

*VII Coloquio del Departamento
de Matemáticas*

29 de julio al 16 agosto de 1991
Realizado en : ESFM-IPN



Conferencias Generales

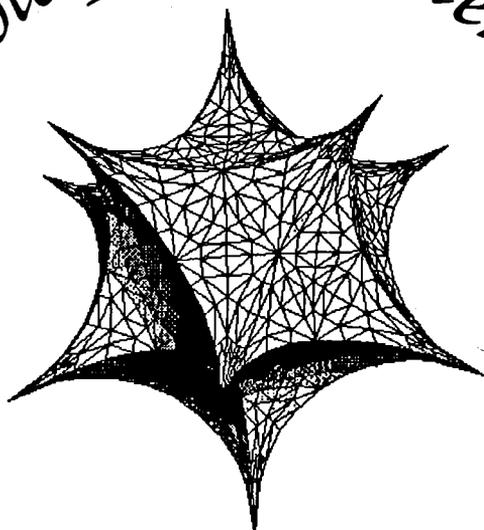


Editor : Luis Gorostiza

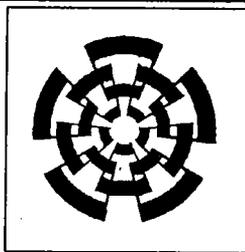


Centro de Investigación y de Estudios Avanzados del I.P.N.

VII Coloquio de Matemáticas



Centro de Investigación y de Estudios Avanzados del I.P.N.
Departamento de Matemáticas.



Prefacio

En este volumen se publican varias de las Conferencias Generales dictadas en el VII Coloquio del Departamento de Matemáticas del Centro de Investigación y de Estudios Avanzados de I. P. N.

Como en los Coloquios anteriores, se procuró que estas conferencias cumplieran el objetivo de divulgar, de manera amena y accesible a los no expertos, algunas ideas básicas de las matemáticas y sus aplicaciones en otras disciplinas.

Agradecemos a los conferencistas por sus interesantes ponencias, y especialmente a los que presentaron los escritos que integran este volumen.

Luis G. Gorostiza

Índice

P. Armendáriz	
Música y Matemáticas	1
J. P. Hennart	
Algunos aspectos de la discretización de problemas de valores iniciales	21
M. C. Jorge	
La forma de la Tierra: historia de su búsqueda y las matemáticas que intervinieron en ella	45
E. A. Lacomba	
Movimiento de los planetas: el sistema solar	103
O. A. Sánchez-Valenzuela	
Teoría de grupos y partículas elementales	117
M. Socolovsky	
Geometría y física	129
F. Valenzuela y P. Gorostiza	
Las matemáticas en la fisiología moderna	155

Música y Matemáticas

Pedro Armendáriz Morales

Universidad Autónoma Metropolitana-Iztapalapa

He querido hablar aquí, un poco acerca de música y muy poco acerca de matemáticas, para aclarar ciertas matemáticas elementales que aparecen en cuestiones musicales y que muchas veces son presentadas como objetos mágicos o en forma por demás oscura.

Voy a suponer que entre la audiencia se encuentran sólo personas interesadas en la música (aficionados, amantes de la música, o como se les quiera llamar) y probablemente muchos músicos frustrados -como es común entre los científicos- o músicos venidos a menos, y por ende, están familiarizados con los conceptos elementales de la teoría musical especialmente en su notación (que ahora se enseña en la escuela secundaria).

Cuando de pequeños empezamos a estudiar música, se nos dice que entre una nota ¹ y su octava -es decir entre un sonido con una frecuencia muy particular y el sonido cuya frecuencia es el doble de la primera- hay 7 notas naturales y otras que durante mucho tiempo fueron consideradas como auxiliares, en total 12 notas. Alguien con inquietudes raras hacia los números se puede preguntar ¿Por qué un total de 12 notas? ¿Hay alguna razón poderosa para que sean 12 y no 13 ó 17 ó 69? ¿Por qué entre las 12 hay 7 tipos y 5 de otro? ¿Por qué no 6 y 6 ó todas iguales? Alguien, una vez, de manera simplista afirmó que 12, porque este es número pequeño y es divisible por 2, 3, 4 y 6 y que la división en 7 y 5 es sólo cuestión de tradición así como el nombre de las notas² o la notación que es usada para representarlas. Veamos que tanto el 7 como el 12 aparecen de una forma natural.

Consideremos ahora un cierto sonido que llamaremos nota y denotaremos por la letra A el cual tiene cierta frecuencia ν y que será nuestro punto ó nota de partida -en nuestros cálculos tomaremos como $\nu = 440$ c/seg. que es

¹Acerca de los nombres de las notas consúltese por ejemplo el artículo de S. Zinker, donde hay muy poca base racional.

²item 1

un estándar internacional³. El sonido que más se asemeja al de A es el que tiene el doble de su frecuencia, 2ν que denotaremos por A' y también todos los demás sonidos con frecuencia de la forma $2^n\nu$ con n un entero y a todas estas notas las identificaremos con el mismo nombre, en realidad estamos dividiendo toda nuestra gama de frecuencias, los números reales entre cero e infinito, en clases, donde identificamos dos de ellas, dándoles el mismo nombre (están en la misma clase) si su cociente es una potencia entera de 2.

Trabajaremos pues, sólo con el intervalo de frecuencias entre ν y 2ν y es ahí donde construiremos nuestra escala musical y si una frecuencia traspone estos límites, la multiplicamos ó dividimos por 2 un número suficiente de veces hasta encuadrarla en nuestro intervalo. A este intervalo, por tradición le llamaremos el intervalo de una octava.

El siguiente sonido en consonancia con el de frecuencia ν , esto es, que al escucharlos juntos suene agradable al oído, es el de frecuencia 3ν , que para transportarlo a nuestro intervalo entre ν y 2ν , dividimos entre 2 y tenemos el sonido de frecuencia $\frac{3}{2}\nu$, el cual será nuestra segunda nota en la escala que queremos construir. Pero entonces también debemos incluir como nota el sonido con frecuencia $\frac{3^2}{2^2}\nu$ que corresponde al triple de frecuencia de $\frac{3}{2}\nu$ y continuar de esta manera. Este proceso terminaría (y empezaría a repetirse) si en algún momento una potencia entera de $3/2$ fuera igual a una potencia entera 2 o lo que es lo mismo, si alguna potencia (positiva) de 3 fuera igual a una potencia entera (positiva) de 2, lo cual es imposible por ser 3 impar.

Continuar indefinidamente con el proceso de asignar notas a la escala triplicando el período cada vez es absurdo y totalmente impráctico. Mejor tratemos de obtener aproximaciones.

Estamos en busca del número x que satisface que $2^x = 3/2$, el cual como vimos no podemos escribir como cociente de dos enteros, pero que queremos aproximar por medio del cociente de dos enteros.

Es conocido en matemáticas que una manera buena para aproximar un número x , es expresarlo⁴ como una fracción continua y tomar las sumas

³Véase Shilov [Sh] pág. 1.

⁴Véase por ejemplo el libro de Chrystal [Ch].

parciales como aproximaciones, en otras palabras escribimos x como

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}}$$

con $a_1, a_2 \dots$ enteros positivos.

Y las aproximaciones son

$$\frac{1}{a_1}, \frac{1}{a_1 + \frac{1}{a_2}}, \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3}}}, \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4}}}}, \text{ etc.}$$

Si hacemos esto para nuestro número x se obtiene⁵,

$$x = \frac{1}{1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{3 + \frac{1}{5 + \frac{1}{2 + \dots}}}}}}}}$$

Esto lo puede hacer "a mano", pero les recomiendo mejor usar una calculadora aunque sea de bolsillo.

⁵Para ver cómo se hace, véase por ejemplo el libro de Shilov [Sh] o el de Christal [Ch].

Las aproximaciones que se obtienen son:

$$\begin{array}{llll}
 i) \frac{1}{1} & ii) \frac{1}{2} & iii) \frac{3}{5} & iv) \frac{7}{12} \\
 v) \frac{24}{41} & vi) \frac{31}{53} & vii) \frac{179}{306} & viii) \frac{389}{665} \\
 ix) \frac{9,126}{15,601} & x) \frac{46,408}{79,335} & \dots &
 \end{array}$$

Hagamos algunas observaciones acerca de estos cocientes.

1. Podemos considerar que las dos primeras aproximaciones son malas, pues nuestro número x es igual a $\log_2(3/2) = \ln(3/2)/\ln(2)$ que es aproximadamente 0.5849625007 ...
2. En la tercera aproximación, tomamos una división del intervalo en 5 notas (una escala de 5 notas) de las cuales 3 son naturales y 2 auxiliares y es llamada escala Pentatónica. Fue utilizada en la antigüedad y aparece en composiciones chinas o escocesas⁶.

Las notas que corresponden en nuestra escala de 12 (a partir de triplicar frecuencias) si partimos de A (440) son: A E B F# C#, siendo las tres primeras las naturales y las dos últimas las auxiliares.

3. La aproximación iv) $7/12$ que es la escala que comunmente usamos era ya conocida por los chinos 1000 años antes de nuestra era⁷ y también por los griegos pitagóricos.

Al cociente $(3/2)^{12}/2^7 = 3^{12}/2^{19} = \frac{531,441}{524,288}$ se le conoce como la Coma de Pitágoras.

4. La aproximación siguiente v) $24/41$ fue encontrada por un estudioso chino llamado King Fang, cuarenta años antes de nuestra era, calculando su coma $3^{41}/2^{65}$, pero al no encontrar una mejora sustancial, continuó con sus cálculos obteniendo la aproximación siguiente vi) $31/53$ y

⁶Véase por ejemplo, los artículos de F.J. Budden [B] y los dos de Silver [S1], [S2] para una información mayor.

⁷Véase el artículo de Hasley y Hewitt [H H] para mayor información.

calculó su coma $3^{53}/2^{84}$. Ambas comas son conocidas como las comas de Fang. Muchos años después, Nicholas Mercator recomendó la escala de 53 notas y se construyeron algunos instrumentos con esta escala, uno de ellos el "*Enharmonic Harmonium*" fue construido por Bosanquet y exhibido en 1876 en el South Kensington Museum.

5. Silver en dos artículos muy parecidos [S1] y [S2] dá las dos aproximaciones siguientes vii) $179/306$ y viii) $389/665$, las cuales según él dice, las calculó sin un ábaco y en la esperanza de alcanzar la inmortalidad por supuesto, calculó también las comas correspondientes, las que llamó las comas de Silver.
6. Es muy fácil, con la ayuda de una computadora, -o hasta con una calculadora de bolsillo, como lo hice yo- tanto, calcular un gran número de los términos en la expresión de x como fracción continua, así como los cocientes correspondientes y las comas correspondientes. Sin embargo, como veremos más adelante, a partir de ix) son de una inutilidad tremenda.

Habiendo ya fijado la escala de 12 notas -ó de 7 si se prefiere llamarla así-, examinemos más de cerca la manera en que están distribuidas.

Recordemos que la segunda nota -o primera según se vea- que pusimos en nuestro intervalo, entre ν y 2ν es la que corresponde a la frecuencia de 3ν que llevada a nuestro intervalo es la de frecuencia $3\nu/2$, la siguiente es la de frecuencia el triple de la anterior es decir $3^2\nu/2$ la cual al llevarla al intervalo entre ν y 2ν es $3^2\nu/2^3$, continuando de esta manera obtenemos las frecuencias de todas las notas siguientes que son:

$$\frac{3^3}{2^4}\nu, \frac{3^4}{2^6}\nu, \frac{3^5}{2^7}\nu, \frac{3^6}{2^9}\nu, \frac{3^7}{2^{11}}\nu, \frac{3^8}{2^{12}}\nu, \frac{3^9}{2^{14}}\nu, \frac{3^{10}}{2^{15}}\nu, \frac{3^{11}}{2^{17}}\nu$$

Así como al intervalo entre ν y 2ν la llamamos una octava, porque corresponde a la 8a. nota natural a partir de la frecuencia ν , al intervalo entre ν y $3\nu/2$ se le llama quinta porque corresponde a la 5a. nota natural contada a partir de la de frecuencia ν .

Las frecuencias encontradas al colocarlas en orden creciente quedan:

$$\nu < \frac{3^7}{2^{11}}\nu < \frac{3^2}{2^3}\nu < \frac{3^9}{2^{14}}\nu < \frac{3^4}{2^6}\nu < \frac{3^{11}}{2^{17}}\nu < \frac{3^6}{2^9}\nu < \frac{3}{2}\nu < \frac{3^8}{2^{12}}\nu < \frac{3^3}{2^4}\nu < \frac{3^{10}}{2^{15}}\nu < \frac{3^5}{2^7}\nu$$

y las denotaremos respectivamente por:

$$A B^b B C C^\# D E^b E F F^\# G G^\#$$

forma de distribuir las notas en el intervalo la llamaremos una *afinación* y en lo que sigue veremos que otras afinaciones son factibles. La afinación anterior es llamada *Pitagórica*.

Al calcular la frecuencia de una nota sólo tomamos en cuenta la frecuencia triple de la anterior de ellas, sin embargo, al principio habíamos dicho que todos los múltiplos enteros de una frecuencia son consonantes con tal frecuencia, el siguiente múltiplo que hemos de considerar es el 5. ¿Hay entre nuestras notas ya elegidas, alguna que corresponda a la frecuencia 5ν ? Es decir, al trasladar nuestro intervalo, una nota de frecuencia $5\nu/4$. Es claro que no. La de frecuencia más cercana a $5\nu/4$ es la frecuencia $3^4\nu/2^6 = 81\nu/64$ que difiere de $5\nu/4 = 80\nu/64$ en $\nu/64$.

Si escuchamos simultáneamente los sonidos de frecuencias ν y $3^4\nu/2^6$ y posteriormente los de frecuencias ν y $5\nu/4$, es más agradable al oído el segundo. ¿Qué sucede si reemplazamos el $C^\#$ de frecuencia $3^4\nu/2^6$ por un $C^\#$ de frecuencia $5\nu/4$? Las distancias entre $C^\#$ y D y entre C y $C^\#$ se alteran, la primera aumenta y la segunda disminuye, si antes para ir de $C^\#$ a D debemos multiplicar la frecuencia de $C^\#$ por $\frac{3^7}{2^{11}} = \left(\frac{3^{11}}{2^{17}}\right) \left(\frac{2^6}{3^4}\right)$ ahora lo debemos hacer por $\frac{3^7}{2^{11}} \cdot \frac{81}{80}$, el cociente $81/80$ es conocido en música como la coma de Dídimo o coma común. Al intervalo entre ν y $5\nu/4$ se le llama una tercera.

Por supuesto, si alteramos la frecuencia de $C^\#$, la quinta nota obtenida, al seguir con el procedimiento de calcular las frecuencias subsecuentes por triplicar la anterior, alteramos todos los intervalos siguientes.

Notemos también que la nota de frecuencia 2ν es la que tiene una frecuencia $4/3$ veces la frecuencia $3\nu/2$ y si es agradable oír simultáneamente los sonidos de frecuencias ν y $3\nu/2$ también lo será para las frecuencias $3\nu/2$

y 2ν , pero no hay entre nuestras notas una de frecuencia $4\nu/3$, la que más se aproxima es la de frecuencia $3^{11}\nu/2^{17}$ pues $\frac{3^{11}}{2^{17}} \div \frac{4}{3} = \frac{3^{12}}{2^{19}}$ la coma de Pitágoras. Al intervalo entre ν y $4\nu/3$ se le llama una cuarta.

Así, pues, si queremos introducir en nuestra escala otras notas con sonidos consonantes con ν debemos prescindir de las quintas perfectas.

Dicho sea de paso, si observamos las proporciones de los intervalos introducidos hasta ahora, todos son de la forma $(n+1)/n$, $2/1$ octava, $3/2$ quinta, $5/4$ tercera, $4/3$ cuarta, hay otros, por ejemplo el intervalo entre A y B es de ν a $9\nu/8$ es decir de $9/8$ que es llamado tono mayor y por supuesto, como sólo hemos tomado frecuencias del doble, triple y quintuple, tanto n como $n+1$ tiene una descomposición en primos donde sólo aparecen el 2, el 3 y el 5. Todas estas proporciones son llamadas en música superparticulares y cada una tiene su nombre. Hay un total de 10 de estas proporciones, como fue probado por Størmer en 1897, al resolver ciertas ecuaciones de Pell⁸.

Volviendo a nuestro problema de afinaciones, una afinación es entonces una división del intervalo entre A y A' digamos, en 12 subdivisiones. Las afinaciones que contienen al menos 7 quintas perfectas son llamadas *justas* en contraposición a las ajustadas ó temperadas, de las demás quintas hasta llegar a 11 son llamadas graves y su intervalo es de $2^3 \cdot 5/3$ y el último intervalo llamada quinta Procrustea es lo que falta para llegar a 2^7 .

En otras palabras, si denotamos por j el número de quintas perfectas, y por g el número de quintas graves y por p la quinta Procrustea se tiene la ecuación

$$\left(\frac{3}{2}\right)^j \left(\frac{2^3 \cdot 5}{3^3}\right)^g p = 2^7$$

con las restricciones $j \geq 7$ y $j + g = 11$.

Por ejemplo para la afinación Pitagórica $j = 11$, $g = 0$ y $p = 2^{18}/3^{11}$.

Las quintas graves aparecen por supuesto debido a la introducción de terceras perfectas.

⁸Véase el artículo de Hasley y Hewitt [III] para mayor información.

iv)



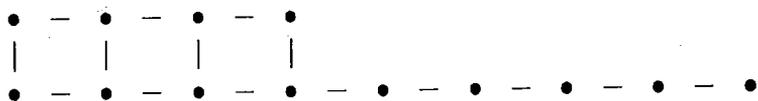
v)



vii)



viii)



ix)



x)



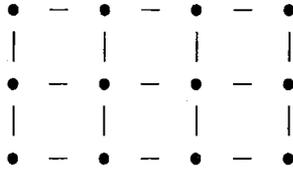
xi)



Para tres líneas horizontales hay 43 arreglos posibles (con 2 quintas graves).

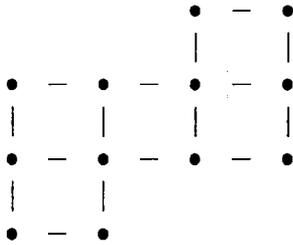
Por ejemplo:

La forma simétrica



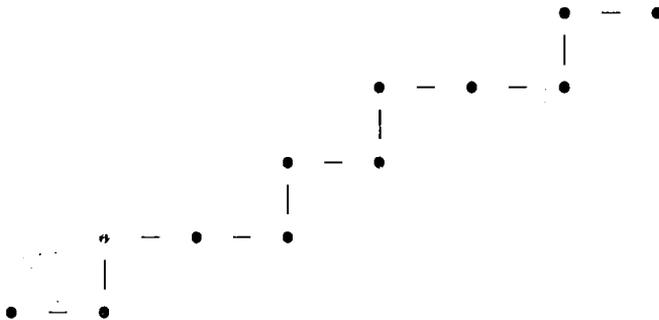
Para cuatro líneas horizontales hay 55 arreglos posibles (con 3 quintas graves).

Por ejemplo el propuesto por Marpurg en 1776



Para cinco líneas horizontales hay 8 arreglos posibles (con 4 quintas graves).

Por ejemplo

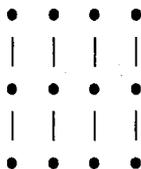


Resumimos todo lo anterior en la tabla siguiente:

Afinaciones justas

Quintas exactas	Quintas graves	No. de arreglos
11	0	1
10	1	11
9	2	43
8	3	55
7	4	8
6	5	0
		<u>118</u>

A través del tiempo han sido propuestas y usadas otras afinaciones. Por ejemplo la llamada Tono medio propuesta por Aron en 1523, la cual consiste en reducir las 11 quintas por $1/4$ de la coma de Dídimo para obtener 8 terceras perfectas.



Esta afinación fue usada desde 1600 hasta principios de este siglo para afinar algunos órganos.

Otra es, la dada por la relación siguiente

C	D	E	F	G	A	B	C
1	$3^2/2^3$	$3^4/2^6$	$4/3$	$3/2$	$3^3/2^4$	$3^5/2^7$	2
	$9/8$	$9/8$	$256/243$	$9/8$	$9/8$	$9/8$	$256/243$

La cual es llamada también Pitagórica, en ella se tiene, entre C y F , una cuarta perfecta, pero entre C y E no hay una tercera perfecta.

Otra llamada también justa es:

C	D	E	F	G	A	B	C
1	$9/8$	$5/4$	$4/3$	$3/2$	$5/3$	$15/8$	2
	$9/8$	$10/9$	$16/15$	$9/8$	$10/9$	$9/8$	$16/15$

Aquí, el intervalo CE es una tercera perfecta, el CF es una cuarta perfecta, el CG una quinta perfecta, etc. Al intervalo de $10/9$ se le llama tono menor y al $16/15$ semitono diatónico.

Observe que en la afinación Pitagórica original el intervalo entre E y F -y también entre B y C - es de $3^7/3^{11}$ mientras que en estas dos últimas, es de $16/15 = 2^4/3 \times 5$.

En estas afinaciones no todas las notas tienen la misma jerarquía, pues por ejemplo, en la última afinación el intervalo entre D y A es de $2^3 \times 5/3^3 < 3/2$, no siendo una quinta perfecta.

De las escalas ajustadas o temperadas, la más usada hoy en día, es la escala temperada igual o igualmente temperada, en ella todas las quintas son reducidas en $1/12$ de la coma de Pitágoras y por tanto todas las notas tienen la misma jerarquía -hay uniformidad de escala-.

Se dice que los chinos ya se habían interesado en una escala igualmente temperada desde al menos mil años antes de nuestra era. El problema era calcular $2^{1/12}$.

Silver menciona⁹ que: un astrónomo chino Ho-Cheng fue acusado de “haber hecho violencia a las cifras” cuando en el año de 420 antes de nuestra era, intentó calcular $2^{1/12}$ y un médico Wang-Po produjo resultados inexactos en 938 antes de nuestra era. No es sino hasta 1528 cuando el príncipe Chu-Tsai-Yu “después de meditar por días y noches antes que la verdad le fuera revelada” produjo cálculos que se dice fueron correctos hasta nueve cifras. Simón Stevin un inspector de canales en Holanda, autor de “*la Disme*” sobre la teoría de fracciones decimales e inventor de buques logró la misma proeza alrededor de 1600. Mientras que un monje de Madrid, Juan Caramuel produjo un refinamiento intitulado “Logaritmi musici”, donde usa logaritmos en base 2.

En 1700 Andreas Werckmeister utilizó esta afinación en un piano.

Silver menciona también que no hay evidencia para suponer que J.S. Bach en su obra *Das Wohltemperirte Clavier* se refiera a esta afinación.

⁹Véanse los artículos de Silver [S1] y [S2].

Hay que hacer notar que la afinación igualmente temperada no fue bien aceptada sino hasta fines del siglo pasado.

Observe también que el semitono igualmente temperado $2^{1/12}$ está entre el semitono cromático $25/24$ y el semitono diatónico $16/15$ esto es: $25/24 < 2^{12} < 16/15$.

En [S1] y [S2] Silver propone la afinación resultante de resolver las 12 ecuaciones siguientes: (donde por A, E, \dots etc. denotamos sus frecuencias).

$$\begin{aligned}\beta &= 3A - 3E \\ \beta &= 3E - 4B \\ \beta &= 3B - 2F^\# \\ \beta &= 3F - 4C^\# \\ \beta &= 3C^\# - 2G^\# \\ \beta &= 3G^\# - 4E^b \\ \beta &= 3E^b - 3B^b \\ \beta &= 3B^b - 2F \\ \beta &= 3F - 4C \\ \beta &= 3C - 2G \\ \beta &= 3G - 4D \\ \beta &= 3D - 4A\end{aligned}$$

Se obtiene fácilmente que $1,568,693\beta = 7153A$ y para $A = 440$, se tiene que $\beta = 2.00633266$; Silver toma $\beta = 1.00316633$, la mitad.

Sin base en lo audible, dado cualquier número $1 < r < 2$ se puede proponer una subdivisión del intervalo entre ν y 2ν en subintervalos de quintas, sólo tomamos enteros α y γ de tal manera que r^α sea próximo a 2^γ y tenemos una escala de α con notas naturales y $\alpha - \gamma$ auxiliares. Al inicio de esta plática tomamos $r = 3/2$ y subsecuentemente se encontraron valores para α y γ . Las frecuencias obtenidas son de la forma $\frac{r^x}{2^y}\nu$ para $2^y < r^x < 2^{y+1}$. En otras palabras, multiplicamos cada frecuencia por r y la dividimos entre 2 el número necesario de veces para que caiga en el intervalo entre ν y 2ν (digamos).

Por supuesto también podemos tomar $0 < r < 1$ y proceder a la inversa.

Por ejemplo, del hecho que para $D = (1 + \sqrt{5})/2$ se tiene que $2^{25}/D^{36}$ sea próximo a 1, se pueden obtener dos afinaciones. Una multiplicando por $p = D^3/4$, obteniendo una división en 12 notas, la segunda que incluye la anterior, al multiplicar por $1/D$, obteniendo una división de 36 notas. Pero también $2^9/D^{13}$ es cercano a uno y al multiplicar por D se obtiene una división en 13 notas. Para ver cómo son las frecuencias aproximadas consúltese las tablas siguientes.

También se menciona que J. Yasser propone una escala de 19 semitonos con 12 notas naturales y 7 auxiliares basándose en la sucesión (de Fibonacci)

$$\begin{aligned} 2, 3 &\rightarrow 5 \\ 5, 2 &\rightarrow 7 \\ 7, 5 &\rightarrow 12 \\ 12, 7 &\rightarrow 19 \\ 19, 12 &\rightarrow 31 \\ 31, 19 &\rightarrow 50 \\ \text{etc.} \end{aligned}$$

y que R. Smith, Master of Trinity y fundador del Smith's Price para matemáticas sugirió una escala de 19 notas "mutables" (7 naturales, 5 auxiliares y 7 mutables) en 1748 y un órgano con esta división se puede ver en la Reid School of Music. El además diseñó un órgano para Handel en el Foundling Hospital el cual tiene una nota extra mutable en cada octava.

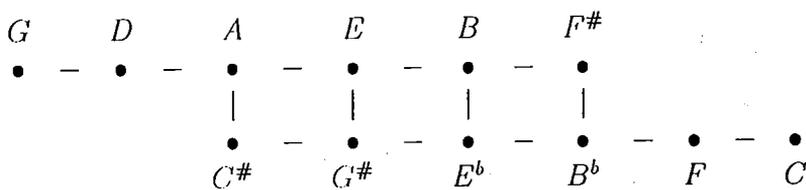
Tabla comparativa de Frecuencias de distintas afinaciones tomando como base $A = 440$ y se aproximó a enteros.

	T.i.	P.	R.P. ⁽¹⁾	S. ⁽²⁾	M. ⁽³⁾	E. ⁽⁴⁾	A. ⁽⁵⁾	P.II ⁽⁶⁾	J. ⁽⁶⁾	S.	$p^{(7)}$	$D^{(8)}$	$41^{(9)}$	$53^{(9)}$
<i>C</i>	262	264	261	258	258	264	257	261	264	262	263	261	261	261
<i>C#</i>	277	278	275	275	275	285	275	277	280	278	278	277	274 ₂₇₉	275 ₂₇₈ 282
<i>D</i>	294	297	293	290	297	297	287	293	297	294	295	288	293	293
<i>E^b</i>	311	313	309	309	309	321	307	311	313	312	312	305	303 ₃₀₉ 314	305 ₃₀₉ 313
<i>E</i>	330	330	330	330	330	330	329	330	330	330	330	323 ₃₄₂	330 ₃₄₁	330 ₃₄₃
<i>F</i>	349	352	348	344	344	356	344	348	352	350	350	356	347 ₃₅₆	357 ₃₄₈ 352
<i>F#</i>	370	371	371	371	371	385	368	369	373	370	370	377	365 ₃₇₂ 378	366 ₃₇₁ 376
<i>G</i>	392	396	391	387	396	396	384	391	396	393	392	399	391	391
<i>G#</i>	415	418	413	413	413	428	411	415	417	416	415	423	411 ₄₁₈ 425	412 ₄₁₈ 423
<i>A</i>	440	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>B^b</i>	466	470	464	464	464	475	460	467	467	467	468	466	463 ₄₇₁	464 ₄₇₀
<i>B</i>	494	495	495	495	495	513	492	495	495	494	496	493	495	495

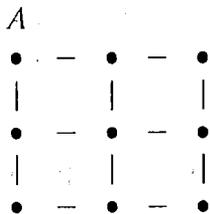
Notas:

$T.i.$	=	Temperada igual	P	=	Pitagórica
$R.P^{(1)}$	=	Ramis de Pareja	$S^{(2)}$	=	Simétrica
$M^{(3)}$	=	Marpug	$E^{(4)}$	=	Escalera
$A^{(5)}$	=	Aron	$P.II^{(6)}$	=	Pitagórica II
$J^{(6)}$	=	Justa	S	=	Silver

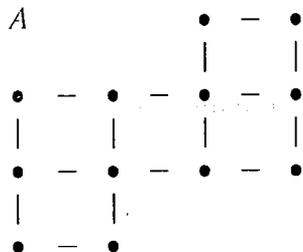
1. Se tomó el diagrama



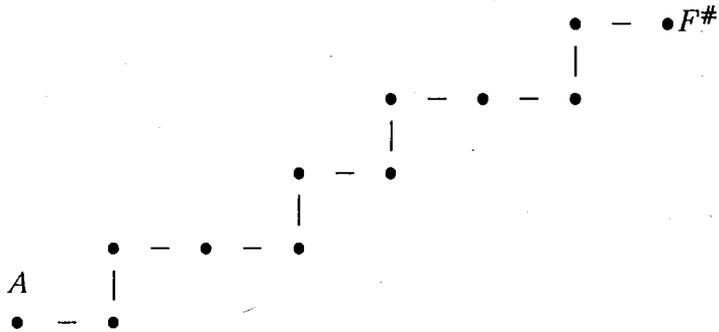
2. Se tomó el diagrama



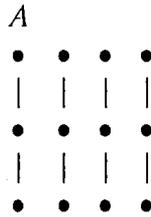
3. Se tomó el diagrama



4. Se tomó el diagrama



5.



6. Los semitonos son igualmente temperados.

7. $p = [(1 + \sqrt{5}) / 2]^{3/4}$.

8. $D = (1 + \sqrt{5}) / 2$.

9. La escala es igualmente temperada.

Comparación de Frecuencias con la escala igualmente temperada

	C	C#	D	E ^b	E	F	F#	G	G#	A	B ^b	B
Pitagórica		x	x	x	x	x	x	x	x	x	x	x
Ramis de Pareja
Simétrica	x	x	x	x	x	x	x	x	x	x	x	x
Marpug
Escalera		x	x	x	x	x	x	x	x	x	x	x
Aron
Pitagórica II	x	x	x	x	x	x	x	x	x	x	x	x
Justa
Silver		x	x	x	x	x	x	x	x	x	x	x
P
D	x	x	x	x	x	x	x	x	x	x	x	x

Silver en el “violín de la monja” termina mencionando que, respecto a los problemas musicales, generalmente los músicos dan las respuestas correctas en base a argumentos ilógicos, mientras que los matemáticos llegan a respuestas incorrectas a través de un proceso de irrefutable razonamiento. ¿Qué tan cierto es esto?

Para terminar trataremos de dar respuesta a algunas de las preguntas que quedaron en el aire al principio de la plática.

Recordemos que cuando aproximábamos por fracciones continuas $\log_2(3/2)$ encontramos las aproximaciones

$$\frac{1}{1}, \frac{1}{2}, \frac{3}{5}, \frac{7}{12}, \frac{24}{41}, \frac{31}{53}, \frac{179}{306}, \frac{389}{665}, \frac{9126}{15601}, \text{ etc.}$$

y decíamos que la última aproximación no sirve a nuestros intereses.

Algo que no dijimos pero que es un hecho muy conocido, es que el oído humano sólo es capaz de percibir sonidos entre cierto rango de frecuencias y que este rango para un oído común y corriente es de 16 Hertz a 20,000 Hertz aproximadamente -esto se puede encontrar en cualquier libro elemental de física-, también es conocido que el oído humano es capaz de distinguir entre dos sonidos cuyas frecuencias difieran en un Hertz si estas frecuencias no

sobrepasan los 4,000 Hertz aproximadamente, por eso restringiremos nuestro sonido a frecuencias de 16 a 4000 Hertz aproximadamente.

Si consideramos sólo escalas igualmente temperadas y primero dividimos el intervalo en 12 partes iguales, para distinguir entre una nota de frecuencia x y la siguiente necesitamos que $x^{2^{1/12}} > x + 1$ es decir que $x > (2^{1/12} - 1)^{-1} \sim 16.82 \sim 17$. Dicho sea de paso, el rango de frecuencias de los pianos que comunmente vemos es de 27 Hertz (27.5) a 4186 Hertz.

Ahora bien, si nuestra división es en 41 partes, un cálculo inmediato ($x^{2^{1/41}} > x + 1$) nos dice que la frecuencia mínima es de aproximadamente 59 Hertz. Para la división en 53 partes la frecuencia mínima es de 76 Hertz, para la de 306 de 441 Hertz, para la de 665 de 959 Hertz y para la de 15,601 es de 22,507 Hertz saliéndose del rango audible.

Observe por último que si construimos un piano con divisiones de 41 hasta una frecuencia próxima de 4000 Hertz necesitamos 250 teclas (casi 3 pianos comunes), como unos pocos cálculos lo demuestran. Si la división es en 53 se necesitan 303 (o 304 teclas), para la división de 306 se necesitan entre 973 a 975 teclas (11 teclados comunes) aunque su rango no vaya mucho más allá de 3 "octavas" y para 665 se necesitan 1415 teclas con un rango de poco más de 2 "octavas".

Por supuesto no necesitábamos restringirnos a 4000 Hertz y pudimos tomar como límite los 20,000 Hertz, aumentando en este caso un poco más de 2 "octavas", por ejemplo para la división de 12 podríamos aumentar 27 teclas a la derecha -y como ya vimos 8 más hacia la izquierda- para el de 665 llegaríamos a 2958 teclas, sin embargo esto acarrea ciertos problemas físicos (de sensibilidad) que no trataremos aquí.

Bibliografía y Referencias

1. [B] Budden, F.J. "*Modern Mathematics and Music*". The Mathematical Gazette pags. 204-215.
2. [F] Fuxs, J.J. "*The Study of Counterpoint*". W.W. Norton & Company Inc. 1965.

3. [HH] Halsey, G.D. y Hewitt, E. "*More on the superparticular ratios in music*". Monthly (1972) pags. 1097-1100.
4. [J] Jeans, J. "*Mathematics of Music*" The World of Mathematics Ed. por Newman, J.R. Simon and Schuster N.Y. 1956 Vol. 4, pags. 2278-2309.
5. [Sch] Scholes, P.A. "*The Beginner's Guide to Harmony*" Oxford University Press London 1977.
6. [Sh] Shilov, G.E. "*Gama Simple*" Lecciones populares de Matemáticas Ed. MIR Moscú 1978.
7. [S1] Silver, A.L. Leigh "*Some Musico-Mathematical Curiosities*" The Mathematical Gazette Vol. XLVIII (1964) No. 363 pags. 1-17.
8. [S2] Silver, A.L. Leigh "*Musimatics or the Nun's Fiddle*" Monthly Vol. 78 (1971) pags. 351-357.
9. [Z] Zinker, S. "*Las bases racionales de la música tonal*" Avance y Perspectiva Nos. 5-6 año 1 (1981) pags. 24-30.

Algunos Aspectos de la Discretización de Problemas de Valores Iniciales

J.P. Hennart †

1 Introducción

En esta conferencia, nos proponemos exponer algunos aspectos básicos de la discretización de problemas de valores iniciales, tanto parabólicos como hiperbólicos. Después de mostrar en la Sección 2 algunas ecuaciones básicas y ejemplos físicos concretos de sus aplicaciones, expondremos en la Sección 3 algunas técnicas clásicas de discretización de la ecuación de difusión del calor que es la ecuación parabólica típica. Esto nos dará la posibilidad de introducir las nociones fundamentales de estabilidad, consistencia y convergencia. En la Sección 4, aplicaremos técnicas semejantes de discretización a la ecuación hiperbólica de advección, lo que nos permitirá ver las diferencias fundamentales entre el caso parabólico y el caso hiperbólico. Finalmente, daremos algunas indicaciones de cómo tratar el caso muy importante en la práctica de las ecuaciones que mezclan los dos aspectos (parabólico e hiperbólico) como la ecuación de difusión-advección.

* Conferencia invitada en el VII Coloquio del CINVESTAV en Agosto de 1991.

† Departamento de Métodos Matemáticos y Numéricos, Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México (IIMAS-UNAM), Apartado Postal 20-726, 01000, México, D.F.

2 Algunas ecuaciones básicas y ejemplos físicos relacionados

La ecuación parabólica de referencia es la ecuación de difusión del calor (o simplemente la ecuación de difusión)

$$\frac{\partial u(x, t)}{\partial t} = \frac{\partial^2 u(x, t)}{\partial x^2}, \quad (2.1)$$

o en notación simplificada

$$u_t = u_{xx}, \quad (2.2)$$

donde $u(x, t)$ debe de cumplir además condiciones iniciales

$$u(x, 0) = u_0(x) \quad \forall x \in (a, b), \quad (2.3)$$

y condiciones de frontera, p. ej.

$$u(a, t) = 0 \quad \text{y} \quad u(b, t) = 0 \quad \forall t \in (0, T]. \quad (2.4)$$

En el caso hiperbólico, el ejemplo clásico es la ecuación de ondas

$$u_{tt} = c^2 u_{xx}, \quad (2.5)$$

o su modelo simplificado, la ecuación de advección

$$u_t + cu_x = 0, \quad (2.6)$$

con la condición inicial

$$u(x, 0) = u_0(x) \quad \text{para} \quad -\infty < x < +\infty. \quad (2.7)$$

En las ecuaciones (2.5) y (2.6), c es una velocidad de onda (acústica, electromagnética, elástica, etc.). La solución general de (2.6) con (2.7) es

$$u(x, t) = u_0(x - ct), \quad (2.8)$$

o sea que hay desplazamiento sin cambio de forma de la condición inicial hacia los x positivos con la velocidad c . Las líneas rectas (si c es una

constante) $x - ct = \text{constante}$ son llamadas las características de la ecuación de advección.

Para entender porqué llamamos a la ecuación de advección un modelo simplificado de la ecuación de ondas, basta recordar que la ecuación de ondas se factoriza en

$$u_{tt} - c^2 u_{xx} = \left(\frac{\partial}{\partial t} - c \frac{\partial}{\partial x}\right) \left(\frac{\partial}{\partial t} + c \frac{\partial}{\partial x}\right) u = 0, \quad (2.9)$$

donde el segundo término corresponde a la ecuación (2.6).

Por consiguiente para la ecuación de ondas, hay dos familias de soluciones y la solución general se escribe

$$u(x, t) = F_1(x + ct) + F_2(x - ct), \quad (2.10)$$

donde el primer término corresponde a un desplazamiento hacia los x negativos mientras el segundo término corresponde a un desplazamiento en sentido contrario. Siendo (2.5) una ecuación con derivadas segundas con respecto al tiempo, necesitamos dos condiciones iniciales, p. ej.

$$u(x, 0) = u_0(x) \text{ y } u_t(x, 0) = v_0(x). \quad (2.11)$$

Con $v_0 = 0$ (ningún movimiento inicial), tendremos que $F_1(x) = F_2(x) = u_0/2$ de tal manera que

$$u(x, t) = \frac{1}{2} u_0(x + ct) + \frac{1}{2} u_0(x - ct), \quad (2.12)$$

mientras que con $u_0 = 0$ (equilibrio inicial), $u_0 = F_1 + F_2 = 0$, $v_0 = 2cF_1'$ y $F_1 = -F_2$ es una integral de $v_0/2c$, es decir

$$u(x, t) = \frac{1}{2c} \int_{x-ct}^{x+ct} v_0(s) ds. \quad (2.13)$$

La solución completa (solución de D'Alembert) se obtiene sumando (2.12) y (2.13).

Veamos ahora algunos ejemplos físicos.

a. La condición inicial es un escalón

En el caso de la ecuación de difusión, esto corresponde a 2 gases con temperaturas diferentes separados hasta $t = 0$, mientras que para la ecuación

de advección es el caso de un muro de agua mantenido hasta el mismo tiempo inicial.

En $t = 0$, removemos la separación entre los gases y se siente inmediatamente la condición inicial en todos los puntos, es decir que la velocidad de propagación es infinita (Fig. 1).

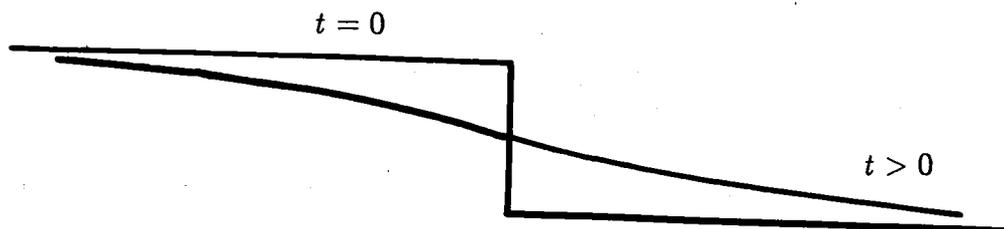


Figura 1 . Solución de la ecuación de difusión con un escalón inicial

El fenómeno es irreversible y de hecho resolver la ecuación de difusión del calor hacia atrás en el tiempo es un problema mal planteado.

Cuando dejamos libre al muro de agua, se mueve con velocidad finita c (Fig. 2).

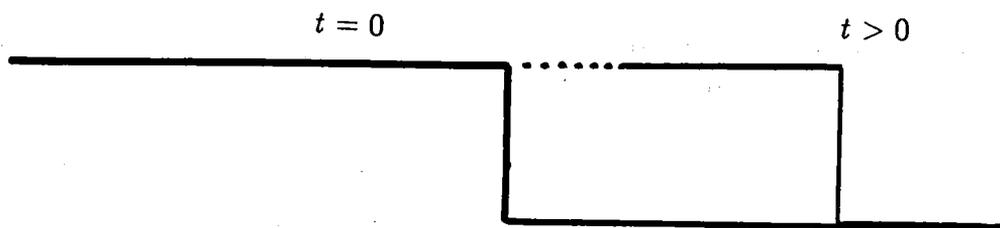


Figura 2 . Solución de la ecuación de advección con un escalón inicial

En este caso, el fenómeno es perfectamente reversible, es decir que cambiando t en $-t$, podemos regresar el muro a su posición inicial. Los esquemas numéricos desarrollados deberían exhibir esta propiedad.

b. La condición inicial es $\text{sen}kx$

En el caso de la ecuación de advección puede ser una onda en un río largo, siendo k el número de onda y $2\pi/k$ la longitud de onda. La solución es

$$u(x, t) = \text{sen}k(x - ct). \quad (2.14)$$

Para la ecuación de difusión es más difícil encontrar una interpretación física pero la solución es

$$u(x, t) = \exp(-k^2t)\text{sen}kx, \quad (2.15)$$

o sea que la onda no se mueve pero decae exponencialmente en tiempo: en otras palabras, la diferencia entre caliente y frío tiende a cero.

En el caso de la ecuación de difusión-advección $u_t + cu_x = u_{xx}$ con las mismas condiciones iniciales, la onda se mueve y decae.

Cabe hacer notar que $\text{sen}kx$ (y $\text{cos}kx$) son casos muy importantes en la práctica si pensamos en descomponer $u_0(x)$ en modos de Fourier.

c. La condición inicial es $\delta(x)$

Para la ecuación de difusión, tendremos

$$u(x, t) = \frac{1}{2\sqrt{\pi t}} e^{-x^2/4t} \quad (2.16)$$

con $\int_{-\infty}^{+\infty} u \, dx = 1$, siendo u máxima en $x = 0$ para todo tiempo, como se ve en la Figura 3.

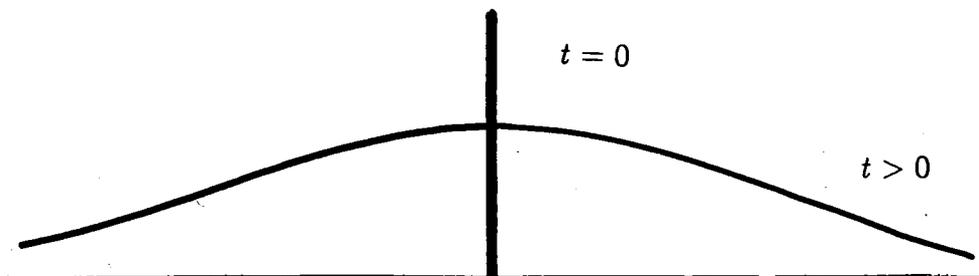


Figura 3. Solución de la ecuación de difusión con una δ inicial

Para la ecuación de advección,

$$u(x, t) = \delta(x - ct). \quad (2.17)$$

El lector interesado en más detalles para estas ecuaciones y otras más debe consultar la Sección 6.4 de Strang (1986).

3 Discretización por diferencias finitas de la ecuación de difusión

Supongamos que queremos integrar la ecuación de difusión desde un tiempo inicial 0 hasta un tiempo final T entre $x_0 \equiv a$ y $x_f \equiv b$. Sobre este dominio $(x, t) \in [a, b] \times [0, T]$, considérese una malla uniforme tanto en x (tamaño h) como en t (tamaño k) (Fig. 4).

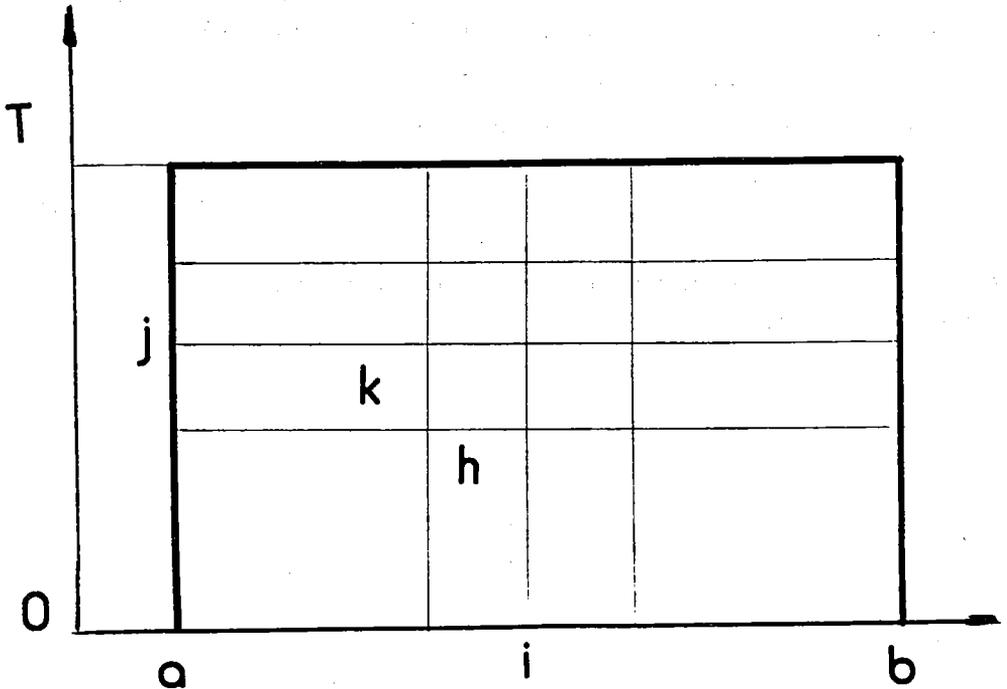


Figura 4. Malla en x y t

La aproximación de $u(x_i, t^j)$ se denotará u_i^j .

Primero se procede a una semidiscretización en x , es decir que para un tiempo t dado $u(x_i, t)$ se aproxima por una función $u_i(t)$ tal que

$$\frac{du_i(t)}{dt} \simeq \frac{\delta_x^2 u_i(t)}{h^2} = \frac{u_{i+1}(t) - 2u_i(t) + u_{i-1}(t)}{h^2}, \quad (3.1)$$

donde δ_x es el operador de diferencia dividida centrada:

$$\frac{du(x)}{dx} \simeq \frac{\delta_x u(x)}{h} = \frac{u(x+h/2) - u(x-h/2)}{h} + O(h^2),$$

y

$$\frac{d^2u(x)}{dx^2} \simeq \frac{\delta_x^2 u(x)}{h^2} = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} + O(h^2).$$

En a y b , tomaremos $u_0 = u_I = 0$. Llamando $U(t) \equiv [u_1(t), \dots, u_{I-1}(t)]^T$ al vector de funciones incógnitas que dependen del tiempo, es fácil ver que hemos transformado nuestro problema inicial en un sistemas de ecuaciones diferenciales ordinarias de la forma

$$\dot{U}(t) = AU(t), \quad (3.2)$$

con $U(t)$ sometido a las condiciones iniciales

$$U(0) = U_0, \quad (3.3)$$

siendo A una matriz de orden $(I-1)$, simétrica y negativa definida. Es además tridiagonal: todos los elementos diagonales son iguales a $(-2/h^2)$, mientras fuera de la diagonal principal valen $(+1/h^2)$.

Este sistema tiene en general una estructura muy particular como la que acabamos de ver y el número de ecuaciones $(I-1)$ puede ser muy grande sobre todo si consideramos problemas con más de una dimensión de espacio. Estamos por lo tanto muy lejos de la situación típica de los textos sobre la solución numérica de ecuaciones diferenciales ordinarias donde usualmente se trata de sistemas pequeños, sin estructura especial y además no lineales en la mayoría de los casos.

Para proceder a la discretización en tiempo de (3.2) con (3.3) y llegar entonces a una discretización completa (en x y t), los esquemas más populares son los esquemas θ cuya forma general es

$$\frac{U^{j+1} - U^j}{k} = A[\theta U^{j+1} + (1 - \theta)U^j] \text{ con } \theta \in [0, 1]. \quad (3.4)$$

Unos casos particulares corresponden a $\theta = 0, 1$ ó 0.5 : cuando $\theta = 0$, tenemos el método de *Euler hacia adelante* o “*Forward Euler*” (FE)

$$\frac{u_i^{j+1} - u_i^j}{k} = \frac{\delta_x^2 u_i^j}{h^2}, \quad (3.5)$$

mientras que para $\theta = 1$, tenemos el método de *Euler hacia atras* o “*Backward Euler*” (BE)

$$\frac{u_i^{j+1} - u_i^j}{k} = \frac{\delta_x^2 u_i^{j+1}}{h^2}. \quad (3.6)$$

Finalmente, con $\theta = 1/2$, tenemos el método de *Crank-Nicolson* (CN)

$$\frac{u_i^{j+1} - u_i^j}{k} = \frac{1}{2} \frac{\delta_x^2 u_i^j + \delta_x^2 u_i^{j+1}}{h^2}. \quad (3.7)$$

Estos métodos se pueden representar en forma esquemática como en la Figura 5.

En el caso de la ecuación de difusión $u_t = u_{xx}$, A no depende del tiempo y la solución formal de (3.2) entre t^j y t^{j+1} es

$$U(t^{j+1}) = \exp(kA)U(t^j), \quad (3.8)$$

con

$$\exp(kA) = \sum_{n=0}^{\infty} \frac{(kA)^n}{n!}. \quad (3.9)$$

Con los esquemas θ , esto se cambia en

$$U^{j+1} = [I - \theta kA]^{-1}[I + (1 - \theta)kA]U^j, \quad (3.10)$$

o sea que $\exp(kA)$ es reemplazada por una aproximación racional. Siendo A simétrica y negativa definitiva, tiene $(I - 1)$ eigenvalores reales negativos λ_i . Llamemos ϕ_i a los eigenvectores correspondientes

$$A\phi_i = \lambda_i\phi_i, \quad i = 1, \dots, I - 1, \quad (3.11)$$

y desarrollemos la condición inicial U_0 como una combinación lineal de estos eigenvectores

$$U_0 = \sum_{i=1}^{I-1} \alpha_{i0} \phi_i. \quad (3.12)$$

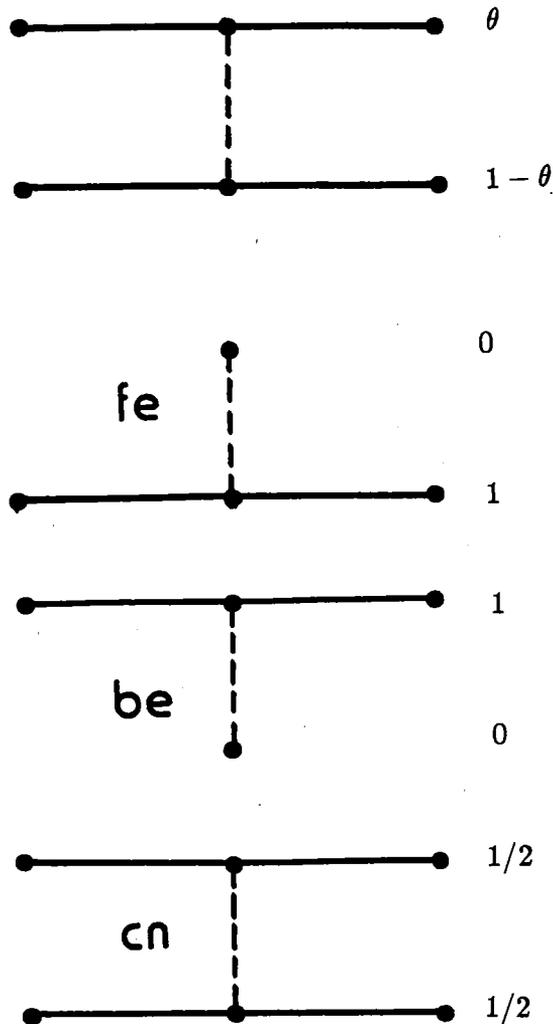


Figura 5. Esquemas θ y sus casos particulares

Los α_i 's representan en (3.12) al peso relativo de los diferentes modos en la condición inicial y posteriormente. Con

$$\dot{U}(t) = \sum_{i=1}^{I-1} \dot{\alpha}_i \phi_i = AU = \sum_{i=1}^{I-1} \alpha_i A \phi_i = \sum_{i=1}^{I-1} \alpha_i \lambda_i \phi_i,$$

vemos que para cualquier i , $\dot{\alpha}_i = \lambda_i \alpha_i$.

Basta entonces estudiar la ecuación escalar de referencia

$$\dot{\phi} = \lambda \phi, \quad (3.13)$$

con la condición inicial

$$\phi(0) = \phi_0, \quad (3.14)$$

cuya solución es

$$\phi(t) = \exp(\lambda t) \phi_0. \quad (3.15)$$

En general, la $\exp(x)$ puede ser aproximada por una aproximación racional de grado p para el numerador y q para el denominador como sigue

$$\exp(x) \simeq R_{pq}(x) = \left(\sum_{k=0}^p a_k x^k \right) / \left(\sum_{\ell=0}^q b_\ell x^\ell \right). \quad (3.16)$$

Los casos particulares considerados anteriormente son derivados del esquema θ general donde

$$\exp(\lambda t) \simeq \frac{1 + (1 - \theta)\lambda t}{1 - \theta \lambda t},$$

que es una aproximación racional del tipo $R_{11}(\lambda t)$. Con $\theta = 0$, obtenemos el esquema FE

$$\exp(\lambda t) \simeq 1 + \lambda t \equiv P_{10}(\lambda t),$$

mientras que con $\theta = 1$, obtenemos el esquema BE

$$\exp(\lambda t) \simeq \frac{1}{1 - \lambda t} \equiv P_{01}(\lambda t).$$

Finalmente, con $\theta = 0.5$, se obtiene el esquema CN

$$\exp(\lambda t) \simeq \frac{1 + \frac{\lambda t}{2}}{1 - \frac{\lambda t}{2}} \equiv P_{11}(\lambda t).$$

En estas expresiones, $P_{pq}(\lambda t)$ en la derecha implica que en los tres últimos casos las aproximaciones racionales (de hecho polinomial para P_{10}) no son cualesquiera: son de hecho aproximaciones de Padé de la exponencial. En (3.16), podemos siempre normalizar a_0 ó b_0 a 1 de tal forma que quedan en total $p + q + 1$ parámetros disponibles: si estos parámetros se escogen para realizar el mejor ajuste de P_{pq} a la exponencial en $x = 0$, R_{pq} se convierte en P_{pq} .

De t^j a t^{j+1} ,

$$\phi(t^{j+1}) = \exp(\lambda k) \phi(t^j), \quad (3.17)$$

que se aproxima por

$$\phi^{j+1} = R_{11}(\lambda k) \phi^j. \quad (3.18)$$

Es fácil darse cuenta que cuando θ es distinto de 0.5 (es decir para FE, BE, y θ en general),

$$\exp(\lambda k) = R_{11}(\lambda k) + O(k^2), \quad (3.19)$$

mientras que para CN

$$\exp(\lambda k) = R_{11}(\lambda k) + O(k^3). \quad (3.20)$$

El error cometido es un error "local" de t^j a t^{j+1} suponiendo que ϕ^j es exacto. Para llegar al tiempo final T , necesitamos $O(k^{-1})$ pasos en tiempo en los cuales se acumulan estos errores locales, de tal manera que en forma heurística un error local en $O(k^{n+1})$ se transforma en un error "global" en $O(k^n)$. FE, BE y R_{11} en general ($\theta \neq \frac{1}{2}$) son por lo tanto esquemas de primer orden mientras CN en un esquema de segundo orden. La Figura 6 trata de dar una idea de cómo aproximan a la exponencial.

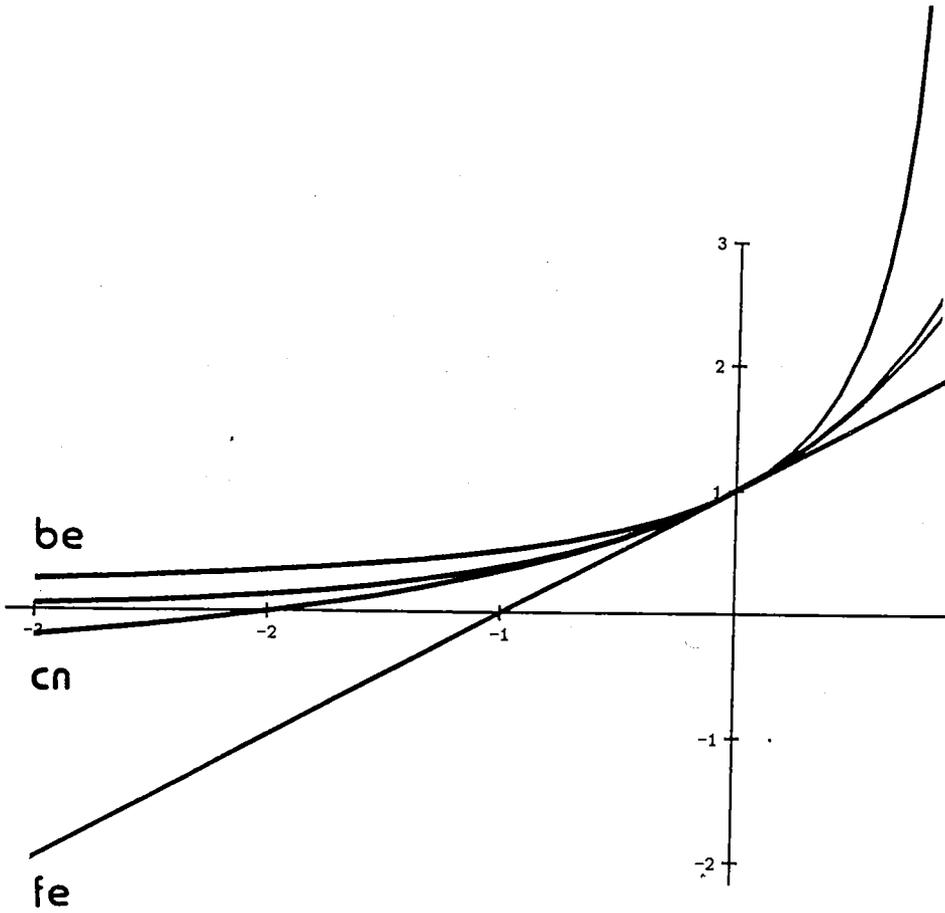


Figura 6. La exponencial y algunas de sus aproximaciones racionales

Ahora bien, hay una pregunta que debemos hacernos: es la de la estabilidad de un esquema. En el caso de FE, $\dot{\phi} = \lambda\phi$ es aproximado por $\phi^{j+1} = (1 + \lambda k)\phi^j$ así que si perturbamos ligeramente ϕ^j en $\phi^j + \delta\phi^j$ (y siempre lo hacemos simplemente por la presencia de errores de redondeo ya que las computadoras tienen una palabra de longitud finita y no pueden representar exactamente los números), esta perturbación en el paso siguiente será $\delta\phi^{j+1} = (1 + \lambda k)\delta\phi^j$. Para que una perturbación introducida

en un tiempo como t^j no pueda crecer (y por consiguiente restarle todo sentido a las aproximaciones posteriores), debemos imponer la condición de estabilidad siguiente

$$|1 + \lambda k| \leq 1. \quad (3.21)$$

k es real positivo pero λ podría ser complejo en casos más generales, lo que es conveniente considerar en (3.21). La región del plano complejo donde se cumple esta condición es un círculo de centro $(-1, 0)$ y de radio 1 en el plano complejo de izquierda ($\Re(\lambda k) \leq 0$) como se muestra en la Figura 7.

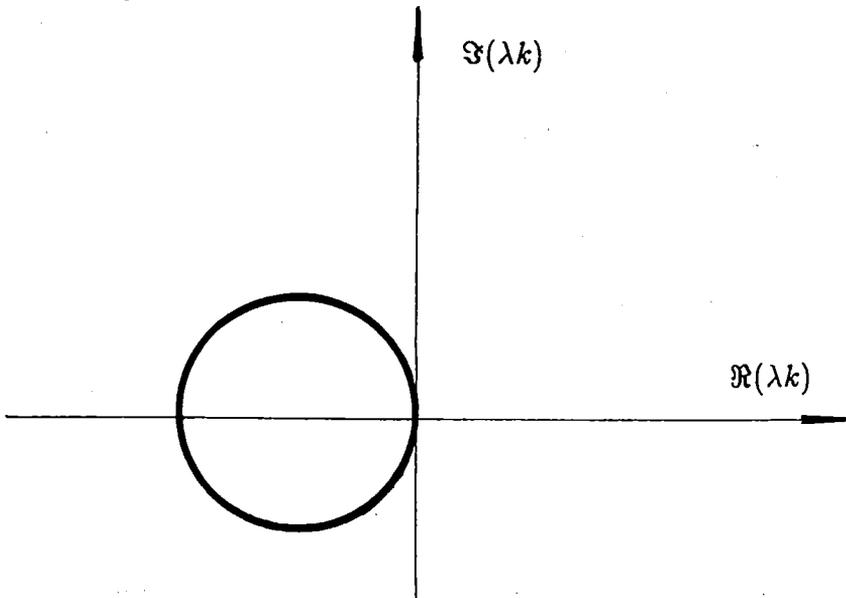


Figura 7. Región de estabilidad absoluta del método FE

Para una ecuación escalar, es muy fácil adaptar la k a la λ para que esto se cumpla, pero para nuestro problema es lejos de ser trivial ya que necesitamos cumplir $|1 + \lambda_i k| \leq 1$ para toda i de 1 a $I - 1$. Los λ_i 's que van a ser cruciales para cumplir la condición de estabilidad son los más grandes en valores absolutos (i grandes): corresponden a transitorios rápidos (exponenciales decrecientes) que desaparecen casi instantáneamente dejando a una solución decreciendo más lentamente en $\exp(\lambda_1 t)$. Un análisis del espectro de A muestra que

$$\left| \frac{\lambda_{I-1}}{\lambda_1} \right| \propto O(h^{-2}), \quad (3.22)$$

siendo h el intervalo de la discretización en espacio. Por consiguiente, tendremos una condición de estabilidad de la forma

$$k \leq Ch^2, \quad (3.23)$$

que resulta ser muy severa ya que si disminuimos h de un factor 2 por ejemplo, tendremos que dividir k entre 4. Muy rápidamente, el número de intervalos en tiempo se vuelve intolerable y se presta a una acumulación importante de los errores de redondeo. Por lo tanto, FE es muy poco recomendable para los problemas parabólicos.

Para determinar la constante C , la técnica más común es la de Von Neumann que consiste en descomponer en serie de Fourier la solución y aplicar el esquema a un modo particular, digamos

$$u(x, t) = \hat{u}(t)e^{i\ell x}. \quad (3.24)$$

Aplicándole FE entre t^j y t^{j+1} , obtenemos

$$\hat{u}^{j+1}e^{i\ell x_i} = \hat{u}^j e^{i\ell x_i} + \frac{k}{h^2} \hat{u}^j (e^{i\ell x_{i+1}} - 2e^{i\ell x_i} + e^{i\ell x_{i-1}}) \quad (3.25)$$

o

$$\hat{u}^{j+1} = \hat{u}^j \left[1 - \frac{4k}{h^2} \text{sen}^2 \frac{\ell h}{2} \right] = \hat{u}^j g(\ell). \quad (3.26)$$

Para obtener la estabilidad, necesitamos que el factor de amplificación $g(\ell)$ sea en módulo inferior o igual a 1 para toda ℓ , lo que nos da

$$k \leq \frac{h^2}{2}, \text{ y por lo tanto } C = \frac{1}{2}. \quad (3.27)$$

Para los demás esquemas θ , conforme θ sube de 0 a $1/2$, la región de estabilidad absoluta de FE va creciendo: es todavía un círculo tocando el eje imaginario pero cuyo centro se va alejando hasta $-\infty$ cuando θ vale $1/2$. En este caso, la región de estabilidad absoluta abarca todo el medio plano $\Re(\lambda k) \leq 0$. De $\theta = 1/2$ a $\theta = 1$, esta región sigue creciendo e invade el medio plano $\Re(\lambda k) > 0$. De hecho es todo el plano complejo menos un

círculo en el medio plano de derecha que se va reduciendo hasta el círculo de centro $(+1, 0)$ y radio 1 para $\theta = 1$.

Entonces con θ en el intervalo $[1/2, 1]$, todos los métodos son estables sin condición, pero con $\theta \neq 0$ son implícitos, es decir que hay que resolver un sistema algebraico a cada paso. En una dimensión, esto no es realmente un problema, pero lo es en 2 ó 3 dimensiones.

Los métodos cuya región de estabilidad absoluta contiene todo el medio plano complejo de izquierda ($\Re(\lambda k) \leq 0$) se llaman *A*-estables. Es el caso de CN que desafortunadamente no tiene un comportamiento asintótico correcto: mientras $\exp(\lambda k) \rightarrow 0$ cuando $\lambda k \rightarrow -\infty$, $P_{11}(\lambda k)$ tiende a (-1) , es decir que para pasos en tiempo k grandes, los transitorios rápidos, en lugar de desaparecer rápidamente, oscilan: en cada paso, son multiplicados por prácticamente (-1) . Al contrario, $P_{01}(\lambda k)$ tiende a 0 cuando $\lambda k \rightarrow -\infty$: los métodos que además de ser *A*-estables presentan este comportamiento asintótico correcto se llaman *L*-estables. BE es solamente de primer orden: por lo tanto la tentación de usar CN de segundo orden es siempre muy fuerte pero si el paso en tiempo es demasiado grande pueden aparecer oscilaciones que podrían pasar desapercibidas si tiene uno la mala idea de imprimir el resultado cada número par de pasos, ya que en estos casos los (-1) se cancelan ($(-1)^2 = 1$).

En lo siguiente definiremos el "error de truncamiento" en el caso de FE. Si $u(x, t)$ es la solución exacta, aplicándole el esquema FE, obtenemos el error de truncamiento e .

$$e \equiv \frac{u(x_i, t^{j+1}) - u(x_i, t^j)}{k} - \frac{\delta_x^2 u(x_i, t^j)}{h^2}. \quad (3.28)$$

Desarrollos en serie de Taylor nos dan

$$\begin{aligned} u(x_i, t^{j+1}) &= u(x_i, t^j) + k u_t(x_i, t^j) + \frac{k^2}{2} u_{tt}(x_i, t^j + \theta_1 k), \\ u(x_{i+1}, t^j) &= u(x_i, t^j) + h u_x(x_i, t^j) + \frac{h^2}{2} u_{xx}(x_i, t^j) \\ &\quad + \frac{h^3}{6} u_{xxx}(x_i, t^j) + \frac{h^4}{24} u_{xxxx}(x_i + \theta_2 h, t^j), \\ u(x_{i-1}, t^j) &= u(x_i, t^j) - h u_x(x_i, t^j) + \frac{h^2}{2} u_{xx}(x_i, t^j) \end{aligned}$$

$$-\frac{h^3}{6}u_{xxx}(x_i, t^j) + \frac{h^4}{24}u_{xxxx}(x_i - \theta_3 h, t^j),$$

donde θ_1, θ_2 y θ_3 son números entre 0 y 1. Sustituyendo estos desarrollos en (3.28), obtenemos que

$$\begin{aligned} e &= \frac{1}{2}ku_{tt}(x_i, t^j + \theta_1 k) - \frac{h^2}{24}[u_{xxxx}(x_i + \theta_2 h, t^j) + u_{xxxx}(x_i - \theta_3 h, t^j)] \\ &= O(k) + O(h^2). \end{aligned}$$

Por lo tanto, el error de truncamiento tiende a cero cuando k y h tienden a cero en forma independiente. Se dice que el esquema FE es consistente. Un teorema fundamental del análisis numérico debido a P. Lax es que consistencia y estabilidad implican convergencia (hacia la solución del problema considerado). Ciertos esquemas deben satisfacer una condición de consistencia que usualmente se presenta como una condición sobre como h y k deben tender a cero uno respectivo al otro.

Para ilustrar este concepto, hablaremos de algunos esquemas en diferencias finitas más exóticos, como el esquema "leapfrog", el esquema de Dufort-Frankel y los esquemas de Saul'ev.

El esquema "leapfrog" es un esquema sobre 3 niveles (t^{j-1}, t^j, t^{j+1}) descrito en la Figura 8 y que se escribe

$$\frac{u_i^{j+1} - u_i^{j-1}}{2k} = \frac{\delta_x^2 u_i^j}{h^2}. \quad (3.29)$$

El análisis de Von Neumann muestra que este esquema es siempre inestable.

Una ligera modificación nos lleva a un esquema siempre estable, el esquema de Dufort-Frankel donde $2u_i^j$ en el miembro de derecha se reemplaza por $(u_i^{j+1} + u_i^{j-1})$, es decir

$$\frac{u_i^{j+1} - u_i^{j-1}}{2k} = \frac{1}{h^2}[u_{i+1}^j - (u_i^{j+1} + u_i^{j-1}) + u_{i-1}^j], \quad (3.30)$$

esquemmatizado en la Figura 9.

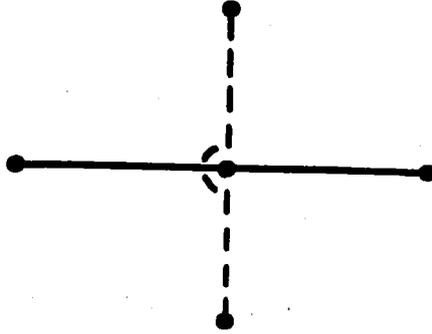


Figura 8. Esquema Leapfrog

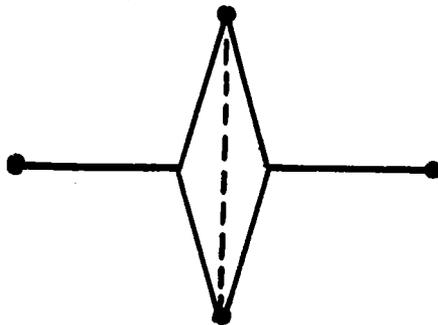


Figura 9. Esquema de Dufort-Frankel

Es explícito: u_i^{j+1} depende únicamente de valores en tiempo anteriores. Además, es siempre estable pero debe satisfacer un condición de consistencia. El error de truncamiento es

$$e = 2\frac{k^2}{h^2}u_{tt} + O(k^2) + O(h^2). \quad (3.31)$$

Si $\frac{k}{h}$ es una constante β , el esquema converge hacia la solución de la ecuación hiperbólica (¡distinta de la ecuación que queremos resolver!)

$$u_t = u_{xx} + 2\beta^2 u_{tt}. \quad (3.32)$$

La condición de consistencia es que β tienda a cero o sea que k tienda a cero más rápidamente que h . Finalmente cabe mencionar que es un esquema de 2 pasos, lo que implica un procedimiento especial de arranque de t^0 a t^1 .

Un esquema que es *indirectamente* de 2 pasos es el esquema de Saul'ev esquematizado en la Figura 10.

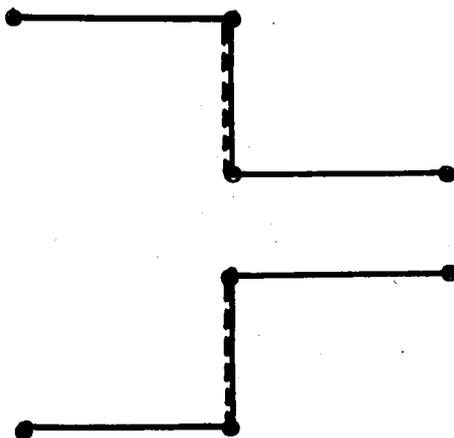


Figura 10. Esquemas de Saul'ev

En su primera etapa, se escribe

$$\frac{u_i^{j+1} - u_i^j}{k} = \frac{1}{h^2} (u_{i-1}^{j+1} - u_i^{j+1} - u_i^j + u_{i+1}^j). \quad (3.33)$$

Si partimos de $x = a$ donde la condición de frontera nos da u_{i-1}^{j+1} , el esquema es explícito; además es estable sin condición pero debe de satisfacer una condición de consistencia. Se usa en forma alternada con el esquema simétrico (segunda etapa) y por lo tanto es un esquema indirectamente de 2 pasos.

4 Discretización por diferencias finitas de la ecuación de advección

En esta sección nos interesaremos en la ecuación de advección (2.6)

$$u_t + cu_x = 0, \quad (4.1)$$

donde $u(x, t)$ debe satisfacer la condición inicial

$$u(x, 0) = u_0(x). \quad (4.2)$$

Una discretización en diferencias divididas que parece muy razonable consiste en reemplazar u_t por una diferencias dividida hacia adelante y u_x por una diferencia dividida centrada dándonos

$$\frac{u_i^{j+1} - u_i^j}{k} + c \frac{u_{i+1}^j - u_{i-1}^j}{2h} = 0. \quad (4.3)$$

Resulta que este esquema es *siempre inestable*.

Lax introdujo una pequeña modificación en donde u_i^j en el primer término es reemplazado por un promedio entre u_{i-1}^j y u_{i+1}^j . (4.3) se convierte en

$$\frac{u_i^{j+1} - \frac{u_{i+1}^j + u_{i-1}^j}{2}}{k} + c \frac{u_{i+1}^j - u_{i-1}^j}{2h} = 0, \quad (4.4)$$

como se esquematiza en la Figura 11. Este esquema es explícito, estable si $\nu \equiv \frac{ck}{h} \leq 1$ y consistente si k es proporcional a h . La condición $\nu \leq 1$ es

conocida como la condición de Courant-Friedrichs-Lewy (CFL) mientras ν es el "Número de Courant".

Cabe hacer notar que esta condición de estabilidad es más fácil de cumplir que las condiciones que encontramos en el caso parabólico ya que al reducir h basta reducir k de manera proporcional. Esto explica indirectamente porque en el caso hiperbólico se siguen usando esquemas explícitos cumpliendo la condición CFL. En el caso parabólico, los esquemas más usados son implícitos como por ejemplo los esquemas θ con $\theta \in [0.5, 1.0]$.

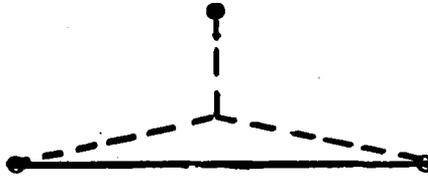


Figura 11. Esquema de Lax para la ecuación de advección

Si buscamos una solución de la ecuación de advección (2.6) del tipo onda armónica

$$u(x, t) = \exp i(kx - \omega t), \quad (4.5)$$

con

ω pulsación

$T = \frac{2\pi}{\omega}$ período

$kx - \omega t$ fase en (x, t)

Claramente $\omega = kc$, la llamada "relación de dispersión". Definamos la velocidad de fase v_f como la velocidad de un punto con fase constante. Entonces

$$kdx - \omega dt = 0, \quad (4.6)$$

y $v_f = c$, es decir una constante que no depende de k . Por lo tanto, los modos de Fourier para la ecuación exacta se desplazan con la misma velocidad (no hay "dispersión") y su amplitud no se amortigua (no hay "disipación").

Un aspecto importante del estudio de los esquemas numéricos desarrollados para la ecuación de advección (2.6) consiste en determinar en que medida hay disipación y/o dispersión numérica. En el caso del esquema de Lax, no hay disipación y dispersión numérica sólo si $\nu \equiv 1$. En general, no es el caso y la manera usual de estudiar un esquema dado consiste en graficar para diferentes ν 's entre 0 y 1, $|g|$ y q en función de h/λ , donde $|g|$ es el amortiguamiento función de $k(\propto \lambda^{-1})$ y $q = v_f/v_{fe}$ es la razón entre la velocidad de fase numérica función de k y la velocidad de fase exacta c . Estas gráficas para los esquemas presentados aquí se encuentran por ejemplo en Brenier y Hennart [1985].

En términos generales, el esquema de Lax es bastante disipativo y dispersivo y una primera idea para mejorarlo consiste en discretizar el término de advección "viento arriba" o "upwind". Esta discretización ya no es centrada pero del punto de vista físico es mucho más lógica porque mira hacia esta parte del eje x de donde viene algo. El esquema "upwind" se muestra en la Figura 12.

Se escribe

$$\frac{u_i^{j+1} - u_i^j}{k} + c \frac{u_i^j - u_{i-1}^j}{h} = 0. \quad (4.7)$$

También debe satisfacer la condición CFL. Por otra parte es menos disipativo y dispersivo que el esquema de Lax. Ambos esquemas son positivos, es decir que una distribución inicial positiva guarda esta propiedad conforme avanza el tiempo. Además conservan la masa

$$\sum_i u_i^j h = \sum_i u_i^0 h$$

y la abscisa promedia \bar{x} se mueve con la velocidad correcta c

$$\sum_i x_i u_i^j h = ct^j + 1$$

suponiendo que $\sum_i u_i^0 h = 1$.



Figura 12. Esquema Upwind

Del lado negativo, tanto el esquema de Lax como como el esquema Upwind son esquemas de primer orden solamente. Al introducir una corrección disipativa (o "viscosidad artificial") al esquema siempre inestable (4.3), se obtiene el esquema de Lax-Wendroff que se escribe

$$\frac{u_i^{j+1} - u_i^j}{k} + c \frac{u_{i+1}^j - u_{i-1}^j}{2h} - \frac{c^2 k}{2} \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{h^2} = 0. \quad (4.8)$$

Es sometido también a la condición CFL. Es más preciso y menos dispersivo que los esquemas anteriores. Sin embargo, no conserva la positividad como se ve en la Figura 13.

Sin dar más detalles, mencionaremos que el esquema Leapfrog de 3 niveles funciona muy bien en el caso hiperbólico. Este esquema no se podía usar en el caso parabólico de la sección anterior. Esto ilustra muy bien el dilema que encuentra uno para discretizar ecuaciones que tienen un carácter mixto parabólico-hiperbólico. El ejemplo típico es la ecuación de advección (o convección)-difusión

$$u_t = u_x + \frac{1}{P} u_{xx}, \quad (4.9)$$

en donde P es el número de Péclet. Si $P \rightarrow 0$, el caso límite es la ecuación de difusión. Al contrario, cuando $P \rightarrow \infty$, obtenemos en el límite la ecuación de advección. Este tipo de ecuaciones se encuentra en muchas aplicaciones, muy importantes como lo son los problemas de difusión y transporte (\equiv advección) de contaminantes en el subsuelo o en la atmósfera. Como lo

hicimos notar, puede ser muy difícil encontrar un esquema que tenga buenas propiedades tanto para la parte difusión como para la parte transporte en un rango amplio de números de Péclet. Una solución muy usada en la práctica consiste en el empleo de *pasos fraccionarios*: se integra la ecuación (4.9) de t^j a t^{j+1} usando (en cualquier orden) el término de difusión solo o el término de advección solo de digamos t^j a $t^{j+1/2}$ y luego de $t^{j+1/2}$ a t^{j+1} . En cada “paso fraccionario”, se usa un esquema discreto bien adaptado al carácter parabólico o hiperbólico de la ecuación considerada. Una referencia reciente (y exhaustiva) es Marchuk [1991].

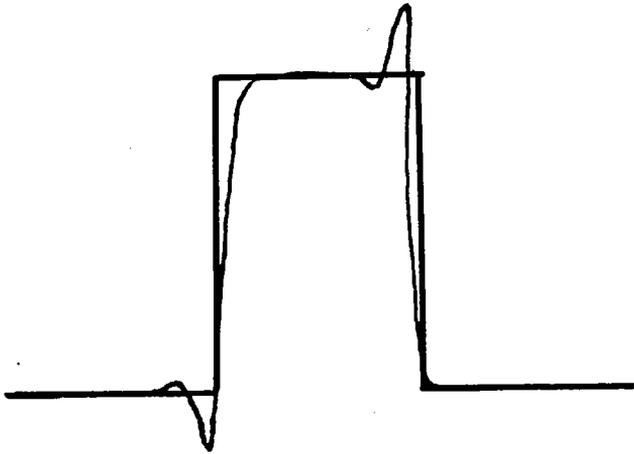


Figura 13. Un ejemplo de no conservación de la positividad con el esquema de Lax-Wendroff

Referencias

- [1] BRENIER, Y. y HENNART, J.P. (1985) Introduction to Numerical Hyperbolic Equations, Com. Técnicas Serie Azul No. 84, IIMAS-UNAM.
- [2] MARCHUK, G.I. (1991) Splitting and alternating direction methods, en Handbook of Numerical Analysis, Vol. I, P.G. Ciarlet y J.L. Lions, pp.197-462, North Holland, Amsterdam.

- [3] STRANG, G. (1986) Introduction to Applied Mathematics, Wellesley-Cambridge Press, Wellesley, MA 02181, USA.

LA FORMA DE LA TIERRA: HISTORIA DE SU BUSQUEDA Y LAS MATEMATICAS QUE INTERVINIERON EN ELLA

MARIA DEL CARMEN JORGE
DEPTO. DE MATEMATICAS Y MECANICA
IIMAS - UNAM
APDO. POSTAL 20-726
DELEG. ALVARO OBREGON
01000 MEXICO, D.F.

INTRODUCCION

Desde tiempos inmemoriales el hombre ha mostrado interés por conocer el lugar que habita y los astros que observa en el cielo.

Para el año 1000 A.C. la astronomía ya estaba bien desarrollada en Babilonia, aunque sus principios se originaron en Mesopotamia y Grecia, en ese orden.

Mientras los babilonios usaron modelos algebraicos para describir el movimiento de los astros, los griegos desarrollaron modelos geométricos.

El desarrollo de la agricultura obligó al hombre a determinar con precisión el inicio de las estaciones. El equinoccio de primavera era apropiado para determinar el inicio de esta estación, de allí el interés de las distintas civilizaciones en estudiarlo con detenimiento.

A fin de determinar con precisión los ciclos agrícolas surge la idea de utilizar el movimiento de la luna como manera de medir los meses. Al ciclo de $29 \frac{1}{2}$ días que es el tiempo promedio entre dos lunas llenas a lunas nuevas, se le conoce como mes sideral; redondeando a 30 días por mes, se tomaron 12 meses como aproximación del año. Sin embargo, esta aproximación hacía que el equinoccio

de primavera se desplazara 10 días cada año. La agricultura no puede tolerar tan evidente error y en aquellos tiempos se optó por ignorar a la luna y basarse en el movimiento aparente del sol para medir los ciclos agrícolas.

En el año 45 A.C. surge el calendario Juliano, introducido por Julio Cesar, en el que establece el sistema de años bisiestos que conocemos en la actualidad, a fin de compensar el error en la aproximación del año astronómico que consta de 365 $1/4$ días. Pero resulta que en realidad el año astronómico cuenta de 365 $1/4$ días menos 11 minutos 14 segundos; aunque esta diferencia no parece significativa, para el año 1582 D.C. el error acumulado era de cerca de 10 días.

El papa Gregorio XIII resuelve el problema ordenando que el día 5 de octubre de 1582 se transforme en 15 de octubre de ese mismo año. Además propone que ciertos años bisiestos del calendario Juliano no se cuenten a fin de evitar nuevas discrepancias. La regla para determinar estos años es que cuando el año bisiesto termine en dos ceros se considerará bisiesto sólo si los dígitos que preceden los ceros son divisibles entre 4. El próximo año bisiesto que se eliminará es el 2100. A este nuevo calendario se le llama Gregoriano y es el que actualmente nos rige.

Aunado a estas discrepancias en la medición del año astronómico estaba la carencia de buenos relojes para medir las horas. Aún durante la época de Newton las longitudes y latitudes no se podían medir con precisión y por tanto los mapas no eran veraces. La invención de buenos cronómetros se dá a finales del siglo XVIII y con ellos la precisión de los mapas mejora considerablemente. La precisión en la medición de las longitudes y latitudes era crucial en la navegación. De todo lo anterior resulta pues evidente la importancia de contar con una manera precisa de medir el tiempo.

El efecto de un pequeño error puede llegar a ser determinante al ir acumulándose, como ilustra el ejemplo del calendario. Algo parecido ocurrió con los primeros medidores de la forma de la Tierra. En consecuencia, durante los siglos XVII y XVIII, se desarrolla una acalorada e interesante polémica sobre esta forma, la cual se reseñará en la primera parte de este trabajo.

La segunda parte comprende la relevancia de la teoría de la gravitación en la determinación de la forma de la Tierra y las matemáticas en ella involucradas.

PARTE I

PRIMERAS HIPOTESIS SOBRE LA FORMA DE LA TIERRA

Para los primeros griegos la Tierra tenía la forma de un disco circular soportado por un enorme océano arriba del cual se encontraba la semi esfera que forma la bóveda celeste. Esta noción se encuentra en los trabajos de Homero y estuvo en boga hasta el siglo VI A.C. El problema con esta figura es que no es claro qué pasa con las estrellas, la Luna y el Sol después de que desaparecen por abajo del oeste del horizonte. De alguna manera, creían que estos astros circulaban hacia el norte, por debajo del horizonte, para luego reaparecer en el este y continuar su curso diurno.

Las observaciones astronómicas acabaron con la idea de que la Tierra era plana pues algunas estrellas que eran visibles, por ejemplo en Egipto, no lo eran en Grecia. Esto afirmaba el que la superficie de la tierra era de alguna manera curva.

Anaximandro (610-547 A.C.) parece ser el primero en tener una idea sobre qué clase de curva tenía la superficie terrestre. Creía que de este a oeste la superficie era plana mientras que de norte a sur se curveaba formando una especie de cilindro. Con esta teoría explicaba la diferencia en las observaciones astronómicas entre Grecia y Egipto, cuya diferencia es esencialmente en latitud, y al mismo tiempo no violaba la idea mitológica de que la región de los muertos se hallaba más allá del oeste.

En las primeras etapas del cristianismo se le atribuía a Pitágoras (550 A.C.) la primicia de concebir a la Tierra esférica. Según Teofrasto (un discípulo de Sócrates), el crédito corresponde a Parménides y según otros historiadores de la época, a Tales de Mileto (600 A.C.). Parménides (finales del siglo VI A.C. y principios del V) plantea un buen argumento para respaldar su idea; argumentó que ninguna otra forma que la esfera mantendría el equilibrio que mantiene la Tierra. La esfericidad de la tierra, además, explica de una manera muy simple

el movimiento de los astros pues una vez que estos llegan al oeste sencillamente continúan su trayectoria circular para reaparecer en el este.

Sin embargo esta idea no ganó popularidad sino hasta más de un siglo después durante la época de Platón. Los argumentos de Platón fueron filosóficos: como la esfera era el cuerpo más perfecto, que tenía la simetría más perfecta y puesto que la Tierra estaba en el centro del universo, ésta debía ser una esfera. A pesar de que este argumento no era tan bueno como el que anticipó Parménides, el poder de convicción de Platón fue suficiente para que a partir de entonces ningún griego suficientemente ilustrado dudara que la Tierra era redonda.

Es Aristóteles (384-322 A.C.) quien presenta un argumento decisivo sobre la redondez de la Tierra. Durante un eclipse de Luna, cuando el Sol, la Tierra y la Luna están alineados, la sombra de la Tierra sobre la Luna es indudablemente circular y a menos de que la Tierra fuera esférica esto no sería posible.

Es muy probable que aún antes de Aristóteles, los navegantes de aquella época hayan llegado a la misma conclusión por la simple observación de cómo se ve emerger a las protuberancias de la tierra conforme uno se va acercando a ellas.

Sin embargo, el argumento de Aristóteles no sería substancial si el eclipse lunar ocurriera a una hora fija pues esto implicaría que tan solo se observa la sombra de una única parte de la Tierra. Como el eclipse ocurre a horas distintas de la noche el argumento es válido. Por otro lado, la observación de los marineros sobre la aparición o desaparición en el horizonte de las protuberancias terrestres, tan solo indica que la Tierra es convexa pero no necesariamente esférica.

Algunas de las obras de los griegos y romanos sobrevivieron durante la edad media y eran bien conocidas pues estaban traducidas al latín. Este es el caso de la obra de Aristóteles "De Coelo" y de la "Historia Natural" de Plinio; esta última contenía los trabajos de Aristóteles y Ptolomeo, en las cuales se sostenía la esfericidad de la Tierra.

A pesar de que algunos ministros de la iglesia sostenían que de acuerdo a las escrituras, la Tierra era plana, no todos compartían las mismas ideas y a través de la edad media no se unificaron las teorías sobre la forma de la Tierra.

Un mito que ha circulado, aún en nuestros días, es que fue Colón quien probó que la Tierra es redonda; en realidad Colón ya estaba convencido de que la Tierra era redonda y su viaje tenía propósitos comerciales y no científicos.

En todo caso fue Magallanes con su famoso viaje de 1522 quien, al darle la vuelta al mundo, confirmó que la Tierra es redonda.

PRIMERAS MEDICIONES

Establecida ya la teoría de la forma esférica de la Tierra, fue natural plantearse el problema de medir su tamaño, es decir su radio. La primicia de esta medición corresponde a Eratóstenes alrededor del año 230 A.C.

Eratóstenes (276-195 A.C.) era el director de la biblioteca de Alejandría, capital de Egipto, la biblioteca más grande e importante de su época; un empleo por cierto muy bien remunerado. Convencido de que la Tierra era redonda, Eratóstenes concibió la idea de calcular su tamaño midiendo un arco del círculo mayor del globo terráqueo. A este tipo de mediciones se les denominan “mediciones de grados”.

Suponiendo a la Tierra esférica, si s_1 es la longitud del arco sobre un círculo mayor correspondiente a un grado, entonces el radio R de la Tierra satisface

$$s_1 = (1^\circ)R = \frac{\pi}{180}R \Rightarrow R = \frac{180}{\pi}s_1.$$

Regresando a Eratóstenes, a unas 500 millas de Alejandría, en la primera catarata del Nilo, se encuentra la ciudad de Asuán, en ella se encontraba un pozo profundo muy particular. Durante el solsticio de verano al medio día cuando el sol estaba más alto y se encontraba en el cenit, su luz iluminaba el fondo del pozo. Esto es equivalente a decir que en Asuán, ese día y a esa hora, el sol no producía sombra alguna sobre una vara vertical.

Al mismo tiempo en Alejandría, el sol producía un ángulo con la vertical estimado en $7^{\circ} 12'$ el cual es exactamente $1/50$ de la circunferencia.

Eratóstenes supuso: 1) la distancia entre Alejandría y Asuán es conocida, 2) ambas ciudades se encuentran sobre un mismo meridiano, de modo que Alejandría está al norte de Asuán pero ambas tienen la misma longitud y la diferencia de latitud de estas dos ciudades es precisamente $7^{\circ} 12'$.

Prosiguió resolviendo el problema geoméricamente. Para su fortuna, desde hacía 50 años el trabajo de Euclides, los "Elementos", era bien conocido. En él se probaba que si dos rectas paralelas son intersectadas por una tercera recta, los ángulos alternos internos son iguales. Ver Fig. 1.

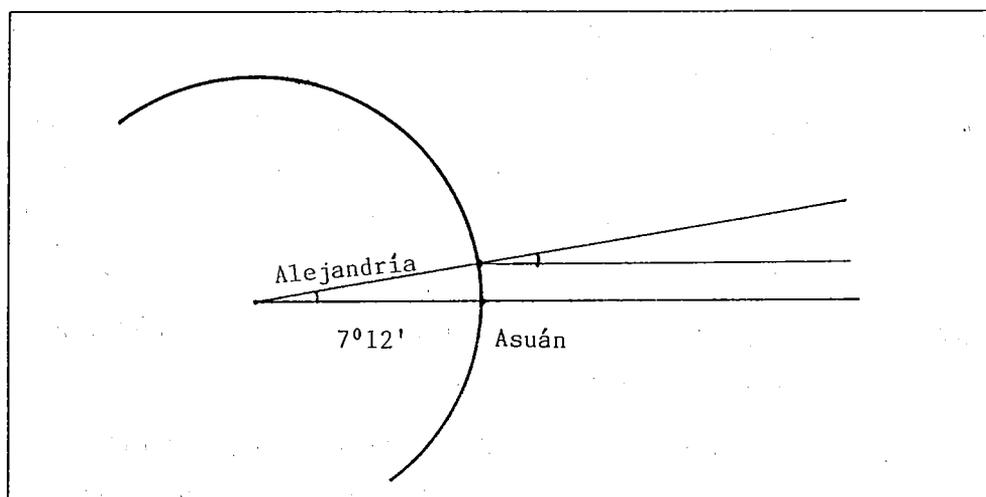


Fig. 1

Luego procedió a calcular la distancia entre las dos ciudades, tal vez utilizando a corredores profesionales (muy comunes en esa época para llevar mensajes urgentes).

Como el ángulo $7^{\circ} 12'$ es $1/50$ de la circunferencia de un círculo mayor de la Tierra, multiplicando por 50 la distancia obtenida se tiene una aproximación de la circunferencia de la Tierra.

De esta sorprendente tarea, Eratóstenes obtuvo un diámetro para la Tierra de 7,850 millas, valor tan solo 70 millas menos que el valor actualmente considerado. Pero según otros autores Eratóstenes falló en un 16% en sus mediciones. La divergencia entre los valores obtenidos por Eratóstenes se debe al tipo de medida que usó, el "estadio", para la cual existían tres versiones distintas: 157 metros para el utilizado en los viajes, 185 metros para el olímpico y 210 metros para el estadio real de Egipto. Según Plinio, Eratóstenes utilizó el primer estadio; si hubiese utilizado el estadio olímpico el error sería de un 17% respecto al valor actual.

Varios errores de medición se conjuntaron favorablemente en el trabajo de Eratóstenes. En primer lugar Alejandría y Asuán no están sobre un mismo meridiano, hay una diferencia de cerca de 3° de longitud. Sin embargo el sol no estaba estrictamente arriba de Asuán sino que producía una sombra con la vertical de aproximadamente $22'$ y además, la diferencia en latitud entre las dos ciudades es en realidad de $7^\circ 5'$. Así, el error en una medición de alguna manera era compensado por el error en las otras en lugar de ir acumulándose, produciendo un error final del uno por ciento.

El método utilizado por Eratóstenes fue excelente y salvo algunas modificaciones se siguió utilizando los 1,000 años siguientes.

En el año 814 D.C. el califa de Bagdad, Al Mamoun, emprendió el famoso proyecto en que midió un grado de latitud. En el desierto llano al norte de Damasco se organizaron dos grupos de jinetes. Después de determinar la altitud de la estrella polar en el punto de partida, un grupo se dirigió hacia el norte y otro hacia el sur. Cuando ambos alcanzaron un punto en el que la altitud de la estrella polar cambió por un grado de la medición previa, ambos grupos se regresaron al punto de partida midiendo la distancia recorrida. El promedio de esta distancia para ambos grupos fue de 71 millas.

En 1525 el francés Jean Fernel calculó la distancia entre París y Amiens contando las revoluciones de una de las ruedas de la carreta en que viajaba. Como ambas

ciudades están casi en el mismo meridiano, después de medir las altitudes de las dos ciudades estimó en 69.1 millas el grado de latitud.

Otras mediciones fueron la de Willebrord Snell en 1617 en Holanda con una cifra de 66.6 millas el grado de latitud. Norwood en Inglaterra en 1635 estimó en 69.5 millas el grado de latitud y Riccioli en Italia en 1650 aseguró que el grado de latitud mide 76.3 millas.

Como dato curioso Delambre y Mechain en los años 1792-1798 volvieron a medir el meridiano de París entre Dunkerque y Barcelona con el propósito de establecer, en 1792, la primera definición del metro como “la diez millonésima parte del cuarto de meridiano que pasa por París”.

La forma exacta de la Tierra se logró establecer sólo después de hacer extensas mediciones de arcos de círculos mayores en varias regiones del globo.

La más cuidadosa de las primeras mediciones de grados la realizó el abate Jean Picard en Francia entre 1669-1670. Además de promover junto con otros astrónomos (Huygens y Auzot) el establecimiento del observatorio de París, también colaboró en el desarrollo de aparatos para medir el diámetro de los cuerpos celestes mediante la adaptación del micrómetro al telescopio. Picard llegó a ser un verdadero experto en el uso de este instrumento. Muchos fueron los proyectos en los que Picard trabajó, entre ellos la elaboración de un nuevo mapa de Francia.

Recordemos que en ese entonces Francia era una potencia económica y cultural en Europa y a la vez escenario de las guerras provocadas por la política expansionista de Luis XIV. Tanto por razones estratégicas o económicas, conocer con mayor exactitud el territorio francés era de vital importancia para la corona. Por ello los proyectos de Picard y más adelante de Cassini, para elaborar un nuevo mapa de Francia, contaron con todo el apoyo económico de parte del gobierno.

A diferencia de su coterráneo Fernel, Picard puso mucho cuidado en la medición del arco de meridiano entre París y Amiens. Fernel simplemente siguió el camino principal entre París y Amiens sin importarle que éste fuera curvas. Picard por el contrario, localizó las partes más llanas del terreno y libres de bosques que

podrían impedir la visibilidad. Por ello decidió utilizar los pueblos de Sourdon y Malvoisine que estaban casi en el mismo meridiano, a una distancia de 80 millas uno del otro, localizados en una franja de terreno muy llano.

Para hacer sus mediciones Picard mejoró el método de triangulación primero introducido por Tycho Brahe en 1589 y luego por Snell en 1697.

Picard comenzó sus mediciones estableciendo primero una línea base de casi siete millas sobre un camino recto pavimentado. Utilizó 4 varillas de 2 toesas (medida francesa que equivale a 1.946 metros) de longitud y una cuerda larga para asegurarse de su rectitud. Luego de medir su línea base AB para asegurarse de su exactitud, llegó a que ésta medía 5,663 toesas. Escogió trece lugares como estaciones para hacer la triangulación en lugares tan diversos como: la torre del reloj de Brie-Comte-Robert, una vara de madera enterrada afuera de la Torre de Monjay, etc. Para mayor precisión utilizó un cuadrante de diez pies de radio en el que montó dos telescopios (uno fijo y otro móvil) equipados cada uno de sus focos por dos hebras cruzadas. Esta es la primera ocasión en que se usa tal aditamento y el crédito de su creación corresponde exclusivamente a Picard. Con el cuadrante en la estación A y en posición horizontal, orientó el telescopio fijo en dirección de la base AB y con el otro telescopio localizó la estación C . Así obtuvo el ángulo BAC . Repitiendo este procedimiento en la estación B obtuvo el ángulo CBA . Conocida la distancia AB y los dos ángulos se pueden obtener los lados AC y BC ya que para cualquier triángulo ABC se cumple

$$\frac{AC}{\text{sen } ABC} = \frac{AB}{\text{sen } BCA} = \frac{BC}{\text{sen } CAB}$$

La medición del primer triángulo arrojó los siguientes datos: ángulo $CAB = 54^\circ 4' 35''$, ángulo $CBA = 95^\circ 6' 55''$, por lo tanto

$$CB = \frac{(5663 \text{ toesas}) \text{ sen } (54^\circ 4' 35'')}{\text{sen } (95^\circ 6' 55'')} \approx 8,953 \text{ toesas.}$$

Aunque la fórmula anterior parece muy simple, la magnitud de las cantidades que tuvo que manejar Picard hacía la tarea muy latosa, a pesar de que hacía ya un siglo que se contaba con tablas de valores de las funciones trigonométricas.

Afortunadamente para Picard, el escocés John Napier habría inventado y publicado las primeras tablas trigonométricas en 1614, cuya versión de 1624 mejorada por Henry Briggs, era de uso común al tiempo que Picard hizo sus mediciones.

La Fig. 2 a continuación muestra las distintas etapas seguidas en las mediciones. Una segunda línea base XY fue erigida por Picard sobre la cual realizó algunas otras triangulaciones como mecanismo de verificación.

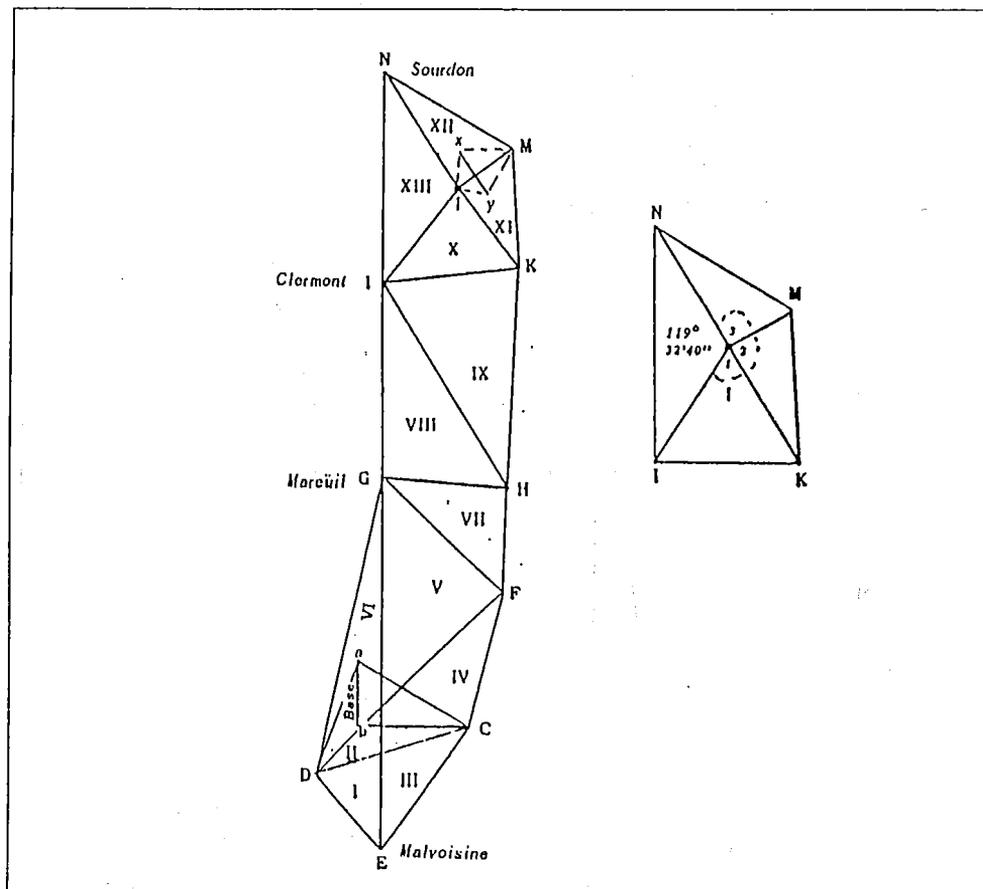


Fig. 2

Finalmente Picard tuvo que hacer unas correcciones debido a que Sourdon y Malvoisine no están en realidad sobre un mismo meridiano, Malvoisine está un poco hacia el este del meridiano. Esto significa que los puntos NIGE no están

sobre una línea recta respecto a la dirección norte-sur sino que tiene la forma mostrada en la Fig. 3.

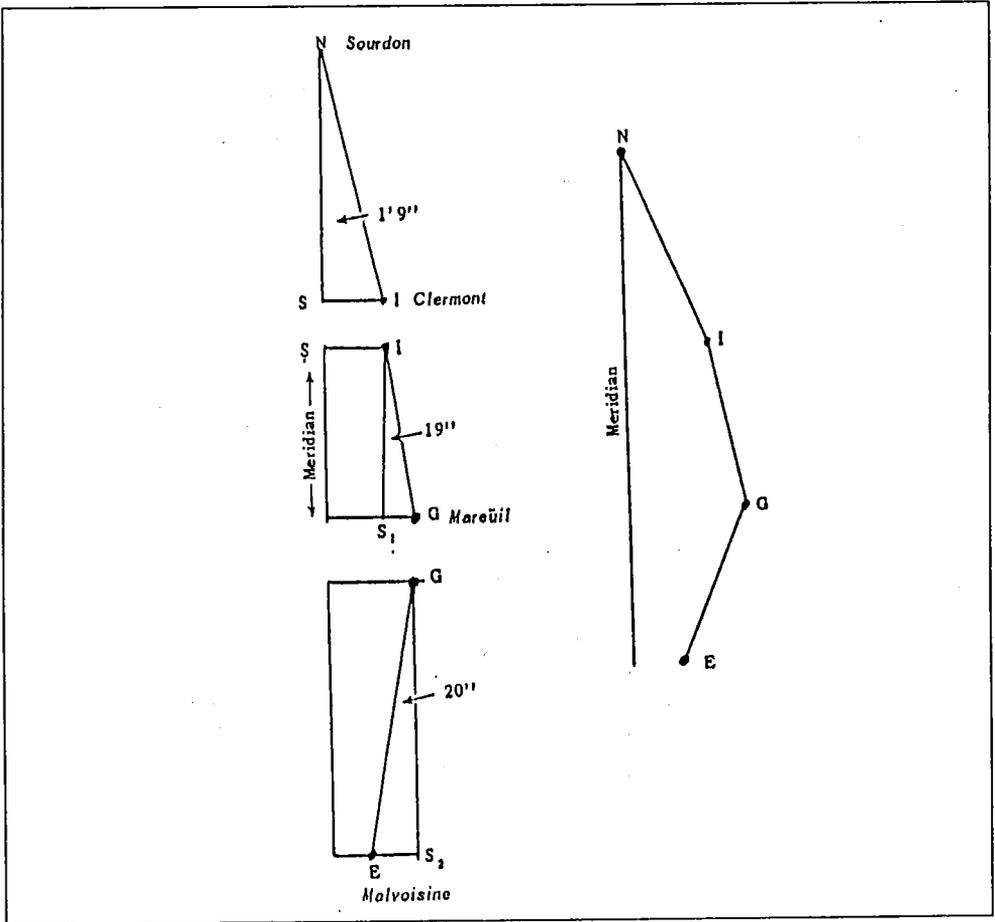


Fig. 3

Después de las correcciones pertinentes, Picard llegó a la cifra 68,347 toesas como la medida del arco de meridiano entre Sourdon y Malvoisine. Como última verificación midió las altitudes de la estrella llamada “Rodilla de Casiopea” desde ambas ciudades, para ello construyó un sector de cenit de 10 pies. Obtuvo una diferencia en latitud de $1^{\circ} 11' 57''$ y de aquí calculó el grado de latitud en 57,060 toesas.

No obstante el cuidado que caracterizó el trabajo de Picard también cometió errores, en parte porque no contaba con las facilidades para hacer mediciones precisas; pero de ningún modo su resultado es desacreditable. Esta medición fue el inicio de un proyecto mayor, el producir un mapa de Francia. El trabajo se inició en 1679 bajo la supervisión de Picard y a su muerte acaecida en 1682, lo continuó G. D. (Giovanni Domenico) Cassini (1625-1712)* y más adelante su hijo Jacques y su nieto, todos ellos astrónomos.

Cuando G.D. Cassini continuó el trabajo de Picard sobre el nuevo mapa de Francia aún no se descubrían los errores cometidos en la medición del meridiano. Al utilizar estos datos, Cassini llegó a la conclusión de que el arco de meridiano se hacía cada vez menor conforme uno se aleja del ecuador rumbo al norte.

Esto sugería que la Tierra era una esfera oblonga, es decir, achatada en el ecuador y alargada en los polos.

Al presentar el nuevo mapa a Luis XIV se cuenta que éste comentó con ironía que había perdido con sus astrónomos más territorios de los que pudo conquistar con sus ejércitos. Lo cierto es que los perfiles del territorio francés elaborados por G.D. Cassini reducían la extensión de los dominios del Rey Sol.

En 1671 el gobierno francés, aconsejado por Cassini, envió a Jean Richer a la isla de Cayenne situada a 5° al norte del ecuador, para efectuar experimentos con el

* Profesor de astronomía en Bolonia en 1650, se hizo famoso por su libro sobre el cometa de 1652. En 1664 logró mayor fama al seguir el curso de un cometa en presencia de la reina de Suecia. Ésta se interesó tanto en el cometa que pasó noches enteras participando en la observación. Todas las predicciones hechas por Cassini sobre el cometa se cumplieron. Escribe su recolección de los hechos en un libro que publica en Roma en 1665 y que dedica a la reina.

Se destacó particularmente por la precisión de sus tablas sobre los períodos de rotación de Marte, Venus, Júpiter y sobre todo de la rotación de los satélites de Júpiter. Esto atrajo la atención de Picard e hizo que lo invitara a Francia donde radicó y trabajó hasta el resto de sus días.

péndulo y realizar observaciones astronómicas. Durante el año que permaneció Richer en la isla, observó que el péndulo del reloj se atrasaba 2'28" por día y descubrió que debía acortar la longitud del péndulo ajustado a la latitud de París para que no se atrasara. El movimiento errático del péndulo ya había sido observado antes; Picard lo notó al estar efectuando sus mediciones pero no le dió la importancia que tiene.

Fué el genio de Isaac Newton quien conjuntó las observaciones de Richer sobre el péndulo, las mediciones de Picard y la forma de esfera achatada en los polos del planeta Júpiter (descubierta por el propio G.D. Cassini), para dar una explicación de los hechos. En 1687 publica su obra "Principia" en la que introduce un concepto revolucionador del universo de su época: la ley de gravitación universal.

NEWTON VS. CASSINI

A partir de Newton se abren nuevas posibilidades en el estudio de la forma de la Tierra. Los métodos puramente geométricos utilizados en los siglos XVI y XVII son ahora enriquecidos por la teoría de la gravitación.

Newton razonó de la siguiente forma: si la Tierra no girara alrededor de su eje entonces todas sus partículas, sometidas a la atracción mutua, deberían formar un cuerpo con forma de globo. A consecuencia de la rotación de la Tierra alrededor de su eje, en cada punto surge una fuerza centrífuga que actúa perpendicularmente al eje de rotación y tiende a alargar a la tierra en dirección del ecuador. Evidentemente esta fuerza será máxima en todos los puntos del ecuador, e igual a cero en los polos. De aquí se deduce, según Newton, que la Tierra a raíz de su diaria rotación debe tomar una forma parecida al elipsoide de revolución achatado en los polos.

Hooke también concibió la forma de la Tierra como una esfera achatada, incluso antes que Newton, y así lo establece en su obra "Sobre Cometas y Gravedad" de 1682.

En lugar de seguir el camino deductivo trazado por Descartes, Newton sigue el inductivo delineado por Galileo. En vez de partir de hipótesis generales acerca de la naturaleza para llegar al conocimiento de hechos particulares, Newton parte del conocimiento de los hechos experimentales para llegar gradualmente a una sistematización más simple y comprensible. Él llegó a su teoría de la gravitación universal, de la sistematización de datos proporcionados por la experiencia; como por ejemplo las leyes de Kepler sobre el movimiento de los cuerpos celestes y las leyes de Galileo sobre la caída de los cuerpos pesados.

La aparición de la obra "Principia" de Newton en 1687, representaba de alguna manera un cuestionamiento al trabajo de Descartes y naturalmente no fue bien acogida por los franceses. Sobre todo cuando acababa de aparecer la obra de Fontenelle (secretario perpetuo de la Academia de Ciencias de París) "Pluralidad de mundos", en la que exponían de manera accesible para el público general las ideas de Descartes. Como las ideas de Newton eran más difíciles de entender, no fue sino hasta la tercera o cuarta década del siglo XVIII en que los franceses aceptaron las hipótesis Newtonianas. Voltaire publica una versión popular sobre la "Filosofía Newtoniana" cincuenta años después de la aparición de "Principia".

Para comprobar la teoría de Newton sobre la forma de la Tierra era necesario determinar la longitud de un arco de meridiano de un grado de latitud en distintas latitudes. Si la longitud del grado de meridiano en la latitud norte resultaba mayor que en la del sur entonces se demostraría que la Tierra estaba achatada en los polos.

Como se mencionó antes, debido a los errores acumulados en las mediciones desde Picard, G.D. Cassini en 1700 llegó a una conclusión contraria a la hipótesis de Newton. Desde ese momento se inicia una larga polémica sobre la forma de la Tierra que habría de concluir muchos años después.

D'Alambert dice en el artículo "Figure de la Terre" (1756) comprendido en la famosa "Encyclopedie": "Se creyó que estaba en juego el honor de la nación dejando a la Tierra tomar una figura extraña, una figura imaginada por un

inglés (Newton) y un holandés (Huygens)... París, y la misma Academia, se dividió entre los dos partidos”.

No es sino hasta 1718 en que Jacques Cassini publica su obra “De la grandeza y de la figura de la Tierra”, donde compendia todas las mediciones hechas en Francia. Sus resultados parecían ser prueba de que Newton estaba equivocado.

En 1735 la Academia de Ciencias Francesa decide poner fin a esta disputa organizando dos expediciones para medir un arco de meridiano en el círculo polar ártico y otro en el ecuador. Las expediciones estuvieron comandadas por dos jóvenes científicos, Maupertuis y La Condamine, de singulares características. Ambos fueron soldados y como tales vivieron experiencias azarosas que de alguna manera moldearon sus personalidades y los prepararon para la aventura. Esta interesante combinación de científico-aventurero, ideal para proyectos como los organizados para medir los arcos de meridiano del ecuador y del círculo polar ártico, fue de suma importancia para concluir con éxito ambas expediciones.

EXPEDICION AL ECUADOR

La primera expedición parte al Perú en 1735. Charles-Marie La Condamine (1701-1774) fue quien propuso primero a la Academia Francesa que mandara una expedición a Cayenne en 1733. Como La Condamine era un miembro muy joven dentro de la Academia su sugerencia no fue escuchada. Sin embargo la idea resultó atractiva a Louis Godin, miembro prestigiado de la Academia, quien logró que se le nombrara jefe de la expedición aunque al final resultó ser quien menos contribuyó en ella. Naturalmente La Condamine también es nombrado por la Academia así como el astrónomo y matemático Pierre Bouguer (1698-1758).

A sus 30 años La Condamine era bien conocido en París. Sus aventuras de su época de soldado y sus viajes por Constantinopla y el Mediterráneo oriental habían hecho de él una figura legendaria. Seguidor de las ideas Newtonianas fue nombrado “químico adjunto” en la Academia Francesa en 1730.

Bouguer, aunque mejor científico que Godin o La Condamine, estaba lejos de ser el compañero ideal para la expedición. De constitución poco saludable y mal carácter, al finalizar la expedición se convierte en enemigo de La Condamine hasta el resto de sus días.

Otros miembros de la expedición fueron el botánico Joseph de Jussieu (1704-1779); Verguin ingeniero naval encargado de hacer los mapas; el relojero y experto en instrumentos Hugo; Morainville responsable de historia natural; dos jóvenes asistentes Couplet y Godin des Odonnais pariente de Godin, y el cirujano Seniergues. La expedición parte de La Rochelle hacia la Martinica a mediados de mayo de 1735. Cuando arriban a Cartagena, Colombia en noviembre de ese año, se unen al grupo dos jóvenes oficiales navales españoles, Jorge Juan y Antonio de Ulloa (de 21 y 19 años respectivamente), nombrados por el rey Felipe V de España como participantes en la expedición por parte de dicha nación (y también a manera de espías del grupo extranjero dentro de los dominios de Felipe).

La expedición estuvo plagada de contratiempos. La burocracia española que reinaba en Ecuador, Perú y Colombia así como la naturaleza de la región, además de conflictos personales de algunos miembros de la expedición, se conjuntaron para hacer el trabajo tremendamente lento y con tragedias que lamentar.

Seniergues se ve involucrado en líos amorosos y termina asesinado por una chusma en una plaza de toros. Couplet, el más joven y vigoroso de los expedicionarios, contrae una fiebre que acaba con él en veinticuatro horas.

La guerra de la oreja de Jenkins entre ingleses y españoles en 1739 que tiene lugar en Cartagena, Colombia y Porto Bello, Panamá, obliga a los españoles Juan y Ulloa a abandonar la expedición para ayudar en la lucha contra los ingleses; regresan a continuar las mediciones un año más tarde.

La triangulación comenzó al noroeste de Quito en el llano de Yaruqui en agosto de 1737 y terminó en julio de 1739. Construyeron alrededor de dos docenas de estaciones algunas de ellas en las altas montañas que estaban a cada lado del valle. La longitud de la triangulación fue de alrededor de docientas millas. Ver

Fig. 4. Al final de la medición en 1740 surgió un conflicto más. Antes de iniciar su regreso a Francia, la Academia Francesa ordenó a La Condamine que erigiera una pirámide en cada extremo de la línea base a manera de signo permanente de la labor allí realizada. Cuando la inscripción conmemorativa se grabó en las bases de piedra que sustentarían a las pirámides, Juan y Ulloa se indignaron al encontrar sus nombres en la parte más baja del texto bajo la categoría de "asistentes". Además aunque se hacía mención de Felipe V, no se decía que el trabajo se había realizado bajo sus auspicios y para colmo las pirámides serían coronadas por una enorme flor de lis, emblema real de Francia. La cuestión fue llevada a la corte; a pesar de que los franceses aceptaron cambiar las inscripciones, en 1742 el gobierno español ordena que las pirámides sean desmanteladas. Casi 100 años después, en 1836, el presidente Rocafuerte de Ecuador mandó erigir las pirámides que hoy pueden verse en el Valle Yaruqui pero no marcan el lugar real donde estuvo la línea base pues con el tiempo se perdió su localización.

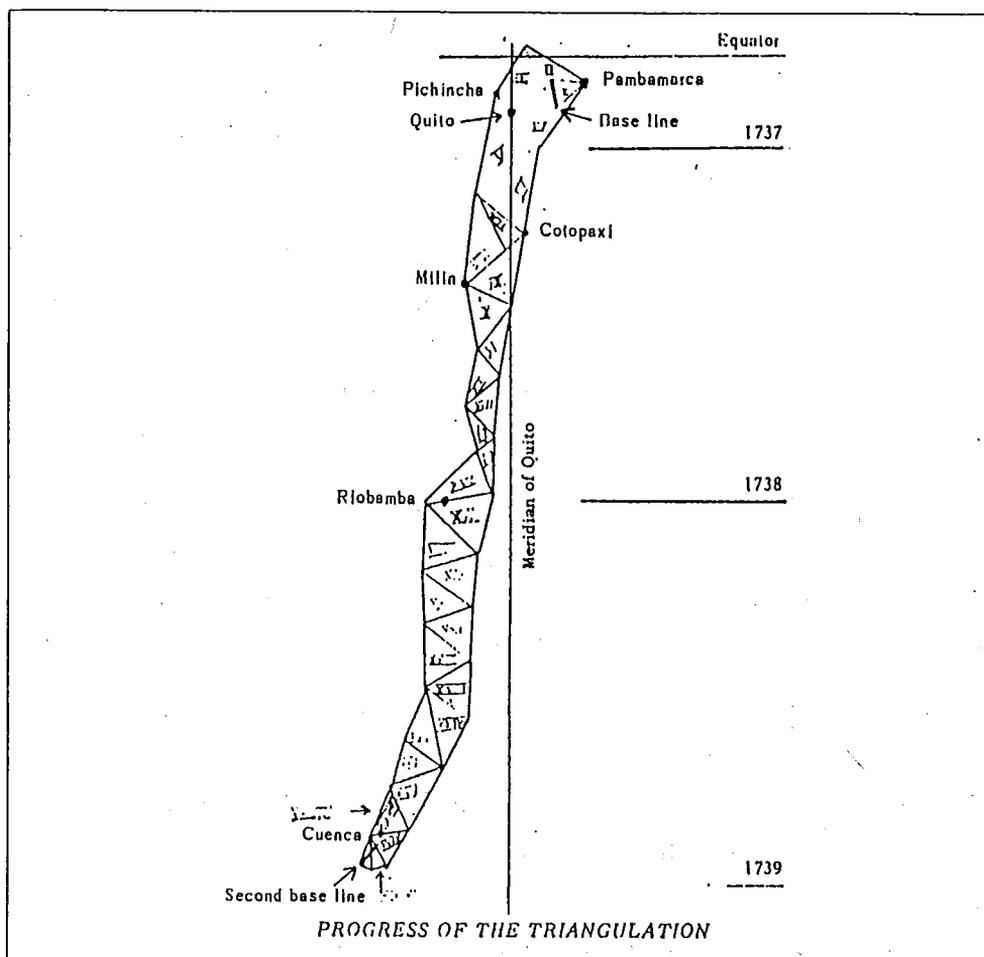


Fig. 4

La república del Ecuador debe su nombre precisamente a este gran proyecto científico. La huella dejada por este grupo de científicos fue tan evidente en el Reino de Quito, que los primeros próceres de la independencia le dieron a las tierras conquistadas al imperio español el nombre de Ecuador.

Además de la triangulación se realizaron observaciones astronómicas y se levantaron mapas; Godin realizó experimentos sobre la velocidad del sonido y Bouguer al experimentar con el péndulo descubrió que la presencia de una gran montaña de masa causa una deflexión en las líneas de fuerza de la gravedad. A este

fenómeno se le conoce hoy en día como "anomalía de Bouguer". Jussieu realizó investigaciones en botánica y mandó a su hermano Bernard, que estaba en Francia y también era botánico, una gran variedad de plantas nuevas. También obtuvo un extracto de quinina.

Bouguer regresa a Francia en 1744. Sus dificultades con La Condamine ya habían comenzado y siendo él mejor científico que La Condamine, resintió la aclamación con que éste fue recibido en Francia acrecentando su rivalidad. Lo que es indudable es que sin el liderazgo de La Condamine la expedición hubiese fracasado. Bouguer siguió investigando en diversas áreas hasta su muerte acaecida en 1758.

Hugo y Morainville se quedaron en Sur América. Godin trabajó un tiempo en la Universidad de San Marcos en Lima hasta 1751 en que va a España para dirigir la Academia Naval en Cadiz. Jussieu se queda en Sur América coleccionando nuevas especies y regresa a París en 1771.

La Condamine fue elegido miembro de la Academia Francesa en 1760 y hasta su muerte en 1774 se involucró en muchos proyectos; uno de ellos fue la campaña de inoculación contra la viruela (de la cual él fue también víctima durante su juventud) en Francia e Italia.

EXPEDICION AL CIRCULO POLAR ARTICO

Pierre-Louis Moreau de Maupertuis (1698-1759) fue el joven científico nombrado por la Academia Francesa para encabezar la expedición a Laponia para medir el arco de meridiano de Kittis. La idea de ir al círculo polar ártico a medir el grado de latitud cerca del polo norte, fue de Maupertuis y surgió después de que la expedición al Perú había ya partido. A pesar de que esta expedición partió un año posterior a la del Ecuador, concluyó mucho antes de ésta.

Ex-oficial de los famosos Mosqueteros Grises, Maupertuis encontraba aburrida la época de paz. Se establece en París como músico y cortesano. Más adelante continúa sus estudios en matemáticas y su avance es notorio. Estudia con los

Bernoullis y produce algunos artículos brillantes en matemáticas; también publica un artículo sobre escorpiones y otro sobre salamandras. En 1731 se vuelve "pensionnaire" en geometría en la "*Academia Real*". En 1728 es elegido miembro de la "Royal Society" de Londres. Entre 1732-35 publica sus obras "Figures des Astres", "Sur le Loix de l'Atraction" y "Sur la Figure de la Terre" en los que defiende las ideas de Newton y la teoría de una Tierra de forma achatada en los polos.

Los acompañantes de Maupertuis nombrados por la Academia fueron: su inseparable amigo Alexis Claude Clairaut (1713-1765), niño prodigio y genio matemático elegido como miembro de la Academia Francesa a los dieciséis años; Pierrè-Charles Le Monnier joven astónomo; Charles Camus, relojero experto; como secretario De Sommereux; el dibujante D'Herbelot y el abate Outhier. Otro interesante personaje se unió a este grupo de científicos franceses, Anders Celsius (1710-1744) profesor de astronomía (al igual que su padre y abuelo) de la Universidad de Upsala, Suecia.

Ya conformada la expedición parten de Dunkerque hacia Suecia el 2 de mayo de 1736. Dos semanas después arriban a Estocolmo donde son recibidos por el rey quien les advierte de lo peligroso que era el viaje que iban a emprender. Después de esta advertencia y de lo accidentado del viaje a Suecia, decide Maupertuis continuar la expedición por tierra y despachar sirvientes y provisiones por barco. La oficina real encargada de hacer mapas les facilitó copias muy precisas de los mapas de las regiones costeras alrededor del Golfo de Bothnia, lugar en que planeaban hacer sus mediciones.

Dos semanas después llegan a Torneo, principal pueblo del sur de Laponia. La primera idea de Maupertuis era la de utilizar las distintas islas del Golfo de Bothnia para hacer sus mediciones pero al darse cuenta que las islas eran muy bajas y boscosas y por tanto la visibilidad muy mala, decidió realizar su triangulación a lo largo del río Torneo cuyo curso iba casi en la dirección norte-sur desde el círculo Polar hasta el Golfo de Torneo. Para fortuna de Maupertuis en ambas riberas del río se elevaban picos que pudieron servir como estaciones para

la triangulación. Aún así el trabajo no fue fácil, lo accidentado del terreno, los insectos y el agobiante calor del verano hicieron enfermar a varios miembros del grupo. El más joven de ellos, Le Monnier (21 años) no soportó y tuvieron que mandarlo a Osver Torneo para que se recuperara. Herbelot a punto de ahogarse en el río se le unió más adelante.

En 63 días realizaron las mediciones de todos los ángulos para lo cual utilizaron un cuadrante de dos pies de radio provisto de un micrómetro, instrumento muy generalizado para entonces. Las técnicas para las mediciones fueron básicamente las utilizadas por Picard con la diferencia de que los instrumentos habían mejorado y, más importante aún, para evitar errores se sistematizaron las mediciones. Maupertuis fue muy cuidadoso a la hora de efectuar las mediciones, distintos grupos efectuaban la misma medición a fin de llegar a una lectura final que no difiriera mucho de las lecturas individuales. La triangulación final se muestra en la Fig. 5.

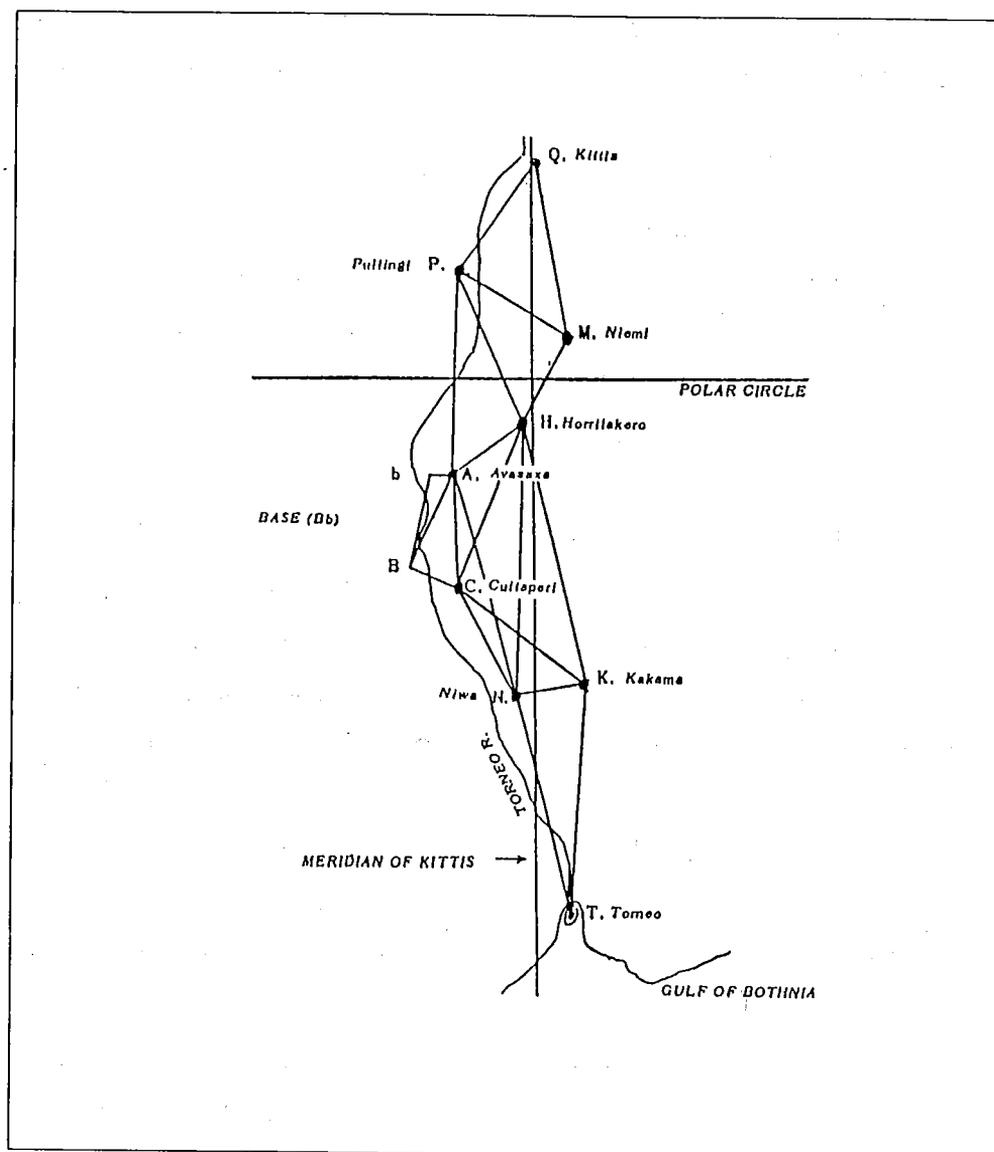


Fig. 5

Ya en vísperas del invierno tuvieron que apresurar sus observaciones astronómicas de verificación en Kittis. La constelación Draco fue utilizada para ello. Lograron llegar a Torneo justo antes de que se congelara el río, donde tomaron la altitud de la estrella escogida al igual que se hizo en Kittis.

En diciembre se efectuó la medición de la línea base y durante el resto del invierno se dedicaron a recheckar las mediciones obtenidas y a montar una línea base auxiliar.

Durante el crudo invierno que vivieron estos científicos en Laponia, observaron que algunos de sus termómetros se congelaban y otros eran poco eficaces para marcar las temperaturas muy bajas. Estas experiencias vividas por Celsius, influyeron para que en 1742 desarrollara su termómetro de mercurio con escala centígrada que utilizamos hoy en día.

Las mediciones eran concluyentes, el arco de meridiano medido en el Círculo Polar era mayor que el medido por Picard, luego Newton estaba en lo cierto. La confirmación final vendría del grupo del Ecuador.

También se utilizó el péndulo para apoyar las mediciones efectuadas. Se observó que el péndulo ajustado para la latitud de París se adelantaba 59 segundos diarios. Esto demostraba, junto con las mediciones de Richer en el ecuador, que la fuerza gravitacional era mayor en los polos que en el ecuador. Así, mientras la longitud del péndulo se debía acortar cerca del ecuador, cerca del polo ésta debía alargarse.

Faltaban aún muchos años para que se conocieran los resultados de la expedición al ecuador; es fácil imaginarnos cómo se sintió La Condamine cuando el éxito de la expedición a Laponia fue de su conocimiento. Para ese entonces apenas comenzaba su triangulación.

Maupertuis y sus compañeros retornan a París, atravesando aún por varias dificultades, y presenta su reporte final a la Academia en noviembre de 1737.

Cassini y sus seguidores acogen la noticia, como era de esperarse, con recelo y amargura (pues destruía el trabajo de su padre y su abuelo) y cuestiona la veracidad de las mediciones de Maupertuis. Celsius defiende sus resultados criticando sin misericordia las mediciones efectuadas por los Cassini. Estos responden con ataques personales sobre "la moral" de Maupertuis y su grupo aludiendo al hecho de que éste regresó a París con la amante finlandesa que le hizo la estancia en Laponia menos penosa.

Para colmo de males Maupertuis se enemista con el cardenal de Fleury debido a que este último, en nombre del rey, distribuyó dinero a los miembros de la expedición, incluyendo una cantidad extra para su líder. Para Maupertuis esto fue un insulto y se negó a aceptar el dinero arguyendo que el honor de sus logros era suficiente recompensa. El tono fue subiendo hasta que Maupertuis le sugiere al cardenal que le dé el dinero a su sirviente que afrontó con él los peligros de la expedición. Sobra decir cómo lo tomó el cardenal.

Cansados de esperar los resultados de La Condamine, Cassini publica en 1744 nuevas mediciones francesas que confirmaban los errores cometidos por sus antepasados. Con esta publicación termina la ya larga disputa entre ambos grupos. Voltaire, amigo de Maupertuis desde hacía varios años, lo bautiza como "el aplanador de la Tierra y de los Cassini".

Sin embargo los problemas de Maupertuis no acaban allá. Fleury interviene en la Academia para que no se le otorgue la membresía. Ya muerto el cardenal se le otorga la tan anhelada membresía en 1743 pero su discurso inaugural no es bien visto por la Academia.

Con todo lo sucedido desde su regreso de Laponia no es de extrañarse que Maupertuis aceptara la invitación de Federico el Grande para unirse a sus planes de hacer de la Academia de Berlín algo grande. En 1744 se va a vivir a Berlín y contrae matrimonio con una dama de la corte de la reina.

Voltaire, amigo de antaño de Maupertuis, se une al grupo de Berlín en 1751 después de años de insistencia por parte del rey. Su decisión llega después de que su amante de veinte años, la matemática Emilie Du Châtelet, muere a consecuencia de un parto.

A corto tiempo de su llegada a Berlín, Voltaire y Maupertuis, tan amigos desde mucho tiempo atrás, se enfrentan por dificultades surgidas entre Voltaire y el rey, y entre Maupertuis y el matemático Koenig. La sátira de Voltaire se vuelca sobre su antiguo amigo y en su libro "Micromegas" de 1752 lo ridiculiza, lo mismo que en su obra "Diatribes du Dr. Akakia". A pesar de que Federico le pidió que no

publicase el libro, Voltaire suprime la edición alemana pero lo publica en Francia. Esto costó a Voltaire su permanencia en Alemania, sin embargo el daño causado a Maupertuis fue ya irreparable. En julio de 1759, Maupertuis muere después de una larga enfermedad, lleno de amargura por el daño ocasionado por su antiguo amigo y admirador.

DESARROLLOS POSTERIORES

Las expediciones de Maupertuis y La Condamine fueron concluyentes para determinar la figura de la Tierra como la de una esfera alargada en el ecuador. Sin embargo muchas cuestiones faltaban por resolver.

El siglo XX produjo las primeras tablas de arcos de meridiano cada 10 minutos de latitud partiendo del ecuador.

La gráfica mostrada en la Fig. 6 fué publicada en "Geographical Journal" en 1941 y compara las mediciones efectuadas en los siglos XVII y XVIII con las de la actualidad.

Resulta impactante observar en la gráfica que a pesar del cuidado con que Maupertuis realizó sus mediciones, su resultado sea tan equivocado. Los resultados de Bouguer y La Condamine por otro lado son sorprendentemente cercanos al actual. También se puede ver que Picard estuvo cerca del resultado correcto mientras que G.D. Cassini muy lejos.

Ninguna de estas equivocaciones merece que el trabajo realizado por estos primeros medidores de la Tierra sea despreciado. El esfuerzo en hacer sus cálculos lo mejor posible, se dió y sus métodos sirvieron como punto de partida de sus sucesores.

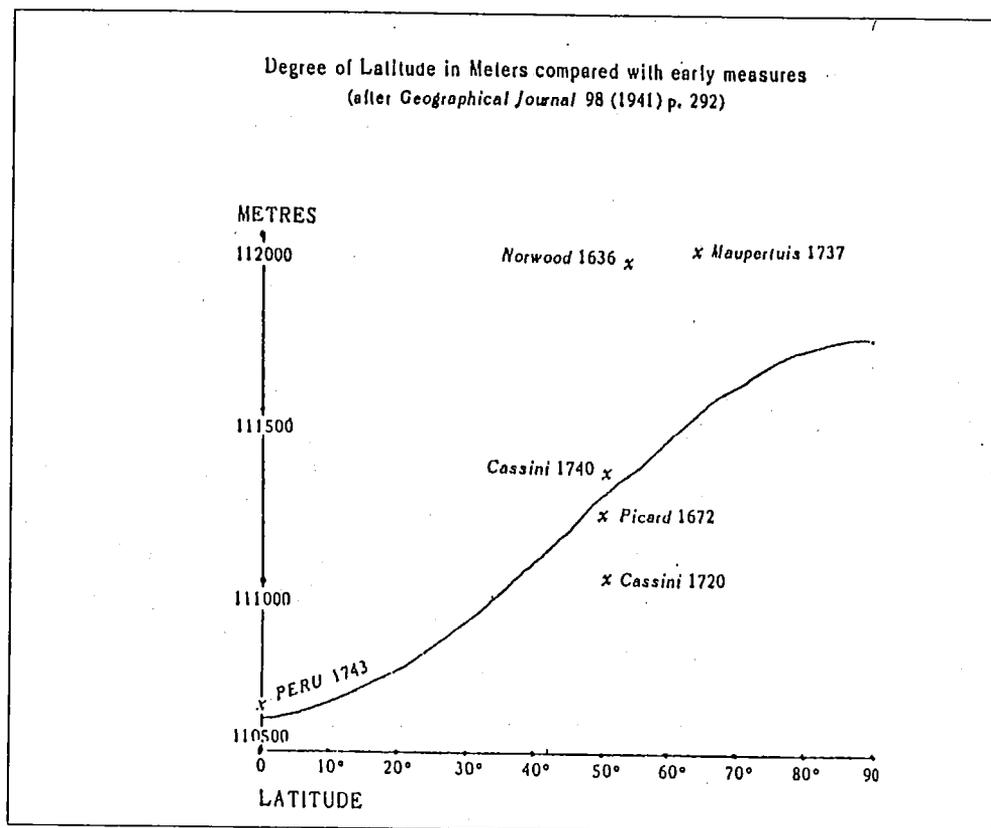


Fig. 6

En la segunda mitad del siglo XX y gracias al fuerte desarrollo de los satélites artificiales, durante 1966-1971 la dependencia U.S. "Coast and Geodetic Survey" implementó un proyecto de triangulación mundial realizada a partir de las observaciones del satélite "Pageos". La red mundial se formó de aproximadamente 45 estaciones, en su mayoría uniformemente distribuidas sobre la superficie del globo, a una distancia promedio de 4,000 Km. Ver Fig. 7.

Se realizaron un total de 14,500 observaciones desde el espacio mediante cámaras balísticas especiales. A fin de determinar la escala de la red, se realizaron mediciones astro-geodésicas convencionales en algunas de las estaciones. Las posiciones tridimensionales de las 45 estaciones fueron determinadas con un error

promedio de ± 4.5 m, con excepción de cuatro estaciones donde el error no sobrepasó los 6 m.

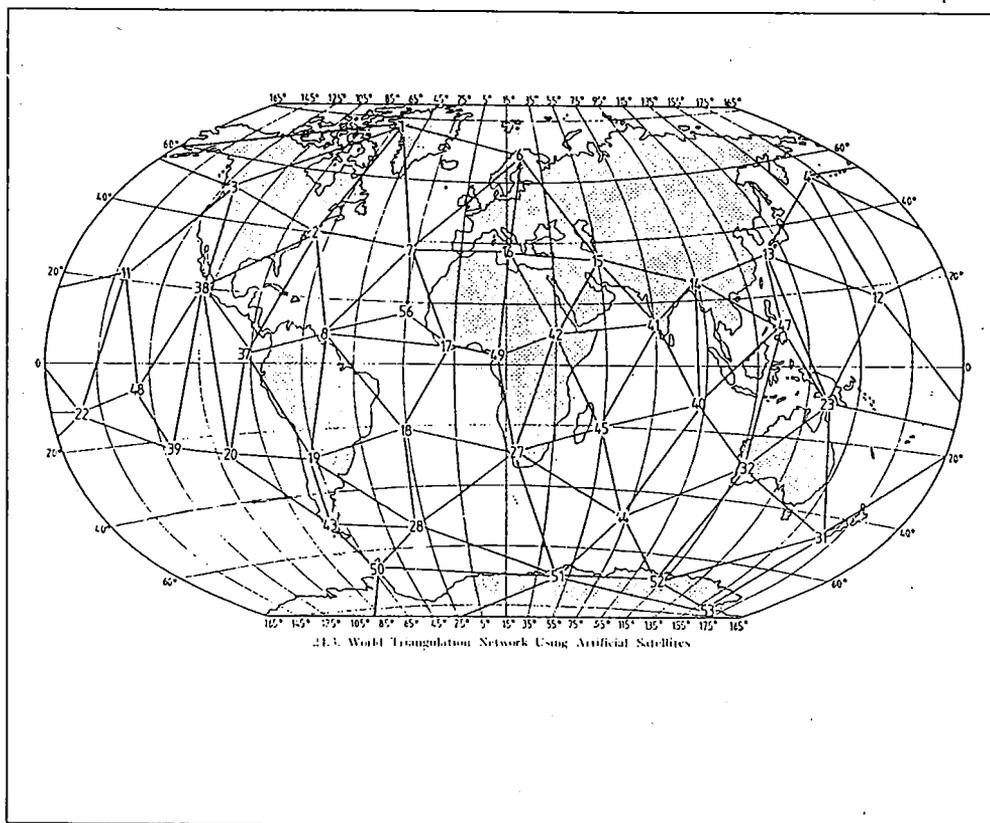


Fig. 7

¡Qué maravillados quedarían Eratóstenes, Snell, Picard, los Cassini, Maupertuis, La Condamine, Clairaut, Bouguer y demás medidores pioneros de la Tierra, si pudieran ver esta figura!

PARTE II

EL PAPEL DE LA GRAVEDAD

Newton establece la ley de gravitación universal en los siguientes términos: "dos puntos materiales se atraen entre sí con una fuerza que es directamente proporcional a sus masas e inversamente proporcional al cuadrado de la distancia entre ellos; la dirección de esta fuerza coincide con la recta que une estos dos puntos".

A partir de esta ley se implementa el método físico para el estudio de la figura de la Tierra. La cuestión se plantea como el estudio del equilibrio de un cuerpo líquido o viscoso que rota, en el que todas sus partículas se atraen según la ley de atracción universal.

La formación del aspecto exterior de la Tierra surgió bajo la acción de las fuerzas gravitacionales, es decir, la fuerza de atracción y la fuerza centrífuga.

Se denomina campo de gravitación de un cuerpo al conjunto de fuerzas de atracción de las partículas elementales del cuerpo en el espacio que lo rodea. La acción total del campo de gravitación de la Tierra y la fuerza centrífuga forman al campo de las fuerzas de gravedad.

El conocimiento del campo gravitacional de la Tierra es imprescindible en los cálculos relacionados con el lanzamiento de los satélites artificiales de la Tierra y el cálculo de sus órbitas. Las mediciones del campo gravitacional se utilizan en geología, astronomía y otras ciencias.

Veamos cómo se calcula la fuerza de gravedad.

Sobre un punto A en la superficie terrestre actúan dos fuerzas: la de atracción terrestre AF y la centrífuga AQ perpendicular al eje de rotación PO como mues-

tra la Fig. 1. Denotando a \vec{g} como la fuerza de gravedad, \vec{F} la fuerza de atracción y \vec{Q} la fuerza centrífuga, se tiene que

$$\vec{g} = \vec{F} + \vec{Q}. \quad (1)$$

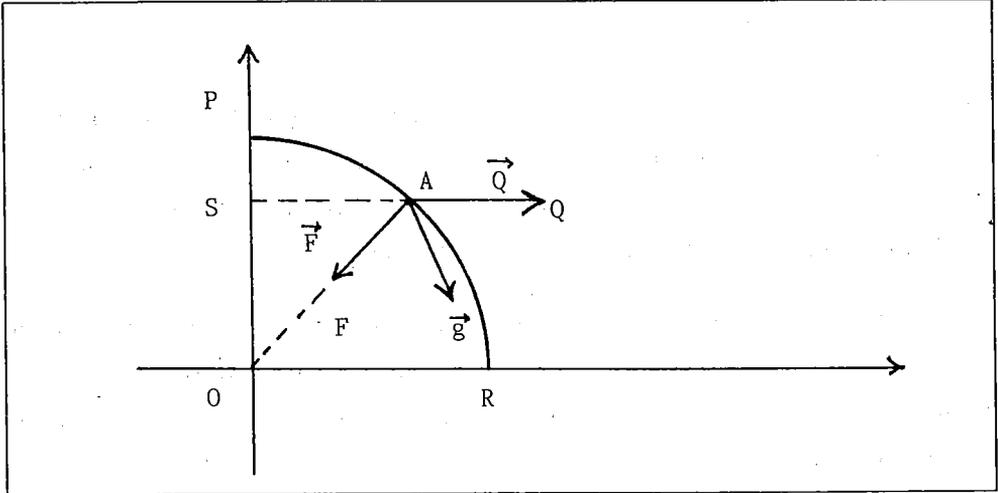


Fig. 1

Denotando por m la masa del punto A , M la masa de la Tierra, G la constante de gravitación de Newton y R el radio de la Tierra

$$F = G \frac{mM}{r^2}, \quad Q = \frac{v^2}{\rho}, \quad (2)$$

donde v es la velocidad lineal del punto A y ρ la distancia de este punto al eje de rotación, es decir $\rho = AS$. Si ω es la velocidad angular de rotación de la Tierra entonces $v = \omega\rho$, de aquí que si $m = 1$

$$Q = \omega^2 \rho. \quad (3)$$

ω ha podido medirse con mucha exactitud a partir de las observaciones astronómicas, de aquí que el estudio del campo gravitacional de la Tierra se reduce más bien al estudio de su campo de atracción.

La fuerza centrífuga Q alcanza su máximo en el ecuador y se hace cero en los polos como puede verse claramente de (3).

Por otro lado la fuerza de atracción F depende de la distribución de las masas en el cuerpo terrestre, o sea, de la densidad de las sustancias del interior de la Tierra.

Para una Tierra de masa homogénea y de forma esférica, la fuerza de atracción sería constante. Si no existiera fuerza centrífuga, entonces la fuerza de gravedad en cualquier punto sobre la superficie, también sería constante. Como esto no ocurre, las partes de la superficie terrestre más alejadas del eje de rotación sienten más la influencia de la fuerza centrífuga y su forma variará más.

Como tampoco la masa es homogénea, la forma de la Tierra también dependerá de la distribución de la densidad de las sustancias que la componen. De manera que es imposible determinar la forma de la Tierra teóricamente sin tener datos determinantes de su campo gravitacional.

Sobre un punto de la superficie terrestre también actúan la fuerza de atracción del Sol y la Luna y la atracción de la masa de la atmósfera que rodea la Tierra. Otras variaciones en F y Q son ocasionadas por los fenómenos de flujos y reflujos de los continentes y océanos, la redistribución de la masa de la Tierra, las oscilaciones de los polos terrestres y el cambio de la velocidad de rotación de la Tierra, entre otros. Aunque el efecto de estos fenómenos puede considerarse pequeño, la precisión actual en la medición de la fuerza de gravedad requiere el tomarlos en cuenta.

La teoría de la relatividad de Einstein también introdujo cambios esenciales en la forma de ver las fuerzas de gravedad. Esta teoría introduce correcciones a los movimientos calculados según la ley de Newton; sin embargo estas correcciones son insignificantes en las cuestiones de la geodesia superior.

GRAVIMETRIA

La fuerza de gravedad se caracteriza por la aceleración que adquiere un cuerpo que cae libremente. El pionero en realizar la determinación de la aceleración de la fuerza de gravedad fue el físico y astrónomo italiano, Galileo Galilei (1564 - 1642). En la actualidad la medición de esta aceleración se realiza con gran exactitud y cuenta con dos clases de métodos: dinámicos y estáticos.

Los métodos dinámicos se basan en la medición del tiempo de movimiento de los diversos cuerpos. En el estático el cuerpo está en reposo y se mide el desplazamiento del cuerpo provocado por su peso.

Algunos de los métodos dinámicos son: la velocidad de caída de un cuerpo y el período de oscilación del péndulo bajo la acción de la fuerza de gravedad. Entre los métodos estáticos se encuentra el cambio de posición de un cuerpo en equilibrio bajo la acción de la fuerza de gravedad y de la fuerza elástica del resorte. Los aparatos basados en los métodos dinámicos se llaman "instrumentos pendulares" y los basados en los métodos estáticos se llaman "gravímetros".

La ecuación de movimiento del péndulo de longitud L viene dada por

$$\ddot{\theta}(t) = -\frac{g}{L} \operatorname{sen} \theta(t) \quad (4)$$

donde θ representa el ángulo del desplazamiento respecto a la posición de equilibrio y g es la aceleración de la fuerza de gravedad.

Para amplitudes pequeñas de oscilación, es decir para θ pequeña, la ecuación se linealiza utilizando el resultado

$$\operatorname{sen} \theta \approx \theta \quad \text{para} \quad \theta \approx 0,$$

de manera que la ecuación (4) se aproxima por la ecuación

$$\ddot{\theta}(t) + \frac{g}{L} \theta(t) = 0$$

cuya solución general viene dada por

$$\theta(t) = A \cos \left(\sqrt{\frac{g}{L}} t \right) + B \operatorname{sen} \left(\sqrt{\frac{g}{L}} t \right), \quad A, B \text{ constantes.} \quad (5)$$

Por lo tanto el péndulo linealizado para pequeñas amplitudes tiene período igual al período de $\theta(t)$ dada por (5), es decir, el período T del péndulo lineal es

$$T = \frac{2\pi}{\sqrt{g/L}} = 2\pi \sqrt{\frac{L}{g}}. \quad (6)$$

T representa dos veces el tiempo que tarda el péndulo en ir de un punto extremo al otro. De la ecuación (6) se obtiene

$$g = \frac{4\pi^2 L}{T^2}$$

de modo que conociendo T , el período de oscilación del péndulo y su longitud L , se calcula g la aceleración de la fuerza de gravedad buscada. A este valor de g se le llama “absoluto” pues no depende de la fuerza de gravedad determinada en otros puntos. Sin embargo éste es un modelo matemático idealizado del péndulo en el cual el largo L del péndulo se considera inextensible e imponderable, cuando la realidad del péndulo físico utilizado para las mediciones es otra. Para ser más realistas se tendría que calcular la “longitud reducida” del péndulo físico, es decir, se tendría que calcular la longitud del péndulo matemático que tiene el mismo período de oscilación que el péndulo físico considerado. Realizar dicho cálculo es una labor de gran envergadura y por ello en lugar de calcular la aceleración “absoluta” se busca la aceleración “relativa” respecto a un punto inicial dado para el cual se conoce la aceleración de la gravedad. Por ejemplo, supongamos que medimos los períodos en tres puntos distintos y los denotamos por

$$T_0 = 2\pi \sqrt{\frac{L}{g_0}}, \quad T_1 = 2\pi \sqrt{\frac{L}{g_1}}, \quad T_2 = 2\pi \sqrt{\frac{L}{g_2}}.$$

Si queremos comparar respecto a la aceleración inicial g_0 tenemos

$$\frac{T_0^2}{T_1^2} = \frac{g_1}{g_0}, \quad \frac{T_0^2}{T_2^2} = \frac{g_2}{g_0},$$

por lo tanto

$$g_1 = \frac{T_0^2}{T_1^2} g_0 \quad , \quad g_2 = \frac{T_0^2}{T_2^2} g_0$$

lo cual permite calcular las aceleraciones de la gravedad g_1, g_2 sin necesidad de conocer la longitud reducida del péndulo. La exactitud de la medición de las g 's ya no va a depender de la longitud del péndulo y aunque depende de la exactitud de la medición de los períodos de oscilación, el error de estas mediciones está cercano a 4×10^{-7} segundos.

Lo anterior ilustra los métodos dinámicos en la medición de la aceleración de la gravedad. Para ilustrar los métodos estáticos se usan aparatos que utilizan la fuerza elástica del resorte para medir el peso de un cuerpo de masa constante. El cambio en el peso del cuerpo se debe al cambio en la aceleración de la fuerza de gravedad. Con estos gravímetros se logra calcular g de manera rápida y precisa.

La unidad de aceleración se llama "gal", en honor a Galileo, y se define:

$$1 \text{ gal} = 1 \text{ cm/seg}^2.$$

Se mencionó antes cómo Bouguer en la expedición al ecuador descubrió que la plomada era desviada en presencia de una gran masa. Durante la segunda mitad del siglo XIX, exploradores ingleses que realizaban mediciones astrónomo-geodésicas en la India, observaron que la desviación de la plomada era menor de la esperada ante la presencia de los montes Himalaya. Era de esperarse que en la India, en cuyo norte se encuentran las masas del potente y extenso sistema de los montes Himalaya, y en cuyo sur tiene al océano Indico con grandes profundidades, fuesen grandes las desviaciones de la línea vertical; sin embargo esto no sucedió.

Alrededor de 1865 y casi simultáneamente, el astrónomo Sir George Airy y el geodesta John Pratt, ambos ingleses, dieron una explicación del fenómeno que sentó las bases de la "isostacia" (equilibrio en griego). Airy propuso que las enormemente pesadas montañas no están soportadas por debajo por una corteza fuertemente rígida sino que "flotan en un manto de roca mas densa". En otras palabras, el exceso de masa de la montaña sobre el nivel del mar, es compensado

por una deficiencia de masa en la base subyacente. Esta base o raíz, proporciona el soporte flotante, al igual que sucede con todo cuerpo que flota como sería un bote o un témpano.

La plomada siente tanto el efecto del exceso de masa por arriba causada por la montaña así como la deficiencia de masa en el interior de la corteza, de aquí que la desviación de la vertical sea menor de la esperada.

Este incidente en los Himalaya no sólo dio lugar al concepto de "isostacia" sino que introdujo el uso de las exploraciones gravimétricas como método para detectar las variaciones de masa en el interior de la corteza a partir de las anomalías observadas.

Como la fuerza de atracción juega un papel preponderante para la determinación de la figura de la Tierra, a continuación se da una introducción al área de las matemáticas que se utiliza como herramienta para calcularla.

TEORIA DEL POTENCIAL

Consideremos dos puntos materiales $A(a, b, c)$, $B(x, y, z)$, A con masa m y B con masa unitaria. La ley de gravitación universal de Newton, nos dice según (2), que

$$F = G \frac{m}{r^2} \quad , \quad r^2 = (a - x)^2 + (b - y)^2 + (c - z)^2$$

Tomenos un sistema cartesiano de coordenadas (X, Y, Z) con el origen en el punto $B(x, y, z)$ y denotemos por α, β, γ los ángulos que forma el vector BA con los ejes coordenados. Sean F_x, F_y, F_z las componentes de la fuerza F sobre los ejes coordenados. Refiriéndonos a la Fig. 2, se tiene

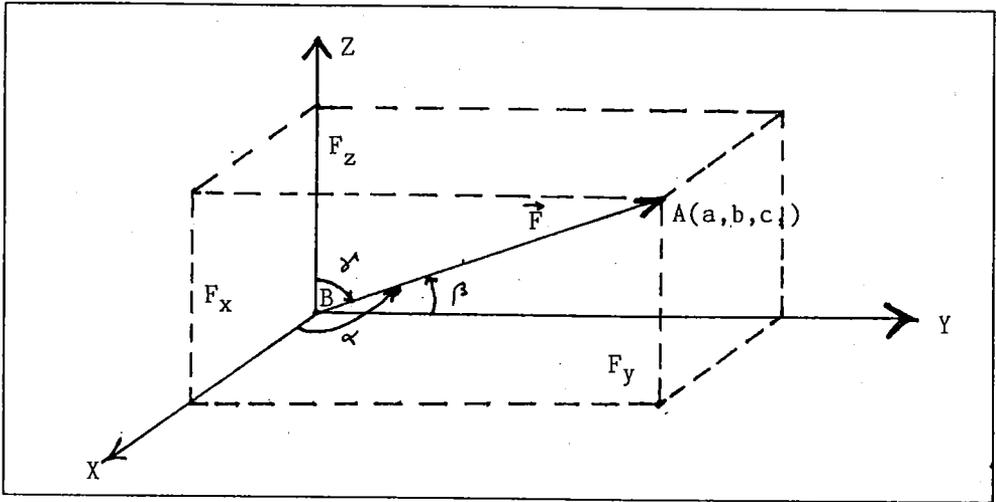


Fig. 2

$$F_x = F \cos \alpha, \quad F_y = F \cos \beta, \quad F_z = F \cos \gamma.$$

Además es claro que

$$\cos \alpha = \frac{a-x}{r}, \quad \cos \beta = \frac{b-y}{r}, \quad \cos \gamma = \frac{c-z}{r},$$

de manera que

$$\begin{aligned} F_x &= F \frac{a-x}{r} = Gm \frac{a-x}{r^3}, \\ F_y &= Gm \frac{b-y}{r^3}, \\ F_z &= Gm \frac{c-z}{r^3}. \end{aligned} \tag{7}$$

El vector \vec{F} que representa a la fuerza de atracción mutua entre A y B está pues definido por

$$\vec{F} = Gm \left(\frac{a-x}{r^3}, \frac{b-y}{r^3}, \frac{c-z}{r^3} \right).$$

Sin embargo observamos que si definimos la función escalar

$$V(x, y, z) = \frac{Gm}{r} \quad (8)$$

se cumple

$$\frac{\partial V}{\partial x} = F_x, \quad \frac{\partial V}{\partial y} = F_y, \quad \frac{\partial V}{\partial z} = F_z,$$

o bien en forma más compacta

$$\vec{F} = Gm\nabla V. \quad (9)$$

La función V , cuyas derivadas parciales respecto a las coordenadas rectangulares del punto de atracción son iguales a las respectivas componentes de la fuerza de atracción, se denomina "*función potencial*", o simplemente "*potencial de atracción*".

Consideremos ahora un cuerpo Ω cualquiera que atrae al punto $A(x, y, z)$. Como la fuerza de atracción es el conjunto de fuerzas de atracción de las partículas elementales que forman al cuerpo, subdividimos Ω en volúmenes elementales $d\Omega = dadbdc$ que contienen en su interior al punto variable $M(a, b, c)$. Sea ρ la densidad del punto M en la unidad de masa, o sea $\rho = dm/d\Omega$. Por tanto

$$dm = \rho da db dc.$$

Denotemos por dF_x, dF_y, dF_z las proyecciones de la fuerza de atracción de las masas elementales situadas en los puntos M y A . Utilizando (7) tenemos

$$dF_x = G \frac{(a-x)}{r^3} dm = G \frac{\rho(a-x)}{r^3} da db dc,$$

y expresiones similares para dF_y y dF_z .

Sumando la acción de todas las partículas del cuerpo Ω

$$\begin{aligned}
 F_x &= G \int \int_{\Omega} \int \rho(a, b, c) \frac{a-x}{r^3} da db dc, \\
 F_y &= G \int \int_{\Omega} \int \rho(a, b, c) \frac{b-y}{r^3} da db dc, \\
 F_z &= G \int \int_{\Omega} \int \rho(a, b, c) \frac{c-z}{r^3} da db dc.
 \end{aligned}
 \tag{10}$$

Luego el potencial V del cuerpo Ω es aquella función cuyo gradiente ∇V , tiene como componentes las expresiones dadas por (10), es decir $\nabla V = \vec{F}$.

De modo que esta $V(x, y, z)$ está definida por

$$V(x, y, z) = G \int \int_{\Omega} \int \frac{dm}{r} = G \int \int_{\Omega} \int \frac{\rho(a, b, c) da db dc}{[(x-a)^2 + (y-b)^2 + (z-c)^2]^{1/2}}
 \tag{11}$$

Si pensamos que el cuerpo Ω representa a la Tierra, (11) es el potencial de atracción de la Tierra en un punto A externo a ella. El potencial V es continuo en todo el espacio y cuando A está muy alejado de Ω , es decir $r \sim \infty$, V se anula como $1/r$. Lo que ocurre es que si r es muy grande, el cuerpo Ω actúa sobre A como si fuera un punto masa para el cual el potencial es de la forma $1/r$. Consecuentemente en la mecánica celeste los planetas se toman como puntos masa.

Las derivadas parciales de V , que corresponden a las componentes de la fuerza dadas por las ecuaciones (10), son continuas en todo el espacio.

Además

$$\frac{\partial^2 V}{\partial x^2} = \frac{\partial F_x}{\partial x} = G \int \int_{\Omega} \int \frac{\partial}{\partial x} \left[\frac{a-x}{r^3} \right] dm = -G \int \int_{\Omega} \int \left[\frac{1}{r^3} - \frac{3(a-x)^2}{r^5} \right] dm
 \tag{12}$$

y expresiones similares para $\partial^2 V / \partial y^2$, $\partial^2 V / \partial z^2$. Sumando estas tres derivadas

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \Delta V = 0.
 \tag{13}$$

La ecuación (13) se conoce como ecuación de Laplace y a las funciones con primeras y segundas derivadas continuas en cierta región donde además satisfacen la ecuación de Laplace, se denominan armónicas en tal región.

Sin embargo la ecuación (13) se satisface sólo para puntos $A(x, y, z)$ exteriores a Ω para los cuales se garantiza que $r \neq 0, a - x \neq 0, b - y \neq 0, c - z \neq 0$, pues de lo contrario en (12) tendríamos la expresión indefinida $0/0$.

Veamos qué ocurre si A está dentro de Ω . Tomemos una esfera S_ϵ de radio ϵ , que contenga al punto A en su interior, con ϵ lo suficientemente pequeño como para que la esfera esté totalmente contenida en Ω y la sustancia dentro de esta esfera pueda considerarse homogénea. Sea Ω_1 la región formada por Ω al remover la esfera de radio ϵ y sean V_1 el potencial de atracción de Ω_1 en el punto A y V_2 el potencial de atracción de la esfera en el punto A . Así, el potencial de Ω en A resulta $V = V_1 + V_2$.

Como A es externo a Ω_1 , la ecuación de Laplace se cumple es decir

$$\Delta V_1 = 0 \quad \text{en} \quad A. \quad (14)$$

El potencial V_2 de la esfera, como puede probarse, es

$$V_2 = \frac{2}{3}\pi G\rho(3\epsilon^2 - \delta^2), \quad \delta^2 = (a-x)^2 + (b-y)^2 + (c-z)^2 \quad (15)$$

ρ es la densidad dentro de la esfera. Derivando V_2 directamente se obtiene

$$\frac{\partial V_2}{\partial x} = -\frac{4}{3}\pi G\rho\delta \frac{\partial \delta}{\partial x} = \frac{4}{3}\pi G\rho(a-x)$$

y fórmulas análogas para las otras derivadas. Por tanto

$$\begin{aligned} \frac{\partial^2 V_2}{\partial x^2} &= -\frac{4}{3}\pi G\rho = \frac{\partial^2 V_2}{\partial y^2} = \frac{\partial^2 V_2}{\partial z^2} \\ \Rightarrow \Delta V_2 &= -4\pi G\rho. \end{aligned} \quad (16)$$

Según (14) y (16), el potencial de atracción V satisface en el punto A

$$\Delta V_1 + \Delta V_2 = \Delta V = -4\pi G\rho, \quad (17)$$

ρ se interpreta como la densidad del cuerpo en un pequeño entorno alrededor del punto atractor A . La ecuación (17) se conoce como la ecuación de Poisson pues fue él quien descubrió que para puntos interiores al cuerpo no se cumplía la ecuación de Laplace. Laplace tomó por un hecho que la ecuación (13) era cierta para todo punto sin importar su posición.

La ecuación de Laplace es en realidad un caso particular de la ecuación de Poisson, pues si en el punto analizado no hay masa atrayente, $\rho = 0$. Esto es lo que ocurre cuando el punto A está fuera de Ω .

En conclusión hemos visto que el potencial de atracción V es una función armónica en el exterior de Ω , mientras que en su interior, V satisface la ecuación de Poisson

$$\Delta V(A) = \begin{cases} 0, & \text{si } A \notin \Omega \\ -4\pi G\rho, & \text{si } A \in \Omega \end{cases}$$

Las ecuaciones de Laplace y Poisson son fundamentales en la teoría del potencial. La determinación de la figura de la Tierra está íntimamente relacionada con la búsqueda del potencial externo de atracción. Este se obtiene de integrar la ecuación de Laplace junto con condiciones complementarias determinadas por el problema particular que se estudie. El potencial gravitacional es armónico en todos los puntos donde no hay masas atrayentes, si descontamos el efecto de la atmósfera y la fuerza centrífuga, el potencial gravitacional externo de la Tierra es armónico. Esta es la razón por la cual las funciones armónicas son importantes en la geodesia física.

Aunque un poco ficticio, en muchas aplicaciones es necesario calcular el potencial de atracción de una "capa simple", es decir, una superficie cerrada S de grueso cero y densidad $\mu = \frac{dm}{d\sigma}$, donde $d\sigma$ es el elemento de superficie. En analogía con la ecuación (11), en este caso tenemos

$$V(A) = G \int_S \int \frac{dm}{r} = G \int_S \int \frac{\mu}{r} d\sigma. \quad (18)$$

Si $A \in S$, a pesar de que $r = 0$, V es continua.

Un caso particularmente importante de capa simple es la esfera. Supongamos que S es la esfera homogénea con centro en el origen y radio R . Identificamos un punto M sobre S por sus coordenadas (φ, θ) donde φ es la longitud y θ es la colatitud. El elemento de superficie resulta $d\sigma = R^2 \operatorname{sen} \theta d\varphi d\theta$.

Para calcular el potencial (18) se consideran dos puntos A, A_1 , uno fuera y otro dentro de la esfera como muestra la Fig. 3.

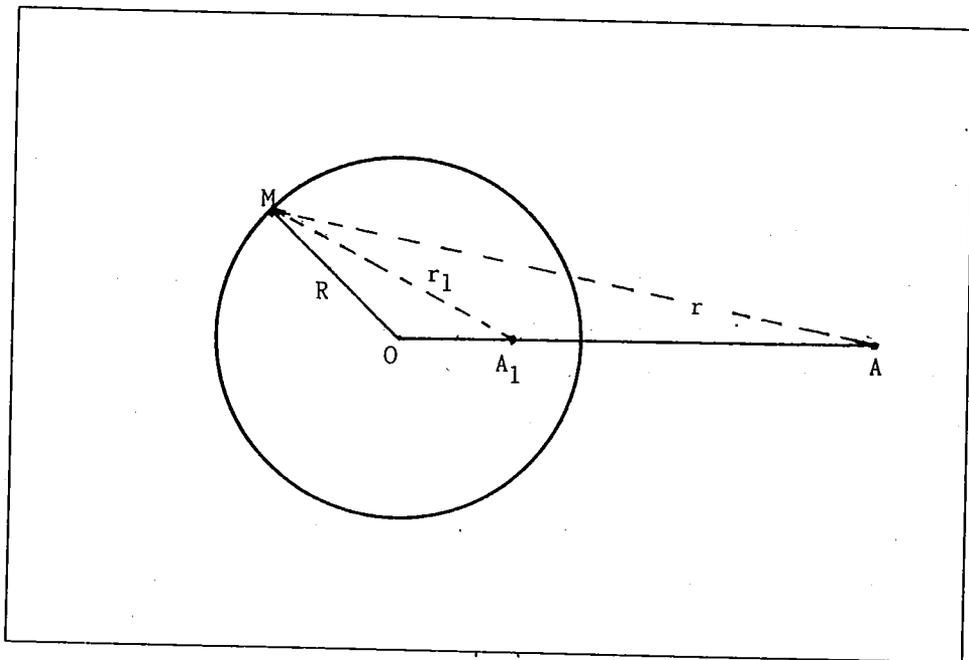


Fig. 3

Sean r, r_1 las distancias MA y MA_1 , respectivamente; ρ, ρ_1 denotarán las distancias OA, OA_1 respectivamente. La ley de los cosenos nos dice que

$$r^2 = R^2 + \rho^2 - 2R\rho \cos \theta,$$

$$r_1^2 = R^2 + \rho_1^2 - 2R\rho_1 \cos \theta.$$

Sustituyendo en (18)

$$V(A) = G\mu \int_0^{2\pi} \int_0^\pi \frac{R^2 \operatorname{sen} \theta d\theta d\varphi}{\sqrt{R^2 + \rho^2 - 2R\rho \cos \theta}} = 2\pi G\mu \int_0^\pi \frac{R^2 \operatorname{sen} \theta d\theta}{\sqrt{R^2 + \rho^2 - 2R\rho \cos \theta}}$$

Para evaluar $V(A_1)$ se cambia ρ por ρ_1 en la fórmula anterior.

Haciendo el cambio de variable $u = R^2 + \rho^2 - 2R\rho \cos \theta$, la integral es

$$\begin{aligned} V(A) &= 2\pi G\mu \left(\frac{2R^2}{\rho} \right) = 4\pi G\mu \frac{R^2}{\rho}, \\ V(A_1) &= 2\pi G\mu (2R) = 4\pi G\mu R. \end{aligned} \quad (19)$$

La masa de la esfera S está dada por $M = 4\pi R^2\mu$, de manera que para puntos exteriores a S el potencial es

$$V = \frac{GM}{\rho}. \quad (20)$$

Comparando (20) con (8) concluimos que la capa esférica homogénea se comporta igual que el punto masa, en el sentido que ambos originan en el espacio exterior el mismo potencial, si consideramos que la masa M está concentrada en el centro de la esfera. Puesto que ρ es la distancia del punto exterior A al centro de la esfera O , la atracción que la capa ejerce sobre A está dirigida hacia el centro de S .

Para el punto interior A_1 el potencial es $V = GM/R$ que es constante y no depende de la posición de A_1 . Así la fuerza de atracción de la capa S sobre un punto en su interior es cero.

En particular la fórmula (20) dice que todas las esferas concéntricas homogéneas de igual masa total M , sin importar su tamaño, generan el mismo potencial. Esto fue descubierto por Newton. En el lenguaje moderno, estas esferas se denominan "cuerpos gravi-equivalentes" ya que originan iguales campos gravitacionales externos.

Si planteamos ahora el problema inverso, es decir, dado el campo gravitacional externo hallar el cuerpo que lo produce, es evidente con este ejemplo que el problema está mal planteado ya que tiene múltiples soluciones. Para el caso de las esferas homogéneas, se obtendrá solución única al problema inverso si se añade otra condición: la esfera homogénea debe tener la mayor densidad posible.

Puesto que $\vec{g} = \vec{F} + \vec{Q}$, el potencial de la fuerza de gravedad es igual a la suma del potencial de atracción, el cual ya calculamos, y del potencial de la fuerza centrífuga Q . Vimos que si la masa $m = 1$, $Q = \omega^2 \rho$. Como la fuerza centrífuga es perpendicular al eje de rotación, haciendo coincidir el eje z con el eje de rotación de la Tierra, la fuerza Q resulta paralela al plano (x, y) y su componente en z es cero.

Sea M un punto en la superficie terrestre, tómesese un plano paralelo al plano (x, y) que pasa por M . Refiriéndonos a la Fig. 4, las componentes de la fuerza \vec{Q} en los ejes x, y son

$$Q_x = Q \cos \alpha = \omega^2 \rho \cos \alpha$$

$$Q_y = Q \cos \beta = \omega^2 \rho \cos \beta$$

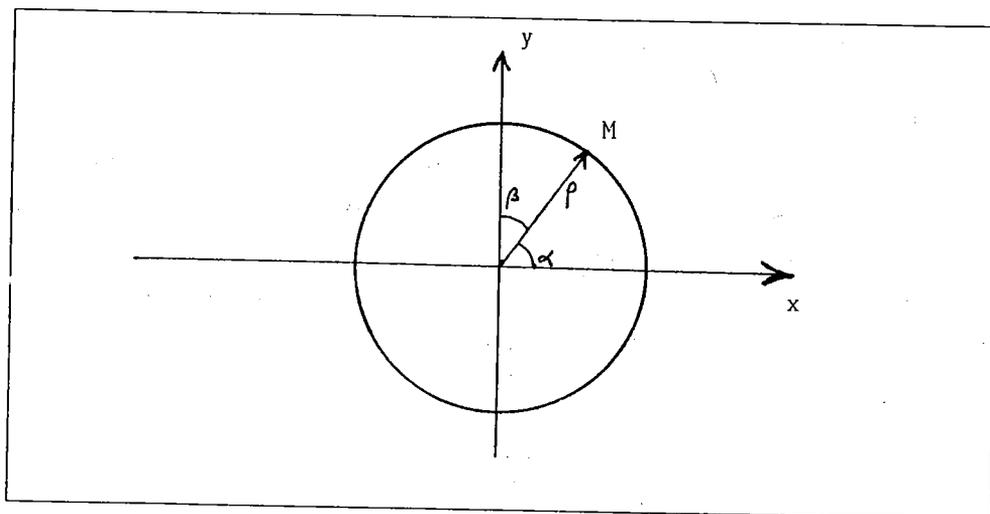


Fig. 4

pero como

$$x = \rho \cos \alpha$$

$$y = \rho \cos \beta$$

entonces

$$Q_x = \omega^2 x,$$

$$Q_y = \omega^2 y,$$

$$Q_z = 0.$$

Estas componentes son las derivadas parciales de la función escalar

$$U = \frac{\omega^2}{2}(x^2 + y^2) \rightarrow \nabla U = (\omega^2 x, \omega^2 y, 0).$$

Luego la función

$$W = U + V = \omega^2(x^2 + y^2) + G \int \frac{dm}{r}$$

es el potencial de la fuerza de gravedad.

ELIPSOIDE DE REFERENCIA

Resolver los problemas teóricos y prácticos de la geodesia significa, de alguna manera, escoger una superficie que bajo condiciones favorables se aproxime bien a la forma de la Tierra, pero que además, matemáticamente pueda expresarse en forma cerrada de modo que tanto el potencial como la gravedad puedan ser calculados. A esta superficie se le denomina la forma "normal" de la Tierra; al correspondiente potencial se le denomina "potencial normal" y a la gravedad "gravedad normal".

El campo gravitacional de un elipsoide es de gran importancia práctica ya que matemáticamente es fácil de manejar. De aquí que se define el "potencial normal" como una superficie de nivel del potencial U que es solución del problema de Dirichlet para el elipsoide de revolución S definido por

$$S = \{(x, y, z) \mid \frac{x^2 + y^2}{a^2} + \frac{z^2}{b^2} = 1\}.$$

Es decir U satisface

$$\Delta U = 0 \quad \text{en el exterior de } S,$$

$$U = U_0 \quad \text{en } S, \quad U \sim 0 \quad \text{en el infinito.}$$

El equipotencial $U(x, y, z) = U_0 = \text{cte.}$ es el "*elipsoide de referencia*", donde U_0 es una constante que depende de la masa M y la velocidad angular ω de la Tierra, así como de las constantes a, b del elipsoide. Hablar de $U(x, y, z) = U_0$ ó de S es equivalente ya que por definición los puntos (x, y, z) de S son los que satisfacen $U(x, y, z) = U_0$. La "*gravedad normal*" es $\vec{\gamma} = \nabla U$. La desviación del campo gravitacional real de Tierra de este campo elipsoidal *normal* es tan pequeña que por ello se considera al potencial gravitacional W como la suma del "*potencial normal*" U más el "*potencial de perturbación*" T : $W = U + T$.

El potencial normal queda totalmente determinado si se conocen M, W, a, b , de manera que el problema geodésico se reduce a la determinación del "*potencial de perturbación*". Sin entrar en mayor detalle hay que mencionar que la razón de escoger a U_0 dependiente de M y W es para lograr que $W - U = T$ sea una magnitud de segundo orden de pequeñez.

El trabajo de Störkés de 1849 fue crucial para la determinación del potencial normal y sigue utilizándose aún en muchos problemas de la geodesia.

GEOIDE

Las superficies de nivel

$$W = C = \text{constante,}$$

satisfacen que la fuerza de gravedad está dirigida en la dirección normal a la superficie y que las componentes de la fuerza de gravedad sobre la tangente a la superficie en cualquier punto, son iguales a cero. Para ver esto sea $W(x, y, z) = W_0$ una superficie de nivel. Derivando W

$$dW = \frac{\partial W}{\partial x} dx + \frac{\partial W}{\partial y} dy + \frac{\partial W}{\partial z} dz = \nabla W \cdot d\vec{g} = \vec{g} \cdot d\vec{x}$$

donde $dx = (dx, dy, dz)$. Si dx se toma a lo largo de $W = W_0$ entonces $dW = 0$ y por tanto $\vec{g} \cdot dx = 0$ luego \vec{g} es perpendicular a $W = W_0$.

La superficie de un líquido en estado de reposo es, idealizadamente, parte de una superficie de nivel y la fuerza de gravedad es perpendicular a ella.

Gauss propuso utilizar la superficie de los océanos, considerada como superficie de nivel, como “la figura matemática de la Tierra”. Pero no fue sino hasta 1873 en Göttingen que el físico alemán Listing introdujo la noción del GEOIDE como: “la superficie de nivel del potencial gravitacional que coincide con la superficie imperturbable de los océanos, continuada por debajo de los continentes, donde será calculada con el nivel del agua en una red de canales conectados con los océanos, siempre y cuando la masa por arriba de estos canales sea constante”.

Listing introdujo este concepto como sustituto de la figura de la Tierra, por esto hasta los últimos tiempos, el problema de determinar la figura de la Tierra se formulaba como el estudio del geoide. En la actualidad se sigue utilizando el geoide para los cálculos geodésicos, sin embargo, esta superficie, que es desconocida debajo de los continentes, es irregular y no se pueden dar simples consideraciones geométricas sobre ella. Las superficies de nivel que se encuentran fuera de la Tierra son analíticas porque el potencial W es analítico allá, pero el geoide se encuentra en parte dentro de la Tierra y por tanto no es analítico. El geoide es continuo y suave pero su curvatura cambia a saltos en las orillas de los mares y océanos y en los lugares donde atraviesan rocas de diferente densidad.

Las dificultades que presenta la determinación del geoide hacen que se utilice en geodesia el llamado “elipsoide de referencia”, el cual es un elipsoide de revolución cuyo centro coincide con el centro de masa de la Tierra y con uno de sus ejes igual al eje de rotación terrestre. Este elipsoide de referencia está determinado por cuatro parámetros

a , semieje mayor; $f = \frac{a-b}{a}$, (b semieje menor) achatamiento;

γ_a , gravedad ecuatorial ; ω , velocidad angular.

En 1924 en la asamblea de la Asociación Internacional de Geodesia que tuvo lugar en Madrid, se definió el "elipsoide internacional" dado por

$$a = 6,378,388 \text{ metros}, \quad f = 1/297,$$

$$\gamma_a = 978.049 \text{ gal}, \quad \omega = (0.72921151) \cdot 10^{-4} \text{ seg}^{-1}.$$

Diversos elipsoides de referencia se consideran en geodesia y astronomía, su elección ha dependido de necesidades específicas y a veces del lugar mismo en que se está trabajando.

Una vez que se tiene el elipsoide de referencia y el geoide, se definen dos coordenadas particulares para cualquier punto sobre la superficie terrestre. Si P es el punto en la superficie, se define H , la "altura ortométrica", como la distancia de P al geoide sobre la trayectoria ortogonal a las superficies de nivel $W = C$. La distancia N entre el geoide y el elipsoide de referencia sobre la normal al elipsoide que pasa por P , se denomina "altura del geoide" u "ondulación del geoide". Ver la Fig. 5.

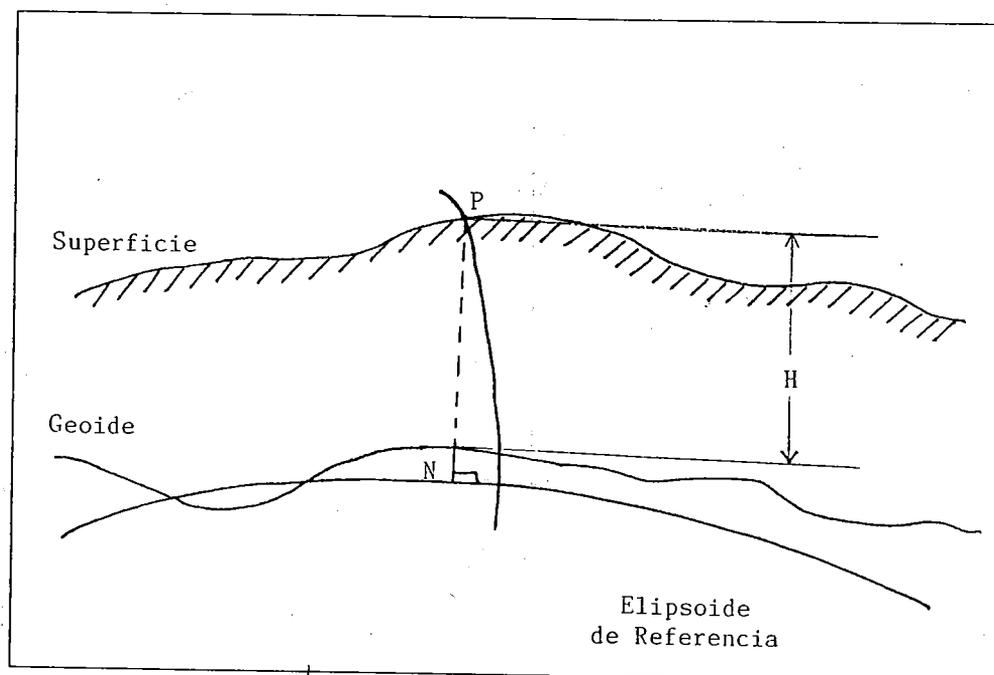


Fig. 5

En realidad N mide la desviación del geoide del elipsoide de revolución. Una manera de determinar el geoide consiste en medir esa desviación N respecto a un elipsoide de referencia dado. Mediante datos gravimétricos y observaciones de satélites, se han elaborado mapas de la altura N del geoide que cubren todo el planeta. En algunas zonas como el sur de la India, esta altura alcanza hasta los 100 metros.

Considérese ahora un punto P sobre el geoide el cual se proyecta sobre un punto Q en el elipsoide de referencia en la dirección de la normal a éste.

Sea g_P el vector gravedad en P y sea γ_Q el vector gravedad normal en Q . Véase la Fig. 6. Se define el "vector anómalo de gravedad" Δg como

$$\Delta g = g_P - \gamma_Q.$$

La diferencia de magnitudes $\Delta g = g_P - \gamma_Q$ es lo que se conoce como "anomalía de la gravedad" y su diferencia en dirección α como "desviación de la vertical".

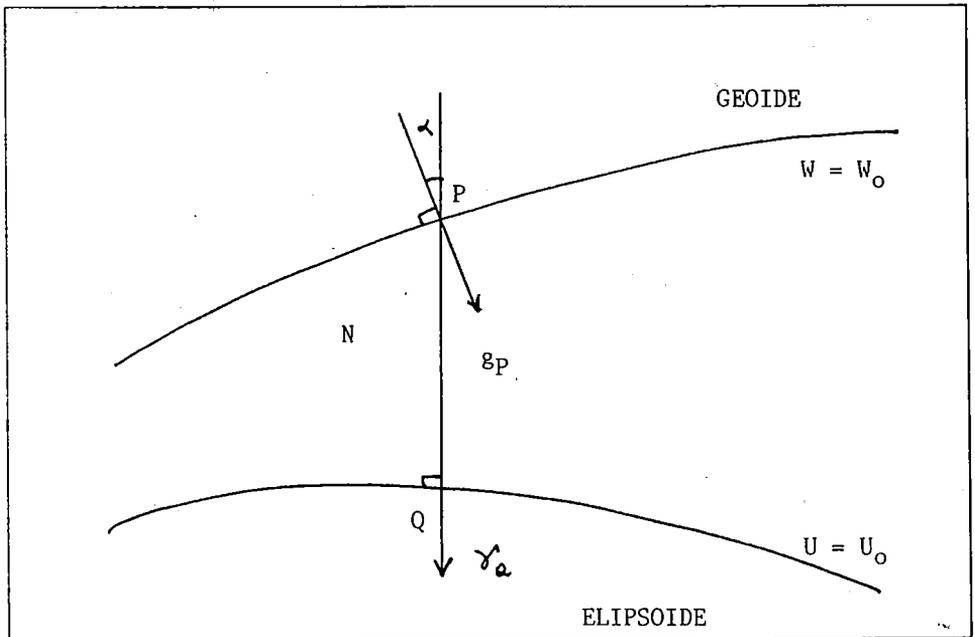


Fig. 6

En los trabajos exploratorios, los datos que se recolectan son Δg y α pues proporcionan mucha información.

Sería interminable continuar describiendo los usos del geoide y del elipsoide de referencia en la geodesia. Para el objeto de este escrito, se han mencionado estas superficies pues representan la concepción moderna de la figura de la Tierra.

ALGUNAS APLICACIONES

Como ya hemos mencionado, si la Tierra fuera una esfera homogénea, la intensidad del campo gravitacional sería constante. Debido a las inhomogeneidades del terreno y a la distribución de las masas de distinta densidad en el subsuelo, la intensidad de la gravedad medida en la superficie terrestre, sufre pequeñas desviaciones de su valor promedio. Aunque pequeñas, estas desviaciones son registradas por los gravímetros. La intensidad promedio del campo gravitacional de la Tierra es de 980 gals. Una anomalía típica tiene una intensidad de 0.5 a 10 mgals y la exactitud de las lecturas gravimétricas es de 0.02 y 0.05 mgals. Estos datos gravimétricos se utilizan en las exploraciones para hallar depósitos minerales.

Cuando el magma derretido se introduce entre las rocas del subsuelo, se forman en la corteza unos cuerpos geológicos, llamados "intrusiones", cuya densidad es notablemente superior a la densidad de los sedimentos de la superficie terrestre. En general, los depósitos minerales se congregan en forma de "intrusiones" de la base de la corteza, mientras que los yacimientos petrolíficos corresponden más bien a cúspides de la base. A fin de distinguir entre una forma u otra del subsuelo, se efectúan mediciones de la gravedad en la superficie a lo largo del terreno y se grafica la componente vertical de las anomalías respecto a la densidad promedio. La interpretación de estas anomalías permite discriminar entre las distintas formas del subsuelo.

Es común que ciertas estructuras geológicas del subsuelo se extiendan a lo largo de cierta dirección, de manera que basta hacer la medición de la gravedad en

el curso de una dirección perpendicular a la dirección de la formación geológica. Esto corresponde a hacer un corte transversal del terreno y considerar tan solo dos dimensiones puesto que la estructura se repite paralelamente en la dirección de la tercera dimensión.

En las figuras siguientes se muestran dos cortes transversales de estructuras geológicas típicas de la corteza terrestre así como las gráficas de las correspondientes anomalías de la gravedad.

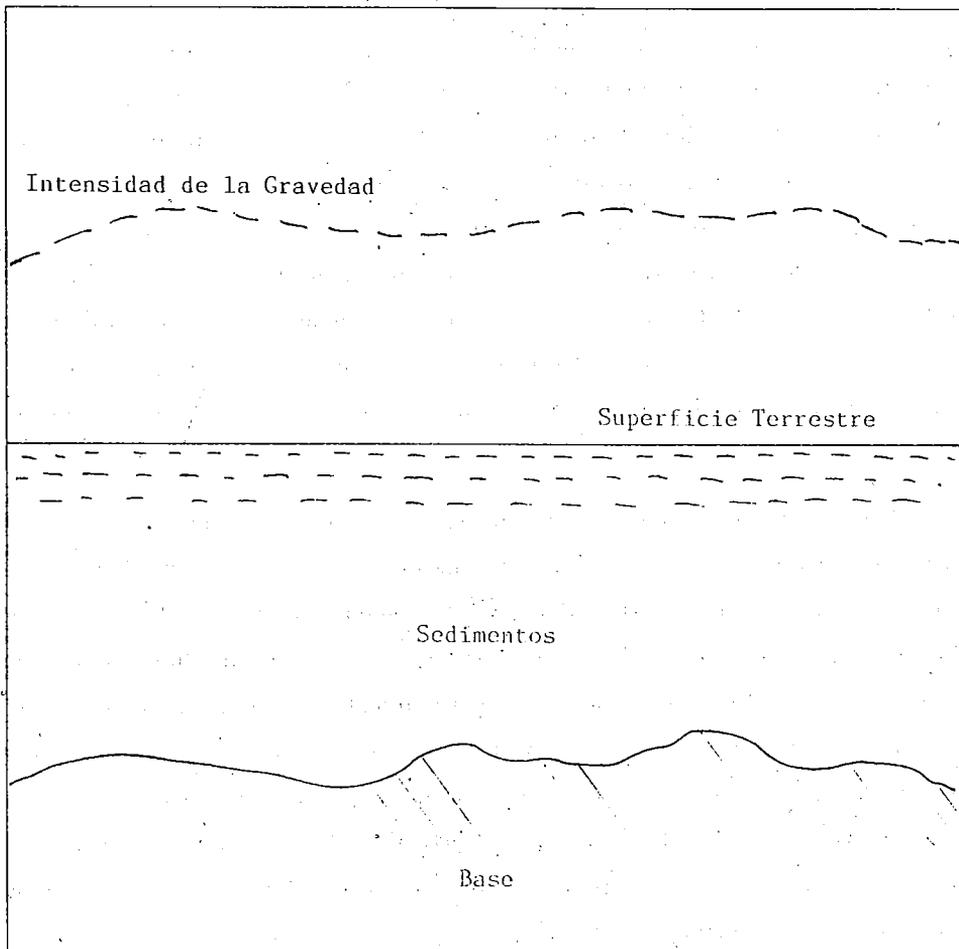


Fig. 7

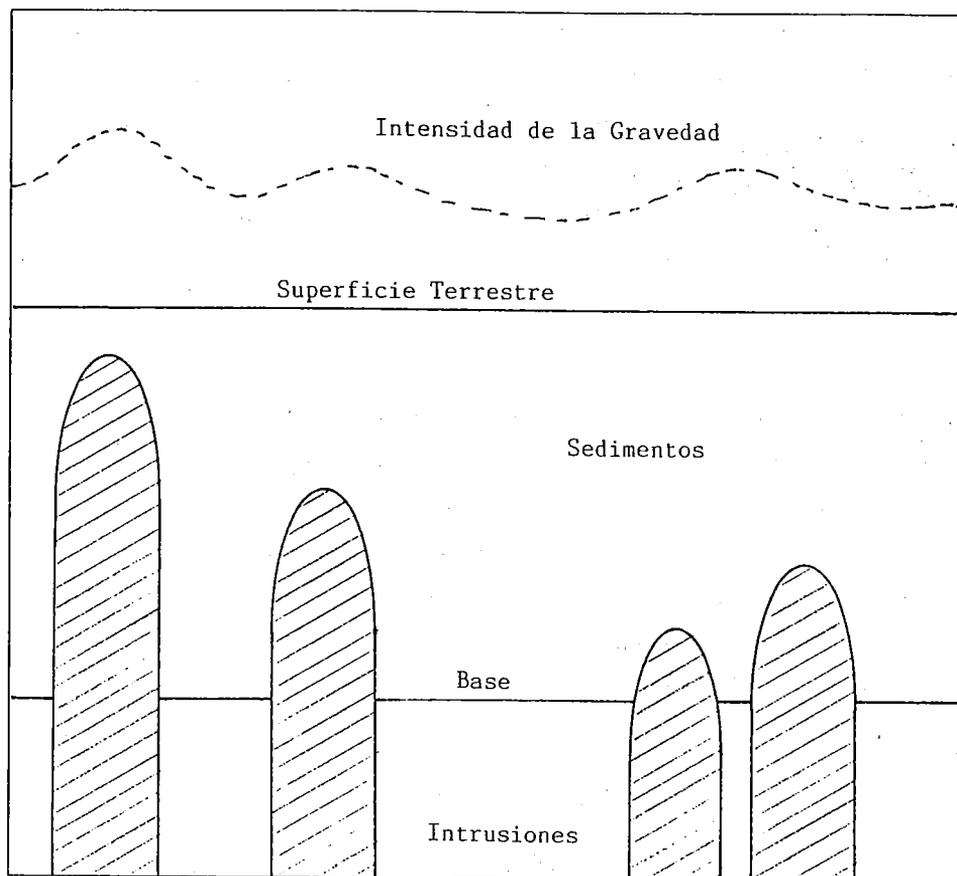


Fig. 8

La Fig. 7 es frecuentemente asociada a yacimientos petrolíferos y la Fig. 8 a depósitos minerales.

Consideremos el caso de la Fig. 8. Se quiere determinar la posición de las intrusiones de la base. Cuando las intrusiones distan entre sí más de lo que distan a la superficie, la gráfica de la intensidad muestra dos máximos consecutivos, de modo que las intrusiones se consideran aisladas. Sin embargo, cuando las intrusiones están muy cercanas entre sí, se produce un solo máximo en la gráfica de la intensidad. Estas interpretaciones de los datos gravimétricos constituyen el primer paso a seguir en la búsqueda de depósitos minerales. El paso más

importante es perforar pozos exploratorios y analizar los datos, la cuestión es pues dónde perforar. Dado un máximo local en la gráfica de la anomalía, tenemos dos posibles situaciones: o es una sola intrusión la que lo produce o son dos muy pegadas. Si decidimos que es una sola, el sitio natural para efectuar la perforación es precisamente en el máximo. Ahora bien, si resulta que las intrusiones son como las del lado derecho de la Fig. 8, entonces el agujero pasará por en medio de ambas y nuestra perforación será inútil. En repetidas ocasiones se ha presentado esta situación en las exploraciones geológicas.

A fin de evitar este problema, los geofísicos proponen lo siguiente: calcule la anomalía de la gravedad a cierta profundidad de la superficie terrestre. Si el máximo local de la anomalía observada en la superficie se preserva a este nivel, se puede concluir con cierta seguridad que el máximo es causado por una sola intrusión. Pero si el máximo local de la superficie se convierte en dos máximos locales en el nivel interior, entonces se trata de dos cuerpos. De esta forma el sitio de la perforación se puede escoger adecuadamente.

Este problema se puede plantear en términos matemáticos utilizando los resultados de la teoría del potencial de la siguiente forma.

Como ya se mencionó, se consideran únicamente dos dimensiones x, y . Suponemos que el eje x coincide con la superficie terrestre y el eje y negativo se extiende hacia el interior. Supongamos también que todas las masas que generan el campo gravitacional se encuentran por debajo del nivel $y = H < 0$. Sea $u(x, y)$ la componente vertical de la intensidad del campo gravitacional generado por las masas del interior ($y < -H$). Vimos ya que en la región externa a las masas generadoras, el potencial gravitacional satisface la ecuación de Laplace, de manera que

$$\Delta u(x, y) = 0, \quad -\infty < x < \infty, \quad y > -H.$$

Ver Fig. 9.

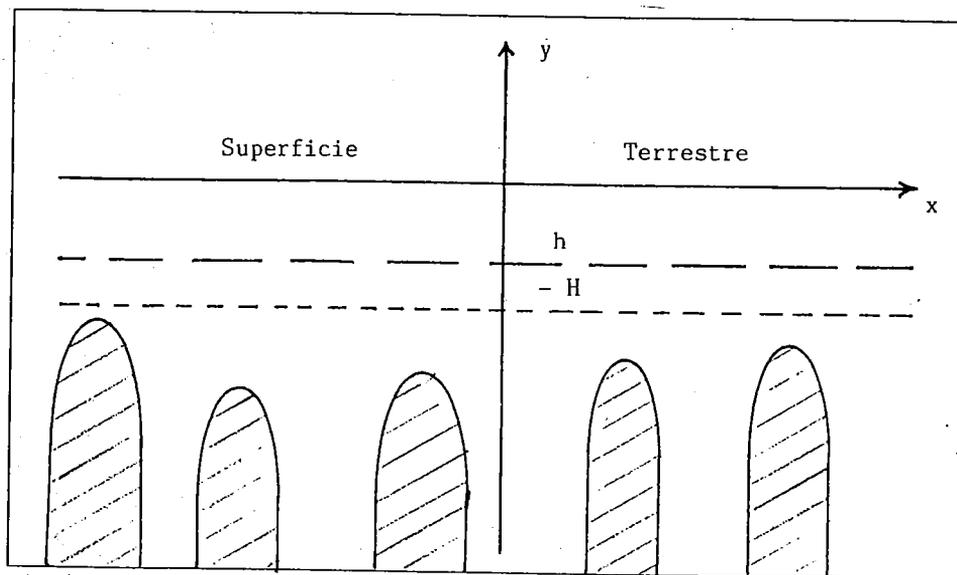


Fig. 9

Como ya se ha medido la gravedad en la superficie entonces

$$u(x, 0) = f(x) \quad , \quad -\infty < x < \infty,$$

para cierta función $f(x)$. Medir la anomalía de la gravedad a cierta profundidad h equivale a determinar el valor de la función $u(x, y)$ para alguna $y = h$, es decir se quiere encontrar $\varphi_h(x)$ tal que

$$u(x, h) = \varphi_h(x) \quad , \quad -\infty < x < \infty \quad , \quad -H < h < 0.$$

El problema de Dirichlet en el semiplano para la teoría del potencial es el siguiente

$$\begin{aligned} \Delta U(x, y) &= 0 \quad , \quad -\infty < x < \infty \quad , \quad y > 0, \\ U(x, 0) &= f(x) \quad , \quad -\infty < x < \infty. \end{aligned} \quad (21)$$

Su solución se puede hallar utilizando transformaciones integrales y se expresa

$$U(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\xi)y}{(x-\xi)^2 + y^2} d\xi \quad , \quad y > 0. \quad (22)$$

Ahora bien, como $u(x, y)$ es armónica en el semiplano $y > -H$ y $h > -H$, se obtiene para φ_h el siguiente problema de Dirichlet

$$\begin{aligned}\Delta u(x, y) &= 0, & -\infty < x < \infty, & y > h, \\ u(x, h) &= \varphi_h(x), & -\infty < x < \infty.\end{aligned}\quad (23)$$

La ecuación de Laplace es invariante bajo traslaciones y en consecuencia se puede obtener la solución al problema (23) para el semiplano $y > h$, realizando una traslación en la ecuación (22). Así

$$u(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\varphi_h(\xi)(y+h)}{(x-\xi)^2 + (y+h)^2} d\xi, \quad y > h \quad (h < 0). \quad (24)$$

La ecuación (24) no nos ha resuelto el problema puesto que φ_h es la incógnita. Sin embargo se tienen los datos en la superficie $y = 0$, por tanto sustituyendo este valor en (24)

$$f(x) = u(x, 0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\varphi_h(\xi)h}{(x-\xi)^2 + h^2} d\xi. \quad (25)$$

Esta es una ecuación integral de primera clase la cual se sabe cómo resolver.

Si en vez de datos de la anomalía de la gravedad se tuvieran datos de la anomalía del campo magnético, el problema se formula análogamente, ya que el potencial magnético también satisface la ecuación de Laplace en regiones externas a las masas que generan el campo.

El problema (23) es un ejemplo de "problema inverso". En un problema inverso se tienen mediciones de ciertos campos o ciertas leyes físicas y se quiere determinar el origen del campo o las características del medio que originan los datos.

Como regla general los problemas inversos son mal planteados en el sentido de que la solución es no única. Un ejemplo de problema inverso mal planteado en geofísica es el siguiente: recuperar el campo gravitacional externo (o el campo magnético externo) de la Tierra dado que se conoce la intensidad del campo en la superficie terrestre. En términos matemáticos este es un problema en teoría del potencial que se plantea como sigue.

Sea $u(x, y, z)$ el potencial gravitacional (magnético) externo de la Tierra. Aproxímese la Tierra por una esfera S_R de radio R . Sea Ω el exterior de S entonces u satisface

$$\begin{aligned}\Delta u(x, y, z) &= 0 \quad \text{en } \Omega, \\ |\nabla u| &= f \quad \text{en } S_R, \\ u &\text{ regular en el infinito.}\end{aligned}\tag{26}$$

La intensidad del campo es precisamente $|\nabla u|$ y es el único dato.

La solución de este problema no es única.

Gauss fue el primero que dió una solución aproximada del problema (26) utilizando la expansión de la solución en serie de funciones armónicas esféricas. En la actualidad estas técnicas se siguen practicando para actualizar el campo magnético, ya que el potencial magnético es aproximado por una suma finita de funciones armónicas esféricas.

Aunque fácil de plantear el problema (26) resulta difícil de estudiar por la no linealidad en la condición de frontera. Utilizando técnicas modernas del análisis y teoría de perturbaciones se han logrado obtener algunos resultados de existencia de la solución y se han logrado calcular ramas de soluciones bifurcadas. Sin embargo aún hay mucho material de estudio en este problema.

Otro problema inverso típico en gravimetría es el calcular la densidad de un material confinado en una región D .

Supongamos que S es una región acotada de la superficie terrestre que limita a un cuerpo D del subsuelo localizado exactamente debajo de S . D es parte del espacio (x, y, z) mientras que S es parte del plano (x, y) . Sea $\rho(x, y, z)$ la desviación de la densidad de su valor promedio en la región D . Sea $u(x, y, z)$ el potencial del campo gravitacional anómalo. Usualmente se mide la componente vertical de la intensidad del campo anómalo es decir, se mide

$$H(x, y) = \frac{\partial}{\partial z} u(x, y, z)|_{z=0}$$

Por lo visto anteriormente el potencial u viene siendo

$$u(x, y, z) = G \int \int_D \int \frac{\xi \rho(\xi, \eta, \zeta) d\xi d\eta d\zeta}{[(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2]^{3/2}}$$

donde G es la constante gravitacional. La función medida $H(x, y)$ resulta

$$H(x, y) = \frac{\partial}{\partial z} u(x, y, 0) = G \int \int_D \int \frac{\xi \rho(\xi, \eta, \zeta) d\xi d\eta d\zeta}{[(x - \xi)^2 + (y - \eta)^2 + \zeta^2]^{3/2}} \quad (27)$$

La ecuación anterior es una ecuación integral de primera clase para la incógnita ρ . Obsérvese en (27) que se quiere determinar la función de tres variables $\rho(x, y, z)$ de la función conocida $H(x, y)$ tan solo de dos variables. De esto resulta que la solución de (27) es no única en la clase de funciones en R^3 con un número finito de derivadas continuas.

RESUMEN Y CONCLUSIONES

La evolución en la concepción de la figura de la Tierra ha sido fascinante. En este recorrido sobre su historia, hemos descrito algunas de las teorías del pasado en las que la mitología a los principios filosóficos dibujaron su figura. Hemos visto también los ingeniosos métodos utilizados para medir su tamaño, desde los corredores profesionales utilizados por Eratóstenes hasta los satélites artificiales de nuestros días.

Vimos cómo el impacto de las teorías de Newton sientan las bases de la geodesia física enriqueciendo los métodos geométricos existentes hasta entonces. La teoría de gravitación universal impone su presencia en la polémica figura e incluso determina los objetivos experimentales a realizar para verificar su forma.

Los conflictos inevitables entre el cartesianismo y el newtonismo, la política hegemónica francesa y los errores de mediciones de grados, dieron lugar a una de las polémicas científicas más apasionadas y aventuradas de la historia aquí descrita. Sus personajes se confunden entre el típico científico de la ilustración,

el soldado y el aventurero. Los proyectos realizados por estos intrépidos newtonianos en escenarios tan exóticos, están plagados de aventuras, contratiempos, accidentes fatales e incluso guerras, pero al mismo tiempo de esa fuerza interna que alimenta la búsqueda del conocimiento.

El desarrollo matemático paralelo a las teorías físicas de Newton, hizo posible establecer formas cada vez más elaboradas que aproximan la figura de la Tierra; desde el esferoide, al elipsoide de referencia, el geoide o el potencial normal, entre otros. Estas son las formas modernas que se usan en geodesia como representativas de la Tierra.

A tres siglos de la aparición de las teorías de Newton y del desarrollo subsecuente de las ciencias, la influencia de la gravitación en innumerables aplicaciones se ha evidenciado cada vez más. Tan solo algunas de ellas se describieron en este trabajo.

En suma, el objetivo de este escrito fue presentar algunas de las cuestiones que trata la geodesia, ciencia que se ocupa de la figura y el campo gravitacional de la Tierra, enfatizando el papel de las matemáticas que la fundamentan.

La intención, más que englobar el mayor número de conceptos posibles, es despertar el interés por estas cuestiones, a través de la descripción de las ideas que se conjuntaron en su desarrollo.

REFERENCIAS

- [1] V. Dragomir, D. Ghitau, et al. *Theory of the Earth's Shape*, Elsevier, Amsterdam - Oxford - New York, 1982.
- [2] G.D. Garland. *Introduction to Geophysics*, W.B. Saunders, Philadelphia, 1979.
- [3] W.A. Heiskanen y H. Moritz. *Physical Geodesy*, W.H. Freeman, San Francisco, 1967.
- [4] F. Hoyle. *Astronomy*, Crescent Books Inc., 1962.
- [5] Tom B. Jones. *The Figure of the Earth*, Coronado Press 1967.
- [6] M.M. Laurent'ev, V.G. Romanov y S.P. Shishat. *Skii. Ill-Posed Problems of Mathematical Physics and Analysis*, AMS Translation of Mathematical Monographs, Vol. 64, 1986.
- [7] F. Press y R. Siever. *Earth*, W.H. Freeman, San Francisco, 1982.
- [8] P.S. Zakatov. *Curso de Geodesia Superior*, Mir, Moscú, 1981.
- [9] D. Zidarov. *Inverse Gravimetric Problem in Geoprospecting and Geodesy*, Elsevier, Amsterdam, 1990.

Movimiento de los Planetas: El Sistema Solar

Ernesto A. Lacomba
Universidad Autónoma Metropolitana
Unidad Iztapalapa
Apdo. Postal 55-534, CP. 09340
México, D.F.

El modelo más sencillo del Sistema Solar es el descrito por las leyes de Kepler. Cada planeta describe una elipse con foco en el Sol, de manera que su radio vector desde el Sol barre áreas iguales en intervalos de tiempo iguales, y su periodo es proporcional a la potencia $3/2$ del semieje mayor de la elipse.

Esto se puede considerar correcto si las masas de los planetas son muy pequeñas con respecto al Sol y despreciando atracción gravitacional mutua, de manera que sólo la atracción del Sol actúa.

Es una visión del Sistema Solar que nos lo presenta como una colección de varios osciladores no lineales desacoplados. El movimiento de cada planeta corresponde a uno de estos osciladores.

La predicción de las posiciones de los planetas resulta elemental y la estabilidad en el sentido de Lagrange es afirmativa, es decir, tenemos ausencia de colisiones o de escapes de cuerpos.

El problema es en realidad más complicado. Las masas de los planetas no son despreciables, hay que considerar la perturbación de cada uno sobre los demás por atracción, etc.

En este trabajo voy a considerar dos aspectos. Uno es el estudio geométrico del problema de Kepler, como modelo simplificado para el movimiento de cada planeta. El otro es una discusión sobre la posibilidad de aparición de caos en el Sistema Solar cuando se consideran las perturbaciones posibles del sistema real.

1 El problema de Kepler.

Procediendo deductivamente, la manera más fácil de obtener la ecuación diferencial para el problema de Kepler es como una reducción de las ecuaciones de Newton para el problema de los dos cuerpos (digamos el Sol y un planeta). Puede probarse que el movimiento tiene lugar siempre en un plano fijo. Usando la coordenada relativa $x \in \mathbb{R}^2$, $x \neq 0$ donde el origen está situado ya sea en el planeta en cuestión o en el Sol, obtenemos la ecuación diferencial de segundo orden

$$\ddot{x} = -\frac{\mu}{|x|^3}x. \quad (1)$$

Aquí $\mu = GM$, donde G es la constante de gravitación universal y M es la suma de las masas de los dos cuerpos. En la terminología moderna, se dice que esta ecuación describe el llamado *problema de Kepler*. Realmente es más rica que lo que describen las leyes de Kepler, pues sus movimientos pueden ser cualquier cónica con foco en el origen o bien movimientos degenerados colineales, como veremos más abajo.

Si introducimos la variable $p = \dot{x} \in \mathbb{R}^2$, podemos escribir (1) como el siguiente sistema de primer orden

$$\begin{aligned} \dot{x} &= p, \\ \dot{p} &= -\frac{\mu}{|x|^3}x. \end{aligned} \quad (2)$$

Si ahora definimos como sigue el Hamiltoniano o *energía total* del sistema

$$H(x, p) = \frac{1}{2}|p|^2 - \frac{\mu}{|x|},$$

se puede verificar que el sistema (2) puede escribirse como las ecuaciones de Hamilton para H , es decir

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p}, \\ \dot{p} &= -\frac{\partial H}{\partial x}, \end{aligned}$$

donde $x = (x_1, x_2)$, $p = (p_1, p_2)$, $(\frac{\partial H}{\partial x_1}, \frac{\partial H}{\partial x_2})$, $\frac{\partial H}{\partial p} = (\frac{\partial H}{\partial p_1}, \frac{\partial H}{\partial p_2})$.

Calculando explícitamente la derivada direccional de H con respecto a t , a lo largo de las soluciones (es decir, sustituyendo \dot{x} y \dot{p} de acuerdo con (2)), obtenemos

$$\begin{aligned} \dot{H} &= p_1\dot{p}_1 + p_2\dot{p}_2 + \frac{\mu}{|x|^3}(x_1\dot{x}_1 + x_2\dot{x}_2) \\ &= p_1(\dot{p}_1 + \frac{\mu}{|x|^3}x_1) + p_2(\dot{p}_2 + \frac{\mu}{|x|^3}x_2) = 0 \end{aligned}$$

es decir que H es una constante de movimiento. Para poder clasificar todas las posibles soluciones de (2), tenemos que considerar la otra constante de movimiento que corresponde a la *segunda ley de Kepler*, es decir, el *momento angular*

$$M(x, p) = x_1p_2 - x_2p_1.$$

Se puede verificar también directamente que

$$\begin{aligned} \dot{M} &= \dot{x}_1 p_2 + x_1 \dot{p}_2 - \dot{x}_2 p_1 - x_2 \dot{p}_1 = \\ &= p_1 p_2 - x_1 \frac{\mu}{|x|^3} x_2 - p_2 p_1 + x_2 \frac{\mu}{|x|^3} x_1 = 0. \end{aligned}$$

Es claro que si $M = 0$, los vectores x y $p = \dot{x}$ son siempre colineales, y el movimiento ocurrirá a lo largo de una recta que pasa por el origen.

Las soluciones posibles son como indica la siguiente tabla. Sean $M = c$, $H = h$ los valores fijos de las constantes de movimiento para la solución considerada

	$h < 0$	$h = 0$	$h > 0$
$c \neq 0$	elipse (1a. ley Kepler)	parábola	hipérbola
$c = 0$	colineal acotada	colineal no acotada	colineal no acotada

De hecho, para $c = 0$ y una energía fija, se obtiene el caso colineal límite de la cónica correspondiente para la misma energía.

Una forma de obtener la *tercera ley de Kepler* para soluciones circulares (elipses con excentricidad nula) es como sigue. Sea $x(t) = a(\cos \omega t, \sin \omega t)$ una tal solución con frecuencia angular $\omega = \frac{2\pi}{T}$, donde T es el periodo correspondiente. Sustituyendo en (1), tenemos

$$-\omega^2 a(\cos \omega t, \sin \omega t) = -\frac{\mu}{a^2}(\cos \omega t, \sin \omega t),$$

así que $a^3 = \mu/\omega^2 = (\mu/4\pi^2)T^2$, como queríamos. Finalmente, la rapidez resulta

$$|p| = |\dot{x}| = a\omega = \sqrt{\frac{\mu}{a}}. \quad (3)$$

Otro caso en que pueden calcularse explícitamente las soluciones es cuando $h = 0$ y $c = 0$. Como el movimiento es colineal, la ecuación $H = h$ se escribe $\frac{1}{2}\dot{x}^2 = \mu/x$ con $x \in \mathbb{R}$, $x \neq 0$. Separando variables e integrando, obtenemos

$$x(t) = K(t - t_0)^{2/3}. \quad (4)$$

Esta dependencia de las posiciones como la potencia $2/3$ del tiempo, corresponde a una expresión más general de la tercera ley de Kepler. Esto depende sólo del hecho de que el término $\mu/|x|$ en H es homogéneo de grado -1 .

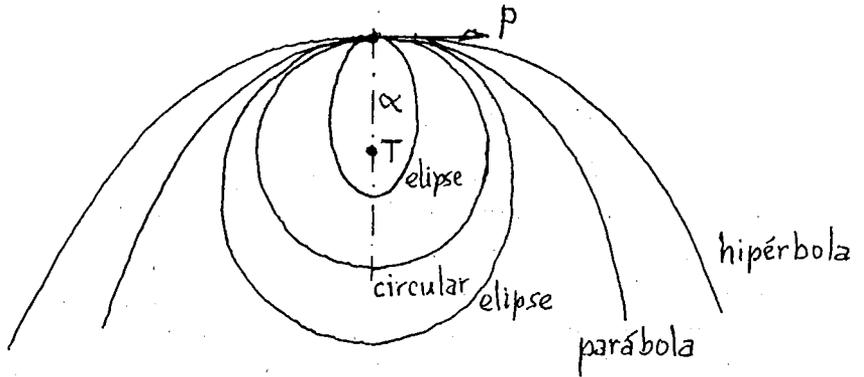


Figura 1: Velocidad de escape de un satélite a una altura α desde el centro de la tierra.

En cualquier otro caso, no es posible dar una descripción general de $x(t)$ en términos de funciones elementales. Por ejemplo, en el caso elíptico se requiere de funciones de Bessel.

Sin embargo, en la sección 2 voy a ilustrar dos maneras en que transformando coordenadas, podemos dar una descripción simple de los movimientos en términos de funciones elementales.

Finalmente, los distintos tipos de curvas cónicas aparecen como soluciones cuando se considera la llamada velocidad de escape de la tierra. Consideremos un cohete o un satélite que se ha logrado colocar a una distancia α , suficientemente grande de la tierra, para que no influya el efecto disipativo de la atmósfera. Su velocidad inicial es perpendicular al radio vector desde la tierra. El valor de la rapidez que produce un movimiento parabólico se obtiene haciendo $H(x, p) = 0$ con $|x| = \alpha$; resolviendo obtenemos $|p_0| = \sqrt{2\mu/\alpha}$. Si $|p| < |p_0|$, el movimiento resultante será de tipo elíptico y si $|p| > |p_0|$, será hiperbólico (ver Fig. 1). Por marcar la transición entre las órbitas acotadas y las no acotadas, p_0 es llamada *velocidad de escape*. Una de las órbitas acotadas es la órbita circular, para la cual se cumple (3): $|p| = \sqrt{\mu/\alpha} = |p_0|/\sqrt{2}$, ya que aquí $\alpha = a$.

2 Soluciones en términos de funciones elementales

Primero que nada, pasando a coordenadas polares (r, θ) en el plano (x_1, x_2) , las ecuaciones de cambio de coordenadas y sus derivadas con respecto al tiempo son

$$\begin{aligned}x_1 &= r \cos \theta, & \dot{x}_1 &= \dot{r} \cos \theta - r \dot{\theta} \sin \theta, \\x_2 &= r \sin \theta, & \dot{x}_2 &= \dot{r} \sin \theta + r \dot{\theta} \cos \theta.\end{aligned}$$

Recordando que $p_1 = \dot{x}_1$, $p_2 = \dot{x}_2$, tenemos

$$|p|^2 = p_1^2 + p_2^2 = \dot{r}^2 + r^2 \dot{\theta}^2, \quad |x|^2 = x_1^2 + x_2^2 = r^2, \quad x_1 p_2 - x_2 p_1 = r^2 \dot{\theta}.$$

Por lo tanto, las ecuaciones para niveles fijos de energía $H = h$ y momento angular $M = c$, toman la siguiente forma

$$\begin{aligned}\frac{1}{2}(\dot{r}^2 + r^2 \dot{\theta}^2) - \frac{\mu}{r} &= h, \\r^2 \dot{\theta} &= c.\end{aligned}\tag{5}$$

Si resolvemos la segunda ecuación para $\dot{\theta} = c/r^2$ y sustituimos en la primera, tenemos

$$\frac{1}{2}\dot{r}^2 + \frac{c^2}{2r^2} - \frac{\mu}{r} = h.\tag{6}$$

Como ya no aparece la variable θ ni su derivada, la ecuación (6) describe la reducción del problema a una dimensión, porque nada más nos describe lo que ocurre en la dirección radial. El potencial $-\mu/r$ debido a la fuerza gravitacional quedó modificado por el término $c^2/2r^2$, correspondiente a la fuerza centrífuga debida al movimiento en la dirección angular.

Notemos que en coordenadas polares es más fácil interpretar la *segunda ley de Kepler*. En efecto, la segunda ecuación de (5) para el momento angular puede escribirse como

$$\frac{c}{2} = \frac{1}{2} r^2 \dot{\theta}.$$

Esta es precisamente la expresión para la rapidez de variación con respecto al tiempo del área barrida por el radio vector, pues $\frac{1}{2} r^2 d\theta$ es el elemento de área.

Ahora sí consideraremos los 2 cambios de coordenadas que nos permitirán estudiar las soluciones de (6). Pueden verse detalles en [3].

A) Método de la hodógrafa.

Recordemos que la ecuación general de una cónica con foco en el origen en coordenadas polares es

$$r = \frac{A}{1 + e \cos \theta}, \quad (7)$$

donde A es una constante y e es la excentricidad ($e = 1$ para parábolas, $e > 1$ para hipérbolas y $0 \leq e < 1$ para elipses). Esto sugiere reemplazar la variable r por $u = 1/r$, lo cual es válido para $r > 0$ (fuera de colisiones). También vamos a reemplazar t por θ como variable independiente, para aprovechar la constante de momento angular. Usando la regla de la cadena,

$$\dot{r} = \frac{dr}{d\theta} \frac{d\theta}{dt} = -\frac{1}{u^2} \frac{du}{d\theta} \dot{\theta}.$$

Pero $\dot{\theta} = \frac{c}{r^2} = cu^2$, así que $\dot{r} = -c \frac{du}{d\theta}$. Sustituyendo en (6), obtenemos finalmente

$$\frac{c^2}{2} \left[\left(\frac{du}{d\theta} \right)^2 + u^2 \right] - \mu u = h. \quad (8)$$

Notemos que aquí debe tenerse que $c \neq 0$, lo cual excluye no sólo la posibilidad de colisiones, sino de soluciones colineales completas, donde θ es constante y no puede ser variable independiente.

La ecuación (8) nos describe circunferencias concéntricas en el plano de coordenadas $u, u' = \frac{du}{d\theta}$, como muestra la Figura 2. Cada una de éstas circunferencias se llama una *hodógrafa* o *curva de velocidades* para el movimiento correspondiente. La razón para este nombre es que comparando (8) con la primera ecuación de (5), tenemos que

$$c^2(u^2 + u'^2) = r^2\dot{\theta}^2 + \dot{r}^2 = |p|^2.$$

Así que la distancia al origen en la Figura 1 corresponde a $|p|/|c|$, para todo punto de la órbita. El ángulo polar en este plano corresponde al ángulo entre la velocidad y el vector de posición.

La región sombreada en esta figura corresponde a órbitas elípticas (acotadas). La frontera corresponde a órbitas parabólicas y el complemento a las hiperbólicas.

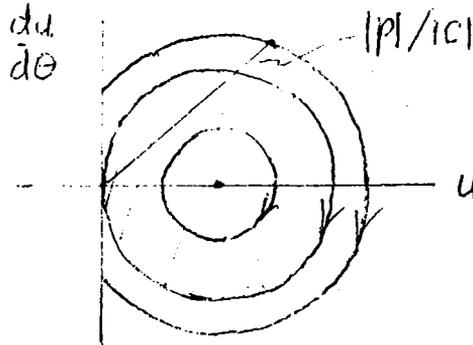


Figura 2: Descripción de la curva de velocidades para el problema de Kepler.

Si derivamos la ecuación (8) con respecto a θ , obtenemos

$$\left(\frac{d^2u}{d\theta^2} + u - \frac{\mu}{c^2}\right) \frac{du}{d\theta} = 0.$$

De aquí, hay dos posibilidades. O bien $\frac{du}{d\theta} \equiv 0$, que se da solo en las soluciones circulares, o si no

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{c^2}. \quad (9)$$

La solución general de (9) es

$$u = K \cos(\theta + \alpha) + \frac{\mu}{c^2},$$

donde K y α son constantes de integración. Finalmente tenemos

$$r = \frac{c^2/\mu}{1 + Kc^2\mu^{-1} \cos(\theta + \alpha)}. \quad (10)$$

Comparando con (7) para $\alpha = 0$ (ángulo de referencia nulo), vemos que $A = c^2/\mu$ y $e = Kc^2/\mu$. De hecho, las soluciones circulares están contenidas en (10), para $K = 0$.

B) Método de la transformación de Sundman.

Multiplicando la ecuación (6) por r^2 , obtenemos

$$\frac{(r\dot{r})^2}{2} + \frac{c^2}{2} = hr^2 + \mu r. \quad (11)$$

Esto sugiere que hagamos una transformación donde $r\dot{r}$ sea la nueva velocidad. Por conveniencia introducimos una constante k que elegiremos abajo de acuerdo con el caso:

$$r\dot{r} = kr'. \quad (12)$$

La ecuación (11) se convierte en

$$\frac{k^2 r'^2}{2} + \frac{c^2}{2} = hr^2 + \mu r. \quad (13)$$

Hemos transformado las coordenadas (r, \dot{r}) a las (r, r') , donde r' se interpreta como la derivada de r con respecto a una nueva variable, que llamaremos E . Si pensamos en (12) como la regla de la cadena, el cambio de derivadas de la variable independiente t a la nueva, está dado por

$$r \frac{d}{dt} = k r' \frac{d}{dE}.$$

Si derivamos (13) con respecto a E como procedimos en A), obtenemos

$$(k^2 r'' - 2hr - \mu)r' = 0.$$

La posibilidad $r' \equiv 0$ de nuevo vale sólo en las órbitas circulares. En cualquier otro caso tenemos que considerar las soluciones de la ecuación diferencial

$$k^2 \frac{d^2 r}{dE^2} - 2hr = \mu.$$

Es análoga a (9), sólo que la aparición de la h hace que tengamos que considerar las soluciones en 3 casos distintos:

a) Si $h < 0$, hagamos $k^2 = -2h$, obteniendo

$$r'' + r = \frac{\mu}{k^2}.$$

Esta ecuación es idéntica a (9). Sin perder generalidad tomamos el desfaseamiento de E como nulo, así que la solución general se puede escribir como

$$r = a(1 - \beta \cos E), \quad (14)$$

donde a y β son constantes que interpretaremos a continuación. Por supuesto que (7) y (14) corresponden a descripciones de la misma órbita

elíptica en términos de 2 parámetros diferentes, θ y E . A θ se le llama la *anomalía verdadera* (porque no es más que el ángulo en coordenadas polares), mientras que E se llama la *anomalía excéntrica*.

Notemos que $\theta = 0$ en la ecuación (7) nos da la mínima distancia del planeta al Sol (llamada *perihelio*), mientras que $\theta = \pi$ nos da la máxima distancia (llamada *afelio*). En la ecuación (14), también $E = 0$ nos da el perihelio y $E = \pi$ nos da el afelio, por aparecer el coseno con signo negativo en el numerador. Igualando los valores mínimo y máximo dado por las 2 ecuaciones, obtenemos el siguiente sistema lineal de 2 ecuaciones en las incógnitas A y β .

$$\begin{aligned} a(1 - \beta) &= \frac{A}{1+e}, \\ a(1 + \beta) &= \frac{A}{1-e}. \end{aligned} \quad (15)$$

Sumándolas obtenemos $2a = \frac{2A}{1-e^2}$, o sea

$$A = a(1 - e^2). \quad (16)$$

Sustituyendo en la primera ecuación de (15), obtenemos

$$\beta = e.$$

Así que

$$r = a(1 - e \cos E), \quad (17)$$

el perihelio y el afelio siendo respectivamente $a(1 - e)$ y $a(1 + e)$. Como la suma de éstos es $2a$, concluimos que a es precisamente el *semieje mayor* de la elipse. Ver la Figura 3, donde O es el foco de atracción y C el centro de la elipse. Si despejamos $r \cos \theta$ de (7) y $a \cos E$ de (17) y restamos, obtenemos

$$a \cos E = ae + r \cos \theta.$$

Esto nos da la interpretación de E como un ángulo para la circunferencia circunscrita a la elipse (de centro C y radio el semieje mayor a). En particular, para la órbita circular se tiene $e = 0$, los puntos O y C coinciden, así que $E = \theta$.

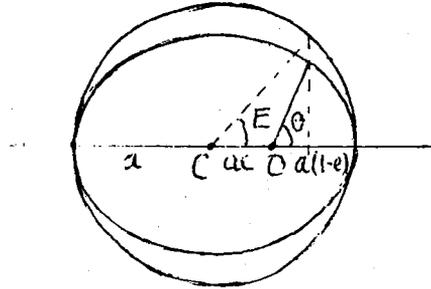


Figura 3: Interpretación de la anomalía excéntrica para una órbita elíptica

Finalmente, t puede obtenerse sustituyendo $r(E)$ de (17), e integrando la expresión

$$t = \int \frac{r(E)}{k} dE.$$

Entonces

$$\omega t = E - e \operatorname{sen} E, \quad (18)$$

si definimos $\omega = k/a = \sqrt{-2h}/a$. El término ωt se llama *anomalía media* (porque cuando $e = 0$, ωt es el ángulo del movimiento circular) y (18) se conoce como la *ecuación de Kepler*. Las ecuaciones (17) y (18) definen implícitamente a r como función de t .

- b) Si $h = 0$, hacemos $k^2 = \mu$, obteniendo simplemente la ecuación diferencial

$$r'' = 1.$$

La solución general $r(E)$ y el valor de $t(E)$ obtenido integrando como en a), resultan

$$\begin{aligned} r &= r_0 + E^2/2 \\ \sqrt{\mu}t &= r_0 E + E^3/6 \end{aligned}$$

- c) Si $h > 0$, hagamos $k^2 = 2h$, obteniendo

$$r'' - r = \frac{\mu}{k^2}.$$

Resolviendo como antes, obtenemos (aunque por supuesto, aquí ω no es una frecuencia angular)

$$\begin{aligned} r &= a(e \operatorname{cosh} E - 1) \\ \omega t &= e \operatorname{senh} E - E. \end{aligned}$$

En cualquiera de los casos a) y c), el valor límite $e = 1$ nos da órbitas colineales que experimentan colisión cuando $E = 0$. Lo mismo ocurre en el caso b) para $r_0 = 0$, donde eliminando E obtenemos de nuevo la ecuación (4).

Observemos que en el método a) tuvimos que excluir explícitamente las soluciones colineales que experimentan colisión. Con el método de Sundman no sólo obtuvimos de nuevo ecuaciones diferenciales lineales análogas, sino que en estas coordenadas tienen sentido las soluciones que pasan por colisión para $E = 0$. Se dice que las ecuaciones resultantes no sólo son lineales, sino que también están regularizadas.

Otra forma común de regularización es mediante el *método de Levi Civita*, que se obtiene simplemente haciendo $r = \chi^2$ en las ecuaciones para el método Sundman. En [4] se puede ver una discusión elemental e histórica de la regularización. Allí se prueba explícitamente que si ocurre una colisión en $t = 0$, siempre se tiene $r(t) = f(t^{1/3})$, donde f es una función analítica de una variable.

3 Caos en el Sistema Solar.

El problema de la estabilidad del Sistema Solar "real" ha recibido consideración desde la época de Laplace. A un nivel rudimentario, tanto él como Lagrange pensaban que era un sistema estable, es decir, no podían ocurrir colisiones ni escapes.

Hasta la década de los sesentas en este siglo fué que se obtuvo un resultado que indicaba que la situación era mucho más complicada de lo que se había supuesto. Un teorema debido a Kolmogorov, Arnold y Moser permitiría concluir que para un conjunto "grande" de posiciones y velocidades iniciales de los planetas no habría colisiones y escapes. De hecho, todo el Sistema Solar se comportaría como un sistema complicado de relojes con frecuencias y semiejes mayores que variarían lentamente. Sin embargo dicho conjunto de condiciones iniciales no es "abierto" es decir, pequeñas variaciones nos hacen salir de él. Para una discusión detallada pero elemental de este resultado, ver Simó [3].

Por otro lado, las posiciones y trayectorias se podían calcular con mucha precisión para decenas y centenas de años, encontrándose siempre estabilidad. Al tener computadoras cada vez más sofisticadas, se pudo extender el intervalo de años en que se hacían los cálculos. Aquí hubo que tomar en cuenta otras correcciones que no eran importantes para intervalos más cortos, como son:

1. Avance de perihelios de los planetas debidos a efectos relativistas.
2. Pérdida de masa del Sol en forma de energías térmica y luminosa.
3. Efecto de mareas, presión de radiación del Sol, etc.
4. Efecto de la galaxia o sistemas próximos.

Con el advenimiento de computadoras cada vez más rápidas y precisas, se comenzaron a hacer cálculos para intervalos más largos de tiempo hacia el futuro y hacia el pasado.

En 1965, dos astrónomos integraron las órbitas de los planetas exteriores (Júpiter, Saturno, Urano, Neptuno y Plutón) por 120,000 años, extendiéndolo posteriormente a 1 millón de años. En 1985 otro grupo hizo cálculos para 5 millones de años. En ambos casos todo parecía correcto, excepto por resonancias de Plutón con el movimiento de Neptuno. Así que hace 7 años parecía que el Sistema Solar permanecería estable durante el tiempo de vida del Sol.

Ahora sabemos que esos cálculos no fueron por intervalos lo suficientemente largos. Los avances más espectaculares en la integración de las ecuaciones de movimiento del Sistema Solar "real", han sido posibles con las supercomputadoras.

J. Wisdom y G. Sussman del MIT usaron una supercomputadora especial, el planetario digital "Orrery" con una velocidad de un tercio de una Cray. Integraron inicialmente las órbitas de los planetas exteriores por 110 millones de años hacia el pasado y hacia el futuro [6]. Sólo Plutón presentaba variaciones grandes. Entonces integraron hasta 900 millones de años, encontrando que Plutón tiene 5 resonancias con Neptuno y además presenta evidencia de caos.

La aparición de caos significa que hay una dependencia muy sensible respecto a las condiciones iniciales. Es decir, si las condiciones iniciales

se cambian ligeramente, la diferencia entre las órbitas crece exponencialmente.

Finalmente, J. Laskar de París, usó un programa TRIP para supercomputadora, modelando los 4 planetas interiores (Mercurio, Venus, Tierra y Marte) durante un período de 200 millones de años hacia el pasado y hacia el futuro, [2]. La integración numérica directa de estos movimientos no era posible en ese momento, debido a los periodos relativamente cortos de dichos planetas. Laskar tuvo que usar nuevas ecuaciones obtenidas aplicando métodos de perturbaciones a las ecuaciones de Newton. Encontró evidencias de caos en cuanto a predicción de las órbitas, incluso en decenas de millones de años. Esto significa simplemente que uno no podría determinar la posición de un planeta en detalle.

En conclusión, se tiene estabilidad (ausencia de colisiones o escapes) para los próximos o pasados cientos de millones de años, pero un movimiento caótico para los planetas interiores y Plutón (imposibilidad de predicción de las posiciones) al cabo de unas decenas de millones de años (véase [1]).

La evidencias más tempranas de caos en el sistema solar fueron observadas hace varios años en el movimiento de algunas lunas (como el satélite Hiperión de Saturno, pequeño y de forma irregular), así como el de algunos asteroides.

La edad aproximada del Sistema Solar es de 5,000 millones de años. Aún si pudieran integrarse numéricamente las ecuaciones hasta esa época, suponiendo además que no hubiera caos, el modelo de Newton ya no serviría. El modelo cosmológico adecuado depende de cómo surgió este sistema.

Bibliografía

- [1] K. Hartley, *Solar system chaos*, Astronomy, May 1990, 34-39.
- [2] J. Laskar, *A numerical experiment on the chaotic behavior of the solar system*, Nature **338** (1989), 237-238.

- [3] H. Pollard, *Celestial Mechanics*, Carus Mathematical Monographs # 18, Math. Asoc. of Amer., 1976.
- [4] D. Saari, *A visit to the Newtonian n-body problem via elementary complex variables*, The American Math. Monthly **97** (1990), 105-119.
- [5] C. Simó, *El Sistema Solar ¿caótico pero estable?*, Contactos **3** (Nueva Epoca) (1991), 7-12.
- [6] G.J. Sussman, J. Wisdom, *Numerical evidence that the motion of Pluto is chaotic*, Science **241** (1988), 433-437.

TEORIA DE GRUPOS Y PARTICULAS ELEMENTALES

O. A. SÁNCHEZ-VALENZUELA

Centro de Investigación en Matemáticas
Guanajuato, México.

Prólogo y Advertencia. Este escrito es una versión extendida del “acordeón” que preparé para mi plática en el Coloquio del CINVESTAV (Julio-Agosto, 1991). Nada—o casi nada—de lo que aquí está escrito debe ser atribuible a mi investigación, ni acreditable a mi persona, excepto—posiblemente—el orden de la presentación y la ponderación de ciertos temas. Todo lo que voy a exponer lo he aprendido de mi maestro, S. Sternberg y él mismo ya ha escrito en diversos lugares, de diversas maneras y con diversos enfoques, las ideas que aquí presento “refritas”. He extraído libremente, de las varias referencias donde el material se encuentra, bloques enteros de texto que también he alterado de manera muy libre, pero sólo con la intención de armar una presentación coherente, sencilla, fácil de leer y expuesta en “una sola pieza”, para la audiencia particular del Coloquio. Deseo aprovechar la ocasión para expresarle mi agradecimiento a S. Sternberg por sus enseñanzas y a los organizadores del Coloquio; especialmente al Profesor Luis Gorostiza por invitarme a participar y al Profesor Juan José Rivaud por las pláticas y discusiones que tuvimos durante estos últimos meses.

¿QUÉ ES UNA PARTÍCULA ELEMENTAL?

Comenzaremos recordando cómo fué evolucionando en la historia de la humanidad el concepto de “elemento”. Los elementos aparecen con su connotación actual—esto es, como objetos básicos a partir de los cuales se componen todos los cuerpos materiales—entre los pueblos de la antigüedad y son los herederos griegos quienes difunden la idea. Para ellos, todos los objetos del mundo material estaban compuestos por cuatro elementos: tierra, aire, agua y fuego. En la cultura helénica, se originó también la concepción atómica de la materia que afirmaba que el proceso de subdividir cualquier objeto material sucesivamente, poseía un límite natural: el *átomo*, o entidad indivisible. Los elementos eran pues, los átomos.

Mucho se dijo e hizo durante la edad media tomando como base las interpretaciones de las enseñanzas filosóficas—degradadas por el paso del tiempo—de los antiguos griegos. La humanidad escuchó a los alquimistas hablar de la “transmutación de los elementos” y muchos se dedicaron a obtener metales preciosos a partir de piedras ociosas. Sin embargo, el gran mérito de la alquimia fué el de proporcionar un número de experimentos tal, que la primera definición razonable de “elemental” cayó operacionalmente por su propio peso

Parcialmente apoyado por el Depto. de Matemáticas del CINVESTAV, con fondos proporcionados por CONACYT.

y fue Thomas Boyle quien la formuló en el siglo XVIII: un *elemento es una substancia indescomponible en ninguna otra, u otras, después de realizar sobre ella una serie (bien determinada) de operaciones químicas.*

A partir de ese momento, el proyecto fundamental para el entendimiento de la composición de los objetos materiales, se tradujo en darse a la búsqueda de los “elementos químicos”. El proyecto culminó con el descubrimiento de la “tabla periódica de los elementos”, por Mendeleev. (Cabe mencionar, sin embargo, que la periodicidad de la tabla permaneció—en ese momento—sin explicación alguna.) La vieja teoría atómica de los antiguos griegos ganó sustento al descubrirse las *leyes de las proporciones* en las reacciones químicas y un gran entendimiento se derivó de pensar que los objetos materiales se componían de moléculas, que a su vez estaban formadas por combinaciones enteras de átomos de los diferentes elementos que ocurrían en su composición. Las ideas alquimistas pasaron a formar parte de las “épocas oscuras” y la tarea de obtener oro a partir del zinc no volvió jamás a ser “ciencia respetable”.

Los descubrimientos del electrón y de la radioactividad, sin embargo, vinieron a modificar nuevamente la concepción de lo “elemental”. Con ellos, la idea de la “inmutabilidad” de los elementos había rodado por tierra. El nuevo proyecto para entender la composición de los objetos materiales consistió en buscar todas las *partículas subatómicas*; éstas ocupaban ahora el lugar de los “elementos indivisibles” de los antiguos griegos. El proyecto terminó en la década de los veinte con la *teoría nuclear* de los átomos; entre otras cosas, esta teoría permitió finalmente explicar en gran medida, la asombrosa periodicidad de la tabla de Mendeleev. Los elementos—que resultaron objetos localizables en el espacio—se redujeron a sólo tres: el electrón, el protón y el neutrón.

Los experimentos de los años siguientes pronto demostraron que esta no podía ser toda la historia. Por ejemplo, se observó que los neutrones no eran buenos candidatos a “elementales” debido a su *inestabilidad*: un neutrón *decae* en un protón, un electrón y un “neutronicito” (*neutrino*, en italiano; término debido a E. Fermi). Rápidamente aparecieron otros *elementos distintos*; proliferaron las “partículas elementales”. El gran proyecto de encontrar *todas* las partículas elementales tuvo mucho de taxonómico y artístico en cuanto a la depuración de técnicas para detectar nuevos objetos y al diseño de dispositivos experimentales se refiere. El proyecto culminó esta vez con el surgimiento del llamado “modelo estándar” de la década de los setentas. El modelo dice que sólo hay dos tipos de constituyentes básicos: *leptones* y *quarks*. Los leptones son, o bien partículas localizables del tipo del electrón, o bien, neutrinos. Los quarks—que los hay en una variedad de *tres colores*—son finalmente los componentes básicos del resto de las “partículas elementales” (e.g., del neutrón, del protón, etc.).

Aquí cabe señalar que existe una clase más de *entidades básicas* que definitivamente tienen que ver con la composición de la materia. A éstas se les ha referido tradicionalmente como “partículas elementales” también, pero su naturaleza es esencialmente distinta: no se trata propiamente de constituyentes de la materia, sino de portadoras de las “fuerzas de interacción” entre las partículas que sí pueden calificarse como *constitutivas*. A esta clase pertenecen los *fotones* y los *bosones masivos*.

Para finalizar esta sección, mencionaremos que las partículas elementales también exhiben una periodicidad: el modelo estándar dice que vienen en *tres generaciones*; cada generación consiste en una partícula tipo electrón, un tipo de neutrino y dos quarks con color. La periodicidad de las generaciones ha conducido nuevamente a la idea de que tanto los leptones como los quarks podrían a su vez ser "objetos compuestos" y que existe un nivel más fundamental aún. Sin embargo, muchos han sido también los argumentos físicos que se han dado en favor de la idea de que los leptones y los quarks son realmente las piezas más básicas de este rompecabezas.

LA TEORÍA DE GRUPOS EN LA DESCRIPCIÓN DE LA NATURALEZA

Desde la década de los cuarentas y tras el trabajo fundamental de E. Wigner, las partículas elementales se entienden operativamente dentro del marco de la teoría cuántica, como *representaciones irreducibles del grupo, G , de simetrías de la Naturaleza*. Desde entonces, el proyecto básico de la teoría de interacciones fundamentales, puede resumirse en la resolución de los siguientes tres problemas:

1. Determinar el grupo G (un problema para la física).
2. Determinar el conjunto, \widehat{G} , de todas las representaciones irreducibles de G (un problema para la matemática).
3. Determinar precisamente qué representaciones del conjunto \widehat{G} realmente ocurren en la Naturaleza como partículas elementales (un problema para la física).

Es claro que la solución de cualquiera de estos problemas tiene una incidencia directa sobre la posible solución de cualquiera de los otros dos. Nosotros no diremos mucho acerca de la complejidad de este formidable proyecto, pero sí queremos mencionar brevemente cuáles son las bases conceptuales de la definición de Wigner, para no dejar al lector con el mal sabor de boca de no saber por qué la teoría de representaciones de grupos juega un papel tan importante en la física como para hacer descansar sobre ella el fundamental concepto de "partícula elemental".

Tal vez, el descubrimiento más asombroso de la Mecánica Cuántica—desde un punto de vista metafísico primero y muy concretamente físico después de que se le formula con precisión—es el hecho de que los "razonamientos cuánticos" obedecen las leyes de una lógica distinta a la clásica aristotélica. El modelo de la *lógica cuántica* es tal que las *proposiciones* se hacen corresponder con *subespacios cerrados de un espacio de Hilbert complejo y separable, \mathcal{H}* (y escribimos la correspondencia así: $P \leftrightarrow W_P$). Las *implicaciones* corresponden a *contenciones de subespacios* en otros ($P \implies Q \leftrightarrow W_P \subset W_Q$). Los conectivos lógicos 'o' e 'y' se entienden respectivamente como *suma* e *intersección* de subespacios y la *negación* consiste en tomar el *complemento ortogonal*. La diferencia con la lógica clásica radica en la no distributividad del 'y' sobre el 'o' ni viceversa.

Cabe preguntarse cuáles son las razones físicas para prestar atención a dicha lógica y la respuesta la podemos referir a una primera explicación dentro del marco de la Mecánica

Clásica: un sistema físico está perfectamente definido por su espacio fase, S , y éste, no es otra cosa más que el conjunto de todos sus estados posibles. Las leyes dinámicas de los sistemas físicos son *deterministas*: conocer el estado de un sistema en un instante de tiempo dado, determina completamente el estado del mismo en cualquier instante posterior. Esto se debe a que las leyes dinámicas son sólo un sistema de ecuaciones diferenciales ordinarias en S . Las cantidades físicas observables se representan mediante funciones reales definidas en el espacio fase. El valor de una tal función en un punto $s \in S$ se interpreta como el valor de la cantidad física en cuestión cuando el sistema se encuentra en el estado s . La aseveración más general que uno puede hacer respecto a un sistema en relación a una cantidad observable—digamos, $f: S \rightarrow \mathbb{R}$ —es que su valor se encuentra dentro de cierto subconjunto, E , de números reales. Ello equivale a decir que el estado del sistema se encuentra dentro del subconjunto $f^{-1}(E)$ del espacio fase. En otras palabras, las proposiciones físicamente significativas, se hacen corresponder con ciertos subconjuntos del espacio fase. La relación de *inclusión* entre dichos subconjuntos, corresponde naturalmente a la *implicación* entre proposiciones. Los conectivos 'o' e 'y' corresponden a la unión e intersección de conjuntos, respectivamente y la negación a la complementación. Ésta es la *lógica del sistema* y se trata de una lógica distributiva.

Ahora bien, si un sistema no obedece las leyes dinámicas de la Mecánica Clásica, no tiene sentido asignarle un espacio fase. Sin embargo, sí es posible considerar la totalidad de aseveraciones físicamente verificables que se pueden hacer de él. El conjunto de tales afirmaciones podría llamarse la *lógica del sistema* si (y sólo si) fuera posible definir operaciones de implicación y negación que lo convirtieran en un conjunto parcialmente ordenado. El punto—sobre el que ya no discutiremos—es que un sistema en Mecánica Cuántica tiene por lógica al conjunto de subespacios cerrados de un espacio de Hilbert separable, \mathcal{H} , como ya se mencionó.

Un teorema fundamental de Wigner, establece que si la dimensión del espacio \mathcal{H} es mayor que dos, cualquier automorfismo de un sistema cuántico (esto es, cualquier automorfismo de su lógica) puede representarse mediante un operador, $U: \mathcal{H} \rightarrow \mathcal{H}$, que es, unitario, o antiunitario y está completamente determinado hasta un factor de fase. Por lo tanto, decir que un grupo G actúa por automorfismos de la lógica de un sistema cuántico dado, significa que nos es dada una representación proyectiva de G en \mathcal{H} por transformaciones unitarias o antiunitarias. Si la representación es altamente reducible—como sería el caso de un sistema muy complicado con muchas interacciones—la diferencia entre una representación proyectiva y una representación ordinaria tiende a desaparecer al aumentar el número de sumandos directos en la descomposición en irreducibles (el caso extremo resulta de permitir solamente un factor de fase arbitrario global). De allí que el objetivo sea fijar la atención en representaciones unitarias o antiunitarias para los sistemas más simples. En particular, la acción de la componente conexa de G da origen a una representación unitaria en \mathcal{H} .

Una pregunta de fundamental importancia es la siguiente: ¿Hasta qué punto está determinado un sistema por su grupo de automorfismos? Nuevamente, podemos dar una primera respuesta empleando argumentos de Mecánica Clásica. Para ello, vamos primero a reformular la pregunta y ponerla en términos más directos: ¿Es posible obtener un es-

espacio fase a partir de un grupo de transformaciones? Lo asombroso es que la respuesta es afirmativa cuando se trata de un grupo de Lie. Este descubrimiento—debido independientemente a Souriau, Kostant y Kirillov en la década de los setentas—es uno de los más significativos para la física teórica y, aunque sea muy por encima de los detalles técnicos, no queremos dejar de mencionar la manera en que se obtiene toda una variedad de espacios fase a partir de un grupo de Lie G .

Todo grupo de Lie actúa en su álgebra de Lie mediante la *representación adjunta* y por lo tanto, en el espacio dual del álgebra de la manera más natural posible: si f es una funcional lineal en el álgebra de Lie de G , la funcional $g \cdot f$, con $g \in G$, se define como la que asigna a cada vector v del álgebra, el valor $f(\rho(g^{-1})(v))$, donde ρ es la representación adjunta (Véase el inciso b de la siguiente sección para recordar la definición de representación). Cada órbita de esta acción es un espacio fase equipado con una manera muy precisa de escribir sus leyes dinámicas.

Cabe recordar que, una característica muy importante que presentan los sistemas de la Naturaleza que varían continuamente con el tiempo es que algunas cantidades observables no cambian a lo largo de la evolución dinámica de los mismos. Estas propiedades parecen ser tan fundamentales que teóricamente ocupan un lugar preponderante entre las leyes de la física; se les conoce como *leyes de conservación*. Por otra parte, existen sistemas físicos que presentan ciertas simetrías (i.e., propiedades de invariancia ante ciertas transformaciones). Un resultado fundamental de la Mecánica Clásica, que fué heredado casi sin cambio alguno por la Mecánica Cuántica, es la estrecha interrelación que existe entre las *leyes de conservación* y las *simetrías* de un sistema físico. Los teoremas en este campo se originan a mediados del siglo pasado con los trabajos de Jacobi y fueron consolidados por Emmy Nöther a principios del actual. Dichos resultados fueron promovidos al reino de la teoría cuántica por Heisenberg, Weyl y Wigner, fundamentalmente.

Los principios de simetría e invariancia resultaron, en la *nueva física*, herramientas imprescindibles; estos principios han permitido predecir algunos de los descubrimientos más asombrosos de la física moderna. Sólo por mencionar algunos ejemplos, la existencia del neutrino mismo viene a salvar al Principio de Conservación de la Energía en el decaimiento del neutrón. Las *reglas de selección de la espectroscopía molecular*, que resultaban de la incapacidad de observar ciertas transiciones entre niveles de energía atómicos o moleculares aparentemente permitidos por la teoría general, pudieron ser explicadas como una propiedad de las representaciones de los presuntos grupos de simetría de los átomos o moléculas bajo consideración. La invariancia (aproximada) en las interacciones dentro del núcleo respecto al intercambio *protón-neutrón* dió origen a la *teoría de Yang-Mills*, que, tras la incorporación de ciertos "mecanismos técnicos" (como el *rompimiento espontáneo de la simetría* y el *mecanismo de Higgs* para la adquisición de masa de ciertos bosones que intermedian las interacciones de las partículas fundamentales), produjo la primera exitosa teoría unificada de interacciones fundamentales: la *teoría electro-débil*. También cabe mencionar la predicción de Gell-Mann de la existencia del *estado singulete del mesón η* (masa aproximada, 550 *MeV*), que resultó de la hipótesis—propuesta también por Y. Ne'eman de manera independiente—de simetría $SU(3)$ en la composición de las partículas susceptibles de interactuar bajo la *fuerza fuerte*.

Los principios de simetría han sido existosamente usados en la teoría cuántica gracias a que con ellos uno puede anticipar respuestas precisas, aún en el caso en que no se cuente con una buena teoría dinámica de las interacciones involucradas. Es seguramente por ello que Wigner se aventurara a afirmar que “si conociéramos todas las leyes de la Naturaleza, . . . , sus propiedades de invariancia ya no proporcionarían información nueva alguna.”

En la siguiente sección ilustraremos con un ejemplo sencillo de la Mecánica Clásica, cómo es que un físico teórico argumenta en favor de tal o cual grupo como un candidato a el grupo de simetría de un sistema natural. Sin embargo, podemos mencionar aquí que hay lineamientos generales que dictan en qué dirección hay que buscar al grupo G de la Naturaleza: por ejemplo, éste debe actuar en el espacio-tiempo e incluir a todas las isometrías del mismo. En el artículo original de Wigner, se tomó a $G = SL(2, \mathbb{C}) \times \mathbb{R}^{3,1}$ —el producto semidirécto del doble cubriente de la componente conexa del grupo de Lorentz con el grupo de translaciones, $\mathbb{R}^{3,1}$, del espacio-tiempo. Lo que Wigner encontró fué que las representaciones físicamente relevantes de este grupo G , estaban parametrizadas por dos “números cuánticos”: (m, s) , siendo m un número real mayor o igual que cero (la masa) y s , un semientero positivo si $m > 0$, o bien, un semientero positivo o negativo si $m = 0$ (el espín).

Dado que la masa y el espín no son los únicos parámetros que caracterizan a las partículas, y dado que sólo ciertos valores reales de m , y para éstos, sólo ciertos valores de s , se observan en la Naturaleza, la historia contada por el grupo de Poincare, $G = SL(2, \mathbb{C}) \times \mathbb{R}^{3,1}$, no puede ser completa. Los modelos más estructurados incluyen los grupos de simetrías internas (de norma o gauge), $U(1) \times SU(2)$, $U(1) \times SU(2) \times SU(3)$, etc., como factores directos del grupo de Poincare.

Antes de dejar esta sección, conviene comentar sobre el por qué debe considerarse, no el grupo de Lorentz directamente, sino su *doble cubriente*. Esta es una de las herencias de Einstein modificada por las ideas de la Mecánica Cuántica: para Einstein, el grupo de los *observadores inerciales* debía contener alguna o algunas de las componentes conexas del grupo de Lorentz, $O(3,1)$. Sin embargo, las representaciones irreducibles de $O(3,1)$ sólo dan lugar a partículas con espín entero; esto es, a los *bosones*. Las partículas con espín semientero, los *fermiones*, sólo se obtienen hasta considerar las representaciones de $SL(2, \mathbb{C})$ que no se pueden factorizar a través de representaciones de $O(3,1)$. Dicho de otra forma, el grupo $SL(2, \mathbb{C})$ admite más representaciones de las que puede admitir (la componente de la identidad de) $O(3,1)_e$ y las diferentes representaciones son físicamente observables y distinguibles entre sí.

LA ESPECTROSCOPÍA MOLECULAR Y EL LEMA DE SCHUR

El propósito de esta sección es explicar cómo es que la espectroscopía molecular—que es una parte de la ciencia experimental—puede entenderse como una aplicación del llamado “Lema de Schur” de la teoría de grupos. A su vez, es un excelente pretexto para ilustrar de manera sencilla, cómo es que se puede argumentar físicamente para distinguir, e entre varios posibles candidatos a grupo de simetría de un sistema dado, a los que definitivamente

no pueden serlo. Para cubrir estos objetivos, vamos a recordar primero algunas definiciones básicas:

a) **Modelos moleculares.** Supóngase que tenemos un sistema de masas puntuales que se mantienen ligadas entre sí mediante fuerzas del tipo que ejercen los resortes (*fuerza* = $-k$ *desplazamiento*). Las constantes k se suponen suficientemente grandes para poder pensar que se trata de un sistema relativamente rígido. Podemos decir que tal sistema es un "modelo clásico"—digamos, en primera aproximación—para una molécula.

Supondremos que conocemos el número de masas puntuales pero no la forma del sistema en equilibrio. El sistema se encuentra dentro de una "caja negra" que podemos agitar y sacudir a diversas frecuencias y determinar, en virtud de las respuestas a estos estímulos, sus frecuencias de resonancia. Estas últimas, se supone que son las frecuencias de libre oscilación alrededor del equilibrio. Queremos mostrar cómo usar esta información, junto con teoría de grupos, para determinar la forma de la "molécula" en equilibrio.

Sea q el vector que describe la desviación del sistema desde la posición de equilibrio, especificando el desplazamiento de cada "átomo". Por lo tanto, si hay N "átomos", q será un vector en un espacio de dimensión $3N$. De la teoría de pequeñas oscilaciones sabemos que el comportamiento del sistema cerca del equilibrio está descrito por una ecuación diferencial de segundo orden de la forma,

$$q'' + Fq = 0, \quad (1)$$

donde F es un operador autoadjunto. Los eigenvalores de F determinan las frecuencias de oscilación y los eigenvectores determinan la configuración asociada a los "modos normales" de vibración.

Como ejemplo, supóngase que la "molécula" está formada por tres "átomos" idénticos, cada uno de masa m , situados a lo largo de una línea y unidos por resortes idénticos de constante de fuerza k .

Por simplicidad supondremos que los átomos pueden ser desplazados sólo en una dimensión, de manera que el vector $q = (q_1, q_2, q_3)$ describe completamente la desviación de la molécula a partir de la posición de equilibrio. Luego, la segunda ley de Newton conduce a,

$$q'' + Fq = 0, \quad F = \frac{k}{m} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}. \quad (2)$$

La simetría de la matriz F indica que ésta representa a un operador autoadjunto. Los eigenvectores de F corresponden a los modos normales de vibración, como se indica a

continuación:

Eigenvalor	Eigenvector	Modo normal
0	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	translación uniforme
$\frac{k}{m}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$	vibración simétrica
$\frac{3k}{m}$	$\begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$	vibración antisimétrica

b) **El Lema de Schur.** Habiendo ya explicado el significado físico del sistema, vamos ahora a recordar las definiciones y los resultados que necesitamos de la teoría de representaciones de grupos.

Si G es un grupo y V un espacio vectorial (que puede por el momento suponerse de dimensión finita), una *representación* de G en V es un homomorfismo del grupo G , al grupo $\text{Aut } V$ de todas las transformaciones lineales invertibles de V en sí mismo. En otras palabras, se trata de una función,

$$\rho: G \rightarrow \text{Aut } V; \quad \text{con,} \quad \rho(gh) = \rho(g) \circ \rho(h).$$

Dada una tal representación, se dice que un subespacio W de V es *invariante*, si para todo $g \in G$, se tiene que $\rho(g)W \subset W$. Una representación es *irreducible*, si los únicos subespacios invariantes son los triviales: $\{0\}$ y V . Finalmente, si $\rho: G \rightarrow \text{Aut } V_\rho$ y $\nu: G \rightarrow \text{Aut } V_\nu$, son dos representaciones distintas del grupo G , cualquier función lineal

$$T: V_\rho \rightarrow V_\nu, \quad \text{con,} \quad T \circ \rho(g) = \nu(g) \circ T,$$

para todo $g \in G$, se llama *equivariante*. El conjunto de todas las transformaciones equivariantes de V_ρ a V_ν se denota por $\text{Hom}_G(V_\rho, V_\nu)$. Con estas definiciones, es bastante sencillo demostrar la siguiente,

Proposición. Si $T \in \text{Hom}_G(V_\rho, V_\nu)$, los subespacios $\text{Im } T \subset V_\nu$ y $\text{Ker } T \subset V_\rho$, son invariantes. Además, cuando $T \neq 0$, se tiene que,

- a) Si ν es irreducible, T es sobre.
- b) Si ρ es irreducible, T es inyectiva.
- c) Si ν y ρ son irreducibles, T es un isomorfismo.

En particular, para cualesquiera dos representaciones irreducibles, ρ y ν , el conjunto $\text{Hom}_G(V_\rho, V_\nu)$, o bien consta únicamente de la transformación lineal que es idénticamente

cero, o bien, es un anillo con división (todo elemento distinto de cero es invertible); esta observación nos conduce a la forma más usada de la proposición anterior (*el lema de Schur*):

$$\rho \text{ irreducible} \quad \text{y} \quad T \in \text{Hom}_G(V_\rho, V_\rho) \quad \implies \quad T = \lambda \text{ id},$$

para algún escalar λ .

c) **El grupo de simetría de una molécula.** Un grupo G es un grupo de simetría del sistema descrito por la ecuación diferencial (1), si existe una representación ρ de G en el espacio de los desplazamientos q (\mathbb{R}^3 en nuestro ejemplo de la molécula lineal y \mathbb{R}^{3N} en general), con la propiedad de que para cada solución q del sistema, $\rho(g)q$ también es solución, para toda $g \in G$. De manera informal, G refleja la simetría existente entre las fuerzas que mantienen unida a la molécula.

Es fácil darse cuenta que si G es un grupo de simetría para el sistema (1), entonces, el operador F que aparece en la ecuación, debe ser equivariante. Esta es la base fundamental de los razonamientos subsiguientes.

En el ejemplo de la molécula lineal, el grupo de simetría, G , debe al menos incluir al elemento g_{13} que intercambia los desplazamientos de los átomos '1' y '3', dado que el sistema es indistinguible del que resulta al intercambiar dichos átomos. Así pues,

$$g_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \in G,$$

y se observa que $g_{13}F = Fg_{13}$, siendo F la matriz (2). La situación es entonces la siguiente: tenemos una representación ρ de G en el espacio vectorial V de los desplazamientos de la molécula y un operador autoadjunto F en V que conmuta con la acción de G . Analicemos primero el caso extremo en el que la acción de G es irreducible:

G actúa irreduciblemente. El lema de Schur dice que F debe ser un múltiplo escalar de la identidad; esto es, F tiene un solo eigenvalor. Por otro lado, obsérvese que genéricamente, al elegir un operador autoadjunto al azar (sin requerir que éste conmute con la acción de G), el número esperado de eigenvalores distintos es igual a la dimensión del espacio vectorial V . La "moraleja" resulta de argumentar en el sentido opuesto: si se observa una sola eigenfrecuencia donde debiera haber habido $\dim V$ eigenfrecuencias distintas, es muy probable que se deba a que existe un grupo G actuando irreduciblemente como simetría del sistema, haciendo coincidir a todos los eigenvalores de F . Cuando se observa más de un eigenvalor, se puede argumentar como en el caso siguiente:

G actúa reduciblemente. Si F conmuta con la acción de G y v es un eigenvector de F , digamos, $Fv = \lambda v$, entonces es fácil deducir que $F\rho(g)(v) = \lambda\rho(g)(v)$, para todo $g \in G$. En otras palabras, el eigenespacio de F correspondiente a un eigenvalor dado, λ , es invariante para la representación ρ . Dado que F es autoadjunto, podemos descomponer V en la suma directa de sus eigenespacios correspondientes a eigenvalores distintos:

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_m}; \quad \lambda_i \neq \lambda_j, \quad \text{if } i \neq j; \quad m \leq \dim V.$$

donde cada subespacio V_{λ_i} es G -invariante. Ahora bién, para cada i , podemos aún descomponer el subespacio V_{λ_i} en una suma directa de subespacios G -irreducibles:

$$V_{\lambda_i} = V_1^{(i)} \oplus V_2^{(i)} \oplus \cdots \oplus V_{r_i}^{(i)}.$$

La "moraleja" esta vez es la siguiente: si G no actúa irreduciblemente, V puede descomponerse en la suma directa de subespacios G -irreducibles, digamos,

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_p, \quad p \leq \dim V,$$

y en cada V_μ , F tiene un sólo eigenvalor. En particular, el número de eigenvalores de F es, a lo más, igual a p ; esto es, al número de componentes G -irreducibles en la descomposición de V . Sin embargo, este número ya no tiene nada que ver con F y depende únicamente de la representación ρ .

¿Cómo puede usarse esta información en la práctica? La respuesta tiene por filosofía la siguiente: Supóngase que existen dos posibilidades para el grupo de simetría del sistema:

$$\rho_1: G_1 \rightarrow \text{Aut } V, \quad \text{y}, \quad \rho_2: G_2 \rightarrow \text{Aut } V.$$

Se buscan a continuación, las respectivas descomposiciones en irreducibles:

$$V = V_1^{\rho_1} \oplus V_2^{\rho_1} \oplus \cdots \oplus V_{p_1}^{\rho_1},$$

$$V = V_1^{\rho_2} \oplus V_2^{\rho_2} \oplus \cdots \oplus V_{p_2}^{\rho_2}.$$

Supóngase que $p_1 > p_2$. Si se observan p_1 eigenfrecuencias distintas, entonces definitivamente G_2 no puede ser el grupo de simetría correcto. Por otro lado, si se observan p_2 frecuencias distintas, se puede argumentar es que existe cierta evidencia de que G_2 sea el grupo correcto.

ESTRUCTURAS FINAS, ESTADOS EXCITADOS Y FILOSOFÍA DEL "WIGNERISMO"

Un principio fundamental en la predicción de observaciones experimentales, empleando los métodos de la teoría de grupos, es el siguiente: si $\rho: G \rightarrow \text{Aut } V_\rho$ es una representación irreducible y H es un subgrupo de G , la restricción, $\rho|_H$, de la representación ρ a $H \subset G$ no es necesariamente irreducible. En otras palabras, considerando que V_ρ es un espacio de representación para H , es posible escribir de manera no trivial,

$$V_\rho = V_1^{(\rho)} \oplus V_2^{(\rho)} \oplus \cdots \oplus V_r^{(\rho)},$$

donde cada $V_\mu^{(\rho)}$ es H -irreducible. Para explicar de qué manera puede emplearse este principio en la práctica, vamos a suponer que un cierto sistema bajo nuestro estudio está descrito por una ley dinámica de la forma,

$$\partial_t \Psi = F_0 \Psi; \quad \Psi \in V_\rho, \quad (3)$$

con F_0 un cierto operador autoadjunto en V_ρ . Supondremos que una primera serie de experimentos revela que se pueden observar tres eigenvalores distintos de F_0 . Supóngase que la hipótesis de que esto se debe a que el grupo G actúa vía ρ en V_ρ como grupo de simetrías del sistema, es válida. De hecho, supóngase—para fines de este argumento—que la representación ρ descompone a V_ρ en tres subespacios G -irreducibles:

$$V_\rho = V^{(1)} \oplus V^{(2)} \oplus V^{(3)}.$$

Supóngase también que F_0 actúa en $V^{(j)}$ como $\lambda_j id$, con $\lambda_j \neq \lambda_k$ si $j \neq k$.

Sin embargo, en una segunda serie de experimentos se demuestra que en realidad hay una estructura *más fina*: lo que en la primera serie de experimentos aparecía como una sola frecuencia λ_1 , es en realidad un agregado de—digamos, sólo por argumentar—tres frecuencias perceptiblemente distintas (o bien, en un lenguaje más usado, que $V^{(1)}$ se *desdobla* en un *tripleto*). Similarmente, la nueva serie de experimentos muestra que en $V^{(2)}$, donde actuaba F_0 como λ_2 , hay en realidad otras tres frecuencias distintas y sólo por argumentar, digamos que en $V^{(3)}$ se observan dos más. El argumento es entonces el siguiente:

El operador F_0 no es el correcto en la descripción de la dinámica; la razón es que en la primera serie de experimentos, no se tomó en cuenta que había interacciones adicionales. Cuando sí se les toma en cuenta, el operador que describe la dinámica correcta no es F_0 , sino $F = F_0 + F_1$, donde F_1 representa una “pequeña perturbación” al operador F_0 original. Esto significa que si bien, $\rho(g)F_0 = F_0\rho(g)$, para todo $g \in G$, la presencia del término perturbativo F_1 hace que no necesariamente sea cierto que $\rho(g)F = F\rho(g)$, para todo $g \in G$. Sin embargo, muy bien puede ser posible que,

$$\rho(h)F = F\rho(h), \quad \text{para todo } h \in H,$$

donde H es algún subgrupo de G . Este argumento es particularmente seductor cuando resulta que la restricción de la representación original ρ de G al subgrupo H , descompone los subespacios $V^{(i)}$ originales en, al menos, el número observado de frecuencias distintas en la segunda serie de experimentos. Y todavía es mejor aún si el nuevo operador $F = F_0 + F_1$ posee los eigenvalores correctos.

El argumento general ya no es entonces difícil de deducir y el punto a notar es que estos argumentos son bastante independientes de las leyes dinámicas. No importa demasiado si se trata de una ecuación diferencial como (1) o como (3). No es siquiera preciso conocer las leyes dinámicas correctas. Lo único que se requiere es que de “primeros principios” tenga sentido argumentar en términos de las observaciones experimentales según alguna correspondencia con ciertas peculiaridades de alguna clase de operadores. Este es particularmente el caso en la Mecánica Cuántica; sus leyes aseguran que los eigenvalores de los operadores autoadjuntos son cantidades físicamente observables. De esta manera, es posible argumentar desde un principio con los grupos G y H directamente y sus representaciones. Bajo ciertas circunstancias, es posible controlar experimentalmente las perturbaciones F_1 y entonces las eigenfrecuencias λ_i observadas según el grupo G con

el operador F_0 corresponden a *estados básicos* y las frecuencias derivadas de F según la *reducción de simetría* de G a H , corresponden a *estados excitados*.

Fué precisamente siguiendo esta filosofía general—usada originalmente por Wigner para explicar las observaciones de la espectroscopía atómica y molecular—como Gell-Mann y Ne'eman llegaron a la conclusión de que tras el estudio de las representaciones irreducibles de $SU(3)$ se podían acomodar en patrones perfectamente coherentes las “partículas elementales” que hasta principios de la década de los sesentas se conocían.

REFERENCES

1. Chew, G. F., Gell-Mann, M., and Rosenfeld, A. H., *Strongly interacting particles*, Scientific American (1964).
2. Ne'eman, Y., and Sternberg, S., *Internal supersymmetry and unification*, Proc. Natl. Acad. Sci. USA **77** (1980), 3127-3131.
3. Ne'eman, Y., and Sternberg, S., *Internal supersymmetry and dimensional reduction*, Gauge theories: fundamental interactions and rigorous results (Dita, P. Georgescu, V., and Purice, R., eds.), Birkhäuser, Boston, 1982, pp. 103-142.
4. Ne'eman, Y., and Sternberg, S., *Superconnections and internal supersymmetry dynamics*, Proc. Natl. Acad. Sci. USA **87** (1990), 7875-7877.
5. Ne'eman, Y., and Sternberg, S., *Internal supersymmetry and superconnections*, Preprint 1991 (to appear).
6. Sánchez-Valenzuela, O. A., *Matemática de las simetrías discretas de la física*, Ciencia **42** (1991), 125-140.
7. Sternberg, S., *Lecture notes on groups and representation theory*, Harvard University, Department of Mathematics, Cambridge, MA, 1984.

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS, APARTADO POSTAL 402, GUANAJUATO, GTO. 36000, MÉXICO

E-mail address: saval@unamvml.bitnet

GEOMETRIA Y FISICA*

M. Socolovsky**

Instituto de Ciencias Nucleares, UNAM, Circuito Exterior, C. U.
Apartado Postal 70-543, 04510 México, D. F.

Resumen

Se describen en la primera parte cuatro ejemplos sencillos de haces fibrados principales (h.f.p.'s) no triviales con fibra discreta: las cubiertas universales del círculo por los reales y del plano complejo menos un punto por la superficie de Riemann del logaritmo, y las cubiertas dobles de los mismos espacios por el círculo y la superficie de Riemann de la raíz cuadrada respectivamente. Se discute brevemente la relación de las cubiertas universales con la mecánica cuántica en la formulación de la integral de camino de Feynman y de las cubiertas dobles con partículas de spin $\frac{1}{2}$. En la segunda parte se presenta una deducción completa de las relaciones de compatibilidad entre las proyecciones locales de la forma de una conexión en un h.f.p. arbitrario (potenciales de norma) y de las transformaciones de norma utilizadas en Física.

1. Introducción

Aproximadamente en los últimos veinte años se ha desarrollado una descripción geométrica de las teorías de campos de norma para las interacciones fundamentales (electromagnéticas (Maxwell-Faraday), nucleares débil y fuerte (Yang-Mills, Weinberg, Salam, Glashow, Gress, Wilczek, Politzer) y gravitacionales (Einstein))¹. Dicha descripción se basa fundamentalmente en la teoría de conexiones en haces fibrados principales (h.f.p.'s) y en sus haces asociados, que fuera desarrollada con mucha anterioridad e independientemente

*Basado en la plática presentada por el autor durante el VII Coloquio del Departamento de Matemáticas del CINVESTAV, ESFM-IPN, México, Julio 29 - Agosto 16, 1991.

**Con licencia en el Depto. de Física, Centro de Investigación y de Estudios Avanzados del IPN.

de las aplicaciones físicas (aproximadamente entre los años '30 y '50²), la que involucra conceptos de topología, geometría diferencial y compleja y topología algebraica³. Mientras que las conexiones representan la "radiación" en un sentido generalizado (fotones en la teoría de Maxwell, bosones intermediarios W^\pm y Z^0 en la teoría débil, gluones en la teoría fuerte y gravitones en la teoría de Einstein), los campos de materia (electrones, neutrinos y quarks) se representan por funciones equivariantes en el espacio total de haces principales (ó equivalentemente por secciones globales en haces asociados), induciéndose así de manera natural el acoplamiento "radiación-materia" vía la derivada covariante de secciones con respecto a una conexión dada⁴.

La teoría cuántica de campos de norma involucra el cálculo de integrales de Feynman⁵ sobre espacios moduli típicamente de dimensión infinita, espacios que resultan del cociente entre el espacio de conexiones sobre un h.f.p. dado y el grupo de norma ó grupo de automorfismos verticales del haz. (Para un estudio detallado de la topología y la geometría Riemanniana de estos espacios véase las Refs. 6 y 7.) Al considerar el acoplamiento con la materia es necesario realizar integraciones sobre álgebras de Grassmann de dimensión infinita (integrales de Berezin, ver Ref. 8); este mismo tipo de integrales se debe realizar sobre los campos "fantasma" ("ghosts") que aparecen al "levantar" integrales sobre el espacio moduli al espacio de conexiones, Refs. 9 y 10.

En este artículo se presenta al lector una breve introducción al tema de haces fibrados principales y conexiones a través de dos secciones que, en principio, pueden ser leídas independientemente. En la sección 2 se describen detalladamente cuatro ejemplos de h.f.p.'s no triviales con fibra discreta: las cubiertas universales $Z \rightarrow \mathcal{R} \rightarrow S^1$ y $Z \rightarrow SR\ell \rightarrow C^*$, y las cubiertas dobles $Z_2 \rightarrow S^1 \rightarrow S^1$ y $Z_2 \rightarrow SRr. \rightarrow C^*$. Se deducen los correspondientes grupos de norma y fórmulas concretas para las proyecciones, atlas y fibras sobre cada punto, y se discuten los sistemas físicos con los que dichos haces están relacionados. Se introduce el concepto de conexión en forma geométrica (que en estos ejemplos resulta ser única) y el de forma de la conexión (igual a cero en estos casos). En la sección 3 se hace una deducción detallada de la fórmula que da las relaciones de compatibilidad⁴ entre las proyecciones locales en el espacio base de una conexión en el espacio total de un h.f.p. (ambos arbitrarios) y la adaptación de esta fórmula para dar las transformaciones de norma usuales en Física¹⁶.

2. Cuatro ejemplos de haces principales

En esta sección describiremos cuatro ejemplos de haces fibrados principales (dos recubrimientos universales y dos cubiertas dobles) que permite en particular dar una idea concreta de los elementos que intervienen en tales estructuras. Es interesante observar que el 1° de los haces considerados, la cubierta universal del círculo (los números reales), juega un papel importante en toda la mecánica cuántica: en esta teoría las amplitudes de transición se expresan en términos de "integrales de camino de Feynman" y los integrandos correspondientes son la exponencial de la acción clásica asociada al camino cuya contribución se considera multiplicada por $2\pi i$ y dividida por la unidad de acción (la constante de Planck \hbar)⁶. Por ejemplo, la amplitud de probabilidad de que una partícula no-relativista de masa m en presencia de un potencial $V(\vec{x})$ pase del punto \vec{x}' en el instante t' al punto \vec{x}'' en el instante t'' en el espacio ordinario de 3 dimensiones (E^3) está dada por $K(\vec{x}'', t''; \vec{x}', t') = \int \exp \frac{2\pi i}{\hbar} \int_{t'}^{t''} dt (\frac{1}{2} m \dot{\vec{x}}^2 - V(\vec{x}))$ donde el dominio de integración son las funciones continuas $\vec{x}: [t', t''] \rightarrow E^3$ que satisfacen $\vec{x}(t') = \vec{x}'$ y $\vec{x}(t'') = \vec{x}''$. El 2° haz, la cubierta universal del plano complejo "agujereado", es el lugar geométrico donde se describe el efecto Aharonov - Bohm de la mecánica cuántica por medio de la integral de Feynman: ésta se "levanta" y se realiza sobre un dominio 1-conexo (la superficie de Riemann del logaritmo) en lugar de realizarla sobre el plano menos un disco (la sección del solenoide que produce el campo magnético)¹¹. La cubierta doble del círculo (el 3° de los haces en discusión) se puede considerar, al menos cualitativamente, como un modelo topológico para los estados físicos de una partícula de spin $\frac{1}{2}$ ya que, del lado geométrico, al recorrer 360° en el espacio base, cualquiera de las dos pre-imagenes en el espacio total recorre 180° y, del lado físico, en la mecánica cuántica la función de onda del electrón cambia de signo por una rotación del laboratorio en 360° . Una discusión más detallada de este modelo se puede encontrar en las Refs. 12 y 13. Finalmente el 4° haz, la cubierta doble de C^* por la superficie de Riemann de la raíz cuadrada, mantiene con el 3° una relación análoga a la que el 2° mantiene con el 1°: se obtiene del anterior por una deformación homotópica de los espacios base y total. Esto induce a pensar que su relación con las partículas de spin $\frac{1}{2}$ es la misma que la del caso iii).

i) La cubierta universal del círculo: los reales

Consideremos el círculo $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ (espacio base), los reales \mathcal{R} (espacio total) y la función $exp : \mathcal{R} \rightarrow S^1$, $exp(t) := e^{i2\pi t}$ (proyección). (Como variedad diferenciable S^1 es cerrada (compacta y sin frontera), orientable, conexa por trayectorias (c.p.t.), de dimensión real = 1 y con grupo fundamental $\pi_1(S^1) \cong \mathbb{Z}$, como grupo de Lie $S^1 < \mathbb{C}^* = GL_1(\mathbb{C})$; como variedad diferenciable \mathcal{R} es no-compacta y sin frontera, orientable, c.p.t., contractible (i.e. del mismo tipo de homotopía que un punto) y por lo tanto 1-conexa (i.e. $\pi_1(\mathcal{R}) \cong 0$) y de dimensión real = 1, con $+$: $\mathcal{R} \times \mathcal{R} \rightarrow \mathcal{R}$, $(t, t') \mapsto t + t'$ \mathcal{R} es grupo de Lie y exp es un homomorfismo de grupos que relaciona en forma natural la estructura aditiva en \mathcal{R} con la multiplicativa en S^1 .) Si $z \in S^1$, $\mathcal{R}_z := exp^{-1}(\{z\}) = \{\frac{\phi}{2\pi} + k\}_{k \in \mathbb{Z}}$ con $\phi = Arg z \in (-\pi, \pi]$ es la fibra sobre z y para cada $z \in S^1$, $\mathcal{R}_z \cong \mathbb{Z}$ (fibra). Notemos que $\mathbb{Z} < \mathcal{R}$ y que $\psi : \mathcal{R} \times \mathbb{Z} \rightarrow \mathcal{R}$, $(t, n) \mapsto \psi(t, n) := t + n$ es una acción (derecha) de \mathbb{Z} sobre \mathcal{R} ($t + 0 = t$ y $(t + n) + m = t + (n + m)$) libre sobre \mathcal{R} ($t + n = t \Rightarrow n = 0$) y transitiva sobre cada fibra ($t = \frac{\phi}{2\pi} + k, t' = \frac{\phi}{2\pi} + k' \in \mathcal{R}_z \Rightarrow \exists n \in \mathbb{Z}$ t.q. $t' = t + n$, a saber $n = k' - k$). Como variedad diferenciable \mathbb{Z} es discreta i.e. su topología coincide con su conjunto potencia ($\tau_{\mathbb{Z}} = \mathcal{P}(\mathbb{Z})$) y de dimensión real = 0. (Nótese que la inclusión $i : \mathbb{Z} \rightarrow \mathcal{R}$, $i(n) := n$ es un monomorfismo de grupos y que por lo tanto se tiene la sucesión exacta corta (s.e.c.) de grupos abelianos $0 \rightarrow \mathbb{Z} \xrightarrow{i} \mathcal{R} \xrightarrow{exp} S^1 \rightarrow 0$.)

Sean $U = S^1 - \{i\}$ y $V = S^1 - \{-i\}$; se tiene $U, V \subseteq S^1$ con $U \cup V = S^1$ y $U \cap V = S^1 - \{i, -i\}$. Es fácil verificar que para $A = U, V$, $\varphi_A : \mathcal{R}_A := exp^{-1}(A) \rightarrow A \times \mathbb{Z}$, $\varphi_A(t) := (exp(t), [t])$ ($[t]$: parte entera de t) es un difeomorfismo de variedades, y que se cumple $\pi_1 \circ \varphi_A = exp_A (= exp|_{\mathcal{R}_A})$ (condición de trivialidad local). Si denotamos por $\mathcal{U} = \{(A, \varphi_A)\}_{A=U, V}$ (atlas), la estructura $\xi := (\mathcal{R}, S^1, exp, \mathbb{Z}, \mathcal{U}; \psi)$ es por definición un haz fibrado principal coordinado (h.f.p.c.). Nótese que φ_A induce $\sigma_A : A \rightarrow \mathcal{R}_A$, $\sigma_A(z) := \varphi_A^{-1}(z, 0)$ que satisface $exp_A \circ \sigma_A = 1_A$ (σ_A : sección local). (Usando la transitividad de ψ en fibras es fácil mostrar que esta correspondencia es biyectiva.)

La clase de equivalencia de h.f.p.c.'s $[\xi] := (\mathcal{R}, S^1, exp, \mathbb{Z}, U; \psi)$, donde $U = \{U, V\}$ y $\xi' \sim \xi$ si ξ' y ξ difieren entre sí a lo más en las funciones φ'_A y φ_A , es un haz fibrado principal (h.f.p.). Por ejemplo, si m es un entero fijo arbitrario, $\mathcal{U}' = \{(A, \varphi'_A)\}_{A=U, V}$ con $\varphi'_A(t) = (exp(t), [t] + m)$ define un h.f.p.c. $\xi' \sim \xi$. En las aplicaciones a la Física, la elección de un sistema de trivializaciones locales (i.e. de un atlas) en un h.f.p. se suele

llamar *fijación de la norma*. Es usual la notación $[\xi] : Z \rightarrow \mathcal{R} \xrightarrow{\text{exp}} S^1$. Nótese que $[\xi]$ es un h.f.p. *no-trivial* i.e. $\mathcal{R} \not\cong S^1 \times Z$ (aunque localmente sí lo es). (En general, si $\mathcal{B}(B, G) (= \{[\xi]\})$ denota las clases de isomorfismo de h.f.p.'s con base B y fibra G y $[B, BG] (= \{[f]\})$ las clases de homotopía de funciones de B a BG : *espacio clasificante del grupo topológico G* (espacio base del haz universal $G \rightarrow EG \xrightarrow{\pi_G} BG$ asociado unívocamente (hasta isomorfismo) a G), se tiene el isomorfismo $\mathcal{B}(BG) \cong [B, BG]$. El espacio total de un representante de la clase $[[\xi]]$ correspondiente a la clase $[f]$ está dado por el "pull-back" $EG \times B \supset f^*(EG) := \{(e, b) | \pi_G(e) = f(b)\}$; se tiene entonces el h.f.p. $G \rightarrow f^*(EG) \xrightarrow{\pi_2} B$ y el diagrama conmutativo²

$$\begin{array}{ccc} f^*(EG) & \xrightarrow{\pi_1} & EG \\ \pi_2 \downarrow & & \downarrow \pi_G \\ B & \xrightarrow{f} & BG \end{array}$$

En particular si $B = S^n, \mathcal{B}(S^n, G) \simeq \pi_n(BG)$ (a menos de puntos base); para $G = Z, EZ \cong \mathcal{R}, BZ \cong S^1$ y $\pi_Z = \text{exp}$ resultando entonces $\mathcal{B}(S^1, Z) \cong \pi_1(S^1) \cong Z$. Se tiene: $\pi_1(S^1) \cong \{[f_n]\}_{n \in \mathbb{Z}}$ con $f_n : S^1 \rightarrow S^1, f_n(e^{it}) := e^{int}$; para $n = 1, f_1 = 1_{S^1}$ y $f_1^*(\mathcal{R}) = \mathcal{R}$, para $n = 0, f_0(e^{it}) = 1$ i.e. $f_0 = \text{cte.}$ y $f_0^*(\mathcal{R}) \cong S^1 \times Z$ (haz trivial): "Z copias del círculo".

Para $t \in \mathcal{R}, \exists! \mathcal{R}_z$ con $t \in \mathcal{R}_z$, a saber $\mathcal{R}_{\text{exp}(t)}; V_t := T_t \mathcal{R}_z$ se llama el *espacio vertical* de \mathcal{R} en t ; de $\dim_{\mathcal{R}} \mathcal{R}_z = 0$ resulta $V_t = 0$ y por lo tanto es único $H_t \stackrel{c}{\subset} T_t \mathcal{R} = \mathcal{R}$ con $T_t \mathcal{R} = V_t \oplus H_t$, a saber $H_t = \mathcal{R}$. Tal *distribución de espacios horizontales* $H := \{H_t\}_{t \in \mathcal{R}}$ es entonces la única *conexión* en el haz $Z \rightarrow \mathcal{R} \xrightarrow{\text{exp}} S^1$. Dado que Z es discreto, $\text{Lie} Z = T_0 Z = \{0\} = \mathcal{R}^0 \stackrel{c}{\subset} \mathcal{R}$ y por lo tanto la *forma de la conexión* ω_H es nula: en efecto, $\omega_H : \mathcal{R} \rightarrow T^* \mathcal{R} \otimes \text{Lie} Z, \omega_H(t) = (t, \omega_t)$ con $\omega_t \in T_t^* \mathcal{R} \otimes 0$ i.e. $\omega_t : \mathcal{R} \rightarrow 0$ lineal, luego $\omega_t(\lambda) = a_\omega \lambda = 0 \forall \lambda \in \mathcal{R}$ que implica $a_\omega = 0$ i.e. $\omega_H = 0$.

Finalmente, determinemos el *grupo de automorfismos verticales* del haz, $\text{Aut}_{S^1} \mathcal{R}(S^1, Z)$, dado por la colección de difeomorfismos $\{f : \mathcal{R} \rightarrow \mathcal{R}\}$ tales que hacen conmutar los diagramas

$$\begin{array}{ccc}
 \mathcal{R} \times Z & \xrightarrow{f \times 1_Z} & \mathcal{R} \times Z \\
 \psi \downarrow & & \downarrow \psi \\
 \mathcal{R} & \xrightarrow{f} & \mathcal{R} \\
 \swarrow \text{exp} & & \swarrow \text{exp} \\
 & S^1 &
 \end{array}$$

Se tienen entonces las ecuaciones $f(t+n) = f(t) + n$ y $e^{i2\pi f(t)} = e^{i2\pi t}$; de la 2da. resulta $e^{i(2\pi f(t)-t)} = 1$ y por lo tanto $f(t) = t + n := f_n(t)$ que trivialmente verifica la 1ra. condición; de $f_n \circ f_m(t) = (t+m) + n = t + (m+n) = f_{n+m}(t)$ resulta $\text{Aut}_{S^1} \mathcal{R}(S^1, Z) \cong Z$. (En la teoría de espacios recubridores $\tilde{X} \xrightarrow{p} X$ este grupo, que se denota por $G(\tilde{X}, p, X)$, se denomina *grupo de transformaciones recubridoras* y se demuestra que si \tilde{X} es 1-conexo y localmente arco-conexo, $G(\tilde{X}, p, X) \cong \pi_1(X, x)$ para cada $x \in X$.¹⁴)

ii) La cubierta universal de C^* : la superficie de Riemann del logaritmo

Consideremos ahora como espacio base al plano complejo menos un punto (el cero) $C^* = GL_1(C)$ (variedad compleja de dimensión compleja = 1, orientable, arco-conexa, no-compacta y sin frontera, del tipo de homotopía del círculo y por lo tanto con grupo fundamental $\pi_1(C^*) \cong Z$, y grupo de Lie abeliano con el producto heredado de C), y como espacio total a la *superficie de Riemann del logaritmo* $SRL = \bigcup_{k \in Z} \cong \bigsqcup_{k \in Z} S_k$ donde la k -sima hoja está dada por $S_k = \{(r, \theta), r > 0, \theta \in [2k\pi, 2(k+1)\pi)\}$ (ver Figura 1).

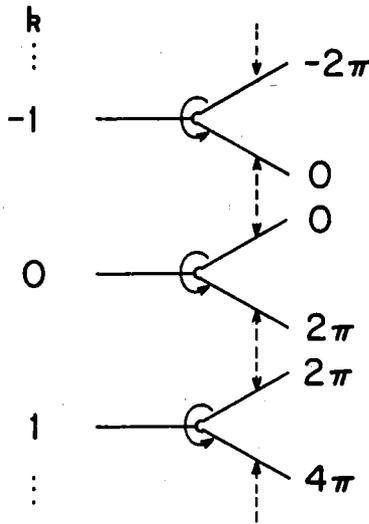


Fig. 1

Esquema de la $SR\ell$; las líneas de trazos indican que se deben identificar los semi-ejes correspondientes.

La forma precisa de definir a la $SR\ell$ y a su topología es utilizando la técnica de “cortar y pegar”, véase por ejemplo la Ref. 15. $SR\ell$ resulta una variedad compleja de dimensión compleja = 1, arco-conexa, 1-conexa i.e. $\pi_1(SR\ell) \cong 0$, no compacta y sin frontera, orientable; con la composición $s * s' = (r, \theta) * (r', \theta') := (rr', \theta + \theta')$ $SR\ell$ tiene la estructura de grupo de Lie abeliano con unidad $(1, 0) \in S_0$ e inversos $s^{-1} = (r, \theta)^{-1} = (\frac{1}{r}, -\theta)$; claramente si $s \in S_k$ y $s' \in S_\ell \Rightarrow s * s' \in S_{k+\ell}$. Se define la proyección $\pi : SR\ell \rightarrow C^*$, $\pi(s) = \pi(r, \theta) := re^{i(\theta - 2k\pi)}$, $s \in S_k$ (que en particular resulta un epimorfismo de grupos: $\pi(s * s') = \pi(rr', \theta + \theta') = rr'e^{i(\theta + \theta' - 2(k+\ell)\pi)} = re^{i(\theta - 2k\pi)} r'e^{i(\theta' - 2\ell\pi)} = \pi(s)\pi(s')$; también se tiene, dado que la inclusión $i : Z \rightarrow SR\ell, i(k) := (1, 2k\pi)$ es un monomorfismo de grupos: $i(k+\ell) = (1, 2(k+\ell)\pi) = (1, 2k\pi + 2\ell\pi) = (1, 2k\pi) * (1, 2\ell\pi) = i(k) * i(\ell)$, la s. e. c. de grupos abelianos $0 \rightarrow Z \xrightarrow{i} SR\ell \xrightarrow{\pi} C^* \rightarrow 0$ y por lo tanto $C^* \cong_{gr} SR\ell/Z$; para $z = re^{i\phi} \in C^* (r > 0, \phi = Arg z \in [0, 2\pi])$ la fibra sobre z está dada por $SR\ell_z = \pi^{-1}(\{re^{i\phi}\}) = \{(r, \phi + 2k\pi)\}_{k \in Z} \cong Z$

$\{(r, \phi + 2k\pi)\}_{k \in \mathbb{Z}} \cong Z$ (fibra). Z actúa libremente sobre $SR\ell$ y transitivamente en fibras, se tiene la función $\psi : SR\ell \times Z \rightarrow SR\ell, \psi(s, n) = \psi((r, \theta), n) := (r, \theta + 2\pi n)$.

Un sistema de trivializaciones locales (atlas) está dado por $\mathcal{U} = \{(U_+, \varphi_+), (U_-, \varphi_-)\}$ con $U_{\pm} = C^* - \mathcal{R}^{\pm}$ y los difeomorfismos $\varphi_{\pm} : SR\ell_{\pm} \rightarrow U_{\pm} \times Z, \varphi_{\pm}(s) = \varphi_{\pm}(r, \theta) = (re^{i(\theta - 2k\pi)}, k)$ donde $SR\ell_{\pm} = \pi^{-1}(U_{\pm})$ y $k \in \mathbb{Z}$ t.q. $s \in S_k$; claramente se cumple $\pi_1 \circ \varphi_{\pm} = \pi_{\pm} (= \pi|_{SR\ell_{\pm}})$ (ver Figura 2).

Se tiene por lo tanto el h.f.p.c. no trivial $\eta = (SR\ell, \pi, C^*, Z, \mathcal{U}; \psi)$. Como en el caso i), $\mathcal{B}(C^*, Z) \cong [C^*, BZ] \cong [S^1, S^1] \cong Z$, y un representante de un elemento de $\mathcal{B}(C^*, Z)$ se obtiene como "pull back" de \mathcal{R} por un representante de la correspondiente clase en $[C^*, S^1]$ (recuérdese que el haz universal para el grupo Z es $Z \rightarrow \mathcal{R} \xrightarrow{\exp} S^1$); así por ejemplo, el haz trivial $\mathcal{R} \times C^* \supset cte.(\mathcal{R}) = \{(t, z) | \exp(t) = cte.(z)\} \cong Z \times C^*$ con $cte. : C^* \rightarrow S^1, cte.(z) = s_0$ fijo en S^1 (por ej. $s_0 = 1$).

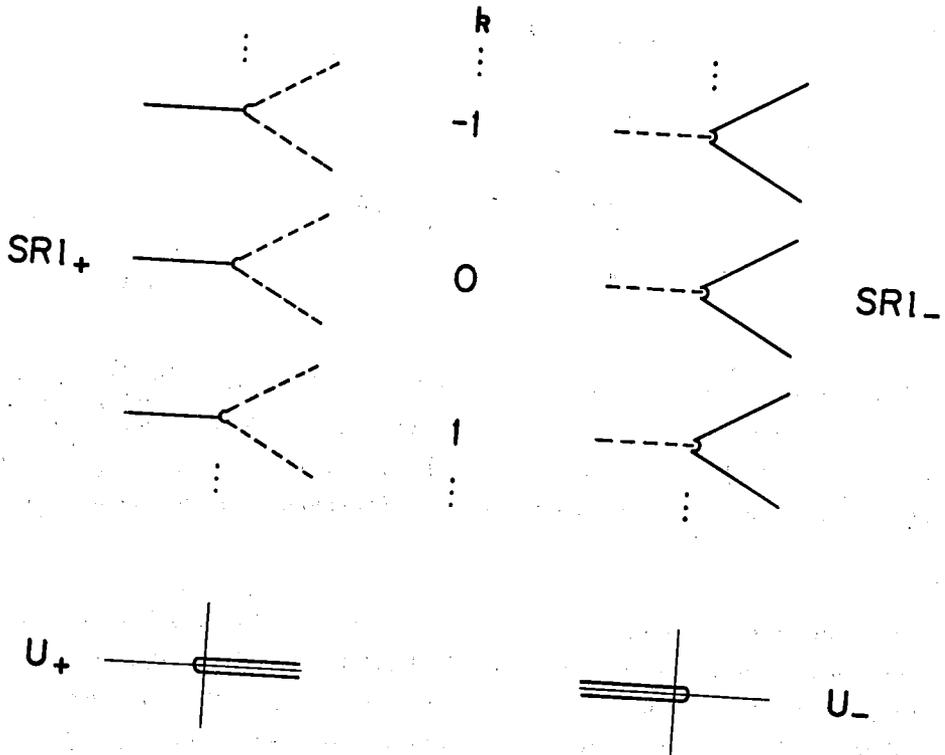


Fig. 2

Atlas del haz $Z \rightarrow SRL \xrightarrow{\pi} C^*$

La determinación de $Aut_C \cdot SRL(C^*, Z)$ es similar al caso i). Sea $\varphi \in Diff(SRL)$ t.q. conmuta el diagrama

$$\begin{array}{ccc}
 S R \ell \times Z & \xrightarrow{\varphi \times 1_Z} & S R \ell \times Z \\
 \psi \downarrow & & \downarrow \psi \\
 S R \ell & \xrightarrow{\varphi} & S R \ell \\
 \pi \searrow & & \swarrow \pi \\
 & C^* &
 \end{array}$$

Si $s = (r, \theta) \in S_k$, $\pi(\varphi(s)) = \pi(r', \theta') = r' e^{i(\theta' - 2k'\pi)} = \pi(r, \theta) = r e^{i(\theta - 2k\pi)}$ (con $k' \in Z$ t.q. $\varphi(s) \in S_{k'}$), luego $r' = r$ y $\theta' - \theta = 2\pi\ell$ con $\ell \in Z$ determinado por φ i.e.: $\varphi(r, \theta) = \varphi_\ell(r, \theta) = (r, \theta + 2\pi\ell)$ que verifica $\psi \circ (\varphi_\ell \times 1_Z) = \varphi_\ell \circ \psi$; de $\varphi_m \circ \varphi_\ell = \varphi_{m+\ell}$ resulta que el grupo de automorfismos verticales del haz es Z . (Mismo comentario que al final del punto i).)

Dado que Z es discreto, $V_s = T_s S R \ell_{\pi(s)} = 0$, y de $T_s S R \ell = \mathcal{R}^2$ resulta que la única distribución de espacios horizontales (conexión) $\{H_s\}_{s \in S R \ell}$ tal que $T_s S R \ell = V_s \oplus H_s$ está dada por $H_s = \mathcal{R}^2$. Para la forma de conexión se tiene $\omega_H : S R \ell \rightarrow T^* S R \ell \otimes 0$, $\omega_H(s) = (s, \omega_s)$ con $\omega_s : T_s S R \ell \rightarrow 0$ lineal y por lo tanto $\omega_s = 0 \forall s \in S R \ell$ i.e. $\omega_H = 0$.

Finalmente mencionemos que el nombre para $S R \ell$ se debe al siguiente hecho: sobre $S R \ell$ y $C^* - \{x > 0\} = \tilde{C}$ se definen respectivamente las funciones analíticas $\log: S R \ell \rightarrow C$, $\log(r, \theta) := \ln r + i\theta$ y $\text{Log}: \tilde{C} \rightarrow C$, $\text{Log}(z) = \text{Log}(r e^{i\phi}) := \ln r + i\phi$ con $\phi = \text{Arg } z \in (0, 2\pi)$. Restringiendo π y \log a $S_0^\circ := \{(r, \theta), r > 0, \theta \in (0, 2\pi)\} \subset S R \ell$ se tiene el diagrama conmutativo

$$\begin{array}{ccc}
 S_0^\circ & \xrightarrow{\log|} & C \\
 \pi| \downarrow & \nearrow & \\
 \tilde{C} & & \text{Log}
 \end{array}$$

$(\pi| = \pi|_{S^0}$ y $\log| = \log|_{S^0}$).

iii) La cubierta doble del círculo: el círculo

Se trata de un haz de esferas (uno de los haces de Hopf) ya que los espacios total y base y la fibra son respectivamente S^1, S^1 y $S^0 \cong Z_2 = \{+1, -1\}$: grupo de Lie discreto de 2 elementos; $\psi : S^1 \times S^0 \rightarrow S^1, \psi(x, g) = zg$ es la acción; $\pi : S^1 \rightarrow S^1, \pi(z) = z^2$ es la proyección donde como en i) $C \supset S^1 = \{z | |z| = 1\}$ i.e. $z = e^{i\varphi}$ con $\varphi = \text{arg} z = \phi + 2k\pi, k \in Z$ arbitrario y $\phi = \text{Arg} z \in [0, 2\pi)$; $S^1_z = \pi^{-1}(\{z\}) = \pi^{-1}(\{e^{i\varphi}\}) = \{\pm e^{i\varphi/2}\}$ es la fibra sobre un punto (claramente $S^1_z \cong Z_2$); $\mathcal{U} = \{(A, \psi_A)\}_{A=U,V}$, donde como en i) $U = S^1 - \{i\}$ y $V = S^1 - \{-i\}$ y $\psi_A : S^1_A = \pi^{-1}(A) \rightarrow A \times Z_2, \psi_A(e^{i\phi}) = (e^{i2\phi}, (-1)^k)$ si $k \in [k\pi, (k+1)\pi), k = 0, 1, \phi \neq \pi/4, 5\pi/4$ para $A = U$ y $\phi \neq 3\pi/4, 7\pi/4$ para $A = V$, es un atlas del haz $\xi = (S^1, S^1, \pi, Z_2, \mathcal{U}; \psi)$ (ver Figura 3).

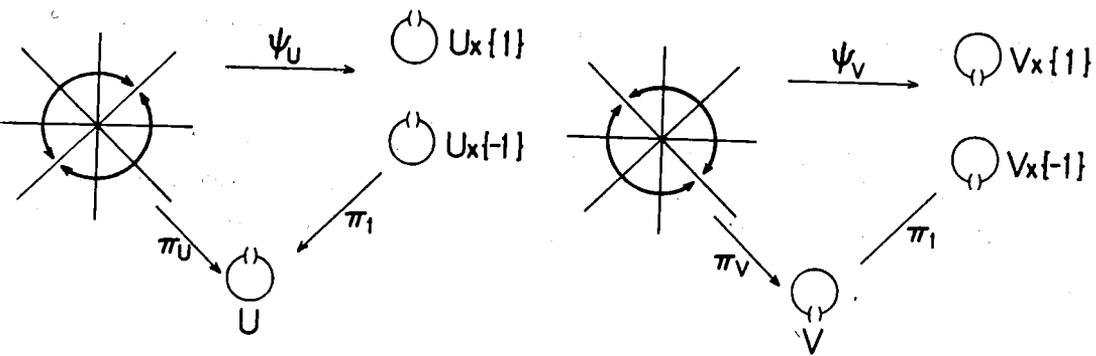


Fig. 3

Atlas para el haz $Z_2 \rightarrow S^1 \xrightarrow{\pi} S^1$

Nótese que el haz es no-trivial y que de $\mathcal{B}(S^1, Z_2) \cong \pi_1(BZ_2)$ y $BZ_2 \cong \mathcal{R}P^\infty$ resulta $\mathcal{B}(S^1, Z_2) \cong Z_2$ (además de ξ se tiene el haz trivial $S^1 \times Z_2$: "2 copias del círculo"). Se tiene la s.e.c. de grupos abelianos $0 \rightarrow Z_2 \xrightarrow{i} S^1 \xrightarrow{\pi} S^1 \rightarrow 0$ donde $i(k) := k$ es la inclusión de S^0 en S^1 , por lo tanto $S^1 \cong_{gr} S^1/Z_2$ y de i), $S^1/Z_2 \cong_{gr} \mathcal{R}/Z$.

El grupo de automorfismos verticales del haz ξ está dado por los difeomorfismos $f : S^1 \rightarrow S^1$ que satisfacen el diagrama conmutativo

$$\begin{array}{ccc}
 S^1 \times Z_2 & \xrightarrow{f \times 1_{Z_2}} & S^1 \times Z_2 \\
 \downarrow & \begin{array}{c} I \quad f \\ \searrow \quad \swarrow \\ II \end{array} & \downarrow \\
 S^1 & & S^1 \\
 \pi & \searrow & \swarrow \pi \\
 & S^1 &
 \end{array}$$

De II, $\pi(f(e^{i\varphi})) = \pi(e^{i\theta}) = e^{i2\theta} = \pi(e^{i\varphi}) = e^{i2\varphi}$ y por lo tanto $\theta = \varphi + k\pi, k \in Z$, luego $f(e^{i\varphi}) = (-1)^k e^{i\varphi} := f_k(e^{i\varphi})$ con $f_0 = 1_{S^1}$ y $f_1 = -1_{S^1}$; ambas soluciones verifican I: $\psi \circ (f_k \times 1_{Z_2})(e^{i\varphi}, g) = \psi(e^{i\theta}, g) = \psi((-1)^k e^{i\varphi}, g) = (-1)^k e^{i\varphi} g$ y $f_k \circ \psi(e^{i\varphi}, g) = f_k(e^{i\varphi} g) = (-1)^k e^{i\varphi} g$. Por lo tanto $Aut_{S^1} S^1(S^1, Z_2) \cong Z_2$.

Como en los casos anteriores, Z_2 discreto implica $Lie Z_2 = 0$ y por lo tanto la única conexión admisible $H = \{H_x = \mathcal{R}\}_x \in S^1$ tiene como forma de conexión la idénticamente nula.

iv) **La cubierta doble de C^* : la superficie de Riemann de la raíz cuadrada**
 Como en ii) el espacio base es C^* ; en cambio el espacio total es la *superficie de Riemann de la raíz cuadrada* $SRr := S_0 \cup S_1 \cong S_0 \amalg S_1$ donde la k -ésima hoja S_k (para $k = 0, 1$) está dada como en ii) (ver Figura 4).

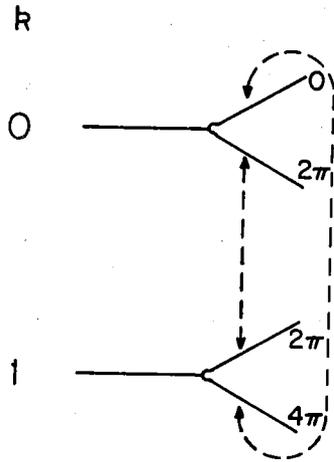


Fig. 4

Esquema de la SRr ; las líneas de trazos indican los semiejes que deben ser identificados

(Para mayores detalles de la construcción de la SRr véase la Ref. 15.) SRr resulta una variedad compleja de dimensión compleja = 1, no compacta, sin frontera, orientable, del mismo tipo de homotopía que S^1 y por lo tanto no-contractible ($\pi_1(SRr) \cong \mathbb{Z}$) y c.p.t. Con la composición $(r, \theta) * (r', \theta') := (rr', \theta + \theta' + 4k\pi)$ (donde $\theta + \theta' + 4k\pi$ es $\theta + \theta'$ si $\theta + \theta' < 4\pi$ y $\theta + \theta' - 4\pi$ si $\theta + \theta' \geq 4\pi$) SRr es un grupo de Lie abeliano con unidad $(1, 0) \in S_0$ e inversos $(r, \theta)^{-1} = (1/r, 4\pi - \theta)$. Se define la proyección $\pi : SRr \rightarrow \mathbb{C}^*$, $\pi(r, \theta) := re^{i(\theta - 2k\pi)}$ con $k \in \{0, 1\}$ t.q. $(r, \theta) \in S_k$ (en particular π resulta un epimorfismo de grupos), y con el monomorfismo inclusión $i : \mathbb{Z}_2 \rightarrow SRr$, $i(1) = (1, 0)$, $i(-1) = (1, 2\pi)$ se tiene la s.e.c. de grupos abelianos $0 \rightarrow \mathbb{Z}_2 \xrightarrow{i} SRr \xrightarrow{\pi} \mathbb{C}^* \rightarrow 0$ y por lo tanto $\mathbb{C}^* \cong_{gr} SRr/\mathbb{Z}_2$; de ii) $SRr/\mathbb{Z}_2 \cong_{gr} SRr/Z$. La fibra sobre cada $z = re^{i\phi} \in \mathbb{C}^*$ está dada por $SRr_z = \pi^{-1}(\{z\}) = \{(r, \phi), (r, \phi + 2\pi)\} \cong \mathbb{Z}_2$ (fibra) donde $\phi \in [0, 2\pi)$. Claramente \mathbb{Z}_2 actúa libremente sobre SRr y transitivamente sobre las fibras con $\psi : SRr \times \mathbb{Z}_2 \rightarrow SRr$, $\psi((r, \theta), 1) = (r, \theta)$ y

$\psi((r, \theta), -1) = (r, \theta + 2\pi(4\pi))$. Análogamente al caso ii), un sistema de trivializaciones locales está dado por $\mathcal{U} = \{U_{\pm}, \varphi_{\pm}\}$ con U_{\pm} como en ii) y $\varphi_{\pm} : SRr_{\pm} \rightarrow U_{\pm} \times Z_2$, $\varphi_{\pm}(r, \theta) = (re^{i\theta}, 1)$ si $(r, \theta) \in S_0$ y $\varphi_{\pm}(r, \theta) = (re^{i(\theta-2\pi)}, -1)$ si $(r, \theta) \in S_1$ ($SRr_{\pm} = \pi^{-1}(U_{\pm})$). Se tiene por lo tanto el h.f.p.c. no-trivial $\xi = (SRr, C^*, \pi, Z_2, \mathcal{U}; \psi)$. Análogamente al caso iii), $B(C^*, Z_2) \cong Z_2$ (además de ξ se tiene el haz trivial $C^* \times Z_2$: "2 copias de C^* "), $Aut_{C^*} SRr(S^1, Z_2) \cong Z_2$ y $\omega_H = 0$ para la única conexión admisible sobre $SRr, H = \{H_{(r, \theta)} = \mathcal{R}^2\}_{(r, \theta) \in SRr}$.

Finalmente, mencionemos que sobre SRr y C^* se definen respectivamente las funciones analíticas $\sqrt{\cdot} : SRr \rightarrow C$, $\sqrt{(r, \theta)} := \sqrt{r}e^{i\theta/2}$ y $F : \tilde{C} \rightarrow C$, $F(\rho e^{i\phi}) := \sqrt{\rho}e^{i\phi/2}$ ($\phi \in (0, 2\pi)$). Restringiendo $\sqrt{\cdot}$ y π a $S_0^{\circ} = \{(r, \theta) | \theta \in (0, 2\pi)\} \subset SRr$ se tiene el diagrama conmutativo

$$\begin{array}{ccc} S_0^{\circ} & \xrightarrow{\sqrt{\cdot}} & C \\ \pi|_{\downarrow} & \nearrow F & \\ & \tilde{C} & \end{array}$$

i.e. es la restricción $\pi|_{S_0^{\circ}}$ de la proyección del haz ξ la que da la relación entre $\sqrt{\cdot}|_{S_0^{\circ}}$ y F .

3. Transformaciones de norma

En esta sección presentamos una deducción exhaustiva de las *relaciones de compatibilidad* entre las proyecciones locales al espacio base de una conexión en el espacio total de un haz fibrado principal coordinado (h.f.p.c.) (ec. (3.4)), y la fórmula de *transformación de norma* utilizada en Física (ec. (3.9)). Detalles concernientes a la teoría de haces fibrados con conexiones que se requieren para su demostración se encuentran en las Refs. 16, 17, 18 y 19.

Sea ω la forma de una conexión H en un h.f.p.c. $\xi = (P, B, \pi, G, \mathcal{U}, \psi)$. ω es entonces la 1-forma diferencial sobre P con valores en el álgebra de Lie \mathcal{G} de G (lo que se denota usualmente por $\omega \in C^{\infty}(T^*P \otimes \mathcal{G})$) dada por $\omega(p) = (p, \omega_p)$, $\omega_p \in T_p^*P \otimes \mathcal{G}$, $\omega_p : T_pP \rightarrow$

\mathcal{G} , $\omega_p(v_p) = \varphi_p^{-1}$ (ver (v_p)) donde ver $(v_p) \in V_p = T_p G_b$, con $b = \pi(p)$ y $G_b = \pi^{-1}(\{b\})$, denota la parte vertical de $v_p \in T_p P = V_p \oplus H_p$ y $\varphi_p : \mathcal{G} \rightarrow V_p, A \mapsto \varphi_p(A)$ es el isomorfismo canónico de espacios vectoriales dado por $\varphi_p(A) = \check{\psi}_b(A)_p$ donde ψ_b es la restricción (transitiva) de la acción (libre) ψ de G sobre P a la fibra sobre b i.e. $\psi_b : G_b \times G \rightarrow G_b, \psi_b(p, g) = \psi(p, g) = \psi_g(p) = pg$ ($\psi_g \in \text{Diff}(P)$) que induce $\psi_{b*(p,g)} : T_{(p,g)}(G_b \times G) \rightarrow T_{pg} G_b$ y en particular $\psi_{b*(p,e)} : T_{(p,e)}(G_b \times G) \rightarrow T_p G_b$, $\check{\psi}_b$ es el homomorfismo (1-1 pues ψ_b es libre) de álgebras de Lie $\check{\psi}_b : \mathcal{G} \rightarrow C^\infty(TG_b), A \mapsto \check{\psi}_b(A) : G_b \rightarrow TG_b, \check{\psi}_b(A)(p) = (p, \check{\psi}_b(A)_p), \check{\psi}_b(A)_p = \psi_{b*(p,e)} \circ \tau_{(p,e)}^{-1}(0_p, A_e)$ y $\tau_{(p,g)}$ es el isomorfismo de espacios vectoriales $\tau_{(p,g)} : T_{(p,g)}(G_b \times G) \rightarrow T_p G_b \oplus T_g G, \tau_{(p,g)}(u_{(p,g)}) = (\pi_{1*}(p,g)(u_{(p,g)}), \pi_{2*}(p,g)(u_{(p,g)}))$ con $\pi_1 : G_b \times G \rightarrow G_b, \pi_1(p, g) = p, \pi_2 : G_b \times G \rightarrow G, \pi_2(p, g) = g$; es fácil verificar que si $i_g : G_b \rightarrow G_b \times G, i_g(p) = (p, g)$ y $j_p : G \rightarrow G_b \times G, j_p(g) = (p, g)$ entonces $\rho_{(p,g)} : T_p G_b \oplus T_g G \rightarrow T_{(p,g)}(G_b \times G), \rho_{(p,g)}(w_p \oplus t_g) = i_{g*p}(w_p) + j_{p*g}(t_g)$ es el isomorfismo de espacios vectoriales inverso i.e. $\rho_{(p,g)} = \tau_{(p,g)}^{-1}$ y por lo tanto resulta $\varphi_p(A) = \psi_{b*(p,e)}(\rho_{(p,e)}(0_p, A_e)) = \psi_{b*(p,e)}(i_{g*p}(0_p) + j_{p*e}(A_e)) = \psi_{b*(p,e)} \circ j_{p*e}(A_e) = (\psi_b \circ j_p)_*(A_e)$. Es fácil verificar las siguientes propiedades de ω : i) $\omega_p(v_p) = 0$ si $v_p \in H_p$ (ω es una 1-forma "vertical"); ii) $\omega_p(\check{\psi}_b(A)_p) = A \forall p \in G_b$; iii) $\psi_{gq}^*(\omega_q) = Ad_{g^{-1}} \circ \omega_p$ si $q = pg$ (" ω se transforma de acuerdo con la representación adjunta de G ": $Ad : G \rightarrow GL(\mathcal{G}), Ad(g) = Ad_g = A_{g*e}, A_g : G \rightarrow G, A_g(h) = ghg^{-1}, A_g = L_g \circ R_{g^{-1}}$ con L_g (traslación izquierda) y R_k (traslación derecha) difeomorfismos de G dados por $L_g(h) = gh$ y $R_k(h) = hk$; ψ_{gq}^* es la función imagen recíproca ó "pull-back" de 1-formas, $\psi_{gq}^* : T_q^* P \otimes \mathcal{G} \rightarrow T_p^* P \otimes \mathcal{G}, \psi_{gq}^*(\omega_q) : T_p P \rightarrow \mathcal{G}, \psi_{gq}^*(\omega_q)(v_p) = \omega_q(\psi_{g*p}(v_p))$).¹

Si $\Phi(\check{\psi}_b(A)) = \{\varphi_t : G_b \rightarrow G_b, \varphi_t(p) = c_p(t)|c_p(0) = p, \dot{c}_p(t) = \check{\psi}_b(A)_{c_p(t)}\}_{t \in \mathcal{R}} \subset \text{Diff}(G_b)$ es el flujo de $\check{\psi}_b(A)$ (ψ_b libre implica que $\check{\psi}_b(A)$ es completo en $C^\infty(TG_b)$) entonces se tiene $c_p(t) = \psi_b(p, \exp(tA)) = p \exp(tA)$ y $c_{p*t}(\frac{d}{d\lambda}|_t) = \check{\psi}_b(A)_{c_p(t)}$; por lo tanto si $f \in C^\infty(G_b), \check{\psi}_b(A)_p \exp(tA)(f) = c_{p*t}(\frac{d}{d\lambda}|_t)(f) = \frac{d}{d\lambda}|_t(f(p \exp(\lambda A)))$, en particular $\check{\psi}_b(A)_p(f) = \frac{d}{dt}|_0(f(p \exp(tA)))$. Dado que G actúa también libremente sobre P , el resultado vale si se cambia G_b por P i.e.

¹En general si $F : M \rightarrow N$ es una función suave, $y \in N$ y $x \in F^{-1}(\{y\}), F_{yx}^* : T_y^* N \rightarrow T_x^* M, \dots \rightarrow F_{yx}^*(\omega_y) : T_x M \rightarrow \mathcal{R}, v_x \mapsto F_{yx}^*(\omega_y)(v_x) = \omega_y(F_{*x}(v_x))$ i.e. $F_{yx}^*(\omega_y) = \omega_y \circ F_{*x}$. Si F es un difeomorfismo se puede escribir $F_{yx}^* = F_y^*$. Es fácil verificar que si $F : M \rightarrow N$ y $G : N \rightarrow S$ son suaves, $y \in G^{-1}(\{z\})$ y $x \in F^{-1}(\{y\})$ entonces $(G \circ F)_{zx}^* = F_{yx}^* \circ G_{zy}^*$ (contravariancia).

$$\psi(A)_p(f) = \varphi_p(A)(f) = \frac{d}{dt} \Big|_0 (f(p \exp(tA))) \quad (3.1)$$

con $f \in C^\infty(P)$.

Sea $\omega_\alpha = \omega|_{P_\alpha} \in C^\infty(T^*P_\alpha \otimes \mathcal{G})$ la restricción de ω a $P_\alpha = \pi^{-1}(U_\alpha)$ con $(U_\alpha, \varphi_\alpha)$ en el atlas $\mathcal{U} = \{(U_\gamma, \varphi_\gamma)\}_{\gamma \in J}$ de ξ ($\varphi_\alpha : P_\alpha \rightarrow U_\alpha \times G$ es un difeomorfismo de variedades con $\pi_1 \circ \varphi_\alpha = \pi_\alpha = \pi|_{P_\alpha}$: *condición de trivialidad local*); la sección local $\sigma_\alpha : U_\alpha \rightarrow P_\alpha$, $\sigma_\alpha(b) = \varphi_\alpha^{-1}(b, e)$ (función suave 1-1) induce el "pull-back" $\sigma_\alpha^* : C^\infty(T^*P_\alpha \otimes \mathcal{G}) \rightarrow C^\infty(T^*U_\alpha \otimes \mathcal{G})$, $A_\alpha = \sigma_\alpha^*(\omega_\alpha)$ con $A_{\alpha b} = \sigma_{\alpha qb}^*(\omega_{\alpha q}) \in T_b^*U_\alpha \otimes \mathcal{G}$, $q = \sigma_\alpha(b)$. Sea $(U_\beta, \varphi_\beta) \in \mathcal{U}$ con $U_\alpha \cap U_\beta \neq \emptyset$ y sea $b \in U_\alpha \cap U_\beta$; como antes se tiene $A_{\beta b} = \sigma_{\beta pb}^*(\omega_{\beta p}) \in T_b^*U_\beta \otimes \mathcal{G}$, $p = \sigma_\beta(b)$. **La relación de compatibilidad es la relación entre $A_{\alpha b}$ y $A_{\beta b}$.** (Nótese que $\omega_{\alpha q} = \omega_q$, $\omega_{\beta p} = \omega_p$ y $T_bU_\alpha \cong T_b(U_\alpha \cap U_\beta) \cong T_bU_\beta$; ver Figura 5.)

Es importante notar que existe una correspondencia biunívoca entre secciones y sistemas de trivializaciones locales de un *haz fibrado principal* (h.f.p.) $[\xi] = \{(P, B, \pi, G, \{U_\alpha\}_{\alpha \in J}, \psi)\}$ es la clase de equivalencia de h.f.p.c.'s que difieren entre sí a lo más por las funciones $\{\varphi_\alpha\}_{\alpha \in J}$ que dan las trivializaciones locales): en efecto, si $\sigma_\alpha : U_\alpha \rightarrow P_\alpha$ es una sección local, entonces $\psi_\alpha : U_\alpha \times G \rightarrow P_\alpha$, $\psi_\alpha(b, g) := \sigma_\alpha(b)g$ es un difeomorfismo con $\pi \circ \psi_\alpha = \pi_1$, en particular $\psi_\alpha(b, e) = \sigma_\alpha(b)$; por lo tanto el atlas \mathcal{U} del haz puede ser dado en la forma $\{(U_\alpha, \sigma_\alpha)\}_{\alpha \in J}$; nótese también que mientras φ_α "coordiniza" los puntos de P_α ($\varphi_\alpha(p) \in U_\alpha \times G$), σ_α "coordiniza" p -formas diferenciales sobre P_α ($\sigma_\alpha^*(\Gamma) \in \Omega^p(U_\alpha)$ si $\Gamma \in \Omega^p(P_\alpha)$).

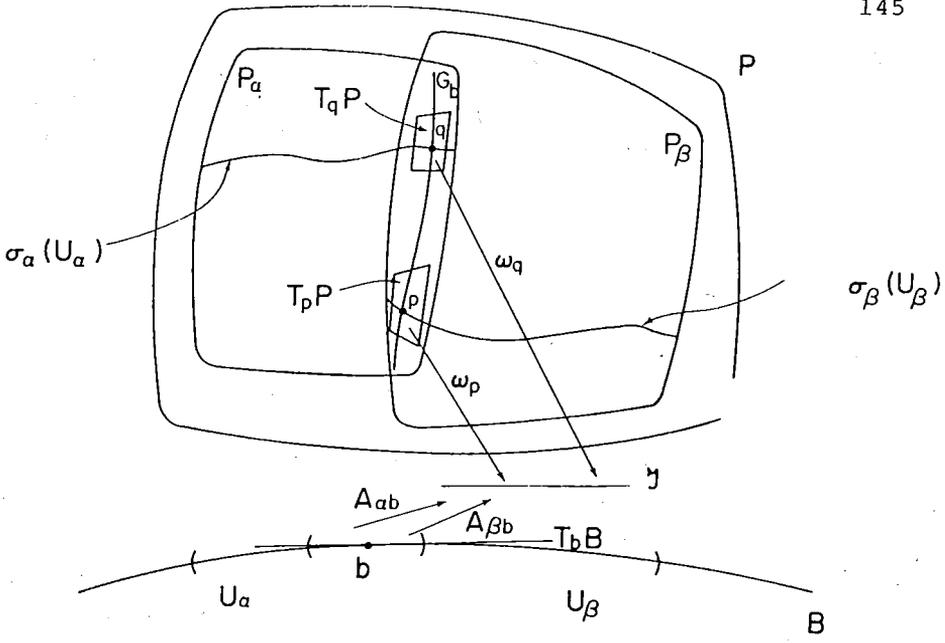


Fig. 5

Dado el carácter libre y transitivo de la acción de G sobre las fibras del haz, existe y es única la función de transición $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow G$ con $\sigma_\beta(b) = \sigma_\alpha(b)g_{\alpha\beta}(b)$; $g_{\alpha\beta}$ cumple $\varphi_\beta \circ \varphi_\alpha^{-1}(b, g) = (b, g_{\beta\alpha}(b)g)$ y $g_{\alpha\beta}g_{\beta\gamma} = g_{\alpha\gamma}$ si $U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset$, de la que se obtiene $g_{\alpha\alpha}(b) = e \forall \alpha \in J$ y $g_{\alpha\beta}(b) = (g_{\beta\alpha}(b))^{-1}$. Se tiene $\sigma_\alpha(b) = \sigma_\beta(b)g_{\beta\alpha}(b) = \psi(\sigma_\beta(b), g_{\beta\alpha}(b)) = \psi \circ (\sigma_\beta \times g_{\beta\alpha})(b)$ i.e. $\sigma_\alpha = \psi \circ (\sigma_\beta \times g_{\beta\alpha})$ (factorización, ver diagrama 1)

$$\begin{array}{ccc}
 & & P \\
 & \nearrow \sigma_\alpha & \uparrow \psi \\
 U_\alpha \cap U_\beta & \longrightarrow & P \times G \\
 & \sigma_\beta \times g_{\beta\alpha} &
 \end{array}$$

Diag. 1

y por lo tanto $A_{\alpha b} = (\psi \circ (\sigma_\beta \times g_{\beta\alpha}))_{q^*}^*(\omega_q) = (\sigma_\beta \times g_{\beta\alpha})_{(p,g)b}^* \circ \psi_{q(p,g)}^*(\omega_q) = (\sigma_\beta \times$

$g\beta\alpha)_{(p,g)b}^*(\psi_{q(p,g)}^*(\omega_q)) = \psi_{q(p,g)}^*(\omega_q) \circ (\sigma_\beta \times g\beta\alpha)_{*b} = (\omega_q \circ \psi_{*(p,g)}) \circ (\sigma_\beta \times g\beta\alpha)_{*b} = (\omega_q \circ \psi_{*(p,g)}) \circ (i_{g*} \circ \sigma_{\beta*b} + j_{p*} \circ g\beta\alpha_{*b})$ donde en la última igualdad se ha usado el resultado del Apéndice. ($g\beta\alpha_{*b} \in T_b^*(U_\alpha \times U_\beta) \otimes T_g G$.) De la linealidad de las 1-formas y las diferenciales resulta

$$\begin{aligned} A_{\alpha b} &= \omega_q \circ (\psi_{*(p,g)} \circ (i_{g*} \circ \sigma_{\beta*b}) + \psi_{*(p,g)} \circ (j_{p*} \circ g\beta\alpha_{*b})) \\ &= \omega_q \circ (\psi_{*(p,g)} \circ (i_{g*} \circ \sigma_{\beta*b})) + \omega_q \circ (\psi_{*(p,g)} \circ (j_{p*} \circ g\beta\alpha_{*b})) \\ &= \omega_q \circ ((\psi \circ i_g)_{*p} \circ \sigma_{\beta*b}) + \omega_q \circ ((\psi \circ j_p)_{*g} \circ g\beta\alpha_{*b}). \end{aligned} \quad (3.2)$$

Analicemos los dos términos por separado: 1er. término: $\psi \circ i_g : P \rightarrow P, \psi \circ i_g(p) = \psi(p, g) = \psi_g(p)$ i.e. $\psi \circ i_g = \psi_g$; luego $(\omega_q \circ \psi_{g*}) \circ \sigma_{\beta*b} = \psi_{gq}^*(\omega_q) \circ \sigma_{\beta*b} = (Ad_{g^{-1}} \circ \omega_p) \circ \sigma_{\beta*b} = Ad_{g^{-1}} \circ (\omega_p \circ \sigma_{\beta*b}) = Ad_{g^{-1}} \circ \sigma_{\beta b}^*(\omega_p) = Ad_{g^{-1}} \circ A_{\beta b}$. 2do. término: si $v_b \in T_b(U_\alpha \cap U_\beta)$ entonces $\exists! v_e \in T_e G$ tal que $g\beta\alpha_{*b}(v_b) = L_{g*} v_e \in T_g G$, a saber $v_e = L_{g^{-1}*} \circ g\beta\alpha_{*b}(v_b) = \Theta_{g\beta\alpha}(b) \circ g\beta\alpha_{*b}(v_b)$ donde $\Theta \in C^\infty(T^*G \otimes \mathcal{G})$ es la forma de Maurer-Cartan sobre G dada por $\Theta(g) = (g, \Theta_g), \Theta_g = L_{g^{-1}*} \circ \Theta_g^* \in T_g^* G \otimes \mathcal{G}; \psi \circ j_p : G \rightarrow P, \psi \circ j_p(g) = \psi(p, g) = \psi_p(g)$ con $\psi_p : G \rightarrow P, \psi_p(g) = pg$; se tiene entonces $(\psi \circ j_p)_{*g} \circ g\beta\alpha_{*b}(v_b) = \psi_{p*} \circ L_{g*} v_e \in T_{pg} P$ y para $f \in C^\infty(P), \psi_{p*} \circ L_{g*} v_e(f) = L_{g*} v_e(f \circ \psi_p) = v_e((f \circ \psi_p) \circ L_g) = v_e(f \circ (\psi_p \circ L_g)) = v_e(f \circ \psi_{pg}) = \frac{d}{dt} \Big|_0 (f \circ \psi_{pg}(\exp(tv_e))) = \frac{d}{dt} \Big|_0 (f(\psi_{pg}(\exp(tv_e)))) = \frac{d}{dt} \Big|_0 (f(pg \exp(tv_e)))$ y comparando con (3.1) después de la identificación $A \equiv v_e$ resulta

$$\psi(A)_{pg} = \psi_{p*} \circ L_{g*} v_e. \quad (3.3)$$

Para el 2do. término de (3.2) se obtiene entonces $\omega_q(\psi_{p*} \circ L_{g*} v_e) = \omega_q(\psi(v_e)_{pg}) = v_e$ donde se ha usado la propiedad ii) de ω . Por lo tanto

$$A_{\alpha b}(v_b) = Ad_{g^{-1}} \circ A_{\beta b}(v_b) + \Theta_g \circ g\beta\alpha_{*b}(v_b)$$

y de la arbitrariedad de v_b

$$\begin{aligned} A_{\alpha b} &= Ad_{g^{-1}} \circ A_{\beta b} + \Theta_g \circ g\beta\alpha_{*b} \\ &= Ad_{g^{-1}} \circ A_{\beta b} + g\beta\alpha_{*b}(\Theta_g) \end{aligned}$$

$$\begin{aligned}
 A_{\alpha b} &= Ad_{g^{-1}} \circ A_{\beta b} + L_{g^{-1}*g} \circ g_{\beta\alpha*b} \\
 &= Ad_{g^{-1}} \circ A_{\beta b} + (L_{g^{-1}} \circ g_{\beta\alpha})_* b \\
 &= Ad_{g^{-1}} \circ A_{\beta b} + d(L_{g^{-1}} \circ g_{\beta\alpha})|_b \in T_b^*(U_\alpha \cap U_\beta) \otimes \mathcal{G}
 \end{aligned}
 \tag{3.4}$$

con $g = g_{\beta\alpha}(b)$ y $g_{\beta\alpha gb} : T_g^*G \otimes \mathcal{G} \rightarrow T_b^*(U_\alpha \cap U_\beta) \otimes \mathcal{G}$, que es la relación de compatibilidad buscada. Usando $Ad_{g^{-1}} = A_{g^{-1}*e} = (L_{g^{-1}} \circ R_g)_*e = L_{g^{-1}*g} \circ R_{g*e}$ y la linealidad de Θ_g , (3.4) se escribe

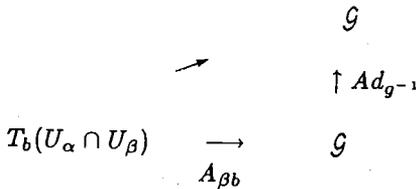
$$A_{\alpha b} = L_{g^{-1}*g} \circ (R_{g*e} \circ A_{\beta b} + g_{\beta\alpha*b})$$

6

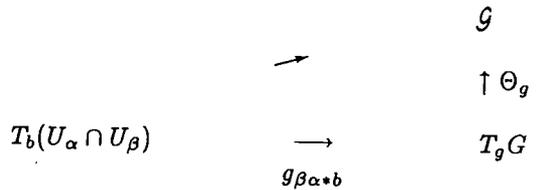
$$L_{g*e} \circ A_{\alpha b} = R_{g*e} \circ A_{\beta b} + g_{\beta\alpha*b}. \tag{3.5}$$

Observaciones:

1) Los dos términos en (3.4) están representados por los siguientes diagramas conmutativos:



Diag. 2 : 1er. término



Diag. 3: 2do. término

2) En (3.4), $L_{g_{\beta\alpha}(b)^{-1}} \circ g_{\beta\alpha} \in C^\infty(U_\alpha \cap U_\beta, G)$, $L_{g_{\beta\alpha}(b)^{-1}}(g_{\beta\alpha}(b')) = g_{\beta\alpha(b)}^{-1} g_{\beta\alpha}(b') (= e$ si $b' = b)$; sin embargo la fórmula es puntual i.e. se tiene una expresión para cada $b \in U_\alpha \cap U_\beta$ y no

tiene sentido una fórmula *local* del tipo $A_\alpha = Ad_{g_{\beta\alpha}^{-1}} \circ A_\beta + (L_{g_{\beta\alpha}^{-1}} \circ g_{\beta\alpha})_* \in C^\infty(T^*(U_\alpha \cap U_\beta) \otimes \mathcal{G})$ ya que $L_{g_{\beta\alpha}^{-1}} = L \circ g_{\beta\alpha}^{-1}$ (y por lo tanto $L_{g_{\beta\alpha}^{-1}} \circ g_{\beta\alpha} = L$) con $L : G \rightarrow Diff(G)$, $L(g) = L_g$ que no es una función suave entre variedades diferenciables de dimensión finita ya que $Diff(G)$ no lo es.

3) Si $\{e_1, \dots, e_r\}$ es una base de \mathcal{G} ($r = \dim G$) y $(U; x^1, \dots, x^s)$ es una carta coordenada sobre B como variedad diferenciable con $b \in U$ ($s = \dim B$), entonces $A_{\alpha b} = \sum_{\mu=1}^s \sum_{i=1}^r A_{\alpha\mu b}^i e_i \otimes dx^\mu|_b$ con $A_{\alpha\mu b}^i \in \mathcal{R}$, y análogo para $A_{\beta b}$. Tomando $v_b = \frac{\partial}{\partial x^\nu}|_b \in T_b(U_\alpha \cap U_\beta)$ resulta

$$A_{\alpha\nu b} = Ad_{g^{-1}}(A_{\beta\nu b}) + \Theta_g(g_{\beta\alpha*b}(\frac{\partial}{\partial x^\nu}|_b)) \tag{3.6}$$

con $A_{\alpha\nu b} = \sum_{i=1}^r A_{\alpha\nu b}^i e^i \in \mathcal{G}$ y $g_{\beta\alpha*b}(\frac{\partial}{\partial x^\nu}|_b) \in T_g G$.

4) Para un grupo de Lie de matrices $G < GL_n(\mathcal{K})$ con $\mathcal{K} = \mathcal{R}, \mathcal{C}$ ó \mathcal{H} , $L_{g*h} = L_g$ y $R_{g*h} = R_g$ son "constantes", independientes de h que se reducen respectivamente a multiplicar por la izquierda y por la derecha por la matriz $g \in G$; para (3.4) se tiene entonces (con $v_b \in T_b(U_\alpha \cap U_\beta)$ y $g = g_{\beta\alpha}(b)$) $A_{\alpha b}(v_b) = L_{g^{-1}*g} \circ R_{g*e}(A_{\beta b}(v_b)) + L_{g^{-1}*g}(g_{\beta\alpha*b}(v_b)) = L_{g^{-1}} \circ R_g(A_{\beta b}(v_b)) + L_{g^{-1}}(g_{\beta\alpha*b}(v_b)) = g^{-1} A_{\beta b}(v_b) g + g^{-1} g_{\beta\alpha*b}(v_b) = g_{\beta\alpha}(b)^{-1} A_{\beta b}(v_b) g_{\beta\alpha}(b) + g_{\beta\alpha}(b)^{-1} dg_{\beta\alpha}|_b(v_b)$; tomando $v_b = \frac{\partial}{\partial x^\nu}|_b$ y usando $A_{\beta b} = A_{\beta\nu b}^i e_i \otimes dx^\nu|_b$ y $dg_{\beta\alpha}|_b = \frac{\partial}{\partial x^\nu}|_b(g_{\beta\alpha}) dx^\nu|_b$ resulta

$$A_{\alpha\mu b} = g_{\beta\alpha}(b)^{-1} A_{\beta\mu b} g_{\beta\alpha}(b) + g_{\beta\alpha}(b)^{-1} \frac{\partial}{\partial x^\mu}|_b(g_{\beta\alpha}) \in \mathcal{G}. \tag{3.7}$$

Si el grupo es abeliano,

$$A_{\alpha\mu b} = A_{\beta\mu b} + g_{\beta\alpha}(b)^{-1} \frac{\partial}{\partial x^\mu}|_b(g_{\beta\alpha}). \tag{3.8}$$

5) Consideremos el h.f.p. $[\xi] = (P, B, \pi, G, U, \psi)$, $U = \{U_\alpha\}_{\alpha \in J}$; la pareja $([\xi], \omega)$ es un objeto en la categoría de h.f.p.'s con conexiones. Sea el difeomorfismo $\varphi : P \rightarrow P$ una transformación de norma global i.e. un elemento del grupo de automorfismos verticales $Aut_B P(B, G)$ de $[\xi]$. Se tiene el siguiente diagrama conmutativo:

$$\begin{array}{ccccccc}
 & & \varphi \times 1_G & & & & \\
 & & \longrightarrow & & & & \\
 P \times G & & & & & & P \times G \\
 & & \psi \downarrow & & & & \downarrow \psi \\
 & & & & & & \\
 T^*P \otimes \mathcal{G} & \xleftarrow{\varphi^*(\omega)} & P & \xrightarrow{\varphi} & P & \xrightarrow{\omega} & T^*P \otimes \mathcal{G} \\
 & & \pi \searrow & & \swarrow \pi & & \\
 & & & & & & \\
 & & & & & & B
 \end{array}$$

Diag. 4

Si C es el conjunto (espacio afín) de conexiones sobre $P, \Psi : C \times \text{Aut}_B P(B, G) \rightarrow C, \Psi(\omega, \varphi) := \varphi^*(\omega)$ es una acción derecha de $\text{Aut}_B P(B, G)$ sobre C y define las transformaciones de norma globales de las conexiones en el espacio total del haz. ($\tilde{\Psi} : \text{Aut}_B P(B, G) \rightarrow \text{Perm}(C), \tilde{\Psi}(\varphi)(\omega) := \varphi^*(\omega)$ es un anti-homomorfismo de grupos y por lo tanto una *representación* de $\text{Aut}_B P(B, G)$ en las permutaciones (biyecciones) de C .) Si tomamos $\xi \in [\xi]$ i.e. si elegimos un sistema de trivializaciones locales, asociadas a las parejas $([\xi], \omega)$ y $([\xi], \varphi^*(\omega))$ se tienen respectivamente las “parejas” $(\xi, \{A_\alpha = \sigma_\alpha^*(\omega)\}_{\alpha \in J} + \text{relaciones (3.4)})$ y $(\xi, \{A'_\alpha = \sigma_\alpha^*(\varphi^*(\omega))\}_{\alpha \in J} + \text{relaciones (3.4)})$; en cada abierto U_α , la relación entre A_α y A'_α es lo que en Física se entiende por **transformación de norma de los potenciales**: dado que $\sigma_\alpha^* \circ \varphi^*(\omega) = (\varphi \circ \sigma_\alpha)^*(\omega) = \sigma'_\alpha^*(\omega)$ con $\sigma'_\alpha : U_\alpha \rightarrow P_\alpha, \exists! g_\alpha : U_\alpha \rightarrow G \mid \sigma'_\alpha = \sigma_\alpha g_\alpha$ i.e. $\sigma'_\alpha(b) = \sigma_\alpha(b)g_\alpha(b) \forall b \in U_\alpha$; luego $A'_\alpha = (\sigma_\alpha g_\alpha)^*(\omega)$ y por lo tanto la fórmula de transformación es análoga a (3.4) pero con $g_{\beta\alpha}$ reemplazado por g_α i.e.

$$A'_{\alpha b} = \text{Ad}_{g_\alpha(b)^{-1}} \circ A_{\alpha b} + \Theta_{g_\alpha(b)} \circ g_{\alpha+b} \in T_b^* U_\alpha \otimes \mathcal{G}. \tag{3.9}$$

Para un grupo de Lie de matrices

$$A'_{\alpha\mu b} = g_\alpha(b)^{-1} A_{\alpha\mu b} g_\alpha(b) + g_\alpha(b)^{-1} \frac{\partial}{\partial x^\mu} \Big|_b (g_\alpha) \in \mathcal{G} \tag{3.10}$$

y en particular para el caso abeliano

$$A'_{\alpha\mu b} = A_{\alpha\mu b} + g_\alpha(b)^{-1} \frac{\partial}{\partial x^\mu} \Big|_b (g_\alpha). \tag{3.11}$$

Un ejemplo importante es el caso del *electromagnetismo* en el que $G = U(1)$, $\mathcal{G} = u(1) = i\mathcal{R}$, $A_{\alpha\beta\mu} = ia_{\alpha\beta\mu} = ia_{\alpha\mu}(b)$ y $g_\alpha = e^{i\Lambda_\alpha}$ con $\Lambda_\alpha : U_\alpha \rightarrow \mathcal{R}$ y $A_{\alpha\mu} : U_\alpha \rightarrow \mathcal{R}$ funciones suaves; se tiene entonces

$$a'_{\alpha\mu} = a_{\alpha\mu} + \frac{\partial}{\partial x^\mu}(\Lambda_\alpha). \quad (3.12)$$

6) Consideremos el caso en que el haz $[\xi]$ es *trivial* i.e. P es isomorfo al haz producto $B \times G$; en particular consideremos $P = B \times G$. Sea $\sigma : B \rightarrow B \times G$, $b \mapsto \sigma(b) := (b, \Sigma(b))$ una sección global ($\Sigma : B \rightarrow G$ es la función suave que determina σ ; en este caso el atlas del h.f.p.c. es $\mathcal{U} = \{(B, \varphi)\}$ con $\varphi^{-1} : B \times G \rightarrow B \times G = P$, $\varphi^{-1}(b, g) = \sigma(b)g = (b, \Sigma(b)g)$, en particular $\varphi^{-1}(b, e) = \sigma(b)$; si $\omega \in C^\infty(T^*(B \times G) \otimes \mathcal{G})$ es la forma de una conexión H en $[\xi]$, $A = \sigma^*(\omega) \in C^\infty(T^*B \otimes \mathcal{G})$. Un elemento f de $Aut_B B \times G(B, G)$ está dado por $f : B \times G \rightarrow B \times G$, $f(b, g) = (b, \varphi_f(b, g)) = (b, \varphi_f(b, e)g) = (b, \alpha(f)(b)g)$ donde $\alpha : Aut_B B \times G(B, G) \rightarrow C^\infty(B, G)$, $\alpha(f) : B \rightarrow G$, $\alpha(f)(b) = \varphi_f(b, e)$ es un isomorfismo de grupos ($(C^\infty(B, G), \cdot)$ con $\alpha \cdot \beta(b) = \alpha(b)\beta(b)$, $\alpha^{-1}(b) = \alpha(b)^{-1}$ (inversa) y $\varepsilon(b) = e$ (identidad) es el *grupo de norma usual en las aplicaciones físicas* en que B , el espacio-tiempo, es un espacio topológicamente contractible i.e. del mismo tipo de homotopía que un punto, por ejemplo el espacio euclideo E^4 ó el de Minkowski M^4). La transformación global de la conexión está dada por $\Psi(\omega, f)f^*(\omega)$ y para la proyección en el espacio base se tiene $A' = \sigma^*(f^*(\omega)) = (f \circ \sigma)^*(\omega) = \sigma'^*(\omega)$ (ver Figura 6) con $\sigma' = f \circ \sigma = \sigma \cdot \gamma$ y $\gamma : B \rightarrow G$ dada por $\sigma'(b) = f(\sigma(b)) = f(b, \Sigma(b)) = (b, \varphi_f(b, e)\Sigma(b)) = (b, \alpha(f)(b)\Sigma(b)) = \sigma \cdot \gamma(b) = (b, \Sigma(b)\gamma(b))$ i.e. $\gamma(b) = \Sigma(b)^{-1}\alpha(f)(b)\Sigma(b) = \Sigma^{-1} \cdot \alpha(f) \cdot \Sigma(b)$.

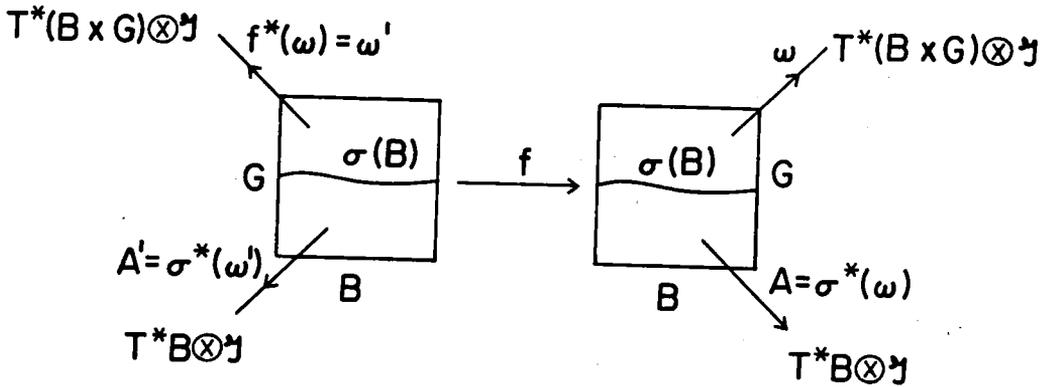


Fig. 6

Análogamente a (3.9) se tiene

$$A'_b = Ad_{(\Sigma^{-1} \cdot \alpha(f) \cdot \Sigma(b))^{-1}} \circ A_b + \Theta_{\Sigma^{-1} \cdot \alpha(f) \cdot \Sigma(b)} \circ (\Sigma^{-1} \cdot \alpha(f) \cdot \Sigma)_{*b} \cdot \quad (3.13)$$

En particular si $\Sigma(b) = e, \mathcal{U} = \{(B, 1_{B \times G})\}$ y

$$A'_b = Ad_{\alpha(f)(b)^{-1}} \circ A_b + \Theta_{\alpha(f)(b)} \circ \alpha(f)_{*b} \cdot \quad (3.14)$$

Agradecimientos

El autor agradece a los organizadores del VII Coloquio del Depto. de Matemáticas del CINVESTAV, y en especial al Dr. Luis Gorostiza, por la oportunidad de presentar en dicho evento la conferencia "Geometría y Física", que dió origen al presente trabajo.

Referencias

1. *Science in the 20th Century*, Scientific American, Special Issue Vol. 3, Nr. 1, 1991; P. Langacker y A. K. Mann, "The unification of electromagnetism with the weak force", *Phys. Today*, Dec. 1989, 22-31.
2. N. Steenrod, *The topology of fibre bundles* (Princeton Univ. Press, 1951); D. Husemoller, *Fiber Bundles*, GTM 20 (Springer-Verlag, 1975); *Fiber Bundles: Their use in Physics*, Eds. J. P. Ezin y A. Verjovsky (World Scientific, 1988).
3. A. Trautman, "Fibre bundles associated with space-time", *Rep. Math. Phys.* 1, 29-62 (1970); W. Drechsler y M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Lecture Notes in Phys. Vol. 67, Springer Verlag (Heidelberg, 1977); M. F. Atiyah, *Geometry of Yang-Mills Fields*, Lezioni Fermiane, Accad. Nazionale dei Lincei, Scuola Normale Superiore (Pisa, 1979); M. Rosenbaum *et al*, "Geometric model for gravitational and electroweak interactions", *J. Math. Phys.* 30, 1579-1590 (1989); E. Nahmad A., "The Magic of Geometry", prep. ICN - UNAM, 1991.
4. M. Daniel y C. M. Viallet, "The geometrical setting of gauge theories of the Yang-Mills type", *Rev. Mod. Phys.* 52, 175 - 197 (1980); C. M. Viallet, "Symmetry and Functional Integration" (pgs. 435 - 460) en *Physics, Geometry and Topology*, ed. H. C. Lee, Plenum Press (New York, 1990).
5. R. P. Feynman y A. R. Hibbs, *Quantum Mechanics and Path Integrals* (Mc Graw - Hill, New York, 1965).
6. I. M. Singer, *Phys. Scripta* 24, 817 (1981).
7. M. Grabiak, B. Müller y W. Greiner, "Geometrical properties of gauge theories", *Ann. of Phys.* 172, 213 - 242 (1986).
8. F. A. Berezin, *The method of Second Quantization*, Academic Press (New York, 1966).
9. I. M. Singer, "Some remarks on the Gribov ambiguity", *Comm. Math. Phys.* 60, 7 - 12 (1978).
10. M. Socolovsky, "Ghost fields in principal bundles" pgs. (204 - 210), en *Relativity and Gravitation : Classical and Quantum*, Proceedings of the 7th Latin-American Symposium on Relativity and Gravitation, eds. J. C. D'Olivo *et al* (World Scientific, 1991).
11. L. S. Schulman, "Approximate topologies", *J. Math. Phys.* 12, 304 - 308 (1971); *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

12. H. J. Bernstein y A. V. Phillips, "Fiber bundles and quantum theory", *Scient. Amer.* **245**, 122-137, July 1981.
13. M. Socolovsky, "Spin, monopole, instanton and Hopf bundles", ICN- UNAM/ CINVESTAV-IPN Prep. 1991, en *Memorias del 1er. Seminario Internacional de Algebra y sus Aplicaciones*, ESFM-IPN, Méx., Ene. 1991.
14. I. M. Singer y J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Scott, Foresman Co., 1967.
15. M. Porter, *Superficies de Riemann*, 3er. Coloquio, Depto. de Matemáticas, CINVESTAV-IPN, 1983.
16. C. J. Isham, *Modern Differential Geometry for Physicists*, (World Scientific, 1989).
17. A. Trautman, *Differential Geometry for Physicists*, Bibliopolis (Napoli, 1984).
18. C. Nash y S. Sen, *Topology and Geometry for Physicists* (Academic Press, 1983).
19. M. Socolovsky, "Gauge transformations and fiber bundle theory", *J. Math. Phys.* **32**, 2522-2531 (1991).

Apéndice

Proposición: Sean $\varphi : X \rightarrow Y$ y $\psi : X \rightarrow Z$ funciones suaves y sea $\varphi \times \psi : X \rightarrow Y \times Z$, $\varphi \times \psi(x) = (\varphi(x), \psi(x))$. Entonces $(\varphi \times \psi)_{*x} : T_x X \rightarrow T_x(\varphi \times \psi)(Y \times Z)$ está dada por $(\varphi \times \psi)_{*x} = (i_{\psi(x)} \circ \varphi_{*x} + j_{\varphi(x)} \circ \psi)_{*x}$ donde $i_z : Y \rightarrow Y \times Z$, $i_z(y) = (y, z)$ y $j_y : Z \rightarrow Y \times Z$, $j_y(z) = (y, z)$.

LAS MATEMATICAS EN LA FISILOGIA MODERNA

Fermín Valenzuela y Patricia Gorostiza

Departamento de Fisiología

Instituto Nacional de Cardiología "Ignacio Chávez"

México D.F.

La fisiología moderna, al igual que el resto de las llamadas ciencias biológicas, ha incorporado desde hace relativamente poco tiempo a las matemáticas dentro de sus quehaceres.

En el proceso de desarrollo de las ciencias biológicas el inicio es el establecer clasificaciones tanto morfológicas como funcionales. Este tipo de investigación fue de carácter fundamentalmente descriptivo. Los seres vivos presentan una gran variabilidad y la actividad taxonómica tuvo que verse apoyada por los métodos estadísticos, para establecer los límites de las clasificaciones. Es éste el primer contacto de los biólogos, en el sentido extenso, con las matemáticas.

En un segundo proceso las preguntas biológicas evolucionaron y se empezaron a establecer relaciones entre los fenómenos. De manera indiscutible, para poder establecer una relación, se debe tener un modelo teórico que se confronte con la realidad. Este modelo teórico no siempre es un modelo formal, y éste fue el caso de los primeros modelos biológicos. Una característica de estos es el hecho de que profundizaban en la caracterización del fenómeno estudiado y se refinaron los métodos de medición. Un ejemplo de este tipo de investigación es la realizada por Bernstein, quien establece a finales del siglo pasado la teoría de la membrana

celular, basado en mediciones muy finas sobre la velocidad de conducción nerviosa y las características del fenómeno eléctrico nervioso. Los postulados de Bernstein dan lugar sin duda, a la nueva electrofisiología.

Junto con el establecimiento de relaciones fenomenológicas se dio también un proceso de penetración en las estructuras, que para muchos estudiosos de la filosofía de la ciencia, fue considerada como la etapa reduccionista de la biología. Durante esta etapa se acumuló y se sigue acumulando un gran cuerpo de conocimientos sobre los procesos vitales. Se pasa del análisis de las características y funciones de los organismos en su conjunto al estudio de sus estructuras constitutivas, llegando hasta los niveles moleculares. Este aumento en la cantidad y la profundidad de los conocimientos ha llevado a una refinación de los modelos. Estos pasaron de ser modelos mecanicistas muy simples a modelos extraordinariamente complejos y en muchos casos, formales.

Los sistemas vivos, son sistemas dinámicos, muy complejos y con un gran número de elementos de control, lo que hace que los modelos formales no puedan elaborarse de manera sencilla con las teorías de control derivadas del álgebra y de la geometría. Esto no impidió que se desarrollaran modelos matemáticos para sistemas biológicos desde hace muchos años, como son los de Lotka (1924) y Volterra (1931). Modelos matemáticos más actuales fueron desarrollados por Rashevsky (1940) y Turing (1952), utilizando muchos de los desarrollos de Poincaré (1882). Otros modelos como los desarrollados por Rosenblueth y Wiener (1945) o por Hodgkin y

Huxley (1952) describían en términos matemáticos una serie de fenómenos biológicos, que aunque cumplían una serie de requisitos desde el punto de vista del modelo, no sucedía lo mismo con la realidad biológica.

No fue sino hasta el advenimiento de los esquemas de bifurcación dinámica desarrollados por Thom (1967)¹ que se contó con un marco adecuado para la realización de modelos biológicos. El desarrollo de la teoría de la dinámica no lineal trajo un problema distinto para la elaboración de los modelos biológicos que ha sido el de la integración de los conocimientos acumulados. La dinámica no lineal y el concepto de oscilación están íntimamente relacionados con el surgimiento del modelaje matemático. Esto requiere explicaciones más sofisticadas que las intuitivas y literarias ofrecidas para conceptos tan complejos como el de la homeostasis, y al mismo tiempo la necesidad de entender que aún cuando los sistemas biológicos están formados por estructuras muy complejas que contienen un gran número de grados de libertad, éstas adoptan siempre formas de baja dimensión con una alta estabilidad, característica principal de estos sistemas. Esta forma estable (atractor) persiste aún cuando pueda haber pérdidas estructurales, aumento de carga, etc.. Estos modos organizativos pueden ser encontrados en cualquiera de los niveles de estudio: desde el nivel molecular hasta el nivel social y se pueden expresar como un sistema que contiene una serie de osciladores estables acoplados, forma que tiene una expresión formal bien establecida en la teoría de los sistemas dinámicos².

Existen varios sistemas sobre los cuales se han establecido modelos basados en la dinámica no-lineal, dos de los más estudiados son el sistema nervioso y el corazón, aunque también existen intentos de modelaje del sistema endócrino y más recientemente del aparato locomotor.

El corazón como sistema a modelar

En lo que respecta al corazón se ha estudiado fundamentalmente con este enfoque lo que concierne a su actividad eléctrica. El corazón tiene una serie de características, que mencionaremos brevemente, que lo convierten en un órgano fascinante para el modelaje. La principal función del corazón es la de bombear sangre oxigenada hacia todos los tejidos y hacer que esta sangre circule también por su sitio de oxigenación (los pulmones). Una característica muy importante de este órgano es su capacidad para adecuar su funcionamiento a los requerimientos del organismo, que varían continuamente según la actividad física; esto se consigue por medio de reajustes de la frecuencia cardíaca y del volumen de sangre que bombea en cada latido.

En un corazón sano, este proceso, que parece sencillo, involucra una gran cantidad de subsistemas interrelacionados con capacidad de reajuste simultánea. Todos estos subsistemas se comportan de manera periódica: el latido del corazón es periódico, lo que implica que su proceso excitatorio (eléctrico) es también un fenómeno periódico. El estímulo eléctrico comienza siempre en un sitio determinado y viaja por vías precisas a manera de estar excitando a todas las células del músculo cardíaco, siguiendo un

patrón de activación bien definido para conseguir una contracción sincrónica, hasta que en menos de un segundo se completa el ciclo e inmediatamente comienza otro, que se repite de manera casi idéntica. Este proceso se repite de manera constante desde antes del nacimiento hasta la muerte del sujeto, sufriendo continuos ajustes. La periodicidad de este fenómeno no se encuentra solamente en los latidos sucesivos, en las vías de conducción de la actividad eléctrica, en el abrir y cerrar de válvulas; dentro de cada célula cardíaca se dan una gran cantidad de eventos cíclicos. La contracción del músculo cardíaco como masa está dada por la contracción sincrónica de cada célula; estas para contraerse llevan a cabo eventos rítmicos: la contracción de una célula muscular esta dada por la interacción de las proteínas contráctiles que están en su interior, dispuestas como un arreglo de bandas paralelas que se acercan mecánicamente entre si, para conseguir un acortamiento en la longitud celular (contracción), y se alejan para alargarla (relajación). Este fenómeno mecánico esta dado por reacciones químicas de formación y ruptura cíclica de enlaces entre átomos, tanto para mover a las proteínas contráctiles, como para producir la energía para que esto se lleve a cabo.

Los mecanismos responsables del inicio y fin de estos procesos también son cíclicos y dependen del flujo de moléculas con carga eléctrica que atraviesan la membrana celular a través de pequeños y numerosos poros o canales. Las características de cada flujo de iones que participan en cada momento de la contracción y la relajación varían de manera idéntica en cada latido.

La excitación del músculo cardíaco es un proceso eléctrico que involucra a todas las células del corazón. Este proceso de conducción del estímulo es posible porque las células se encuentran conectadas entre si, por puentes de baja resistencia eléctrica, que permiten que se propague la onda excitatoria, viajando de una célula a otra. La velocidad con la que esta onda se propaga no es homogénea, ya que existen vías preferenciales que poseen una alta velocidad de conducción. Esta heterogeneidad en el proceso de conducción hace que existan isócronas heterogéneas. La excitación rítmica del corazón comienza en la parte superior de la aurícula derecha en una zona denominada nodo sinusal. Esta porción posee la cualidad de ser automática, lo que significa que no necesita de ningún estímulo externo para generar sus impulsos. Del nodo sinusal la onda excitatoria viaja a las aurículas, lo que provoca que estas se contraigan y vacíen su contenido hacia los ventrículos, para que estos se llenen de manera completa. Este proceso tiene que suceder antes que los ventrículos se contraigan.

Entre las aurículas y los ventrículos el único punto de contacto eléctrico es una estructura llamada nodo aurículo-ventricular, que se caracteriza por tener una velocidad de conducción muy lenta. Esta lentitud del proceso excitatorio permite que se produzca un retardo entre la conducción auricular y la ventricular.

La primera descripción del proceso excitatorio de las aurículas fue dada por Lewis, quien propuso que la excitación auricular eran como "ondas que se producen en un estanque al dejar

caer una piedra sobre su superficie". Este modelo de propagación que implica la existencia de isócronas homogéneas, fue corregido posteriormente, ya que mediciones de la temporalidad del proceso mostraron la heterogeneidad de las isócronas, debido, como ya se mencionó, a la existencia de vías preferenciales de conducción, llamados haces internodales, porque parten del nodo sinusal y se dirigen hacia el nodo aurículo-ventricular.

El nodo aurículo-ventricular es una estructura que se ha propuesto funciona como un sumador. Esto es va sumando los estímulos que le llegan hasta alcanzar un punto en el que se excita. A este punto se le conoce como umbral de excitación.

La actividad eléctrica de los tejidos cardíacos depende de la transferencia de cargas entre el interior y el exterior de las células. Las partículas que transfieren la carga son iones, que se mueven de acuerdo a gradientes electroquímicos que se establecen a ambos lados de la membrana celular y de acuerdo a las características de la misma membrana, que contiene canales que poseen la propiedad de abrirse o cerrarse en función del voltaje y del tiempo y que permiten el paso selectivo de los iones.

Si todos estos mecanismos oscilatorios, no lo fueran o no estuvieran en perfecta coordinación entre ellos y presentes en todas las células, el corazón no podría cumplir su función de bombeo rítmico de la sangre hacia el organismo. Existen numerosas patologías, a todos los niveles de organización del sistema en donde la coordinación entre los mecanismos oscilatorios se pierde. Un ejemplo de la pérdida de sincronía del proceso, lo constituyen

las arritmias cardíacas, en donde el fenómeno eléctrico que es oscilatorio periódico se convierte en aperiódico (caos)³.

Uno de los primeros modelos matemáticos para explicar la actividad eléctrica de los tejidos, fue la propuesta por Hodgkin y Huxley⁴ en 1952, en la que se establecía un modelo que explicaba la apertura y cierre de los canales de la membrana celular que permiten el paso de los iones a través de la membrana y que son en última instancia los responsables de que se produzca un cambio de potencial y se genere una señal eléctrica. En este modelo se propone la existencia de poros con compuertas de activación e inactivación, cuya actividad está determinada por el voltaje de la membrana y por el tiempo, con cinéticas de primer orden, siendo la cinética global del poro la suma de las cinéticas de cada compuerta. Este modelo contiene una serie de ecuaciones diferenciales cuya solución solo es posible mediante integración numérica.

Basados en el modelo de Hodgkin y Huxley, se desarrollaron una serie de modelos para la actividad eléctrica del corazón^{5,6}, que han prevalecido por más de 30 años e incluso algunos siguen siendo utilizados para la interpretación de algunos datos. Sin embargo con el advenimiento de técnicas de registro más sensibles, estos modelos han mostrado una serie de errores, siendo quizá el principal el considerar la actividad de los canales iónicos de una manera determinística. Los nuevos modelos, basados en técnicas experimentales que permiten el registro de la actividad de un solo canal y no únicamente la resultante final de la actividad de todos

los canales⁷, han mostrado que los poros o canales se comportan de forma estocástica, con cambios en la probabilidad de apertura y cierre de acuerdo al voltaje, pero no se puede ya seguir hablando de una determinación del estado del canal por voltaje o tiempo. Cabe mencionar que aún cuando los modelos originales han sido rebasados, estos permitieron la postulación y la posterior búsqueda de los canales iónicos, los cuales hasta hace relativamente poco tiempo, no pasaban de ser una entelequia.

Una de las críticas más constante a los modelos biofísicos es que son "reduccionistas" y que no contemplan la función del todo (holismo), sin embargo han permitido la penetración de la estructura tanto funcional como morfológica e incrementado el entendimiento de la función de las partes. Un modelo holístico no puede ser elaborado sin el conocimiento de la función de las partes. Es cierto también que muchos de los modelos biofísicos no permiten la extrapolación de resultados a niveles superiores de integración.

Un fenómeno similar a lo que se observa en la función eléctrica, se puede observar en lo referente a la actividad contráctil, en el sentido de que existen modelos subcelulares que incorporan la función de las proteínas contráctiles, el papel de algunos iones y enzimas. Por otro lado también existen modelos que contemplan la contracción global del corazón, de tipo topológico que no incorporan ni los eventos eléctricos que dan lugar al proceso de excitación del proceso contráctil ni los elementos extracardíacos que modulan la función mecánica.

En resumen se podría argumentar la existencia de dos tipos de modelos distintos: los biofísicos o reduccionistas, cuyo propósito ha sido y sigue siendo el de ofrecer un entendimiento detallado de los mecanismos subyacentes a las funciones biológicas además de tener un propósito heurístico que oriente los pasos de nuevas investigaciones. Por otro lado también existen modelos globales, que en su propósito totalizador (holístico) dejan de lado una serie de elementos importantes para el entendimiento de la función y cuya correspondencia matemática es alta pero la biológica está muy por debajo de lo deseado.

Modelos recientes basados en la matemática no lineal, han incorporado elementos tales como la complejidad y el establecimiento de ligaduras entre todos los niveles desde el subcelular hasta el social⁸. Esto se ha logrado mediante la identificación de los parámetros fundamentales de cada nivel, que permiten establecer modelos de bajo perfil.

Uno de los propósitos fundamentales de estos modelos globales, de bajo perfil es el de fijar ideas básicas, clarificar los elementos principales de cada nivel y por sobre todo ha permitido, mediante el establecimiento de un nuevo enfoque (no un nuevo método), plantear preguntas con postulados distintos. Esto último ha abierto un nuevo panorama en el estudio de los fenómenos biológicos, en particular de los fenómenos cardíacos, que como ya se mencionó tienen una conducta oscilatoria. Este nuevo enfoque permite modelar e incluso predecir muchas de las características observadas tanto en el paciente en su conjunto como en la célula

individual aislada. Se puede afirmar que estos modelos han provocado una verdadera ruptura epistemológica, en el sentido de que han modificado la visión que de los modelos se tenía con respecto a la incorporación de tantos elementos como fuera posible manejar. En la actualidad en lugar de incorporar todos los elementos se pretende identificar los elementos más significativos de cada nivel y las ligaduras interniveles.

Se puede afirmar que las tendencias actuales del modelaje matemático de los sistemas biológicos en general pretenden establecer modelos sencillos, que consideren principalmente la conducta de los seres vivos, basados en sus características esenciales, sin incorporar cada uno de los elementos de cada nivel. La consideración básica es que los sistemas biológicos son altamente complejos y que la incorporación de todos sus elementos hacen imposible un modelaje adecuado. Esto es lo que se conoce como la búsqueda de modelos de bajo perfil.

Por otro lado se buscan también los elementos de correspondencia entre los distintos niveles del conjunto, de modo tal que se establezcan las ligas entre niveles y los sistemas moduladores. Este elemento en el modelaje matemático de los sistemas vivos se convierte en esencial, porque las conductas de los sistemas se encuentran controladas por un gran número de factores tanto internos como externos al sistema mismo. En la teoría de las oscilaciones esto es lo que obliga al uso de más de un oscilador de ciclo límite estable acoplado. Es el acople de fase entre dos o más osciladores lo que permite incorporar algunos de

los mecanismos controladores dentro de los modelos⁹. Por otro lado y debido a la variación biológica presente en estos sistemas, (variaciones estocásticas) muchas de las predicciones establecidas por los modelos matemáticos son difíciles de probar experimentalmente, lo que obliga a simplificar los modelos e incorporar técnicas experimentales que permitan eliminar elementos de ruido externo. La eliminación de ruido permite establecer la existencia de osciladores, bifurcaciones, transición de fases entre estados y la identificación de las variaciones estocásticas. Si el ruido externo no es eliminado el establecimiento de los elementos de análisis no se puede efectuar y el modelo contendrá necesariamente gran cantidad de elementos de incertidumbre.

En términos de dinámica compleja, podemos afirmar que lo que interesa al establecer un modelo es la determinación de los vectores, identificar el estado espacio y descomponer éste en sus conjuntos abiertos, determinando el atractor para cada uno. Este atractor es lo que conocemos como el estado de equilibrio dinámico. El otro elemento importante es el sitio("locus") de atracción del sistema en su conjunto. Así como existen conjuntos de atractores que determinan el comportamiento de un sistema, existen también conjuntos de sistemas que determinan a una población mayor. Esta característica de los sistemas dinámicos es lo que la hace muy atractiva para el tratamiento formal de los fenómenos biológicos. El proceso de acoplamiento de dos o más sistemas es uno de los problemas que existen en este modelaje, pero se avanza rápidamente conform .dentifican los elementos esenciales de los sistemas y

se determinan las ligaduras; estos procesos de identificación son esenciales para la determinación del factor dinámico común, que es quien permitirá el establecimiento de un ciclo serial de acuerdo a la teoría de las dinámicas complejas.

Modelaje ¿para qué?

Desde el punto de vista del empleo "práctico" de los conocimientos que emergen de estos conceptos en medicina, está el estudio de los procesos patológicos. Hasta hace relativamente poco tiempo, se estudiaba el estado de salud (normal) y el estado de enfermedad (anormal), pero poco se conocía e incluso poco se estudiaba el proceso mediante el cual un sistema pasaba de su estado normal a sus estado de anormalidad o enfermedad. El empleo de la teoría de los sistemas dinámicos ofrece la posibilidad del estudio de la transición y de los elementos que pueden perturbar un sistema de forma tal que pasen de un estado dinámico estable a un proceso caótico. Un ejemplo de esto es el estudio de las perturbaciones de la actividad eléctrica rítmica del corazón que provoca que se dé la transición de un fenómeno periódico a un fenómeno caótico (arritmia). Este mismo tratamiento se está empleando para establecer los patrones de progresión de enfermedades relacionadas con el sistema endócrino y algunas alteraciones del sueño.

Uno de los problemas fundamentales en el modelaje de los sistemas biológicos es la carencia de interacción entre el matemático y el fisiólogo o el médico. Esta falta de comunicación ha generado una serie de problemas derivados de la falta de

información, fisiológica por un lado y matemática por el otro. Otro problema imposible de resolver sin la interacción de las dos ramas es el concerniente a la validación de los modelos propuestos. Debe tenerse presente que al formular un modelo se hace con un propósito bien definido, definición que es labor del profesional de las ciencias biológicas, que se convertirá en el usuario del modelo. El objetivo del modelo debe tener límites bien definidos, lo que permite asegurar que el modelo será no una representación fiel, en el sentido amplio del término, sino una formulación adecuada para un conjunto bien definido de objetivos.

Validación de los modelos

Un modelo no está completo sin un proceso de validación, para que este proceso se realice los criterios validatorios deben ser explícitos. Entre muchos criterios se puede mencionar a la consistencia y a la validez de los algoritmos empleados. Estos dos criterios aseguran la ausencia de contradicciones funcionales, de acuerdo a la realidad biológica y el empleo de algoritmos que produzcan soluciones aceptables para el problema a modelar.

Por otro lado, se debe buscar también la validez teórica y empírica del modelo, esto es, que sea consistente con las teorías aceptadas y que sus resultados sean concordantes con los datos experimentales disponibles. Otro criterio importante es el heurístico, que permite el proponer en forma de hipótesis, para ser probadas por el modelo propuesto, la existencia y comportamiento de subsistemas, difíciles de explorar en forma experimental.

Por último, un criterio que no se puede pasar por alto es la

necesidad de que el modelo propuesto ofrezca información útil, tanto en la predicción como en el análisis de fenómeno reales, de una manera más o menos certera, de forma tal que el modelo se convierta en una herramienta útil en el proceso de toma de decisiones. A esto se le podría denominar como el criterio pragmático del modelaje.

En conclusión podemos afirmar que en el momento actual se cuenta con los elementos suficientes para iniciar un proceso más global de modelaje de los fenómenos biológicos. Que este proceso de modelaje, aunque no exento de problemas, puede llevar a planteamientos distintos de los tradicionales, que abran nuevos caminos para el entendimiento y estudio de fenómenos complejos, de una manera simple e identifiquen parámetros desconocidos. Esto incrementaría nuestro entendimiento y modificaría los enfoques de abordaje de los problemas biológicos.

REFERENCIAS BIBLIOGRAFICAS

- 1) Thom R: Structural stability and morphogenesis, an outline of a general theory of models. Reading MA, Benjamin Cummings, 1972.
- 2) Rosen R: Dynamical system theory in biology. New York, Wiley, 1970. vol 1.
- 3) Glass L, Mackey MC: Pathological conditions resulting from instabilities in physiological control systems. Ann NY Acad Sci 316:214-235, 1979.
- 4) Hodgkin A, Huxley AF: A quantitative description of membrane current and application to conduction and excitation in nerve.

- J Physiol (London) 117:500-544, 1952.
- 5) McAllister RE, Noble D, Tsien RW: Reconstruction of the electrical activity of cardiac Purkinje fibres. J Physiol (London) 251:1-59, 1975.
 - 6) DiFrancesco D, Noble D: A model of cardiac electrical activity incorporating ionic pumps and concentration changes. Phil Trans R Soc (London) B 307:353-398, 1985.
 - 7) Hamill OP, Marty A, Neher E, Sakmann B, Sigworth FJ: Improved patch clamp techniques for high resolution current recording from cells and cell free membrane patches. Pflügers Arch 391:85-100, 1981.
 - 8) Thompson JMT, Stewart HB: Non-linear dynamics and chaos. Wiley, New York, 1988.
 - 9) Glass L, Mackey MC: From clocks to chaos: the rhythms of life. Princeton University Press. Princeton NJ. 1988.
 - 10) Cobelli C, Carson ER, Finkelstein L, Leaning MS: Validation of simple and complex models in physiology and medicine. Am J Physiol 246:R259-R266, 1984.