that we must make room for different – that is, expanding and productive – ways of thinking.

Leaving aside these changes that are taking place before our eyes, we must recognize that even from time immemorial all concepts – especially primitive ideas – undergo uncountable transformations simply by their being placed in different contexts. Like a substance subject to combinations in a given chemical process, when a category functions semantically in a different medium it becomes a different concept. We like to think of categories as fixed points of reference, but this is far from being the case. Even in mathematics we shall never be able to fix the concept of set, given that we cannot know fully all its possible connotations: first, because it is impossible to build a complete axiomatizable set theory, and second, because new kinds of sets

are still being produced. Non-well-founded sets have now been recognized as useful, and in the appropriate context "gluttonous sets" have appeared with some remarkable properties such as the paradoxical one of having the membership of some of their elements being true and false at the same time.[6]

This is the case also regarding categories involved in philosophical postulates. We "can never hope to finally formulate metaphysical first principles. Weakness of insight and deficiencies of language stand in the way inexorably. Words and phrases must be stretched towards a generality foreign to their ordinary usage; and however such elements of language be stabilized as technicalities, they remain metaphors mutely appealing for an imaginative leap. There is no first principle which is in itself unknowable, not to be captured by a flash of insight. But, putting aside the difficulties of language, deficiency in imaginative penetration forbids progress in any form other than that of an asymptotic approach to a scheme of principles only definable in terms of the ideal which they should satisfy."[7]

Today, in view of the inexorable fragmentation that computers have brought to our way of thinking mathematically, we must conclude that axiomatics and deducibility are becoming greatly limited. This inevitably affects the role of primitive ideas in the process. As for "metaphysical principles," they have been notoriously changeable always, and so were the categories on which they were based. Take, for instance, the notion of substance. It was a fundamental concept from Aristotle to Spinoza. Nowadays, however, function – the primary way interaction affects the nature of things – is more fundamental than substance; it actually explains substance as it makes of the latter a complex of functions, i.e., a defined concept, no longer a category.[8] Such an overturn makes the old conceptions of materialism incompatible with the physical world as we know it today. Matter is much more like a biological organism than what we were able to imagine before.

To sum up, primitive ideas are far from being mental bricks with which to build meaningful sentences. The mind has no atoms; instead, primitive ideas come attached to larger fragments of consciousness and can change with vertiginous speed as these fragments keep interacting, often beyond our control.

§4. Complex categories

The mere juxtaposition of two ideas, related or not, changes the meaning of both. This is particularly true when the two ideas are in opposition to one another. The relative weight we give to each item in the pair leads to a special way in which we view things. In a paper titled "One and Many,"[9] I dealt with the issue of how different the world looks as we attach to the relations between these two categories different directions and dynamics. We can think of (i) a

world without many, a monistic world in which differences are unreal; (ii) a world without one in which, in contrast, divisibility is the rule and unities are mere illusions; (iii) a world of many ones, the standard atomistic view based on the belief in the existence of some real atoms that are indivisible and constitute the all-important building blocks of the universe; etc. However, no matter how hard we try to exclude or relegate the companion opposite idea in order to concentrate on one single notion, the companion of such notion still projects its shadow on our thoughts. Many concepts – categories included – cannot help but evoke their opposite. In these cases, we subconsciously think in pairs: we cannot entertain the idea of creativity without placing it against a background of inertial entities. Exclusive creativity would be tantamount to absolute chaos – some things must endure to have an orderly changeable cosmos.

But I want now to move beyond the mere juxtaposition of concepts and the inevitable interactions that such juxtaposition creates in order to deal with a more profound symbiosis of ideas: the emergence of complex notions. One of the most recent examples of this intellectual phenomenon is provided by the mathematical theory of deterministic chaos, whose principal conceptual by-product is the complex category "chaos-and-order." It is essential to understand that this complex is not the mere juxtaposition of order and chaos, two opposite single notions; it is instead a third item that stands on its own, not a unitary synthesis in any sense, but an inseparable multiplicity that is mathematically determined with total precision.

Complex categories have been well known before. Kant talked about multiplicities being taken as unities and called this categoreal complex "totalities." Cantor took this idea from Kant, added Bolzano's treatment of aggregates, and called multiplicities taken as unities "sets." To be sure, he did not consider sets as being formally defined by such complex expression – the axioms of set theory were to do the mathematically proper, indirect definition. Yet Cantor rightly thought that Kant's "totalities" were a good way to introduce informally the notion of set in an intuitively understandable manner.

I have proposed elsewhere to take Cantor's method of informal characterization seriously and to consider "to take as ..." as a formal operator that applies to neutral objects to begin with, in order to give them a connotation that was not in the object before. This operator effects an internal transformation of the object. Thus we can take objects as multiplicities, and then such multiplicities as unities, or we can take the objects as parts, and then such parts as wholes, etc. The successive application of pairs of opposite kinds of "taking as ..." generates different complex categories – and there is an indefinite number of such pairs. "Set" is a name for one such complex category. But let this be well understood: a set is not in any way a defined concept; it is a complex primitive idea composed of two single ideas in an ordered, internally

related conjugation. The infinite variety of possible ways in which entities can be taken provides an inexhaustible general method with which to generate complex ideas using different operations of "taking as ..." applied in orderly succession. The order, of course, is most important. Were we to take an object as a unity first, and then as an inseparable multiplicity present in it, we would be obtaining the mirror image of a set, something we can appropriately call a "tes."[10] Momentous changes in perception – and hence in understanding – follow such seemingly minor reversals. In fact, new objects are thus being created.

The transformations effected by the operations of "taking as..." are then internal ones in the sense that, unlike most mathematical operations, they are not mere single-valued correspondences; instead, they actually change the internal character of the objects that are being taken in one way or another. The creation of new, ever more complex, categories leads to a mental transubstantiation of the objects to which the categories are applied in successive acts of comprehension. We are, then, obviously very far removed from the notion that primitive ideas are unchangeable eternal entities, and much closer to Occam's view that categories are "the signs of things." Kant said that categories are the outcome of the relations between subject and object, but interpreting this statement in the spirit of Occam indicates that these relations move primarily from object to subject, not the other way around. Categories are not determinations of thought, as Hegel said, but rather determinations of reality, special forms that reality assumes when it is present in the mind. And since – to paraphrase Whitehead – the complexity of reality is inexhaustible, so is the potential growth in complexity and variety of primitive ideas truly unstoppable.

§5. Local and global categories

The meaning of every primitive idea has a life span as well as a limited scope in space. The connotation of concepts applied to the universe at large breaks down thoroughly in ultramicroscopic observations. Light years and nanoseconds – huge spaces and infinitesimal lapses – radically change the meaning of many of our most fundamental primitive ideas, as physics has well demonstrated. There are concepts even that must be put aside altogether, either because they do not apply in a given realm or because they lose their perspicuity and usefulness after completion of the immediate applications for which they were intended. A computer program, for example, may start with undefined concepts that are indispensable in order to reach a given objective but for which there is no use whatsoever afterwards. On the other hand, the notions of number and line have endured for centuries, but what changes have they undergone throughout history! Even categories with the broadest global scope in both space and time are

inevitably subject to remarkable metamorphoses as we become progressively dissatisfied with prevailing frames of mind that are evidently unsuitable to apprehend the facts.

III

EXPERIMENTS IN THE CREATION OF PRIMITIVE IDEAS

§6. Reversing the figure-ground contrast

From an atomistic point of view, terms stand our clearly against the background of relations: relations are supposed to be nothing without the terms that they relate. If, as I have done in another work,[11] we reverse this contrast by turning relations into the primary figure, then a number of semantic and perceptual consequences follow. A new primitive idea emerges, that of *in-between*, which encompasses our thinking of relations as both vectors and media. In turn, terms become clusters of relations, derived crystallizations of moving networks of pure connectedness. The terms' apparent independence is then nothing more than a deceiving mirage. We clearly see, for example, that the meaning of a sentence really lies between its words, in the intuitive currents of comprehension with which we utter or read the sentence. Semantically speaking, words are nothing other than handles of relations, meeting points of flows of continuous understanding.

From this example we can develop a general experimental method for the creation of new categories. Reversals of the figure-ground contrast – making the figure the background of the background – is always possible; the question, of course, is how fecund the new categories can be. Fruitful examples already exist, some of which go back to the Greeks. Parts are often seen as the figure against the whole that gathers these parts as ground. Proclus occasionally focused on the whole as being the figure whose ground is the part; he also talked about the whole being a part of the part. Gestalt psychology has also made a special theme of the whole against its parts, and has dealt extensively and imaginatively with the dominant role of totalities in thinking.

§7. Making new complexes out of contrasting pairs

I have already talked about complex categories, using chaos-and-order as my example. Now I want to add a sequence of experimental cases in which pairs of contrasting ideas are set together into an inseparable complex, not as a synthesis of undefined concepts in any dialectic sense, but as an irreducible multiplicity to be preserved, used, and treasured as such. This new primitive idea is certainly not defined by its components, nor is it a mere juxtaposition: it is a true creative conjugation.

Absolute-and-relative: Einstein's principles of relativity are stated as absolutes. It is an absolute statement to assert that space and time are functions of the relative motion of each

center of reference vis-à-vis other moving centers. This dependence does not vary from region to region of the universe. Hence the absoluteness of relativity cannot be dismissed as a mere play on words, or explained away in terms of levels of language. If I say, "I believe in the absoluteness of relativity," I am not uttering a metastatement that considers relative motions from a nonexistent outside. Physical laws that are not absolutely intended are not truly laws, just as a detached belief, a "metabelief," is not a serious belief, only mere opinion.

Abstract-and-concrete: Just as categories are signs of things, abstractions are signs of concreteness. Abstract and concrete are often taken each by itself as independent concepts separated by a thick mental wall. Yet the abstract emerges from the concrete; it consists of selections of the concrete obtained in an effort to get fixed snapshots of a moving, complex, and elusive reality. "The concreteness of the abstract" is not just a way of speaking. There is nothing more concrete than to view concreteness as the sum total of all the abstract perspectives to which it lends itself. (How creative the conjugation of abstract and concrete can be is easily seen in the area of aesthetics. In one of his periods, the painter Gustav Klimt created works that seem at first sight an example of pure abstract art: patches of color exhibiting a certain pattern. A more attentive look very soon discovers realistic figures embedded in the pattern – as in his famous "The Kiss." This embedding, the wedding of abstraction and realism, creates a thoroughly new artistic quality. We do not perceive the pattern of colors and the human figures as arbitrarily set adjacent to one another but as a whole whose parts encapsulate the entire picture to create a new kind of object. Not to see these paintings of Klimt this way, insisting on dissecting the abstract design from the concrete representation, is to miss altogether his aesthetic point.)

Active-and-inert: The Bhagavad Gita gives the ultimate insight into how these categories conjugate. I paraphrase: it is a mark of wisdom to be able to reach the profoundest inaction in the heart of action, and to engage in action in the heart of inaction. Christian mystics also extol the intense drive that emerges in the midst of a calm contemplation. The two aspects – the exertion and the stillness – do not reduce one to the other; their staying in intimate confluence is precisely the existential condition of the mystic experience.

Affirmative-and-negative: Classical logic teaches us to draw an absolute line between assertion and negation. Yet, useful as this division is in daily situations, the fact remains that even in ordinary language there is often no absolute cleavage between the two: they actually coexist in many real-life experiences. In addition, negations are often alternative assertions. Only the complex category can convey the semantic, epistemological, and even ontological conjugation we are faced with in such situations, which are merely examples of the generalized presence of legitimate logical and existential antinomies.

And-and-or: When "or" is taken in the exclusive sense – either this or that but not both – "and" and "or" stay each on its own course. When "or" is taken inclusively – either this or that or both – as in mathematics, then we have an antinomy: the primitive concept of "or" blends with the primitive concept of "and" in intimate coexistence. In other words, the inclusive "or" is a complex, contradictory category. To choose inclusively becomes a *gathering choice*, not a simple disjunction.

Attract-and-repel: Psychology and literature offer many descriptions of the love-hate complex, a composite mental state whose reality is indisputable. In the dynamics of the mind, the field of psychological forces that pervade it – and that Kurt Lewin so intelligently described[12] – is everywhere crisscrossed by concurrent attractions and repulsions. It is indeed naive to think of the mind as a succession of simple states with only external interaction. Some states are inherently complex: everyone is occasionally torn, impelled, paralyzed, exalted, or crushed by the action of contradictory forces working simultaneously in the mind. We are often in favor of something, and at the same time have reservations, or are even repelled by some aspects of what we promote. This is merely the stuff of ordinary life.

Finite-and-infinite: Many mathematicians would think these are two simple, absolutely incompatible categories. Yet not only are there several different conceptions of finite and infinite, but also, if we drop the axiom of choice, then the so-called "mediate sets" emerge that share the characteristics of finiteness and infiniteness. These sets are more than just in the middle; they are finite-and-infinite, infinite in that they are not equinumerous to any finite set, finite in that they are not equinumerous to any regular infinite set.

Question-and-answer: How often an answer poses more questions than the one it responds to! Of course, we know then immediately not only that the answer elicits a spectrum of new queries, but that the answer itself is part and frame of such a spectrum. Moreover, how many inquiring suggestions include their own implicit answers!

Discrete-and-continuous: The continuum is usually dealt with in mathematics through the devices of set theory – essentially, divisibility of a class into disjoint subclasses. Yet the most characteristic property of the continuum is not its divisibility but its connectedness. Of course, regions can be distinguished in any continuum, but they are not necessarily separable; they are the continuum's discrete aspects, a discreteness that does not imply severance. At any rate, independently of mathematics, continuity remains a primitive idea that neither Dedekind's nor Huntington's approaches managed to capture fully. Using some ideas of Theodore de Laguna, Whitehead outlined a theory of extensive connection independent of set theory;[13] his description

of the extensive continuum is the outcome of this theory. Regions are "the relata which are involved in the scheme of 'extensive connection.' Thus, regions are the things that are connected."[14] Connection is a symmetrical relation, but it is neither transitive nor reflexive, and no region is connected with all the other regions; however, any two regions are mediately connected, i.e., there exists a third region with which both are connected. A region *B* is part of a region *A* when every region connected with *B* is connected with *A*. Two regions are "externally" connected when they are connected but do not overlap, i.e., when there is no third region that is a part of both, and so on. Clearly, the extensive continuum is an exemplification of the discrete-and-continuous complex, *regions* being the identifiable discrete entities and *connectedness* the primary characteristic of continuity. Both de Laguna and Whitehead provide diagrams to represent their respective constructions. These diagrams – as Whitehead admits – are misleading: they can all too easily be confused with the set-theoretic Venn diagrams with which they have nothing in common. It certainly would be good to be able to design accurate representations of the extensive continuum that would be as intuitive as Venn's; this does not, however, seem to be easy. Having written a book titled *In-Between* in which this category is explored, I was delighted to learn in 1990 that to improve on what he already had published "Whitehead told his Harvard classes that it would be best to begin the theory of extension with the relation of betweenness among regions."[15]

More complex categories obtainable from contrasting pairs could be added indefinitely, and clear-cut, useful examples provided for each of them: analysis-and-synthesis, appearance-and-reality, arrival-and-departure, this-and-that, etc. Each of these complexes is a witness to the fact that, contrary to analytic prejudices, primitive ideas are not pigeonholes. But I want now to go beyond pairs, jump to the other end of the spectrum, and look into those categories that emerge as the integration of an infinite series of simple items.

IV

LIMIT IDEAS

§8. A brief history of the notion

In his *Theory of Definition* Heinrich Rickert gives a very apt description of how concepts emerge in the mind, and of their nature and role in the flow of thought. I have already quoted the following fragments elsewhere, but this excellent piece of concrete phenomenology bears repetition and is relevant to my purpose.

"Ordinarily the concept is considered as a preliminary stage to thought, and a judgment as a relation between two concepts. ... [Yet] the content of a concept ... is a series of judgments.

We do not realize this very clearly because we never have occasion to complete verbally such act of concept formation, expressing it in a sentence.... We can then compare the content of our knowledge with a spread of threads in which nodal fixed points are the concepts, while the threads that go from one node to another would represent the relations between concepts, that is, the judgments. If we conceive the threads in their direction toward the nodes, we have an analogy of the synthetic definition, for here the judgments meet in the concept. ... The concept divides into its judgments. ... In a strict sense, thought only moves ... in the level of judgments, and this fact throws light on the theory of the concept."[16]

Rickert was a follower of Kant, and one can perceive in the above quotations a metamorphosis of Kant's notion of limit idea (*Grenzbegriff*).[17] Kant introduced this expression in connection with his *noumena* – a word he used to refer to what lies beyond our sensibility. The notion was taken and transformed by several authors, but it was Husserl who used it in a most positive and systematic way. A limit idea in Husserl's sense parallels the mathematical notion of limit of an infinite sequence or of an infinite series. Thus a sequence of perceptions of an object gives us a sequence of incomplete perspectives whose conceptual integration generates the limit idea of the specific concrete object, never fully accessible to our senses in its totality – as when we look at a sculpture from successive angles. Although the concrete object as a whole is not a given, we have a concept of it as the point toward which a sequence of mental determinations converges. In Husserl's words, slightly condensed: "Perfect givenness is an idea in the Kantian sense – a system of endless processes of continuous appearings, a determined continuum of such appearances. This continuum is infinite in all sides, consisting of appearances in all its phases of a determinable X so ordered in its concatenations and so determined with respect to the essential contents that any of its *lines* yields, in its continuous course, a harmonious concatenation."[18] Rickert's concrete concepts themselves are limit ideas in the same sense, gradually apprehensible in their full concreteness, places to which our thoughts keep converging – in contrast with Kant's never reachable noumena.

It is interesting that Whitehead, not influenced by the authors just mentioned, came to the following, similar, description. I abbreviate: "The chief error in philosophy is overstatement. One form of overstatement is what I have termed elsewhere the 'fallacy of misplaced concreteness.' This fallacy consists of neglecting the degree of abstraction involved when an actual entity is considered merely so far as it exemplifies certain categories of thought. There are aspects of actualities which are simply ignored so long as we restrict thought to these categories."[19] In other words, the notion itself of a specific actual entity is a limit idea, not a mere building block of a finite number of sentences but the mental completion of an unlimited

sequence of movements of thought. As such an integration, the notion plays the role of a part that has as parts the wholes within which it functions.

§9. Convergent and divergent limit ideas

The "fallacy of misplaced concreteness" could also be referred to as "the error of stopping too soon," or of "being satisfied with incomplete considerations," or of "not making the effort to touch the limit." Now this limit is not necessarily unitary; it can be and remain an irreducible multiplicity – Kant talked of noumena in the plural. Furthermore, as with the mathematical parallel, some thought sequences converge to a single notion, others to several notions, others still diverge to infinity. In all cases, again, the limit is not a given but is constructed, reached through specific approaches. Recursive functions, for example, are defined by induction, but inductive definitions never deliver the defined object in its entirety; they are only instructions for progressive computation.

Oscillating sequences were considered divergent in mathematics until the notion of convergence was extended,[20] although still with the restriction that convergence should lead to a single limit. Nevertheless, nothing stops us from stating that the oscillating sequence 1, -1, 2, -2, 1, -1, 2, -2, ... converges to four distinct limits, 1, -1, 2, and –2. Similarly, we collect successive impressions of a person through the years, some good, some bad, some beautiful, some ugly, etc. Our limit idea of such a person is, then, not single but multiple, obtained by the coexistence of several convergent sub-sequences, each telescoping a different picture. There is no way to integrate these experiences, to concretely subsume them all together into a unitary image.

It is clear, then, that we are very far from taking limit ideas as abstractions, or as containers of meaning: they are, rather, approximations to a target, shortcuts we handle as we do any physical object. This is particularly true of divergent sequences, sequences whose terms keep distancing from one another. The limit notion of the universe as a whole is the end result of all our disparate experiences of reality, experiences not only inconsistent at times but also unrelated, even random. Terms of a divergent sequence cannot be seen as foreshortenings of a single or multiple destination. Yet they are tied together by a single process, be it deterministic or aleatory.

Now, independently of infinite sequences, the notion of objects whose existence is the outcome of a fiat, such as in an application of the axiom of choice, also shares in the nature of limit ideas, but more in the sense of Kantian noumena. A choice set of a given arbitrary family of sets exists because the axiom of choice says so. We can never know in general what its elements are. However, we treat such entities as though we knew them well; actually, we build large segments of mathematics on these unknown entities. Mathematicians believe in them

despite their totally nonconstructive nature. They are not in any way approximations; they are planted in a lump through a single act of mathematical faith. Computers cannot by themselves make such a leap into the unknown. The mind must have already established a pattern of steps that the computers can take up in order to mimic in an approximate way what the imagination of the programmer has dreamt nonconstructively.

§10. Categoreal attitudes and moods

If from thinking we move now to other faculties of the mind, then we find processes that parallel the way we use primitive ideas in our rational discourse. Our actions, for example, spring from our attitudes in a logical way. But our attitudes are often self-created, categoreal in an extended sense. Predispositions, or prejudices, are examples of concrete attitudinal categories from which actions follow as a conclusion follows a premise in a theorem. An "Ah!" of approval or a "Bah!" of disapproval are beginnings that frame our immediate inclinations and our subsequent behavior.

This is similar to the world of feelings. Moods, the diffuse condition from which our directed feelings emerge, are the concrete "categories" on which our emotional life is based. Felix Krüger talked about the *bewusstseinerfüllende Breite* of primary dispositions from which even the subject-object complex belatedly derives.[21] Moods are the "primitive ideas" on which love and hate, sentimental attraction and repulsion, are founded with logical inevitability. A change in mood substantially changes our existential position, the way we live our lives. But this is no place to elaborate on these extensions. Let us, therefore, stop our considerations at this point.

NOTES

[1] A. N. Whitehead, *Process and Reality: An Essay in Cosmology*, corrected edition, New York: The Free Press, 1978.
[2] Ibid., p. xiii.
[3] Cf. F. Brentano, *Aristotle and His World View*, trans. by R. George and R. Chisholm, Berkeley, CA: University of California Press, 1978, pp. 28-33.
[4] Cf., for example, F. G. Asenjo, "Toward an Antinomic Mathematics," Chapter 15 in R. Routley, G. Priest, and J. Norman (eds.), *Paraconsistent Logic*, Munich: Philosophia Verlag, 1989, pp. 394-414; also, C. Mortensen, *Inconsistent Mathematics*, Dordrecht: Kluwer, 1995.
[5] *Process and Reality*, p. 3.
[6] P. Aczel, *Non-Well-Founded Sets*, Stanford, CA: Center for the Study of Language and Information, 1988. Also, M. Crabbé, "Soyons positifs: la complétude de la théorie naïve des ensembles," *L'anti-fondation en logique et en théorie des ensembles*, edited by R. Hinnion, Louvain-la-Neuve: Academia, 1992, pp. 51-68.
[7] *Process and Reality*, p. 4.
[8] Cf. E. Cassirer, *Substance and Function*, trans. by W. C. and M. C. Swabey, New York: Dover, 1953.

[9] F. G. Asenjo, "One and Many," *Philosophy and Phenomenological Research*, XXVI, 1966, pp. 361-370.
[10] F. G. Asenjo, "The Logic of Opposition" and "The Present Situation of Mathematics and the Mathematics of Opposition," both to appear, expand on the method just briefly described.
[11] F. G. Asenjo, *In-Between: An Essay on Categories*, Washington, DC, and London: University Press of America, 1988.
[12] K. Lewin, *Principles of Topological Psychology*, New York: McGraw-Hill, 1936; also, *The Conceptual Representation and the Measurement of Psychological Forces*, Durham, NC: Duke University Press, 1938.
[13] T. de Laguna, "Point, Line and Surface as Sets of Solids," *The Journal of Philosophy*, Vol. 19, 1922, pp. 449-461. Whitehead's presentation is in *Process and Reality*, pp. 294-309. See also K. Menger, "Topology Without Points," Rice Institute Pamphlet 27, 1, 1940, p. 107; J. von Neumann, *Continuous Geometry*, Princeton, NJ: Princeton University Press, 1960; and F. G. Asenjo, "Continua Without Sets," *Logic and Logical Philosophy*, Toruń, Poland: Nicholas Copernicus University, 1993, pp. 95-128.
[14] *Process and Reality*, p. 294.
[15] V. Lowe, *Alfred North Whitehead: The Man and His Work*, vol. II, Baltimore, MD: The Johns Hopkins University Press, 1990, p. 236.
[16] These fragments are all from Chapter 5 of H. Rickert's *Zur Lehre von der Definition*, Tübingen: Mohr Verlag, 1929; quoted in my "Continua Without Sets."
[17] I. Kant, *Critique of Pure Reason*, trans. by P. Guyer and A. Wood, Cambridge: Cambridge University Press, 1998, p. 362.
[18] E. Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, First Book, trans. by F. Kersten, The Hague: Martinus Nijhoff, 1982, p. 342.
[19] *Process and Reality*, pp. 7-8.
[20] Cf. G. Hardy, *Divergent Series*, Providence, RI: American Mathematical Society, 2000.
[21] See *In-Between*, p. 142. The best study of directed feelings is still Alexander Pfänder's treatment in his "Zur Psychologie der Gesinnungen," *Jahrbuch für Philosophie und phänomenologische Forschung*, I, 1913, pp. 325-404, and III, 1916, pp. 1-125.

Rev. Academia de Ciencias. Zaragoza. 56 (2001)

SOBRE LOS HECHOS

Eladio Domínguez

Dept. Informática e Ingeniería de Sistemas
Facultad de Ciencias. Universidad de Zaragoza
50009 Zaragoza (Spain)
e-mail: noesis@posta.unizar.es

Para que una teoría sea aceptada como teoría científica es necesario que se establezcan sus tipos de verdad. Una teoría científica no sólo debe establecer sus objetivos, sus formas de comunicación, sus tipos de proposiciones sino que además debe establecer cuándo una proposición es aceptada como verdadera o falsa. Esto último es lo que determina el método que es propio de la teoría.

Aunque en todos los trabajos publicados sobre Fenomática he abordado en alguna medida el tema de la verdad fenomática aún es necesario desarrollar una profunda investigación sobre dicha noción para alcanzar un cierto grado de formalización aceptable. En este artículo presento algunas reflexiones sobre los hechos como las verdades primeras, como las verdades que sentimos como más incuestionables. Estas reflexiones fueron la base fenomenológica de las ideas que mostré en la conferencia que impartí en el EACA'2001 en relación al constructivismo.

1. Hecho y verdad. *Un hecho es una cosa* (física, abstracta, general o concreta) *que aceptamos como percibida directamente*, como percibida a través de nuestros órganos sensoriales (tanto externos como internos). Si veo algo como una mesa que se encuentra ante mí, esa cosa, la mesa, la siento como un hecho, la siento como una verdad irrefutable, como algo indiscutible, como algo cuya verdad es inalterable mientras permanezca el hecho, mientras permanezca la mesa en su posición.

Sentimos también que esa verdad, -que una mesa se encuentra ante mí- como hecho que es, debe ser aceptada por toda otra persona presente en el caso en que no tenga dañada la

vista, y también sentimos que esa verdad debe ser aceptada por toda persona ausente a la que se lo comunique, siempre y cuando acepte la no existencia de error en mi propia percepción y acepte también que se comunica con verdad, sin engaño.

Sin embargo pudiera ser que otra persona me indicara que estaba en un error, que lo que siento como una mesa realmente es un taburete. Pudiera ser que, ante mi incredulidad, me acercara al objeto y, pudiera ser entonces, que me diera cuenta de mi error, que pudiera ser que sintiera ahora aquello como un taburete.

He aceptado el objeto como una mesa, como un hecho, como verdad mientras no se me ha mostrado lo contrario. Aunque la he sentido como una verdad indiscutible, no he tenido inconveniente en aceptarlo como falso, como algo que no siento como un hecho, cuando ha existido una muestra directa de ello. Este tipo de casos nos mueve a pensar que una verdad absoluta lo es sólo en el mundo subjetivo, en relación a la percepción del individuo; nos mueve a pensar también que algo sentido como un hecho, como *una verdad absoluta se siente así mientras no se nos muestre lo contrario.*

Un hecho del que conocemos su aceptación como tal por una comunidad, por todas las personas presentes en ese momento, es una verdad que sentimos como más cierta, como más irrefutable. La sentimos como una verdad intersubjetiva, como una verdad sentida como absoluta por una cierta comunidad de personas. A pesar que *los hechos intersubjetivados se sienten como reafirmados en su propia verdad, también* se sienten, aunque sea de modo inconsciente, *como verdades mientras no se nos muestre lo contrario.*

Un hecho es sentido como que así lo será para toda persona que se encuentre en nuestras propias circunstancias. Ese *es nuestro sentido* real *de la verdad incuestionable, como la verdad de hecho, como verdad subjetiva, como verdad intersubjetivable, como verdad objetivable* por cualquier persona que se encuentre en nuestras circunstancias.

2. Hecho y percepción. En el apartado anterior hemos señalado un hecho como una cosa que aceptamos como percibida directamente. Ese es nuestro sentir natural ante un hecho cuando estamos en actitud natural. En actitud teorética, sin embargo, es necesario analizar con más detalle las situaciones mostradas anteriormente.

Un hecho refiere a una cosa que se siente como una verdad mientras no se nos muestre lo contrario. Pero la cosa referida en un hecho es, generalmente, independiente de nosotros como observadores, y, por consiguiente, independiente de nuestro acto del percibir. En ello se fundamenta precisamente el que la misma cosa pueda ser sentida como distinta en circunstancias diferentes o que la misma cosa pueda ser sentida como la misma. Eso nos

induce a pensar que *el hecho es realmente el efecto de percibir directamente la cosa. La cosa es el objeto al que se dirige la atención de nuestro acto. Nosotros somos el sujeto de ese acto y el hecho es el efecto de ese percibir.*

3. Hecho e intencionalidad. Cuando sentimos esa cosa como una mesa, *el acto de percibir está guiado por una intención.* Quizás sea sólo la intención de observar lo que hay a mi alrededor, o bien la de descubrir posibles obstáculos en mi caminar, o bien la de buscar una mesa para comer en ella.

La intencionalidad que forma parte, como ingrediente, del percibir directo *guía el* propio *acto llevando ciertos aspectos sentidos al plano de la conciencia y reduciendo otros, reduciéndolos, olvidándolos de modo inconsciente para, quizás, recordarlos, elevarlos a la conciencia si fuera necesario.*

De ese modo, si mi intención es un simple observar lo que hay a mi alrededor, serán aspectos relevantes aquellos que me permiten reconocer de qué objeto se trata –por ejemplo, que tiene cuatro patas que soportan una superficie lisa,...-. Si mi intención es la de descubrir posibles obstáculos, serán aspectos de interés los de su magnitud y situación. Y, por último, si mi intención es buscar una mesa para comer en ella, intentaré descubrir en la cosa los aspectos que me la revelen como mesa así como aquellos que me indiquen si puede ser utilizada con el fin deseado.

A través del último ejemplo podemos observar que las intenciones se superponen, matizándose unas a otras. De hecho, en el segundo caso, en el descubrir obstáculos, quizás sienta también la cosa como una mesa, quizás sienta lo que es el obstáculo. En esta situación, aunque la cosa como mesa queda en mi consciencia, ese conocimiento queda subyugado a un segundo plano, lo sentimos como algo no esencial en relación a la intencionalidad que ha conducido el acto del percibir, en relación a la de descubrir obstáculos.

En este punto considero importante señalar, aunque sea someramente, el sentido de la reducción al plano de la inconsciencia de los aspectos no sentidos como esenciales. *La reducción no es un olvido total, la reducción es un desviar los aspectos, extraerlos del plano consciente y dejarlos ahí, en la inconsciencia para que puedan ser recordados si fuera necesario.* Ese reducir como acto cognitivo *está guiado por esa intencionalidad*, tiene como intención el olvido temporal de esos aspectos con objeto de que puedan ser recordados posteriormente.

El reducir anterior es el que se realiza en general salvo cuando en la propia intención se pretende el olvido total, el olvido sin posibilidad de recuerdo.

4. Hecho y humanidad. *La intencionalidad que conduce al hecho* –al hecho percibido- *es lo que señala al sujeto de la acción como ser humano.*

En general, cuando un animal mira a su alrededor para distinguir algo que pueda ser comido, es un acto conducido por el instinto, por el instinto de supervivencia. Cuando decimos que 'ese animal mata para comer', en actitud natural sentimos que el acto de matar conduce al de comer, utilizamos la partícula 'para' con el fin de señalar que el acto de matar conduce el acto de comer el animal que ha matado. Pero, en esa actitud natural no damos ningún sentido expreso de fin o intención. En actitud reflexiva deberíamos decir que el sentido de la preposición 'para' en la frase 'ese animal mata para comer' es el de *conducción instintiva.*

En el sentido expresado admitimos que *la intencionalidad, y, con ella, los hechos son propios de la humanidad.*

El que la intencionalidad forme parte inseparable de los actos ordinarios ejecutados por seres humanos es una señal de su consciencia.

Aunque una persona, como animal que es, puede realizar también actos instintivos, lo que lo *señala como ser humano es* que puede *realizar*, y de hecho lo hará ordinariamente, *actos que se* le *revelan como conscientes, revelación que* le *viene dada por la intención del acto.*

5. Hecho y viabilidad. *Un perceptible es una cosa que puede ser objeto de un percibir*, directo o indirecto. Sobre toda cosa admitimos esa propiedad, la de ser perceptible. Pero aunque nos obliguemos a ese principio, *no debemos identificar cosa con perceptible.* Utilizamos el término 'perceptible' cuando sobre la cosa sentimos el carácter de ser objeto de un percibir, y reservamos el término 'cosa' para cuando reducimos ese aspecto.

No todo perceptible puede ser objeto de un percibir directo. Puedo percibir lo inexistente pero siempre será a través de un percibir indirecto. Sin embargo, en el caso particular de un hecho, no sólo damos carácter de perceptible sino que además sentimos que es propio de su naturaleza el poder ser directamente percibido. Así sentimos y afirmamos, *por principio,* que *todo hecho puede ser objeto de un percibir directo, puede ser referido, como objeto, por otro hecho.*

Ese rasgo asociado a todo hecho de sentirlo en el universo de lo que directamente puede ser percibido es lo que llamamos *viabilidad.*

6. Hecho como algo construido. El percibir no lo sentimos como un mero sentir. *En el percibir siempre apreciamos* por lo menos un 'tener consciencia de un sentir' y, precisando más, *un tener consciencia de un sentir como el sentir de algo*[1].

El efecto de un sentir distinguido en un acto del percibir –es decir, la sensación distinguida- no es nunca algo irreducible. *Una sensación distinguida se obtiene a partir de un cúmulo de sensaciones más primitivas que nuestra consciencia* –y en definitiva, el acto del percibir- *distingue como una sensación*. Este aspecto señalado por *el 'tener consciencia' y el que la sensación distinguida sea 'como el sentir de algo' invocan al conocimiento previo que tenemos del mundo*, conocimiento que nos proporciona el método utilizado en el percibir. De ese modo *sentimos todo percibir como un construir*[2].

El conocimiento del mundo no es solo necesario para la intencionalidad del acto sino que es necesario también para la construcción de la sensación distinguida. Tenemos que tener una idea de magnitud para poder percibir algo como un obstáculo, es necesario tener previamente una idea de color para poder percibir algo como un color concreto. Sin el conocimiento previo, sin la consciencia actuando en el percibir, tendríamos una mera sensación que no podríamos reconocer como la sensación de algo.

7. Hecho y validez. Los hechos, como hemos señalado anteriormente, se sienten como verdaderos. Sin embargo, en actitud reflexiva, podríamos preguntarnos sobre la corrección de lo percibido, sobre su validación y su validez en relación a la intención del acto.

Usualmente volvemos nuestra atención hacia el objeto y lo comparamos, teniendo presente nuestra intención, con el hecho percibido. Pero ¿cómo debemos sentir esa comparación?, ¿podemos comparar directamente y entre sí esas dos realidades dadas en el objeto, una, y en el hecho, la otra?

Según nuestra perspectiva, *toda comparación se realiza a través de nuestras propias percepciones de esas realidades que deseamos comparar.*

La vuelta al objeto siempre se realiza una vez efectuada la primera percepción. Por ello, la segunda no sólo tendrá matices distinguidos de la primera por realizarse en distinto momento sino que también se distinguirá por el hecho de que, en el segundo caso, el acto se realiza con una intención añadida, la de validar la propia percepción.

---

(1) *Un paseo fenomático*
(2) *Constructivismo en Informática: ¿Es constructivista el constructivismo?*

Cuando volvemos nuestra atención hacia el hecho, hacia el efecto del primer acto del percibir, realizamos otro percibir, esta vez en nuestro interior, un percibir que señalamos como un recordar, que está matizado por la intención de comparar para validar.

Las dos últimas percepciones, ambas en el mismo campo de nuestra consciencia, son las que pueden compararse entre sí. *Estos dos hechos, la vuelta al objeto y el recuerdo del inicialmente percibido, son los que comparamos para establecer la corrección o falsedad del primer hecho o su validez en relación a nuestra intención perceptiva inicial.*

En algunas ocasiones la comparación es sencilla, simplemente se indistinguen. En otras es necesario realizar una tarea de análisis de esos hechos. *En todo caso es una comparación entre los hechos, no entre las realidades mismas.*

8. Hecho y complejidad. *Un hecho simple es un hecho en el que no se distingue, como parte, ningún otro hecho.* Si veo la silla en un acto intencional de buscar una para sentarme en ella, para descansar, usualmente no distinguiré en la percepción ningún otro hecho. Sentiré la silla como un todo sin partes.

Ahora bien, si busco una silla para sentarme, pero busco la que tiene el respaldo recto, cuando la encuentre se me dará un hecho como parte, se me dará, en lo que percibo como una silla, lo que percibo como su respaldo recto. Se trata en este caso de *un hecho complejo, hecho, en el que, como parte, se distingue por lo menos un hecho.*

Los dos ejemplos anteriores nos muestran hechos distinguidos, uno como simple y el otro como complejo. Aunque pudiera ser que en ambos casos se tratara de la misma silla, las dos percepciones, los dos hechos se distinguen. La distinción proviene de que *en uno de los actos del buscar existe un matiz en su intencionalidad* –buscar una de respaldo recto- *que distingue el correspondiente acto de buscar.*

Otro ejemplo nos lo proporciona la frase 'Juan tiene un coche'. Esa cosa puede ser percibida como una expresión en la que no se quiere distinguir ninguna parte, o puede ser percibida como una frase en la que se distinguen las partes -como el sujeto, el verbo y el predicado- . Pero también podría ser percibida como que, además, cada parte esta constituida por ciertas letras.

9. Hecho y comunicación. Un hecho, la percepción directa de una cosa, es un elemento subjetivo; *es una cosa en el nivel de consciencia de un individuo. Para que* esa verdad sentida *se perciba como una verdad intersubjetiva es necesario que sea comunicada y así pueda ser*

comparada con la cosa que se nos presenta como la percepción directa del mismo objeto realizada por otro sujeto.

Por ello, para sentir las verdades como intersubjetivas *es necesario sentir sobre los hechos el carácter de ser comunicables*, de ser transmitidos a través de algún lenguaje.

Para llevar a efecto la comunicación de un hecho se realiza un proceso mental consistente en construir la forma que expresará el hecho.

Es importante resaltar que dicha expresión, *el hecho realizado como expresión, expresa al propio hecho pero no expresa el objeto que ha sido percibido. Dicho objeto es señalado intencionalmente a través de la expresión construida mediante la significación que se da a ésta.* De este modo, cuando al ver una persona decimos 'Juan' estamos señalando al objeto, a la persona concreta, a través de su nombre, nombre que la señala, nombre que es la realización del hecho percibido como expresión; 'Juan' es el hecho en forma de expresión, la persona que señala es el objeto sobre el que hemos construido el hecho, la forma de señalar –a través del nombre- nos indica el hecho percibido –la percepción de una persona llamada Juan-; finalmente, el hecho sentido en actitud natural es el propio Juan.

Según sea el contexto 'Juan' puede referir, como objeto que ha sido percibido, a una persona concreta, a la persona que estamos señalando. Aunque también pudiera ser que refiriera al propio nombre que se expresa; caso éste en el que el objeto que se percibe y lo percibido en él se indistinguen.

Una vez se ha construido en nuestra mente la expresión que expresa el hecho, con objeto de que sea comunicado el hecho, *es necesario realizar* otra construcción consistente en *la materialización de la expresión.* Esta cosa, la materialización de la expresión que, en nuestra mente, expresa el hecho, es *el hecho transcendido, es el hecho convertido en perceptible para la pluralidad de los sujetos.*

10. Hecho comunicado y hecho. *El hecho comunicado es un hecho con el que se pretende* comunicar el inicialmente percibido, con el que se pretende que *el que perciba la comunicación* perciba, a través de ella, el objeto de la comunicación, *perciba el hecho que se comunica.* Si es así, si el interlocutor percibe el objeto, es cuando decimos que ha comprendido la comunicación. Sin embargo hay que tener presente que la percepción del objeto, *la percepción del hecho inicial, a través de la comunicación, es una percepción indirecta generalmente*; es decir, ese hecho inicial, que lo es para el comunicador, no es, en general, un hecho para el comunicado.

11. Hecho aceptado como hecho. La mayoría de las personas no ha visto directamente el planeta Venus. Una minoría lo distingue en el cielo del resto de astros, alguien les ha señalado esa estrella brillante como Venus. Muchos sólo lo han visto en algún documental gráfico. Pero todos lo aceptan como un hecho, todos aceptan la existencia de Venus como un hecho.

Una persona, dadas sus naturales limitaciones físicas, no puede percibir todas y cada una de las cosas que en alguna ocasión han sido distinguidas, como hechos y que siguen permaneciendo ahí como para ser percibidas directamente. Pero *para que una persona se sienta integrada socialmente es necesario que acepte como hecho la existencia de una amplia familia de cosas percibidas directamente por otras personas y que no han sido percibidas directamente por él mismo*. Su grado de integración depende en cierto modo de la magnitud de dicha familia.

Un matemático no puede demostrar todo lo demostrable en matemáticas pero, sin embargo, para ser matemático, para comprender y construir matemáticas es necesario que acepte como hechos la mayoría de las demostraciones matemáticas. Notemos aquí que me refiero a las demostraciones como hechos, no me refiero a lo que demuestran. Un matemático constructivista aceptará como hecho una demostración no constructiva pero no aceptará como hecho lo que se demuestra a través de ella mientras no perciba una demostración constructiva.

12. Hecho e información. Un comunicar es un comunicar algo, es comunicar el objeto de la percepción. De este modo 'Juan' es una comunicación, se comunica un nombre de persona.

Si el anterior nombre se comunica oralmente con la intención de llamar la atención de una persona llamándola por su nombre, es sólo la comunicación de algo -el nombre Juan-; pero no se comunica algo sobre una cosa, a la persona no se le comunica su nombre, sólo se expresa su nombre con la intención de llamar su atención.

Cuando señalamos a una persona y dirigiéndonos a otra decimos 'Juan', estamos comunicando el nombre de esa persona señalada, estamos comunicando algo -Juan- sobre una cosa -la persona-. El hecho comunicado -Juan- es en este caso, y no lo es en el anterior, un dato de información.

Cuando nos dirigimos a una persona y le decimos 'éste es mi amigo Juan', estamos comunicando algo -es mi amigo Juan- sobre algo -lo señalado por éste-. En este caso lo comunicado y sobre qué o quién se comunican están presentes en la propia comunicación.

Decimos entonces que se trata de una información y se llama *informar* a este tipo de acto del comunicar.

En este punto es muy importante señalar la diferencia existente entre los dos últimos casos.

En ambos se informa.

En el primero de los dos la información está dada mediante el dato comunicado -Juan- y el gesto que señala a la persona con ese nombre; es una información mixta constituida por dos comunicaciones relacionadas, ambas expresadas en lenguajes distinguidos. La comunicación del nombre, por sí misma, es sólo comunicación. No es información pues en la propia comunicación no se puede percibir de quién es el nombre. Sin embargo le llamamos dato por comunicar algo sobre lo que se informa.

Y, por último, en el segundo caso es una información si en la frase se percibe que lo que se afirma -es mi amigo Juan- se afirma de -éste—Ambos se perciben como datos de información.

Fijémonos ahora en la diferencia entre los datos 'es mi amigo Juan' y éste'. En el primero podemos distinguir como partes relacionadas 'mi', 'amigo' y 'Juan'. Los tres son comunicaciones de algo sobre la persona señalada por 'Juan'. Es por ello que 'es mi amigo Juan' se percibe como un dato complejo. Por el contrario la partícula 'es' no se siente como un dato de la persona señalada, se trata de una partícula verbal que establece la relación entre los otros datos. Como parte de la frase 'es mi amigo Juan', la partícula 'es' señala sobre alguien, no expresado en esta frase, el hecho de que se siente como 'mi amigo Juan'.

Otra forma distinguida de percibir la información 'éste es mi amigo Juan' es considerando como datos 'éste' y ' mi amigo Juan', y considerando entre ellos la partícula 'es' como la que estructura dichos datos para constituirse en información.

13. Hechos frente a las verdades de razón. Si oigo una voz, la reconozco como la de mi amigo Juan y ubico a la persona que habla en la habitación contigua, deduzco que mi amigo Juan se encuentra en ella. Es una percepción indirecta dada por un método deductivo. Se trata de una verdad de razón. No percibo directamente que mi amigo Juan está en esa habitación, no se trata de un hecho.

Las verdades de razón se sienten como más débiles que las verdades de hecho y la credibilidad en ellas es dependiente de la credibilidad en el proceso que nos conduce indirectamente a su percepción.

La credibilidad de las verdades de hecho también depende del propio acto del percibir directo (que realmente es también un proceso complejo aunque, en nuestro contexto, no nos detenemos a analizar los procesos más primitivos que forman parte de él).

Nos merecen menos credibilidad las verdades de hecho sentidas por personas con deficiencias sensoriales y, por el contrario, en las percepciones abstractas, aquellas en las que no son fundamentales nuestros órganos sensoriales externos, dichas deficiencias tienen poca influencia, generalmente, en la credibilidad de las verdades de razón.

14. Hecho y cosa factible. En actitud reflexiva nos podemos preguntar si cierto perceptible es un hecho. El sentido preciso de esa pregunta es el de conocer si en cierto perceptible se aprecia la característica de que pueda ser percibido directamente. Por ejemplo, podemos preguntarnos si sabremos construir el menor número primo mayor que 2^{1000}. Puesto que sabemos, como verdad de razón, que ese número existe y disponemos de un algoritmo para su cálculo, percibimos que ese numero es factible; es decir, en él se percibe la característica de poder ser construido y, de ese modo, percibirlo como una verdad de hecho.

Quizás no dispongamos del tiempo, de la paciencia o de la máquina adecuada para ejecutar el algoritmo mencionado anteriormente. Quizás no podamos percibir dicho número en la práctica. Si pensamos lo anterior estaremos sintiendo que estamos ante una cosa que pudiendo ser un hecho no se nos dará nunca como tal.

Lo *factible*, como *aquello que percibimos como que puede ser percibido como un hecho*, no sólo aparece en actitud reflexiva sino que es necesario para hablar, para comunicar en actitud teorética.

Referencias

[1] Eladio Domínguez. *Un paseo fenomático*. Publicaciones de la Academia de Ciencias de Zaragoza, 1999.

[2] Eladio Domínguez. *Reflexiones fenomáticas sobre la conferencia del Prof. Asenjo*. Revista de la Academia de Ciencias de Zaragoza, 55(2000)23-41.

[3] Eladio Domínguez. Constructivismo en Informática: ¿Es constructivista el constructivismo? *Actas del Encuentro de Álgebra Computacional y Aplicaciones EACA'2001*, editor J. Rubio, Universidad de la Rioja, 2001, pp. 25-34.

Rev. Academia de Ciencias. Zaragoza. **56** (2001)

SUPERFICIE DE MÖBIUS: APLICACIONES

Concepción Longás Monguilod

Departamento de Matemática Aplicada

Universidad Complutende. Madrid.

Abstract

The Möbius strip, is, from its beginning, one of the mathematical discoveries that has contributed more to provide different examples and counter-examples from different geometric and topologyc properties. In this article we gather several of these applications, such as theorems and properties, in which the Möbius strip has special meaning.

1. Banda de Möbius: aplicaciones topológicas y geométricas

Desde que en 1858, J. B. Listing y A. F. Möbius, descubrieron, por diferentes caminos, la banda de Möbius, han sido numerosas las aplicaciones, que, de sus propiedades geométricas y topológicas, se han estudiado.

Para construir una banda de Möbius se necesita, un rectángulo de papel. Si indicamos sus vértices con las letras A, B, C, y D, y luego efectuamos una torsión de 180°, en cualquiera de los dos lados del rectángulo, de tal forma que hagamos coincidir el vértice C con el A, y el vértice D con el B, obtenemos una banda de Möbius de primer orden.

Una banda de Möbius se puede dibujar en tres dimensiones, utilizando la siguiente parametrización:

$$\begin{aligned} X &= \cos s - t\cos s/2 + \cos s, \\ Y &= \operatorname{sen} s + t\cos s/2 + \operatorname{sen} s, \qquad \text{con} \qquad -1 \le -t \le 1, \quad 0 \le s \le 2\pi. \\ X &= t \operatorname{sen} s/2 \end{aligned}$$

Otra forma de parametrizar la superficie de la banda es, mediante la siguiente expresión:

$$R(u,v) = ((4-v\sin u)\cos 2u, (4-v\sin u)\sin 2u, v\cos u), \qquad \text{con} \quad 0 \le u \le \pi, \quad -1 \le v \le .$$

La banda de Möbius, que está construida a partir de una superficie formada por dos caras, tiene la propiedad siguiente:

Todo punto P de esta superficie, puede ser unido con otro punto Q cualquiera de la misma, mediante una curva contenida en ella y sin tener que pasar por la frontera de la superficie.

La superficie así obtenida tiene una sola cara y no es orientable. En su descripción inicial, Möbius ya demostró esta propiedad. Su método fue considerar que la superficie rectangular, que da lugar a la banda de Möbius, está construida por figuras poligonales planas (por ejemplo, triángulos) situadas una al lado de la otra. A continuación define la diferencia entre giro de sentido directo y de sentido inverso. Se toma un triángulo T, de la superficie, y se define el orden cíclico de sus vértices para que la rotación sea en el sentido contrario a las agujas del reloj. En el triángulo contiguo T' se debe elegir, para sus vértices, el sentido que sea compatible con la rotación.

Para que dos triángulos contiguos verifiquen la compatibilidad con la rotación, es necesario que la arista que comparten esté orientada en sentidos contrarios respecto a las rotaciones de los triángulos (es decir, que en el primer triángulo se recorre en un sentido, y en el otro triángulo en sentido contrario).

En la superficie formada por los triángulos, éstos son compatibles con un conjunto de rotaciones, incluso aunque se prolongue la superficie. Pero cuando se construye la banda de Möbius, juntando, de forma adecuada, las aristas del primer y último triángulos de la superficie anterior, dicha condición de compatibilidad no se cumple. Es decir, la arista común del primer y último triángulos, está orientada en ambos, en el mismo sentido.

August F. Möbius indicó que no todos los triángulos que forman la banda, son compatibles con las rotaciones. Es, por lo tanto, un método para describir la propiedad de no orientabilidad de la superficie de Möbius.

Existe otra caracterización de superficies orientables, que viene dada por el siguiente Teorema:

Teorema 1 *Sea X una superficie de dimensión n y sea (X, ψ) una subvariedad $\subset \mathbb{R}^{n+1}$. Si (X, ψ) admite un campo vectorial normal no nulo, es decir, existe una aplicación continua: $V : X \longmapsto T(\mathbb{R}^{n+1}) : \quad x \to V(x) \neq 0$, donde $V(x) \in T(\mathbb{R}^{n+1}, \psi(x))$, y es perpendicular a $d\psi(T(X, x))$. Entonces X es orientable.*

La banda de Möbius no cumple dicho teorema, puesto que no existe dicha aplicación V, y, por lo tanto, como hemos indicado anteriormente, es una superficie no orientable. Una superficie no orientable de dimensión 2, es llamada superficie unilátera.

Otra propiedad característica de la banda de Möbius, es que no verifica el siguiente teorema, que es una generalización, de la segunda forma alternativa, del Teorema de

Green, a tres dimensiones.

Teorema 2 (De la Divergencia) *Sea W una región sólida acotada por una superficie cerrada S, orientada por vectores normales unitarios dirigidos hacia el exterior de W. Si $\boldsymbol{F} = P\boldsymbol{i} + Q\boldsymbol{i} + R\boldsymbol{k}$ es un campo vectorial cuyas funciones P, Q y R tienen derivadas parciales continuas en W, entonces*

$$\int\int_S \boldsymbol{F} \cdot \boldsymbol{n} dS = \int\int\int div\boldsymbol{F}\, dV,$$

donde la integral triple, se extiende a la región W.

Aunque la banda de Möbius es una superficie cerrada, no se puede dibujar ningún vector normal a su superficie, dirigido hacia el exterior, que sea válido en cualquier punto de la banda.

2. La banda de Möbius y los haces de fibras

Las propiedades globales de los haces de fibras, están relacionadas con el concepto de espacio producto.

Si M y N son superficies, se puede definir su espacio producto, $M \times N$, formado por todos los pares ordenados (a, b), tales que $a \in M$ y $b \in N$.

Un haz fibrado es, al menos localmente, un espacio producto UxF, donde U es un conjunto abierto de la superficie base $B(U \subset B)$, y el espacio F, representa una fibra característica. Esta condición de espacio producto local, forma parte de la definición de un haz fibrado: es 'trivial localmente', es decir, es un espacio producto para una región local de B. La cuestión que se plantea es, si también es 'trivial globalmente', es decir, si todo el haz de fibra puede ser representado como un espacio producto $B \times F$.

La respuesta es **no** generalmente. Vamos a indicar un ejemplo, proporcionado por la banda de Möbius, de que se puede construir un haz no trivial globalmente, incluso cuando el espacio base B permite un haz trivial globalmente.

Consideremos TS^1, el haz tangente de la circunferencia S^1, (Fig. 2.a). Esta circunferencia permite un campo vectorial continuo no nulo (se podría dibujar como un cilindro), y $TS^1 \equiv S^2 \times \mathbb{R}$ (es decir, es el espacio producto de la circunferencia TS^1, por la fibra $\mathbb{R}$, formada por las líneas verticales que pasan por S^1). Este haz fibrado, es, por lo tanto, 'trivial globalmente'.

Si cortamos la circunferencia por un punto P y desarrollamos el haz en forma análoga a la superficie lateral de un cilindro, podríamos reconstruir la figura inicial juntando los extremos superiores entre sí, y análogamente, los inferiores, y tendríamos un haz trivial

globalmente. Pero podemos reagrupar el haz fibrado en una forma diferente, (Fig. 2.b), formando una banda de Möbius.

Localmente, el haz, representado en la Fig. 2.b, tiene una aplicación inyectiva continua sobre alguna parte del haz inicialmente construido. Pero globalmente, no existe una aplicación continua inyectiva que haga corresponder un haz con el otro. Por lo tanto, la banda de Möbius no es un espacio producto, y, como consecuencia, el segundo haz no es trivial globalmente.

El ejemplo de Möbius nos indica que, no es suficiente decir que existe una base y la fibra de un haz, porque puede haber más de una forma de construir dicho haz. Se necesita por tanto, para definir un haz de fibra, la introducción de grupos.

Por lo tanto, un haz fibrado es un espacio E, para el que existen: una superficie base B; una proyección $\pi : E \mapsto B$; una fibra característica F; un grupo G de homeomorfismos de F en sí misma, y una familia $\{U_k\}$ de conjuntos abiertos que cubren B, de tal forma que verifiquen:

1. Localmente el haz es trivial
2. Cuando dos abiertos U_i, U_j de la familia $\{U_k\}$ se superponen, un punto dado x en su intersección, tiene dos homeomorfismos distintos $h_i(x)$ y $h_j(x)$ de su fibra en F.

3. La banda de Möbius en Geometría proyectiva

Cuando una asíntota de una hipérbola se toma como una línea proyectiva, podemos considerarla como un camino cerrado, pues podemos ir hacia el infinito a través de ella y volver del infinito también a través de ella, por la otra parte del plano. De esta forma, vemos que la rama de la asíntota que va hacia infinito, lo hace a la derecha de la hipérbola, y la rama de la asíntota que vuelve de infinito lo hace a la izquierda de la hipérbola.

Ahora bien, como sabemos, la asíntota es una tangente a la hipérbola en el infinito, pero no corta nunca a la curva. Si ahora ensanchamos la asíntota, (Fig. 3), como una banda, y observamos su frontera, sucede al igual que antes, que cuando la asíntota va hacia infinito, la frontera de la superficie queda a derecha de la curva, y cuando la asíntota vuelve de infinito, la frontera de la superficie aparece a la izquierda de la hipérbola.

Es decir, que un recorrido de dicha asíntota proporciona solamente media vuelta de la frontera de la banda. Por lo tanto, es una banda de Möbius. De aquí se puede deducir que el plano real proyectivo, es no orientable.

4. Banda de Möbius de orden n

Si en la construcción de la banda de Möbius, en lugar de realizar una torsión de 180 al rectángulo inicial, efectuamos n semivueltas antes de unir los vértices en la forma indicada, obtendremos una cinta de Möbius de orden n-ésimo. Las torsiones pueden ser realizadas en un extremo o en otro del rectángulo inicial, o en ambos. Pero aunque la figura obtenida, con estas distintas posibilidades de efectuar los giros, sea algo diferente de aspecto, las propiedades que vamos a enunciar a continuación, se verifican en cada una de las mismas.

Propiedades: Si n es impar, la superficie así obtenida tiene una sola cara y una única frontera. Dicha frontera, además, está anudada cuando $n > 3$.

Cuando n es par, la superficie resultante tiene dos caras y dos fronteras, las cuales están entrelazadas cuando $n \geq 2$.

Este resultado significa que cuando el número de semivueltas es impar, la banda sigue manteniendo la condición de unilateralidad, y las propiedades que de dicha condición se derivan.

Otras aplicaciones.

Además de las aplicaciones de las propiedades de la banda de Möbius, en las especialidades de la Geometría y Topología, existen ramas de la Física en las que, sus propiedades topológicas, o las diferentes ideas que, el significado de las mismas, sugieren, sirven como base a una distinta variedad de interpretaciones.

Así, en Física Cuántica, no existe separación entre comportamiento corpuscular (propio de las partículas componentes de la Materia) y ondulatorio (asociado a fenómenos de propagación). Un quantum puede manifestar un comportamiento ya corpuscular, ya ondulatorio.

La banda de Möbius puede ser utilizada para modelar el mundo físico de los quantum. Al igual que un quantum, unifica dos superficies: partícula (objeto) y onda (sujeto). Si se escribe en un lado del rectángulo 'onda' y en la otra cara del mismo 'partícula' y efectuamos una torsión de 180°, para obtener una banda de Möbius, se tiene un quantum onda-partícula.

Los quantum se clasifican en Fermiones y Bosones. Los primeros, que verifican el Principio de exclusión de Pauli, tienen un comportamiento estadístico según la Estadística de Fermi-Dirac, y su momento cinético intrínseco, llamado "spin", puede ser 1/2 ó -1/2, dependiendo de la orientación respecto a las líneas de fuerza del campo correspondiente.

Los Bosones, tienen un comportamiento estadístico según la Estadística de Bose-Einstein, y su "spin", es un número entero: 0 en los mesones, 1 ó -1 para los fotones, 2 ó -2

cuantos de gravitación, siendo éste el valor máximo obtenido para partículas elementales.

Si en vez de realizar una torsión de 180°, realizamos dos torsiones, es decir 360°, entre las diferentes formas obtenidas de la banda de Möbius de orden 2, existe una con un lóbulo hacia fuera y otro dirigido hacia adentro, en forma de dos cuasi-cilindros cuyas fronteras se interfieren. Esta banda es un modelo de un sistema fermiónico, considerando uno de los lóbulos como el spin positivo, y el otro, como el negativo.

La banda de Möbius, de torsión 360°, también puede modelar aspectos de los Bosones, de los Fermiones y de ambos. Y es posible todavía, conseguir modelos más complejos de Bosones 'wobble' (cambiantes), cuando se toma una torsión de 720°.

Aunque Bosones y Fermiones obedecen a estadísticas cuánticas distintas, existiría un nexo entre ellos, si hubiese una transformación que pasase de un Fermión a un Bosón (y viceversa). Concretamente, de una partícula a otra tal que sus spines difiriesen en 1/2. Así, de un electrón (spin 1/2) a una partícula de spin igual a 0 (es-electrón).

La simetría a la que daría lugar la anterior transformación, se conoce con el nombre de Super-simetría. Esta transformación no se ha conseguido todavía en la realidad física.

Referencias

[1] Atwater, H. A.: "Introduction to General Relativity". International Series in Natural Philosophy, ed. D. Ter Haar, 1974. New York.

[2] Benz, Walter: "Vorlesungen ber Geometrie der Algebren: Geometrie von Möbius, Laguerre-Lie, Minkowski, in einheitlicher und grundlogengeometrischer Behanding". Springer, 1973. Berlin.

[3] Dr, Arne: "Möbius functions, incidence algebras and Power series representations". Springer cop., 1986. Berlín.

[4] Fauvel, J. y otros: "Möbius and his band. Mathematics and Astronomy in ninetenneth-century". Oxford University Press, 1993. New York.

[5] Larson, R. E. y otros: "Cálculo". McGraw-Hill/ Interamericana de España, S. A. U., 1999. Madrid.

[6] Möbius, A. F.: "Gesammelte werke: Herausgegeben auf Veranlassung der Kniglich Schsischen Gesellschaft der Wissenschaften" hearausgegeben von R. Baltzer. Verlag von S. Hiszel, 1855. Berlín.

[7] Rey Pastor, J. y Babini, J.: "Historia de la Matemática". Gedisa, 2000. Barcelona.

[8] Sánchez del Río, C. (coord.): "Física cuántica". Eudema Universidad, 1991. Madrid.

[9] Schrder, E. M.: "Vorlesungen ber geometrie. Band 1, Möbiussche, elliptische und Hyperbolische ebenen". B I. Wissennchaftsverlag, cop 1991. Mannheim.

[10] Schutz, B.: "Geometrical methods of mathematical physics". Cambridge University Press, 1987. Cambridge.

[11] Singer, I. M. Y Thorpe, J. A.: "Lecture Notes on Elementary Topology and Geometry". V. R. Ramideran pub., 1996. University of Bangalore Press.

[12] Taton, R.: "Historia general de las Ciencias". Ediciones Orbis, 1988. Barcelona.

A continuación indicamos una serie de artículos sobre este tema, que pueden ser consultados en Internet.

- The Möbius strip. (Bellevue Community College. Science Division).
  `http://www.scidiv.bcc.ctc.edu/Math/mobius.html`
- A 3D Möbius Strip. (Doug Renselle).
  `http://www.quantonics.com/Level_4_Strip_Quanton_Latched_Left.html`
- The Möbius Strip. (Mark E. Soulson). `http://www.meson.org/topology/mobius.html`
- The Möbius Strip: an introduction to no orientable surfaces. (Math 655: Introduction to Topology. Univ. Ohio).
  `http://www.math.ohio-state.edu/~fiedorow`
- Index of history de Mathematicians. (School of Mathematical and Computtional Sciences. University of St. Andrews)
  `http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians`
- Geometry and the Imagination (John Conway, Peter Doyle, Jane Gilman, and Bill Thurston. Course en University of Minnesota Geometry Center. Junio 1991).
  `http://www.math.dartmouth.edu/~doyle/docs/gi/gi/gi.html`
- Particle and Astro-physics Challenge Hants Phenomenolism. (Lawrence H. Starkey. Twentieth World Congress of Philosophy. Boston.1998).
  `http://www.bu.edu/wcp/`

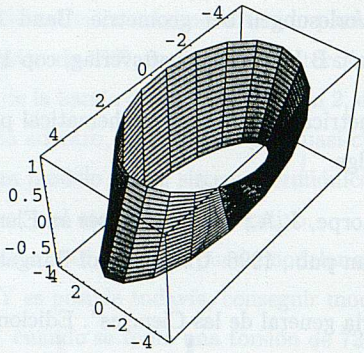

Figura 1.—

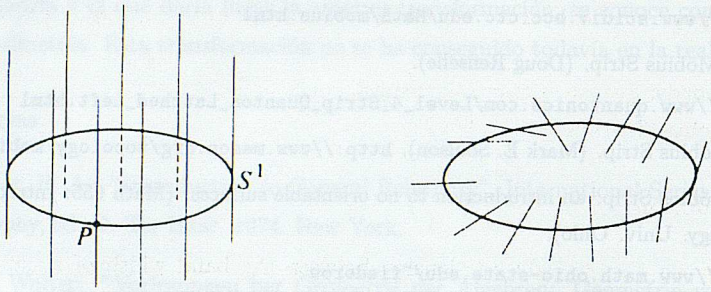

Figura 2.—

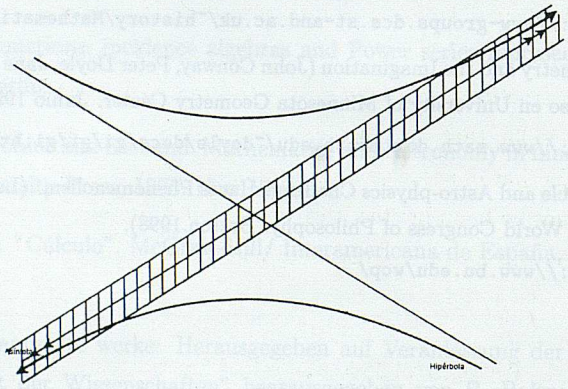

Figura 3.—

Rev. Academia de Ciencias. Zaragoza. 56 (2001)

New Shannon Sampling Recomposition

Lucía Agud

Departamento de Matemática Aplicada. Universidad de Zaragoza.

Raquel G. Catalán *

Departamento de Matemática e Informática. Universidad Pública de Navarra

Abstract

We will see that we can apply a new sampling theorem based on Shannon-Whittaker-Kotel'nikov's theorem to some kinds of band-limited signals using less samples by a time-unity than the rate given by Nyquist frequency associated to the signal.

Key words: Shannon's sampling theorem, family of unicity.

AMS classification: 94A12, 94A20.

1. Introduction

It is well known the strongly relationship between the band width of a band limited signal and the minimal frequency of sampling necessary for recomposing the signal from its values on a uniformly distributed sequence of points.

We will see (in theorem 2) that in a particular case of band limited signals the sampling frequency can be taken smaller than the Nyquist rate associated to the signal. This sampling reconstruction will need a unicity theorem based on the definition of unicity family introduced in [5], that we will adapt to our particular case. The practical importance of this result is that opens a way for making a Shannon type reconstruction of signals that even may not be band limited, through a limit process.

All along this paper, the couple of Fourier transforms in $L^1(\mathbb{R})$ [1]:

$$X(\nu) = \int_{-\infty}^{+\infty} x(t)e^{-j\nu t}dt, \qquad x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\nu)e^{j\nu t}dt,$$

will be used. We will call $\mathcal{B}(S)$ the set of functions in $L^2(\mathbb{R})$ which Fourier Transform support is in S. If S is bounded then we will say that the function is band-limited.

We remind the Shannon-Whittaker-Kotel'nikov sampling theorem [7], in a reformulation within our notation:

* Partially supported by the Spanish DGES PB97-1013, and PB98-0551

Theorem S-W-K. *Let $B > \sigma > 0$. If $x(t)$ is a signal (function) band-limited to $[-B, B]$ it can be reconstructed from its samples values at the points $\{t_k = k/(2\sigma); k \in \mathbb{Z}\}$ via the formula*

$$x(t) = \sum_{k=-\infty}^{\infty} x(t_k)\mathrm{sinc}(\frac{t}{\tau} - k) \tag{1}$$

where $\tau = 1/(2\sigma)$, with the series being absolutely and uniformly convergent on compact sets. $\tau_0 = \frac{1}{2B}$ is called the Nyquist rate.

In what follows, if the sums indexes vary in $\mathbb{Z}$, they will be omitted.

Definition. *It will be said that a family $\Lambda = \{\lambda_k\}_{k\in K} \subset \mathbb{R}$ is of unicity for $\mathcal{B}(S)$ [5], if for any signal $h(t) \in \mathcal{B}(S)$ we have that if $h(\lambda_k) = 0$ for all $k \in K$, then $h \equiv 0$.*

In [2] it is proved that given a set $\Lambda = \{\lambda_k\}_{k\in K} \subset \mathbb{R}$ there exists an upper bound ρ such that the family of trigonometric polynomials $\{e^{j\lambda_k\nu}\}$ spans the functions space $L^2((-r, r))$ $\forall r \leq \rho$.

Theorem 1. *Let $\omega \in \mathbb{R}$, $\tau > 0$, and $\rho = 1/(2\tau)$. The family of points $\{\lambda_k = \omega + k\tau; k \in \mathbb{Z}\}$, is of unicity for the signals in $\mathcal{B}((-r, r))$ $\forall r \leq \rho$.*

Proof:

Let us consider a band limited signal $x(t)$ with band width $r \leq \rho$, such that $x(\lambda_k) = 0$ $\forall k \in \mathbb{Z}$. By [2], there exist $(\alpha_k)_{k\in\mathbb{Z}} \subset \mathbb{R}$ such that $X(\nu)$ can be expressed in $L^2(\mathbb{R})$ as:

$$X(\nu) = \begin{cases} \sum \alpha_k e^{-j\lambda_k\nu} & \nu \in (-r, r) \\ 0 & \text{elsewhere.} \end{cases} \tag{2}$$

Denoting with $\Pi_r(\nu)$ the characteristic function of the interval $(-r, r)$, and $Y(\nu) = \sum \alpha_k e^{-j\lambda_k\nu}$, the Fourier Transform of the signal $x(t)$ will be expressed as $X(\nu) = Y(\nu)\,\Pi_r(\nu)$. Using (2), the signal $x(t)$ can be seen as the convolution product

$$x(t) = (y * 2r\mathrm{sinc}(2r\cdot))\,(t)$$

where $y(t) = \sum \alpha_k \delta_{\lambda_k}$. Therefore

$$x(t) = 2r \sum_k \alpha_k \mathrm{sinc}(2r(t - \lambda_k)).$$

As $x(\lambda_p) = 0$ $\forall p \in \mathbb{Z} \Longrightarrow$, $\sum_k \alpha_k \mathrm{sinc}(2r(\lambda_p - \lambda_k)) = 0$ for all $p \in \mathbb{Z}$.
Calling $\beta(t) = \mathrm{sinc}(2r\tau t)$ and using that $\lambda_k = \omega + k\tau$, the upwards expression can be written as

$$0 = \sum_k \alpha_k \beta(p - k) \tag{3}$$

$\forall p \in \mathbb{Z}$.

This expression can be seen as a convolution product of the sequences β and α, where $\alpha(k) = \alpha_k$, $\forall k \in \mathbb{Z}$, whose Z-transforms [4][6] will be denoted $B(z)$ and $A(z)$ respectively:

$$\begin{aligned} B(z) &= \sum_k \beta(k) z^{-k}, \\ A(z) &= \sum_k \alpha(k) z^{-k}. \end{aligned}$$

By (2), in the convolution product of (3), we have that $A(z)B(z) = 0 \; \forall z \in \mathbb{C}$. $B(z)$ is not the null function because $\beta(0) = 1$. So, by the analytic conti-nuation principle, $A \equiv 0$, what means that $\alpha(k) = 0 \;\; \forall k \in \mathbb{Z}$. Replacing in the reconstruction of x it is got that $x \equiv 0$. ∎

Observe that the recomposition function given by the sampling theorem is the only one verifying that $f(n\tau) = x(n\tau)$, i.e., x is the only signal that is recomposed in this way.

2. A Special Recomposition

Let us take a signal $x(t) \in \mathcal{B}([-B, B])$. Clearly the signal $y = x^p$ with p odd, is pB-BL. This means that we can recompose y directly applying Shannon's sampling theorem, or to recompose it through the signal x as

$$y(t) = \left[\sum_n (y(n\tau))^{\frac{1}{p}} \operatorname{sinc}\left(\frac{t}{\tau} - n\right) \right]^p. \tag{4}$$

Generally, taking a signal y band-limited such that $y^{\frac{1}{p}}(t)$ is also band-limited, the first signal can be recomposed using less samples by time-unity, through the second one.

As the new signal y has to be an analytic real function, all the samples have to be in the same 1-dimensional subspace. So, all the p roots in $\mathbb{C}$ will be taken in $\mathbb{R}$ without lost of generality. We see the importance of this fact in the following theorem.

Theorem 2. *Let $(s_n)_{n\in\mathbb{Z}} \in l^{2/p}(\mathbb{Z})$, $B \in \mathbb{R}^+$, $0 < \tau \leq \frac{1}{2B}$ and p odd. There exist exactly p signals x_k $0 \leq k \leq p-1$, such that $x_k^p \in \mathcal{B}([-pB, pB])$ and $x_k^p(n\tau) = s_n$.*

Proof: Let us call

$$x_0(t) = \sum_n s_n^{1/p} \operatorname{sinc}(2Bt - n) \tag{5}$$

where $s_n^{1/p}$ is the real p root of s_n.

From (5) we have that $x_0(n\tau) = s_n^{1/p}$, and $x_0 \in \mathcal{B}([-B, B])$. Take $x_k = \varepsilon_k x_0$, with $\{\varepsilon_k; 0 \leq k \leq p-1\}$ the p-roots of unity. Clearly these x_k are p different signals verifying the theorem.

Let us now suppose that there exists $y(t) \in \mathcal{B}([-B, B])$ such that $y^p \in \mathcal{B}([-pB, pB])$ and $y^p(n\tau) = s_n$. Then $y(n\tau) = s_n^{1/p}\varepsilon_k$ as $y(t)$ is an analytic function. Thus,

$$y(t) = \sum_n y(n\tau)\text{sinc}(2Bt - n) = \varepsilon_k \sum_n s_n^{1/p}\text{sinc}(2Bt - n) = \varepsilon_k x_0(t). \quad \blacksquare$$

The upper result is a particular case of signals u that can be expressed as $f(x(t))$, being $f \in L^2(\mathbb{R})$ biyective, x band limited, and with the band width of x smaller than the band width of u (Clearly all polynomials $p(t) = t^{2n+1}$, verify these properties). Then, using (1) for x, u can be expressed as

$$u(t) = f\left(\sum_{n=-\infty}^{\infty} f^{-1}(u(n\tau))\text{sinc}(\frac{t}{\tau} - n)\right). \tag{6}$$

This recomposition has the advantage of being got from more spaced samples (in the case of polynomials it is immediately verified as the spectrum of $p(x)$ is got by successive convolutions of the spectrum of x with itself). Even more, in general, given $x(t)$ a band-limited signal, f biyective, we have seen that a recomposition result can be found from samples of $f(x)$ although may be $f(x)$ was not band limited.

References

[1] L.C. Andrews, B.K. Shivamoggi. *Integral Transforms for Engineers and applied Mathematicians.* Ed. MacMillan (1988).

[2] A. Beurling, P. Malliavi *On the closure of characters and the zeros of entire functions.* The Institute for Advanced Study, Princeton, New Jersey. Vol. 118, pp. 79-93 (1967).

[3] R. Bracewell. *The Fourier Transform and its Applications.* Ed. McGraw-Hill (1978).

[4] C. Gasquet. *Analyse de Fourier et Applications.* Ed. Mason (1990).

[5] H.L. Landau. *Sampling, data transmision and the Nyquist rate.* IEEE Vol. 55, n. 10, pp. 1701-1706 (1967).

[6] B. Picinbono. *Theorie des signaux et des Systemes.* Ed. Dunod Université (1989).

[7] A. Zayed. *Advances in Shannon's Sampling Theory.* Ed. CRC Press (1993).

Rev. Academia de Ciencias. Zaragoza. 56 (2001)

ON PARTIAL QUASI BILATERAL GENERATING FUNCTION INVOLVING LAGUERRE POLYNOMIAL

M.C. MUKHERJEE
Netaji Nagar Vidyamandir
Calcuta - 700 092, INDIA

Abstrat:

In [1] the author introduced the term "partial quasi bilateral generating function" as follows

$$G(x,z,w)=\sum_{n=0}^{\infty} a_n p_{m+n}^{(\alpha)}(x) q_1^{(m+n)}(z) w^n$$

where $p_{m+n}^{(\alpha)}(x)$ and $q_1^{(m+n)}(z)$ are two special functions of orders m+n and 1 and of parameters α and m+n. In the present paper we shall show that the existence of a partial quasibilinear generating function involving Laguerre polynomial implies the existence of a more general generating function by means of one parameter group of continuous tranformations.

1.-Introduction:

In [2] the authors have proved the following theorem as bilateral generating function involving Laguerre polynomial.

Theorem 1: If there exists a bilateral generating relation of the form

$$G(x,w)=\sum_{n=0}^{\infty} a_n L_n^{(\alpha)}(x) w^n \tag{1.1}$$

then

$$(1+w)^{\alpha}\exp(-wx)G(x(1+w),wv) = \sum_{n=0}^{\infty} w^n \sigma_n(x,v) \tag{1.2}$$

where $\sigma_n(x,v)=\sum_{q=0}^{n} a_q \binom{n}{q} L_n^{(\alpha-n+q)}(x) v^q$

The above mencioned theorem 1 can be extended by the present author [3] in the following way:

Theorem 2: If there exists a quasi bilinear generating relation of the form

$$G(x,u,w)=\sum_{n=0}^{\infty} a_n L_n^{(\alpha)}(x) L_m^{(n)}(u) w^n \tag{1.3}$$

then

$$(1.4)\quad (1+wy^{-1}z)^{\alpha}\exp(-w(xy^{-1}z+t))G(x(1+wy^{-1}z),u+wt,wztv)$$

$$=\sum_{n=0}^{\infty}\sum_{q=0}^{\infty}a_n(wv)^n\frac{(wy^{-1}z)^q}{q!}(n+1)_q L_{n+q}^{(\alpha-q)}(x)(zt)^n\sum_{p=0}^{\infty}L_m^{(n+p)}(u)\frac{(-tw)^p}{p!}$$

In this present paper, the present author has generalised the <u>theorem 2</u> in the following way:

<u>Theorem 3</u>: If there exists a partial quasi bilinear generating function of the form

$$(1.5)\quad G(x,u,w)=\sum_{n=0}^{\infty}a_n w_n L_{m+n}^{(\alpha)}(x)L_p^{(m+n)}(u)$$

then the following more general generating function can be obtained as follows

$$(1.6)\quad (1+wy^{-1}z)^{\alpha}\exp(-w(t+xy^{-1}z))G(x(1+wy^{-1}z),u+wt,wvzt)$$

$$=\sum_{n=0}^{\infty}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty}\frac{a_n(m+n+1)_q}{q!r!}(wvzt)^n(wy^{-1}z)^q L_{m+n+q}^{(\alpha-q)}(x)L_p^{(m+n+r)}(u)(-wt)^r$$

2.- Proof of theorem 3.

We now consider the following operators [4]

$$(2.1)\quad R_1=xy^{-1}z\frac{\partial}{\partial x}+z\frac{\partial}{\partial y}-xy^{-1}z \qquad \text{and} \qquad (2.2)\quad R_2=t\frac{\partial}{\partial x}-t.$$

Also $\quad (2.3)\quad R_1\left(L_{m+n}^{\alpha}(x)y^{\alpha}z^n\right)=(m+n+1)L_{m+n+1}^{(\alpha-1)}(x)y^{\alpha-1}z^{n+1}$

and $\quad (2.4)\quad R_2\left(L_p^{(m+n)}(n)t^n\right)=-L_p^{(m+n+1)}(u)t^{n+1}$

such that $\quad (2.5)\quad e^{wR_1}f(x,y,z)=\exp(-wxy^{-1}z)f(x(1+wy^{-1}z),y+wz,z)$

and $\quad (2.6)\quad e^{wR_2}f(u,t)=\exp(-wt)f(u+wt,t)$

We consider the following generating relation

$$(2.7)\quad G(x,u,w)=\sum_{n=0}^{\infty}a_n w^n L_{m+n}^{(\alpha)}(x)L_p^{(m+n)}(u)$$

Replacing w by $wztv$ and then multiplying both sides of (2.7) by y^{α} we get

$$(2.8)\quad y^{\alpha}G(x,u,wvzt)=\sum_{n=0}^{\infty}a_n(wv)^n\left(L_{m+n}^{(\alpha)}(x)y^d z^n\right)\left(L_p^{(m+n)}(u)t^n\right)$$

Operating $e^{wR_1}\ e^{wR_2}$ on both sides of (2.8) we obtain

$$(2.9)\quad e^{wR_1}e^{wR_2}\left(y^{\alpha}G(x,v,wvzt)\right) = e^{wR_1}e^{wR_2}\left[\sum_{n=0}^{\infty}a_n(wv)^n\left(L_{m+n}^{(\alpha)}(x)y^{\alpha}z^n\right)\left(L_p^{(m+n)}(u)t^n\right)\right]$$

The lest member of (2.9), with the help of (2.5) and (2.6), becomes

$$(2.10)\quad e^{wR_1}e^{wR_2}\left[y^{\alpha}G(x,u,wvzt)\right] = e^{wR_1}\left[\exp(-wt)\left(y^{\alpha}G(x,u+wt,wvzt)\right)\right]$$

$$= y^{\alpha}\left(1+wy^{-1}z\right)^{\alpha}e^{-w(t+xy^{-1}z)}G\left(x\left(1+wy^{-1}z\right),u+wt,wvzt\right)$$

On the other hand the right member of (2.9), with the help of (2.3) and (2.4), becomes

$$(2.11)\quad e^{wR_1}e^{wR_2}\left[\sum_{n=0}^{\infty}a_n(wv)^n\left(L_{m+n}^{(\alpha)}(x)y^{\alpha}z^n\right)\left(L_p^{(m+n)}(u)t^n\right)\right]$$

$$= \sum_{n=0}^{\infty}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty}a_n(wv)^n\frac{w^q}{q!}.\frac{w^r}{r!}(m+n+1)_q L_{m+n+q}^{(\alpha-q)}(x)y^{\alpha-q}z^{n+q}(-1)^r L_p^{(m+n+r)}(u)t^{n+r}$$

Equating (2.10) and (2.11), we obtain

$$(2.12)\quad \left(1+wy^{-1}z\right)^{\alpha}\exp\left(-w\left(t+xy^{-1}z\right)\right)G\left(x\left(1+wy^{-1}z\right),u+wt,wvzt\right)$$

$$= \sum_{n=0}^{\infty}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty}\frac{(-1)^r a_n(m+n+1)_q}{q!r!}(wvzt)^n\left(wy^{-1}z\right)^q L_{m+n+q}^{(\alpha-q)}(x)L_p^{(m+n+r)}(u)(wt)^r$$

which is our desired result.

Special cases.

Case(i): Putting m=0 in the above mentioned result (2.12), we get

If $\quad G(x,u,w) = \sum_{n=0}^{\infty}a_n L_n^{(\alpha)}(x)L_p^{(n)}(u)w^n$

then $\quad \left(1+wy^{-1}z\right)^{\alpha}\exp\left(-w\left(t+xy^{-1}z\right)\right)G\left(x\left(1+wy^{-1}z\right),u+wt,wvzt\right)$

$$= \sum_{n=0}^{\infty}\sum_{q=0}^{\infty}a_n(wvzt)^n\frac{\left(wy^{-1}z\right)^q}{q!}(n+1)_q L_{n+q}^{(\alpha-q)}(x)\sum_{r=0}^{\infty}\frac{(-wt)^r}{r!}L_p^{(n+r)}(u),$$

which is theorem 2.

Case (ii): Putting $y = z = t = 1$ and $p = 0$ in the above mentioned result (2.12), we get the following result:

If $\quad G(x,w) = \sum_{n=0}^{\infty}a_n L_n^{(\alpha)}(x)w^n$

then

$$(1+w)^{\alpha}\exp(-w(x+1))G(x(1+w),wv)$$

$$=\sum_{n=0}^{\infty}\sum_{q=0}^{\infty}a_n(wv)^n.\frac{w^q}{q!}(n+1)_q L_{n+q}^{(\alpha-q)}(x)\sum_{r=0}^{\infty}\frac{(-w)^r}{r!}L_0^{(n+r)}(u).$$

Setting $L_0^{(n+1)}(u)=1$ and simplifying we get

$$(1+w)^{\alpha}\exp(-wx)\,G(x(1+w),wv)$$

$$=\sum_{n=0}^{\infty}\sum_{q=0}^{n}w^n a_{n-q}\frac{(n-q+1)_q}{q!}L_n^{(\alpha-n+q)}(x)v^{n-q}=\sum_{n=0}^{\infty}w^n\sigma_n(x,v)$$

where

$$\sigma_n(x,v)=\sum_{q=0}^{n}a_q\binom{n}{q}L_n^{(\alpha-n+q)}(x)v^q$$

which is the <u>theorem 1</u>.

References:

[1] A.K. Mandal (1992-93): Pure Math. Manuscript Vol. 10, 73-79.

[2] R. Sarma and A.K. Chongdar (1990: Some generating functions of Laguerre polynomial from Lie group view point. Bull. Cal. Math. Soc. 82, p. 527.

[3] M.C. Mukherjee: An extension of bilateral generating function of certain special function 1. Com.

[4] A.K. Chongdar (1989): Some generating functions involving Laguerre polynomial. Bull. Cal. Math. Soc. 76, p. 262.

Rev. Academia de Ciencias. Zaragoza. **56** *(2001)*

Une approche algébrique du problème de l'idéal fermé

R. Choukri

Abstract

We give some algebraic properties of topologically simple Banach algebras

Introduction. Une algèbre de Banach commutative est dite topologiquement simple si elle n'admet pas d'idéaux fermés propres ([2], définition 18, p.166). Deux exemples "triviaux" de telles algèbres sont: l'algèbre de Banach usuelle $\mathbf{C}$ et l'espace de Banach usuel $\mathbf{C}$ muni du produit trivial. La question se pose alors sur l'existence d'algèbres de Banach commutatives topologiquement simples non triviales. Plusieurs auteurs se sont intéréssés à la question et ils l'ont traitée de manières différentes (voir [4], [6],...). Le problème de l'existence d'une telle algèbre (dit problème de l'idéal femé) est toujours ouvert. Dans cette note, nous nous intéréssons aux algèbres topologiquement simples d'un point de vue différent. Plus précisement, nous dégageons certaines propriétés algébriques de telles algèbres.

Préliminaires. Soit A une algèbre commutative unitaire. On appelle hauteur d'un idéal premier P de A le plus grand des entiers n, eventuellement infini, tel qu'il existe une suite finie $(P_i)_{0\leq i\leq n}$ d'idéaux premiers de A, deux à deux distincts, vérifiant $P_0 \subset ... \subset P_{n-1} \subset P_n = P$. La dimension de Krull de A est le supermum des hauteurs de ses idéaux premiers ([8], p.71). On suppose que A est intègre et on note par K son corps des fractions. Un élément x de K est dit quasi-entier sur A si le sous-A-module $A[x]$ de K est fractionnaire, i.e., il existe $d \in A\setminus\{0\}$ vérifiant $dA[x] \subset A$ ([3], p.195). L'ensemble de tels éléments, noté A^*, est appelé la quasi-clôture de A. C'est une sous-algèbre de K contenant A. Si $A = A^*$, on dira que A est complétement intégralement close . L'algèbre A est dite de valuation si la famille de ses idéaux est totalement ordonnée pour l'inclusion ([3], p.85). La dimension de Krull d'une telle algèbre est appelée aussi hauteur. Si P est un idéal premier de A, on notera par A_P la sous-algèbre de K formée des $x \in K$ pour lesquels il existe $s \notin P$, dépendant de x, tel que $sx \in A$. L'idéal de A_P engendré par P est noté PA_P. Si B est une algèbre de Banach commutative intègre, on notera:

i) $S(B) = \{(x_n)_n \subset B,\ \exists d \in B \setminus \{0\},\ (dx_n)_n$ convergente$\}$.
ii) $S_0(B) = \{(x_n)_n \subset B,\ \exists d \in B \setminus \{0\},\ (dx_n)_n$ convergente vers zéro$\}$.
Il est facile de voir que $S(B)$ est une algèbre et que $S_0(B)$ en est un idéal.

Propriétés algébriques des algèbres topologiquement simples.

Le théorème suivant regroupe certaines propriétés algébriques des algèbres de Banach commutatives topologiquement simples non triviales.

Théorème. Soit A une algèbre de Banach commutative topologiquement simple non triviale et $A^{\neq}$ l'algèbre obtenue par adjonction d'une unité à A. Alors:

1) $A^{\neq}$ est locale, i.e., admet un seul idéal maximal et, de plus, elle est intègre.
2) Pour toute suite $(I_n)_{n\geq 0}$ d'idéaux non nuls de $A^{\neq}$, l'idéal $\bigcap_{n\geq 0} I_n$ est non nul.
3) La quasi-clôture de $A^{\neq}$ est égale à son corps des fractions K.
4) Tout idéal premier non nul de $A^{\neq}$ est de hauteur infinie. En particulier, $A^{\neq}$ est de dimension de Krull infinie.
5) $A^{\neq}$ n'est contenue dans aucune algèbre de valuation, contenue dans K, de dimension de Krull finie autre que K.
6) La famille des idéaux premiers non nuls de $A^{\neq}$ est non dénombrable.
7) L'algèbre quotient $S(A)/S_0(A)$ est isomorphe au corps des fractions L de A. En particulier, l'algèbre $S(A)/S_0(A)$ est un corps.

Preuve. 1) L'algèbre A est nécessairement radicale, car, sinon, le noyau d'un caractère non nul constituera un idéal fermé propre. D'où la localité de $A^{\neq}$. Comme A est nécessairement non unitaire, pour montrer que $A^{\neq}$ est intègre, il suffit de montrer que A l'est. Supposons qu'ils existent $x, y \in A$, non nuls, vérifiant $xy = 0$. Alors $Annx = \{a \in A,\ ax = 0\}$ est un idéal fermé non nul de A. Il est donc égal à A. Ainsi on a $Ax = \{0\}$. L'idéal $I = \{a \in A,\ Aa = \{0\}\}$, qui est fermé, est alors non nul; il est également égal à A. D'où $A^2 = \{0\}$; et par conséquent, A doit être de dimension 1 et A sera triviale.

2) Soit $(I_n)_{n\geq 0}$ une suite d'idéaux non nuls de $A^{\neq}$. Evidemment, on peut supposer que les I_n sont propres. Comme $A^{\neq}$ est locale, les I_n sont contenus dans A. Pour $n \geq 0$, soit $a_n \in I_n$ non nul et f_n l'application définie dans A par $f_n(x) = a_n x$. Le théorème de Mittag-Lefler ([5], théorème 5.3, p.147) appliqué à la suite (f_n) montre que l'idéal $\bigcap_{n\geq 0} a_0 ... a_n A$ est non nul. Donc $\bigcap_n I_n$ n'est pas nul.

3) Soit $a \in A^{\neq}$, non nul. Par l'assertion 2), l'idéal $\bigcap_n a^n A^{\neq}$ est non nul. Considérons d dans $a^n A^{\neq} \setminus \{0\}$. Pour tout $n \geq 0$, il existe x_n appartenant à $A^{\neq}$ tel que $d = a^n x_n$. Alors, il est clair que a^{-1} est dans $(A^{\neq})^*$. Par ailleurs, $A^{\neq}$ est

contenue dans $(A^{\neq})^*$. D'où $(A^{\neq})^* = K$, vu que $(A^{\neq})^*$ est une algèbre.

4) Supposons que $A^{\neq}$ admet un idéal premier non nul P de hauteur finie n. D'abord $n \geq 1$ puisque l'idéal premier P contient strictement l'idéal premier $\{0\}$. Considérons une suite $(P_i)_{0 \leq i \leq n}$ d'idéaux premiers de $A^{\neq}$, deux à deux distincts, tel que $P_0 \subset ... \subset P_{n-1} \subset P_n = P$. Alors, il est clair que P_1 est de hauteur 1. Considérons $a \in P_1$ non nul. D'après ([1], p.11), on a $\bigcap_n a^n A^{\neq}_{P_1} = \{0\}$. Par ailleurs, on a $\{0\} \neq \bigcap_n a^n A^{\neq} \subset \bigcap_n a^n A^{\neq}_{P_1} = \{0\}$; ce qui est contradictoire.

5) Supposons qu'il existe une algèbre de valuation V contenue dans K, autre que K, contenant $A^{\neq}$ de dimension de Krull finie. Comme V n'est pas un corps, sa dimension de Krull n'est pas nulle. En raisonnant de la même façon que dans 4), on met en evidence un idéal premier Q de V de hauteur 1. Par ailleurs, QV_Q est l'unique idéal premier non nul de V_Q ([7], corollaire, p.21). Et, par conséquent, V_Q est de dimension de Krull égale à 1. D'autre part, par ([3], proposition 1, p.106), elle est de valuation. D'après ([3], proposition 9, p.113), elle est complétement intégralement close. Cependant, on a $K = \left(A^{\neq}\right)^* \subset V_Q^* = V_Q$; ce qui est contradictoire.

6) Supposons que la famille des idéaux premiers non nuls de $A^{\neq}$ est dénombrable et considérons un élément x non nul dans leur intersection. Alors $A^{\neq}[x^{-1}]$ n'admet pas d'idéaux premiers non nuls, car, sinon, la trace sur $A^{\neq}$ d'un tel idéal est un idéal premier non nul de $A^{\neq}$ et contiendrait alors x qui est inversible dans $A^{\neq}[x^{-1}]$. Ainsi $A^{\neq}[x^{-1}]$ est un corps et par suite on a $K = A^{\neq}[x^{-1}]$. Considérons maintenant Ω la famille des sous-algèbres de K, distinctes de K, contenant $A^{\neq}$. On montre facilement qu'elle est inductive pour l'inclusion. Soit V un élément maximal de Ω. Alors V est une algèbre de valuation de dimension de Krull égale à 1 ([3], proposition 6, p.111). Ce qui est en contradiction avec l'assertion 5).

7) Soient $u, v, u', v' \in A$, où v, v' non nuls tels que $uv' = u'v$ et $(x_n)_n, (y_n)_n$ deux suites d'éléments de A telles que $vx_n \to u$ et $v'y_n \to u'$. Alors on a $vv'(x_n - y_n) \to 0$. Il en résulte que $(x_n - y_n)_n \in S_0(A)$. Posons , pour tout $u/v \in L$, $\varphi(u/v) = \widetilde{(x_n)}$, où (x_n) est une suite d'éléments de A telle que $vx_n \to u$. Par ce qui précède, φ est bien une application. On montre alors aisément qu'elle est un morphisme d'algèbres (non nul). Comme L est un corps, le noyau de φ, qui est un idéal de L, est nul. Quant à la surjection, elle découle du fait que tout idéal non nul de A est dense.

Remarque. De point de vue topologique, l'absence des idéaux fermés propres est "peu appréciée". De point vue algébrique, certaines assertions du théorème précédent, notamment 3) et 4), sont également "peu appréciées".

Remerciement. L'auteur remercie vivement le Professeur M. Oudadess pour ses remarques et commentaires.

Références

[1] **D. D. Anderson**. "The Krull intersection theorem". Pacific Journal of Mathematics, vol 57, n°1, 1975, p.11.

[2] **F. F. Bonsall et J. Duncan**. "Complete Normed Algebras". Berlin Heidelberg, New- York, 1973.

[3] **N. Bourbaki**. "Algèbre Commutative". Chapitres 5 à 7. Masson, Paris. 1985.

[4] **J. Cusack**. "Automatic continuity and topologically simple Banach algebras." J. London. Math. Soc. (2). 16 (1977). 493-500.

[5] **H. G. Dalles**. "Automatic continuity: A survey". Bull. Lond. Math. Soc. 10. 129-183. (1978).

[6] **J. Esterle**. "Radical Banach Algebras and Automatic Continuity". Lecture Notes in Mathematics. 975. 1983. p.66.

[7] **J. P. Lafon**. "Algèbre Commutative." Langages géometrique et algébrique. Hermann. Paris. 1977.

[8] **H. Matsumura**. "Commutative Algebra". Second Edition. 1980.

Adresse. E.N.S. Takaddoum. Département de Mathématiques. B.P. 5118. Rabat. 10105. Maroc.

Rev. Academia de Ciencias. Zaragoza. **56** *(2001)*

Sur deux théorèmes de Katznelson

A. EL KINANI

Abstract

We show that if A is a Banach algebra, then all functions operate spectraly on A if, and only if, $A/RadA$ is finite dimensional. We also prove that a Banach algebra has the property that every function opearte weakly if, and only if, it is isomorphic to C^n, $n \in N$. Furtheremore, we show that C^*-algebras are exactly Banach algebras in which all continuous functions operate weakly.

Mots clés et phrases: Transformée de Fourier Gelfand, Fonction opérant sur une algèbre, C^*-algèbre.
1999 Mathematics Subject Classification. 46J35, 46K05.

I. Introduction. Soient $(A, \|.\|)$ une algèbre de Banach complexe unitaire commutative semi-simple et D une partie de C. On dit qu'une fonction $f\colon D \longrightarrow C$ opère, au sens de Katznelson ([3]) sur A, si $f \circ \hat{x}$ est une transformée de Fourier-Gelfand d'un élément de A, pour tout x élément de A dont le spectre est contenu dans D, où $\hat{x}$ désigne la transformée de Fourier-Gelfand de x.

Dans [3], Y. Katznelson a montré les deux résultats suivants.

Théorème 1. Soit $(A, \|.\|)$ une algèbre de Banach complexe unitaire commutative et semi-simple. Une condition nécessaire et suffisante pour que toute fonction opère, sur A, est que A soit de dimension finie.

Théorème 2. Soit $(A, \|.\|)$ une algèbre de Banach complexe unitaire commutative et semi-simple. Une condition nécessaire et suffisante pour que toute fonction continue opère, sur A, est que A soit l'algèbre de toutes les fonctions continues sur son espace des caractères (non nuls).

Dans cette article, nous étendons les deux théorèmes précédents aux algèbres de Banach qui ne sont pas nécessairement commutatives et semi-simples. Pour ce faire, nous introduisons les deux notions, de fonction qui opère, suivantes.

Définitions I.1. Soient $(A, \|.\|)$ une algèbre de Banach complexe unitaire, D une partie de C et $f\colon D \longrightarrow C$.

1) On dit que f opère spectralement sur A si, pour tout $a \in A$ dont le spectre est contenu dans D, il existe un élément dans A, noté $f(a)$, tel que $Spf(a) = f(Spa)$.

2) On dit que f opère faiblement sur A si, pour tout $a \in A$ dont le spectre est contenu dans D, il existe un élément unique dans A, noté $f(a)$, contenu dans la bicommutante, notée $\Gamma(\Gamma(a))$, de $\{a\}$, dans A, tel que $\widehat{f(a)} = f \circ \hat{a}$ sur $\mathcal{M}(\Gamma(\Gamma(a)))$, où $\mathcal{M}(\Gamma(\Gamma(a)))$ est l'ensemble des caractères de $\Gamma(\Gamma(a))$.

Remarque I.2. Si $f\colon D \longrightarrow C$ opère faiblement sur A, alors f opère spectralement sur A. Si de plus $(A, \|.\|)$ est une algèbre de Banach complexe unitaire commutative et semi-simple, alors $f\colon D \longrightarrow C$ opère faiblement sur A si, et seulement si, f opère au sens de Katznelson ([3]) sur A.

Nous montrons qu'une condition nécessaire et suffisante pour que toute fonction opère spectralement sur une algèbre de Banach unitaire est que cette algèbre soit de dimension finie modulo son radical de Jacobson. Nous prouvons aussi que les algèbres C^n, $n \in N$, sont exactement les algèbres de Banach unitaires sur lesquelles toute fonction opère faiblement. Enfin, nous caractérisons les C^*-algèbres, parmi les algèbres de Banach unitaires, par le fait que toute fonction continue opère faiblement sur de telles algèbres.

Dans toute la suite, les algèbres considérées sont complexes, unitaires et non nécessairement commutatives. Pour tout élément x d'une algèbre unitaire A, le spectre de x, noté $Sp_A x$ (ou tout simplement Spx si aucune confusion n'est à craindre) est l'ensemble $Spx = \{\lambda \in C : \lambda e - x \text{ est non inversible dans } A\}$. Le rayon spectral de x est donné par $\rho(x) = \sup\{|\lambda| : \lambda \in Spx\}$.

II. Caractérisations des algèbres par les fonctions qui opèrent sur elles.

Soit $(A, \|.\|)$ une algèbre de Banach commutative et semi-simple. Y. Katznelson ([3]) a montré que si toute fonction opère au sens de [3], sur A, alors l'espace $\mathcal{M}$ des idéaux maximaux de A doit être compact et discret. Dans le cas d'une algèbre de Banach quelconque, on obtient ce qui suit.

Théorème II.1. Soit $(A, \|.\|)$ une algèbre de Banach. Les assertions suivantes sont équivalentes.

1) Toute fonction opère spectralement sur A.

2) $A/RadA$ est de dimension finie.

Preuve. 1)$\Longrightarrow$2). On va montrer que Spa est fini, pour tout $a \in A$ et on conclut par un résultat de Kaplansky ([2]). Supposons qu'il existe $a \in A$ tel que Spa soit infini; et soit $(\lambda_n)_n$ une suite d'éléments, de Spa, distincts deux à deux. On considère la fonction f: $C \longrightarrow C$ définie par $f(\lambda_n) = n$, pour tout n, et f est nulle ailleurs. Comme f opère spectralement sur A, il existe $f(a) \in A$ tel que $Spf(a) = f(Spa)$. Par suite $\rho(f(a)) \geq n$, pour tout n; contradiction.

2)$\Longrightarrow$1). Soient f: $D \longrightarrow C$ et $a \in A$ tel que $Sp_A a \subset D$. Comme $Sp_A a = Sp_{A/RadA} s(a)$, où s: $A \longrightarrow A/RadA$ est la surjection canonique, il existe λ_1, λ_2, ..., $\lambda_n \in C$ tel que $Sp_A a = \{\lambda_1, \lambda_2, ..., \lambda_n\}$. Soit P un polynôme tel que $P(\lambda_k) = f((\lambda_k)$, pour tout $k = 1, 2, ..., n$. Posons $f(a) = P(a)$. Alors $Spf(a) = f(Spa)$.

Comme conséquence, nous avons les corollaires suivants.

Corollaire II.2. Soit $(A, \|.\|)$ une algèbre de Banach commutative. Une condition nécessaire et suffisante pour que toute fonction opère spectralement, sur A, est que $A/RadA = C^n$, où $n \in N^*$.

Corollaire II.3. ([3], Théorème 1). Soit $(A, \|.\|)$ une algèbre de Banach commutative et semi-simple. Une condition nécessaire et suffisante pour que toute fonction opère au sens de [3], sur A, est que A soit de dimension finie.

Preuve. Si toute fonction opère au sens de [3] sur A, alors, par la remarque I.2 et le théorème II.1, A est de dimension finie vue qu'elle est semi-simple. Réciproquement, si A est commutative, semi-simple et de dimension finie, alors elle est isomorphe à C^n, où $n = \dim A$. Si $\{e_1, e_2, ..., e_n\}$ désigne la base canonique de C^n, alors ses caractères sont exactement χ_1, $\chi_2, ..., \chi_n$ tels que $\chi_i(e_j) = \delta_{ij}$. Soient maintenant $a \in A$ et f une fonction définie dans une partie D, de C, tel que $Spa \subset D$. Soient $\lambda_1, \lambda_2, ..., \lambda_n$ dans C tels que $a = \sum_{i=1}^{n} \lambda_i e_i$. On pose $f(a) = f(\chi_1(a))e_1 + f(\chi_2(a))e_2 + ... + f(\chi_n a))e_n$. Il est clair que $f(a) \in A$ et que $\hat{f(a)} = f \circ \hat{a}$. Ainsi f opère au sens de [3] sur A.

Les algèbres de Banach unitaires sur lesquelles toute fonction opère faiblement sont caractérisées comme suit.

Théorème II.4. Soit $(A, \|.\|)$ une algèbre de Banach. Les assertions suivantes sont équivalentes.

1) Toute fonction opère faiblement sur A.

2) A est isomorphe (algébriquement et topologiquement) à C^n, où $n \in N^*$.

Preuve. 1)$\Longrightarrow$2). Par la remarque I.2 et le théorème II.1, $A/RadA$ est de dimension finie. Ensuite, en utilisant le fait que la fonction $Z : \lambda \longrightarrow \lambda$ opère faiblement, sur A, on montre que A est sans éléments quasi-nilpotents. Elle

est donc semi-simple. De plus elle est commutative par [4] car la compacité de la boule unité de A et la semi-continuité supérieure de la fonction $x \longrightarrow \rho(x)$ montrent qu'il existe $\alpha > 0$ tel que $\|x\| \leq \alpha\rho(x)$, pour tout $x \in A$. Ainsi A est commutative, semi-simple et de dimension finie. Elle est donc isomorphe à C^n, où $n = \dim A$. Comme C^n ne posséde qu'une seule topologie d'algèbre de Banach, $A = C^n$ (algébriquement et topologiquement).

2)$\Longrightarrow$1). Soient f une fonction définie dans une partie D de C et $a \in A$ tel que $Spa \subset D$. Le raisonnement du corollaire II.3 dans l'algèbre $\Gamma(\Gamma(a))$ montre qu'il existe un élément $f(a)$ dans $\Gamma(\Gamma(a))$ tel que $\widehat{f(a)} = f \circ \hat{a}$ sur $\mathcal{M}(\Gamma(\Gamma(a)))$. L'unicité de $f(a)$ découle du fait que l'algèbre $\Gamma(\Gamma(a))$ est semi-simple vue que A est sans éléments quasi-nilpotents.

Remarque II.5. Soit $(A, \|.\|)$ une algèbre de Banach commutative et semi-simple. Les assertions suivantes sont équivalentes.

1) A est isomorphe à C^n, où $n \in N^*$.

2) Toute fonction opère au sens de [3] sur A.

3) Toute fonction opère faiblement sur A.

4) Toute fonction opère spectralement sur A.

On sait ([1]) que les fonctions continues opèrent faiblement sur les C^*-algèbres unitaires commutatives. Ces dernières sont exactement les algèbres de Banach unitaires sur lesquelles toute fonction continue opère faiblement comme le montre ce qui suit.

Théorème II.6. Soit $(A, \|.\|)$ une algèbre de Banach. Les assertions suivantes sont équivalentes.

1) Toute fonction continue opère faiblement sur A

2) (A, ρ) est une C^*-algèbre commutative.

Preuve. Il reste à montrer **1)$\Longrightarrow$2).** Montrons tout d'abord que A est hermitienne. En effet soit $a \in A$. Comme la fonction $\overline{z} : \lambda \longrightarrow \lambda$ opère faiblement, sur A, il existe un élément unique $\overline{z}(a)$, contenu dans $\Gamma(\Gamma(a))$, tel que $\widehat{\overline{z}(a)} = \overline{z} \circ \hat{a}$ sur $\mathcal{M}(\Gamma(\Gamma(a)))$. On a donc une aplication $*$ de A dans A définie par $a \longrightarrow a^* = \overline{z}(a)$. Un calcul simple montre que $*$ est une involution sur A. De plus, pour tout $\chi \in \mathcal{M}(\Gamma(\Gamma(a)))$, on a $\chi(x^*) = \widehat{\overline{z}(x)}(\chi) = \overline{\chi(x)}$. Si maintenant h est un élément hermitien de A, alors

$$Sp_A h = Sp_{\Gamma(\Gamma(h))} h \subset R.$$

Ainsi A est hermitienne. Par ailleurs A est commutative car tous ses éléments sont normaux vu que $x^* \in \Gamma(\Gamma(x))$, pour tout $x \in A$. De plus, comme dans la preuve du théorème II.4, on montre que A est semi-simple. Ainsi A est une algèbre hermitienne, commutative et semi-simple. Soient maintenant $a \in A$ et

Φ_a l'application, de $C(Sp_A a)$ dans A, définie par $\Phi_a(h) = h(a)$. La linéarité de la transformation de Gelfand et le fait que A est semi-simple montre que Φ_a est un morphisme d'algèbres. Montrons que Φ_a est continu. Pour cela, montrons que son graphe est fermé. Soit $(g_n)_n$ une suite d'éléments de $C(Sp_A a)$ telle que $g_n \underset{n}{\longrightarrow} g$ uniformément dans $Sp_A a$ et $g_n(a) \underset{n}{\longrightarrow} y$. Si on pose $h_n = g_n - g$, alors $h_n \circ \widehat{a} \underset{n}{\longrightarrow} 0$ car $\|h_n\| \underset{n}{\longrightarrow} 0$ et $\|h_n\| = \|h_n \circ \widehat{a}\|$. Or $h_n \circ \widehat{a} = \widehat{h_n(a)}$. Donc $\widehat{h_n(a)} \underset{n}{\longrightarrow} 0$. Et comme la transformation de Gelfand est continue et A est semi-simple, on a $y = g(x)$. Soient $a \in A$ et $PL(a, a^*)$ la sous-algèbre involutive pleine fermée engendrée par a. D'aprés ce qui précede, $PL(a, a^*)$ est une sous algèbre commutative semi-simple et hermitienne. Montrons que $PL(a, a^*)$ est une C^*-algèbre, pour tout $a \in A$. En effet l'homéomorphisme

$$\widehat{a} : \mathcal{M}(PL(a, a^*)) \longrightarrow Sp_A a$$

définit, par transposition, un $*$-morphisme d'algèbres unitaires, isométrique et surjectif

$$\psi : C(Sp_A a) \longrightarrow C(\mathcal{M}(PL(a, a^*)))$$

donné par $\psi(f) = f \circ \widehat{a}$. Par ailleurs la transformation de Gelfand

$$J_a : PL(a, a^*) \longrightarrow C(\mathcal{M}(PL(a, a^*)))$$

est $*$-morphisme d'algèbres. Elle est injective du fait que $PL(a, a^*)$ est semi-simple. Montrons qu'elle est surjective. Pour $g \in C(\mathcal{M}(PL(a, a^*)))$, il existe $f \in C(Sp_A a)$ telle que $f \circ \widehat{a} = g$. Comme f est continue sur $Sp_A a$, il existe une suite de polynômes $(p_n(z, \overline{z}))_n$ telle que $f = \lim_n p_n$ uniformément sur $Sp_A a$. D'où, d'aprés ce qui précède, $\lim_n p_n(a, a^*) = f(a)$. Par ailleurs $(p_n(a, a^*))_n \subset PL(a, a^*)$. D'ou $f(a) \in PL(a, a^*)$ car $PL(a, a^*)$ est fermé. De plus, pour tout $\chi \in \mathcal{M}(PL(a, a^*))$, on a $g(\chi) = \chi(f(a))$. D'où $g = J_a(f(a))$, et par suite $PL(a, a^*) = C(\mathcal{M}(PL(a, a^*)))$. Ainsi $(PL(a, a^*), \rho)$ est une C^*-algèbre. Il s'ensuit que, pour tout élément hermitien h de A, il existe une constante $M_h > 0$ telle que

$$\|k\| \leq M_h \rho(k), \text{ pour tout } k \text{ élément hermitien de } PL(h).$$

D'où, pour tout $t \in R$ et pour tout k hermitien de $PL(h)$, on a

$$\left\|e^{itk}\right\| \leq M_h \rho(e^{itk}) = M_h.$$

Pour $n \in N^*$, posons

$$H_n = \left\{h \in Sym(A) : \left\|e^{ith}\right\| \leq n, \text{ pour tout } t \in R\right\},$$

où $Sym(A)$ désigne l'ensemble des éléments hermitiens de A. Alors $(H_n)_n$ est une suite de fermés de $Sym(A)$. D'où, d'aprés le théorème de Baire, il existe n_0

tel que l'intérieur de H_{n_0} soit non vide, et par conséquent contient une boule de centre h_0 et de rayon r. Soit maintenant h un élément hermitien de A. Alors, pour tout $t \in R$, on a

$$\left\| e^{itr\frac{h}{\|h\|+1}} \right\| \leq \left\| e^{it\left(r\frac{h}{\|h\|+1}+h_0\right)} \right\| \left\| e^{-ith_0} \right\| \leq n_0^2;$$

donc en particulier

$$\left\| e^{ih} \right\| \leq n_0^2, \text{ pour tout } h \in Sym(A).$$

D'où, d'aprés [5], la norme $\|.\|$ est équivalente, sur A, à une norme de C^*-algèbre. Cette dèrnière norme est exactement ρ puisque A est commutative.

Comme conséquence, nous avons le résultat suivant.

Corollaire II.7. ([3], Théorème 2). Soit $(A, \|.\|)$ une algèbre de Banach commutative et semi-simple. Une condition nécessaire et suffisante pour que toute fonction continue opère au sens de [3], sur A, est que A soit l'algèbre de toutes les fonctions continues sur son espace des caractères (non nuls).

Références

[1] J. Dixmier, Les C^*-algèbres et leurs représentations. Gauthiers Villars 2ème édition (1969).

[2] I. Kaplansky, Ring isomorphisms of Banach algebras. Canad. J. Math. 6(1954).

[3] Y. Katznelson, Algèbres caractérisées par les fonctions qui opèrent sur elles. C. R. Acad. Sc. , Paris, A-B 247(1958), 903-905.

[4] C. Le Page, Sur quelques conditions impliquant la commutativité dans les algèbres de Banach. C. R. Acad. Sc., Paris, A-B 265(1967), A 235-A 237.

[5] V. Ptàk, Banach algebras with involution. Manuscripta Math., 6(1972), 245-290.

Ecole Normale Supérieure, B.P. 5118, Takaddoum, 10105 Rabat, Maroc.

Keplerian problems in Frenet variables

Javier Ribera* y Antonio Elipe†

*Depto. Matemática Aplicada. Universidad de Zaragoza. 50009 Zaragoza

†Grupo de Mecánica Espacial. Universidad de Zaragoza. 50009 Zaragoza

Abstract

In this paper we introduce a set of symplectic variables, based on the intrinsic coordinates or Frenet frame. Applications to Keplerian motions are made.

Keywords: intrinsic variables, Keplerian systems, canonical transformations

1. Introduction

The two body problem is, most likely, the problem corresponding to a dynamical system better studied, and it is the origin of celestial mechanics, since motion of planets, satellites, both natural and artificial, may be considered as a perturbation of the basic two body problem. Hence the interest in obtaining appropriate sets of variables, canonical or not, that— on the one hand— give a good picture of the motion and— on the other— one can handle the problem when perturbations are included. Let us mention, among the most popular, the several anomalies, true, eccentric or mean, and canonical sets of variables, like the polar-nodal variables, the Delaunay variables, parabolic or Levi-Cività transformation, and many more.

In some practical cases, for instance in the orbit of an artificial satellite, it is necessary to monitor the motion in the along-track, in-track and out-of-plane directions, in order to detect the errors produced in numerical integrations. But these directions are the tangential, normal and binormal ones, that is to say, the Frenet frame in classical differential geometry [3]. Thus, it seems that *a set* of variables defined on the basis of these directions is of interest in astrodynamics. Based on this fact, in this paper we define a set of intrinsic variables, analogous to the Cartesian ones, but with respect to the Frenet frame. These coordinates are completed with their conjugate momenta, that we name *Frenet variables*, in such a way that the transformation from Cartesian canonical variables to the Frenet ones is symplectic, more precisely, it is a Mathieu transformation.

As a first illustration, we formulate the two body problem in this set of variables (Section 2), and we establish some relations between our coordinates and the true anomaly.

In so doing, some properties of the Keplerian motion (in particular the hyperbolic motion) are easily obtained.

2. Frenet variables

Let us consider the motion of a particle P of unit mass moving around another point O under a Newtonian gravity force. With respect to an inertial frame with origin at the point O, the equations of motion are

$$\ddot{\boldsymbol{x}} = -\mu \frac{\boldsymbol{x}}{r^3},$$

where $\boldsymbol{x}$ is the position vector of the particle, $r = \|\boldsymbol{x}\|$ its norm (the radial distance) and μ the Gaussian constant. As it is well known, the motion is planar. Let us define the angular momentum vector

$$\boldsymbol{G} = \boldsymbol{x} \times \dot{\boldsymbol{x}},$$

and the Laplace vector

$$\boldsymbol{A} = \frac{1}{\mu}(\dot{\boldsymbol{x}} \times \boldsymbol{G}) - \frac{\boldsymbol{x}}{r},$$

that are first integrals of the motion (see [2, 1]). Thus, we can introduce a new orthonormal inertial frame $O\,\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$ defined as

$$\boldsymbol{e}_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|}, \quad \boldsymbol{e}_3 = \frac{\boldsymbol{G}}{\|\boldsymbol{G}\|}, \quad \boldsymbol{e}_2 = \boldsymbol{e}_3 \times \boldsymbol{e}_1.$$

In this system, $\boldsymbol{e}_1$ points towards the pericenter, and $\boldsymbol{e}_3$ is perpendicular to the motion plane.

Let us now define another reference frame, $P\,\boldsymbol{f}_1\boldsymbol{f}_2\boldsymbol{f}_3$, but now a *moving* one, such that

$$\boldsymbol{f}_3 = \boldsymbol{e}_3, \quad \boldsymbol{f}_2 = \frac{\dot{\boldsymbol{x}}}{\|\dot{\boldsymbol{x}}\|}, \quad \text{and} \quad \boldsymbol{f}_1 = \boldsymbol{f}_3 \times \boldsymbol{f}_2,$$

that is to say, $\boldsymbol{f}_2$ is the tangent vector to the orbit, $\boldsymbol{f}_1$ the binormal and $\boldsymbol{f}_3$ is the normal, or the classic *Frenet's frame* (see e.g. the textbook of Struik [3]).

Since both frames have a common axis ($\boldsymbol{f}_3 = \boldsymbol{e}_3$), they are related by means of a rotation about this axis and a certain angle α, hence, we have

$$\begin{aligned} \boldsymbol{f}_1 &= \boldsymbol{e}_1 \cos\alpha + \boldsymbol{e}_2 \sin\alpha, \\ \boldsymbol{f}_2 &= -\boldsymbol{e}_1 \sin\alpha + \boldsymbol{e}_2 \cos\alpha, \end{aligned} \tag{1}$$

or, conversely,

$$\begin{aligned} \boldsymbol{e}_1 &= \boldsymbol{f}_1 \cos\alpha - \boldsymbol{f}_2 \sin\alpha, \\ \boldsymbol{e}_2 &= \boldsymbol{f}_1 \sin\alpha + \boldsymbol{f}_2 \cos\alpha. \end{aligned} \tag{2}$$

Let us denote by

$$\boldsymbol{x} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2, \qquad \dot{\boldsymbol{x}} = \boldsymbol{X} = X\boldsymbol{e}_1 + Y\boldsymbol{e}_2, \tag{3}$$

the expression of the position and velocity vector with respect to the fixed frame, and

$$\boldsymbol{x} = u\boldsymbol{f}_1 + v\boldsymbol{f}_2, \qquad \dot{\boldsymbol{x}} = \boldsymbol{X} = V\boldsymbol{f}_2, \tag{4}$$

with respect to the moving frame.

The angular momentum $\boldsymbol{G}$, when expressed in this moving basis, is

$$\boldsymbol{G} = \boldsymbol{x} \times \boldsymbol{X} = uV\boldsymbol{f}_3,$$

hence, since it is an integral of the motion, its norm,

$$\|\boldsymbol{G}\| = uV = \Phi \tag{5}$$

is constant along the motion.

Let us take the angle $\varphi = \alpha + \pi/2$, that is, the angle between $\boldsymbol{e}_1$ and $\boldsymbol{f}_2$, or in words, the angle between the pericenter and the tangent to the trajectory. From Eqs. (3) and (4), and the relations (1), there results that

$$\begin{aligned} x &= v\cos\varphi + u\sin\varphi, \quad X = V\cos\varphi, \\ y &= v\sin\varphi - u\cos\varphi, \quad Y = V\sin\varphi. \end{aligned}$$

But taking into account the integral of the angular momentum as given in relation (5), we can put the previous expression as

$$\begin{aligned} x &= v\cos\varphi + \frac{\Phi}{V}\sin\varphi, \quad X = V\cos\varphi, \\ y &= v\sin\varphi - \frac{\Phi}{V}\cos\varphi, \quad Y = V\sin\varphi. \end{aligned} \tag{6}$$

Thus, we can think of these relations as a transformation of variables

$$(x, y, X, Y) \longmapsto (v, \varphi, V, \Phi)$$

in such a way that the variables (v, φ) are the coordinates and (V, Φ) their conjugate moments. This is justified, since the above transformation (6) is symplectic. Indeed, it is a matter of computing Poisson brackets to see that this is the case. Thus, we can state

Proposition 1 *The transformation* $(x, y, X, Y) \longmapsto (v, \varphi, V, \Phi)$ *given by Eqs. (6) is completely canonical.*

Besides, we can check easily that

$$X\,dx + Y\,dy = \Phi d\varphi + V\,dv,$$

which shows the following

Proposition 2 *The transformation* $(x, y, X, Y) \longmapsto (v, \varphi, V, \Phi)$ *given by Eqs. (6) is a Mathieu transformation.*

In Frenet variables, the radial distance is

$$r^2 = x^2 + y^2 = \frac{\Phi^2}{V^2} + v^2,$$

and the velocity

$$X^2 + Y^2 = V^2,$$

then there results that the Hamiltonian of the Keplerian motion is

$$\mathcal{H} = \frac{1}{2}V^2 - \mu\left(\frac{\Phi^2}{V^2} + v^2\right)^{-1/2}. \tag{7}$$

It is worth noting that there is a kind of duality between Frenet variables and the classical polar–nodal variables $(r, \vartheta, R, \Theta)$, since this set is related to the Cartesian ones by

$$\begin{aligned} x &= r\cos\vartheta, \quad X = R\cos\vartheta - \frac{\Theta}{r}\sin\vartheta, \\ y &= r\sin\vartheta, \quad Y = R\sin\vartheta + \frac{\Theta}{r}\cos\vartheta. \end{aligned} \tag{8}$$

Indeed, in polar–nodal coordinates, r, ϑ are the polar coordinates for the position vector, while in Frenet coordinates, V, φ are the polar coordinates for the velocity vector.

3. Relation with the true anomaly

In a Keplerian orbit or in perturbed Keplerian orbits, the polar coordinates, namely the radial distance r and the true anomaly f, play an essential role; thus, it should be of interest to analyze the relations between the polar coordinates and the Frenet ones.

First of all, let us denote by $\boldsymbol{e}_r = \boldsymbol{x}/r$ the unit vector in the radial direction, and by $\boldsymbol{e}_t = \boldsymbol{e}_3 \times \boldsymbol{e}_r$ the unit vector in the transversal direction. Thus, it is well known that the velocity may be expressed in this system as

$$\boldsymbol{X} = \boldsymbol{e}_r\,\dot{r} + \boldsymbol{e}_t\, r\dot{f}.$$

Let us designate by ψ the angle between the radial vector $\boldsymbol{e}_r$ and the vector $\boldsymbol{f}_2$, that is, $\psi \in [0, 2\pi)$ is the angle such that

$$\boldsymbol{f}_1 = \boldsymbol{e}_r \sin\psi - \boldsymbol{e}_t\cos\psi, \qquad \boldsymbol{f}_2 = \boldsymbol{e}_r\cos\psi + \boldsymbol{e}_t\sin\psi \tag{9}$$

Hence, since $\boldsymbol{X} = V\boldsymbol{f}_2$ (Eq.(4)), there results that

$$\boldsymbol{X} = \boldsymbol{e}_r\,\dot{r} + \boldsymbol{e}_t\, r\dot{f} = V\boldsymbol{f}_2 = V(\boldsymbol{e}_r\cos\psi + \boldsymbol{e}_t\sin\psi),$$

and therefore,

$$\dot{r} = V\cos\psi, \quad \text{and} \quad r\dot{f} = V\sin\psi,$$

hence the angle ψ above defined may be computed as

$$\tan\psi = \frac{r\dot{f}}{\dot{r}} = \frac{r}{r'} = \frac{p}{re\sin f}, \tag{10}$$

where $(') = d/df$, and $r = p/(1 + e\cos f)$.

On the other hand, taking into account the relations (9), we have that

$$\boldsymbol{x} = r\boldsymbol{e}_r = r(\boldsymbol{f}_1\sin\psi + \boldsymbol{f}_2\cos\psi).$$

Then by identifying coefficients with Eq. (4), there results that

$$u = r\sin\psi, \qquad v = r\cos\psi,$$

and by using (10), we obtain

$$v = \frac{re\sin f}{p}u. \tag{11}$$

On account of the integral relation (5), $\dot{u}V = -u\dot{V}$, but from the Hamilton equations derived from Hamiltonian (7),

$$\dot{V} = -\frac{\partial\mathcal{H}}{\partial v} = -\mu\frac{v}{(u^2+v^2)^{-3/2}} = -\mu\frac{v}{r^3}.$$

With this, and Eq. (11), there results

$$\dot{u} = \frac{\mu uv}{r^3V} = \frac{\mu u^3}{\Phi pr^2}e\sin f;$$

but remembering that $\dot{u} = u'\dot{f} = u'\Phi/r^2$, the above equation is converted into

$$\frac{du}{df} = \frac{u^3}{p^2}e\sin f,$$

that is the sought differential equation. Note that in order to arrive at this expression, we replaced the well know relation $\mu = \Phi^2/p$. Thus, for having explicitly the function $u = u(f)$ available, we have to integrate the above equation, which is solved by the quadratures

$$\int_{u_0}^{u}\frac{du}{u^3} = \frac{e}{p^2}\int_0^f \sin f\,df,$$

whose solution is

$$\frac{1}{u^2} = \frac{1}{u_0^2} - \frac{2e}{p^2}(1-\cos f). \tag{12}$$

Note that when $f = 0$, the Frenet variable $v = 0$ (Eq. (11)) and then $u_0 = r_0$, the value of the radial distance at the pericenter.

Obtaining the other Frenet variable, v, is very simple, since both u and v are such that $r^2 = u^2 + v^2$.

3.1 *Hyperbolic case*

For the hyperbolic motion, $p = a(e^2 - 1)$, and $r_0 = a(e - 1)$. After some algebra, from Eq. (12), there follows that

$$u(f) = \frac{p}{\sqrt{1 + e^2 + 2e\cos f}}. \tag{13}$$

From this equation, we can check that at the pericenter, $u(0) = a(e-1)$. On the other hand, it is well known for the hyperbolic motion that $\lim_{r\to\infty} f = f_\infty = \mathrm{acos}(-1/e)$, and then

$$\lim_{f\to f_\infty} u = a\sqrt{e^2 - 1} = b.$$

That is, at the asymptote, the value of u is the semiminor axis b, a quantity known as the *impact parameter*.

Another known property of the hyperbolic motion at the infinity, namely, that the product $b\,V_\infty = G$, now is easy to prove. Indeed, since the product $uV = \Phi = G$ is constant along the motion, when evaluated at the infinity, there results that

$$u(f_\infty)\,V_\infty = b\,V_\infty = \Phi = G.$$

3.2 *Elliptic case*

For the elliptic case, $e < 1$, the parameter is $p = a(1 - e^2)$ and the minimum distance is $r_0 = a(1 - e)$, then Eq. (12) is converted into

$$u(f) = \frac{p}{\sqrt{1 + e^2 + 2e\cos f}}, \tag{14}$$

which coincides, formally, with the one (13) obtained for the hyperbolic motion.

3.3 *Parabolic case*

When the motion is parabolic, $e = 1$, then the pericenter distance is $r_0 = p/2$, thanks to which the expression (12) of $u(f)$ reduces simply to

$$u(f) = \frac{p}{\sqrt{2(1 + \cos f)}}. \tag{15}$$

Acknowledgments. Authors are indebted to Dr. Floría for his criticism and helpful comments. This paper has been supported by Projects #PB98-1576 and #ESP99-1074-CO2-01 and by the Centre National d'Études Spatiales at Toulouse.

References

[1] Danby, J.M.A.: 1988, *Fundamentals of Celestial Mechanics*, Wilmann-Bell, Inc., Richmond, VA.

[2] Goldstein, H.: 1989, *Classical Mechanics*. 2nd ed. Addison-Wesley, Reading. MA.

[3] Struik, D. J.: *Lectures on Classical Differential Geometry*, 2nd ed. Addison-Wesley, Reading. MA. Dover reprint.

LENGTH OF ORBITAL ARC AND CANONICAL KEPLERIAN ELEMENTS

Luis Floría * and Ignacio Aparicio **

* Grupo de Mecánica Espacial. Universidad de Zaragoza.
Facultad de Ciencias (Matemáticas). E – 50 009 Zaragoza. Spain.
** Grupo de Mecánica Celeste I. Universidad de Valladolid. ETSII.
Dept. Matemática Aplicada a la Ingeniería. E – 47 011 Valladolid. Spain.

Abstract

In this paper the use of the *orbital arc length* as an independent variable (that is, as a reparametrizing fictitious time) in the study of *Keplerian motion* is revisited, on this occasion in the context of Hamiltonian Celestial Mechanics in *extended phase space*. In particular, we focus our attention on the introduction of this parameter both as an independent argument and as a *canonical element*. It is shown how all the different sets of DS (Delaunay–Similar or Delaunay–Scheifele) canonical variables, with the length of orbital arc as the *pseudo–time*, can be easily encompassed within a single elegant framework inspired in that considered by Bond & Broucke and Bond & Janin in the late 1970s and early 1980s.

Key words: Keplerian systems, extended phase space, canonical transformations, Delaunay and Delaunay–Similar (DS) elements, time transformation, length of arc.

AMS (MOS) Subject Classification: 70 H 15, 70 F 15, 70 M 20.

1. Introduction

For analytical step–size regulation in numerical integration of highly eccentric elliptic orbits, *E. V. Brumberg* (1992) proposed the use of the *length of orbital arc* as the independent variable. His derivation of this parameter as a pseudo–time, explicitly based on geometrical and dynamical properties pertaining to elliptic Keplerian motion, was thus originally limited to the case of elliptic orbits in the two–body problem. Brumberg's reparametrizing transformation, establishing a differential relation between the physical

time t and the orbital arc length, was systematized and generalized (Floría 1997) within a universal formulation and uniform treatment of the two–body problem (Stiefel & Scheifele 1971, §11).

The reckoning and analytical work required by the above developments does not resort to concepts belonging to the general framework of Hamiltonian Mechanics. However, for our purposes, the *homogeneous canonical formalism* is the natural stage on which we shall perform our study. Thanks to a device due to *Poincaré* (1905, vol. I, Chapter 1, §12, pp. 13–16), the ordinary phase space is enlarged by means of two additional dimensions, giving rise to the so–called *extended phase space*: the physical time t is incorporated as a dependent variable, an additional coordinate–like canonical variable, whose conjugate momentum is related to the original Hamiltonian (after a change of sign). The extended–phase–space approach facilitates the introduction of new independent variables other than the physical time. Details concerning this matter can be found, e.g., in Stiefel & Scheifele (1971, §30, §34, §37), and Scheifele (1970a, 1972).

By means of a generating function *of the second type*, depending on the old coordinates and the new momenta (Goldstein 1980, §9.1, pp. 383–384), *Bond & Broucke* (1980) developed a completely canonical transformation in extended phase space that relates the (enlarged) Delaunay elliptic Keplerian elements to the set of Delaunay–Similar (or Delaunay–Scheifele, DS) canonical variables proposed by *Scheifele & Graf* (1974). These Scheifele–Graf canonical variables are Keplerian orbital elements with respect to the true anomaly as the independent variable. Here, we adopt the concept of *elements of the motion* in the sense of Stiefel & Scheifele (1971, §18, pp. 83–84).

The generating function, and the transformation, proposed by Bond & Broucke was also employed by *Bond & Janin* (1981) to obtain analogous canonical orbital elements (of the Scheifele–Graf kind) with respect to any anomaly–like independent variable.

For the formulation and approximate analytical solution of perturbed Keplerian systems arising in the theory of motion of artificial Earth satellites, *Scheifele* and his ETH Zürich co–workers (1970a §§2.2; 1970b; 1972, Part B) had already studied the construction and application of *other* different sets of eight Delaunay–Similar orbital variables and elements, putting special emphasis on the use of Keplerian true– and eccentric–like anomalies as the independent variable. Their approach resorts to the formulation of the Keplerian Hamiltonian in polar spherical coordinates in the extended phase space and its subsequent solution by the Hamilton–Jacobi technique via separation of variables.

Such a lengthy and cumbersome treatment can be considerably simplified if, form the outset, (enlarged) polar nodal variables are adopted to formulate the Kepler problem (*Deprit* 1981). As a consequence, the *true anomaly* is easily made a coordinate–type

canonical variable, incorporating it as a DS element. On the basis of Deprit's derivation of Scheifele's (1970b, 1972) Delaunay–Similar set of Keplerian elements, *Floría* (1994) presented a *unified pattern* for the *general and systematic construction* of canonical sets of DS variables in a true–like anomaly as the pseudo–time.

It should be remembered that the developments and results contained in the articles by Bond & Broucke and Bond & Janin are drastically restricted to the consideration of *only one specific set* of DS elements, just the one due to Scheifele & Graf (1974). Since *other different DS sets* (Floría 1994) are of interest in the study of perturbed Keplerian systems in extended–phase–space formulation, in the present paper we intend to *modify* the Bond–Broucke generating function, on obtaining its *general expression* in order to derive the transformation to *any possible DS set* of (unspecified) canonical orbital variables from (extended) Delaunay elements. Accordingly, the diverse canonical sets of dependent variables introduced by Scheifele and his collaborators are embedded in our unified scheme. In particular, the new Keplerian orbital elements will use the *length of orbital arc* as the independent variable (fictitious time).

2. The Generating Function. Transformation from Delaunay Elements

After a review and summary of some well–known results concerning the standard Delaunay variables, the enlarged set (with the time t and the negative of the total energy as an additional conjugate couple of canonical variables) provides us with the starting point for the passage to a new canonical set of generic (unspecified) DS variables in the 8–dimensional phase space. The approach and steps taken by Bond & Broucke (1980) and Bond & Janin (1981) will be *adapted* to our purposes, in order to produce *unspecified DS orbital elements* for the Kepler problem.

The list of symbols (l_D , g_D , h_D ; L_D , G_D , H_D) will denote the canonical set of Delaunay variables. Obviously, the subscript D means "*Delaunay*". Some formulae relating these variables to the usual Keplerian orbital elements $a \equiv a(L_D)$, $e \equiv e(L_D, G_D)$, $p \equiv p(G_D)$, $I(G_D, H_D)$, $\omega(g_D)$ and $\Omega(h_D)$, the radial distance r (Euclidean norm of the position vector of the moving point), and the *auxiliary variables* $E \equiv E(r; L_D, G_D)$ and $f \equiv f(r; L_D, G_D)$, Keplerian eccentric and true anomalies, are

$$L_D = \sqrt{\mu a}\,, \quad G_D^2 = \mu a \left(1 - e^2\right) = \mu p\,, \quad H_D = G_D \cos I\,, \tag{1}$$

$$e^2 = 1 - \left(G_D^2/L_D^2\right)\,, \quad p = G_D^2/\mu\,, \tag{2}$$

$$l_D = \Phi(E) = E - e \sin E, \quad g_D = \omega\,, \quad h_D = \Omega\,, \tag{3}$$

$$r = a\,(1 - e \cos E)\,, \quad r = p/\,(1 + e \cos f)\,, \tag{4}$$

$$d\,l_D/d\,E = d\,\Phi(E)/d\,E = 1 - e \cos E = r/a\,. \tag{5}$$

The Hamiltonian of a Keplerian system, formulated in this chart, reduces to

$$\mathcal{H}_0 \equiv \mathcal{H}_0\,(-,-,-;L_D,-,-) = -\mu^2/(2\,L_D^2)\,. \tag{6}$$

The canonical system of differential equations of motion in the Delaunay chart, as derived from Hamiltonian (6) with t as the independent variable, leads to a *solution* whose structure shows that the Delaunay variables are a *set of canonical elements of the motion* for Keplerian systems (in the sense of Stiefel & Scheifele 1971, §18): they are quantities which, in the unperturbed Kepler problem, are constant or linear functions of the independent variable.

The enlarged Delaunay set, in extended, 8–dimensional phase space, is formed from the six classical Delaunay variables by appending to them the *time t as a coordinate* whose canonically conjugate momentum T is the energy of the problem after a change of sign. A completely canonical transformation

$$(t\,,\,l_D\,,\,g_D\,,\,h_D\,;\,T\,,\,L_D\,,\,G_D\,,\,H_D) \longrightarrow (\psi\,,\,l\,,\,g\,,\,h\,;\,\Psi\,,\,L\,,\,G\,,\,H)\,,$$

from the Delaunay elements (in extended phase space) to a set of unspecified Delaunay–Similar variables, is performed by means of the *generating function* of the second type $S \equiv S_{DS}(t\,,\,l_D\,,\,g_D\,,\,h_D\,;\,\Psi\,,\,L\,,\,G\,,\,H)$,

$$S = t\,L + \left[\mu/\sqrt{2\,L}\right] l_D + \mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H)\,Z(t) + g_D\,G + h_D\,H\,, \tag{7}$$

which is a *generalization* of the one utilized by Bond & Broucke (1980) and Bond & Janin (1981) to obtain the Scheifele–Graf (1974) canonical (dependent) variables. Now $\mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H)$ is, in principle, an *arbitrary function* of the new canonical momenta, and $Z(t)$ is a certain *unspecified* function of t.

With the notations

$$Z'(t) \equiv dZ(t)/dt\,, \quad \partial\mathcal{F}/\partial(\Psi,L,G,H) \equiv \mathcal{F}_{(\Psi,L,G,H)}\,, \tag{8}$$

the generating relations derived from S yield the *implicit transformation equations*

$$\psi = \partial S/\partial\Psi = \mathcal{F}_\Psi\,Z(t) \;\Rightarrow\; Z(t) = \psi/\mathcal{F}_\Psi\,, \tag{9}$$

$$l = \partial S/\partial L = t - \left[\mu/(2\,L)^{3/2}\right] l_D + \mathcal{F}_L\,Z(t)\,, \tag{10}$$

$$g = \partial S/\partial G = g_D + \mathcal{F}_G\,Z(t) = g_D + (\mathcal{F}_G/\mathcal{F}_\Psi)\,\psi\,, \tag{11}$$

$$h = \partial S/\partial H = h_D + \mathcal{F}_H\,Z(t) = h_D + (\mathcal{F}_H/\mathcal{F}_\Psi)\,\psi\,, \tag{12}$$

$$T = \partial S/\partial t = L + \mathcal{F}(\Psi,L,G,H)\,(dZ/dt) \tag{13}$$

$$\Rightarrow\; \mathcal{F}(\Psi,L,G,H) = \left\{T - \left[\mu^2/\left(2\,L_D^2\right)\right]\right\}/Z'(t)\,, \tag{14}$$

$$L_D = \partial S/\partial l_D = \mu/\sqrt{2\,L} \;\Rightarrow\; L = \mu^2/\left(2\,L_D^2\right), \tag{15}$$

$$G_D = \partial S/\partial g_D = G\,, \quad H_D = \partial S/\partial h_D = H\,. \tag{16}$$

Notice that the preceding equation for l is *a general expression of the Kepler equation.*

In terms of the enlarged Delaunay set, and then –after the transformation– in the corresponding DS formulation, Hamiltonian (6) is converted into

$$\mathcal{H}_0 \;\rightarrow\; (\mathcal{H}_0)_h = T - \left[\mu^2/\left(2\,L_D^2\right)\right] \Longrightarrow \widetilde{\mathcal{H}}_h = \mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H)\;Z'(t)\,, \qquad (17)$$

with t as the independent variable. By definition of T (Bond & Broucke 1980, p. 358), in line with the result due to Poincaré, the numerical value of this Hamiltonian is zero.

3. Simplification of the Hamiltonian. Change of Time Parameter

Hamiltonian $\widetilde{\mathcal{H}}_h$ can be simplified if one defines a *fictitious time* σ (for our purposes, the time–parameter of interest will be the *length of orbital arc*) as the new independent variable, given by means of a *differential relation* $dt = \tilde{f}\,d\sigma$, the function $\tilde{f}$ being appropriately chosen as a function of the *new* canonical DS variables. For convenience in the derivation of results, the specific form of $\tilde{f}$ will be presented at a later stage (see below).

Consequently, according to the general theory of reparametrization of motion in terms of new independent variables (see, e.g., Scheifele 1970a; or Stiefel & Scheifele 1971, §34), the above Hamiltonian becomes

$$\mathcal{K}_0 \;=\; \tilde{f}\,\widetilde{\mathcal{H}}_h \;=\; \mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H)\;(dZ/dt)\;\tilde{f}\,. \qquad (18)$$

This Hamiltonian takes σ as the independent variable.

Remember that $Z(t)$ has not been specified yet: the way to further simplification in the functional form and dependence of (18) is still open, since some suitable condition can be imposed on the time–related functions $\tilde{f}$ and $Z(t)$. *For the Keplerian system at issue*, $\tilde{f}$ and $Z(t)$ are taken such that

$$\tilde{f}\;(dZ/dt) \;=\; 1 \;\Longrightarrow\; \mathcal{K}_0 \;=\; \mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H) \;\equiv\; 0 \qquad (19)$$

in the phase space of the new variables. Notice also that $\mathcal{F}(\Psi\,,\,L\,,\,G\,,\,H)$ still remains unspecified. Special *choices* of $\mathcal{F}$ (for instance, as in Floría 1994) lead to specific DS–like sets: that is, *families of DS sets of canonical orbital elements* can be defined.

A simple *parametrical solution, in terms of* the new pseudo–time σ, to the canonical system of equations of motion issued from $\mathcal{K}_0$ is

$$(\Psi, L, G, H)' \;=\; -\,\partial\mathcal{F}/\partial\,(\psi, l, g, h) \;=\; \mathbf{0} \qquad (20)$$

$$\Longrightarrow\; (\Psi, L, G, H) \;=\; (\Psi_0, L_0, G_0, H_0)\,, \qquad (21)$$

$$(\psi, l, g, h)' \;=\; \partial\mathcal{F}/\partial\,(\Psi, L, G, H) \;=\; \mathcal{F}_{(\Psi,L,G,H)} \qquad (22)$$

$$\Longrightarrow\; (\psi, l, g, h) \;=\; (\mathcal{F}_\Psi, \mathcal{F}_L, \mathcal{F}_G, \mathcal{F}_H)\;\sigma \;+\; (\psi_0, l_0, g_0, h_0)\;, \qquad (23)$$

where $(\psi_0, l_0, g_0, h_0, \Psi_0, L_0, G_0, H_0)$ represent integration constants. In view of this solution, these DS canonical variables constitute a set of Keplerian *elements of the motion* with respect to σ (Stiefel & Scheifele 1971, §18). In particular, observe that

$$\psi' \equiv d\psi/d\sigma = \partial\mathcal{K}_0/\partial\Psi = \mathcal{F}_\Psi \implies \psi = \mathcal{F}_\Psi\,\sigma + \text{const.} \tag{24}$$

By examining the derivatives with respect to t, from Eq. (24) for ψ it follows that

$$d\psi/dt = (d\psi/d\sigma)\,(d\sigma/dt) = \mathcal{F}_\Psi\,(d\sigma/dt) = \mathcal{F}_\Psi/\tilde{f}\,, \tag{25}$$

and the generalized anomaly ψ must be *consistent* with the choice of the time transformation given by $\tilde{f}$ (Bond & Janin 1981, p. 161). Some useful formulae for Keplerian elements are:

$$a = L_D^2/\mu = \mu/(2\,L)\,, \quad n = \sqrt{\mu/a^3} = (2\,L)^{3/2}/\mu\,, \tag{26}$$

$$e^2 = 1 - (p/a) = 1 - \left(G_D^2/L_D^2\right) = 1 - \left(2\,L\,G^2/\mu^2\right). \tag{27}$$

In the presence of the equations of the transformation defined by S, and taking into account the expressions for the Delaunay variables, there results the relation

$$t = l + \left[\mu/(2\,L)^{3/2}\right][E - e\sin E] - (\mathcal{F}_L/\mathcal{F}_\Psi)\,\psi\,, \tag{28}$$

a *generalized Kepler equation*, essential to develop the reparametrization which replaces t by the pseudo–time σ, and to obtain the expression for $\tilde{f}$ under a special choice of ψ.

4. Time Transformation to the New Independent Variable

According to Bond & Janin (1981, §4), we consider the total derivative of Eq. (28) with respect to σ,

$$\begin{aligned}\tilde{f} = t' \equiv dt/d\sigma = l' + \left[\mu/(2\,L)^{3/2}\right]'\Phi(E) + \left[\mu/(2\,L)^{3/2}\right]\Phi'(E)\\ - (\mathcal{F}_L/\mathcal{F}_\Psi)'\,\psi - (\mathcal{F}_L/\mathcal{F}_\Psi)\,\psi'.\end{aligned} \tag{29}$$

With the help of the canonical equations of motion [Formulae (20) and (22)], bearing in mind that the orbital eccentricity e remains constant in the unperturbed Keplerian motion, Formula (26) for the semi–major axis a, and Eq. (5), the preceding formula can be expressed in the form

$$\tilde{f} = \left[\mu/(2\,L)^{3/2}\right][1 - e\cos E]\,(dE/d\sigma) = \left[r/\sqrt{2\,L}\right](dE/d\sigma)\,. \tag{30}$$

Since $\psi' = \mathcal{F}_\Psi \Rightarrow d\psi = \mathcal{F}_\Psi d\sigma$, derivatives with respect to σ are expressed as derivatives with respect to ψ, and the *reparametrizing function* in the time transformation becomes

$$dE/d\sigma = (dE/d\psi)\,(d\psi/d\sigma) \Rightarrow \tilde{f} = \left(r/\sqrt{2L}\right)[\partial\mathcal{F}(\Psi, L, G, H)/\partial\Psi]\,(dE/d\psi) \tag{31}$$

Thus, given a particular anomaly ψ, in order to calculate $\tilde{f}$ *the eccentric anomaly* E *must be expressed as a function* $E = E(\psi)$ *of* ψ (Bond & Janin 1981, p. 165). Moreover,

$$Z(t) \;=\; \psi/\mathcal{F}_\Psi \Longrightarrow dZ(t)/d\sigma \;=\; \psi'/\mathcal{F}_\Psi \;=\; 1 \Longrightarrow Z(t) \;=\; \sigma + \text{const.} \tag{32}$$

We can summarize the preceding results as follows:

- The fundamental relations $\psi \longrightarrow E = E(\psi)\,,\; dE/d\psi\,,$

$$t \;=\; l + \frac{\mu}{(2\,L)^{3/2}}\,[E - e\sin E] - \left(\frac{\mathcal{F}_L}{\mathcal{F}_\Psi}\right)\psi\,, \quad \text{(Kepler's equation)}\,, \tag{33}$$

$$dt \;=\; \tilde{f}\,d\sigma\,,\; \tilde{f} = \frac{r}{\sqrt{2L}}\,\frac{\partial\mathcal{F}\,(\Psi, L, G, H)}{\partial\Psi}\,\frac{dE}{d\psi}\,, \quad \text{(reparametrizing function)}, \tag{34}$$

complete the introduction of canonical orbital elements, *depending on the new independent variable*, for the Keplerian system generated by (6). In addition to this, special choices for $\mathcal{F}(\Psi, L, G, H) = \mathcal{K}_0$ can be adduced for elliptic DS sets (Floría 1994).

In particular, if we want to convert the *length of orbital arc* into a canonical element, say $\psi = \sigma$, we can take advantage of formulae established by Brumberg (1992; see also Floría 1997): fixing Cartesian coordinates in the orbital plane,

$$x(E) \;=\; a\,(\cos E - e)\,, \quad y(E) \;=\; a\sqrt{1 - e^2}\,\sin E\,, \tag{35}$$

$$(d\,\sigma)^2 \;=\; (d\,x)^2 + (d\,y)^2 \;=\; a^2\left(1 - e^2\cos^2 E\right)(d\,E)^2\,, \tag{36}$$

$$dE/d\sigma \;=\; 1/a\sqrt{1 - e^2\cos^2 E}\,, \tag{37}$$

$$\tilde{f} \;=\; \mathcal{F}_\Psi\, r^{1/2}/\sqrt{2\,\mu - 2\,L\,r} \;=\; \mathcal{F}_\Psi\, r^{1/2}/\sqrt{\mu\,(1 + e\cos E)}\,, \tag{38}$$

$$t \;=\; l + \left[\mu/\,(2L)^{3/2}\right][E - e\sin E] - (\mathcal{F}_L/\mathcal{F}_\Psi)\,\sigma\,. \tag{39}$$

Acknowledgements

This research has been partially supported by the Dirección General de Enseñanza Superior (DGES) of Spain, Project PB. 98–1576.

References

[1] Bond, V. and Broucke, R.: 1980, 'Analytical Satellite Theory in Extended Phase Space', *Celest. Mech.* **21**, 357–360.

[2] Bond, V. R. and Janin, G.: 1981, 'Canonical Orbital Elements in Terms of an Arbitrary Independent Variable', *Celest. Mech.* **23**, 159–172.

[3] Brumberg, E. V.: 1992, 'Length of Arc as Independent Argument for Highly Eccentric Orbits', *Celest. Mech. and Dyn. Astron.* **53**, 323–328.

[4] Deprit, A.: 1981, 'A Note Concerning the TR–Transformation', *Celest. Mech.* **23**, 299–305.

[5] Floría, L.: 1994, 'On the Definition of the Delaunay–Similar Canonical Variables of Scheifele', *Mechanics Research Communications* **21**, 409–414.

[6] Floría, L.: 1997, "Orbital Arc Length as a Universal Independent Variable". In: I. M. Wytrzyszczak, J. H. Lieske and R. A. Feldman (Eds.), *Dynamics and Astrometry of Natural and Artificial Celestial Bodies* (IAU Colloq. 165), 405–410. Kluwer.

[7] Goldstein, H.: 1980, *Classical Mechanics* (Second Edition). Addison–Wesley.

[8] Poincaré, H.: 1905, *Leçons de Mécanique Céleste (professées à la Sorbonne)*, vol. I (Théorie générale des perturbations planetaires). Gauthier–Villars, Paris.

[9] Scheifele, G.: 1970a, 'On Nonclassical Canonical Systems', *Celest. Mech.* **2**, 296–310.

[10] Scheifele, G.: 1970b, 'Généralisation des éléments de Delaunay en Mécanique Céleste. Application au mouvement d'un satellite artificiel', *C.R. Acad. Sci. Paris* **271**, 729-732.

[11] Scheifele, G. & Graf, O.: 1974, *Analytical Satellite Theories Based on a New Set of Canonical Elements.* AIAA Paper No. 74–838.

[12] Scheifele, G. and Stiefel, E.: 1972, *Canonical Satellite Theory Based on Independent Variables Different from Time.* Report to ESRO. ESOC-contract 219/70/AR, ETH.

[13] Stiefel, E. L. and Scheifele, G.: 1971, *Linear and Regular Celestial Mechanics.* Springer–Verlag, Berlin–Heidelberg–New York.

Movilidad electroencefalográfica mediante interpolación fractal

M. A. Navascués

Depto. de Matemática Aplicada - CPS. Universidad de Zaragoza.

M. V. Sebastián

Dpto. de Matemáticas - Facultad de Ciencias. Universidad de Zaragoza

J. R. Valdizán

Servicio de Neurofisiología Clínica - Hospital Miguel Servet. Zaragoza

Abstract

Fractal interpolation functions provide new methods of approximation of experimental data. In the present paper, a fractal technique generalizing cubic spline functions is applied. Under some hypothesis about the original signal, error bounds of interpolation are given. Additionally, a quadrature formula to calcule the Hjorth mobility of a electroencephalographic recording is proposed, that complements the description of the signal in the frequency domain. The results about the convergence of interpolation functions allow to deduce an expression for the computation error of the quantifying parameter by the method described.

1. Introducción

La interpretación de un electroencefalograma (EEG) es tarea complicada por la falta de un modelo adecuado que explique cómo se reflejan las características del sistema nervioso central en las observaciones. Se necesitan, por tanto, métodos cuantitativos para la descripción del EEG, que permitan asignar valores numéricos a las características básicas del sistema observado para poder discriminar entre diferentes estados del sistema (o relacionar estos estados con variables endógenas y exógenas). El problema consiste en definir descriptores de calidad para la caracterización general de un modelo de amplitud/tiempo/frecuencia.

El neurofisiólogo B. Hjorth en 1970 definió los parámetros denominados descriptores normalizados de pendiente (o parámetros de Hjorth) en la revista Electroencephalography and Clinical Neurophysiology (Hjorth, 1970). Estos parámetros, basados en la desviación estándar de la señal y sus derivadas, se usan como herramienta clínica en la descripción cuantitativa del EEG, ya que representan la señal EEG en el dominio del tiempo y en

el de la frecuencia. Entre ellos se encuentra la movilidad, que da una medida de la desviación estándar de la derivada de la señal en referencia a la desviación estándar de la amplitud. La movilidad electroencefalográfica es un estimador de la frecuencia media del EEG. Además, por tratarse de un parámetro adimensional permite realizar comparaciones y normalizaciones entre distintos equipos.

Para obtener este indicador, se realiza una interpolación sobre la señal discretizada. La interpolación polinómica supone que la señal es demasiado "lisa" y no recoge la estructura fractal de la misma. Si el muestreo no es muy refinado, la interpolación por splines realiza sobre ésta un suavizado o filtro de paso bajo, que omite las frecuencias altas. En el presente artículo se propone el uso de funciones de interpolación fractal cúbica para obtener una fórmula de cuadratura de la movilidad.

Las funciones de interpolación fractal son de creación reciente (1986) y constituyen una herramienta útil en la aproximación de datos experimentales. Estas funciones pueden computarse de manera rápida y poseen propiedades geométricas que permiten representaciones gráficas adecuadas de fenómenos complejos y un cálculo sencillo de la dimensión fractal del gráfico de las mismas. En el caso particular de las funciones de interpolación fractal polinómica, el método representa una generalización de las funciones splines de este tipo.

Se presentan a continuación una serie de resultados que acotan el error de interpolación cometido en la aproximación por funciones de interpolación fractal spline. Posteriormente, se utilizan estas funciones para interpolar una conjunto de puntos correspondientes a una señal electroencefalográfica muestreada y calcular la movilidad de ésta. Para determinados valores de los parámetros que definen la función de interpolación fractal, se obtiene como caso particular la cuadratura mediante splines cúbicos.

2. Movilidad electroencefalográfica

Una señal electroencefalográfica se puede expresar en una cierta época como una función del tiempo $x(t)$ y, por medio de su transformada de Fourier puede describirse en función de la frecuencia, $\widehat{x}(w)$. Multiplicando la transformada $\widehat{x}(w)$ por su conjugada se obtiene el espectro de potencia de la señal $S(w) = \widehat{x}(w)\ \widehat{x}^*(w)$.

Se define el momento espectral de orden n como:

$$m_n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} w^n S(w)\ dw \qquad [2.1]$$

La descripción de las frecuencias completas obtenida por medio de la transformada de Fourier es simétrica con respecto a la frecuencia 0. Por ser una señal real $\widehat{x}(-w) = \widehat{x}^*(w)$, entonces $S(-w) = S(w)$ y $w^n\ S(w)$ es una función impar si n lo es. Como consecuencia

todos los momentos impares son 0.

Los parámetros de Hjorth o descriptores normalizados de pendiente se definen en función de los momentos espectrales (Hjorth, 1970). La movilidad es la raíz cuadrada del cociente entre el momento de orden 2 y el de orden 0.

$$m_0 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S(w)\ dw \qquad [2.2]$$

$$m_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} w^2 S(w)\ dw \qquad [2.3]$$

$$M = (m_2/m_0)^{1/2} \qquad [2.4]$$

En la práctica se trabaja con señales reales de duración finita. Para determinados cálculos de tipo espectral, es necesaria la condición de periodicidad. Se denotará $L^2(T)$ el espacio de funciones periódicas de periodo T de cuadrado integrable en $I = [0, T]$. Se define la norma de $f \in L^2(T)$ como:

$$\|f\|^2_{L^2} = (f, f)_{L^2} = \frac{1}{T}\int_I f(t)\ f^*(t)\ dt \qquad [2.5]$$

La expresión de los momentos de orden 0 y 2 en el dominio temporal se basa en la igualdad de la energía, es decir, que la potencia total en el dominio de la frecuencia es idéntica a la potencia media en el dominio del tiempo. Se tiene, por la fórmula de Plancherel (Parseval) para una señal de este tipo (Hsu, 1987) la siguiente igualdad:

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\widehat{x}(w)|^2\ dw = \frac{1}{T}\int_I |x(t)|^2\ dt \qquad [2.6]$$

Considerando que $S(w) = \widehat{x}(w)\widehat{x}^*(w) = |\widehat{x}(w)|^2$ y que x es una señal real, y utilizando las propiedades de la transformada de Fourier de la derivada de la señal, se pueden expresar los momentos de orden 0 y 2 de la siguiente forma:

$$m_0 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S(w)\ dw = \frac{1}{T}\int_I x^2(t)\ dt \qquad [2.7]$$

$$m_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} w^2\ S(w)\ dw = \frac{1}{T}\int_I (\frac{dx}{dt})^2\ dt \qquad [2.8]$$

Obtenidos estos momentos es inmediato calcular la movilidad de la señal mediante la expresión [2.4].

3. Funciones de interpolación fractal cúbica

3.1 Sistemas de funciones iteradas (SFI).

Sea K un espacio métrico completo respecto de la distancia $d(x, y)\ \forall\ x, y \in K$. Sea $\mathcal{H}$ el conjunto de todos los subconjuntos de K compactos no vacíos. $\mathcal{H}$ es un espacio métrico completo con la distancia de Hausdorff (Barnsley, 1988):

$$h(A, B) = \max\ \{\underset{x\in A}{Sup}\ \underset{y\in B}{Inf}\ d(x, y),\ \underset{x\in B}{Sup}\ \underset{y\in A}{Inf}\ d(x, y)\}$$

definida para cualesquiera $A,\ B\ \in\ \mathcal{H}$.

Sean $w_n : K \to K \quad n = 1, 2, ..., N$ un conjunto de transformaciones continuas. Entonces la $N+1-$tupla $\{K, w_n :\ n = 1, 2, ..., N\}$ se denomina sistema de funciones iteradas (SFI)(Edgar, 1990). Se define la transformación $W : \mathcal{H} \to \mathcal{H}$ mediante la igualdad:

$$W(A) = \bigcup_n w_n(A) \text{ para } A \in \mathcal{H}$$

Cualquier conjunto $G \in \mathcal{H}$ tal que $W(G) = G$ se dice atractor del SFI (es decir, el atractor es un punto fijo de W). Si K es un compacto, entonces cualquier SFI admite al menos un atractor.

Si para algún $0 \leq s < 1$ y todo $n \in \{1, 2, ..., N\}$ se verifica la desigualdad:

$$d(w_n(x), w_n(y)) \leq s\ d(x, y) \quad \forall\ x, y\ \in K$$

entonces el SFI se dice hiperbólico. En este caso W es una aplicación contractiva respecto la métrica de Hausdorff, es decir,

$$h(W(A), W(B)) \leq s\ h(A, B) \quad \forall\ A, B\ \in \mathcal{H}$$

Como consecuencia del teorema de la aplicación contractiva, W admite un único punto fijo, es decir, existe un único atractor G que verifica

$$G = \lim_{m \to \infty} W^{om}(S) \quad \forall\ S\ \in \mathcal{H}$$

donde W^{om} denota la composición de W consigo misma m veces.

3.2 Interpolación fractal cúbica.

Sean $t_0 < t_1 < ... < t_N$ un conjunto de números reales, se denota por $I = [t_0, t_N] \subset \mathbb{R}$ el intervalo cerrado que los contiene. Sea dado el conjunto de puntos de interpolación $\{(t_n, x_n) \in I \times \mathbb{R} :\ n = 0, 1, 2, ..., N\}$. Para cada subintervalo $I_n = [t_{n-1}, t_n]$ se define la aplicación $L_n : I \to I_n,\ n \in \{1, 2, ..., N\}$ de modo que:

$$L_n(t_0) = t_{n-1},\ L_n(t_N) = t_n \qquad [3.1]$$

y L_n sea un homeomorfismo contractivo:

$$|L_n(c_1) - L_n(c_2)| \leq l\ |c_1 - c_2| \quad \forall\ c_1, c_2\ \in I \qquad [3.2]$$

para algún $0 \leq l < 1$.

Sea $-1 < \alpha_n < 1$; $n = 1, 2, ..., N$, y $F = I \times \mathbb{R}$. Se consideran N aplicaciones continuas, $F_n : F \to \mathbb{R}$ satisfaciendo:

$$F_n(t_0, x_0) = x_{n-1},\ \ F_n(t_N, x_N) = x_n,\ \ n = 1, 2, ..., N \qquad [3.3]$$

$$|F_n(t,x) - F_n(t,y)| \leq \alpha_n |x-y|, \quad t \in I, \quad x,y \in \mathbb{R} \qquad [3.4]$$

Sea $w_n(t,x) = (L_n(t), F_n(t,x))$, $\forall\, n = 1,2,...,N$, se puede enunciar el siguiente resultado:

Teorema (Barnsley, 1988): El SFI $\{F, w_n : n = 1,2,...,N\}$ descrito anteriormente admite un único atractor G. G es el gráfico de una función continua $f : I \to \mathbb{R}$ que verifica $f(t_n) = x_n$ para $n = 0,1,2,...,N$.

La función anterior recibe el nombre de función de interpolación fractal (FIF) asociada con $\{(L_n(t), F_n(t,x))\}_{n=1}^{N}$. Además, es la única función $f : I \to \mathbb{R}$ que satisface la ecuación funcional

$$f(L_n(t)) = F_n(t, f(t)), \quad n = 1,2,...,N, \quad t \in I$$

o, lo que es lo mismo,

$$f(t) = F_n(L_n^{-1}(t), f \circ L_n^{-1}(t)), \quad n = 1,2,...,N, \quad t \in I_n = [t_{n-1}, t_n] \qquad [3.5]$$

Se define el conjunto $\mathcal{F}$ de funciones continuas $f : [t_0, t_N] \to \mathbb{R}$ tales que $f(t_0) = x_0$; $f(t_N) = x_N$. En $\mathcal{F}$ se define la métrica d asociada a la norma del supremo:

$$d(f,g) = \max\{|f(t) - g(t)| : t \in [t_0, t_N]\} \quad \forall\, f,g \in \mathcal{F}.$$

Entonces $(\mathcal{F}, d)$ es un espacio métrico completo.

Dado el SFI anterior, se define la aplicación $T : \mathcal{F} \to \mathcal{F}$ por:

$$(Tf)(t) = F_n(L_n^{-1}(t), f \circ L_n^{-1}(t)) \quad \forall\, t \in [t_{n-1}, t_n], \quad n = 1,2,...,N \qquad [3.6]$$

Utilizando las condiciones [3.1]-[3.4], se comprueba que $(Tf)(t)$ es continua en el intervalo $[t_{n-1}, t_n]$ para $n = 1,2,...,N$ y además lo es en $t_1, t_2, ..., t_{N-1}$. En cada punto, $(Tf)(t_n) = x_n$. Además T es una aplicación contractiva en el espacio métrico $(\mathcal{F}, d)$

$$d(Tf, Tg) \leq |\alpha|_\infty\, d(f,g)$$

donde $|\alpha|_\infty = \max\{|\alpha_n|;\ n = 1,2,...,N\}$. Suponiendo que $|\alpha|_\infty < 1$, el teorema de la aplicación contractiva implica que T posee un único punto fijo en $\mathcal{F}$, es decir existe $f \in \mathcal{F}$ tal que $(Tf)(t) = f(t)$ $\forall\, t \in [t_0, t_N]$. Además, f pasa a través de los puntos de interpolación. Esta función f es la FIF asociada a w_n.

Las funciones de interpolación fractal más estudiadas hasta ahora han sido del tipo

$$\begin{cases} L_n(t) = a_n t + b_n \\ F_n(t,x) = \alpha_n x + q_n(t) \end{cases} \qquad [3.7]$$

donde $q_n(t)$ es una aplicación afín, llamadas FIF afines (Barnsley, 1988; Hardin et al, 1992). En el presente trabajo se estudiará el caso en el que el grado de q_n es 3, que puede considerarse una generalización de los splines cúbicos.

Se trabajará con puntos equidistantes y además, sin pérdida de generalidad, se considerará el intervalo de interpolación en $t_0 = 0$, por lo que si se denota $t_n - t_{n-1} = h$, entonces $t_N - t_0 = Nh$ y por lo tanto

$$a_n = \frac{1}{N} \quad \text{y} \quad b_n = t_{n-1} \qquad [3.8]$$

Si $\alpha_n = 0 \ \forall \ n = 1, 2, ..., N$, entonces $F_n(t, x) = q_n(t)$ ([3.6]) y por lo tanto $f(t) = q_n \circ L_n^{-1}(t) \ \ \forall \ t \in I_n$ ([3.5]). Entonces $f(t)$ es un polinomio cúbico a trozos continuo.

3.3 Funciones de interpolación fractal spline.

Para el cálculo del momento espectral de orden 2 es necesario usar la derivada de la función de interpolación fractal. El siguiente teorema asegura la existencia de FIF's diferenciables.

Teorema (Barnsley y Harrington, 1989): Sea $t_0 < t_1 < t_2 < ... < t_N$. Para $n = 1, 2, ..., N$ se consideran las aplicaciones afines $L_n(t) = a_n t + b_n$ verificando [3.1], [3.2]. Sea $F_n(t, x) = \alpha_n x + q_n(t)$, $n = 1, 2, ..., N$ verificando [3.3] y [3.4]. Se supone que para algún entero $p \geq 0$, $|\alpha_n| \leq a_n^p$ y $q_n \in C^p[t_0, t_N]$; $n = 1, 2, ..., N$. Sea

$$F_{nk}(t, x) = \frac{\alpha_n x + q_n^{(k)}(t)}{a_n^k} \qquad k = 1, 2, ..., p \qquad [3.9]$$

$$x_{0,k} = \frac{q_1^{(k)}(t_0)}{a_1^k - \alpha_1} \qquad x_{N,k} = \frac{q_N^{(k)}(t_N)}{a_N^k - \alpha_N} \qquad k = 1, 2, ..., p$$

Si

$$F_{n-1,k}(t_N, x_{N,k}) = F_{nk}(t_0, x_{0,k}) \qquad [3.10]$$

con $n = 2, 3, ..., N$ y $k = 1, 2, ..., p$, entonces $\{(L_n(t), F_n(t, x))\}_{n=1}^N$ determina una FIF $f \in C^p[t_0, t_N]$ y $f^{(k)}$ es la FIF determinada por $\{(L_n(t), F_{nk}(t, x))\}_{n=1}^N$, para $k = 1, 2, ..., p$.

En nuestro caso, q_n son polinomios de grado 3. Se estudiará en el presente trabajo el caso en el que $f \in C^{(2}$ por ser la generalización de los splines cúbicos. En este caso ha de cumplirse que $|\alpha_n| < a_n^2$ y basta imponer la condición de que $|\alpha_n| < (1/N)^2$ ([3.8]), $n = 1, 2, ..., N$ y construir los F_n según [3.9] y [3.10] para f' y f'':

$$F_{n1}(t, x) = N\alpha_n x + N q_n'(t)$$

$$F_{n2}(t, x) = N^2 \alpha_n x + N^2 q_n''(t)$$

Al imponer estas condiciones, los polinomios q_n quedan expresados en términos del único parámetro α_n (Barnsley y Harrington, 1989). Si $\alpha_n = 0$, $f(t) = q_n \circ L_n^{-1}(t) \ \forall \ t \in I_n$ (FIF) es un polinomio cúbico a trozos y $f \in C^{(2}$, por tanto es un spline cúbico.

4. Cuadratura de la movilidad de una señal electroencefalográfica

Se pretende calcular la movilidad de una señal, de la cual en la práctica sólo se tienen datos muestreados. Se proponen las técnicas de reconstrucción y ajuste de dicha señal mediante las funciones de interpolación fractal expuestas en el apartado anterior. A partir de aquí se calculan los momentos espectrales de la señal mediante fórmulas explícitas que vienen dadas en términos de los coeficientes del sistema de funciones iteradas que definen la función de interpolación fractal y de los momentos temporales de la señal, que se definen a continuación.

4.1 Momentos temporales de la señal.

Se comienza hallando los momentos temporales en función de los coeficientes del sistema de funciones iteradas que definen la función de interpolación de interpolación fractal cúbica.

Dada una señal $f(t)$ definida en el intervalo $I = [0, T]$, se definen los momentos temporales como:

$$M_n = \int_0^T t^n f(t) dt$$

Para calcular M_0 se utiliza la función de interpolación fractal asociada a un SFI general del tipo [3.7]. En el siguiente desarrollo se tiene en cuenta que la FIF es el punto fijo de la aplicación $T : \mathcal{F} \to \mathcal{F}$ definida en [3.6].

$$M_0 = \int_{t_0}^{t_N} f(t)\ dt = \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (Tf)(t)\ dt = \sum_{n=1}^{N} \int_{t_0}^{t_N} [\alpha_n f(t) + q_n(t)]\ d(a_n t + b_n) =$$

$$= \sum_{n=1}^{N} a_n \alpha_n \int_{t_0}^{t_N} f(t)\ dt + \sum_{n=1}^{N} a_n \int_{t_0}^{t_N} q_n(t)\ dt =$$

$$= \sum_{n=1}^{N} a_n \alpha_n M_0 + \sum_{n=1}^{N} a_n \int_{t_0}^{t_N} q_n(t)\ dt$$

donde se ha realizado el cambio de variable $t = L_n(\tilde{t})$. Despejando de aquí el momento M_0 se obtiene:

$$M_0 = \frac{\sum_{n=1}^{N} a_n \int_{t_0}^{t_N} q_n(t)\ dt}{1 - \sum_{n=1}^{N} a_n \alpha_n} \qquad [4.1]$$

Considerando que la partición del intervalo $[t_0, t_N]$ se hace de modo equidistante, entonces $a_n = 1/N$. Llamando $\alpha = \sum_{n=1}^{N} \alpha_n$ y $\beta = \sum_{n=1}^{N} \int_{t_0}^{t_N} q_n(t)\ dt$ se obtiene la expresión:

$$M_0 = \frac{\beta}{N - \alpha} \qquad [4.2]$$

Los momentos temporales de orden superior pueden calcularse usando la fórmula obtenida en el teorma 3 de la referencia (Barnsley, 1986)

$$M_m = \left(\sum_{k=0}^{m-1} M_k \binom{m}{k} \sum_{n=1}^{N} a_n^{k+1} \alpha_n b_n^{m-k} + Q_m \right) \Big/ \left(1 - \sum_{n=1}^{N} a_n^{m+1} \alpha_n \right) \qquad [4.3]$$

siendo $Q_m = \int_I t^m Q(t)\ dt$ y $Q(t) = q_n \circ L_n^{-1}(t)$ si $t \in I_n$

Considerando que en el caso en estudio $a_n = 1/N$ y $b_n = t_{n-1}$ se simplifica como:

$$M_m = \Big(\sum_{k=0}^{m-1} M_k \binom{m}{k} (\frac{1}{N})^{k+1} \sum_{n=1}^{N} \alpha_n t_{n-1}^{m-k} + Q_m\Big) \Big/ \Big(1 - (\frac{1}{N})^{m+1} \sum_{n=1}^{N} \alpha_n\Big) \qquad [4.4]$$

4.2 Momentos espectrales de la señal.

Se denota por $\mathcal{F}$ el conjunto de funciones continuas $f : [t_0, t_N] \to \mathbb{R}$ tales que $f(t_0) = x_0$; $f(t_N) = x_N$, y se considera la función $T : \mathcal{F} \to \mathcal{F}$ definida por:

$$(Tf)(t) = \alpha_n f(L_n^{-1}(t)) + q_n(L_n^{-1}(t)) \quad \forall t \in [t_{n-1}, t_n], \quad n = 1, 2, ..., N$$

Según se ha expuesto en los apartados anteriores ([3.6]), T posee un único punto fijo $f \in \mathcal{F}$ de modo que $(Tf)(t) = f(t)\ \forall\ t \in [t_0, t_N]$, y f es la FIF asociada con el SFI.

Para la obtención de $m_0 = \frac{1}{t_N - t_0} \int_{t_0}^{t_N} (f(t))^2\ dt$ se considera la integral $I_0 = \int_{t_0}^{t_N} (f(t))^2\ dt$. Realizando un cambio de variable afín, se tiene:

$$I_0 = \int_{t_0}^{t_N} (f(t))(f(t))\ dt = \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (Tf(t))(Tf(t))\ dt =$$

$$= \sum_{n=1}^{N} \int_{t_0}^{t_N} [\alpha_n f(t) + q_n(t)][\alpha_n f(t) + q_n(t)]\ d(a_n t + b_n)$$

Se considera a partir de ahora $a_n = 1/N$:

$$I_0 = \frac{1}{N} \sum_{n=1}^{N} \int_{t_0}^{t_N} [\alpha_n f(t) + q_n(t)][\alpha_n f(t) + q_n(t)]\ dt =$$

$$= \frac{1}{N} \Big[(\sum_{n=1}^{N} \alpha_n^2) I_0 + 2 \sum_{n=1}^{N} (\alpha_n \int_{t_0}^{t_N} f(t) q_n(t)\ dt) + N J_0 \Big] \qquad [4.5]$$

El tercer sumando puede expresarse como el momento de orden 0 de otra función

$$J_0 = \sum_{n=1}^{N} \frac{1}{N} \int_{t_0}^{t_N} (q_n(\tilde{t}))^2\ d\tilde{t} = \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (q_n \circ L_n^{-1}(t))^2\ dt$$

(con $t = L_n(\tilde{t})$), por tanto $J_0 = \int_{t_0}^{t_N} (Q(t))^2\ dt$ siendo $Q(t) = q_n \circ L_n^{-1}(t)$ si $t \in I_n$.

En el caso de que $\alpha_n = 0\ \forall\ n = 1, 2, ..., N$, $Q(t)$ es la FIF y, además es un spline cúbico, por tanto J_0 sería la integral correspondiente al momento espectral de orden 0 del spline cúbico que interpola los datos.

Sustituyendo $q_n(t) = q_{3n} t^3 + q_{2n} t^2 + q_{1n} t + q_{0n}$ se puede expresar I_0 en términos de los momentos temporales. Se calcula primero el sumando

$$\sum_{n=1}^{N} \alpha_n \int_{t_0}^{t_N} f(t) q_n(t) dt = \sum_{n=1}^{N} \alpha_n \int_{t_0}^{t_N} [q_{3n} t^3 f(t) + q_{2n} t^2 f(t) + q_{1n} t f(t) + q_{0n} f(t)] dt$$

$$= \sum_{n=1}^{N} \alpha_n q_{3n} \int_{t_0}^{t_N} t^3 f(t)\, dt + \sum_{n=1}^{N} \alpha_n q_{2n} \int_{t_0}^{t_N} t^2 f(t)\, dt +$$

$$+ \sum_{n=1}^{N} \alpha_n q_{1n} \int_{t_0}^{t_N} t f(t)\, dt + \sum_{n=1}^{N} \alpha_n q_{0n} \int_{t_0}^{t_N} f(t)\, dt =$$

$$= M_3(\sum_{n=1}^{N} \alpha_n q_{3n}) + M_2(\sum_{n=1}^{N} \alpha_n q_{2n}) + M_1(\sum_{n=1}^{N} \alpha_n q_{1n}) + M_0(\sum_{n=1}^{N} \alpha_n q_{0n})$$

Denotando por $\theta = \sum_{n=1}^{N} \alpha_n^2$ y despejando I_0 se tiene:

$$I_0 = (\frac{1}{N-\theta})\Big[2M_3(\sum_{n=1}^{N} \alpha_n q_{3n}) + 2M_2(\sum_{n=1}^{N} \alpha_n q_{2n}) +$$

$$+2M_1(\sum_{n=1}^{N} \alpha_n q_{1n}) + 2M_0(\sum_{n=1}^{N} \alpha_n q_{0n}) + N J_0\Big] \qquad [4.6]$$

de donde se obtiene inmediatamente el valor de m_0. Para calcular m_1 se puede utilizar el procedimiento anterior, sabiendo que en este caso la FIF es f' y que el SFI que la determina según [3.9] es $\{(L_n(t), F_{n1}(t,x))\}_{n=1}^{N}$ donde

$$\begin{cases} L_n(t) = a_n t + b_n \\ \\ F_{n1}(t,x) = N\alpha_n x + N q_n'(t) \end{cases} \qquad [4.7]$$

con $q_n'(t) = 3q_{3n}t^2 + 2q_{2n}t + q_{1n}$. En este caso la función $T_1 : \mathcal{F} \to \mathcal{F}$ dada por

$$(T_1 g)(t) = N(\alpha_n g(L_n^{-1}(t)) + q_n'(L_n^{-1}(t))) \;\; \forall\, t \in [t_{n-1}, t_n], \;\; n = 1, 2, ..., N$$

posee un único punto fijo que coincide con f':

$$(T_1 f')(t) = f'(t) \;\; \forall\, t \in [t_0, t_N]$$

Se puede calcular m_2 como:

$$m_2 = \frac{1}{t_N - t_0} \int_{t_0}^{t_N} (f'(t))^2\, dt = \frac{1}{Nh} I_1$$

donde I_1 se obtiene de la expresión [4.6] cambiando las constantes en función de la nueva FIF asociada a este caso y teniendo en cuenta que los polinomios son de segundo grado, $q_n^1(t) = N q_n'(t)$ y $\alpha_n^1 = N\alpha_n$. Mediante cálculos análogos a los de I_0, se tiene que:

$$I_1 = \frac{N}{1 - N\theta}[6M_2^1(\sum_{n=1}^{N} \alpha_n q_{3n}) + 4M_1^1(\sum_{n=1}^{N} \alpha_n q_{2n}) + 2M_0^1(\sum_{n=1}^{N} \alpha_n q_{1n}) + \frac{J_1}{N}] \qquad [4.8]$$

donde M_0^1, M_1^1, M_2^1 son los momentos temporales de f', $J_1 = \int_{t_0}^{t_N} (Q^1(t))^2\, dt$ y $Q^1(t) = q_n^1 \circ L_n^{-1}(t)$ si $t \in I_n$, de donde se deduce inmediatamente el valor de m_2. Por lo dicho anteriormente, si $\alpha_n = 0$, J_1 sería la integral que define el momento espectral de orden 2 del spline cúbico que interpola los datos. El cálculo detallado puede encontrarse en la referencia (Sebastián, 2001).

5. Acotación de los errores de aproximación

Se pretende dar una aproximación del error cometido al calcular los momentos espectrales de la señal usando funciones de interpolación fractal. En primer lugar se acotará el error cometido al sustituir la señal $x(t)$ por la FIF $f_\alpha(t)$ que tiene como factores de escala vertical $\alpha = (\alpha_1,\ \alpha_2, ..., \alpha_n)$, suponiendo que f_α es la FIF cúbica correspondiente a un SFI verificando [3.9], [3.10] para $p = 2$. Posteriormente se acotará el error cometido al integrar la FIF y su derivada para hallar los momentos espectral de orden 0 y 2.

5.1 Acotaciones iniciales.

Para la acotación de la aproximación se va a utilizar un resultado concerniente a las funciones spline cúbicas.

Teorema (Hall y Meyer, 1976):

Sea $f \in C[a,b]$ se define $\|f\|_\infty = sup|f(t)|$ cuando $t \in [a,b]$. Sea $f \in C^4[a,b]$ y $|f^{(4)}(t)| \leq L$ para todo $t \in [a,b]$. Sea una partición del intervalo $[a,b]$, $\Delta = \{a = t_0 < t_1 < ... < t_N = b\}$ equiespaciada de paso h. Sea S_Δ la función spline que interpola los valores de la función f en los nodos $t_o, t_1, ..., t_N \in \Delta$ y S_Δ es del tipo I o del tipo II. Entonces se verifica

$$\|f^{(r)} - S_\Delta^{(r)}\|_\infty \leq C_r\ L\ h^{4-r} \quad (r = 0, 1, 2)$$

con $C_0 = 5/384$, $\quad C_1 = 1/24$, $\quad C_2 = 3/8$. Las constantes C_0 y C_1 son óptimas.

NOTA: Un spline es de tipo I si sus derivadas primeras en a y b son conocidas. Un spline de tipo II es aquel que se representa explícitamente mediante las segundas derivadas en a y b.

Se considera la aplicación

$$T : J \times \mathcal{F} \to \mathcal{F}$$

$$(\alpha, f) \to T_\alpha f$$

con $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$ y J el intervalo $J = [0,r] \times [0,r] \times ... \times [0,r]$; $0 \leq r < 1$; r fijo y $[t_0, t_N] = I$.

$$T_\alpha f(t) = F_n^{\alpha_n}(L_n^{-1}(t), f \circ L_n^{-1}(t)) = \alpha_n f \circ L_n^{-1}(t) + q_n^{\alpha_n} \circ L_n^{-1}(t) \qquad [5.1]$$

El superíndice α_n representa la dependencia de F_n respecto el factor de escala vertical. El polinomio $q_n(t) = q_{3n}t^3 + q_{2n}t^2 + q_{1n}t + q_{0n}$, y $t \in [t_{n-1}, t_n] = I_n$ y queda expresado en función de α_n, es decir, $q_n^{\alpha_n}(t) = q_n(\alpha_n, t)$ (Barnsley y Harrington, 1989). Se sabe que el punto fijo de T_α es la FIF (Teorema de Barnsley).

Proposición 5.1. Dadas $f_1,\ f_2\ \in \mathcal{F}$ se verifica que:

$$\|T_\alpha f_1 - T_\alpha f_2\|_\infty \leq |\alpha|_\infty\ \|f_1 - f_2\|_\infty$$

siendo $|\alpha|_\infty = \max_n\{|\alpha_n|\}$.

Demostración

Para $t \in I_n = [t_{n-1}, t_n]$ se dan las siguientes desigualdades:

$$|T_\alpha f_1(t) - T_\alpha f_2(t)| \underset{t\in I_n}{=} |\alpha_n f_1 \circ L_n^{-1}(t) + q_n \circ L_n^{-1}(t) - \alpha_n f_2 \circ L_n^{-1}(t) - q_n \circ L_n^{-1}(t)| =$$

$$= |\alpha_n||f_1 \circ L_n^{-1}(t) - f_2 \circ L_n^{-1}(t)| \leq$$

$$\leq \max_n\{|\alpha_n|\}\ |f_1 \circ L_n^{-1}(t) - f_2 \circ L_n^{-1}(t)| \leq |\alpha|_\infty\ \|f_1 - f_2\|_\infty$$

de modo que

$$\|T_\alpha f_1 - T_\alpha f_2\|_\infty \leq |\alpha|_\infty\ \|f_1 - f_2\|_\infty \qquad [5.2]$$

Proposición 5.2. Sean $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, $\beta = (\beta_1, \beta_2, ..., \beta_n)$, $f \in \mathcal{F}$ y supongamos que $q_n(\alpha_n, t)$ es diferenciable y que $\exists\ C \geq 0$ tal que $|\frac{\partial q_n}{\partial \alpha_n}(\xi, t)| \leq C\ \forall\ (\xi, t) \in J \times I$ y $\forall\ n = 1, 2, ..., N$. Entonces se tiene:

$$\|T_\alpha f - T_\beta f\|_\infty \leq |\alpha - \beta|_\infty(\|f\|_\infty + C)$$

Demostración

Se pretende acotar la diferencia $|T_\alpha f(t) - T_\beta f(t)|$. Dada $f \in \mathcal{F}$ para cada valor $t\ \in I_n$ se tiene:

$$|T_\alpha f(t) - T_\beta f(t)| = |\alpha_n f \circ L_n^{-1}(t) + q_n^{\alpha_n} \circ L_n^{-1}(t) - \beta_n f \circ L_n^{-1}(t) - q_n^{\beta_n} \circ L_n^{-1}(t)| \leq$$

$$|\alpha_n f \circ L_n^{-1}(t) - \beta_n f \circ L_n^{-1}(t)| + |q_n^{\alpha_n} \circ L_n^{-1}(t) - q_n^{\beta_n} \circ L_n^{-1}(t)|$$

Para el primer sumando se verifica la desigualdad:

$$|\alpha_n f \circ L_n^{-1}(t) - \beta_n f \circ L_n^{-1}(t)| \leq |\alpha_n - \beta_n|\ |f \circ L_n^{-1}(t)| \leq |\alpha - \beta|_\infty\ \|f\|_\infty \qquad [5.3]$$

Para acotar el segundo sumando se aplica el teorema del valor medio para funciones de varias variables. Con las hipótesis del enunciado: $\exists\ (\xi, \tilde{t}) \in\ J \times I$ tal que

$$q_n(\alpha_n, \tilde{t}) - q_n(\beta_n, \tilde{t}) = \frac{\partial q_n}{\partial \alpha_n}(\xi, \tilde{t})(\alpha_n - \beta_n)$$

y, por tanto,

$$|q_n^{\alpha_n} \circ L_n^{-1}(t) - q_n^{\beta_n} \circ L_n^{-1}(t)| \leq C|\alpha_n - \beta_n| \leq C|\alpha - \beta|_\infty \qquad [5.4]$$

De las desigualdades [5.3] y [5.4] se obtiene el resultado.

Proposición 5.3. Sean $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, $\beta = (\beta_1, \beta_2, ..., \beta_n)$, f_α, f_β las correspondientes funciones de interpolación fractal con factores de escala vertical α y β. Con las hipótesis de la proposición 5.2, se tiene:

$$\|f_\alpha - f_\beta\|_\infty \leq \frac{1}{1-|\alpha|_\infty}|\alpha - \beta|_\infty(\|f_\beta\|_\infty + C)$$

Demostración

Por definición f_α, f_β son los puntos fijos de T_α y T_β, respectivamente. Por tanto $T_\alpha(f_\alpha) = f_\alpha$, $T_\beta(f_\beta) = f_\beta$. Aplicando las proposiciones 5.1 y 5.2, se tiene:

$$\|f_\alpha - f_\beta\|_\infty = \|T_\alpha f_\alpha - T_\alpha f_\beta + T_\alpha f_\beta - T_\beta f_\beta\|_\infty \leq$$

$$\leq \|T_\alpha f_\alpha - T_\alpha f_\beta\|_\infty + \|T_\alpha f_\beta - T_\beta f_\beta\|_\infty \leq$$

$$\leq |\alpha|_\infty \|f_\alpha - f_\beta\|_\infty + |\alpha - \beta|_\infty(\|f_\beta\|_\infty + C)$$

Despejando se tiene:

$$\|f_\alpha - f_\beta\|_\infty \leq \frac{1}{1-|\alpha|_\infty}|\alpha - \beta|_\infty(\|f_\beta\|_\infty + C) \qquad [5.5]$$

Consecuencia

Haciendo $\beta = 0$ en la proposición anterior

$$\|f_\alpha - f_0\|_\infty \leq \frac{1}{1-|\alpha|_\infty}|\alpha|_\infty(\|f_0\|_\infty + C) \qquad [5.6]$$

f_0 es un spline cúbico que interpola los datos. Se puede mayorar $\|f_0\|_\infty$ aplicando el teorema de Hall y Meyer:

$$\|f_0\|_\infty \leq k_0 h^4 + \|x\|_\infty$$

Llamando $\|x\|_\infty = L_0$:

$$\|f_\alpha - f_0\|_\infty \leq \frac{1}{1-|\alpha|_\infty}|\alpha|_\infty(k_0 h^4 + L_0 + C) \qquad [5.7]$$

5.2 Acotación del error de interpolación.

Teorema 5.4. Sea $x(t)$ una señal verificando $x(t) \in C^4[t_0, t_N]$ y $|x^{(4)}(t)| \leq L \ \forall \ t \in [t_0, t_N]$. Supongamos que $q_n(\alpha_n, t)$ es diferenciable y que $\exists \ C \geq 0$ tal que $|\frac{\partial q_n}{\partial \alpha_n}(\xi, t)| \leq C$ $\forall \ (\xi, t) \in J \times I$, $\forall \ n = 1, 2, ..., N$. Entonces

$$\|x - f_\alpha\|_\infty \leq \frac{N^2}{N^2 - 1}[k_0 h^4 + \frac{(L_0 + C)}{T^2}h^2]$$

siendo k_0 la constante del teorema de Hall y Meyer y $L_0 = \|x\|_\infty$.

Demostración

$$\|x - f_\alpha\|_\infty \leq \|x - f_0\|_\infty + \|f_0 - f_\alpha\|_\infty$$

El primer sumando puede acotarse aplicando el teorema de Hall y Meyer:

$$\|x - f_0\|_\infty \leq k_0 h^4 \qquad [5.8]$$

Para la acotación del segundo sumando, se utiliza la consecuencia de la proposición 5.3:

$$\|f_0 - f_\alpha\|_\infty \leq \frac{1}{1-|\alpha|_\infty}|\alpha|_\infty(k_0 h^4 + L_0 + C) \qquad [5.9]$$

De [5.8] y [5.9] se obtiene:

$$\|x - f_\alpha\|_\infty \leq \frac{1}{1-|\alpha|_\infty}[k_0 h^4 + |\alpha|_\infty(L_0 + C)]$$

Además, por la hipótesis del teorema de diferenciabilidad de las funciones de interpolación fractal de este tipo: $|\alpha|_\infty < \frac{1}{N^2} = \frac{h^2}{T^2}$ y, por tanto, $\frac{1}{1-|\alpha|_\infty} < \frac{N^2}{N^2-1}$, de modo que la desigualdad anterior se transforma en:

$$\|x - f_\alpha\|_\infty \leq \frac{N^2}{N^2-1}[k_0 h^4 + \frac{(L_0 + C)}{T^2}h^2]$$

5.3 Acotaciones de la derivada.

A continuación se obtendrá un resultado similar para la derivada de la función de interpolación fractal. En este caso $\mathcal{F}^1$ es el conjunto de funciones continuas g tales que

$$g(t_0) = \frac{q_1'(t_0)}{a_1 - \alpha_1} \quad g(t_N) = \frac{q_N'(t_N)}{a_N - \alpha_N}$$

$F_{n1}(t,x) = N\alpha_n x + Nq_n'(t)$ y

$$T_\alpha^1 f(t) = N\alpha_n f \circ L_n^{-1}(t) + Nq_n' \circ L_n^{-1}(t) \quad \forall t \in I_n \qquad [5.10]$$

siendo $q_n'(\alpha_n, t) = \frac{\partial q_n}{\partial t}(\alpha_n, t)$.

Proposición 5.5. Dadas $f_1,\ f_2 \in \mathcal{F}^1$ se verifica que:

$$\|T_\alpha^1 f_1 - T_\alpha^1 f_2\|_\infty \leq N|\alpha|_\infty \ \|f_1 - f_2\|_\infty$$

Demostración

Se deduce inmediatamente de la proposición 5.1 teniendo en cuenta que en este caso, el factor de escala vertical es $N\alpha_n$.

Proposición 5.6. Sean $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, $\beta = (\beta_1, \beta_2, ..., \beta_n)$, $f \in \mathcal{F}^1$ y supongamos que $\frac{\partial q_n}{\partial t}(\alpha_n, t)$ es diferenciable y que $\exists\ C_1 \geq 0$ tal que $|\frac{\partial^2 q_n}{\partial \alpha_n \partial t}(\xi, t)| \leq C_1\ \forall\ (\xi, t) \in J \times I$ y $\forall\ n = 1, 2, ..., N$. Entonces:

$$\|T_\alpha^1 f - T_\beta^1 f\|_\infty \leq N|\alpha - \beta|_\infty(\|f\|_\infty + C_1)$$

Demostración

Se trata del mismo resultado de la proposición 5.2, aplicado a la función de interpolación fractal f' con factor de escala vertical $N\alpha_n$ y polinomio $Nq'_n(t)$.

Proposición 5.7. Sean $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, $\beta = (\beta_1, \beta_2, ..., \beta_n)$, y sean f'_α, f'_β las funciones de interpolación fractal correspondientes a las aplicaciones T^1_α, T^1_β. Con las hipótesis de la proposición 5.6, se tiene:

$$\|f'_\alpha - f'_\beta\|_\infty \leq \frac{1}{1 - N|\alpha|_\infty}|\alpha - \beta|_\infty(\|f'_\beta\|_\infty + C_1)$$

Demostración

Análoga a la de la proposición 5.3.

Consecuencia

Haciendo $\beta = 0$ en la proposición anterior

$$\|f'_\alpha - f'_0\|_\infty \leq \frac{N|\alpha|_\infty}{1 - N|\alpha|_\infty}(\|f'_0\|_\infty + C_1)$$

f'_0 es la derivada del spline cúbico que interpola los datos. Según el teorema de Hall y Meyer $\|f'_0\|_\infty \leq k_1 h^3 + \|x'\|_\infty$ y, por tanto,

$$\|f'_\alpha - f'_0\|_\infty \leq \frac{N|\alpha|_\infty}{1 - N|\alpha|_\infty}(k_1 h^3 + \|x'\|_\infty + C_1)$$

5.4 Acotación del error de interpolación de la derivada.

Teorema 5.8. Sea $x(t)$ una señal verificando $x(t) \in C^4[t_0, t_N]$ y $|x^{(4)}(t)| \leq L$ para todo $t \in [t_0, t_N]$. Supongamos que $\frac{\partial q_n}{\partial t}(\alpha_n, t)$ es diferenciable y que $\exists\ C_1 \geq 0$ tal que $|\frac{\partial^2 q_n}{\partial \alpha_n \partial t}(\xi, t)| \leq C_1\ \forall\ (\xi, t) \in J \times I$ y $\forall\ n = 1, 2, ..., N$. Entonces:

$$\|x' - f'_\alpha\|_\infty \leq \frac{N}{N-1}[k_1 h^3 + \frac{(L_1 + C_1)}{T}h]$$

siendo k_1 la constante del teorema de Hall y Meyer y $L_1 = \|x'\|_\infty$

Demostración

$$\|x' - f'_\alpha\|_\infty \leq \|x' - f'_0\|_\infty + \|f'_0 - f'_\alpha\|_\infty$$

El primer sumando puede acotarse utilizando el teorema de Hall y Meyer:

$$\|x' - f'_0|_\infty \leq k_1 h^3 \qquad [5.11]$$

siendo $k_1 = L\frac{1}{24}$.

Para la acotación del segundo sumando, se utiliza la consecuencia de la proposición 5.7:

$$\|f'_0 - f'_\alpha\|_\infty \leq \frac{N|\alpha|_\infty}{1 - N|\alpha|_\infty}(k_1 h^3 + \|x'\|_\infty + C_1) \qquad [5.12]$$

Denotando $\|x'\|_\infty = L_1$, de [5.11] y [5.12], se obtiene:

$$\|x' - f'_\alpha\|_\infty \leq \frac{1}{1 - N|\alpha|_\infty}(k_1 h^3 + N|\alpha|_\infty(L_1 + C_1))$$

Como $|\alpha|_\infty < \frac{1}{N^2}$, $\frac{1}{1-N|\alpha|_\infty} < \frac{N}{N-1}$. Además $\frac{1}{N} = \frac{h}{T}$, por tanto:

$$\|x' - f'_\alpha\|_\infty \leq \frac{N}{N-1}[k_1 h^3 + \frac{(L_1 + C_1)}{T}h]$$

5.5 Acotación de los errores en la raíz de los momentos de orden 2 y 4.

Teorema 5.9. Sea $x(t)$ una señal verificando $x(t) \in C^4[t_0, t_N]$ y $|x^{(4)}(t)| \leq L$ para todo $t \in [t_0, t_N]$. Supongamos que se verifican las hipótesis de las proposiciones 5.2 y 5.6 para las funciones $q_n(\alpha_n, t)$. Sean

$$m_{0_{exac}} = \frac{1}{T}\int_{t_0}^{t_N} (x(t))^2\, dt$$

$$m_{2_{exac}} = \frac{1}{T}\int_{t_0}^{t_N} (x'(t))^2\, dt$$

el momento espectral de orden 0 y 2 de la señal inicial; y

$$m_{0_{aprox}} = \frac{1}{T}\int_{t_0}^{t_N} (f_\alpha(t))^2\, dt$$

$$m_{2_{aprox}} = \frac{1}{T}\int_{t_0}^{t_N} (f'_\alpha(t))^2\, dt$$

Se tienen las siguientes desigualdades para E_0 y E_2, errores en la raíz de los momentos de orden 0 y 2 respectivamente:

$$|E_0| \leq \frac{N^2}{N^2 - 1}[k_0 h^4 + \frac{(L_0 + C)}{T^2}h^2]$$

$$|E_2| \leq \frac{N}{N-1}[k_1 h^3 + \frac{(L_1 + C_1)}{T}h]$$

siendo k_0 y k_1 las constantes del teorema de Hall y Meyer, $L_0 = \|x\|_\infty$, $L_1 = \|x'\|_\infty$ y C, C_1 las constantes de las proposiciones 5.2 y 5.6.

Demostración

Se define la norma

$$\|f\| = \frac{1}{\sqrt{T}}\left[\int_I (f(t))^2\, dt\right]^{1/2}$$

y la seminorma

$$\|f\|_1 = \frac{1}{\sqrt{T}}\left[\int_I (f'(t))^2\, dt\right]^{1/2}$$

$$|E_0| = |m_{0_{exac}}^{1/2} - m_{0_{aprox}}^{1/2}| = |\ \|x\| - \|f_\alpha\|\| \leq$$

$$\leq \|x - f_\alpha\| = \frac{1}{\sqrt{T}}\left[\int_{t_0}^{t_N} [x(t) - f_\alpha(t)]^2\, dt\right]^{1/2}$$

Aplicando en la integral anterior el teorema 5.4

$$|E_0| \leq \frac{1}{\sqrt{T}} \sqrt{T} \frac{N^2}{N^2-1} [k_0 h^4 + \frac{(L_0 + C)}{T^2} h^2]$$

Utilizando un desarrollo similar, se obtiene la segunda expresión.

Referencias

[1] Barnsley M F. Fractal functions and interpolation. Constr. Approx. 2. 1986; 4: 303-329.

[2] Barnsley M. Fractals Everywhere. Academic Press, Inc. 1988.

[3] Barnsley M F, Harrington AN. The calculus of fractal interpolation functions. J. Approx. Theory. 1989; 57: 14-34.

[4] Edgar GA. Measure, Topology and Fractal Geometry. Springer-Verlag, New York, 1990.

[5] Hall CA, Meyer WW. Optimal error bounds for cubic splines interpolation. J. Approx. Theory. 1976; 16: 105-122.

[6] Hardin DP, Kessler B, Massopust PR. Multiresolution analyses based on fractal functions. J. Approx. Theory. 1992; 71: 104-120.

[7] Hjorth, B. EEG analysis based on time domain properties. Electroen. Clin. Neuro. 1970; 29: 306-310.

[8] Hsu HP. Análisis de Fourier. Addison-Wesley Iberoamericana, S.A. Wilmington, 1987.

[9] Sebastián MV. Dinámica no lineal de registros electrofisiológicos. (Tesis Doctoral). Universidad de Zaragoza. 2001.

[10] Stoer J, Bulirsch R. Introduction to Numerical Analysis. Springer-Verlag. New York, 1980.

Rev. Academia de Ciencias. Zaragoza. **56** *(2001)*

OPTIMAL HOMOGENEOUS LINEAR ESTIMATION FOR A SUPERPOPULATION MODEL

MARIANO RUIZ ESPEJO

UNED and UPSAM, Apartado 19207, 28080 Madrid, Spain

HOUSILA P. SINGH, RAJESH SINGH

School of Studies in Statistics, Vikram University, Ujjain 456010, M.P., India

and

SARALEES NADARAJAH

Dept. of Statistics, University of California, Riverside, CA 92521-0138, USA

Summary. A class of homogeneous linear estimators has been studied. For simple random sampling without replacement design there does not exist an optimal estimator for a finite population mean. However, for an uncorrelated general model, there is a unique optimal estimator (predictor) which is identifiable when the coefficient of variation of the superpopulation model is known.

Key words: Coefficient of variation, optimal estimation, superpopulation model.

AMS Classification: 62D05.

1. Introduction

Let $U = \{1, 2, ..., i, ..., N\}$ denote a finite population of N distinct and identifiable units, uniquely labelled from 1 to N. Let y denote the study variable taking positive values, y_i for the ith unit of the population U. For estimating the finite population mean

$$\bar{Y} = \frac{1}{N}\sum_{i \in U} y_i, \qquad 1\bullet1$$

a simple random sample $s \subset U$ of size n is drawn without replacement from U. Under simple random sampling without replacement (SRSWOR) design p, we define a class of estimators for $\bar{Y}$ as

$$t = \sum_{i \in s} t_i y_i, \qquad 1\bullet2$$

where t_i's are real constants.

It will be shown that there does not exist optimal estimator (uniformly minimum mean squared error (UMMSE) estimator) in the class of estimators 't' given in (1•2). However under an uncorrelated and positive model for which $y_i(> 0)$ is selected from an identically distributed and positive Y_i with

$$E_M(Y_i) = \mu(> 0), \quad V_M(Y_i) = \sigma^2(> 0) \quad \text{and} \quad Cov_M(Y_i, Y_j) = 0 \quad \text{if} \quad i \neq j \in U. \qquad 1\bullet3$$

It will be shown with the usual criterion of uniformly minimum average mean squared error (UMAMSE) that there exists a unique optimal predictor in the class 't' in (1•2), for instance, see Hedayat and Sinha (1991, Chapter 10). This average mean squared error (*AMSE*) has been considered by Chaubey, Singh and Dwivedi (1984) and recently by Singh, Singh and Ruiz Espejo (1998).

The optimal predictor depends on the coefficient of variation $c(= \sigma/\mu)$, and for this reason

coefficient of variation should be known for the construction of 'optimal predictor'. On the other hand, when 'c' is unknown, the optimal predictor does exist, but it is not implementable.

We will explain these assertions with technical justifications in the subsequent sections.

2. Design-based optimality

It is known that there does not exist an UMMSE estimator for the finite population mean $\bar{Y}$, for any non-census design, in the class of all estimators (Ruiz Espejo, 1987).

In this section we will prove that there does not exist UMMSE estimator for $\bar{Y}$ in the class of all homogeneous linear estimators 't' given in (1•2) under SRSWOR design p, the parametric space being

$$\Omega = \mathbb{R}_+^N = \{(y_1, \ldots, y_N) \in \mathbb{R}^N : y_i > 0, \forall i \in U\}.$$

Let s be a sample of size n ($2 \le n < N$) such that

$$S = \{s \subset U : \text{card}(s) = n\},$$

and for all samples $s \in S$, $p(s) = 1/\binom{N}{n}$.

The p-mean squared error (MSE) of a generic estimator of the type 't' in (1•2) is

$$MSE_p(t) = E_p\left[(t - \bar{Y})^2\right] = E_p(t^2) - 2\bar{Y}E_p(t) + \bar{Y}^2. \qquad 2\bullet1$$

Now,

$$E_p(t^2) = \sum_{s\in S}\left(\sum_{i\in s} t_i y_i\right)^2 p(s) = \sum_{s\in S}\left(\sum_{i\in s}\sum_{j\in s} t_i t_j y_i y_j\right)\frac{1}{\binom{N}{n}} =$$

$$= \sum_{i\in U}\sum_{j\in U} t_i t_j y_i y_j \text{card}\{s : i,j \in s\}\frac{1}{\binom{N}{n}} =$$

$$= \sum_{i\in U} t_i^2 y_i^2 \text{card}\{s : i \in s\}\frac{1}{\binom{N}{n}} + \sum_{i\neq j}\sum_{\in U} t_i t_j y_i y_j \text{card}\{s : i,j \in s\}\frac{1}{\binom{N}{n}} =$$

$$= \sum_{i\in U} t_i^2 y_i^2 \frac{\binom{N-1}{n-1}}{\binom{N}{n}} + \sum_{i\neq j}\sum_{\in U} t_i t_j y_i y_j \frac{\binom{N-2}{n-2}}{\binom{N}{n}} =$$

$$= \frac{n}{N}\sum_{i\in U} t_i^2 y_i^2 + \frac{n(n-1)}{N(N-1)}\sum_{i\neq j}\sum_{\in U} t_i t_j y_i y_j, \qquad 2\bullet2$$

$$E_p(t) = \sum_{s\in S}\left(\sum_{i\in s} t_i y_i\right)p(s) = \sum_{i\in U} t_i y_i \text{card}\{s : i \in s\}\frac{1}{\binom{N}{n}} =$$

$$= \frac{n}{N}\sum_{i\in U} t_i y_i. \qquad 2\bullet3$$

From (2•1), (2•2) and (2•3),

$$MSE_p(t) = E_p(t^2) - 2\bar{Y}E_p(t) + \bar{Y}^2 =$$

$$= \frac{n}{N}\sum_{i\in U} t_i^2 y_i^2 + \frac{n(n-1)}{N(N-1)}\sum_{i\neq j}\sum_{\in U} t_i t_j y_i y_j - 2\left(\frac{1}{N}\sum_{i\in U} y_i\right)\frac{n}{N}\left(\sum_{i\in U} t_i y_i\right) +$$

$$+ \frac{1}{N^2}\left(\sum_{i\in U} y_i\right)^2 = \frac{n}{N}\sum_{i\in U} t_i^2 y_i^2 + \frac{n(n-1)}{N(N-1)}\sum_{i\neq j}\sum_{\in U} t_i t_j y_i y_j$$

$$- \frac{2n}{N^2}\sum_{i\in U}\sum_{j\in U} y_i t_j y_j + \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U} y_i y_j =$$

$$= \left[\frac{n}{N} - \frac{n(n-1)}{N(N-1)} \right] \sum_{i \in U} t_i^2 y_i^2 +$$

$$+ \sum_{i \in U} \sum_{j \in U} y_i y_j \left[\frac{n(n-1)}{N(N-1)} t_i t_j - \frac{2n}{N^2} t_j + \frac{1}{N^2} \right]. \qquad 2\bullet 4$$

If we minimize the $MSE_p(t)$, given in (2•4), w.r.t. t_i ($i \in U$), we have the system of equations

$$\begin{cases} \dfrac{\partial MSE_p(t)}{\partial t_i} = 0 \\ i \in U \end{cases}$$

or

$$\begin{cases} 2t_i y_i^2 \dfrac{n}{N} - \dfrac{2n}{N^2} y_i^2 + \dfrac{2n(n-1)}{N(N-1)} y_i y_j \sum\limits_{j \in U: j \neq i} t_j = 0 \\ i \in U \end{cases}$$

or

$$\begin{cases} t_i = \dfrac{1}{N} - \dfrac{n-1}{N-1} \dfrac{y_j}{y_i} \sum\limits_{j \in U: j \neq i} t_j \\ i \in U. \end{cases} \qquad 2\bullet 5$$

From (2•5),

$$T = \sum_{i \in U} t_i = 1 - \frac{(n-1)y_j}{N-1} \sum_{i \in U} \sum_{j \in U: j \neq i} \frac{t_j}{y_i} =$$

$$= 1 - \frac{(n-1)y_j}{N-1} \sum_{i \in U} \frac{1}{y_i} (T - t_i) =$$

$$= 1 - \frac{(n-1)y_j}{N-1} \left(T \sum_{i \in U} \frac{1}{y_i} - \sum_{i \in U} \frac{t_i}{y_i} \right),$$

or

$$T \left[1 + \frac{(n-1)y_j}{N-1} \sum_{i \in U} \frac{1}{y_i} \right] = 1 + \frac{(n-1)y_j}{N-1} \sum_{i \in U} \frac{t_i}{y_i}.$$

If we sum for $j \in U$, and simplifying, we get

$$T \left[1 + \frac{(n-1)\bar{Y}}{N-1} \sum_{i \in U} \frac{1}{y_i} \right] = 1 + \frac{(n-1)\bar{Y}}{N-1} \sum_{i \in U} \frac{t_i}{y_i}. \qquad 2\bullet 6$$

Now, using the notation of population harmonic means

$$H_y = \frac{N}{\sum\limits_{i \in U} \frac{1}{y_i}} \quad \text{and} \quad H_{y/t} = \frac{N}{\sum\limits_{i \in U} \frac{t_i}{y_i}}$$

in (2•6), we have the following necessary condition of minimum:

$$T = \frac{[(N-1)H_{y/t} + (n-1)N\bar{Y}]H_y}{[(N-1)H_y + (n-1)N\bar{Y}]H_{y/t}}. \qquad 2\bullet 7$$

This condition is also sufficient because

$$\frac{\partial^2 [MSE_p(t)]}{\partial t_i^2} = \frac{2n}{N} y_i^2 > 0 \quad (i \in U)$$

and

$$\frac{\partial^2[MSE_p(t)]}{\partial t_i \partial t_j} = 2\frac{n(n-1)}{N(N-1)} y_i y_j > 0 \quad (i \neq j \in U).$$

However, as T depends on $(y_1, ..., y_N)$, the values of t_i's are not known constants. Thus we can not get a solution for $(t_1, ..., t_N)$, which yields UMMSE of t.

3. Model-based optimality

We consider that the observed value y_i is the realization of a random variable $Y_i(> 0)$ which is uncorrelated and identically distributed with

$$E_M(Y_i) = \mu, \quad V_M(Y_i) = \sigma^2 \quad \text{and} \quad Cov_M(Y_i, Y_j) = 0 \quad (i \neq j).$$

Using the criterion of minimum "average mean squared error" ($AMSE$) of t for predicting

$$\bar{Y} = \frac{1}{N}\sum_{i \in U} y_i,$$

we have

$$AMSE(t) = E_M[MSE_p(t)] =$$

$$= E_M[E_p(t^2)] - 2E_M[\bar{Y}E_p(t)] + E_M(\bar{Y}^2), \qquad 3\bullet1$$

where

$$A = E_M[E_p(t^2)] =$$

$$= \frac{n}{N}(\sigma^2 + \mu^2)\sum_{i \in U} t_i^2 + \frac{n(n-1)}{N(N-1)}\mu^2 \sum_{i \neq j}\sum_{\in U} t_i t_j, \qquad 3\bullet2$$

$$B = -2E_M[\bar{Y}E_p(t)] = -2\frac{n}{N^2}(\sigma^2 + N\mu^2)\sum_{i \in U} t_i \qquad 3\bullet3$$

and

$$C = E_M(\bar{Y}^2) = \frac{\sigma^2}{N} + \mu^2. \qquad 3\bullet4$$

From (3•1), (3•2), (3•3) and (3•4), the

$$AMSE(t) = A + B + C. \qquad 3\bullet5$$

Minimizing $AMSE(t)$, we have the necessary and sufficient conditions:

$$\begin{cases} \dfrac{\partial AMSE(t)}{\partial t_i} = 0 \\ i \in U \end{cases}$$

which gives the optimum values of t_i's as

$$\begin{cases} t_i = \dfrac{1}{\sigma^2 + \mu^2}\left[\dfrac{\sigma^2 + N\mu^2}{N} - \dfrac{(n-1)\mu^2}{N-1}(T - t_i)\right] \\ i \in U \end{cases}$$

or

$$\begin{cases} t_i = \dfrac{(N-1)(\sigma^2 + N\mu^2) - N(n-1)\mu^2 T}{N(N-1)(\sigma^2 + \mu^2) - N(n-1)\mu^2} \\ i \in U, \end{cases} \qquad 3\bullet6$$

where

$$T = \sum_{i \in U} t_i \Rightarrow T = \frac{\sigma^2 + N\mu^2}{\sigma^2 + n\mu^2}. \qquad 3\bullet7$$

We note from (3•7) that T is not usually equal to N/n unless when $\sigma^2 = 0$ (value excluded), and thus the sample mean is not the optimal predictor of the population mean $\bar{Y}$.

Substituting (3•7) in (3•6) we get the optimum values of constants

$$\begin{cases} t_i = \dfrac{(\sigma^2 + N\mu^2)\left[N - 1 - \dfrac{N(n-1)\mu^2}{\sigma^2 + n\mu^2}\right]}{N(N-1)\sigma^2 + N(N-n)\mu^2} \\ i \in U \end{cases} \qquad 3\bullet7$$

or

$$\begin{cases} t_i = \dfrac{\sigma^2 + N\mu^2}{N(\sigma^2 + n\mu^2)} \\ i \in U \end{cases}$$

or

$$\begin{cases} t_i = \dfrac{N + c^2}{N(n + c^2)} = t^* \quad \text{(say)} \\ i \in U. \end{cases} \qquad 3\bullet8$$

It follows from (1•2) and (3•8) that the optimal predictor of $\bar{Y}$ is

$$t_{opt} = t^* \sum_{i \in s} y_i$$

or

$$t_{opt} = \frac{N + c^2}{N(n + c^2)} \sum_{i \in s} y_i \qquad 3\bullet9$$

which is not p-unbiased.

The predictor (3•9) can be used in practice only when the coefficient of variation is known in advance. In many situations, the experimenter may have true figure of the coefficient of variation (particularly, in life sciences and biological experiments) from long association with the experimental material or empirical evidence gathered from repetead experiments or from extraneous source, for instance see Searls (1964), Murthy (1967, p. 96), Khan (1968), Gleser and Healy (1976) and Sen (1978).

Substituting (3•8) in (3•5) we get the *AMSE* of the 'optimal predictor' t_{opt} as

$$AMSE(t_{opt}) = \frac{(N-n)(N+c^2)}{N^2(n+c^2)}\sigma^2. \qquad 3\bullet10$$

It can easily be proved under model (1•3) that the average variance (*AVar*) of the sample mean $\bar{y} = (1/n)\sum_{i \in s} y_i$ (or $t = \sum_{i \in s} t_i y_i$ with $t_i = 1/n$) is

$$AVar(\bar{y}) = \frac{N-n}{nN}\sigma^2. \qquad 3\bullet11$$

It follows from (3•10) and (3•11) that the relative efficiency (*RE*) of 'optimal predictor' t_{opt} with respect to sample mean $\bar{y}$ is

$$RE(t_{opt}, \bar{y}) = \frac{N(n+c^2)}{n(N+c^2)} \qquad 3\bullet12$$

which is always greather than 'unity'. Thus the 'optimal predictor' t_{opt} is more efficient than the sample mean $\bar{y}$.

Remark 3.1. When the population size N is very large or population is infinite (i.e. $N \to \infty$), the 'optimal predictor' t_{opt} in (3•9) reduces to

$$t^*_{opt} = \frac{1}{n + c^2} \sum_{i \in s} y_i \qquad 3\bullet13$$

which is due to Searls (1964).

Remark 3.2. The 'optimal predictor' t_{opt} in (3•9) will be known when coefficient of variation 'c' is known in advance. In other cases, the 'optimal predictor' would be unimplementable.

References

[1] **Chaubey, Y.P., Dwivedi, T.D.** and **Singh, M.** (1984). An efficiency comparison of product and ratio estimator. *Communications in Statistics - Theory and Methods*, **13**, 699-709.

[2] **Gleser, L.J.** and **Healy, J.D.** (1976). Estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **71**, 977-981.

[3] **Hedayat, A.S.** and **Sinha, B.K.** (1991). *Design and Inference in Finite Population Sampling*. John Wiley. New York.

[4] **Khan, R.A.** (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **63**, 1039-1041.

[5] **Mukerjee, R.** and **Sengupta, S.** (1989). Optimal estimation of a finite population total under a general correlated model. *Biometrika*, **76**, 789-794.

[6] **Murthy, M.N.** (1967). *Sampling Theory and Methods*. Statistical Publishing Society. Calcutta.

[7] **Ruiz Espejo, M.** (1987). On UMV and UMMSE estimators in finite populations. *Estadística Española*, **29**, No. 115, 105-111.

[8] **Searls, D.T.** (1964). The utilization of a known coefficient of variation in estimation procedure. *Journal of the American Statistical Association*, **59**, 1225-1226.

[9] **Sen, A.R.** (1978). Estimation of the population mean when the coefficient of variation is known. *Communications in Statistics - Theory and Methods A*, **7**, No. 7, 657-672.

[10] **Singh, R., Singh, H.P.** and **Ruiz Espejo, M.** (1998). The efficiency of an alternative to ratio estimator under a super population model. *Journal of Statistical Planning and Inference*, **71**, 287-301.

Rev. Academia de Ciencias. Zaragoza. 56 (2001)

UNBIASED AND OPTIMAL LINEAR ESTIMATION FOR SOME SUPERPOPULATION MODELS

MARIANO RUIZ ESPEJO
UNED and UPSAM, Apartado 19207, 28080 Madrid, Spain
and
HOUSILA P. SINGH
School of Studies in Statistics, Vikram University, Ujjain 456010, M.P., India

Summary. In this paper, the characterization of the unbiased linear estimators under some superpopulation models has been treated for simple random sampling without replacement (SRSWOR) design of fixed size using auxiliary information. Some illustrations of such estimators have been given. An optimal linear estimator is suggested for some superpopulation models and for specific parameters.

Key words: Optimal estimation, simple random sampling without replacement design, superpopulation models, unbiasedness.

AMS Classification: 62D05.

1. Introduction

Consider a finite population $U = \{1, 2, ..., N\}$ of N identifiable units. Let y and x denote the study variable and the auxiliary variable taking values y_i and x_i ($i = 1, 2, ..., N$) respectively on the ith unit of U. Suppose $\mathbf{y} = (y_1, y_2, ..., y_N) \in \mathbb{R}^N$ is the parameter of the variable y under investigation and $\mathbf{x} = (x_1, x_2, ..., x_N) \in \mathbb{R}_+^N$ is the parameter of the auxiliary variable x. A simple random sample s of size n is drawn without replacement in order to estimate the parameter $\bar{y} = (1/N)\sum_{i\in U} y_i$ (in the design-unbiasedness) or $E(\bar{y})$ under the model (in the model-unbiasedness).

Let t be a linear estimator of the type

$$t = \sum_{i\in s} t_i y_i, \tag{1.1}$$

where t_i's are real fixed numbers ($i \in U$) and p is the simple random sampling without replacement design of fixed size n. Let S be the set of samples

$$S = \{s\colon\ s \subset U,\ \mathrm{card}\,(s) = n\}\,.$$

Then $p\,(s) = 1/\binom{N}{n}$ for all $s \in S$. Further, we denote

$$\bar{x}_s = \frac{1}{n}\sum_{i\in s} x_i,\ \bar{y}_s = \frac{1}{n}\sum_{i\in s} y_i,\ \bar{x} = \frac{1}{N}\sum_{i\in U} x_i \text{ and } \bar{y} = \frac{1}{N}\sum_{i\in U} y_i.$$

We assume that y_i is the realization of a random variable Y_i defined on $\mathbb{R}$.

In this paper we have discussed the characterization of the unbiased linear estimators of the type t under three superpopulation models. Our theoretical foundations are based on the books of Cassel *et al.* (1977), Hedayat *et al.* (1991) and Mukhopadhyay (1996).

For a general superpopulation model, the uniqueness of an optimal estimator is provided for specific parameters.

2. Design unbiasedness

This section deals with the problem of design unbiasedness of linear estimator t in (1.1). The results are given in the Theorem 2.1 and Corollary 2.1, where E_p stands design expectation.

THEOREM 2.1. *The linear estimator t is p-unbiased for the population mean $\bar{y}$, if and only if*

$$n\sum_{i\in U} t_i y_i = \sum_{i\in U} y_i \quad \text{for all} \quad \mathbf{y} \in \mathbb{R}^N. \ \square \tag{2.1}$$

PROOF. Taking design expectation of both sides in (1.1), we get

$$E_p\,(t) = E_p\left(\sum_{i\in s} t_i y_i\right) = \sum_{s\in S}\left(\sum_{i\in s} t_i y_i\right) p\,(s)$$

$$= \sum_{i\in U} t_i y_i \mathrm{card}\,\{s \in S\colon\ i \in s\}\,\frac{1}{\binom{N}{n}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}}\sum_{i\in U} t_i y_i = \frac{n}{N}\sum_{i\in U} t_i y_i. \tag{2.2}$$

The estimator t would be design unbiased if and only if

$$E_p(t) = \bar{y} \text{ or iff } \frac{n}{N}\sum_{i\in U} t_i y_i = \frac{1}{N}\sum_{i\in U} y_i \text{ or iff } n\sum_{i\in U} t_i y_i = \sum_{i\in U} y_i,$$

which proves the theorem. $\square$

COROLLARY 2.1. *As equation* (2.1) *is the necessary and sufficient condition for t to be p-unbiased,* (2.1) *equivales to the conditions* (2.1.a) *or* (2.1.b):

(2.1.a) $t_i = 1/n$ *for all* $i \in U$.

(2.1.b) $t = \bar{y}_s$. $\square$

In other words, if we consider t is an estimator of the population mean $\bar{y}$, then t is said to be p-unbiased if and only if t is equal to the sample mean $\bar{y}_s$.

3. Superpopulation model M_1

In this section, we have characterized the unbiasedness of the linear estimator t under the superpopulation model M_1:

$$Y_i = \beta X_i + e_i \text{ with } \beta \text{ unknow fixed}, \quad E(e_i|X_i) = 0 \text{ and } X_i \neq 0. \tag{3.1}$$

The results are given in the form of theorems and corollaries.

THEOREM 3.1. *The linear estimator t is M_1-unbiased for $E_{M_1}(\bar{y})$ if and only if*

$$\sum_{i \in s} t_i x_i = \bar{x} \text{ for all } s \in S. \ \square \tag{3.2}$$

PROOF. Taking the expectation of t in (1.1) under linear model $M = M_1$, we get

$$E_M(t) = E_M\left(\sum_{i \in s} t_i y_i\right) = \sum_{i \in s} t_i E_M(y_i) = \beta \sum_{i \in s} t_i x_i, \text{ for all } s \in S, \tag{3.3}$$

and

$$E_M(\bar{y}) = E_M\left(\frac{1}{N}\sum_{i \in U} y_i\right) = \frac{1}{N}\sum_{i \in U} E_M(y_i) = \beta \frac{1}{N}\sum_{i \in U} x_i = \beta \bar{x}. \tag{3.4}$$

It follows from (3.3) and (3.4) that $E_M(t) = E_M(\bar{y})$ if and only if

$$\sum_{i \in s} t_i x_i = \bar{x}.$$

Hence the theorem is proved. $\square$

COROLLARY 3.1. *When all the x_i $(i \in U)$ are known, the equation (3.2) is given for a sample of size $n = \text{card}(s) = 1$ for all $s \in S$, if and only if*

$$t_i x_i = \bar{x} \text{ for all } i \in U, \text{ or } t_i = \bar{x}/x_i \text{ for all } i \in U,$$

or also the unique (M_1-unbiased for $E_{M_1}(\bar{y})$) linear estimator is the ratio estimator (with $s = \{i\}$) $t = (\bar{x}/x_i)\, y_i$. $\square$

This result is useful and clear, but limited to $n = 1$.

REMARK 3.1. When equation (3.2) is given for a sample of size $n \geq 2$, there exists an infinite number of solutions for $(t_{i_1}, t_{i_2}, ..., t_{i_n})$; some examples are for $n = 2$:

$$t = \frac{\bar{x}}{2}\left(\frac{y_{i_1}}{x_{i_1}} + \frac{y_{i_2}}{x_{i_2}}\right) \text{ and } t = \frac{\bar{x}}{3}\left(\frac{2y_{i_1}}{x_{i_1}} + \frac{y_{i_2}}{x_{i_2}}\right).$$

THEOREM 3.2. *The linear estimator t is design as well as model unbiased (pM_1-unbiased) if and only if*

$$n\sum_{i\in U} t_i x_i = N\bar{x}. \ \square \tag{3.5}$$

PROOF. Taking expectation of t in (1.1), we have (as in Theorem 2.1)

$$E(t) = \beta\frac{n}{N}\sum_{i\in U} t_i x_i, \tag{3.6}$$

and

$$E_M(\bar{y}) = \beta\frac{1}{N}\sum_{i\in U} x_i. \tag{3.7}$$

The estimator t would be unbiased if and only if (3.6) and (3.7) are same

$$E(t) = E_M(\bar{y}) \ \text{ or iff } \ \beta\frac{n}{N}\sum_{i\in U} t_i x_i = \beta\frac{1}{N}\sum_{i\in U} x_i$$

or iff

$$n\sum_{i\in U} t_i x_i = \sum_{i\in U} x_i \ \text{ or iff } \ n\sum_{i\in U} t_i x_i = N\bar{x}.$$

This proves the theorem. □

COROLLARY 3.2. *A sufficient condition to verify equation* (3.5) *is given by*

$$nt_i x_i = x_i \ \textit{ for all } \ i \in U,$$

or equivalently

$$t_i = \frac{1}{n} \ \textit{ for all } \ i \in U, \ \textit{ or } \ t = \bar{y}_s \ \textit{(the sample mean).} \ \square$$

COROLLARY 3.3. *Another sufficient condition to verify the equation* (3.5) *is given by*

$$nt_i x_i = \bar{x} \ \textit{ for all } \ i \in U,$$

or equivalently

$$t_i = \frac{\bar{x}}{nx_i} \ \textit{ for all } \ i \in U, \ \textit{ or } \ t = \frac{\bar{x}}{n}\sum_{i\in s}\frac{y_i}{x_i},$$

the sample mean-of-the-ratios per $\bar{x}$. □

REMARK 3.2. Another pM_1-unbiased estimator is the ratio estimator (for $n \geq 2$) given by

$$t = \bar{y}_s\frac{\bar{x}}{\bar{x}_s} = \frac{\bar{x}}{n\bar{x}_s}\sum_{i\in s} y_i,$$

but this estimator belongs to some more general classes of linear estimators given by

$$t = t_s\sum_{i\in s} y_i$$

(the class e_3 from Hedayat *et al.* (1991), p. 23, Table 2.1) or

$$t = \sum_{i \in s} t_{si} y_i$$

(the class most general of a homogeneous linear estimator considered in Hedayat *et al.* (1991), p. 23, Equation (2.1)), for which the coefficient t_s depends on s, and t_{si} depends on s and i, not only on i as we have considered in this article.

4. Superpopulation model M_2

Now we consider the unbiasedness of the linear estimator t under the following superpopulation model $M = M_2$:

$$Y_i = \alpha + \beta X_i + e_i \text{ with } E(e_i | X_i) = 0 \text{ and}$$

$$\alpha \neq 0,\ \beta > 0,\ X_i > 0 \text{ for all } i \in U \text{ and } X_i \neq X_j \text{ for } i \neq j \in U. \tag{4.1}$$

The results are given in Theorems 4.1 and 4.2 and Corollaries 4.1 and 4.2.

THEOREM 4.1. *The linear estimator t is M_2-unbiased for $E_{M_2}(\bar{y})$ if and only if*

$$\sum_{i \in s} t_i (\alpha + \beta x_i) = \alpha + \beta \bar{x} \text{ for all } s \in S. \ \square \tag{4.2}$$

PROOF. Taking expectation of t in (1.1) under the superpopulation model $M = M_2$, we have

$$E_M(t) = E_M\left(\sum_{i \in s} t_i y_i\right) = \sum_{i \in s} t_i E_M(y_i) = \sum_{i \in s} t_i (\alpha + \beta x_i), \text{ for all } s \in S \tag{4.3}$$

and

$$E_M(\bar{y}) = E_M\left(\frac{1}{N}\sum_{i \in U} y_i\right) = \frac{1}{N}\sum_{i \in U} E_M(y_i) = \frac{1}{N}\sum_{i \in U}(\alpha + \beta x_i) = \alpha + \beta \bar{x}. \tag{4.4}$$

It follows from (4.3) and (4.4) that $E_M(t) = E_M(\bar{y})$ if and only if

$$\sum_{i \in s} t_i (\alpha + \beta x_i) = \alpha + \beta \bar{x}.$$

Thus the theorem is proved. $\square$

COROLLARY 4.1. *The equation* (4.2) *can be characterized by the system*

$$\begin{cases} \alpha \sum_{i \in s} t_i = \alpha, \text{ and} \\ \beta \sum_{i \in s} t_i x_i = \beta \bar{x} \end{cases} \quad \text{or} \quad \begin{cases} \sum_{i \in s} t_i = 1, \text{ and} \\ \sum_{i \in s} t_i x_i = \bar{x}, \end{cases}$$

which gives the unique solution for $n = \text{card}(s) = 2$ *and* $s = \{i_1, i_2\}$:

$$t_{i_1} = \frac{\bar{x} - x_{i_2}}{x_{i_1} - x_{i_2}} \quad \textit{and} \quad t_{i_2} = \frac{\bar{x} - x_{i_1}}{x_{i_2} - x_{i_1}},$$

or the unique M_2*-unbiased linear estimator is*

$$t = \frac{\bar{x} - x_{i_2}}{x_{i_1} - x_{i_2}} y_{i_1} + \frac{\bar{x} - x_{i_1}}{x_{i_2} - x_{i_1}} y_{i_2}. \ \square$$

For $n \geq 3$, there exist infinite M_2-unbiased linear estimators.

THEOREM 4.2. *The linear estimator* t *is* pM_2*-unbiased if and only if*

$$n \sum_{i \in U} t_i (\alpha + \beta x_i) = N (\alpha + \beta \bar{x}). \ \square \tag{4.5}$$

PROOF. For superpopulation model $M = M_2$, we have

$$E(t) = \frac{n}{N} \sum_{i \in U} t_i (\alpha + \beta x_i), \tag{4.6}$$

and

$$E_M(\bar{y}) = \alpha + \beta \bar{x}. \tag{4.7}$$

From the equations (4.6) and (4.7) we have the necessary and sufficient condition (4.5). $\square$

COROLLARY 4.2. *A sufficient condition for the equation* (4.5) *—when* $\gamma = \alpha/\beta$ *is known— is given by*

$$n t_i (\alpha + \beta x_i) = \alpha + \beta \bar{x} \quad \textit{for all} \quad i \in U,$$

or

$$t_i = \frac{\alpha + \beta \bar{x}}{n (\alpha + \beta x_i)} \quad \textit{for all} \quad i \in U,$$

or a pM_2*-unbiased estimator will be*

$$t = \frac{1}{n} \sum_{i \in s} \frac{\alpha + \beta \bar{x}}{\alpha + \beta x_i} y_i = \frac{1}{n} \sum_{i \in s} \frac{\gamma + \bar{x}}{\gamma + x_i} y_i, \tag{4.8}$$

where $\gamma = \alpha/\beta$. *The estimator* (4.8) *is implementable when* γ *is known in advance.* $\square$

COROLLARY 4.3. *Another sufficient condition for the equation* (4.5) *is given by*

$$n t_i (\alpha + \beta x_i) = \alpha + \beta x_i \quad \textit{for all} \quad i \in U, \quad \textit{or} \quad t_i = \frac{1}{n} \quad \textit{for all} \quad i \in U,$$

from which we get the pM_2-unbiased sample mean estimator

$$t = \frac{1}{n}\sum_{i\in s} y_i = \bar{y}_s$$

which is always implementable. □

REMARK 4.1. Note that the regression estimator, for $n \geq 2$,

$$t_{reg} = \bar{y}_s + \frac{\sum_{i\in s}(y_i - \bar{y}_s)(x_i - \bar{x}_s)}{\sum_{i\in s}(x_i - \bar{x}_s)^2}(\bar{x} - \bar{x}_s)$$

is not p-unbiased, but it is M_2-unbiased and pM_2-unbiased, since for $M = M_2$,

$$E_M(t_{reg}) = \alpha + \beta\bar{x}_s + \frac{\sum_{i\in s}\beta(x_i - \bar{x}_s)^2}{\sum_{i\in s}(x_i - \bar{x}_s)^2}(\bar{x} - \bar{x}_s) =$$

$$= \alpha + \beta\bar{x}_s + \beta(\bar{x} - \bar{x}_s) = \alpha + \beta\bar{x} = E_M(\bar{y}).$$

REMARK 4.2. It can easily be shown for model $M = M_1$ that the regression estimator t_{reg} is M_1-unbiased and pM_1-unbiased.

REMARK 4.3. However t_{reg} (for $n \geq 2$) is a linear estimator of the type (Hedayat *et al.* (1991), p. 23, Equation (2.1))

$$t = \sum_{i\in s} t_{si} y_i,$$

which is more general than the type of linear estimator

$$t = \sum_{i\in s} t_i y_i$$

defined in (1.1).

The justification is given below:

$$t_{reg} = \bar{y}_s + \frac{\sum_{i\in s} y_i(x_i - \bar{x}_s)}{\sum_{i\in s}(x_i - \bar{x}_s)^2}(\bar{x} - \bar{x}_s) = \sum_{i\in s}\frac{1}{n}y_i + \sum_{i\in s}\frac{(x_i - \bar{x}_s)(\bar{x} - \bar{x}_s)}{\sum_{j\in s}(x_j - \bar{x}_s)^2}y_i =$$

$$= \sum_{i \in s} \left\{ \frac{1}{n} + \frac{(x_i - \bar{x}_s)(\bar{x} - \bar{x}_s)}{\sum_{j \in s} (x_j - \bar{x}_s)^2} \right\} y_i,$$

where

$$t_{si} = \frac{1}{n} + \frac{(x_i - \bar{x}_s)(\bar{x} - \bar{x}_s)}{\sum_{j \in s} (x_j - \bar{x}_s)^2}$$

depends on i and s, not only on i.

5. Superpopulation model M_3

Consider the superpopulation model $M = M_3$ as

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \cdots + \alpha_{n-1} X_i^{n-1} + e_i$$

where $\alpha_0, \alpha_1, ..., \alpha_{n-1}$ are constants prefixed, $X_i \neq X_j$ if $i \neq j$ (all the x_i from X_i, $i \in U$, are known), where $E(e_i | X_i) = 0$, and the sample s has size m.

THEOREM 5.1. *The linear estimator $t = \sum_{i \in s} t_i y_i$ is M_3-unbiased for $E_M(\bar{y})$ (with $M = M_3$) and unique when $m = n$.* □

PROOF. The model $M = M_3$ is

$$y_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_i^j + e_i \text{ with } E(e_i | x_i) = 0.$$

The model expectation of t is

$$E_M(t) = E_M\left(\sum_{i \in s} t_i y_i\right) = \sum_{i \in s} t_i E_M(y_i) = \sum_{i \in s} t_i \left(\alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_i^j\right). \tag{5.1}$$

The expectation of $\bar{y}$ under model $M = M_3$ is

$$E_M(\bar{y}) = E_M\left(\frac{1}{N} \sum_{i \in U} y_i\right) = \frac{1}{N} \sum_{i \in U} E_M(y_i)$$

$$= \frac{1}{N} \sum_{i \in U} \left(\alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_i^j\right) = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j \left(\frac{1}{N} \sum_{i \in U} x_i^j\right). \tag{5.2}$$

It follows from (5.1) and (5.2) that $E_M(t) = E_M(\bar{y})$ if the system of n equations:

$$\sum_{i\in s} t_i = 1, \quad \sum_{i\in s} t_i x_i = \frac{1}{N}\sum_{i\in U} x_i, \quad \cdots \quad \sum_{i\in s} t_i x_i^{n-1} = \frac{1}{N}\sum_{i\in U} x_i^{n-1}, \tag{5.3}$$

are verified, where the constants t_i (for $i = i_1, i_2, ..., i_n$) are the n unknown quantities. The uniqueness of solution of the system is due to Rouché-Fröbenius theorem.

Using Cramer's rule, we get

$$t_{i_1} = D_1/D, ..., t_{i_n} = D_n/D$$

where

$$D_1 = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{N}\sum_{i=1}^{N} x_i & x_{i_2} & \cdots & x_{i_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{i=1}^{N} x_i^{n-1} & x_{i_2}^{n-1} & \cdots & x_{i_n}^{n-1} \end{vmatrix}, \ldots, D_n = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_{i_1} & x_{i_2} & \cdots & \frac{1}{N}\sum_{i=1}^{N} x_i \\ \vdots & \vdots & \ddots & \vdots \\ x_{i_1}^{n-1} & x_{i_2}^{n-1} & \cdots & \frac{1}{N}\sum_{i=1}^{N} x_i^{n-1} \end{vmatrix},$$

and D is the van der Monde determinant

$$D = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_{i_1} & x_{i_2} & \cdots & x_{i_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i_1}^{n-1} & x_{i_2}^{n-1} & \cdots & x_{i_n}^{n-1} \end{vmatrix}$$

which is non null since $x_i \neq x_j$ if $i \neq j$. □

REMARK 5.1. The polynomial model, with the degree same as the sample size minus one, is of limited interest. For this case, the system (5.3) gives a unique solution for the vector

$$(t_{i_1}, t_{i_2}, ..., t_{i_n}).$$

If the sample size is $m < n$, there does not exist a solution, and if $m > n$ there exists an infinite number of solutions for the vector $(t_{i_1}, t_{i_2}, ..., t_{i_n})$.

REMARK 5.2. If there exist $x_i = x_j$ with $i \neq j \in s$, then $D = 0$, and hence there does not exist solution for the vector $(t_{i_1}, t_{i_2}, ..., t_{i_n})$.

THEOREM 5.2. *The linear estimator $t = \sum_{i\in s} t_i y_i$ is pM_3-unbiased if and only if*

$$\frac{n}{N}\sum_{i\in U} t_i \left(\alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_i^j\right) = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j \left(\frac{1}{N}\sum_{i\in U} x_i^j\right). \ \square \tag{5.4}$$

The proof is simple, so it is omitted.

COROLLARY 5.1. *The condition* (5.4) *can be expressed as the system* (*when* $\alpha_j \neq 0$ *for all* $j = 0, 1, 2, ..., n-1$):

$$\begin{cases} n\sum_{i\in U} t_i & = & N \\ n\sum_{i\in U} t_i x_i & = & \sum_{i\in U} x_i \\ & \vdots & \\ n\sum_{i\in U} t_i x_i^{n-1} & = & \sum_{i\in U} x_i^{n-1} \end{cases} . \square$$

6. Optimal estimation

We consider the superpopulation model M as

$$Y_i = \varphi(X_i) + e_i \quad (i \in U),$$

with φ a function known and $E(e_i|X_i) = 0$, and $Cov(e_i, e_j|X_i, X_j)$ is a constant known for i and j fixed. We assume the class of estimators given in (1.1) for estimating the parameter

$$f(\mathbf{y}) = \sum_{i\in U} p_i y_i \text{ when the } p_i \ (i \in U) \text{ are prefixed and known.}$$

The criterion of optimality is minimum 'expected mean squared error' with p the simple random sampling without replacement design, min $E_M\{MSE_p(t)\}$. As

$$MSE_p(t) = V_p(t) + B_p^2(t) =$$

$$= \frac{n(N-n)}{N^2}\sum_{i\in U} t_i^2 y_i^2 - \frac{n(N-n)}{N^2(N-1)}\sum_{i\in U}\sum_{j\in U-\{i\}} t_i t_j y_i y_j + \left\{\frac{1}{N}\sum_{i\in U}(nt_i - Np_i)\, y_i\right\}^2$$

$$= \sum_{i\in U} A_i y_i^2 + \sum_{i\in U}\sum_{j\in U-\{i\}} A_{ij} y_i y_j,$$

where for $i \in U$,

$$A_i = \frac{n(N-n)}{N^2} t_i^2 + \frac{(nt_i - Np_i)^2}{N^2}$$

and for all i and j of U with $i \neq j$,

$$A_{ij} = \frac{-n(N-n)}{N^2(N-1)} t_i t_j + \frac{(nt_i - Np_i)(nt_j - Np_j)}{N^2}.$$

Since for $i \in U$,

$$E_M\left(y_i^2|x_i\right) = \varphi^2(x_i) + V_M(e_i|x_i) = B_i \text{ (say)}$$

and for $i \neq j \in U$,

$$E_M\left(y_i y_j | x_i, x_j\right) = \varphi\left(x_i\right)\varphi\left(x_j\right) + E_M\left(e_i e_j | x_i, x_j\right) = B_{ij} \ \text{(say)},$$

we get

$$F = E_M\{MSE_p(t)\} = \sum_{i\in U} A_i B_i + \sum_{i\in U}\sum_{j\in U-\{i\}} A_{ij} B_{ij}.$$

To minimize F we have the necessary conditions (when the constants B_i and B_{ij} are known)

$$\frac{\partial F}{\partial t_i} = \frac{\partial A_i}{\partial t_i} B_i + 2 \sum_{j\in U-\{i\}} \frac{\partial A_{ij}}{\partial t_i} B_{ij} = 0 \ \text{ for all } \ i \in U,$$

or the system of N equations with N unknown quantities $(t_i,\ i \in U)$:

$$K_i t_i + \sum_{j\in U-\{i\}} K_{ij} t_j = C_i \ \ (i \in U), \tag{6.1}$$

where

$$K_i = \frac{2nB_i}{N}, \quad K_{ij} = 2\left\{\frac{-n(N-n)}{N^2(N-1)} + \frac{n^2}{N^2}\right\} B_{ij}$$

and

$$C_i = K_i p_i + \frac{2n}{N} \sum_{j\in U-\{i\}} B_{ij} p_j$$

are constants known. The system (6.1) have a unique optimal solution for the linear estimator $t = \sum_{i\in s} t_i y_i$, of the parameter $f(\mathbf{y}) = \sum_{i\in U} p_i y_i$, when the determinant

$$\begin{vmatrix} K_1 & K_{12} & \cdots & K_{1n} & \cdots & K_{1N} \\ K_{21} & K_2 & \cdots & K_{2n} & \cdots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ K_{n1} & K_{n2} & \cdots & K_n & \cdots & K_{nN} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \cdots & K_{Nn} & \cdots & K_N \end{vmatrix} \neq 0, \tag{6.2}$$

due to Rouché-Fröbenius theorem. For the optimal linear estimator, the concrete values t_i $(i \in s)$ are provided via Cramer's rule.

References

[1] **Cassel, C.M., Särndal, C.E.** and **Wretman, J.H.** (1977). *Foundations of Inference in Survey Sampling.* John Wiley. New York.

[2] **Hedayat, A.S.** and **Sinha, B.K.** (1991). *Design and Inference in Finite Population Sampling.* John Wiley. New York.

[3] **Mukhopadhyay, P.** (1996). *Inferential Problems in Survey Sampling.* New Age International. New Delhi.

[4] **Royall, R.M.** (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

Nota necrológica.

El pasado 17 de Mayo de 2001 falleció en Zaragoza el Profesor Dr. D. José María Savirón de Cidón, Académico Numerario por la Sección de Físicas, de ésta Academia de Ciencias Exactas, Físicas, Químicas y Naturales y Catedrático de Mecánica de Fluidos de la Universidad de Zaragoza.

Nos ha dejado demasiado pronto quien, para nosotros, fue un gran amigo y compañero y un profesor y maestro para decenas de generaciones de estudiantes.

Nació Pepe, como se hacía llamar y lo era para todos los que con él tratábamos, en Madrid debido a los avatares de la historia de la España de aquella época, 1937, y las circunstancias familiares, pero en el seno de una familia de raíz aragonesa y universitaria. Su abuelo Paulino Savirón había llegado a ser Rector de la Universidad de Zaragoza. Su niñez y adolescencia transcurrió en tierras asturianas y su formación preuniversitaria y universitaria la realizó en Zaragoza. Él mismo se consideraba de las tres regiones y frecuentemente le oíamos decir, hoy vengo de asturiano o madrileño, nunca de aragonés porque eso lo era siempre.

Cursó la Licenciatura de Física Matemática e hizo su tesis doctoral bajo la dirección del Profesor Justiniano Casas Peláez, tesis que constituyó uno de los primeros trabajos españoles sobre teoría de separación de isótopos por termodifusión, iniciándo así una prolífica escuela de científicos en el tema. El Laboratorio de Zaragoza, que el dirigió durante años en la Facultad de Ciencias, llegó a ser referencia internacional en el tema y de él ha salido un nutrido plantel de catedráticos y profesores de universidad y más de 25 tesis doctorales.

Tras una breve estancia como Catedrático en la Universidad de La Laguna pasó a ser Catedrático de la de Zaragoza, donde ha permanecido durante más de 35 años.

Científicamente, Savirón, era un hombre muy original, con gran sentido físico y didáctico de los temas que trataba, a los que dotaba de una rigurosidad física y matemática impecable, aunque siempre con un

enfoque nuevo, sencillo y elegante. Realizó y dirigió muchos trabajos de investigación de muy diversa índole, como los de su primera etapa en separación isotópica, tanto por espectrometría de masas como por termodifusión, los posteriores para la Confederación Hidrográfica del Ebro sobre la fluidodinámica del propio río, los realizados para empresas sobre propiedades termodinámica de hidrocarburos y gases licuados del petróleo y últimamente sobre la termohidráulica del Plomo fundido. A esa labor hay que añadir la de divulgación de alto nivel, casi podíamos llamar docente, en la que cabe mencionar sus libros de problemas de física, su colaboración en textos de Física y Química de la enseñanza secundaria y sobre la física y el deporte.

Humanamente, Pepe era generoso y entrañable, siempre dispuesto a ayudar a quien lo necesitaba sin necesidad de pedírselo y a "desfacer entuertos" de sus amigos, aún a costa de sus relaciones personales. Quizá la personalidad singular de Pepe quede muy bien reflejada en los párrafos que le dedica D. Justiniano Casas en su discurso de contestación al de su ingreso en esta Academia el 25 de Mayo de 1992.

"*Si tuviera que caracterizar al Dr. Savirón como persona diría simplemente que era un joven en extremo inteligente. Físicamente, en consonancia con sus exaltadas actividades deportivas, elástico como una ballesta, y, en paralelo, con una agilidad mental asombrosa, a la vez que absolutamente desordenado en el tiempo y en el espacio. Creo que nunca fue posible contar con él en día, hora y lugar determinado. Nunca supe cuándo comía, trabajaba y descansaba; dónde estaba cuando desaparecía por días del laboratorio sin dejar rastro. Lo único que se podía asegurar es que era un hombre comprometido con su tarea y que a la hora en punto surgía como una aparición con todo el trabajo esmeradamente terminado y derramando torrentes de nuevas ideas.*

¿Y qué es ahora de nuestro recipiendario? Pues lo mismo que antes con todas sus características congénitas más o menos, más bien menos, atenuadas por el paso del tiempo, pero imperecederas. Y además de eso, cargado con un excepcional curriculum científico en forma de decenas de relevantes trabajos científicos y libros publicados.".

La personalidad del Profesor Savirón dejó huellas en muchos ambientes y sectores tanto locales como regionales y nacionales adquiriendo compromisos donde solicitaron su apoyo o ayuda.

Fué director del Departamento de Física Fundamental, el primer coordinador del Curso de Orientación Universitaria, organizó y fue el primer director del curso selectivo de la Academia General Militar de

Zaragoza, que permitía a los alumnos realizar a la vez el primer año de Físicas recibiendo en 1977 la Gran Cruz de la Orden del Mérito Militar con distintivo blanco de primera clase. En 1980 fue elegido Decano de la facultad de Ciencias, periodo muy difícil en el mundo universitario que él supo sacar adelante sin graves problemas. Fue presidente de la ponencia de Física de la Comisión Asesora de Investigación Científica y Técnica en Madrid (CAICYT), lo que le valió un gran conocimiento de la situación de la física y los físicos en España y que sirvió para que casi todos ellos, en aquellos años, conocieran de la amistad y el buen hacer de Pepe. Posteriormente, en 1989, pasó a ser Presidente del Consejo Asesor de Investigación del Gobierno de Aragón, y por ello miembro del Consejo Rector del Instituto Tecnológico de Aragón (CONAI). Dentro de sus actividades dedicadas a la "res publica" también fue Coordinador de la ponencia de Innovación Tecnológica de la Comunidad de Trabajo de los Pirineos. Recibió en 1991 la Medalla de la Real Sociedad Española de Física de la que fue presidente en el periodo 1993-95 y motor y organizador de Conferencias y Congresos como las Jornadas de Física de Zaragoza organizadas en IberCaja y la XXIII Reunión Bienal de la Real Sociedad Española de Física de Jaca. En muchos de esos proyectos tuvimos el placer y el honor de colaborar muy estrechamente con él.

Descansa en paz Pepe, para muchos seguirás vivo en nuestra memoria, quizá hasta que te acompañemos. Para todos seguirás presente mucho tiempo a través de tus obras y tu buen hacer.

Rafael Nuñez-Lagos Roglá
y
Luis Joaquín Boya Balet