ISBN: 978-607-487-338-2

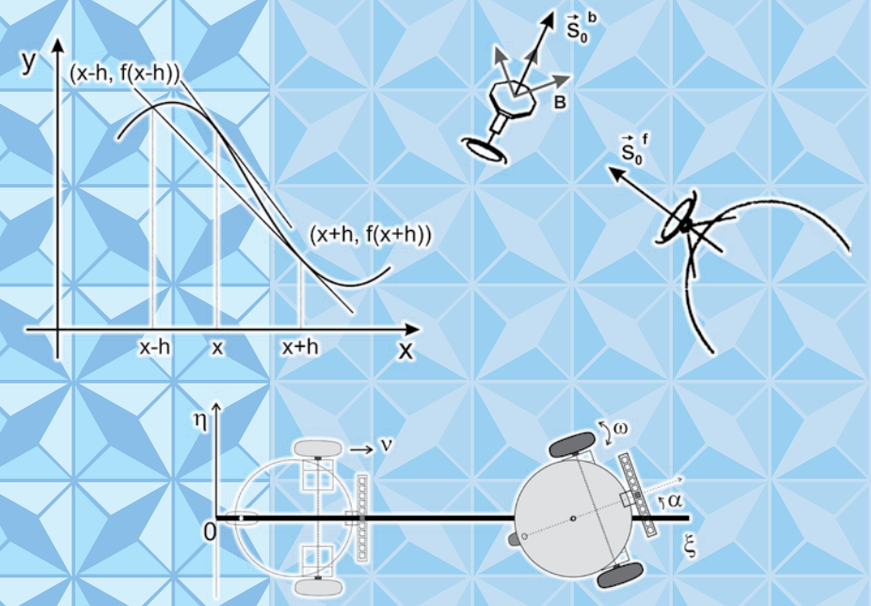
Es este libro la expresión de un esfuerzo por dejar constancia de la riqueza matemática de la Sexta Gran Semana Nacional de la Matemática (6GSNM). Los trabajos aquí presentados fueron sometidos a estricto arbitraje.

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias Físico matemáticas
Dirección de Fomento Editorial

Miguel Ángel García Ariza
Fernando Macías Romero
José Jacobo Oliveros Oliveros
editores

Matemáticas y sus Aplicaciones I

Matemáticas y sus Aplicaciones I



Miguel Ángel García Ariza
Fernando Macías Romero
José Jacobo Oliveros Oliveros
editores

Textos Científicos
Benemérita Universidad Autónoma de Puebla

Matemáticas y sus Aplicaciones I

Facultad de Ciencias Físico Matemáticas

Benemérita Universidad Autónoma de Puebla

Matemáticas y sus Aplicaciones I

Facultad de Ciencias Físico Matemáticas

Benemérita Universidad Autónoma de Puebla

EDITORES:
Miguel Ángel García Ariza,
Fernando Macías Romero,
José Jacobo Oliveros Oliveros

Primera edición, 2011
ISBN: 978-607-487-338-2

Dirección de Fomento Editorial
2 Norte 1404, C. P. 72000
Puebla, Pue.
Teléfono y fax: 01 222 246 8559

Impreso y hecho en México
Printed and made in Mexico

Matemáticas y sus Aplicaciones I

Selección bajo arbitraje riguroso de algunos trabajos presentados en la Sexta Gran Semana Nacional de la Matemática (6GSNM) realizada del 6 al 10 de septiembre de 2010 en la FCFM, BUAP.

Editores:
Miguel Ángel García Ariza, Fernando Macías Romero, Jacobo Oliveros Oliveros.

Contenido

Presentación V

Análisis Matemático 1

Capítulo 1. La $m-$Convergencia en el Espacio de Funciones: Integración. 3
Juan Alberto Escamilla Reyna
María Guadalupe Raggi Cárdenas.
Capítulo 2. Un Teorema del Valor Medio para la Derivada Simétrica. 13
Juan Alberto Escamilla Reyna
Victor Hugo Rodríguez Ávila
Anel Vázquez Martínez.
Capítulo 3. El Lema de Cousin Aplicaciones al Análisis Real. 23
María Guadalupe Raggi Cárdenas
Ericka Tlatilpa Guarneros
Teresa Torres Calzada.

Ecuaciones Diferenciales y Modelación Matemática 33

Capítulo 4. Estimación de Parámetros de un Motor DC con Interfaz para la Construcción del Modelo de Referencia de un Control Adaptativo. 35
Vladimir Vasilievich Alexandrov
Wuiyebaldo Fermín Guerrero Sánchez
Rigoberto Juárez Salazar
José Jacobo Oliveros Oliveros.
Capítulo 5. Inestabilidad de la Convección Natural en Cavidades Verticales y Horizontales Llenas de Aire. 45
Elsa Báez Juárez
María Blanca del Carmen Bermúdez Juárez
Alfredo Nicolás Carrizosa.
Capítulo 6. La Sumabilidad de Borel en la Solución de Ecuaciones Diferenciales. 57
Laura Angélica Cano Cordero.
Capítulo 7. Alcances y Limitaciones del Cómputo Científico: un Ejemplo. 69
Mario Alberto Carballo Flores
Reynaldo Domínguez Castillo
Francisco Sergio Salem Silva.

Capítulo 8. La Construcción de Rectas Tangentes antes de la Invención de la Derivada. 81
Lucía Cervantes Gómez
Ana Luisa González Pérez
Griselda Sánchez Denicia.
Capítulo 9. Cotas de Error para la Tercera Derivada de Esplines Cúbicos Empleando Cálculo Diferencial. 95
Lucía Cervantes Gómez
Valentín Jornet Plá
José Jacobo Oliveros Oliveros.
Capítulo 10. Estabilización de la Orientación de un Satélite por medio de Leyes de Control No Lineales con Retroalimentación de Salida. 109
Rafael Cruz José
José Fermi Guerrero Castellanos
Wuiyebaldo Fermín Guerrero Sánchez
José Jacobo Oliveros Oliveros.
Capítulo 11. Propuesta de Algoritmo Estable para la Identificación de Fuentes Bioeléctricas Tipo Dipolar. 123
Eladio Flores Mena
Andrés Fraguela Collar
José Eligio Moisés Gutiérrez Arias
Gabriela Morales Timal
María Monserrat Morín Castillo
José Jacobo Oliveros Oliveros.
Capítulo 12. Programa para el Análisis del Crecimiento de Tomates en Ambiente Controlado. 135
José Eligio Moisés Gutiérrez Arias
Irineo López Cruz
María Monserrat Morín Castillo
Ricardo Darío Peña Moreno
Eduardo Ríos Silva
Gabriel Romero Rodríguez
Juan Carlos Torres Monsivais.
Capítulo 13. Diseño Automático de Circuitos Electrónicos Analógicos usando una Estrategia General. 147
José Eligio Moisés Gutiérrez Arias
María Monserrat Morín Castillo
Ricardo Darío Peña Moreno
Eduardo Ríos Silva
Gabriel Romero Rodríguez
Juan Carlos Torres Monsivais
Alexandre Zemliak.
Capítulo 14. Validación Numérica de un Control Óptimo Discreto para un Robot Móvil. 163
José Eligio Moisés Gutiérrez Arias
María Monserrat Morín Castillo
Gelacio Salas Ortega.

Capítulo 15. Algunas Consideraciones sobre el Control del Caos Determinista. 175
Evodio Muñoz Aguirre.

Enseñanza, Historia y Divulgación de las Matemáticas 187

Capítulo 16. Antecedentes sobre las Teorías Paraconsistentes. 189
Eduardo Ariza Pérez
Pedro García Juárez
Rosa García Tamayo.
Capítulo 17. Un Bosquejo Histórico de Algunas Relaciones entre las Ciencias y la Milicia. 199
Juan Francisco Estrada García.
Capítulo 18. Estrategias para resolver Problemas de Matemáticas de Nivel Preuniversitario. 209
Lidia Aurora Hernández Rebollar
María Araceli Juárez Ramírez
Francisco Javier Rodríguez Martínez.
Capítulo 19. Acerca del Abuso de la Proporcionalidad por Estudiantes del Nivel Medio Superior. 219
Lidia Aurora Hernández Rebollar
Araceli Juárez Ramírez
Josip Sliško Ignjatov
Josué Vázquez Rodríguez.
Capítulo 20. Física y Matemática desde Arquímedes. 229
Raúl Linares Gracia
Juan Armando Reyes Flores.
Capítulo 21. El Tratamiento de los Infinitesimales según L'Hospital. 237
Raúl Linares Gracia
Josué Vázquez Rodríguez.

Lógica Matemática 247

Capítulo 22. Breve Reseña de Model Checking. 249
José Arrazola Ramírez
Iván Cortés
Jesús Lavalle Martínez.
Capítulo 23. Algunos Sistemas Lógicos. 257
José Arrazola Ramírez
Oscar Estrada Estrada
Jesús Lavalle Martínez.
Capítulo 24. Lógica Posibilista Estándar. 267
José Arrazola Ramírez
Oscar Estrada Estrada
Jesús Lavalle Martínez
Felipe Mazón Cambrón.

Capítulo 25. Demostración Automática de Teoremas. 281
José Arrazola Ramírez
Jesús Lavalle Martínez
Juan Pablo Muñoz Toriz.

Topología 289

Capítulo 26. La Topología de los Hiperespacios. 291
Vianey Córdova Salazar
David Herrera Carrasco
Fernando Macías Romero.
Capítulo 27. Dendritas Locales. 301
Luis Alberto Guerrero Méndez
David Herrera Carrasco
Fernando Macías Romero.
Capítulo 28. ¿Tienen las Dendritas Locales Producto Simétrico Único?. 313
David Herrera Carrasco
Fernando Macías Romero
Francisco Vázquez Juárez.
Capítulo 29. Subespacios en Espacios Ordenados. 327
Manuel Ibarra Contreras
Armando Martínez García.

Presentación

Las Grandes Semanas Nacionales de la Matemática son un ejercicio anual del colectivo matemático nacional que, organizado bajo la dirección de la Academia de Matemáticas de la Facultad de Ciencias Físico Matemáticas, presenta una gran gama de actividades que forman parte del quehacer matemático.

Es este libro la expresión de un esfuerzo por dejar constancia de la riqueza matemática de la Sexta Gran Semana Nacional de la Matemática (6GSNM). Toda obra editorial se realizó con la esperanza de tener numerosos lectores; si por lo menos los asistentes se convierten en lectores y éstos propagan este volumen, estaremos satisfechos.

En este libro se recogen algunos trabajos de la 6GSNM, agrupándolos de acuerdo a las sesiones de la misma, los cuales fueron sometidos a arbitraje riguroso. Agradecemos sinceramente a todos los árbitros su dedicación y profesionalismo así como a los encargados de las mencionadas sesiones: José Enrique Ramón Arrazola Ramírez, Lidia Aurora Hernández Rebollar, David Herrera Carrasco, Raúl Linares Gracia, Francisco Javier Mendoza Torres, María Monserrat Morín Castillo y José Jacobo Oliveros Oliveros.

Análisis Matemático

Capítulo 1

LA m-CONVERGENCIA EN EL ESPACIO DE FUNCIONES: INTEGRACIÓN

JUAN ALBERTO ESCAMILLA REYNA
MARÍA GUADALUPE RAGGI CÁRDENAS

Resumen. Dada una sucesión $\{f_n\}$ de funciones integrables en un intervalo cerrado y no acotado, sucesión m-convergente a f, probaremos con la m-convergencia, un teorema que nos asegura la igualdad entre el límite de las integrales de estas funciones y la integral del límite f.

1. Introducción

Consideremos una sucesión de funciones $\{f_n\}_{n\in\mathbb{N}}$, donde $f_n : [a,\infty) \to \mathbb{R}$. Nos preguntamos,

1.1. Pregunta. ¿Para qué tipo de convergencia se cumple

$$(*) \qquad \text{Si } \{f_n\} \overset{??}{\to} f, \text{ entonces } \int_a^\infty f_n \to \int_a^\infty f\,?$$

Para responder esto, presentaremos el concepto de m-convergencia para sucesiones de funciones reales de variable real, su relación con la convergencia puntual y la uniforme. Además analizaremos el comportamiento de la integral de estas sucesiones con respecto a la m-convergencia.

Consideremos una sucesión de funciones $\{f_n\}_{n\in\mathbb{N}}$, donde $f_n : I \subset \mathbb{R} \to \mathbb{R}$. Hablaremos sobre la convergencia de este tipo de sucesiones con diferentes tipos de convergencia: cuando I es un intervalo cerrado acotado $[a,b]$ y cuando es un intervalo no acotado $[a,\infty)$.

Analizaremos las igualdades:

$$(1) \qquad \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{dx} = \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,.$$

para funciones definidas en intervalos cerrados y acotados, y

$$(2) \qquad \lim_{n\to\infty} \int_a^\infty f_n(x)\,\mathrm{dx} = \int_a^\infty \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,.$$

para funciones definidas en intervalos no acotados.

2. Convergencia de Funciones

Empezaremos recordando algunas definiciones:

2.1. Definición (Convergencia Puntual). Para cada $n \in \mathbb{N}$ sean $f_n, f : A \subset \mathbb{R} \to \mathbb{R}$. Diremos que la sucesión $\{f_n\}$ converge puntualmente a f, $(f_n \overset{c.p.}{\to} f)$ si para cada $x_o \in A$ y cada $\epsilon > 0$, existe $N \in \mathbb{N}$, $N = N(x_o, \epsilon) > 0$ tal que

$$\text{para toda } n \geq N, \text{ se cumple que } |f_n(x_o) - f(x_o)| < \epsilon.$$

o de otra manera, la sucesión de números reales

$$\{f_n(x_o)\} \text{ converge a } f(x_o) \text{ cuando } n \to \infty.$$

Una definición similar, pero más fuerte es la de la convergencia uniforme:

2.2. Definición (Convergencia Uniforme). Para cada $n \in \mathbb{N}$, sea $A \subset \mathbb{R}$ y $f_n, f : A \to \mathbb{R}$. Diremos que la sucesión $\{f_n\}$ converge uniformemente a f, $(f_n \overset{c.u.}{\to} f)$ si para cada $\epsilon > 0$, existe $N \in \mathbb{N}$, $N = N(\epsilon) > 0$ tal que

$$\text{para toda } n \geq N, \text{ se cumple que } |f_n(x) - f(x)| < \epsilon, \text{ para toda } x \in A$$

Es casi inmediato que

$$\text{Si } f_n \overset{c.u.}{\to} f, \text{ entonces } f_n \overset{c.p.}{\to} f.$$

Pero la implicación recíproca, no, como lo podemos ver con el siguiente ejemplo:

2.3. Ejemplo. Sean $n \in \mathbb{N}$, y $f_n : [0,1] \to \mathbb{R}$ definidas como $f_n(x) = n^2 x(1-x^2)^n$. Es fácil de comprobar que, para cada $x_o \in [0,1]$ se tiene que

$$\lim_{n\to\infty} f_n(x_o) = \lim_{n\to\infty} n^2 x_o (1 - x_o^2)^n = 0.$$

Sin embargo, es fácil comprobar que esta sucesión no converge uniformemente a 0.

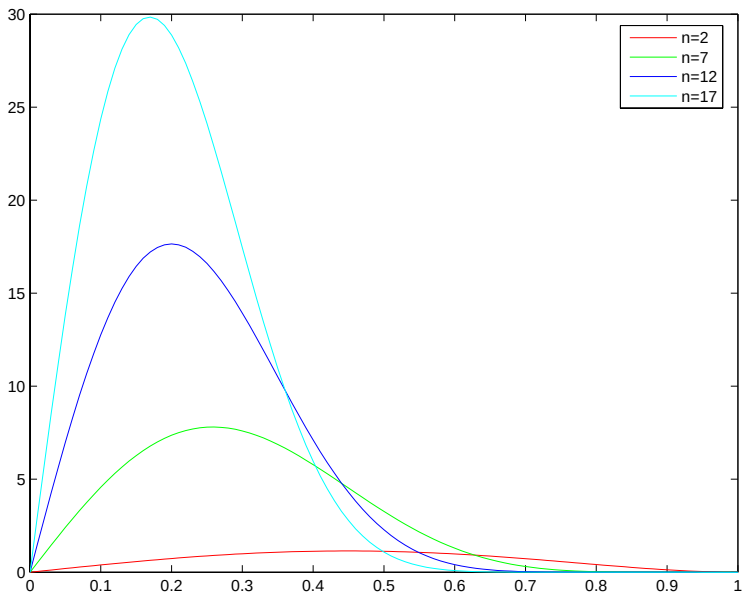


Figura 1. $f_n(x) = n^2 x(1-x^2)^n$

Ahora, considerando que la sucesión converge puntualmente a 0, analicemos el comportamiento de ésta, con respecto a la igualdad (*).

Si integramos

$$\int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{dx} = \int_0^1 \lim_{n\to\infty} n^2x(1-x^2)^n\,\mathrm{dx} = \int_0^1 0\,\mathrm{dx} = 0.$$

pero, por otro lado, tenemos que

$$\begin{aligned}\lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{dx} &= \lim_{n\to\infty}\int_0^1 n^2x(1-x^2)^n\,\mathrm{dx}\\ &= \lim_{n\to\infty}\frac{n^2}{2n+2}\\ &= \lim_{n\to\infty}\frac{1}{2/n+2/n^2} = \infty.\end{aligned}$$

Sin embargo, si una sucesión de funciones integrables, definidas en un intervalo cerrado y acotado, converge uniformemente a una función integrable, definida en el mismo intervalo, sí se cumple la igualdad (*), esto es:

2.4. Teorema (Integración de una Sucesión Uniformemente Convergente). Para cada $n \in \mathbb{N}$, sean $f_n : [a,b] \subset \mathbb{R} \to \mathbb{R}$, una sucesión de funciones Riemann integrables en $[a,b]$ y f una función definida en $[a,b]$. Supongamos que $f_n \overset{c.u.}{\to} f$ en $[a,b]$. Entonces f es Riemann integrable en $[a,b]$ y

$$\lim_{n\to\infty}\int_a^b f_n(x)\,\mathrm{dx} = \int_b^a \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,. \tag{1}$$

Demostración.

La prueba consiste en demostrar primero que f es Riemann integrable utilizando sumas inferiores y superiores y, después con las propiedades de las integrales, se demuestra la igualdad:

Sea $\epsilon > 0$, entonces

(a): existe $N = N(\epsilon) \in \mathbb{N}$, tal que para toda $n \geq N$

$$|f_n(x) - f(x)| < \frac{\epsilon}{3(b-a)} \quad \text{para toda} \quad x \in [a,b].$$

(b): Para esta N, como f_N es Riemann integrable, existe una partición $P = \{a = x_o, x_1, \ldots, x_m = b\}$ de $[a,b]$ tal que

$$|S(f_N, P) - s(f_N, P)| < \frac{\epsilon}{3}.$$

De **(a)** se sigue que

$$|\sup_{(x_{i-1},x_i)} f_n(x) - \sup_{(x_{i-1},x_i)} f(x)| \leq \frac{\epsilon}{3(b-a)}.$$

y que

$$|S(f_N,P)-S(f,P)| = \left|\sum_{i=1}^{m}(\sup_{(x_{i-1},x_i)} f_N(x) - \sup_{(x_{i-1},x_i)} f(x))(x_i - x_{i-1})\right|$$
$$\leq \sum_{i=1}^{m} |\sup_{(x_{i-1},x_i)} f_N(x) - \sup_{(x_{i-1},x_i)} f(x)|\,(x_i - x_{i-1})$$
$$\leq \sum_{i=1}^{m}\left[\frac{\epsilon}{3(b-a)}\right](x_i - x_{i-1}) = \left[\frac{\epsilon}{3(b-a)}\right](b-a).$$

Por lo tanto

$$|S(f_n,P)-S(f,P)| \leq \frac{\epsilon}{3}.$$

Análogamente se obtiene que

$$|s(f_n,P)-s(f,P)| \leq \frac{\epsilon}{3}.$$

De esto y de **(b)**, concluímos que

$$|S(f,P)-s(f,P)| \leq |S(f,P)-S(f_n,P)|+|S(f_n,P)-s(f_n,P)|+|s(f_n,P)-s(f,P)|$$
$$< \frac{\epsilon}{3}+\frac{\epsilon}{3}+\frac{\epsilon}{3} = \epsilon.$$

Luego, f es Riemann integrable en $[a,b]$.

Veamos ahora que

$$\lim_{n\to\infty}\int_a^b f_n(x)\,\mathrm{dx} = \int_b^a \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,.$$

Sea $\epsilon > 0$, entonces existe $N = N(\epsilon) \in \mathbb{N}$, tal que, para toda $n \geq N$

$$|f_n(x) - f(x)| < \frac{\epsilon}{(b-a)} \quad \text{para toda} \quad x \in [a,b].$$

De aquí que

$$\left|\int_a^b (f_n(x)-f(x))\,\mathrm{dx}\right| \leq \int_a^b |f_n(x)-f(x)|\,\mathrm{dx} < \int_a^b \frac{\epsilon}{(b-a)}\,\mathrm{dx} = \frac{\epsilon}{(b-a)}(b-a) = \epsilon.$$

Luego entonces,

$$\lim_{n\to\infty}\int_a^b f_n(x)\,\mathrm{dx} = \int_b^a \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,.$$

□

Pero, si cambiamos el intervalo $[a,b]$ por un intervalo de la forma $[a,\infty)$, el teorema **2.4**, ya no se cumple. Veamos el siguiente ejemplo.

2.5. Ejemplo. Para cada $n \in \mathbb{N}$, $n \geq 2$ sean las funciones $f_n : [1,\infty) \to \mathbb{R}$, definidas como

$$f_n(x) = \begin{cases} 1/x, & \text{si } \ 1 \leq x \leq n \\ n/x^2, & \text{si } \ x \geq n. \end{cases}$$

Se demuestra que:

- $f_n(x) \overset{c.u.}{\to} 1/x$ en $[1,\infty)$.

- Para cada $n \in \mathbb{N}$, la integral impropia $\int_1^\infty f_n(x)\,\mathrm{dx} = 1 + \log(n)$.
- Pero, sabemos que la función límite $f(x) = 1/x$ no es Riemann integrable en el intervalo $(1, \infty)$.

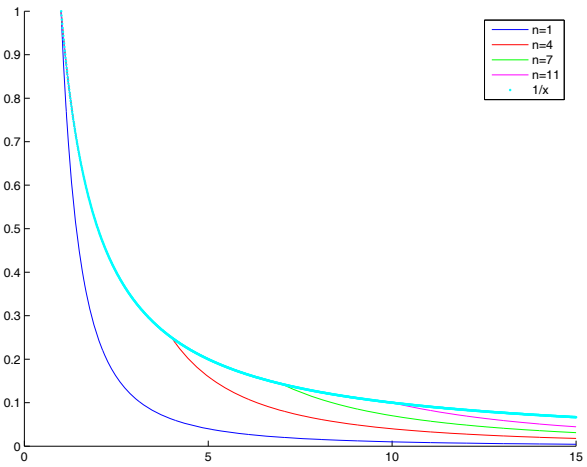


FIGURA 2. $f_n(x) = n/x^2,\ x \geq n$

Pero aún cuando se tenga una sucesión de funciones cuyos elementos sean Riemann integrables en un intervalo no acotado, que converja uniformemente sobre ese intervalo y que la función límite también sea Riemann integrable en ese intervalo, tampoco se cumple necesariamente la igualdad (2) como lo mostramos en el siguiente ejemplo.

2.6. EJEMPLO. Para cada $n \in \mathbb{N},\ n \geq 2$ sean las funciones $f_n : [1, \infty) \to \mathbb{R}$, definidas como

$$f_n(x) = \begin{cases} 1/x^2, & \text{si } \ 1 \leq x \leq n \\ n/x^2, & \text{si } \ x > n. \end{cases}$$

Se demuestra que:

- $f_n \overset{c.u.}{\to} 1/x^2$ en $[1,\infty)$.
- Para cada $n \in \mathbb{N}$, la integral impropia $\int_1^\infty f_n(x)\,\mathrm{dx} = 2 - 1/n$.

Además

$$\lim_{n\to\infty} \int_1^\infty f_n(x)\,\mathrm{dx} = \lim_{n\to\infty} \left(2 - \frac{1}{n}\right) = 2.$$

Pero, por otro lado,

$$\int_1^\infty \lim_{n\to\infty} f_n(x)\,\mathrm{dx} = \int_1^\infty \frac{1}{x^2}\,\mathrm{dx} = \lim_{b\to\infty} \int_1^b \frac{1}{x^2}\,\mathrm{dx} = \lim_{b\to\infty} \left(1 - \frac{1}{b}\right) = 1.$$

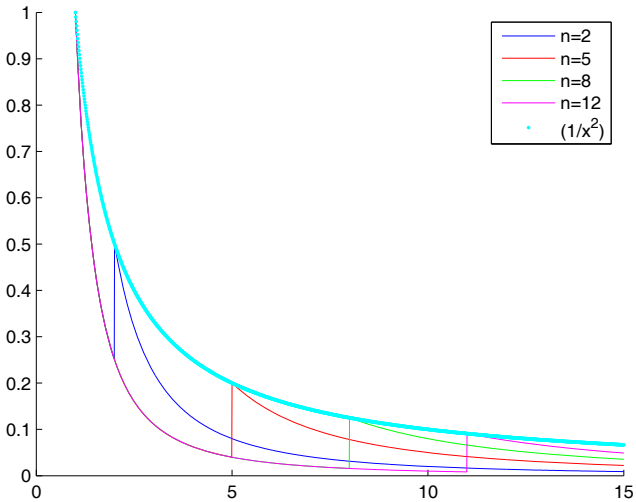


FIGURA 3. $f_n(x) = n/x^2,\ x \geq n$

Podemos agregarle condiciones a la sucesión de funciones para que con la convergencia uniforme se cumpla la igualdad (2), ver [1], o. . . , desearíamos tener otro tipo de convergencia donde la igualdad (2) sea verdadera, es decir, si la sucesión $\{f_n\}$ "converge" a la función f en el intervalo $[a,\infty)$ y cada f_n tiene integral impropia entonces f tiene integral impropia y

$$\lim_{n\to\infty} \int_a^\infty f_n(x)\,\mathrm{dx} = \int_a^\infty \lim_{n\to\infty} f_n(x)\,\mathrm{dx}\,.$$

y que, por supuesto, coincida con la convergencia uniforme en intervalos cerrados y acotados.

3. *m*-Convergencia

Eliakim Hastings Moore, matemático americano, (1862–1932), definió otro tipo de convergencia, ver [5], en la cual, la igualdad (2) es válida.

Veamos un concepto necesario para la definición de la convergencia propuesta:

$$\mathbb{C}^+(A) = \{\epsilon : A \to \mathbb{R} \mid \epsilon(x) > 0\ \forall x \in A,\ \epsilon \text{ continua}\}\,.$$

3.1. Definición (m-convergencia). Sean f y $\{f_n\}$ una sucesión de funciones, con $f, f_n : A \subset \mathbb{R} \to \mathbb{R}$, entonces $\{f_n\}$ es $m-$convergente a f ($f_n \overset{m.c.}{\to} f$) si, para toda $\epsilon \in \mathbb{C}^+(\mathrm{A})$, existe $N = N(\epsilon)$ tal que, para toda $n \geq N$,

$$|f_n(x) - f(x)| < \epsilon(x), \quad \text{para toda } x \in A.$$

El primer resultado que tenemos sobre la relación entre la m-convergencia y la convergencia uniforme es un teorema muy sencillo de demostrar:

3.2. Teorema. Sean f y $\{f_n\}$ una sucesión de funciones tales que $f_n, f : A \subset \mathbb{R} \to \mathbb{R}$. Si $f_n \overset{m.c.}{\to} f$ en A entonces $f_n \overset{c.u.}{\to} f$ en A.

Si el dominio de las funciones es un intervalo cerrado y acotado entonces las dos convergencias son equivalentes, esto es:

3.3. Teorema. Sean f y $\{f_n\}$ una sucesión de funciones tales que $f_n, f : [a, b] \to \mathbb{R}$ son continuas y $f_n \overset{c.u.}{\to} f$ en $[a, b]$. Entonces $f_n \overset{m.c.}{\to} f$ en $[a, b]$.

Demostración.

La demostración se basa en el hecho de que ϵ es una función continua positiva definida en un intervalo cerrado $[a, b]$ y que, por lo tanto, en ese intervalo alcanza su mínimo.

Sea $\epsilon^* \in \mathbb{C}^+(\mathrm{A})$. Por ser ϵ^* continua en A y A cerrado y acotado, existe $a \in A$ tal que $0 < M = \epsilon^*(a) \leq \epsilon^*(x)$ para toda $x \in A$.

Como $f_n \overset{c.u.}{\to} f$ en A, existe $N \in \mathbb{N}$ tal que para toda $n \geq N$

$$|f_n(x) - f(x)| < M, \text{ para toda } x \in A.$$

Por lo tanto, $f_n \overset{m.c.}{\to} f$ en A.

□

Pero si el dominio no es acotado, la convergencia uniforme de una sucesión de funciones no garantiza la m-convergencia.

El ejemplo 2.5 nos muestra una sucesión de funciones continuas cuya convergencia es uniforme, sin embargo, la sucesión no es m-convergente. La demostración de esto, es por contradicción: suponemos que es m-convergente y tomamos $\epsilon(x) = 1/x^2 > 0$.

Sin embargo, existe una relación entre la convergencia uniforme y la m-convergencia en intervalos no acotados, como lo veremos en el siguiente teorema. Además este teorema es fundamental para la demostración de la igualdad (2).

3.4. Teorema. 0 Sean $f_n, f : [a, \infty) \to \mathbb{R}$. Entonces $f_n \overset{m.c.}{\to} f$ en $[a, \infty)$, si y sólo si, existe $M > a$, tal que $f_n \overset{c.u.}{\to} f$ en $[a, M]$ y $f_n - f \equiv \mathbf{0}$ en $[M, \infty)$ excepto para un número finito de subíndices n.

Demostración.

[⇒](Contradicción)

Supongamos que $f_n \overset{m.c.}{\to} f$ en $[a, \infty)$ y que la conclusión no se cumple, es decir, para toda $M > a$, $f_n \overset{c.u.}{\nrightarrow} f$ en $[a, M]$ ó $f_n - f \not\equiv \mathbf{0}$ en $[M, \infty)$ para un número infinito de subíndices n.

Por el teorema **3.2**, se tiene que para toda $M > a$, $f_n \overset{c.u.}{\to} f$ en $[a, M]$, luego $f_n - f \not\equiv \mathbf{0}$ en $[M, \infty)$ para un número infinito de subíndices n.

Sea $M = |a|+1$, existe $n_1 \in \mathbb{N}$ tal que $f_{n_1} - f \not\equiv \mathbf{0}$ en $[|a|+1, \infty)$, es decir, existe $x_1 \geq |a|+1$ tal que $f_{n_1}(x_1) - f(x_1) \neq 0$.

Sea $M = x_1 + 1$, existe $n_2 \in \mathbb{N}$, $n_2 > n_1$ tal que $f_{n_2} - f \not\equiv \mathbf{0}$ en $[x_1 + |a|, \infty)$, es decir, existe $x_2 > x_1$ tal que $f_{n_2}(x_2) - f(x_2) \neq 0$.

Podemos encontrar dos subsucesiónes $\{f_{n_k}\}$ y $\{x_k\}$ con $x_{k+1} > x_k$, $\{x_k\}$ convergente a ∞ y tal que $(f_{n_k} - f)(x_k) \neq 0$. Llamemos $c_k = |f_{n_k}(x_k) - f(x_k)|/2$

Consideremos la función $\epsilon : [a, \infty) \to \mathbb{R}$ con:

$$\epsilon(x) = \begin{cases} c_1, & \text{si } x \in [a, x_1], \\ c_k, & \text{si } x = x_k,\ k \in \mathbb{N}, \\ \frac{c_{k+1}-c_k}{x_{k+1}-x_k}(x - x_k) + c_k, & \text{si } x \in [x_k, x_{k+1}],\ k \in \mathbb{N}. \end{cases}$$

Claramente ϵ es continua, luego, existe $N \in \mathbb{N}$ tal que para toda $n \geq N$,

$$|f_n(x) - f(x)| < \epsilon(x), \text{ para toda } x \in [a, \infty).$$

Sea $n_k > N$, entonces

$$|f_{n_k}(x_k) - f(x_k)| < \epsilon(x_k), \text{ para toda } k \in \mathbb{N},$$

lo cual es una contradicción.

[⇐]

Sean $\{n_j\}_{j=1}^{k}$ tal que $f_{n_j} - f \not\equiv \mathbf{0}$ y $\epsilon(x) > 0$, por el teorema **??** existe $N \in \mathbb{N}$, tal que, si $n \geq N$, entonces

$$|f_n(x) - f(x)| < \epsilon(x), \text{ para cada } x \in [a, M].$$

Sea $n \geq N' = \text{máx}\{N, n_1, n_2, \ldots, n_k\}$, y $x \in [a, \infty)$, entonces

$$|f_n(x) - f(x)| < \epsilon(x).$$

Luego, $f_n \overset{m.c.}{\to} f$ en $[a, \infty)$.

□

Finalmente, con el teorema anterior, podemos asegurar la validez de la igualdad (2):

3.5. Teorema. Sean $f_n, f : [a, \infty) \to \mathbb{R},\ n \in \mathbb{N}$. Si $f_n \overset{m.c.}{\to} f$ en $[a, \infty)$ y $\int_a^\infty f_n$ existe, entonces $\int_a^\infty f$ existe y

$$\lim_{n\to\infty} \int_a^\infty f_n(x)\,\mathrm{dx} = \int_a^\infty f(x)\,\mathrm{dx}\,. \tag{2}$$

Demostración.

Por el teorema **3.4**, existe $M > a$ tal que $f_n \overset{c.u.}{\to} f$ en $[a, M]$ y $f_n - f \equiv \mathbf{0}$ en $[M, \infty)$ excepto para un número finito de subíndices n.

Como

$$\int_a^\infty f = \int_a^M f + \int_M^\infty f,$$

entonces, $\int_a^\infty f$ existe, pues cada integral del segundo miembro de la igualdad anterior existe.

Demostremos ahora que $\int_a^\infty f_n \underset{n}{\to} \int_a^\infty f$.

$$\left|\int_a^\infty f_n - \int_a^\infty f\right| = \left|\int_a^M f_n + \int_M^\infty f_n - \int_a^M f - \int_M^\infty f\right|$$
$$= \left|\left(\int_a^M f_n - \int_a^M f\right) + \left(\int_M^\infty f_n - \int_M^\infty f\right)\right|.$$

Como cada integral $\int_M^\infty f_n$, $\int_M^\infty f$ existe y $\int_M^\infty f_n - \int_M^\infty f = \int_M^\infty (f_n - f) = 0$, entonces

$$\left|\int_a^\infty f_n - \int_a^\infty f\right| = \left|\int_a^M f_n - \int_a^M f\right|.$$

De esta igualdad y por el teorema **2.4**, podemos concluir el resultado, esto es

$$\lim_{n\to\infty} \int_a^\infty f_n = \int_a^\infty \lim_{n\to\infty} f_n.$$

□

Referencias

[1] Apostol Tom M. *Análisis Matemático*, California Institute of Technology. Editorial Reverté, S.A. 1976.

[2] Bartle Robert G. University of Illions Urbana-Champaign *The Elements of Real Analysis*, Editorial John Wiley & Sons, Inc.

[3] Fernández Muñiz José Luis, De La Torre Molné Graciela, *Análisis Matemático, Tomo V*, Editorial Pueblo y Educación, 1987.

[4] Khinchin Aleksandr, *A Course of Mathematical Analisys*, Moscow University U.S.S.R. Hindustan Publishing Corp 1960 (India) DELHI.

[5] Moore E. H. *On The Fundations of The Theory of Linear Integral Equations*, Bull. Amer. math., 46, pp. 151-161, (1911-1912).

[6] Raggi Cárdenas Ma. Guadalupe, *La m-Topología en Espacios de Funciones* Puebla, Pue., 1998.

[7] Rudin Walter, *Principles of Mathematical Analisys*, University of Wisconsin-Madison 1976.

[8] Takeuchi Yu, Universidad Nacional de Colombia *Sucesiones y Series Tomo II*, Editorial Limusa México, 1976.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

jescami@fcfm.buap.mx, gperaggi@fcfm.buap.mx

Capítulo 2

UN TEOREMA DEL VALOR MEDIO PARA LA DERIVADA SIMÉTRICA

JUAN ALBERTO ESCAMILLA REYNA
VICTOR HUGO RODRÍGUEZ ÁVILA
ANEL VÁZQUEZ MARTÍNEZ
FCFM - BUAP

Resumen. Presentaremos, para la derivada simétrica, un resultado análogo al Teorema del Valor Medio clásico. Usaremos este resultado para probar una equivalencia entre la derivada simétrica y la derivada clásica.

1. Introducción

Como ya hemos visto en nuestros cursos de cálculo, una función diferenciable, cumple ciertos teoremas y propiedades, en esta memoria veremos una generalización de la derivada clásica, la cual llamaremos derivada simétrica.

Gráficamente podemos observar que la interpretación de la derivada clásica y la interpretación de la derivada simétrica parecen ser la misma, pero mostraremos que desde el punto de vista analítico ambos conceptos no son equivalentes.

2. derivada clásica y derivada simétrica

2.1. Definición. Sean I un intervalo abierto, $x \in I$ y $f : I \to \mathbb{R}$ una función, se dice que:

a) f tiene derivada clásica (DC) en x, si

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

existe. A este límite lo denotaremos como $f'(x)$.

En este caso, diremos también que la derivada de f existe en x. Si este límite no existe, diremos que f no tiene derivada en x o que la derivada de f no existe en x.

b) f tiene derivada simétrica (DS) en x, si

$$\lim_{h\to 0} \frac{f(x+h) - f(x-h)}{2h}$$

existe. A este límite lo denotaremos como $f_s(x)$. Tenemos las mismas observaciones que en a) para la derivada simétrica.

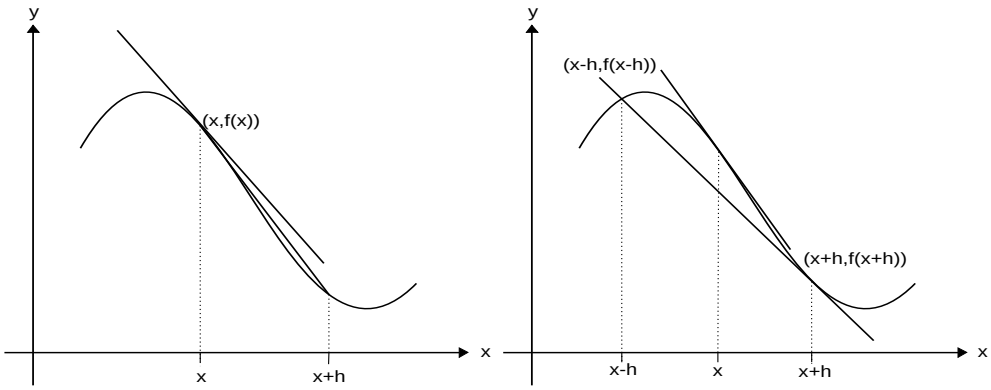


FIGURA 1. Representación gráfica de la derivada clásica y derivada simétrica, respectivamente.

2.2. OBSERVACIÓN. Si una función es simétricamente diferenciable en todo punto del intervalo, entonces diremos que la función tiene DS en ese intervalo.

2.3. TEOREMA. Si una función f es diferenciable en $x \in I$, entonces f es simétricamente diferenciable en x. Más aún, $f'(x) = f_s(x)$.

DEMOSTRACIÓN. Mostraremos que $f_s(x)$ existe.

$$\begin{aligned} f_s(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x-h)}{2h} \\ &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{2h} + \lim_{h\to 0} \frac{f(x)-f(x-h)}{2h} \\ &= \frac{1}{2}f'(x) + \frac{1}{2}f'(x) \\ &= f'(x), \end{aligned}$$

□

El Teorema 2.3 afirma que DC implica DS, pero es importante notar que el recíproco de este teorema no siempre se cumple, es decir, que si f tiene derivada simétrica en un punto no necesariamente tiene derivada clásica.

2.4. EJEMPLO. Consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x)=|x|$, esta función es simétricamente diferenciable en cero, pero no es diferenciable en cero.

DEMOSTRACIÓN. Si tomamos los límites laterales de f en cero, el límite lateral izquierdo es -1 y el límite lateral derecho es 1, observemos que como estos límites laterales son distintos, entonces el límite de f en cero no existe, por lo tanto, la función no tiene derivada clásica en cero.

Ahora veamos que f tiene derivada simétrica en cero.

$$\begin{aligned} f_s(0) &= \lim_{h\to 0} \frac{f(0+h)-f(0-h)}{2h} \\ &= \lim_{h\to 0} \frac{|\, h \,| - |\, -h \,|}{2h} \\ &= \lim_{h\to 0} 0 \\ &= 0, \end{aligned}$$

por lo tanto $f_s(0)= 0$. □

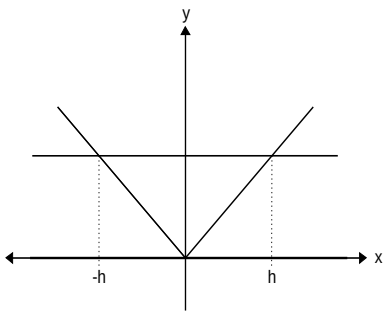


FIGURA 2. Gráfica de la función $f(x) = |x|$ y gráfica de la recta tangente a esta función en el punto $(0, 0)$. (en el sentido de la derivada simétrica)

2.5. OBSERVACIÓN. Al generalizar la definición de DC a DS se pierden algunas propiedades de DC como son: la continuidad y el Teorema del Valor Medio, entre otras.

2.6. EJEMPLO. Veamos una función con derivada simétrica en un punto x_o que no es continua en x_o, consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida como

$$f(x) = \begin{cases} 2, & \text{si} \quad x \neq 0, \\ 5, & \text{si} \quad x = 0. \end{cases}$$

La función f tiene DS en cero, ¿Será f continua en cero?

DEMOSTRACIÓN. Veamos que f es simétricamente diferenciable en cero.

$$\begin{aligned} f_s(0) &= \lim_{h\to 0} \frac{f(0+h)-f(0-h)}{2h} \\ &= \lim_{h\to 0} \frac{2-2}{2h} \\ &= 0, \end{aligned}$$

por lo tanto $f_s(0) = 0$.
Para responder la pregunta anterior, notemos que el límite lateral izquierdo de f en 0 es igual a 2 y el límite lateral derecho de f en 0 es igual a 2, por lo tanto el límite de f en 0 existe y es igual a 2, como $f(0) = 5$, entonces f no es continua en

cero. □

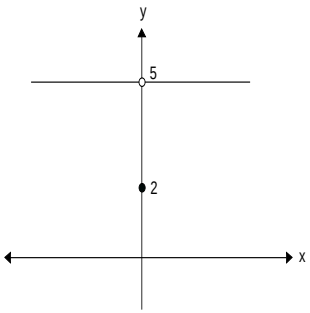


FIGURA 3. Gráfica de la función f

2.7. EJEMPLO. Veamos una función que no tiene derivada simétrica en un punto. Consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida como

$$f(x) = \begin{cases} \frac{|x|}{x}, & \text{si} \quad x \neq 0, \\ 0, & \text{si} \quad x = 0. \end{cases}$$

La función f no es simétricamente diferenciable en cero.

DEMOSTRACIÓN. Esto lo podemos demostrar calculando el siguiente límite

$$\begin{aligned} f_s(0) &= \lim_{h\to 0} \frac{f(0+h) - f(0-h)}{2h} \\ &= \lim_{h\to 0} \frac{\frac{|h|}{h} - \frac{|-h|}{-h}}{2h} \\ &= \lim_{h\to 0} \frac{|h|}{h^2}, \end{aligned}$$

este último límite no existe, por lo tanto f no es simétricamente diferenciable en cero.

□

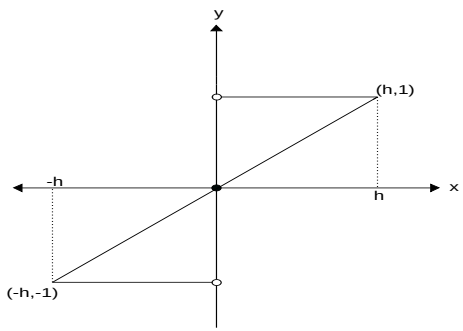


FIGURA 4. Gráfica de la función $f(x) = \frac{|x|}{x}$.

Ahora recordemos lo que dice el Teorema del Valor Medio para la derivada clásica y veamos si se tiene un teorema del valor medio para la derivada simétrica.

2.8. Teorema. Sea f una función continua en $[a, b]$ y derivable en (a, b), entonces existe $\kappa \in (a, b)$ tal que:

$$f'(\kappa) = \frac{f(b) - f(a)}{b - a}.$$

2.9. Ejemplo. Consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida como $f(x)=|x|$. Esta función no satisface el Teorema del Valor Medio en el intervalo [-1,2], si se considera la derivada simétrica en lugar de la derivada clásica.

$$f_s(x) = \begin{cases} \frac{|x|}{x}, & \text{si} \quad x \neq 0, \\ 0, & \text{si} \quad x = 0. \end{cases}$$

Demostración. Fijémonos en la imagen de f_s que es el conjunto $\{0, -1, 1\}$, y en la pendiente de la recta secante que une los puntos $(-1, 1)$ y $(2, 2)$ de la gráfica de f que está dada por:

$$\frac{f(2) - f(-1)}{3} = \frac{1}{3}.$$

Notemos que $\frac{1}{3} \notin \{0, -1, 1\}$, es decir, no existe $x \in$ [-1,2] tal que $f_s(x) = \frac{1}{3}$. □

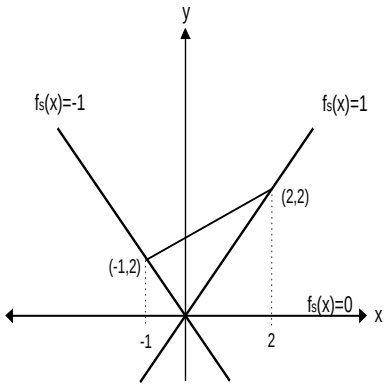


Figura 5. Gráfica de una función que no cumple con el Teorema del Valor Medio para DS.

3. Un Teorema del Valor Medio para la derivada simétrica

Esta sección es la parte central de esta memoria, ya que estableceremos un Casi Teorema del Valor Medio para funciones con derivada simétrica. Además mostraremos que toda función continua en un intervalo cerrado, cuya derivada simétrica tiene la propiedad de Darboux cumple con la conclusión del Teorema del Valor Medio.

3.1. Lema. Sea f una función continua en el intervalo $[a,b]$ y simétricamente diferenciable en (a,b).

a) Si $f(b) > f(a)$, entonces existe $\alpha \in (a,b)$ tal que
$$f_s(\alpha) \geq 0.$$
b) Si $f(b) < f(a)$, entonces existe $\beta \in (a,b)$ tal que
$$f_s(\beta) \leq 0.$$

Demostración. Suponemos que $f(b) > f(a)$, sea k un número real tal que $f(a) < k < f(b)$. Consideremos el conjunto $A = \{x \in [a,b] \mid f(x) > k\}$. A está acotado inferiormente por a, además como es un subconjunto de $\mathbb{R}$ distinto del vacío, ya que $b \in A$, entonces tiene ínfimo, digamos η. Primero probaremos que η es distinto de a y b, como η es el ínfimo de A, existe una sucesión $\{x_n\}$ de elementos de A que converge a η, luego como f es continua en η y $x_n \in A$, $(n \in \mathbb{N})$, entonces $f(\eta) \geq k$, por lo tanto $\eta > a$. Como f es continua en b y $f(b) > k$, existe $\delta > 0$ que cumple
$$\text{si } x \in (b-\delta, b] \text{ y } x \in [a,b], \text{ entonces } f(x) > k.$$

Sea $x \in (b-\delta, b]$ y $x \in [a,b)$, entonces $x \in A$ y por lo tanto $\eta \leq x < b$. Sea $(\eta - r, \eta + r)$ una vecindad de η contenida en $[a,b]$, Ahora probaremos que $f_s(\eta) \geq 0$. Lo haremos por contradicción. Como $f_s(\eta) < 0$, existe $r > r_1 > 0$ tal que

$$\text{si } 0 < h < r_1, \text{ entonces } \frac{f(\eta+h) - f(\eta-h)}{2h} < 0. \tag{1}$$

$$\text{Para cada } 0 < h < r_1, \text{ se tiene que } f(\eta - h) \leq k. \tag{2}$$

Para cada $0 < h < r_1$, existe $0 < h_1 < r_1$, tal que

$$\eta + h_1 \in A, \text{ es decir, } f(\eta + h_1) > k. \tag{3}$$

De (2) y (3)
$$\frac{f(\eta+h_1) - f(\eta-h_1)}{2h} \geq 0.$$
Lo cual es una contradicción con (1).

Análogamente se demuestra que si $f(a) > f(b)$, entonces existe $\xi \in (a,b)$ tal que $f_s(\eta) \geq 0$. □

El siguiente teorema es considerado como una versión del Teorema de Rolle para funciones simétricamente diferenciables.

3.2. Teorema. Sea f una función continua en $[a,b]$ y simétricamente diferenciable en (a,b). Supongamos que $f(a) = f(b) = 0$, entonces existen α y $\beta \in (a,b)$ tal que

$$f_s(\alpha) \geq 0 \quad y \quad f_s(\beta) \leq 0.$$

Ahora, una vez mencionadas las dos herramientas anteriores, probaremos el Casi Teorema del Valor Medio para funciones simétricamente diferenciables.

3.3. Teorema. Sea f una función continua en $[a,b]$ y simétricamente diferenciable en (a,b), entonces existen α y $\beta \in (a,b)$ tales que

$$f_s(\alpha) \leq \frac{f(b)-f(a)}{b-a} \leq f_s(\beta).$$

Demostración. Consideremos la función $g : \mathbb{R} \to \mathbb{R}$ definida como

$$g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a).$$

La derivada simétrica de $g(x)$ está dada por:

$$g_s(x) = f_s(x) - \frac{f(b)-f(a)}{b-a}.$$

Veamos que se cumplen las condiciones del teorema 3.2

$$\begin{aligned} g(a) &= f(a) - f(a) - \frac{f(b)-f(a)}{b-a}(a-a), \\ g(a) &= 0, \end{aligned}$$

y análogamente pasa con b, es decir $g(a) = 0 = g(b)$. Aplicando el teorema 3.2 a g en el intervalo $[a,b]$ obtenemos:

$$g_s(\alpha) = f_s(\alpha) - \frac{f(b)-f(a)}{b-a} \geq 0,$$

$$g_s(\beta) = f_s(\beta) - \frac{f(b)-f(a)}{b-a} \leq 0,$$

es decir

$$f_s(\alpha) \geq \frac{f(b)-f(a)}{b-a},$$

$$f_s(\beta) \leq \frac{f(b)-f(a)}{b-a},$$

entonces

$$f_s(\beta) \leq \frac{f(b)-f(a)}{b-a} \leq f_s(\alpha).$$

□

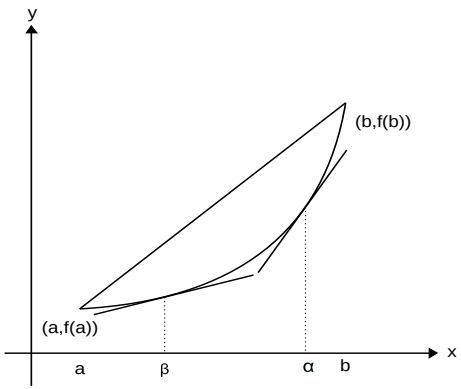


Figura 6. Ilustración del Teorema 3.3

Observemos que el Teorema 3.3 es un teorema análogo al Teorema del Valor Medio para funciones simétricamente diferenciables. Una pregunta "natural" que surge es, ¿Qué condición o qué condiciones deben ser impuestas a la derivada simétrica de f para que la conclusión del Teorema del Valor Medio se cumpla?, la respuesta a esta pregunta es, si la derivada simétrica de f tiene la Propiedad de Darboux, entonces se cumple dicha conclusión.

3.4. Definición. Sea una función $f : [a, b] \to \mathbb{R}$, decimos que f tiene la propiedad de Darboux si para cualesquiera α y $\beta \in [a, b]$ y un número k entre $f(\alpha)$ y $f(\beta)$, se tiene que existe un número γ entre α y β tal que

$$k = f(\gamma).$$

3.5. Observación. Si una función f es continua, entonces tiene la Propiedad de Darboux. Si f tiene la propiedad de Darboux puede ser discontinua, ver [1].

A continuación, enunciaremos y demostraremos el Teorema del Valor Medio para funciones simétricamente diferenciables.

3.6. Teorema. Sea f una función continua en $[a, b]$ y simétricamente diferenciable en (a, b). Si f_s tiene la Propiedad de Darboux, entonces existe $\gamma \in (a, b)$ tal que:

$$f_s(\gamma) = \frac{f(b) - f(a)}{b - a}.$$

Demostración. Por hipótesis, sabemos que f es continua en $[a, b]$ y simétricamente diferenciable en (a, b), entonces por el Casi Teorema del Valor Medio, existen α y $\beta \in (a, b)$ tal que:

$$f_s(\alpha) \leq \frac{f(b) - f(a)}{b - a} \leq f_s(\beta),$$

y como f_s tiene la Propiedad de Darboux, para $k = \frac{f(b)-f(a)}{b-a}$ existe un número γ entre α y β tal que $k = f_s(\gamma)$, por lo tanto

$$f_s(\gamma) = \frac{f(b) - f(a)}{b - a}.$$

□

4. Una Aplicación

Como las funciones simétricamente diferenciables no son necesariamente diferenciables, la pregunta a seguir sería, ¿Qué condiciones adicionales deben ser impuestas en la derivada simétrica para que sea igual a la derivada clásica? En esta sección mostraremos, por medio del Teorema del Valor Medio, que si $f(x)$ y f_s ambas son continuas en el mismo intervalo, entonces f es diferenciable.

4.1. Teorema. Sea f una función continua y simétricamente diferenciable en (a, b). Si f_s es continua en (a, b), entonces f' existe y

$$f'(x) = f_s(x).$$

Demostración. Elegimos h lo suficientemente pequeña para $a < x + h < b$. Además, sabemos que f es continua y simétricamente diferenciable en (a, b), y f_s es continua, entonces tiene la propiedad de Darboux, es decir, existe $\gamma \in (a, b)$ tal que:

$$f_s(\gamma) = \frac{f(b) - f(a)}{b - a},$$

Luego, sea $a = x + h \quad y \quad b = x$, tenemos que:

$$f_s(\gamma) = \frac{f(x + h) - f(x)}{h},$$

para alguna $\gamma \in (x, x + h)$. Ahora aplicando el límite a ambos lados de la igualdad cuando $h \to 0$, y sabiendo que el límite del lado izquierdo de la igualdad existe, obtenemos:

$$f_s(x) = f'(x).$$

□

Referencias

[1] Gordon, Russell A., *The Integrals of Lebesgue, Denjoy, Perron and Henstock*, Graduate Studies in Mathematics Volumen 4., American Mathematical Society, Rhode Isaland, (1994)

[2] Shoo P.K., Riedel T., *Mean Value and Funcional Equations*, World Scientific, Publishing Co. Pte. Ltd., Printed in Singapore by Uto-print (1998)

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

jescami@fcfm.buap.mx, vhr1012_87@hotmail.com, anel_ferro@hotmail.com

Capítulo 3

EL LEMA DE COUSIN APLICACIONES AL ANÁLISIS REAL

MARÍA GUADALUPE RAGGI CÁRDENAS
ERICKA TLATILPA GUARNEROS
TERESA TORRES CALZADA

Resumen. Nuestro propósito es usar el Lema de Cousin para dar demostraciones alternativas a los Teoremas siguientes: el Teorema del Valor Intermedio, el Teorema de Continuidad Uniforme, un Teorema de Weierstrass para funciones reales definidas en un intervalo cerrado finito.

1. INTRODUCCIÓN

Alrededor de los años 60's del siglo pasado, el matemático checo J. Kurzweil y el matemático inglés R. Henstock construyeron una integral que generaliza a la integral de Riemann, la de Lebesgue, la de Newton, las integrales impropias, permite integrar todas las derivadas y tiene buenos teoremas de convergencia, ver [1, 3, 5].

Para demostrar la unicidad de la integral de Henstock-Kurzweil de una función se usa un lema conocido como *Lema de Cousin*. La demostración de este lema se le atribuye al matemático belga P. Cousin, ver [5]. Este lema permite dar demostraciones alternativas, entre otros, a varios resultados del análisis real como el Teorema del Valor Intermedio, el Teorema de Weierstrass, el Teorema de Continuidad Uniforme, etc. Nuestro propósito es presentar estas demostraciones, ya que son poco conocidas.

Iniciaremos la siguiente sección, introduciendo algunos conceptos básicos necesarios para la demostrar estos resultados.

2. PRELIMINARES

2.1. Definición . Sea $I = [a, b]$. Una partición es una colección finita de subintervalos de I (denotada $\{I_i\}_{i=1}^n$) con cada $I_i = [x_{i-1}, x_i]$ para $i = 1, ..., n$, donde

$$a = x_0 < x_1 < x_2 < ... < x_n = b.$$

2.2. Definición . Una partición etiquetada de I es una partición a la cual, para cada subintervalo I_i se le asigna un punto $t_i \in I_i$ que se llama etiqueta. Así, decimos que la partición está etiquetada y se denota por:

$$P = \{(I_i, t_i) \mid i = 1, ..., n\} = \{(I_i, t_i)\}_{i=1}^n.$$

2.3. Definición . Sean $I = [a, b]$ y $\delta : I \to \mathbb{R}$ es una función medidora, si $\delta(t) > 0$ para todo $t \in I$.

2.4. Definición . Sean $P = \{(I_i, t_i)\}_{i=1}^{n}$ una partición etiquetada y δ una función medidora de I, se dice que P es δ-fina si:

$$I_i \subseteq [t_i - \delta(t_i), t_i + \delta(t_i)], \qquad i = 1, ..., n$$

Notación: Si la partición P es δ-fina, se denotará como $P \ll \delta$.

2.5. Definición . Sea $f : [a, b] \longrightarrow \mathbb{R}$. f es uniformemente continua en [a,b], si para todo $\epsilon > 0$ existe $\delta = \delta_\epsilon > 0$ tal que

$$\text{si } |x - y| < \delta, \text{ entonces } |f(x) - f(y)| < \epsilon.$$

2.6. Observación. Si f es uniformemente continua en $[a, b]$, entonces f es continua en $[a, b]$. Pero el recíproco no siempre se cumple.

2.7. Ejemplo. Sea $f : (0, 1) \longrightarrow \mathbb{R}$, $f(x) = \frac{1}{x}$

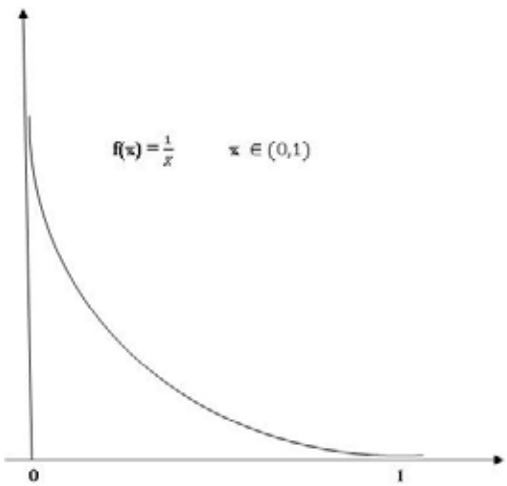


Figure 1. Gráfica de $f(x) = 1/x$

f es continua, pero no es uniformemente continua en (0,1) como lo demostraremos a continuación.

Sea $\epsilon_0 = \frac{1}{2}$ y sea $\delta > 0$, por la propiedad arquimediana, existe $n \in \mathbb{N}$, tal que $\frac{1}{n} < \delta/2$, de donde, existen $n_1, n_2 \in \mathbb{N}$, tal que $\left|\frac{1}{n_1} - \frac{1}{n_2}\right| < \delta$, entonces

$$\left|f\left(\frac{1}{n_1}\right) - f\left(\frac{1}{n_2}\right)\right| = \left|\frac{1}{1/n_1} - \frac{1}{1/n_2}\right| = |n_1 - n_2| \geq 1 > \frac{1}{2}.$$

Por lo tanto f no es uniformemente continua.

2.8. Proposición. Sea $f : [a, b] \longrightarrow \mathbb{R}$, $t \in [a, b]$, $L \in \mathbb{R}$. Si f continua en t y $f(t) < L$, entonces existe $\delta_t > 0$ tal que

$$\text{si } x \in [a, b] \text{ y } \mid x - t \mid < \delta_t, \text{ entonces } f(x) < L.$$

Demostración. Sea $\epsilon = L - f(t) > 0$, como f es continua en t, entonces existe $\delta_t > 0$ que cumple

$$\text{si } x \in [a, b] \text{ y } |x - t| < \delta_t, \text{ entonces } |f(x) - f(t)| < \epsilon.$$

Es decir, tenemos que

$$|f(x) - f(t)| < L - f(t)$$

por lo tanto $f(x) < L$.

□

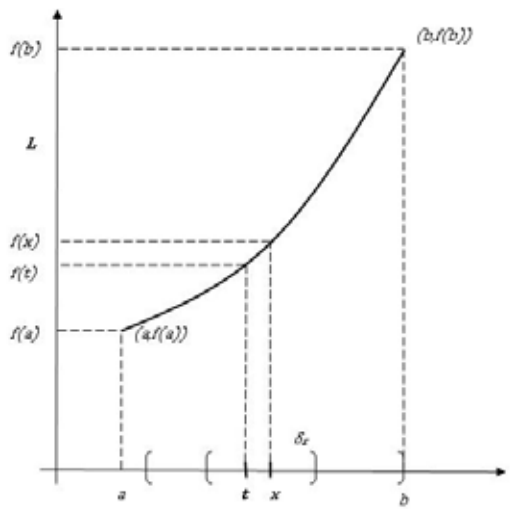


Figure 2. Interpretación Geométrica de la Prop. **2.8**

2.9. Proposición. Sea $f : [a, b] \longrightarrow \mathbb{R}$, $t \in [a, b]$, $L \in \mathbb{R}$. Si f continua en t y $f(t) > L$, entonces existe $\delta_t > 0$ tal que

$$\text{si } x \in [a, b] \text{ y } |x - t| < \delta_t, \text{ entonces } f(x) > L.$$

Demostración. La demostración es análoga a la de la **Propósición 2.8.**

□

2.10. Proposición. Sea $f : [a, b] \longrightarrow \mathbb{R}$. Si f es continua en t, entonces existen $\delta_t > 0$ y $M_t > 0$ tales que

$$\text{si } x \in [a, b] \text{ y } |x - t| < \delta_t, \text{ entonces } |f(x)| \leq M_t.$$

Demostración. Sea $\epsilon = 1$, como f es continua en t, entonces existe $\delta_t > 0$ tal que si $| x - t | < \delta_t$ y $x \in [a, b]$, entonces $| f(x) - f(t) | < 1$.

Pero además, sabemos que

$$| f(x) | - | f(t) | \leq | f(x) - f(t) | \text{ y } | f(x) - f(t) | < 1,$$

entonces

$$|f(x)| - |f(t)| < 1.$$

Por lo tanto

$$| f(x) | < 1 + | f(t) | = M_t.$$

□

Sea una función $f : [a, b] \longrightarrow \mathbb{R}$ y una partición etiquetada $P = \{([x_{i-1}, x_i], t_i) \mid i = 1, \ldots, n\}$. A

$$S(f, P) = \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}).$$

le llamamos la suma de Riemann para la función f con respecto a la partición etiquetada P.

2.11. Definición . Sea $f : [a, b] \longrightarrow \mathbb{R}$ una función. Decimos que f es una función **Henstock-Kurzweil integrable** sobre $[a, b]$, si existe $A \in \mathbb{R}$, tal que para todo $\epsilon > 0$, existe δ una función medidora en [a,b], tal que

$$\text{si } (P \ll \delta), \text{ entonces } |S(f, P) - A| < \epsilon.$$

3. LEMA DE COUSIN

Puede parecer un poco sorprendente que una partición δ-fina siempre exista. Sin importar "lo mal" que la función δ se comporte. Sin embargo, es importante que δ sea mayor que cero y que el intervalo sea compacto. El descubrimiento de la existencia de una partición δ-fina para cualquier δ positiva se remonta al siglo XIX por el matemático belga Cousin.

3.1. Lema (Cousin). Si $I = [a, b]$ y $\delta : I \longrightarrow (0.\infty)$, entonces existe una partición etiquetada δ-fina del intervalo $[a, b]$.

Demostración. Consideremos el conjunto

$$S = \{x \in [a, b] \mid \text{ existe una partición } \delta\text{-fina en } [a, x]\}.$$

Es claro que $S \neq \emptyset$ pues $a \in S$ y además está acotado superiormente por b. Así que por el Axioma del Supremo existe $\alpha = \sup S$, además $\alpha \in [a, b]$.

Vamos a demostrar que $\alpha \in S$, es decir que el intervalo $[a, \alpha]$ tiene una partición etiquetada δ-fina.

Como $\delta(\alpha) > 0$, $\alpha - \frac{\delta(\alpha)}{2}$ no es cota superior de S, entonces existe $x_0 \in S$ tal que

$$\alpha - \frac{\delta(\alpha)}{2} < x_0.$$

El intervalo $[a, x_0]$ tiene una partición $P \ll \delta$ y

$$\overline{P} = P \cup ([x_0, \alpha], \alpha)$$

es δ-fina en $[a, \alpha]$. Por lo tanto $\alpha \in S$.

Para terminar la demostración, probaremos que $\alpha = b$. Por contradicción, supongamos que $\alpha < b$, sea $z \in [a, b]$ tal que

$$\alpha < z < \min\{\alpha + \delta(\alpha), b\}.$$

Entonces $\overline{P} \cup \{([\alpha, z], \alpha)\}$ es una partición etiquetada δ-fina del intervalo $[a, z]$, es decir $z \in S$, lo cual contradice que α es el supremo de S.

□

4. APLICACIONES DEL LEMA DE COUSIN

4.1. Teorema. Sea $f : [a, b] \longrightarrow \mathbb{R}$. Si f es Henstock-Kurzweil integrable sobre $[a, b]$, entonces existe un único A que cumple con:

$$\text{si } (P \ll \delta), \text{ entonces } |S(f, P) - A| < \epsilon.$$

Demostración. Supongamos que existen A_1 y A_2 con $A_1 \neq A_2$ que cumplen la conclusión del teorema. Sea $\epsilon = \frac{|A_1 - A_2|}{2}$, existen δ_1 y δ_2 funciones medidoras, tales que

$$|S(f,P) - A_1| < \epsilon \tag{1}$$

$$|S(f,P) - A_2| < \epsilon \tag{2}$$

cada vez que $P \ll \delta_1$ y $P \ll \delta_2$.

Sea $\delta = \min\{\delta_1, \delta_2\}$ y sea P una partición etiquetada δ-fina (el Lema de Cousin nos garantiza la existencia de al menos una). Entonces

$$|A_1 - A_2| \leq |A_1 - S(f,P)| + |S(f,P) - A_2| < 2\epsilon = |A_1 - A_2|.$$

Esto es una contradición. Por lo tanto $\mathrm{A}_1 = A_2$.

□

Las pruebas de los siguientes teoremas fueron obtenidas de [2, 4].

4.2. Teorema (Teorema del Valor Intermedio (TVM)). Sea $f : [a,b] \longrightarrow \mathbb{R}$ una función continua y L un número entre $f(a)$ y $f(b)$, entonces existe $c \in [a,b]$ tal que $f(c) = L$.

Demostración. Supongamos que $f(a) < f(b)$ y que $f(x) \neq L$ para toda $x \in [a,b]$.

Si $x \in [a,b]$,sea $x \in [a,b]$ se tiene $f(x) > L$ o bien $f(x) < L$,

(1) si $f(x) > L$, por la Proposición 2.9, existe $\delta_x > 0$ tal que $f(t) > L$ para $t \in [a,b]$ que cumpla $|t - x| < \delta_x$.

(2) si $f(x) < L$, por la Proposición 2.8, existe $\delta_x > 0$ tal que $f(t) < L$ para $t \in [a,b]$ que cumpla $|t - x| < \delta_x$.

Sea la función $\delta : [a,b] \longrightarrow (0,\infty)$ definida como $\delta(x) = \delta_x$. Por el Lema de Cousin existe una partición δ-fina

$$P = \{([x_{i-1}, x_i], c_i) \mid i = 1, \ldots, n\},$$

donde, para cada i, $f(t) > L$ para cada $t \in [x_{i-1}, x_i]$, o bien, $f(t) < L$ para cada $t \in [x_{i-1}, x_i]$.

Como $f(a) < L$, entonces $f(t) < L$ para cada $t \in [x_0, x_1]$, f es menor que L, en particular en x_1, esto es $f(x_1) < L$.

Como $f(x_1) < L$, entonces $f(x) < L$ para cada $t \in [x_1, x_2]$, en particular $f(x_2) < L$ y de manera análoga continuamos con el proceso hasta llegar al intervalo $[x_{n-1}, x_n]$, donde $f(x_{n-1}) < L$, entonces $f(t) < L$ para cada $t \in [x_{n-1}, x_n]$, en particular $f(x_n) < L$, pero $x_n = b$ de modo que $f(b) < L$ lo cual contradice la hipótesis. Podemos concluir que existe $c \in [a,b]$ tal que $f(c) = L$.

De manera análaloga se puede ver en el caso de que $f(x) > L$. □

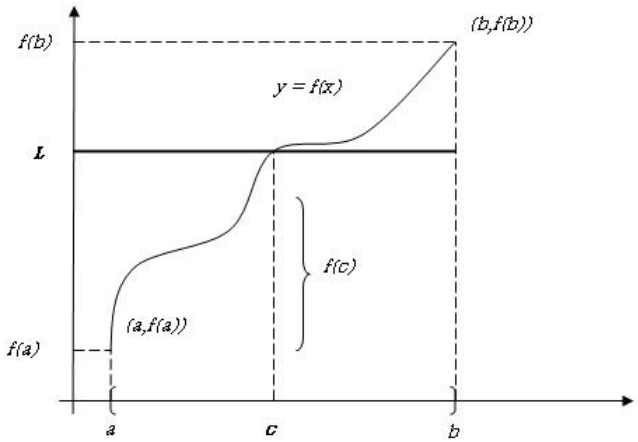


FIGURE 3. Interpretación Geométrica del TVM

4.3. OBSERVACIÓN. Una consecuencia del Teorema del Valor Intermedio es que si $f(a)$ y $f(b)$ tienen signos opuestos, hay cuando menos un número c entre a y b, tal que $f(c) = 0$. Así, si el punto $(a, f(a))$ está abajo del eje x y el punto $(b, f(b))$ está arriba del eje x, o viceversa, la gráfica cruza el eje x cuando menos una vez entre $x = a$ y $x = b$, como se ve en la siguientes figuras.

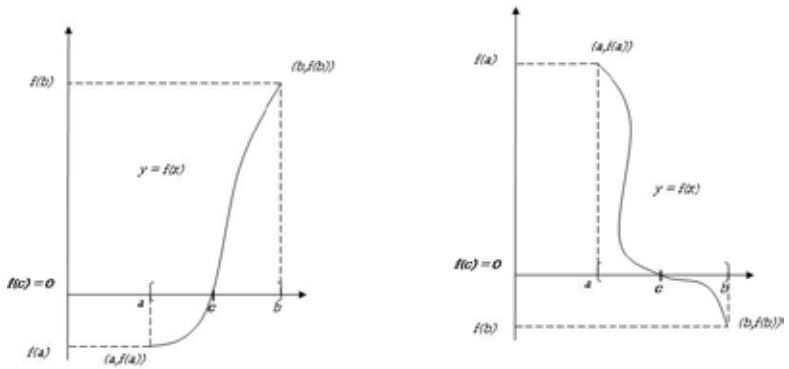


FIGURE 4. Interpretación Geométrica de la Observación

4.4. TEOREMA (Teorema de Weierstrass). Sea $f : [a, b] \longrightarrow \mathbb{R}$ continua en $[a, b]$, entonces existe $M > 0$ tal que $| f(x) | \leq M$ para cada $x \in [a, b]$

DEMOSTRACIÓN. Por la Proposición 2.10 tenemos que para cada $t \in [a, b]$ existe $\delta_t > 0$ y $M_t > 0$ tal que

$$\text{si } x \in [a, b] \text{ y } |x - t| < \delta_t, \text{ entonces } |f(x)| \leq M_t. \tag{3}$$

Definamos $\delta : [a,b] \to (0,\infty)$ como $\delta(t) = \delta_t$, entonces por el Lema de Cousin existe

$$P = \{([x_{i-1}, x_i], t_i) \mid i = 1, 2, \ldots, n\} \ll \delta).$$

Sea $M = \max\{M_{t_i} \mid i = 1, 2, \ldots, n\}$. Sea $x \in [a,b]$, existe $i_o \in \{1, ..., n\}$ tal que $x \in [x_{i_0-1}, x_{i_o}]$, entonces $|x - t_{i_o}| < \delta(t_{i_o})$ entonces, por (3),

$$|f(x)| \leq M_{t_{i_o}} \leq M.$$

□

4.5. Teorema (Continuidad Uniforme). Si $f : [a,b] \longrightarrow \mathbb{R}$ es continua en $[a,b]$, entonces f es uniformemente continua en $[a,b]$.

Demostración. Sea $\epsilon > 0$, como f es continua en $t \in [a,b]$, existe $\delta_t > 0$ tal que

$$\text{si } x \in [a,b] \text{ y } |x - t| < \delta_t, \text{ entonces } |f(x) - f(t)| < \tfrac{\epsilon}{2}.$$

Definimos $\delta : [a,b] \longrightarrow \mathbb{R}$ como $\delta(t) = \frac{\delta_t}{2}$. Por el Lema de Cousin existe

$$P = \{([x_{i-1}, x_i], t_i) \mid i = 1, \ldots, n\} \text{ , } (P \ll \delta).$$

Sea $\delta = \min\{\delta(t_i) \mid i = 1, \ldots, n\}$.

Sean $x, y \in [a,b]$ tal que $|x - y| < \delta$,por demostrar que $|f(x) - f(y)| < \epsilon$.

Existe $1 \leq i \leq n$,tal que $x \in [x_{i-1}, x_i]$,entonces

$$|x - t_i| \quad < \quad \delta(t_i) \quad = \quad \tfrac{\delta_{t_i}}{2}.$$

Probaremos que

$$|y - t_i| < \delta_{t_i}$$

Sabemos que $|x - y| < \delta$, en particular se tiene que:

$$|x - y| \quad < \quad \delta(t_i) \quad = \quad \tfrac{\delta_{t_i}}{2}$$

Luego

$$|y - t_i| \quad = \quad |y - x + x - t_i| \quad \leq \quad |y - x| + |x - t_i| \quad < \quad \tfrac{\delta_{t_i}}{2} + \tfrac{\delta_{t_i}}{2} \quad = \quad \delta_{t_i}.$$

Entonces

$$|f(x) - f(y)| \quad \leq \quad |f(x) - f(t_i)| + |f(t_i) - f(y)| \quad < \quad \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} \quad = \quad \epsilon.$$

Por lo tanto f es uniformemente continua en $[a,b]$.

□

4.6. Observación. Existen otros resultados del análisis real, que se pueden demostrar usando el Lema de Cousin. El contenido de esta memoria será parte de un estudio más extenso sobre este tema que será la base del trabajo de tesis de la licenciatura en matemáticas de la estudiante Ericka Tlatilpa Guarneros.

Referencias

[1] Bartle R.G., *A Modern Theory of Integration*, Grad. Studies Math., Vol. .32, American Math. Soc., Providence, Rhode Island, 2001.
[2] Bosch Girald C., *Las Particiones y el Teorema de Bolzano*, Miscelánea Matemática, 41 (2005), 1-7.
[3] Gordon Russell A., *The Integrals of Lebesgue, Denjoy,Perron, and Henstock*, Graduate Studies in Mathematics, Vol. 4 American Math. Soc., Providence, Rhode Island, 1994.
[4] Gordon Russell A., *The Use of Tagget Partitions in Elementary Real Analysis*, Amer. Math. Monthly 105 (1998), 105-117 and 886.
[5] Vyborny Rudolf, *The Integral:An Easy Approach after Kursweil and Henstock*,Cambridge University Press, 2000.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

gperaggi@fcfm.buap.mx, akcire_eri@hotmail.com, shalybad@hotmail.com

Ecuaciones Diferenciales y Modelación Matemática

Capítulo 4

ESTIMACIÓN DE PARÁMETROS DE UN MOTOR DC CON INTERFAZ PARA LA CONSTRUCCIÓN DEL MODELO DE REFERENCIA DE UN CONTROL ADAPTATIVO

VLADIMIR VASILIEVICH ALEXANDROV
WUIYEVALDO FERMÍN GUERRERO SÁNCHEZ
RIGOBERTO JUÁREZ SALAZAR
JOSÉ JACOBO OLIVEROS OLIVEROS
FCFM - BUAP

Resumen. En este trabajo se presenta la descripción de un motor de corriente directa (DC) al cual se tiene acceso vía una interfaz electrónica, se presenta un modelo que describe tanto al motor DC como la interfaz con la finalidad de estimar solo un conjunto de parámetros suficientes para fines de control y que incluyen tanto los parámetros del motor DC como los de la interfaz; posteriormente se estiman los parámetros aplicando el método recursivo de mínimos cuadrados con factor de olvido. La estimación de los parámetros se realiza en el tiempo que se obtienen los datos (en linea) considerando el problema del desfase en tiempo de las señales involucradas y los problemas de derivación provocados por ruido en las señales. Finalmente con los parámetros estimados se construye un modelo de referencia para la implementación de un control adaptativo.

1. Introducción

Las configuraciones típicas de sistemas de control pueden dividirse principalmente en tres partes: la planta, el controlador y la interfaz entre los dos últimos [1]. En la figura 1 se ilustra una configuración típica de sistemas dinámicos y los sistemas de control con interfaz entre estos; debido a esta configuración, no se puede operar directamente sobre la entrada del sistema, al igual que no se pueden realizar mediciones directamente.

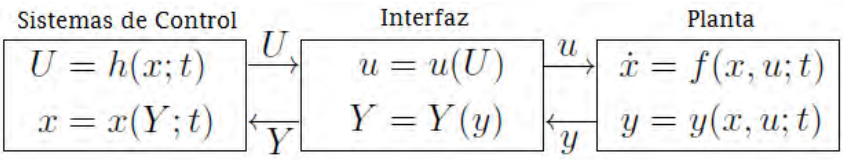


Figura 1. Configuración típica de los sistemas de control.

Las interfaces juegan el papel de acoplamiento entre las señales del controlador (digitales y de baja magnitud[1]) y las del sistema (continuas y de alta magnitud[2]) para ser lo suficientemente potentes para la planta y lo suficientemente bajas para poder ser manipulables por el controlador.

Una de las características de estas interfaces es que son diseñadas de tal manera que los formatos de señales continuas y/o discretas sean transparentes tanto para el controlador como para el sistema y que las operaciones realizables por la interfaz sean estrictamente escalados y traslaciones lineales e invariantes. Por esta razón, se considera una interfaz para un sistema arbitrario que modifica las señales sólo con escalado y traslación de señales como se ilustra en la figura 2.

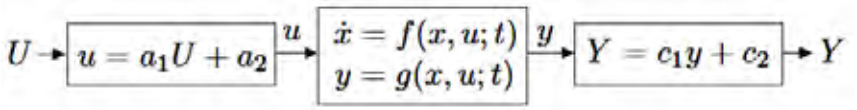


FIGURA 2. Conjunto interfaz de entrada, sistema e interfaz de salida.

Se aplica el método de mínimos cuadrados con factor de olvido [2, 3, 4, 5, 6] para la estimación de los parámetros de un motor de corriente directa (DC) de laboratorio fabricado por la empresa Feedback Instruments Limited [7] diseñado para ser operado desde la plataforma de Matlab Simulink; en la figura 3 se muestra un dibujo del motor con el que se realizaron los experimentos. Éste motor está equipado con un tacómetro analógico, encoders como medidores de posición angular con resolución de $2\pi/32$ rad., un sensor de corriente de armadura, engranaje de reducción de velocidad y freno magnético de operación manual. La tarjeta de adquisición de datos es una PCI-1711 con 16 canales para entradas analógicas, 2 canales para salidas analógicas, 2 canales de 8 bits para entradas digitales, 2 canales de 8 bits para salidas analógicas, un contador de eventos y temporizador programable. Opera con una frecuencia de hasta 100 kHz y los convertidores A/D son de 12 bits.

Se considera modelo lineal de primer orden para el motor DC

$$\dot{\omega} = -b_1\omega + b_2v, \tag{1}$$

donde ω es la velocidad angular del eje del motor, b_1 es el coeficiente de fricción, b_2 una ganancia de amplificación de la entrada y v es el voltaje aplicado. Suponiendo que las interfaces de entrada y salida son invariantes en el tiempo, aportan únicamente amplificación lineal (a_1, c_1) y traslación constante de la entrada (conocida como *offset*) (a_2, c_2) como:

$$v = a_1V + a_2, \tag{2a}$$

$$W = c_1\omega + c_2, \tag{2b}$$

donde W y V son las señales que entrega la interfaz que se corresponden con ω y v respectivamente.

Por la ecuación (2b) se tiene que $\omega = \frac{1}{c_1}(W - c_2)$ y $\dot{\omega} = \frac{1}{c_1}\dot{W}$ entonces:

$$\dot{W} = -b_1W + a_1b_2c_1V + a_2b_2c_1 + b_1c_2 = \eta_1W + \eta_2V + \eta_3, \tag{3}$$

[1]Se expresa "baja magnitud"para generalizar tanto a voltaje, corriente o ambos.

[2]Se expresa "alta magnitud"para generalizar adecuadamente niveles elevados de corriente, voltaje, temperatura, fuerza, etc., dependiendo del sistema.

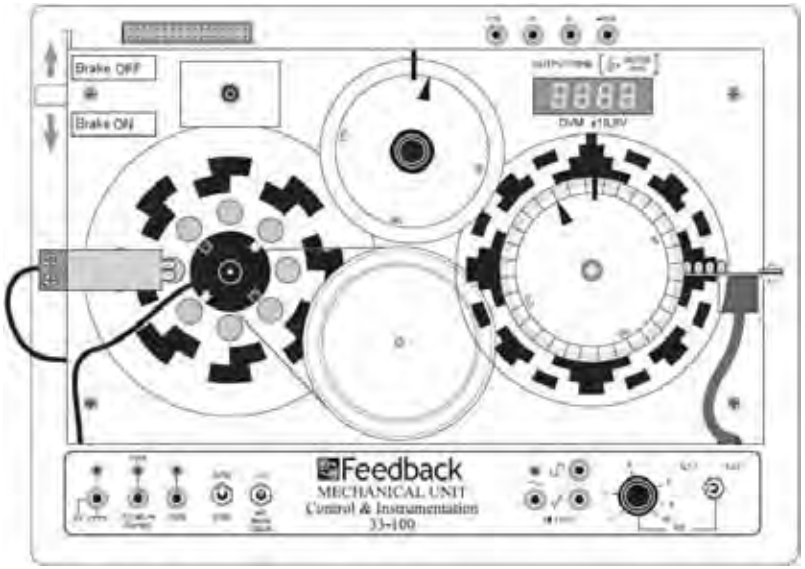


FIGURA 3. Motor de laboratorio empleado para la realización de los experimentos mostrados en este trabajo.

donde $\eta_1 = -b_1$, $\eta_2 = a_1 b_2 c_1$ y $\eta_3 = a_2 b_2 c_1 + b_1 c_2$.

1.1. Pre-filtrado. El sistema sobre el que se trabaja cuenta con un sensor que mide la velocidad a la que gira el eje del motor; sin embargo, la medición incluye error no estacionario. Para obtener $\dot{W}$ no se puede derivar W directamente por la alta sensibilidad de la derivada al ruido en la medición.

Si el ruido en la medición fuera estacionario, entonces se podría derivar directamente; suponga que la medición de la variable $x(t)$ es $\tilde{x}(t) = x(t) + \delta$ donde δ es el término de error constante tan grande como uno quiera, en ese caso, $x' = \tilde{x}'$. Por otro lado, suponga que $\delta(t) = \sin(\phi^2 t)/\phi$, en este caso $\tilde{x}' = x' + \phi\cos(\phi^2 t)$; note que si $\phi \to \infty$ entonces $\delta \to 0$, sin embargo aún siendo el término de error δ muy "pequeño"(o baja amplitud; del orden de $1/\phi$), la alta frecuencia (del orden de ϕ^2) provoca que la derivada $\tilde{x}'$ sea muy distinta de x' (del orden de $\|\phi\cos(\phi^2 t)\|$). Con ésto se puede concluir que, más que la amplitud, es la componente frecuencial del término de error lo que hace que la derivada de la medición con error sea muy distinta respecto a mediciones sin error.

En vista de que el ruido en la señal W es suficiente para que la derivada numérica sea errónea; se aplica un filtro pasa-bajas del tipo Butterworth [8] debido a que presenta respuesta plana; es decir, en el rango de operación se mantiene constante casi hasta la frecuencia de corte, punto donde decrece con cierta pendiente función del orden del filtro; para este caso en particular el filtro de Butterworth es de tercer orden con frecuencia de corte $c_f = 60$ rad/s [9]. Se define la señal $\hat{W} = F(W)$, donde $F(\cdot)$ indica la operación del filtrado.

Las dos principales características de la señal $\hat{W}$ son primeramente la atenuación de las componentes de frecuencia mayor a la frecuencia de corte c_f del filtro, y la segunda es que $\hat{W}$ presenta un retardo τ_F respecto a W que está en función del orden del filtro empleado. Para el proceso de estimación de los parámetros, se requiere

Datos: Θ_{k-1}, P_{k-1}, y_k, φ_k, λ.
Resultado: Θ_k, P_k.

$$K_k = \frac{P_{k-1}\varphi_k^T}{\lambda + \varphi_k P_{k-1}\varphi_k^T};$$
$$P_k = \frac{1}{\lambda}\left(P_{k-1} - K_k\varphi_k P_{k-1}\right);$$
$$e_k = y_k - \varphi_k\hat{\Theta}_{k-1};$$
$$\hat{\Theta}_k = \hat{\Theta}_{k-1} + K_k e_k;$$

Retorna: Θ_k, P_k.

CUADRO 1. Algoritmo de estimación recursiva por mínimos cuadrados con factor de olvido.

que los datos estén en fase; debido al retardo τ_F de $\hat{W}$, es necesario proporcionar el mismo retardo a la señal V para que ambas sigan en fase, por lo tanto se usará la señal $\hat{V} = F(V)$.

1.2. Derivación numérica. Como la señal $\hat{W}$ es discretizada con periodo de muestreo constante T, entonces $\hat{W}(t_k) = \hat{W}(kT)$ y puede calcularse numéricamente la derivada aplicando la forma centrada de orden $O(h^4)$ como [10]:

$$f_0' \approx \frac{-f_2 + 8f_1 - 8f_{-1} + f_{-2}}{12h}, \tag{4}$$

donde $f_i = \hat{W}(t_{k-i})$, $i = \overline{-2, 2}$ y $h = T$.

1.3. Estimación de parámetros. La ecuación (3) es lineal respecto a los parámetros y se puede escribir como

$$y = \varphi\Theta, \tag{5}$$

donde $y = \dot{W}$, $\varphi = \begin{bmatrix} W & V & 1 \end{bmatrix}$ y $\theta = \begin{bmatrix} \eta_1 & \eta_2 & \eta_3 \end{bmatrix}^T$. Discretizando (5), se puede aplicar el método de estimación recursiva de parámetros por mínimos cuadrados con factor de olvido $\lambda \in (0, 1]$ y matriz de covarianza inicial $P_0 = \sigma\mathbb{I}_3$ para λ, $\sigma \in [10^3, 10^6]$, adecuados. Este algoritmo se resume en el cuadro 1.

1.4. Implementación. La aplicación de la derivada numérica (4) requiere retardar $\tau_d = 2h$ la señal $\hat{W}$ para tener disponibles $\hat{W}(t_{k+1})$ y $\hat{W}(t_{k+2})$ en el instante t_k cuando se calcula la derivada. Como el estimador de parámetros no considera desfase de los datos, se propone la referencia temporal $\tilde{t}$ que está retrasada $2h + \tau_F$ respecto a la referencia original t como

$$\tilde{t} = t - \tau_F - \tau_d, \tag{6}$$

y en $\tilde{t}$ se realiza la estimación de los parámetros como

$$\tilde{y} = \tilde{\varphi}\tilde{\theta}, \tag{7}$$

donde

$$\begin{aligned}
\tilde{y} &= \dot{\tilde{W}} = z^{-2}\dot{\hat{W}} = z^{-2}\frac{d}{dt}F(W),\\
\tilde{\varphi}_1 &= \tilde{W}^2 = z^{-2}\hat{W}^2 = z^{-2}F(W),\\
\tilde{\varphi}_2 &= \tilde{W} = z^{-2}\hat{W} = z^{-2}F(W),\\
\tilde{\varphi}_3 &= \tilde{V} = z^{-2}\hat{V} = z^{-2}F(V),\\
\tilde{\varphi}_4 &= 1.
\end{aligned} \tag{8}$$

Una representación gráfica de esta implementación se muestra en la figura 4.

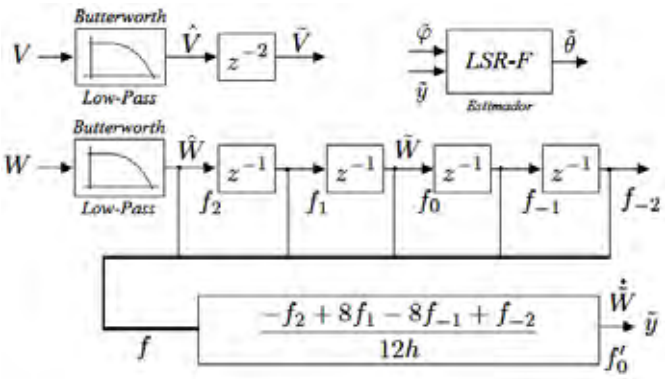


FIGURA 4. Estructura del estimador de parámetros con retardo.

1.5. Resultados experimentales. A continuación, se muestra el resultado de la estimación de parámetros del motor DC. En la figura 5 se muestran las gráficas de evolución de la estimación de parámetros realizando el experimento durante un intervalo $[0, 100]$ seg. con un factor de olvido $\lambda = 0{,}95$ y matriz de covarianza inicial $P_0 = 10^4\mathbb{I}_3$. En la figura 6 se muestra la validación del modelo por comparación directa con los datos experimentales ($\eta_1 = -4{,}13$, $\eta_2 = 3{,}91$, $\eta_3 = -0{,}162$) y una gráfica del error de predicción.

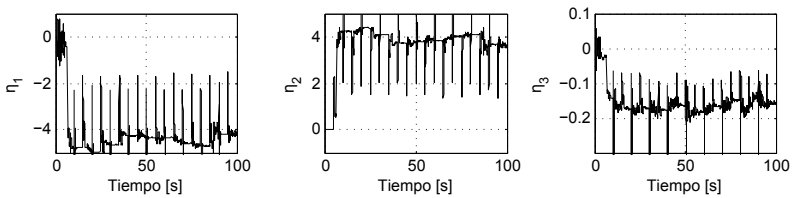


FIGURA 5. Evolución de la estimación de los parámetros η_1, η_2, η_3 considerando periodo de muestreo $T = 0{,}01$ seg., $\lambda = 0{,}95$ y $P_0 = 10^4\mathbb{I}_3$.

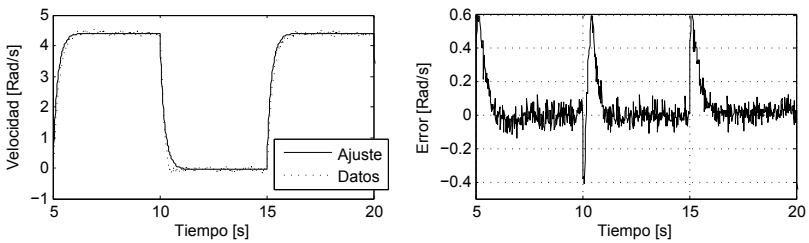


FIGURA 6. Izq.) Validación del modelo por comparación directa con los datos experimentales. Derecha) Error de predicción $e = \omega - \hat{\omega}$.

Es importante remarcar que el parámetro η_3 es pequeño en comparación con los parámetros η_1 y η_2. Por la ecuación (3) se sabe que $\eta_3 = a_2 b_2 c_1 + b_1 c_2$, de la ecuación de la interfaz en la entrada (2a) y la interfaz en la salida (2b) vemos que los términos a_2 y c_2 corresponden a los *offset* que las interfaces aportan; por lo tanto, que η_3 sea "pequeño"sugiere que los *offset* de las interfaces son pequeños o nulos. Para simplificar el modelo, basándonos en este razonamiento consideramos que $a_2 = c_2 = 0 \Rightarrow \eta_3 = 0$. Lo anterior no puede ser una pérdida de generalidad, ya que siempre es posible lograr que los *offset* sean nulos con la traslación (9a) donde se eligen a (9b) como cantidades de desplazamiento.

$$\tilde{v} = v - \delta^v, \quad \tilde{W} = W - \delta^W, \tag{9a}$$

$$\delta^v = a_2, \quad \delta^W = c_2. \tag{9b}$$

2. Estimación de parámetros de un motor DC con interfaz simplificada de entrada y salida

Considerando nuevamente las ecuaciones (2) y suponiendo que los *offset* de la interfaz de entrada y salida son nulos por la aplicación previa de la traslación (9); es posible simplificar el modelo (3) como:

$$\dot{W} = -b_1 W + a_1 b_2 c_1 V = \eta_1 W + \eta_2 V, \tag{10}$$

donde $\eta_1 = -b$ y $\eta_2 = a_1 b_2 c_1$. Aplicando nuevamente todo lo descrito en la sección anterior, pero ésta vez definiendo $y = \dot{W}$, $\varphi = \begin{bmatrix} W & V \end{bmatrix}$ y $\theta = \begin{bmatrix} \eta_1 & \eta_2 \end{bmatrix}^T$. La evolución de la estimación de parámetros se muestra en la figura 7 y en la figura 8 se presenta una gráfica del error de estimación y una comparación directa del modelo ajustado con $\eta_1 = -3{,}53$, $\eta_2 = 3{,}06$ y los datos experimentales.

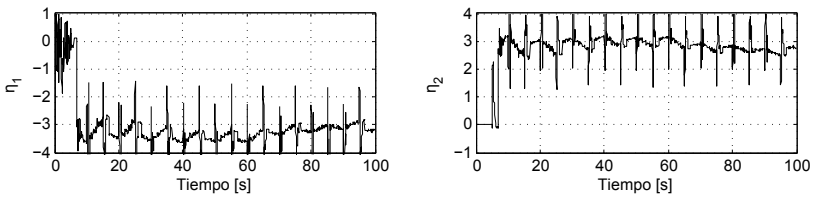


Figura 7. Evolución de la estimación de los parámetros η_1, η_2, considerando periodo de muestreo $T = 0{,}01$ seg., $\lambda = 0{,}95$ y $P_0 = 10^4 \mathbb{I}_3$.

Observe que en esta estimación la pérdida de grados de libertad pone de manifiesto una deficiencia en el modelo (1). En la gráfica del error de la figura 8, en el tiempo $t = 20\ s$ inicia un pico con magnitud cerca de 1, a diferencia de la gráfica del error de la figura 6 donde el error se mantiene acotado en 0,6, lo que hace evidente que el ajuste del modelo se hace más pobre debido a una no linealidad no considerada en el modelo: la planta presenta efectos de un torque elástico que provoca que la velocidad caiga rápidamente a cero (incluso por debajo de él) mientras que el modelo cae exponencialmente lo que provoca que en ese intervalo de transición el error se dispare. Sin embargo, el precio que se ha pagado puede compensarse con la facilidad con la que puede obtenerse una función de trasferencia del modelo

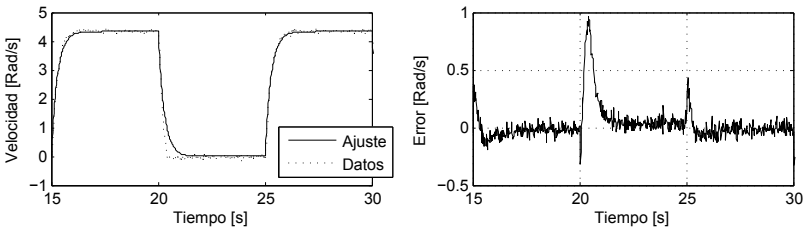


FIGURA 8. Izq.) Validación del modelo por comparación directa con los datos experimentales. Derecha) Error de predicción $e = \omega - \hat{\omega}$.

$\dot{W}(t) = \eta_1 W(t) + \eta_2 V(t) \rightarrow w(s)/v(s) = \eta_2/(s - \eta_1)$, algo que no puede realizarse tan sencillamente para el modelo $\dot{W}(t) = \eta_1 W(t) + \eta_2 V(t) + \eta_3$.

3. Control Adaptativo

En el contexto del control automático el término *adaptativo* se refiere a la facultad de cambiar el comportamiento o parámetros del control en respuesta a cambios en las circunstancias del sistema controlado. Un regulador adaptativo será aquel que pueda modificar su comportamiento en respuesta a cambios en la dinámica del sistema y/o las perturbaciones [4, 5, 6].

Existen varios enfoques para el diseño de control adaptativo, en los que figuran principalmente: ganancia tabulada, modelo de referencia y reguladores autosintonizables. En este trabajo se presenta un control adaptativo por modelo de referencia que consiste en el ajuste de parámetros del controlador tal que la diferencia entre la salida del proceso y la de un modelo prescrito se minimice; esto es conocido como *seguimiento de modelo*. Para lograr éste objetivo existen varios métodos entre los que figuran el método de gradiente negativo del error, teoría de estabilidad de Lyapunov y teoría de de pasividad.

En este trabajo se aplica el método de teoría de estabilidad de Lyapunov. En esta clasificación, podemos encontrar principalmente dos tipos de reguladores: reguladores con parámetros ajustables proporcional al vector de estados y reguladores con adaptación de ganancia, éste último es el que se aplica en este trabajo. Particularmente, este controlador, en [11] es formulado como *control de acción integral* y se desarrolla como solución al problema de perturbaciones en el proceso, derivas e inexactitud en los parámetros estimados e incertidumbres (dinámicas no modeladas).

3.1. Control adaptativo por modelo de referencia mediante el método de estabilidad de Lyapunov para adaptación de ganancia. Se considera el problema de ajuste de una ganancia θ del sistema (11a) y al modelo de referencia (11b).

$$\omega(t) = G(p)\theta v, \tag{11a}$$

$$\omega_m(t) = G(p)\theta_m v, \tag{11b}$$

donde

$$G(p) = \frac{1}{p + b_1};\ p = \frac{d}{dt}, \tag{12}$$

y p es el operador derivada temporal. Note que si las condiciones iniciales con las que se resuelve (11) son cero, entonces $p = s$ donde s es la variable compleja del dominio de la transformada de Laplace; con lo que $G(p) = G(s)$ es una *función de transferencia*.

Se define al error $e = y - y_m = G(p)\theta u_c - G(p)\theta_m u_c = G(p)(\theta - \theta_m)u_c = G(p)\varphi u_c$, donde $\varphi = \theta - \theta_m$. En el espacio de estados tenemos que $\dot{x} = Ax + B\varphi u_c$; $e = Cx$. Eligiendo (13) como función candidata de Lyapunov donde la matriz $P > 0$, obtenemos la regla de ajuste de ganancia como (15) donde γ es conocida como la ganancia de adaptación.

$$V = \frac{1}{2}\left(\gamma x^T P x + \varphi^2\right), \tag{13}$$

$$\dot{V} = -\frac{\gamma}{2} x^T Q x + \varphi\left(\dot{\theta} + \gamma u_c^T B^T P x\right), \tag{14}$$

$$\dot{\theta} = -\gamma u_c^T B^T P x. \tag{15}$$

Si adicionalmente $G(p)$ es *estrictamente positiva real*, entonces existe una matriz P tal que $B^T P = C$, entonces la regla de ajuste de los parámetros puede ser expresada como (16) cuya representación gráfica se muestra en la figura 9.

$$\dot{\theta} = -\gamma u_c^T e. \tag{16}$$

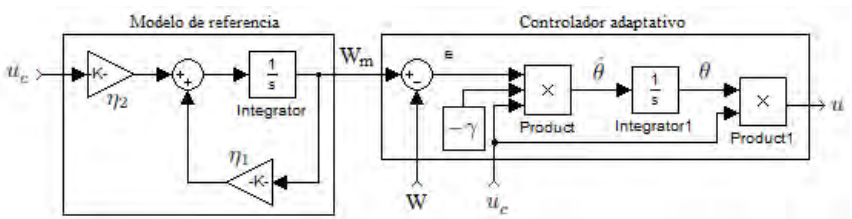


FIGURA 9. Diagrama de bloques de la implementación del control adaptativo por modelo de referencia (16).

Lema 1: (Kalman-Yakubovich [4]**)** Sea el sistema completamente observable y completamente controlable, lineal e invariante en el tiempo (17a).

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx, \end{aligned} \tag{17a}$$

$$G(s) = C(s\mathbb{I} - A)^{-1}B. \tag{17b}$$

La función de transferencia (17b) es estrictamente positiva real si y solo si existen matrices definidas positivas tales que:

$$\begin{aligned} A^T P + PA &= -Q, \\ B^T P &= C. \end{aligned} \tag{18}$$

4. Resultados

La controlabilidad y observabilidad para sistemas lineales se prueban verificando que las matrices $Q = [B, AB, A^2B, \cdots, A^{n-1}]$ y $O = [C^T, A^TC^T, \cdots, (A^{n-1})^TC^T]$ tengan rango máximo, es decir, rank(Q) =rank$(O) = n = 1$, lo cual se satisface pues $A = [\eta_1]$, $B = [\eta_2]$, $C = [1]$ y rank$([\eta_1])$ =rank$([1]) = 1$. Respecto a la estabilidad, se puede aplicar el criterio de Hurwitz, donde el polinomio característico de A es $\lambda = \eta_1$ y como $\eta_1 < 0$ se prueba que la estabilidad es asintótica y global; por el lema 1 se concluye que $G(p)$ es estrictamente positiva real y se puede aplicar la señal de control (16).

Se propone sea el modelo de referencia con la misma estructura que (10) con los parámetros estimados. Se excita al sistema con una señal cuadrada con amplitud igual a 5 c_1^{-1} V y periodo igual a 10 seg., la ganancia de adaptación $\gamma = 0{,}05$ y un periodo de muestreo de 0,001 seg. Se muestran los resultados obtenidos tras un experimento que consiste en la activación y desactivación alternada del freno acoplado al mecanismo del motor. Aplicando la regla de adaptación de ganancia (16) se obtienen los resultados que se muestran en la figura 10.

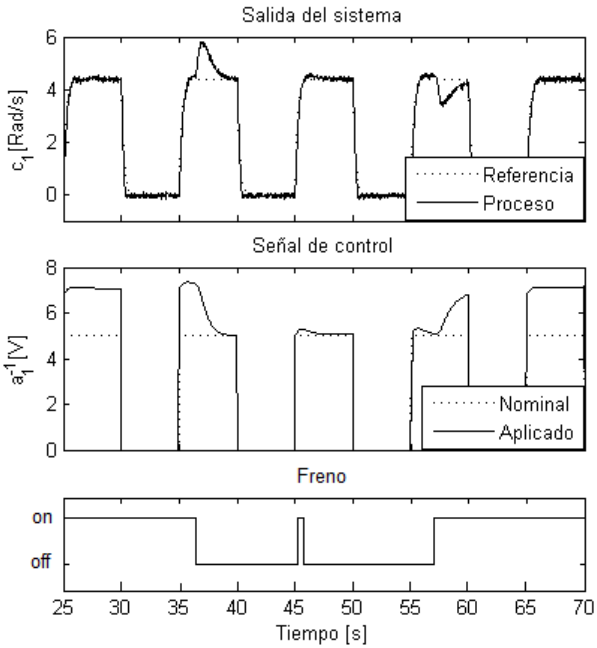


Figura 10. Arriba) Comparación entre la salida del modelo de referencia y la respuesta del sistema. Medio) Comparación entre el control nominal y el control aplicado al sistema. Abajo) Activación y desactivación del freno.

5. Conclusiones

Los parámetros del motor DC considerado varían con el tiempo como se pudo observar en las figuras 5 y 7; sin embargo, el control adaptativo que se aplicó logra mantener al sistema dentro de los márgenes de operación preestablecidos por el

modelo de referencia, aún cuando los parámetros son modificados bruscamente por la aplicación del freno en el mecanismo del motor como se ve en la figura 10. Ésto comprueba la principal propiedad del control adaptativo de *adaptarse* a los cambios de los parámetros del sistema.

Para resolver un problema de control, en general, no se recomienda emplear solo un controlador adaptativo, pues es preferible que el sistema sea, de antemano, asintóticamente estable. Esto no es una pérdida de generalidad, ya que para todo sistema completamente controlable y completamente observable, siempre se puede estabilizar con algún control clásico (conocido como controlador fijo de *primer nivel*) tal como retroalimentación proporcional al vector de estados y, al sistema resultante (asintóticamente estable), aplicarle un controlador adaptativo como control de *segundo nivel* para *robustecer* al controlador de primer nivel.

De la figura 9, el control adaptativo (16) que se aplicó, consiste solo de un sumador, cuatro productos escalares y una operación de integración numérica; por lo que el controlador es muy económico computacionalmente; ésto permite ser implementado en sistemas computacionales mínimos o empotrados. Sin embargo, el modelo de referencia también se debe resolver en tiempo real lo que, para sistemas muy complejos, puede significar una carga computacional considerable; sin embargo, es posible considerar un modelo de referencia con estructura diferente a la del proceso (incluso no ser un sistema de ecuaciones diferenciales), con salida suficientemente suave para mantenerse en la región de atracción del controlador; ésta estructura puede ser un *spline* de grado adecuado.

Referencias

[1] Muhammad H. Rashid, *Electrónica de potencia. Circuitos, dispositivos y aplicaciones*, Prentice Hall, 2002.
[2] Lennart Ljung, *System Identification Toolbox for use with Matlab*, User's guideMathWorks, Inc. 1995.
[3] Lennart Ljung, *System identification: theory for the user*, Prentice Hall. 1999.
[4] Åström, K.J. and Wittenmark, Björn, *Adaptive Control*, Dover Publications. Mineola, N.Y. 2008.
[5] F. Rodríguez R. and M. J. López Sánchez, *Control adaptativo y robusto*, Universidad de Sevilla, 2005.
[6] Petros A. Ioannou and Jing Sun, *Robust Adaptive Control*, Prentice-Hall, Inc., 2003.
[7] User's Manual, *Analogue Servo - Fundamentals Trainer 33-100*, Feedback Instruments Limited., 2002.
[8] Mya Thandar Kyu and Zaw Min Aung and Zaw Min Naing, *Design and Implementation of Active Filter for Data Acquisition System*, International MultiConference of Engineers and Computer Scientists, 2009.
[9] Marcello L. R. de Campos, *Butterworth Filters*, Encyclopedia of Electrical and Electronics Engineering, 1999.
[10] Richard L. Burden and J. Douglas Faires, *Análisis Numérico*, Grupo Editorial Iberoamérica, 1985.
[11] Hassan K. Khalil, *Nonlinear Systems*, rentice Hall. Third Edition, 2002.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

rjuarezsalazar@gmail.com, willi@fcfm.buap.mx, oliveros@fcfm.buap.mx,
vladimiralexandrov366@hotmail.com

Capítulo 5

INESTABILIDAD DE LA CONVECCIÓN NATURAL EN CAVIDADES VERTICALES Y HORIZONTALES LLENAS DE AIRE

ELSA BÁEZ JUÁREZ[3]
MARÍA BLANCA DEL CARMEN BERMÚDEZ JUÁREZ[1]
ALFREDO NICOLÁS CARRIZOSA[2]
[1]FACULTAD DE CIENCIAS DE LA COMPUTACIÓN-BUAP
[2]DEPARTAMENTO DE MATEMATICAS-UAM-I
[3]DEPARTAMENTO DE MATEMÁTICAS APLICADAS Y SISTEMAS-UAM-C

Resumen. En este trabajo se presentan resultados numéricos de problemas de convección natural en cavidades verticales (altas) y horizontales llenas de aire. Se estudia el fenómeno de los ojos de gato a medida que algunos parámetros como la razón geométrica (A) y el ángulo de inclinación de la cavidad varían. Los ojos de gato son una serie de celdas co-rotantes similares sucesivas debidas a perturbaciones en un flujo base. En particular, [1] muestra que el régimen de los ojos de gato sólo puede ser observado en cavidades llenas de aire de razón geométrica mayor a un valor crítico entre 11 y 12.

Los flujos puede ser modelados mediante la aproximación no estacionaria de Boussinesq en la formulación función corriente-vorticidad, la cual es resuelta mediante un proceso iterativo de punto fijo aplicado a un sistema elíptico no lineal que resulta después de una discretización en el tiempo.

Los experimentos se llevan a cabo en cavidades con razones geométricas $A = 16$ y $A = \frac{1}{16}$; los flujos térmicos son convergentes al estado estacionario o bien corresponden a cierto tiempo final T_f. Se consideran números de Rayleigh $Ra = 1{,}1 \times 10^4$, $Ra = 1{,}4 \times 10^4$ y mayores. Matemáticamente, el problema de convección natural se puede modelar mediante la aproximación de Boussinesq no estacionaria en variables función corriente-vorticidad. Los resultados se obtienen usando un método numérico, el cual, después de una discretización apropiada de segundo orden en el tiempo, nos lleva a la solución de un sistema de ecuaciones elípticas no lineales, el cual a su vez es resuelto mediante un proceso iterativo de punto fijo. Entonces, en cada iteración, se tienen que resolver problemas elípticos, lineales, simétricos, bien condicionados y desacoplados.

1. Introducción

El estudio de flujos de convección natural tiene considerable importancia, tanto teórica como práctica y se ha convertido en un problema clásico en mecánica de fluidos [2]. La física involucrada en flujos de convección natural modela muchas aplicaciones en ingeniería: sistemas de almacenamiento de energía, ventilación de edificios, enfriamiento de dispositivos electrónicos, invernaderos, sistemas de energía solar. Existe interés, no sólo en la dinámica y la evolución del fluido, sino también en la transferencia de calor, y cómo éstos se ven afectados por las características del dominio del flujo.

Esta clase de flujos, acoplados térmicamente a fluidos viscosos en un sistema gravitacional, pueden ser modelados mediante la aproximación de Boussinesq no

estacionaria, la cual está basada en el hecho de que las variaciones de temperatura son suficientemente pequeñas, lo cual implica nula variación en la densidad, excepto por la fuerza de flotación en la ecuación de momento, llevando a una estructura incompresible. Además, en este trabajo, se considera la formulación en $2D$ en variables función corriente-vorticidad; la restricción de incompresibilidad se satisface automáticamente y se evita el cálculo de la presión. Los resultados se obtienen mediante el uso de un método numérico que después de una discretización conveniente de segundo orden en el tiempo, lleva a la solución de un sistema de ecuaciones elípticas, no lineal, el cual, a su vez, es resuelto mediante un proceso iterativo de punto fijo. Con este método iterativo se tienen que resolver problemas elípticos, lineales, simétricos, bien condicionados y desacoplados. Para este tipo de problemas existen ya resolvedores eficientes, como Fishpack [9], y Modulef [10] independientemente de la discretización espacial. Este método numérico, previamente reportado en [3] para convección mixta y en [4] para convección natural en cavidades inclinadas, ha mostrado ser suficientemente robusto como para permitir hacer un estudio de los efectos, para flujos de convección natural en cavidades bidimensionales verticales (altas) y horizontales, llenas con aire y calentadas por un lado. Los números de Rayleigh (Ra) considerados, corresponden a $Ra = 1{,}1 \times 10^4$, $Ra = 1{,}4 \times 10^4$ y mayores; la razón geométrica A (A=razón de la altura al ancho) corresponde a $A = 16$ y $A = \frac{1}{16}$.

2. Modelo matemático y método numérico

Sea $\Omega \subset R^N$ $(N = 2, 3)$ la región de un fluido no estacionario, térmico y viscoso, y Γ su frontera. La derivación de las ecuaciones de Boussinesq está basada en cuatro suposiciones acerca de los efectos térmicos y termodinámicos del flujo. La primera suposición es que las variaciones de la densidad son despreciables, excepto por el término de fuerza en la ecuación de momento, el cual está dado por $\rho\mathbf{g}$, donde ρ denota la densidad y $\mathbf{g}$ denota la constante de aceleración debida a la gravedad. Como una simplificación, se puede suponer que la densidad en el término $\rho\mathbf{g}$ está dada por $\rho = \rho_0[1 - \beta(T - T_0)]$, la cual es una aproximación lineal de la ecuación de estado $\rho = \rho(T, p)$ alrededor de una temperatura de referencia T_0 a presión constante; ρ_0 es la densidad en T_0 y β es el coeficiente de expansión térmica. También podemos suponer que, en la ecuación de energía, uno puede despreciar el término de disipación de energía mecánica; y que la viscosidad ν el coeficiente de expansión térmica β, la conductividad térmica κ, el calor específico a presión constante c_p son constantes. Las ecuaciones que resultan de estas suposiciones, son las ecuaciones de Boussinesq.

Si suponemos que hay una escala de longitud l y una temperatura $T_l - T_0$, características, $T_0 < T_l$, inherentes al problema, por ejemplo, la distancia y la diferencia de temperatura entre dos paredes, se define el número de Prandtl adimensional $Pr = \kappa/c_P$, el número de Rayleigh $Ra = \frac{\beta l^3 \kappa g \rho_0^2}{\mu^3 c_p}(T_l - T_0)$. Si además adimensionalizamos de acuerdo a $x \leftarrow x/l$, $u \leftarrow u/U$, $T \leftarrow (T - T_0)/(T_l - T_0)$, $p \leftarrow (p - g \cdot x)/(\rho_0^2 U^2)$ obtenemos el siguiente sistema incompresible adimensional

$$
\begin{aligned}
\mathbf{u}_t - \nabla^2\mathbf{u} + \nabla p + (\mathbf{u}\cdot\nabla)\mathbf{u} &= \frac{Ra}{Pr}\theta\mathbf{e}, \qquad (a)\\
\nabla\cdot\mathbf{u} &= 0, \qquad (b)\\
\theta_t - \frac{1}{Pr}\nabla^2\theta + \mathbf{u}\cdot\nabla\theta &= 0, \qquad (c)
\end{aligned} \tag{1}
$$

en Ω, $t > 0$; donde $\mathbf{u}$, p y θ son la velocidad, presión, y temperatura del fluido respectivamente, $\mathbf{e}$ es el vector unitario en la dirección gravitacional.

El sistema debe ser provisto con condiciones iniciales $\mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x})$ y $\theta(\mathbf{x},0) = \theta_0(\mathbf{x})$ en Ω; y condiciones de frontera, por ejemplo, $\mathbf{u} = \mathbf{f}$ y $B\theta = 0$ en Γ, $t \geq 0$, donde B es un operador de frontera para la temperatura que puede involucrar condiciones de tipo Dirichlet, Neumann y mixto.

Restringiendo las ecuaciones (1a-c) a una región bidimensional Ω, tomando el rotacional en ambos lados de la ecuación (1a) y tomando en cuenta que

$$
u_1 = \frac{\partial\psi}{\partial y}, \quad u_2 = -\frac{\partial\psi}{\partial x}, \tag{2}
$$

lo cual se sigue de (1b), con ψ función corriente y $(u_1, u_2) = \mathbf{u}$; la componente en la dirección $\mathbf{k} = (0,0,1)$ da el sistema escalar

$$
\begin{aligned}
\nabla^2\psi &= -\omega, \qquad (a)\\
\omega_t - \nabla^2\omega + \mathbf{u}\cdot\nabla\omega &= \frac{Ra}{Pr}\frac{\partial\theta}{\partial x}, \qquad (b)\\
\theta_t - \gamma\nabla^2\theta + \mathbf{u}\cdot\nabla\theta &= 0, \qquad (c)
\end{aligned} \tag{3}
$$

donde $\gamma = 1/Pr$ y ω es la vorticidad, la cual, de $\omega\mathbf{k} = \nabla\times\mathbf{u} = -\nabla^2\psi\mathbf{k}$, da $(3a)$ y $\omega = \frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y}$ también. Entonces, el sistema (3) resulta ser la aproximación de Boussinesq en variables función corriente y vorticidad. La condición de incompresibilidad $(1b)$, por (2), se satisface automáticamente y la presión p ha sido eliminada.

Este trabajo trata con convección natural en cavidades rectangulares, entonces las ecuaciones son válidas en $\Omega = (0,a)\times(0,b)$; $a > 0$, $b > 0$. Para construir la condición de frontera para ω, lo cual no es trivial , véase por ejemplo [5], usamos la propuesta dada en [6], extendida a problemas de convección natural en cavidades rectangulares: por expansión de Taylor de ψ en la frontera y usando $(3a)$, se obtienen las siguientes relaciones $O(h_x^2)$(las primeras dos) y $O(h_y^2)$ (las últimas dos), donde h_x y h_y son los tamaños de paso en x y y respectivamente.

$$
\begin{aligned}
\omega(0,y,t) &= -\frac{1}{2h_x^2}[8\psi(h_x,y,t)-\psi(2h_x,y,t)],\\
\omega(a,y,t) &= -\frac{1}{2h_x^2}[8\psi(a-h_x,y,t)-\psi(a-2h_x,y,t)],\\
\omega(x,0,t) &= -\frac{1}{2h_y^2}[8\psi(x,h_y,t)-\psi(x,2h_y,t)],\\
\omega(x,b,t) &= -\frac{1}{2h_y^2}[8\psi(x,b-h_y,t)-\psi(xb-2h_y,t)].
\end{aligned}
\tag{4}
$$

Debe observarse que los valores de frontera para ω son valores dados en Ω y $t>0$, todavía desconocidos de la función corriente ψ. Este problema será resuelto como parte de un proceso iterativo de punto fijo.

En transferencia de calor entre en una frontera (superficie) y un fluido, el número de Nusselt (parámetro adimensional) es la razón de transferencia de calor convectiva a transferencia de calor conductiva através de (o normal a) la frontera. Un número de Nusselt cercano a la unidad es característico de flujo laminar. Un número de Nusselt grande corresponde a convección más activa, con flujo turbulento típicamente en el rango de 100 y 1000.

El número de Nusselt local Nu mide la transferencia de calor en cada punto de la pared donde la temperatura mayor es dada y el número de Nusselt global $\overline{Nu}$ mide la transferencia de calor promedio en la pared. Estos parámetros adimensionales están definidos por

Número de Nusselt local:

$$Nu(x) = -\frac{\partial\theta}{\partial y}|_{y=0,b} \quad \text{y} \quad Nu(y) = -\frac{\partial\theta}{\partial x}|_{x=0,a}$$

Número de Nusselt global:

$$\overline{Nu}\,|_{y=0,b} = \frac{1}{A}\int_0^a Nu(x)\,dx \quad \text{ó} \quad \overline{Nu}\,|_{x=0,a} = \frac{1}{A}\int_0^b Nu(y)\,dy$$

Las derivadas temporales de ω y θ en (3) son aproximadas mediante la aproximación de segundo orden siguiente

$$f_t(\mathbf{x},(n+1)\Delta t) \approx \frac{3f^{n+1}-4f^n+f^{n-1}}{2\Delta t}, \tag{5}$$

donde $n \geq 1$, $\mathbf{x} \in \Omega$, $\Delta t > 0$ es el paso de tiempo, y $f^r \approx f(\mathbf{x}, r\Delta t)$; en cada nivel de tiempo $t=(n+1)\Delta t$ se obtiene un sistema semidiscreto, en Ω, con sus correspondientes condiciones de frontera en Γ, el cual resulta ser

$$
\begin{aligned}
\nabla^2\psi^{n+1} &= -\omega^{n+1}, & \psi^{n+1}|_\Gamma &= 0,\\
\alpha\omega^{n+1} - \nabla^2\omega^{n+1} + \mathbf{u}^{n+1}\cdot\nabla\omega^{n+1} &= \frac{R_a}{Pr}\frac{\partial\theta^{n+1}}{\partial x} + f_\omega, & \omega^{n+1}|_\Gamma &= \omega_{bc}^{n+1},\\
\alpha\theta^{n+1} - \gamma\nabla^2\theta^{n+1} + \mathbf{u^{n+1}}\cdot\nabla\theta^{n+1} &= f_\theta, & B\theta^{n+1}|_\Gamma &= 0.
\end{aligned}
\tag{6}
$$

donde $\alpha = \frac{3}{2\Delta t}$, $f_\omega = \frac{4\omega^n - \omega^{n-1}}{2\Delta t}$, y $f_\theta = \frac{4\theta^n - \theta^{n-1}}{2\Delta t}$; ω_{bc} denota la condición de frontera de ω dada en (4), B denota el operador de frontera para θ mencionado arriba, y las componentes u_1 and u_2 of $\mathbf{u}$, en términos de ψ, están dadas por (2).

Después de renombrar $(\psi^{n+1}, \omega^{n+1}, \theta^{n+1})$ por (ψ, ω, θ) se obtiene un sistema elíptico no lineal de la siguiente forma

$$
\begin{aligned}
\nabla^2\psi &= -\omega, & \psi|_\Gamma &= 0 & (a),\\
\alpha\omega - \nabla^2\omega + \mathbf{u}\cdot\nabla\omega &= \frac{Ra}{Pr}\frac{\partial\theta}{\partial x} + f_\omega, & \omega|_\Gamma &= \omega_{bc} & (b),\\
\alpha\theta - \gamma\nabla^2\theta + \mathbf{u}\cdot\nabla\theta &= f_\theta, & B\theta|_\Gamma &= 0 & (c).
\end{aligned} \tag{7}
$$

Para obtener $(\omega^1, \theta^1, \psi^1)$ en (6), se puede usar una aproximación de primer orden para las derivadas a través de una subsucesión con un Δt más pequeño; también se obtiene un sistema estacionario de la forma (7).

Denotando

$$
\begin{aligned}
R_\omega(\omega,\psi) &\equiv \alpha\omega - \nabla^2\omega + \mathbf{u}\cdot\nabla\omega - \frac{Ra}{Pr}\frac{\partial\theta}{\partial x} - f_\omega,\\
R_\theta(\theta,\psi) &\equiv \alpha\theta - \gamma\nabla^2\theta + \mathbf{u}\cdot\nabla\theta - f_\theta.
\end{aligned}
$$

Entonces, el sistema (7) es equivalente, en Ω, a

$$
\begin{aligned}
\nabla^2\psi &= -\omega, & \psi &= 0 \quad \text{sobre} \quad \Gamma,\\
R_\theta(\theta,\psi) &= 0, & B\theta &= 0 \quad \text{sobre} \quad \Gamma,\\
R_\omega(\omega,\psi) &= 0, & \omega &= \omega_{bc} \quad \text{sobre} \quad \Gamma.
\end{aligned} \tag{8}
$$

Para resolver (8), en cada nivel de tiempo $(n+1)\Delta t$, se aplica el siguiente proceso iterativo de punto fijo, en Ω:

Con $\{\theta^0, \omega^0\} = \{\theta^n, \omega^n\}$ dados, se resuelve, hasta tener convergencia en θ y ω,

$$
\begin{aligned}
\nabla^2\psi^{m+1} &= -\omega^m, \quad \psi^{m+1} = 0 \ \text{ sobre } \ \Gamma,\\
\theta^{m+1} &= \theta^m - \rho_\theta(\alpha I - \gamma\nabla^2)^{-1} R_\theta(\theta^m, \psi^{m+1}),\\
B\theta^{m+1} &= 0 \ \text{ sobre } \ \Gamma, \quad \rho_\theta > 0,\\
\omega^{m+1} &= \omega^m - \rho_\omega(\alpha I - \nabla^2)^{-1} R_\omega(\omega^m, \psi^{m+1}),\\
\omega^{m+1} &= \omega_{bc}^{m+1} \ \text{ sobre } \ \Gamma, \quad \rho_\omega > 0,
\end{aligned} \tag{9}
$$

y se toma $(\omega^{n+1}, \psi^{n+1}, \theta^{n+1}) = (\omega^{m+1}, \psi^{m+1}, \theta^{m+1})$.

Cuando decimos, "hasta tener convergencia", nos referimos a que dos valores consecutivos de θ (y ω), o sea, θ^{m+1} y θ^m (y de ω^{m+1} y ω^m respectivamente), no difieran en más de una cierta tolerancia, tol, dada, por ejemplo, $tol = 10^{-7}$.

Debe hacerse notar que la construción de la condición de frontera de ω, ω_{bc} en (4), dada implícitamente por valores desconocidos de ψ en Ω, se realiza como parte del proceso iterativo en (9). Finalmente, el sistema (9) es equivalente a

$$
\begin{aligned}
\nabla^2\psi^{m+1} &= -\omega^m, \quad \psi^{m+1}=0 \text{ en } \Gamma,\\
(\alpha I-\gamma\nabla^2)\theta^{m+1} &= (\alpha I-\gamma\nabla^2)\theta^m-\rho_\theta R_\theta(\theta^m,\psi^{m+1}),\\
B\theta^{m+1} &= 0 \text{ en } \Gamma, \quad \rho_\theta>0,\\
(\alpha I-\nabla^2)\omega^{m+1} &= (\alpha I-\nabla^2)\omega^m-\rho_\omega R_\omega(\omega^m,\psi^{m+1}),\\
\omega^{m+1} &= \omega_{bc}^{m+1} \text{ en } \Gamma, \quad \rho_\omega>0.
\end{aligned} \tag{10}
$$

Luego entonces, en cada iteración, se tienen que resolver en Ω, *tres* problemas elípticos lineales y desacoplados asociados con los operadores ∇^2, $\alpha I-\gamma\nabla^2$, y $\alpha I-\nabla^2$.

Para la discretización de los problemas elípticos, como los de (10), puede utilizarse o bien diferencias finitas o elemento finito, si se consideren dominios rectangulares.

Para el caso de elemento finito, deben escogerse formulaciones variacionales y restringirlas a los espacios de elementos finitos de dimensión finita, como aquellos en [7] y [8]. Para los resultados específicos que se muestran en este trabajo, se usa la aproximación de segundo orden de Fishpack[9]. Luego, dicha aproximación de segundo orden en espacio, combinada con la aproximación de segundo orden en (5) para las primeras derivadas en tiempo, la aproximación de diferencias centrales de segundo orden en los puntos interiores, y con (5) en la frontera, para todas las primeras derivadas en el espacio, y la regla trapezoidal de segundo orden para calcular el Número de Nusselt global $\overline{N}u$ implica que el problema discreto completo se basa en discretizaciones de segundo orden solamente.

3. Resultados numéricos

Los resultados se llevan a cabo en cavidades verticales (altas) y horizontales, con razones geométricas A=16 y $A=\frac{1}{16}$ llenas con aire ($Pr=0{,}71$) y corresponden a flujos convergentes al estado estacionario o bien a un cierto tiempo final T_f, mostrando, en estas condiciones, que el flujo es dependiente del tiempo.

En este trabajo se consideran números de Rayleigh $Ra=1{,}1\times10^4$, $Ra=1{,}4\times10^4$ (y mayores). Se usan los siguientes valores de h_x y h_y:

1. $(h_x,h_y)=(1/32,16/512)$,
2. $(h_x,h_y)=(1/48,1/768)$,
3. $(h_x,h_y)=(1/64,1/1024)$;

para 1) y 2), $\Delta t=0{,}0001$, mientras que para 3), $\Delta t=0{,}00001$.

Para justificar que los resultados obtenidos son correctos, se realizaron estudios sobre la independencia del tamaño de la malla en términos del error relativo L_∞ discreto, punto a punto, en la cerradura de la cavidad $\overline{\Omega}$, o sea, se busca el máximo, en valor absoluto, de la diferencia entre los valores obtenidos usando una malla y la otra. Los resultados con las mallas anteriores se muestran en la siguiente tabla.

malla	$\psi - error$	$\theta - error$
1) Vs 2)	0.46 %	0.16 %
1) Vs 3)	0.6 %	0.2 %
2) Vs 3)	0.17 %	0.06 %

Tabla 1. Independencia de la malla: $Ra = 1{,}1 \times 10^4$ con $A = 16$

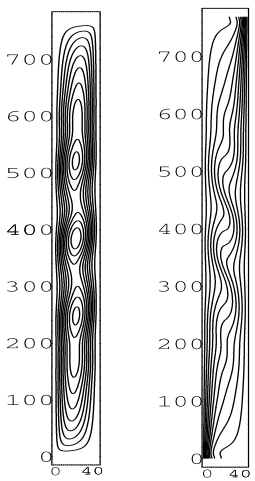


FIGURA 1. $Ra = 1{,}1 \times 10^4$; $(h_x, h_y) = (1/48, 16/768)$, $\Delta t = 0{,}0001$ a $T_{ss} = 5{,}68$

En la figura (1) se muestran resultados para $Ra = 1{,}1 \times 10^4$, y razón geométrica $A = 16$, con $h_x = 1/48$ y $h_y = 16/768$. El resultado converge al estado estacionario $T_{ee} = 5{,}68$. Como puede observarse, para este número de Rayleigh aparecen tres ojos de gato en las líneas de corriente, a la izquierda, mientras que en las isotermas, derecha, aparecen ondulaciones, asociadas a los ojos de gato.

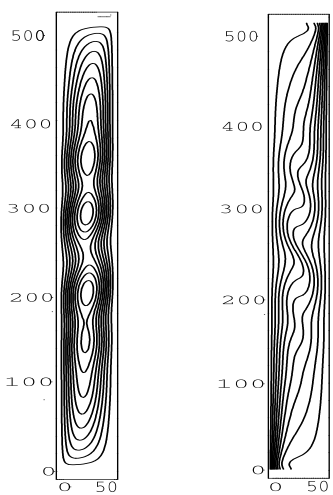


FIGURA 2. $Ra = 1{,}4 \times 10^4$; $(h_x, h_y) = (1/32, 16/512)$, $\Delta t = 0{,}0001$, a $T_f = 30$.

Ahora, en la figura (2), para un número de Rayleigh mayor, $1{,}4 \times 10^4$, con $h_x = 1/32$ y $h_y = 16/512$, al tiempo final $T_f = 30$, aparece un ojo de gato bien formado, y dos que tienden a desaparecer. Las ondulaciones, en las isotermas también tienden a desaparecer. Entonces, a medida que el Ra crece, podemos inferir, que los ojos de gato tienden a desaparecer.

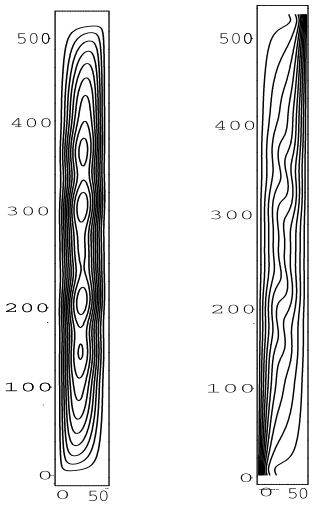


FIGURA 3. $Ra = 1{,}4 \times 10^4$; $(h_x, h_y) = (1/32, 16/512)$, $\Delta t = 0{,}0001$, a $T_f = 50$.

En la figura (3) se muestran los resultados obtenidos para el mismo número de Rayleigh que en la figura anterior, Fig. 2, pero ahora a $T_f = 50$. Los ojos de gato y las ondulaciones tienden a desaparecer aún más.

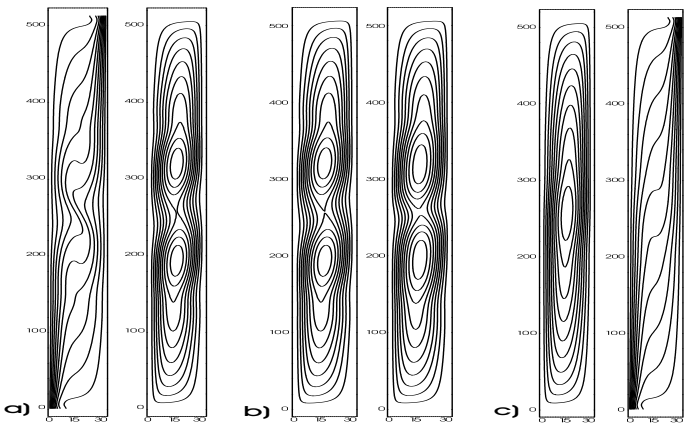


FIGURA 4. a) $Ra = 2{,}775 \times 10^4$ (IS-SL), b) $Ra = 2{,}790625 \times 10^4$ (SL)-$Ra = 3{,}60501747 \times 10^4$ (SL), c) $Ra = 3{,}60501747375 \times 10^4$ (SL-IS)

En la figura (4), se muestran en a), las isotermas (izquierda) y las líneas de corriente (derecha), para $Ra = 2{,}775 \times 10^4$, un número de Rayleigh mayor al de

la figura anterior. Pueden observarse sólo dos ojos de gato, y las oscilaciones de las isotermas, asociadas a los ojos de gato, también tienden a desaparecer. En b), el número de Rayleigh es mayor, $Ra = 2{,}790625$ (izquierda) y $Ra = 3{,}60501747$ (derecha). Para estos casos se muestran sólo las líneas de corriente (SL), y pueden verse también sólo dos ojos de gato. Para c), $Ra = 3{,}60501747375 \times 10^4$ se muestran las líneas de corriente (izquierda) y las isotermas (derecha), y como puede verse, los ojos de gato desaparecen y también las ondulaciones en las isotermas.

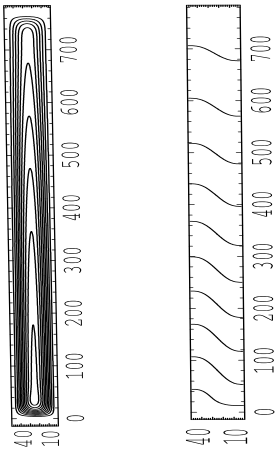


FIGURA 5. $Ra = 1{,}1 \times 10^4$; $(h_x, h_y) = (16/768, 1/48)$, $\Delta t = 0{,}0001$ a $T_f = 9$.

En la figura (5), se muestran los resultados para $Ra = 1{,}1 \times 10^4$ y una cavidad horizontal, $A = \frac{1}{16}$ a $T_f = 9$. En este caso, no se observa el fenómeno anterior de los ojos de gato en las líneas de corriente (izquierda) ni las ondulaciones en las isotermas (derecha).

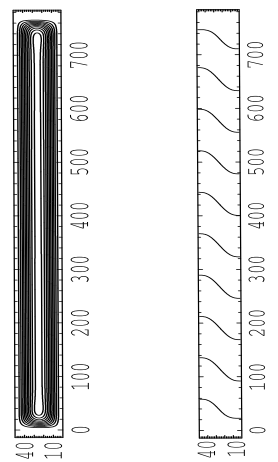


FIGURA 6. $Ra = 1{,}1 \times 10^4$; $(h_x, h_y) = (16/768, 1/48)$, $\Delta t = 0{,}0001$ a $T_f = 30$.

En la figura (6), se muestran los resultados para el mismo número de Rayleigh que en la figura anterior y la misma razón geométrica. En las líneas de corriente

(izquierda), puede observarse una especie de cotonete, y las isotermas (derecha) no presentan ondulaciones.

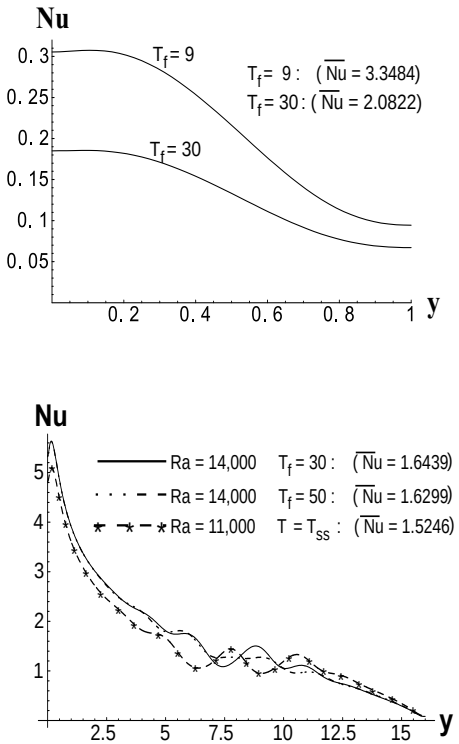


FIGURA 7. $Ra = 1{,}1 \times 10^4$ con $A = 1/16$ (arriba). $A = 16$: $Ra = 1{,}4 \times 10^4$ y $Ra = 1{,}1 \times 10^4$ a $Tss = 5{,}68$ (abajo)

Por último, en la figura (7), se presentan (arriba) las gráficas de los números de Nusselt locales, y los números de Nusselt globales para $Ra = 1{,}1 \times 10^4$ con $A = \frac{1}{16}$ a $T_f = 9$ y 30 respectivamente. Puede observarse que la transferencia de calor es mayor para $T_f = 9$. Abajo aparecen los números de Nusselt globales y las gráficas de los números de Nusselt globales para $Ra = 1{,}4 \times 10^4$ a $T_f = 30$ y 50, y para $Ra = 1{,}1 \times 10^4$ al estado estacionario. Puede verse que para $Ra = 1{,}4 \times 10^4$, el número de Nusselt global es mayor que para $T_f = 50$ y que para $Ra = 1{,}1 \times 10^4$.

4. Conclusiones

Se han presentado resultados en cavidades verticales (altas) y horizontales con razones geométricas $A = 16$ y $\frac{1}{16}$ para números de Rayleigh $Ra = 1{,}1 \times 10^4$, $Ra = 1{,}4 \times 10^4$, y mayores. Se trata de ver como la estructura de ojos de gato en cavidades altas permanece, cambia o desaparece. Los resultados muestran que para algunos Ra's el flujo es dependiente del tiempo cuando la razón geométrica A cambia de vertical a horizontal, o bien, si Ra se incrementa.

Cálculos preliminares muestran para Ra más grandes, con $A = 16$, que una vez que la estructura de ojos de gato desaparece en el estado estacionario, se obtiene

una estructura bien formada, lo cual no tiene nada que ver con una estructura periódica mencionada por otros autores [11].

Con respecto a la sensibilidad a errores del esquema que se propone en este trabajo, cabe hacer notar que al discretizar los problemas y resolver el sistema de ecuaciones resultante, si el Δt es pequeño, se trabaja con una matriz que resulta ser diagonalmente dominante y como se sabe, para este tipo de matrices, no se tiene ningún problema al resolver el sistema de ecuaciones asociado. Entonces, el esquema planteado, trabajando con un Δt adecuado y también con un tamaño de discretización espacial (h_x y h_y) adecuado no presenta problemas de convergencia, aunque sí hay que ser cuidadoso en la elección de los parámetros antes mencionados.

Referencias

[1] Roux, B., Grondlin, J., Bontoux, P., and Vahl Davis, G., Reverse Transition from Multicellular to Monocellular Motion in Vertical Fluid Layers, Phys. Chem. Hydro., Vol 3F, (1980) 292-297.

[2] P. Le Quéré, T. Alziary de Roquefort, Computation of Natural Convection in Two-Dimensional Cavities with Chebyshev Polynomials, J. of Computational Physics 57 (1985) 210-228.

[3] A. Nicolás, B. Bermúdez, 2D Thermal/Isothermal Incompressible Viscous Flows, Int. J. for Num. Meth. in Fluids 48 (2005) 349-366.

[4] E. Báez, A. Nicolás, 2D natural convection flows in tilted cavities: porous media and homogeneous fluids, Int. J. of Heat and Mass Transfer 49 (2006) 4773-4785.

[5] R. Peyret, T.D. Taylor, Computational Methods for Fluid Flow, Springer-Verlag, NY (1983).

[6] A. Nicolás, B. Bermúdez, 2D Incompressible Viscous Flows at Moderate and High Reynolds Numbers, Computer Methods in Engineering and Sciences, 16 (5) (2004) 441-451.

[7] M. D. Gunzburger, Finite Element Methods for Viscous Incompressible Flows: A guide to theory, practice, and algorithms, Academic Press, INC. (1989).

[8] R. Glowinski, Hanbook of Numerical Analysis: Numerical Methods for Fluids (Part 3), North-Holland Ed. (2003).

[9] . Adams, P. Swarztrauber and R. Sweet, FISHPACK: A Package of Fortran Subprograms for the Solution of Separable Elliptic PDE's, The National Center for Atmospheric Research, Boulder, CO, USA (1980).

[10] H.F. Du Toit, P.L. George, P. Laug, P. Paté, D. Steer, M. Vidrascu, An Introduction to MODULEF, MODULEF User Guide n 1, INRIA (1991).

[11] P. Le Quéré, A note on multilpe and unsteady solutions in two-dimensional convection in a tall cavity, J. of Heat Transfer, 112 (1990) 965-974.

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN-BUAP
Av. San Claudio y 18 Sur, Ciudad Universitaria, Col. Jardines de San Manuel
CP. 72570, Puebla, Pue. México.

bbj@solarium.cs.buap.mx, anc@xanum.uam.mx, ebj@xanum.uam.mx

Capítulo 6

LA SUMABILIDAD DE BOREL EN LA SOLUCIÓN DE ECUACIONES DIFERENCIALES

LAURA ANGELICA CANO CORDERO
FCFM - BUAP

Resumen. El presente manuscrito pretende dar un esbozo sobre el uso de los criterios de sumabilidad, en particular, el de Borel, en la resolución de ecuaciones diferenciales de segundo orden. Así como mostrar algunas de sus aplicaciones.

1. Introducción

Comencemos el presente artículo planteando una pregunta clásica de nuestros cursos de Cálculo Integral. Dada una serie $\alpha_n = \sum_i^n a_i$, $a_i \in \mathbb{R}$ ó $a_i \in \mathbb{C}$, ¿es convergente y a qué valor converge, es decir, $\lim_{n\to\infty} \alpha_n = L$, para algún $L \in \mathbb{C}$ ó $L \in \mathbb{R}$?.

Esta pregunta nos causa muchos estragos y en el proceso de entender criterios de convergencia, también aprendemos algunas series con nombre propio. Algunos ejemplos de series que nos son familiares son:

1.1. Ejemplo.

(1) (Serie Armónica)

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n},$$

(2) (Serie alternada)

$$1 - \frac{1}{2} + \frac{1}{3} - \cdots = \sum_{n=1}^{\infty} (-1)^n \frac{1}{n},$$

(3) (Función Gamma)

$$1^{-s} + 2^{-s} + 3^{-s} + \cdots = \sum_{n=1}^{\infty} n^{-s}, \text{ donde } s \in \mathbb{C} \text{ y } Re(s) > 0,$$

(4) (Heaviside)

$$\sum_{-\infty}^{\infty} \frac{x^{c-r}}{\Gamma(c-r+1)}, \text{con } c,\ r > 0.$$

Para establecer la convergencia de ésta y otras series, aprendemos algunos criterios de convergencia.

1.2. Ejemplo. Criterios de convergencia Sea $\sum_{n=0}^{\infty} \alpha_k$

(1) **Condición del resto.** Para que una serie sea divergente, una condición suficiente es que

$$\lim_{k\to\infty} \alpha_k = 0.$$

Esta afirmación es muy útil, ya que nos ahorra trabajo en los criterios cuando el límite es distinto de cero.

(2) **Criterio de la razón.** Supongamos que los términos de la serie son positivos. Si existe

$$\lim_{k\to\infty} \frac{a_k}{a_{k+1}} = L \text{ con } L \in [0,\infty)$$

el Criterio la razón establece que:

(a) si $L < 1$, la serie converge.
(b) si $L > 1$, entonces la serie diverge.
(c) si $L = 1$, no es posible decir algo sobre el comportamiento de la serie.

(3) **Criterio de Cauchy** Supongamos que los términos a_k son positivos que

$$\lim_{k\to\infty} \sqrt{k}a_k = L,\ L \in [0,\infty).$$

Entonces, si:
(a) $L < 1$, la serie es convergente.
(b) $L > 1$ entonces la serie es divergente.
(c) $L = 1$, no podemos concluir nada a priori y tenemos que recurrir al criterio de Raabe, o de comparación, para ver si podemos llegar a alguna conclusión.

(4) **Criterio de Raabe.** En algunas series, puede ocurrir que ni el criterio de la razón ni el de la raíz nos permitan determinar la convergencia o divergencia de la serie, entonces recurrimos al **criterio de Raabe.**

Sea una serie tal que $a_k > 0$ (serie de términos positivos). Y supongamos que existe

$$\lim_{k\to\infty} k(1 - \frac{a_k}{a_{k+1}}) = L,$$

con $L \in [0,\infty)$ Por tanto, si $L > 1$, entonces la serie es convergente y si $L < 1$, la serie es divergente.

(5) Una serie de la forma $\sum_{n=1}^{\infty} (-1)^n a_n$ se llama alternada. Tal serie converge si se cumplen las siguientes condiciones:
(a) $\lim_{k\to\infty} (-1)^n a_n = 0$ para n par y n impar.
(b) La serie tiene que ser absolutamente decreciente, es decir:

$$\mid a_k \mid \geq \mid a_{k+1} \mid .$$

Si esto se cumple, la serie es condicionalmente convergente, de lo contrario la serie diverge.

Lo dicho hasta aquí, sugiere que nuestro tema se ha agotado ahora que tenemos criterios de convergencia. Sin embargo, este estudio solo es una parte puesto que la mayor parte de la series que se obtienen de las aplicaciones, son divergentes en todos los puntos, un ejemplo de ellas en la serie de Heaviside del Ejemplo 1.1 inciso (4), la cual Heaviside obtuvo en sus estudios en Electromagnetismo. Más aún, los siguientes teoremas nos muestran la importancia de tener series convergentes.

1.3. Teorema.

(1) (Fourier.) Toda función periódica real se puede escribir como una combinación lineal infinita de senos y cosenos.
(2) Toda función analítica es holomorfa, y recíprocamente.
(3) La solución de ecuaciones diferenciales de segundo orden se puede dar por medio de serie de potencias.

Esta problemática fue causa de estudio de matemáticos como **A. Cauchy,** entre otros, quienes inmersos en esta problemática se plantearon la necesidad de replantear el concepto de convergencia de una serie. Bajo esta luz, surgen las siguientes definiciones de sumación de una serie, conocidas hoy en día como condiciones de **sumabilidad**.

2. Tipos de Sumación

A continuación definiremos formalmente los conceptos de sumación más importantes.

2.1. Definición.

(1) **Sumación de Cesàro.** Sea $\{a_k\}$ una sucesión, siendo

$$s_k = \sum_{n=1}^{k} a_n,$$

la suma k-ésima de los primeros k términos de la serie.

La sucesión $\{a_n\}$ se denomina **sumable Cesàro,** con una suma de Cesàro a α, si

$$\lim_{k\to\infty} \frac{1}{k} s_k = \alpha.$$

(2) **Sumación de Abel.** Sea $\sum_{n=0}^{k} a_n$ una serie y sea (r_n) una sucesión de números tal que $r_n \to 1^-$ y sea $k \leq 0$. Decimos que la serie es **Abel sumable** con suma de Abel igual a α si:

$$\lim_{n\to\infty} \sum_{n\geq 0} r_n^k s_k = \alpha.$$

(3) **Sumación de Euler.** Sea $\sum_{n=0}^{k} a_n$ una serie, la sumación de Euler de esta serie se define como:

$$\sum_{j=0}^{\infty} a_j := \sum_{i=0}^{\infty} \frac{1}{(1+y)^{i+1}} \sum_{j=0}^{i} \binom{i}{j} y^{j+1} a_j.$$

2.2. Observación. Un tipo de sumación de vital importancia para nuestro propósito es la **sumación de Borel**. Sin embargo hablaremos con mayor detalle de la misma más adelante.

Las definiciones anteriores nos llevan a plantearnos de manera natural las siguientes preguntas:

¿Al considerar estas definiciones de sumaciones en una serie, las series divergentes en el sentido usual serán convergentes en alguno de los sentidos anteriores?

¿Las series convergentes en el sentido usual lo serán según el sentido de sumabilidad anteriores?

La respuesta a ambas preguntas es afirmativa; respecto a la primera los siguientes ejemplos nos serán útiles.

2.3. Ejemplo.

(1) La sucesión de Grandi, definida por $a_n = (-1)^{n+1}$
(2) (Teorema de Féjer) Sea $f \in L^1[-\pi, \pi]$ una función 2π periódica entonces en cada punto t donde $f(t^-)$ y $f(t^+)$ existe, la serie de Fourier para f es Césaro sumable a $\frac{f(t^-)+f(t^+)}{2}$.
(3) La serie de Fourier de una función $f \in L[-\pi, \pi]$ 2π periódica es Abel sumable a $f(t)$ para casi todo número real t, es decir, si $u(r,t) = \frac{a_0}{2} + +(a_n \cos nt + b_n \sin nt)r^n$, $0 < r < 1$, entonces $u(r,t) \to f(t)$ cuando $r \to 1$ para casi todo t.

Respecto a la segunda pregunta, se deja al lector como un ejercicio, el cual le permitirá resfrescar sus conocimientos de cálculo.

2.4. Observación. Basta mostrar la afirmación anterior para el caso de la sumabilidad de Cesáro pues la sumabilidad de Cesàro implica la sumabilidad de Abel, el lector interesado puede consultar [2] para tener una demostración detallada.

3. Ecuaciones Diferenciales

En [5] Euler investigó el problema de la suma de la serie formal

$$s = 1 - 2 + 6 - 24 + 120 - ...$$

y de hecho se este estudio lo generalizó a la siguiente serie

$$\hat{f} := \sum_{k=0}^{\infty} k!(z)^{k+1},\ z > 0. \tag{1}$$

Más aún, Euler nota que $\hat{f}$ es la solución para de la ecuación diferencial de segundo grado:

$$z^2 y' + y = z. \tag{2}$$

Sin embargo, (1) es una serie divergente para $x \in \mathbb{R}^+ \backslash \{0\}$ (esta última afirmación se puede verificar usando el criterio de la razón). Entonces para estos valores de x

no es posible tener una aproximación de la solución.

Un primer intento para resolver este problema es utilizar el **método de cuadratura**, es decir, dar una solución de la ecuación diferencial mediante la combinación finita de integrales de funciones simples. Como lo muestra la siguiente proposición.

3.1. Proposición. Para todo $x \geq 0$

$$| f(x) - f_k(x) | \leq k! x^{k+1},$$

donde

$$f(x) = e^{\frac{1}{x}} \int_0^x \frac{e^{\frac{1}{-y}}}{y} dy,$$

y $f_k(x) = \sum_{n=0}^{k+1} (-1)^n n! x^{n+1}$, f es una solución para la ecuación diferencial de Euler.

Demostración. Consideramos la siguiente fórmula, la cual es sencilla de verificar,

$$\frac{1}{1+\psi} = \sum_{n=0}^{k-1} (-1)^n \psi^n + (-1)^k \frac{\psi^k}{1+\psi}.$$

Entonces

$$\begin{aligned} f(x) &= \int_0^\infty e^{\frac{-\psi}{x}} \left(\sum_{n=0}^{k-1} (-1)^n \psi^n + (-1)^k \tfrac{\psi^k}{1+\psi}\right) d\psi \\ &= \sum_{n=0}^{k-1} \int_0^\infty e^{\frac{-\psi}{x}} \psi^n + \int_0^\infty (-1)^k \tfrac{\psi^k e^{\frac{-\psi}{x}}}{1+\psi} d\psi. \end{aligned}$$

Usando la siguiente igualdad

$$\Gamma(n+1) = n! = \int_0^\infty z^n e^{-z} dz, \tag{3}$$

la cual implica:

$$\int_0^\infty \psi^n e^{\frac{-\psi}{x}} d\psi = n! x^{n+1}, \tag{4}$$

así

$$f(x) = \sum_{n=0}^{k-1} (-1)^n x^{n+1} n! + \int_0^\infty (-1)^k \frac{\psi^k e^{\frac{-\psi}{x}}}{1+\psi} d\psi = f_k(x) + R_k(x), \tag{5}$$

donde $R_k = f - f_k$. De esto,

$$| R_k(x) | = \int_0^\infty \frac{\psi^k e^{\frac{-\psi}{x}}}{1+\psi} d\psi \leq \int_0^\infty \psi^k e^{\frac{-\psi}{x}} = k! x^{k+1}. \blacksquare \tag{6}$$

Sin embargo, esta solución parece muy artificial, por lo que este proceso resulta ser poco viable para el caso de una perturbación de la ecuación de Euler. Por lo que de manera natural podríamos preguntarnos ¿cómo obtener una solución de la forma (5) para la ecuación generalizada de Euler? Para este caso ¿por qué no aplicar algún criterio de sumabilidad a nuestro problema? Los tipos de sumabilidad

mencionados anteriormente no proporcionan una sumabilidad que de una solución a nuestro problema; por ello es necesario introducir un nuevo criterio de sumabilidad.

3.2. Definición. Sumación de Borel
Sea $\sum_{n=0}^{\infty} a_n$ una serie entonces la **sumación de Borel S** de la serie se define como:

$$S = \sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \frac{a_n}{n!} n! = \sum_{n=0}^{\infty} \frac{a_n}{n!} \int_0^{\infty} z^n e^{-z} dz = \int_0^z (\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n) e^{-z} dz \tag{7}$$

Con la ayuda de (7) podemos introducir el concepto de sumabilidad de Borel.

3.3. Definición. Una serie $\sum_{n=0}^{\infty} a_n$ se dice **Borel sumable** si:

(1) La serie $\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$ es convergente.

(2) El radio de convergencia se puede extender a todo el eje real positivo.

(3) La integral $\int_0^z (\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n) e^{-z} dz$ es convergente.

Anteriormente vimos que la sumabilidad de Abel es más débil que la sumabilidad de Cesàro, por lo que podemos preguntarnos ¿la sumabilidad de Borel es más débil que la sumabilidad de Cesàro, o será una generalización de la misma?

Para responder a esta pregunta necesitamos la siguiente definición.

3.4. Definición. Sean $\{v_n\}_{n\in\mathbb{N}}$ y $\{p_n\}_{n\in\mathbb{N}}$ dos sucesiones de números reales. Entonces la **sumación de Cesàro de** v_n **con pesos** p_n se define como:

$$w_n := \frac{p_1 v_1 + p_2 v_2 + ... + p_n v_n}{p_1 + p2 + ... + p_n}. \tag{8}$$

3.5. Observación. Claramente si consideramos $p_n = 1,\ \forall n \in \mathbb{N}$ tenemos la sumabilidad de Cesàro.

3.6. Proposición. La sumabilidad de Borel es la de Cesàro con pesos $p_n = \frac{\lambda^n}{n!}$.

Demostración. La veracidad de esta afirmación se sigue del hecho que la serie $\sum_{n=0}^{\infty} p_n$ es convergente y converge a e^{λ}. ■

Pero cómo verificar que realmente este método resuelve nuestro problema para el caso de la ecuación de Euler. Para ello son necesarios algunos cálculos que a continuación hacemos.

3.7. Ejemplo. Sea $a_n := (-1)^n n! x^{n+1}$. Entonces

$$\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n = x \sum_{n=0}^{\infty} (-1)^n \frac{a_n}{n!} (xz)^n = \frac{x}{1+xz}.$$

Así

$$\int_0^z \left(\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n \right) e^{-z} dz = \int_0^{\infty} \frac{xe^{-z}}{1+xz} dz = \int_0^{\infty} \frac{e^{\frac{-\psi}{z}}}{1+x\psi} d\psi,$$

la cual es la solución de la Proposición 3.1.

Más aún, los Teoremas Tauberianos nos permiten hablar de la extensión del dominio de convergencia de una serie en direcciones; concepto que a continuación definimos.

3.8. Definición.

(1) Una serie de potencias $\sum_{n=0} a_n z^{n+1}$ es **1-sumable en la dirección** d, donde d es la semirrecta del origen en el plano complejo si la serie $\sum_{n=0}^{\infty} a_n \frac{\psi^n}{n!}$ es convergente y la integral

$$\int_0^z \left(\sum_{n=0}^{\infty} \frac{a_n \psi^n}{n!} \right) e^{\frac{-\psi}{x}} d\psi,$$

es convergente con valor $S(x)$.

(2) Si una serie no es 1-sumable en la dirección d, entonces la dirección d se llama **dirección excepcional**.

(3) Una serie es **1- sumable** si es 1-sumable en todas las direcciones d excepto en número finito de direcciones excepcionales.

Hasta lo expuesto pareciera que la sumación de Borel es adecuada para encontrar una expansión asintótica para la solución de la Ecuación de Euler, sin embargo, la proposición siguiente nos muestra que la misma técnica se puede aplicar para un espacio de funciones.

3.9. Teorema. Consideremos una ecuación algebraica diferencial

$$F(x, y, y', ..., y^{(m)}) = 0,$$

donde F es un polinomio de varias variables. Si $\hat{f}(x) = a_n x^n$ es una solución formal de la ecuación diferencial y $\hat{f}$ es borel absolutamente sumable con suma de Borel $f(x)$, entonces $f(x)$ es una solución de la ecuación diferencial y tiene expansión asintótica a $\hat{f}$.

El lector interesado puede encontrar en [7] una demostración de esta proposición.

Nuestro análisis nos has llevado a una solución asintótica de la ecuación de Euler en el eje real, el cual corresponde en el plano complejo a $\{z \in \mathbb{C} : Re(z) > 0\}$, por lo que al considerar dicha solución como un subconjunto del plano complejo es natural plantearse la posibilidad de aplicar las herramientas del Análisis Complejo para dar una prolongación analítica de dicha solución. Para ello realicemos el siguiente análisis.

Una forma simple de obtener una prolongación analítica de

$$\int_0^{\infty} e^{x\psi} \phi(\psi) d\psi \tag{9}$$

es considerar una rotación por ángulo θ, de la trayectoria de integración, como se muestra en la figura.

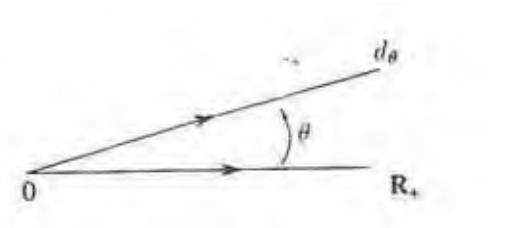


FIGURA 1. Rotación por ángulo θ de la trayectoria de integración

La integral $\int_d \theta e^{-x\psi}\phi(\psi)d\psi$ define una función $\psi^\theta(x)$ en el semiplano P_θ, donde P_θ se define como

$$P_\theta := \{x \mid Re(x \cdot \psi) > 0,\ \forall \psi \in d_\theta\},$$

y $Re(x \cdot \psi)$ no es otro que el producto escalar hermitiano $< x, \overline{\psi} >$.

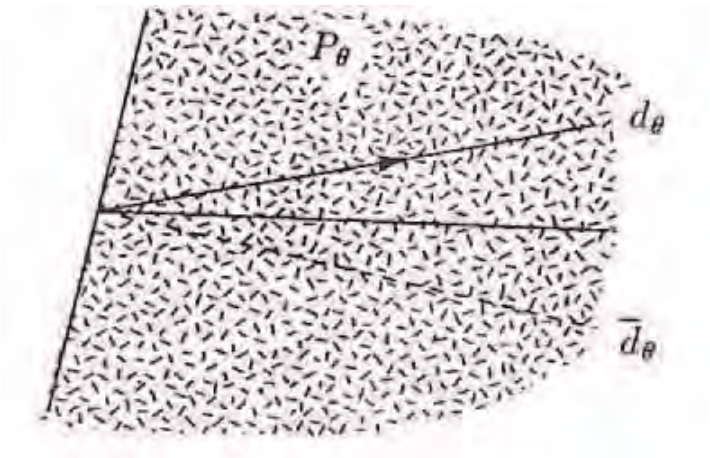


FIGURA 2. Semiplano P_θ

Por el teorema de Cauchy, $\hat{f}$ y ϕ^θ coinciden en la intersección de los semiplanos P_θ y $P_0 := \{Re(z) > 0\}$. Así cuando realizamos una rotación en sentido contrario a las manecillas del reloj obtenemos una prolongación analítica de $\hat{f}$.

Por ende, esta construcción nos provee de una prolongación analítica de $\hat{f}$ en todo $\mathbb{C}$ excepto cuando $\theta = \pi$, pues en $\psi = -1$ $\hat{f}$ tiene una singularidad. Así, excepto en una dirección podemos hablar de una prolongación de la solución de $\hat{f}$. Sin embargo, de manera casi natural podemos plantearnos lo siguiente: Si para obtener una prolongación analítica de $\hat{f}$ en el plano complejo una herramienta útil para hacerlo fue la deformación de la trayectoria en donde conocíamos la solución asintótica, ¿podemos deformarla de manera tal que evitemos la singularidad y tener

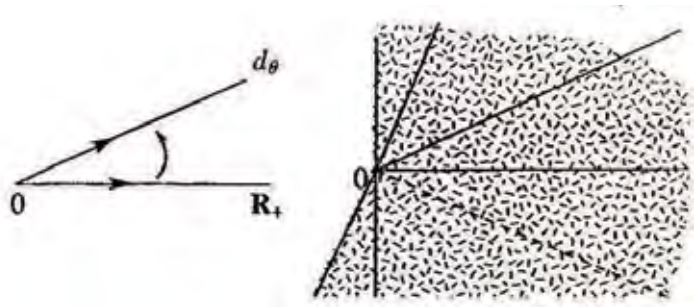


FIGURA 3. Dominio de la prolongación analítica de $\hat{f}$

información en esa dirección? Una tentativa inicial sería considerar la trayectoria inicial y solo deformarla en -1 ¿cómo? pues mediante una rotación en la misma, como se muestra en la siguiente figura.

FIGURA 4. Trayectoria de la integral cuando existe una singularidad

Si continuamos este procedimiento, de manera iterada obtenemos que para el semiplano $Re(z) > 0$ una prolongación analítica de $\hat{f}$ dada por

$$\phi_+^\pi(x) = \int_{d_\pi^+} e^{-x\psi}\phi(\psi)d\psi,$$

donde d_π^+ es el camino que evita la singularidad de la ecuación, siendo como se muestra a continuación .

FIGURA 5. Trayectoria para la prolongación analítica de $\hat{f}$

Más aún, para todo x en el semiplano $Re\ (z) > 0$ tenemos dos prolongaciones analíticas de $\hat{f}$ dados por las figuras (3) y (4). Por lo que debemos considerar el hecho de que estas prolongaciones no sean iguales, cuestión que resolvemos a continuación.

$$(\phi_+^\pi - \phi_-^\pi)(x) = \int_\gamma e^{-\psi x}\frac{1}{1+\psi}d\psi = -2i\pi e^x, \ \ \forall\, x : Re\ (x) > 0$$

FIGURA 6. Trayectoria $d_\pi^+ - d_\pi^-$

donde γ es la trayectoria $d_\pi^+ - d_\pi^-$, donde $d_\pi^+ - d_\pi^-$ es un camino de la forma:
Por lo tanto, $\phi_+^\pi \neq \phi_-^\pi$.

Lo anterior lo podemos resumir en la siguiente proposición.

3.10. PROPOSICIÓN.

(1) Las sumas de series 1-sumables en las diferentes direcciones d dan funciones las cuales son continuaciones analíticas una de la otra conforme movemos la línea d continuamente sobre las direcciones en las cuales las series son 1-sumables. Esto provee una función definida en un sector con vértice en el origen.
(2) La suma de Borel de una serie de potencias divergente puede no ser uniforme en una vecindad del origen. Es necesariamente ramificada. Este comportamiento se conoce en la literatura como el **fenómeno de Stokes**.
(3) Si una serie $\sum_{n=0}^{\infty} a_n x^{n+1}$ tiene un radio de convergencia r y su suma es una función f para $\mid x \mid < r$, entonces un teorema de análisis complejo establece que la función f tiene al menos una singularidad en el círculo $\mid x \mid = r$. La idea de Borel es que una serie divergente es una serie con radio de convergencia $r = 0$. Así tenemos al menos una singularidad escondida en una dirección: para la ecuación de Euler es la dirección $\mathbb{R}^-$.

4. POLÍGONO DE BOREL

El inciso (3) de la Proposición (3.10) no da la información sobre la existencia de una singularidad para la suma f, y hemos estudiado a dicha función utilizando trayectorias que no pasan por la singularidad. Por lo que debemos considerar el comportamiento de f en una vecindad de f.

Para ello, supongamos que c es una singularidad de f, y que dicha función tiene radio de convergencia $\mid z \mid = c$, alrededor de un punto P. En este caso,

$$J(z) := c^{-1} \int e^{-t(1-\frac{z}{c})} dt,$$

la cual es convergente si y sólo si $Re(\frac{z}{c}) < 1$, i.e., si z y el origen están en el mismo lado de la línea LP que pasa a través de P perpendicular a OP. La región así definida están al interior de un polígono convexo, el cual puede ser cerrado o abierto y puede ser homotópico a un ángulo, una banda o un medio plano, dicho polígono es llamado un **polígono de Borel**.

4.1. Ejemplo. Sea $f(z) = \frac{1}{1-z^2}$ entonces la frontera del polígono de Borel está formado por las líneas $x = \pm 1$.

A continuación explicaremos una de las bondades que presentan las funciones Borel sumables, en el polígono de Borel.

Supongamos que $f(u)$ es regular, i.e., u no es singularidad de f, supongamos además que se encuentra al interior de una cerrada K que contiene al 0 tal que para todos los puntos $z \in K$ se cumpla que

$$Re(\frac{z}{u}) \leq 1 - \delta < 1. \tag{10}$$

Entonces

$$f(z) = \frac{1}{2i\pi} \int \frac{f(u)}{u-z} du = \frac{1}{2i\pi} \int \frac{f(u)}{u} du \int e^{-t+t\frac{z}{u}} dt, \tag{11}$$

la última igualdad está dominada por

$$\frac{1}{2\pi} \int \frac{\mid f(u) \mid}{\mid u \mid} \mid du \mid \int e^{-\delta t} dt.$$

Por lo que podemos invertir el orden de las integrales; y obtenemos

$$f(z) = \int e^{-t} dt \frac{1}{2\pi i} \int \frac{f(u)}{u} e^{t\frac{z}{u}} du = \int e^{-t} I(t,z) dt. \tag{12}$$

Dado que $f(u)$ es regular al interior de K y $e^{t\frac{z}{u}}$ es regular excepto en 0, podemos calcular $I(t,z)$ contrayendo K a una curva $K^{'}$ al interior del círculo de convergencia de $f(u)$. Las series para $f(u)$ y $e^{t\frac{z}{u}}$ son uniformemente convergentes en $K^{'}$, y entonces

$$I(t,z) = \frac{1}{2\pi i} \int_{K'} \sum a_n u^n \sum \frac{1}{n!} \Big(\frac{tz}{u}\Big]\Big)^n \frac{du}{u} = \sum a_n \frac{(tz)^n}{n!} = a(tz).$$

Así

$$f(z) = \int e^t a(tz) dt.$$

El lector interesado puede consultar [6] para conocer las propiedades que presentan las funciones Borel sumables al interior del polígono de Borel, así como su relación con la sumación de Borel de la cual solo dimos la definición.

5. Aplicaciones

El estudio de la sumabilidad de Borel no solo se restringe a la solución de ecuaciones diferenciales, también ha encontrado un campo fértil en el Análisis Matemático, rama en la que se realiza un estudio abstracto análogo al de la teoría de Fourier para funciones 2π periódicas.

5.1. Definición. La transformación

$$\phi(x) = \sum_{n=0}^{\infty} a_n x^{n+1} \ \mapsto = \sum_{n=0}^{\infty} \frac{a_n \xi^n}{n!} := \phi(\psi)$$

que asocia a una serie formal en términos de $\frac{1}{x}$ una serie entera en términos de ψ es llamada **transformación de Borel**.

Más aún, en dicha transformación podemos definir la operación multiplicación, la cual se define como:

$$\phi \cdot \psi \mapsto \phi * \psi,$$

donde

$$(\psi * \phi)(\alpha) = \int_0^{\alpha} \psi(u)\phi(\alpha - u)du,$$

y se denomina la **convolución de la transformada de Borel** o bien **multiplicación para la transformación de Borel**.

5.2. Observación. Una de las peculiaridades de la transformada de Borel es que con esta operación, la multiplicación entre dos funciones se define como la multiplicación entre gérmenes analíticos alrededor de 0, [1].

Este tema es muy extenso y dar una exposición detallada va más allá de este artículo, pues requiere de temas más avanzados de Análisis que los aquí expuestos. El lector interesado en profundizar en este tema puede consultar [3].

En la Física ha repercutido en el estudio de la siguientes series, cuyo significado físico no discutiremos pues rebasa el contenido del presente trabajo.

(1) Series Lindstedt
(2) Solución de la ecuación 1D de Schröndinger.

Pero lo más interesante de este tema, es que la sumación de Borel comenzamos estudiándola en la modelación de un sistema dinámico continuo, que se estudia mediante la ecuación de Euler, ésta misma tiene aplicación en los sistemas dinámicos discretos en el estudio de un fenómeno conocido como **universalidad de las funciones unimodales** que fuera descubierto por Feigenbaum en 1978, [4].

Referencias

[1] Allan Clark, *Elements of abstract algebra*, Dover, 1984.
[2] Bachman George, Narici Lawrence, Beckenstein Lawrence *Fourier and Wavelet Anaylisis*, Springer- Verlag, 2000.
[3] Costin Ovidiu, *Asymptotics and Borel summability (Monographs and surveys in pure and applied mathematics ; 141)*, CRC Press, 2008.
[4] Eckmann Jean Pierre, *Computer Methods and Borel Summability Applied to Feigenbaum's Equation* (with P. Wittwer). Lecture Notes in Physics, Springer-Verlag, Berlin Heidelberg New York, Vol. 227 (1985).
[5] Euler L., *De seriebus divergentibus, Novi Commentarii academiae scientiarum Petropolitanae*(1754-55) 1760,pp. 205-237, reprinted in Opera Omnia Series, I vol. 14, pp. 585-617.
[6] Hardy Godfrey Harold, *Divergent series,* 2 Edition, AMS Bookstore, 2000.
[7] Rousseau Christiane, *Divergent series: past, present, future . . .*, Internal Communication Département de mathématiques et de statistique and CRM Université de Montréal .

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.
caclmx@yahoo.com.mx

Capítulo 7

ALCANCES Y LIMITACIONES DEL CÓMPUTO CIENTÍFICO: UN EJEMPLO

MARIO ALBERTO CARBALLO FLORES
REYNALDO DOMÍNGUEZ CASTILLO
FRANCISCO SERGIO SALEM SILVA
FACULTAD DE MATEMÁTICAS - UNIVERSIDAD VERACRUZANA

Resumen. En este trabajo hecharemos un vistazo a algunas capacidades y limitaciones del computo científico. En Partícular veremos que no se puede invertir una matriz de dimensión 10000 por 10000 aún siendo esta tridiagonal con valores constantes en la diagonal, en la subdiagonal y la superdiagonal (matriz que resulta al resolver la ecuación de Poisson de dimensión 2), esto si tratamos de resolverlo con un programa ingenuo y no tomamos en cuenta la estructura de la matriz. Por otro lado daremos una impresionante aplicación de la transformada rápida de Fourier que implicitamente resuelve un sistema de ecuaciones lineales del orden de 11000 por 11000. Este sistema resulta cuando queremos interpolar una cierta función y queremos que el error sea del mismo orden que la epsilon de la máquina. Los cálculos serán realizados en Python y Matlab.

1. Introducción

Actualmente la existencia de equipos de cómputo, con grandes capacidades de almacenamiento y de cálculo, hace que en ocasiones pensemos que el cómputo científico es algo mágico. Cuántas veces no hemos oído decir a alguien: Qué chiste tiene resolver un sistema de ecuaciones, si con un simple programa puedo obtener la solución, hay personas supuestamente con buen bagaje matemático que comentan a alumnos -pues está muy bien que hayas analizado este algoritmo, pero actualmente Mathematica puede calcular el 50,000-avo primo siguiente-, o decir -que maple puede elevar un número a la 512-ava potencia modulo 345, sin ningún problema-, lo que no dicen es que esto puede tomar horas o que los algoritmos para lograr esto tienen que ser tremendamente bien diseñados y recordamos que muchos de los algoritmos que hoy en día usamos han sido el resultado de un proceso inventivo de mucha gente, que evitaban en lo posible acumular errores de redondeo y lograr un ahorro considerable de almacenamiento, así como una optimización del uso del procesador. Este trabajo es una mirada un tanto retrospectiva de lo que ha sido el cómputo científico, pues aquí presentamos dos ejemplos en los que si los algoritmos se emplean de manera ingenua aún con computadoras con gran capacidad de almacenamiento y un buen procesador podemos no tener buenos resultados. En particular usamos dos computadoras, la computadora 1(Memoria RAM de 2GB, versión 7.8.0.347 de MATLAB, Windows7.) y la computadora 2 (Memoria RAM de 4GB, versión 2.6 de Python, Ubuntu 10.04). En el primer ejemplo observaremos que en Matlab no es posible cargar una matriz de 10,000 x 10,000 mucho menos resolver un sistema de ecuaciones lineales que determina la matriz propuesta. Cuando usamos la computadora 2 si podemos almacenar la matriz y resolver una sistema

de ecuaciones lineales, usando Python, pero aún tenemos el problema de que el tiempo de ejecución es muy grande, en ambos casos (Matlab y Python) atacamos el problema aprovechando la estructura de nuestra matriz -tridiagonal-. El problema de almacenamiento como el tiempo de proceso se solucionan, aprovechando que la matriz es dispersa. En el segundo ejemplo atacamos el problema de la interpolación polinomial y recordamos que las matrices de Vandermonde que surgen de este problema aún de dimensiones pequeñas están mal condicionadas[2], en este caso abordaremos el problema usando polinomios de Chebychev y el polinomio trigonométrico equivalente para poder usar el algoritmo de la transformada rápida de Fourier, y así poder obtener un polinomio de grado de más de 11000 (necesario para lograr una precisión del orden del épsilon de la máquina), en un tiempo impresionantemente corto .

2. Falta de Almacenamiento

A continuación, mostramos que las capacidades del espacio de trabajo como la capacidad de almacenamiento y el sistema de programación pueden resultar inútiles si no sabemos aprovechar la estructura de las matrices que resultan del proceso de discretización de un problema dado. Un primer ejemplo que resulta interesante explorar es el siguiente:

2.1. Ejemplo. La ecuación de Poisson.
Sea $L > 0$

$$-y''(x) = f(x), \quad x \in (0, L). \tag{1}$$

Esta ecuación diferencial ordinaria de segundo orden con valores en la frontera describe la distribución de la temperatura y en una barra de longitud L, f es una función que representa la fuente de calor a lo largo de la barra. Además, $y(0) = y_1$ y $y(L) = y_2$.

Resolvemos el problema usando el método de diferencias finitas. Dividimos el intervalo $[0, L]$ en $n + 1$ subintervalos de longitud $h = \frac{L}{n+1}$, así los extremos de los subintervalos son $x_j = j \cdot h$, $j = 0, 1, ..., n + 1$. Buscamos una solución aproximada en los puntos x_j, es decir, $y_j \approx y(x_j)$ con $j = 0, 1, ..., n + 1$. Para aproximar y'' usamos la siguiente fórmula conocida como diferencias centrales. Debemos aclarar que los elementos de la solución sólo contiene los valores en el intervalo $(0, L)$, en los extremos los valores son proporcionados por las condiciones de frontera.

$$y''(x) \approx \frac{y(x + h) - 2y(x) + y(x - h)}{h^2}. \tag{2}$$

Entonces el problema se reduce a resolver un sistema de ecuaciones lineales $Ay = r$ de la forma, donde los r_i son los valores de f evaluados en x_i.

$$\begin{pmatrix} d_1 & a_1 & 0 & ... & 0 \\ b_2 & d_2 & a_2 & ... & 0 \\ & & \vdots & & \\ 0 & ... & b_{n-1} & d_{n-1} & a_{n-1} \\ 0 & ... & 0 & b_n & d_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{n-1} \\ r_n \end{pmatrix} \tag{3}$$

La matriz A resulta ser tridiagonal, donde las entradas de las diagonales son los únicos elementos distintos de cero (A también es llamada matriz banda). En la mayoría de los casos la matriz A resulta ser muy grande si queremos una buena aproximación a la solución del problema. Como caso particular, resolveremos el siguiente sistema que está definido por una matriz tridiagonal que es dominante diagonalmente, la matriz A es una modificación al sistema de ecuaciones que se obtiene al resolver la Ecuación de Poisson, trabajaremos con este sistema en particular pues como mencionamos anteriormente la matriz A es dominante diagonalmente

$$\begin{pmatrix} 3 & 1 & 0 & \dots & 0 \\ 1 & 3 & 1 & \dots & 0 \\ & & \vdots & & \\ 0 & \dots & 1 & 3 & 1 \\ 0 & \dots & 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

Primero resolvemos el problema directamente usando MATLAB con las características de la computadora 1. Como queremos una buena aproximación usamos el valor de $n = 10000$, luego cargamos la matriz A con la siguiente serie de comandos:

```
n=10000; A1=eye(n);
A2=diag(ones(1,n-1),1); A3=diag(ones(1,n-1),-1);
A=A2+3*A1+A3;
```

En este punto tenemos un problema, no podemos cargar una matriz con estas dimensiones, así que buscamos un límite de almacenamiento y encontramos un valor de $n = 6500$, con el que si se puede cargar al menos una matriz. Sin embargo, surge otro problema: con este nuevo valor ya ocupamos toda la memoria de la computadora y aún nos falta resolver el sistema. Procedemos con otro método [1] aprovechando que A es una matriz en banda. El propósito es ignorar los elementos cero. Para resolver el sistema (4) usamos el método de Thomas, el cual sólo almacena las entradas de las diagonales. El método de Thomas [8] es el siguiente:

1. Iniciamos con $a_1 = \frac{a_1}{d_1}, \qquad r_1 = \frac{r_1}{d_1}$.
2. Para $i = 2, ..., n-1$

$$a_i = \frac{a_i}{d_i - b_i a_{i-1}}, \qquad r_i = \frac{r_i - b_i r_{i-1}}{d_i - b_i a_{i-1}}.$$

3. Para la última ecuación:

$$r_n = \frac{r_n - b_n r_{n-1}}{d_n - b_n a_{n-1}}$$

4. Sustitución hacia atrás

$$y_n = r_n$$
$$y_i = r_i - a_i y_{i+1}, \qquad i = n-1, n-2, ..., 1.$$

La siguiente función de MATLAB implementa Thomas:

```
function y = Thomas(a, d, b, r)
  n=length(d);
  a(1)=a(1)/d(1);
  r(1) = r(1)/d(1);
  for i = 2:n-1
      denom = d(i) - b(i)*a(i-1);
      if (denom == 0), error('denominador cero'), end
      a(i) = a(i)/denom;
      r(i) = (r(i)-b(i)*r(i-1))/denom;
  end
  r(n) = (r(n)-b(n)*r(n-1))/(d(n) - b(n)*a(n-1));
  y(n) = r(n);
  for i = n-1:-1:1
      y(i) = r(i) - a(i)*y(i+1);
  end
  y=y(:);
```

La ventaja de ocupar Thomas es muy notoria, pues sólo ocupa $3n$ entradas de las n^2 que contiene A.

Finalmente, el tiempo de solución del sistema de ecuaciones del problema particular de Poisson es:

```
n = 10000;
a = [ones(1, n-1) 0];
d = 3*ones(1, n);
b = [0 ones(1, n-1)];
r = [1 zeros(1, n-1)];
y = Thomas(a, d, b, r);
tic, y; toc
  Elapsed time is 0.000283 seconds.
```

Los primeros 5 elementos de la solución y_j son:

```
y(1:5)
    ans =
     0.381966011250105
    -0.145898033750315
     0.055728090000841
    -0.021286236252208
     0.008130618755783
```

Ahora observamos que sucede si los cálculos anteriores se realizan en Python con las características de la computadora 2.

En Python[2] no tenemos problemas para cargar la matriz, en parte por las capacidades de la computadora 2. Con las siguientes instrucciones podemos cargar la matriz mencionada anteriormente, además intentamos invertirla.

```
from numpy import eye, diag, ones
from numpy.linalg import inv
n=10000
A= 3*eye(n) + diag(ones(n-1), 1) + diag(ones(n-1), -1)
I= inv(A)
```

Este proceso tarda aproximadamente 15 minutos. El proceso de invertir una matriz de tales dimensiones representa una sobrecarga para la memoria y el procesador de la computadora, es por ello que no vamos a invertirla, veamos que pasa si intentamos resolver un sistema de ecuaciones lineales de este tamaño con una implementación simple del método de eliminación gaussiana.

```
from numpy import dot, array,eye,diag,ones,concatenate,zeros

def gaussElim(a,b):
    n=len(b)
    for k in range(0,n-1):
        for i in range(k+1,n):
            if a[i,k] != 0.0:
                lam = a[i,k]/a[k,k]
                a[i,k+1:n]=a[i,k+1:n]-lam*a[k,k+1:n]
                b[i]= b[i]-lam*b[k]
    for k in range(n-1,-1,-1):
        b[k]= (b[k]- dot(a[k,k+1:n],b[k+1:n]))/a[k,k]
    return b


n=10000
a = eye(n)*3 + diag(ones(n-1),1) + diag(ones(n-1),-1)
I=gaussElim(a, concatenate((ones(1),zeros(n-1))))
print I
```

lo anterior representa un problema muy sencillo, sin embargo por las dimensiones de la matriz este proceso es muy tardado, por lo que no es muy práctico usarlo.

Igual que antes usamos el método de Thomas para aprovechar la estructura de la matriz(tridiagonal), a continuación presentamos una implementación de Thomas en Python:

```
def thomas(a,d,b,r):
    n=len(d)
    a[0]=a[0]/d[0]
    r[0]=r[0]/d[0]
    for i in range(1,n-2):
        denom = d[i] - b[i]*a[i-1]
        if (denom==0):
```

```
            print "DENOMINADOR CERO"
        else:
            a[i]= a[i]/denom
            r[i]=(r[i] - b[i]*r[i-1])/denom
    r[n-1]=(r[n-1]-b[n-1]*r[n-2])/(d[n-1] - b[n-1]*a[n-2])
    #FIN DEL METODO DE THOMAS
    x=zeros(n)
    x[n-1]=r[n-1]

    for i in range(n-2,-1, -1):
        x[i]= r[i]-a[i]*x[i+1]
    return x
```

Recordemos que este método solo usa las entradas de la matriz que son distintas de cero, así logramos liberar la memoria utilizada por las entradas que son ceros. Con las siguientes instrucciones resolvemos el sistema $Ax = b$, para este caso b es la primer columna de la matriz identidad, x resulta ser la primer columna de A^{-1}. Repetimos este proceso con cada una de las columnas de la matriz identidad, y de este modo logramos calcular A^{-1} con una reducción aproximada del 50 % del tiempo respecto al comando **inv** . En el código creamos las diagonales de la matriz A; otra opción es extraer las diagonales de dicha matriz. Para definir las diagonales de la matriz A y resolver el sistema tenemos lo siguiente:

```
d=3*ones(10000)
b=concatenate((ones(9999),zeros(1)))
a=concatenate((zeros(1),ones(9999) ))
r=zeros(10000)
r[0]=1
I1 = thomas(b,d,a,r)
```

3. Problema de Interpolación

Como ha sido comentado, aprovechar la estructura de las matrices es la mejor opción cuando queremos una buena exactitud a la solución de un problema dado. Sin embargo, en general nos topamos con problemas que involucran matrices densas, en estos casos debemos recurrir a métodos de factorización o replantear el problema para poder aprovechar otros métodos de solución. Un ejemplo que involucra una matriz densa es el siguiente.

3.1. Ejemplo. Interpolación polinomial. Si sabemos los valores de una función $y_1, y_2, ..., y_n$ en los puntos de interpolación $x_1, x_2..., x_n$, podemos encontrar un polinomio de interpolación usando la siguiente forma de potencias.

$$P(x) = c_1 x^{n-1} + c_2 x^{n-2} + ... + c_{n-1} x + c_n, \tag{5}$$

cuyos coeficientes se pueden obtener resolviendo un sistema de la forma $Vc = f$ donde las entradas de la matriz V están dadas por $V_{k,j} = x_k^{n-j}$ y obtenemos el siguiente sistema.

$$\begin{pmatrix} x_1^{n-1} & x_1^{n-2} & \dots & x_1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \dots & x_2 & 1 \\ & & \vdots & & \\ x_n^{n-1} & x_n^{n-2} & \dots & x_n & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{6}$$

La matriz V de este sistema se define como una Matriz de Vandermonde[1], donde cada x_i de la matriz es: $v_{k,j} = x_k^{n-j}$. Es bien conocido que esta matriz está mal condicionada[2]lo cual nos hace tener errores al resolver este sistema.

Observemos un caso particular, donde usaremos como puntos de interpolación los puntos de Chebyshev, resolveremos el sistema que involucra la matriz V con un lado derecho f con el siguiente código

```
from numpy import arange, cos, pi, vander, zeros, ones,
concatenate, dot
from numpy.linalg import solve

n=1000
p=arange(0,n+1)
pTchev = cos(pi*p/n)
M=vander(pTchev)
c=solve(M, concatenate((4*ones(1), zeros(n))))
print(dot(M,c))
```

para este caso estamos usando como lado derecho el vector columna $c = [4, 0, 0, ..., 0]$ que tiene dimensión $1 \times n + 1$; la matriz de Vandemonde es una que definimos a partir del siguiente vector $[1, 0{,}99999507, 0{,}99998026, ..., -0{,}99998026, -0{,}99999507, -1.]$, y usando el comando **vander** construimos la matriz, luego usando **solve** resolvemos el sistema, finalmente calculamos el producto de Vc comparamos con nuestro lado derecho y podemos ver que el c que calculamos nos da un resultado incorrecto.

A continuación usando un método en [6] replanteamos el problema (6), para aprovechar el algoritmo de la Transformada Rápida de Fourier [3] y reducir el número de operaciones.

Para la interpolación de los datos ocupamos una serie truncada en polinomios de Chebyshev [4] ahora el problema consiste en calcular los coeficientes de la serie truncada. Los datos de entrada es un conjunto de números $f_0, ..., f_N$ que pueden ser muestras de una función $f(x)$ en los puntos de Chebyshev. Antes de seguir con la formulación del problema se presentan algunas definiciones.

3.2. Definición. Los puntos de Chebyshev están definidos por

$$x_j = cos(\pi j/N), \quad 0 \le j \le N. \tag{7}$$

3.3. Definición. Se define el j-ésimo Polinomio de Chebyshev.

$$T_j(x) = cos(j \ \ arcos(x)), \quad x \in [-1, 1]. \tag{8}$$

3.4. Definición. El polinomio de interpolación equivalente a un polinomio de Chebyshev, puede ser considerado como una serie truncada de Chebyshev, esto es

$$f(x) \approx p(x) = \sum_{j=0}^{N} a_j T_j(x). \tag{9}$$

El método para calcular dichos coeficientes se basa en tres identidades entre una variable real x, una variable angular θ, y una variable compleja $z = e^{i\theta}$ sobre el círculo unitario del plano complejo.

3.5. Observación. Para $x \in [-1, 1]$ se cumple lo siguiente:

$$x = Rez = \frac{1}{2}(z + z^{-1}) = cos\theta \tag{10}$$

donde $x \in [-1, 1]$, $\theta \in [0, 2\pi]$ y $|z| = 1$.

3.6. Definición. El j-ésimo polinomio de Chebyshev se puede escribir de acuerdo a la observación anterior como:

$$T_j(x) = Rez^j = \frac{1}{2}(z^j + z^{-j}) = cosj\theta \tag{11}$$

Se pueden hacer algunos cálculos para ver que efectivamente se trata de un polinomio en x. Recursivamente se obtiene

$$T_{j+1}(x) = 2xT_j(x) - T_{j-1}(x) \tag{12}$$

para cada $j \geq 0$.

3.7. Observación. De la recursividad anterior se deduce que T_N es un polinomio de grado N para cada $N \geq 0$, con coeficiente principal 2^{N-1} para $N \geq 1$.

3.8. Observación. Como que T_N es de grado N para cada N, cualquier polinomio de grado N se puede escribir como una serie truncada en polinomios de Chebyshev (unicidad del polinomio de interpolación)[7].

Las observaciones anteriores permiten las siguientes equivalencias que son la base del método para calcular los coeficientes de una serie truncada de Chebyshev.

- En la variable $x \in [-1, 1]$, $p(x)$ es un **polinomio algebraico** determinado por sus valores en los $N + 1$ puntos de Chebyshev $x_0, ..., x_N$.

$$f(x) \approx p(x) = \sum_{j=0}^{N} a_j T_j(x). \tag{13}$$

- En la variable $\theta \in [0, 2\pi]$, la misma función es un **polinomio trigonométrico** $P(\theta)$ determinado por sus valores en los $2N$ puntos equidistantes $\theta_0, ..., \theta_{2N+1}$ con $\theta_j = \pi j/N$, $j = 0, ..., 2N + 1$; en este caso la función toma valores iguales en θ y $2\pi - \theta$. Así podemos escribir

$$f(\cos\theta) \approx P(\theta) = \sum_{j=0}^{N} a_j cosj\theta. \tag{14}$$

La equivalencia anterior ayuda a replantear el problema del cálculo de coeficientes de Chebyshev en el eje real al problema del cálculo de coeficientes de un polinomio trigonométrico en un intervalo regular $[0, 2\pi]$. Las series truncadas $p(x)$ y $P(\theta)$ se tratan como interpolaciones de las funciones $f(x)$ y $F(\theta)$, respectivamente. Los puntos de interpolación son $x_j = \cos\theta_j = Re(z_j)$ con $0 \leq j \leq N$ y $\theta_j = \frac{j\cdot\pi}{N}$

$$\theta_j = \frac{j\pi}{N},$$

Ahora se describe el siguiente método para calcular los coeficientes de una serie truncada en polinomios de Chebyshev:

1. Sean x_j puntos de Chebyshev con $0 \leq j \leq N$ y los siguientes valores

$$f_j = f(x_j) = p(x_j) = p_j.$$

Estos valores pueden ser evaluaciones en los puntos de Chebyshev de una función $f(x)$. Estos valores se extienden a un vector $\overrightarrow{V}$ de longitud $2N$ con la siguiente condición. Es decir

$$\overrightarrow{V} = \{p_0, p_1, p_2, ..., p_N, p_{N-1}, p_{N-2}, ..., p_2, p_1\}. \tag{15}$$

Esto se hace porque el conjunto de puntos de Chebyshev toma valores iguales en los correspondientes puntos sobre el círculo unitario, la distribución de los puntos de Chebyshev. Los valores son etiquetados como sigue:

$$V_0 = p_0, V_1 = p_1, ..., V_N = p_N, V_{N+1} = p_{N-1}..., V_{2N-1} = p_1.$$

2. El problema ahora consiste en calcular los coeficientes de un polinomio trigonométrico de interpolación cuyos datos son $\overrightarrow{V}$, los nodos igualmente distribuidos (puntos de interpolación) $\theta_j = j\pi/N$ con $j = 1, 2, ..., 2N - 1$, esto se traduce a un sistema de ecuaciones lineales de la siguiente forma

$$V_j = \sum_{k=0}^{2N-1} c_k e^{ik\theta_j}, \quad j = 0, 1, ..., 2N - 1. \tag{16}$$

Esto también se puede escribir matricialmente, es decir de la forma

$$\overrightarrow{V} = F_{2N}\,\overrightarrow{c}. \tag{17}$$

que explicítamente se escribe

$$\begin{pmatrix} V_0 \\ V_1 \\ \vdots \\ V_N \\ \vdots \\ V_{2N-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & e^{i\theta_1} & \dots & e^{i(N)\theta_1} & \dots & e^{i(2N-1)\theta_1} \\ \vdots & & & & & \\ 1 & e^{i\theta_N} & \dots & e^{i(N)\theta_N} & \dots & e^{i(2N-1)\theta_N} \\ \vdots & & & & & \\ 1 & e^{i\theta_{2N-1}} & \dots & e^{i(N)\theta_{2N-1}} & \dots & e^{i(2N-1)\theta_{2N-1}} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_N \\ \vdots \\ c_{2N-1} \end{pmatrix}$$

3.9. Observación. La matriz F_{2N} es conocida como la matriz de Fourier. Es fácil ver que las columnas de la matriz F_{2N} son ortogonales y que las columas de $\frac{1}{\sqrt{2N}}F_{2N}$ y $\frac{1}{\sqrt{2N}}\overline{F}^T_{2N}$ son ortonormales[5], por lo que el producto de estas matrices produce I_{2N}, donde $\overline{F}^T_{2N}$ es la matriz transpuesta conjugada de F_{2N}.

3.10. Observación. La inversa de F_{2N} está dada por

$$F^{-1}_{2N} = \frac{1}{2N}\overline{F}^T_{2N}. \tag{18}$$

3. Ahora que se trabaja con un conjunto de puntos igualmente distribuidos θ_j con $j = 0, 1, ..., 2N-1$ en el intervalo $[0, 2\pi]$, se procede a calcular los coeficientes c_k, para la cual se reescribe el problema como sigue:

$$\overrightarrow{c} = F^{-1}_{2N}\overrightarrow{V} = \frac{1}{2N}\overline{F}^T_{2N}\overrightarrow{V}. \tag{19}$$

En apariencia los coeficientes c_k son fáciles de obtener ya que todos los valores θ_j y V_j son conocidos, la matriz F^{-1}_{2N} es fácil de obtener y su tamaño depende del número de puntos de Chebyshev x_j que sean tomados sobre el intervalo $[-1, 1]$. Si se requiere una buena aproximación del polinomio de interpolación se deben resolver sistemas de ecuaciones muy grandes que implican resolver matrices de Fourier de doble dimensión, lo que computacionalmente parece ser más costoso. Sin embargo podemos calcular los valores c_k usando el algoritmo de la Transformada Rápida de Fourier, logrando que el número de operaciones del orden de $O((2N)^2)$ (si el sistema se resuelve de manera directa) se realicen en un número de operaciones del orden de $O(2N \log_2 2N)$.

Una vez obtenidos los coeficientes c_k se manejan en el siguiente orden:

$$c_N, ..., c_1, c_0, c_{N+1}, ..., c_{2N-1}. \tag{20}$$

La razón de esto es que los valores $c_N, ..., c_0$ corresponden a los coeficientes $a_N, ..., a_0$ de la serie truncada de Chebyshev, como se muestra a continuación.

4. Para obtener los coeficientes a_j de la serie truncada es necesaria la siguiente igualdad [3]:

$$P(\theta) = \sum_{k=0}^{2N-1} e^{ik\theta}c_k = \sum_{j=0}^{N} a_j \cos j\theta. \tag{21}$$

Realizando algunos cálculos sencillos se obtiene que los valores a_j están dados por

$$a_N = \frac{1}{2N}c_N, ..., a_1 = \frac{1}{N}c_1, a_0 = \frac{1}{2N}c_0. \tag{22}$$

A continuación presentamos un ejemplo del problema de interpolación usando el método anterior codificado en Python.

3.11. Ejemplo. Calcular los coeficientes de una serie truncada en polinomios de Chebyshev para la siguiente función

$$f(x) = x^2 \cos 20x + x^5 \sin 50\pi x^2, \quad x \in [-\pi, 3\pi].$$

Para trabajar con los puntos de Chebyshev sobre el intervalo $[-\pi, 3\pi]$ usamos el siguiente cambio de variable. Sean $t_k \in [-1, 1]$ puntos de Chebysev, en general para un intervalo $[a, b]$ tenemos

$$x_k = \frac{b-a}{2}t_k + \frac{a+b}{2}. \tag{23}$$

Proponemos un valor de $N = 11207$, para alcanzar una buena exactitud.

```
N = 11207; a=-pi; b = pi; y = arange(0,N+1);
x = cos(pi*y/N) * (b-a)/2 + (a+b)/2
f = (x**2)*(cos(2*x)) + (x**5)*(sin(50*pi*(x**2)))
n = len(f)

q= arange(n-2,0,-1)
v = concatenate((f ,f[q]))

c = fft(v).real
m = len(c)
y=arange(n-2,0,-1)

print c[n-1]/m
print c[y]/N
print c[0]/m
```

Con el programa anterior podemos ver que obtenemos los coeficientes de un polonomio que es de grado mayor a 11000, en aproximadamente menos de .3 segundos.

4. Conclusiones

En la actualidad el hardware con el que cuentan las computadoras llega a ser impresionante, si lo comparamos con el que se usaba hace algunos años, por ello algunas veces pensamos que es posible resolver problemas muy complejos; sin embargo aun con las supercomputadoras existentes y los modernos lenguaje de

programación, un problema aparentemente sencillo puede llegar a exhibir las limitaciones del hardware y de los lenguajes de programación. Por ello es importante desarrollar algoritmos eficientes e inteligentes para resolver problemas, capaces de aprovechar el hardware porque, si no, aún con los avances tecnológicos podemos llevarnos desagradables sorpresas

Referencias

[1] C. B. Moler. *Numerical Computing with MATLAB*. Society for Industrial and Applied Mathematics, Philadelphia, 2008.
[2] J. Kiusalaas. *Numerical Methods in Engineering with Python*. Cambridge, 2010.
[3] G. Strang. *Computational science and engineering*. Wellesley-Cambridge Press, Massachusetts, 2007.
[4] Zachary Battles and Lloyd N. Trefethen. *An Extension of Matlab to Continuos Functions and Operators*. SIAM J. SCI Compt. Vol 25, No. 5, pp 1743-1770.
[5] G. Strang. *Linear algebra and its applications 3 Edición*. 1998.
[6] John P. Boyd. *Chebyshev and Fourier Spectral Methods: Second Revised*
[7] Van Loan, Charles F. *Introduction to scientific computing : a Matrix-Vector approach using MATLAB*, Prentice Hall, 1997.
[8] Laurene Fausett. *Applied Numerical Analysis Using Matlab*. Prentice-Hall, 1999.

FACULTAD DE MATEMÁTICAS-UNIVERSIDAD VERACRUZANA
Lomas del estadio s/n, Zona Universitaria. C.P. 91000, Xalapa Veracruz México.

fsergios@gmail.com, malboy123@gmail.com, elreyborrego@gmail.com

Capítulo 8

LA CONSTRUCCIÓN DE RECTAS TANGENTES ANTES DE LA INVENCIÓN DE LA DERIVADA

LUCÍA CERVANTES GÓMEZ
ANA LUISA GONZÁLEZ PÉREZ
GRISELDA SÁNCHEZ DENICIA
FCFM - BUAP

Resumen. Mostramos algunos métodos que se utilizaban para construir las rectas tangentes antes de la invención de la derivada, tomando como ejemplos la circunferencia, la parábola y la cicloide. La intención de este trabajo es ilustrar una de las grandes ventajas de la derivación: el hecho de que constituye un método general que permite definir y construir las rectas tangentes para una gran clase de curvas.

1. Introducción

Es un hecho conocido que áreas muy importantes de la Física están escritas en el lenguaje de las ecuaciones diferenciales y además, que actualmente muchos fenómenos de diversas áreas como la Biología, Ingeniería, Economía, etc, por sus características y los objetivos que se pretenden, siguen teniendo este lenguaje determinista como mejor opción para modelarse, sin embargo, hemos encontrado que para que los estudiantes lleguen a interesarse, se involucren y fluyan mejor en la elaboración de estos modelos deterministas, es deseable que hayan comprendido de una manera más profunda el concepto de derivada y su importancia, por lo que consideramos importante poner a su disposición información de la derivada complementaria al material que se cubre normalmente en los cursos de Cálculo Diferencial, con la intención de que amplíen su panorama.

Como estudiantes, es común que al concluir un primer curso de Cálculo nos hayamos quedado con la idea de que derivar consiste sólo en una serie de reglas que hay que memorizar y aplicar, difícilmente nos ubicamos en que somos herederos de una rica tradición matemática que nos permite resolver ahora, de manera casi rutinaria, problemas que en la antigüedad representaban enormes desafíos incluso para los mejores matemáticos.

En este trabajo trataremos de ilustrar el poder de la derivación para la obtención de las tangentes a curvas que pertenecen a la clase de funciones derivables, contrastando con la manera en que era necesario construirlas cuando el cálculo no existía como tal.

Por supuesto, podemos preguntarnos: ¿para qué trazar rectas tangentes? en realidad hay varias razones, mencionaremos sólo una superficialmente. Tomemos como ejemplo una partícula moviéndose rápido sobre una circunferencia y que se soltara

repentinamente (o desaparecieran las fuerzas que la mantenían allí), la partícula continuaría moviéndose (al menos durante un momento, mientras no ganaran otras fuerzas actuantes) sobre la recta tangente (es muy fácil diseñar varios experimentos que te permitan convencerte de esto, te invitamos a inventar algunos), en realidad, si la partícula estuviera moviéndose describiendo como trayectoria cualquier curva y se soltara en un punto de la curva en la cual puede trazarse una recta tangente, la partícula continuaría su movimiento sobre la tangente, por lo que si estamos interesados en la comprensión y descripción del movimiento, nos conviene poder trazar la recta tangente (cuando exista) en cualquier punto de una curva dada.

Primero construiremos las rectas tangentes a la circunferencia y la parábola con los métodos inventados por los griegos desde hace más de 2,000 años, sus construcciones era principalmente geométricas, usando regla (no graduada) y compás. Puedes revisar bellos ejemplos de este tipo de construcciones en el libro IV de *Los elementos* de Euclides, que se calcula fue escrito aprox. en el año 300 a.C., leyendo la traducción al español [7] publicada en 1576.

Posteriormente mostraremos otros métodos generados en el siglo XVII para construir las rectas tangentes a la cicloide.

2. Tangente a la circunferencia

Debido a que la idea que dimos de recta tangente en la introducción puede no ser muy cómoda para las pruebas que requeriremos, al principio vamos a llamar la caracterización que seguramente tu recuerdas de la prepa o bachillerato: empezamos recordando una de las propiedades importantes de la recta tangente a la circunferencia: *La recta tangente a la circunferencia en un punto dado la toca sólo en ese punto.*

Además esta propiedad caracteriza a una recta tangente de la circunferencia, esto es, si una recta toca a la circunferencia en un sólo punto, entonces es tangente en ese punto; nota que si trazáramos una recta en una hoja donde estuviera el dibujo de una circunferencia, lo más probable es que no la tocara o la cruzara en dos lugares distintos (ver fig. 1).

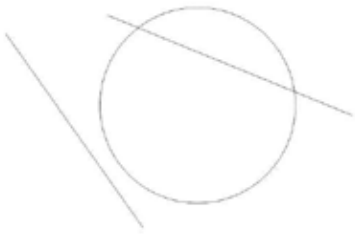


Figura 1. La mayoría de las rectas no tocan la circunferencia o la cruzan dos veces

Antes de describir el procedimiento conviene recordar la manera de trazar mediatrices[1] a un segmento dado, para trazar la mediatriz a un segmento de recta se

[1]La *mediatriz* a un segmento es la recta perpendicular que pasa por el punto medio del segmento

trazan dos circunferencias con centro en los extremos e igual radio (toma un radio que sea mayor que la mitad de la longitud del segmento, puedes ver animaciones de las construcciones en [6]), la recta que se obtiene uniendo los dos puntos de intersección de las circunferencias es la mediatriz (te invitamos a practicar esta construcción y que pruebes que en efecto obtienes una mediatriz, ver fig. 2).

FIGURA 2. Construcción de la mediatriz de un segmento dado

Construcción de la recta tangente a una circunferencia en un punto dado P usando la regla y el compás:

Primero encuentra el centro de la circunferencia, este paso es necesario ya que no siempre está claro cual es el centro, por ejemplo, cuando tienes la circunferencia que obtuviste dibujando el contorno de una moneda, para encontrar el centro escoge tres puntos sobre la circunferencia que te permitirán obtener dos cuerdas juntas; a continuación traza las mediatrices a ambas cuerdas y donde se intersecan será el centro. En segundo lugar dibuja el radio que pasa por P y por último traza la perpendicular al radio por P, para esto prolonga el radio duplicando su longitud y encuentra la mediatriz del segmento. (ver fig. 3):

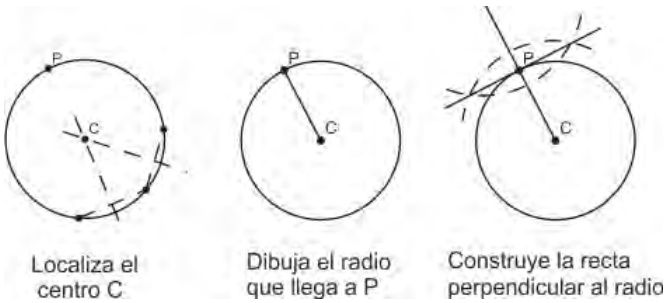


FIGURA 3. Construcción de la tangente a la circunferencia

Es fácil demostrar que la perpendicular construida con este procedimiento debe ser tangente a la circunferencia en el punto deseado P.

La demostración consiste en probar que este punto P es el único lugar donde la recta construida toca a la circunferencia.

Para esto supongamos que S es algún punto diferente del punto P ubicado sobre la recta, entonces los tres puntos C, S y P serán los vértices de un triángulo rectángulo, con el segmento CS como la hipotenusa (ver fig. 4).

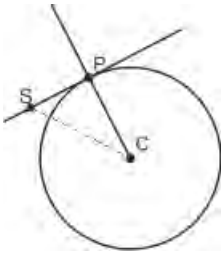


FIGURA 4. Dado S diferente de P sobre la recta, S debe está fuera del círculo

Dado que la hipotenusa de un triángulo rectángulo siempre es más larga que cualquiera de sus dos lados, vemos que el segmento CS será más largo que el segmento CP. Pero CP es el radio del círculo dado, por lo tanto, el punto S no puede estar sobre la circunferencia porque está a una distancia mayor que el radio del centro del circulo: CP. Esto nos dice que el punto P es el único punto donde la recta y la circunferencia se encuentran, y muestra, por lo tanto, que la recta es tangente a la circunferencia en P.

Notemos que este sencillo procedimiento que hemos descrito para construir una recta tangente a una circunferencia tiene la desventaja de que no funcionará con otras curvas, es fácil convencernos de ésto, ya que podemos tomar una curva y simplemente ver que las perpendiculares a sus tangentes no necesariamente se intersecan en un punto central, ésto es porque una curva no circular no tiene un centro, la construcción usa una propiedad exclusiva de las circunferencias o sus arcos, por eso no funcionará en otros casos (ver fig. 5).

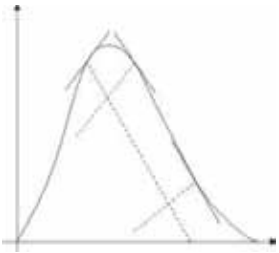


FIGURA 5. Esta curva no tiene un centro

3. Tangente a la parábola

Después de la circunferencia, la parábola es una de las curvas más sencillas, ésta consta de todos los puntos del plano que se encuentran a igual distancia de un punto fijo y de una recta fija. Al punto fijo se le conoce como el *foco* de la parábola, y a la recta fija como la *directriz* (ver fig. 6).

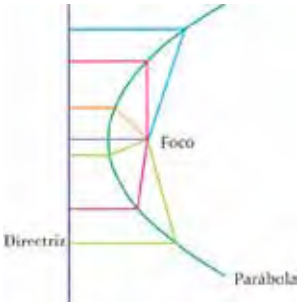


Figura 6. Parábola

Claro que hay muchas parábolas diferentes, dependiendo de la manera en que se escogen el foco y la directriz, igual que existen muchos círculos diferentes, obtenidos al variar la ubicación del centro y el tamaño del radio.

Cualquier parábola, sin embargo, automáticamente será simétrica con respecto a la recta perpendicular a la directriz y que pasa a través de su foco.

Esta recta de simetría, llamada el *eje* de la parábola, debe cruzar la parábola en exactamente un punto, conocido como el vértice.

Las diferentes partes de la parábola, su foco, directriz, eje, y el vértice, se muestran en la figura (ver fig. 7).

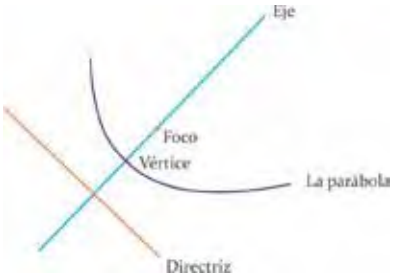


Figura 7. La parábola, su foco, directriz, eje y vértice

Podemos describir el método griego para dibujar rectas tangentes a la parábola en tres pasos:

Supongamos que P denota un punto sobre la parábola a través del cual deseamos dibujar una tangente. Primero trazamos una perpendicular de este punto P al eje

de la parábola, sea Q el punto en la base de esta perpendicular (ver fig. 8).

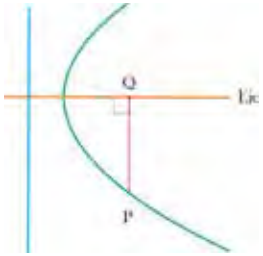


FIGURA 8. Traza la perpendicular del punto Q al eje

En segundo lugar, dibujamos un círculo a través del punto Q, teniendo el vértice V como su centro. Este círculo interseca al eje de la parábola en un nuevo punto, diferente de P, al cual denotaremos como R(ver fig. 9).

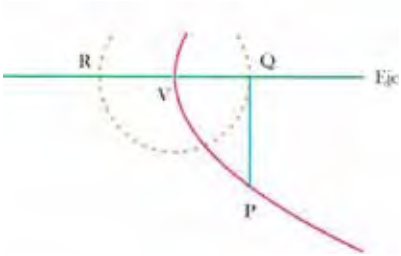


FIGURA 9. Localiza R equidistante de V

Finalmente, dibujemos una recta que pase a través de los puntos P y R. Podemos probar que esta recta PR toca la parábola solo en el punto P, y ésto debe convencernos de que la recta PR es la tangente que estamos buscando (ver fig. 10).

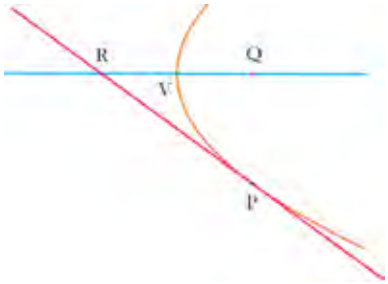


FIGURA 10. Traza la recta que une R y P

Para llevar a cabo la demostración, vamos a suponer que S denota un punto sobre la recta PR diferente de P. Probaremos que ese punto S debe estar más lejos

del foco que de la directriz, y por lo tanto que S no está sobre la parábola.

Para que esta prueba sea más clara nos apoyaremos en la siguiente notación: para denotar un segmento lo haremos con una rayita arriba, por ej.: $\overline{BP}$ = segmentoBP y la longitud del segmento acotándolo entre dos rayitas verticales, por ejemplo, $|\overline{BP}|$ = longitud del segmento BP.

Comenzaremos esta demostración trazando perpendiculares desde los puntos S y P hacia la directriz de la parábola, denotando sus bases A y B, respectivamente, es claro que el segmento $\overline{SB}$ es más largo que el segmento $\overline{SA}$, porque los tres puntos A, B y S son los vértices de un triángulo rectángulo que tiene como hipotenusa $\overline{SB}$. y sabemos que la hipotenusa de un triángulo rectángulo, siempre es más grande que cada uno de los lados (ver fig. 11).

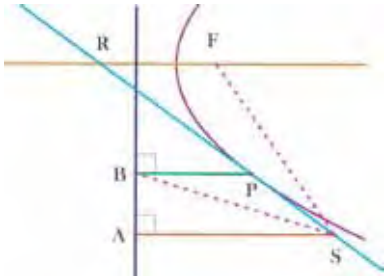


FIGURA 11. La longitud de $\overline{SB}$ es mayor que la de $\overline{SA}$

Lo que probaremos ahora es que los segmentos $\overline{SF}$ y $\overline{SB}$ tienen la misma longitud, una vez que hayamos establecido esta igualdad, obtendremos que $\overline{SF}$ debe ser mayor que $\overline{SA}$.

$$|\overline{SF}| = |\overline{SB}| > |\overline{SA}| \tag{1}$$

En consecuencia, S no podrá estar sobre la parábola ya que la ésta, por definición, consiste sólo de los puntos que son equidistantes del foco y la directriz. Esto deja a P como el único punto posible que pertenece tanto a la parábola como la recta RP.

Así, el argumento completo se reduce a mostrar que S está a igual distancia de B y F, lo cual vamos a establecer probando que el segmento de recta $\overline{RP}$ es la mediatriz del segmento $\overline{BF}$ y usando el hecho conocido de que todos los puntos sobre la mediatriz están a igual distancia de cada extremo del segmento original (en este caso B y F) (ver fig. 12).

Probaremos que el cuadrilátero [2] $BPFR$ es en realidad un rombo.

Sea C el punto donde el eje de la parábola se interseca con la directriz, sabemos que $\overline{BP}$ y $\overline{CQ}$ son paralelos y tienen la misma longitud, ya que son los lados opuestos de un rectángulo (por la manera como se determinaron Q y B sobre el eje y

[2]Un *cuadrilátero* es un polígono con cuatro lados

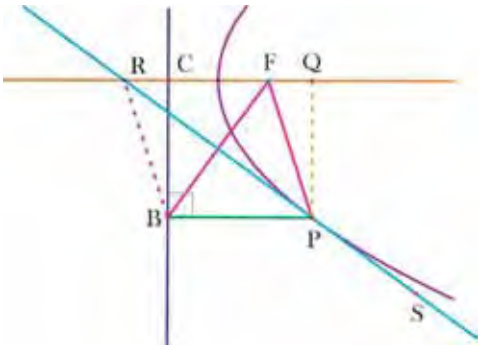


FIGURA 12. La longitud de $\overline{SB}$ es igual a la de $\overline{SF}$ ya que S está en la mediatriz de $\overline{BF}$

la directriz, las cuales son perpendiculares)

$$|\overline{BP}| = |\overline{CQ}|, \quad \overline{BP} \,||\, \overline{CQ} \tag{2}$$

Probemos además que la longitud de $\overline{CQ}$ es igual a la longitud de $\overline{RF}$, de la figura 12 es fácil notar que[3]

$$\overline{FQ} = \overline{VQ} - \overline{VF} \tag{3}$$

$$\overline{RC} = \overline{RV} - \overline{CV} \tag{4}$$

además, por la manera en que se generó R:

$$|\overline{VQ}| = |\overline{RV}| \tag{5}$$

por la definición de parábola y que V pertenece a la misma:

$$|\overline{VF}| = |\overline{CV}| \tag{6}$$

Tomando la ecuación (3) y sustituyendo en ella las longitudes $|\overline{VQ}|$ y $|\overline{VF}|$ por las longitudes equivalentes descritas en las ecuaciones (4) y (5), obtenemos:

$$\overline{FQ} = \overline{VQ} - \overline{VF} \;=\; \overline{RV} - \overline{CV} = \overline{RC}$$

con lo que hemos probado que la longitud del segmento FQ es igual a la longitud del segmento RC:

$$|\overline{FQ}| = |\overline{RC}| \tag{7}$$

Por otra parte, regresando nuevamente a la figura 12 podemos notar que

$$|\overline{RF}| = |\overline{RC}| + |\overline{CF}| \tag{8}$$

Usando la ecuación (7) para sustituir $|\overline{RC}|$ por $|\overline{FQ}|$ en la ecuación (8),

[3]Expresa la manera en que deben modificarse las ecs. cuando el punto R queda entre C y V

$$|\overline{RF}| = |\overline{RC}| + |\overline{CF}| \;=\; |\overline{FQ}| + |\overline{CF}| \;=\; |\overline{CF}| + |\overline{FQ}| = |\overline{CQ}|$$

Con lo que obtenemos:

$$|\overline{RF}| = |\overline{CQ}| \tag{9}$$

Combinando las ecs. (2) y (9) obtenemos:

$$|\overline{RF}| = |\overline{BP}|, \quad \overline{RF} \,||\, \overline{BP} \tag{10}$$

Usando el resultado que garantiza que un cuadrilátero que tiene dos lados opuestos paralelos de la misma magnitud, necesariamente es un paralelogramo[4] y la ecuación (10), obtenemos que los otros dos lados del cuadrilátero $BPFR$ deben ser también paralelos y con magnitudes iguales entre sí, esto es:

$$|\overline{RB}| = |\overline{FP}|, \quad \overline{RB} \,||\, \overline{FP} \tag{11}$$

Por otra parte, como P pertenece a la parábola, debe cumplirse que

$$|\overline{BP}| = |\overline{FP}| \tag{12}$$

Considerando la información de las ecuaciones (10), (11) y (12) obtenemos que nuestro paralelogramo $RFPB$ es en realidad un **rombo**.

Finalmente, empleando el teorema de geometría que afirma que las diagonales de un rombo son mediatrices entre sí (puede ver la animación del resultado en [6] y una demostración en el apéndice de [1]), queda concluida la demostración, ya que el segmento BF es una de las diagonales de nuestro rombo $BPFR$, y la recta que pasa sobre $\overline{RP}$ coincide con la otra diagonal; así acorde al teorema mencionado, la recta que pasa sobre $\overline{RP}$ debe ser la mediatriz de BF y, como mencionamos antes, esto implica que las longitudes de los segmentos $\overline{SB}$ y $\overline{SF}$ son iguales, lo cual a su vez implica que $|\overline{SA}|$ es menor que $|\overline{SF}|$ (ec. 1), lo cual significa que cualquier otro punto S de la recta RP no pertenece a la parábola y por lo tanto, completa nuestra demostración de que la recta RP toca a la parábola dada sólo en el punto deseado P.

4. PLANTEAMIENTO DEL PROBLEMA DE LAS TANGENTES

El procedimiento anterior de tres pasos ideado por los griegos para construir tangentes a una parábola no funcionará con otros tipos de curvas debido a que está basado en propiedades geométricas especiales que solo posee la parábola. Es necesario usar un procedimiento diferente si deseamos construir tangentes a otro tipo de curvas (una construcción para la elipse pude verse en [2]).

Al principio encontramos un método sencillo para trazar la recta tangente de la circunferencia en un punto dado, la construcción para la parábola tampoco fue complicada, sin embargo, la demostración de que en efecto era la recta tangente

[4]Un *paralelogramo* es un cuadrilátero cuyos lados opuestos son paralelos y tienen longitudes iguales

fue un poco más laboriosa. Por otra parte, en ambos casos, para demostrar que las rectas construidas efectivamente eran rectas tangentes usamos la propiedad de que tocaban a las curvas sólo en el punto dado (P).

La propiedad que caracteriza la recta tangente a la circunferencia de tocarla sólo en el punto de tangencia, puede guiarnos en algunos casos para identificar tangentes, por ejemplo en la figura 13.

Figura 13.

Pero al observar otro tipos de curvas, nos damos cuenta de que esa caracterización de recta tangente no nos permite definirla, al menos no coincide con la idea intuitiva expresada en la introducción, de que es la recta por la que continuaría el movimiento de la partícula si al encontrarse en el punto P repentinamente se soltara o desaparecieran las fuerzas que la obligaban a continuar en la curva dada (ver fig. 14).

Figura 14.

Por otra parte, una recta puede tocar a una curva en más de un punto y seguir siendo considerada una tangente (ver fig 15).

Figura 15.

Así vemos que incluso obtener la definición de recta tangente a una curva es más complicado que generalizar el concepto especial de recta tangente a una circunferencia, la cual podía definirse como la recta que la tocaba sólo en un punto.

Durante varios siglos, matemáticos importantes dedicaron sus esfuerzos a resolver el problema de encontrar las rectas tangentes a curvas dadas e inventaron varios métodos para construir tangentes a ciertas curvas especiales.

Incluso alrededor de 1640 todavía no había una definición de tangente aceptada por los principales matemáticos de la época [3]; de hecho, se manejaba la tangente concibiéndola de diferentes maneras, algunas hacían más énfasis en aspectos geométricos, otras en aspectos dinámicos y otras en la idea de límite, la cual ya estaba en el ambiente como la concebimos actualmente pero todavía no se había formalizado; presentamos sólo dos ejemplos de construcción de tangentes a la cicloide con los principales argumentos, los detalles requieren profundizar en la concepción de límite y los posponemos para una siguiente oportunidad; por otra parte, para profundizar en construcciones de la tangente a la cicloide con este tipo de métodos, te invitamos a leer [5]).

5. La cicloide

La forma geométrica usada actualmente definir la cicloide es la siguiente: *La cicloide es la curva descrita por un punto de una circunferencia que rueda sin resbalar sobre una recta*, esta definición es la que manejaba Descartes (ver fig. 16).

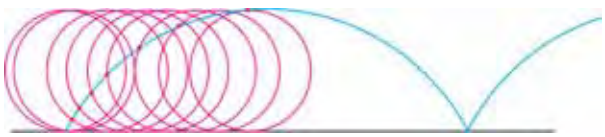


Figura 16. Cicloide de Descartes

El método de Descartes para construir la tangente a la cicloide se basaba en los 'centros de rotación instantáneos'. Consideremos un polígono, por ejemplo un triángulo, que rueda sobre una recta y fijémonos en la trayectoria que realiza un punto fijo P del triángulo, observemos que la curva descrita por el punto consiste de la unión de arcos de circunferencia cuyos centros son los puntos sobre la recta que tocan los vértices del triángulo sobre los cuales se apoya para girar; en consecuencia, la tangente a un punto de esta 'cicloide'generada por el triángulo será la perpendicular a la recta que une el punto P con el centro de la circunferencia del arco en el cual se encuentra el punto (ver fig. 17).

En la figura 17 se muestran las curvas descritas por el vértice A y por el punto P del triángulo ABC al rodar éste sobre la recta, es claro que ambas están formadas por la unión de arcos de circunferencias, notemos cómo los puntos de rotación son C, después B, posteriormente A y luego otra vez C, repitiéndose la imagen. En cualquiera de los arcos descritos por P, la recta tangente al punto P en una posición dada, será la recta perpendicular al radio que une el punto con el centro de rotación correspondiente, por tratarse de un arco de circunferencia. La idea clave

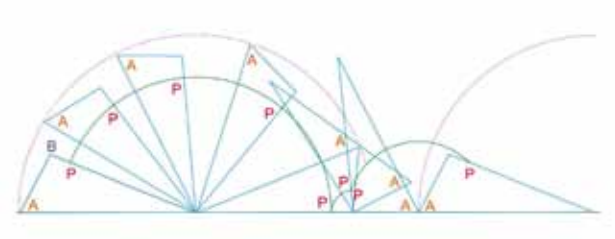


FIGURA 17. 'Cicloide'generada por un triángulo

que usaba Descartes era considerar la circunferencia como la figura límite de los polígonos regulares inscritos, pudiendo determinar así la tangente a un punto P de la cicloide como la perpendicular al centro de rotación 'límite'que correspondería sólo al punto P al que podemos visualizar como el límite del arquito generado por un polígono y en el caso límite se convirtió en el punto P (para una explicación más amplia puedes consultar [1]).

Basada en la idea anterior, la construcción para la tangente se muestra a continuación (ver fig. 18)

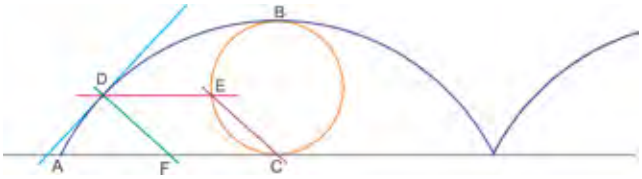


FIGURA 18. Construcción de Descartes

Sea D cualquier punto del arco de la cicloide AB, para construir la tangente primero hay que trazar la paralela a $\overline{AC}$ que pasa por el punto D; sea E el punto de intersección de esta paralela con la circunferencia, traza la recta que une a C y E y luego la paralela a ésta que pasa por D, la perpendicular a esta última recta que pasa por D es la tangente a la cicloide en D (¿cómo se podría justificar que es la tangente?).

Roberval definía la cicloide de la siguiente manera:

Consideremos que el diámetro $\overline{AB}$ del círculo se desplaza paralelamente a su posición inicial con el punto A sobre la recta $\overline{AC}$ hasta que llega a la posición $\overline{CD}$. Simultáneamente hagamos que el punto A se mueva sobre la circunferencia de tal forma que la velocidad del punto A sobre la circunferencia sea igual a la velocidad del diámetro $\overline{AB}$ a lo largo de $\overline{AC}$; en particular se tendrá que el punto A alcanzará la posición D en el momento que el diámetro alcance la posición $\overline{CD}$. Esto significa que el punto A es conducido por dos movimientos, uno el del propio punto a lo largo de la circunferencia y el otro, el de traslación de la semicircunferencia (ver fig. 19).

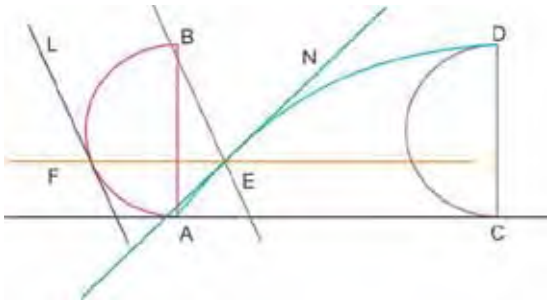


FIGURA 19. Cicloide de Roberval

El método que usaba Roberval para encontrar la tangente a la cicloide en el punto E es el siguiente: traza la paralela a AC que pasa por E; ésta interseca la semicircunferencia AB en F, luego considera la tangente L a la semicircunferencia en F y su paralela en E, esta recta forma cierto ángulo con la recta que pasa por F y E, cuya bisectriz es la tangente buscada, ya que es el resultado de dos 'movimientos iguales'; esto es, uno de ellos, el desplazamiento lateral daría como tangente a la recta que pasa por F y E, el otro, el movimiento a lo largo de la circunferencia daría como tangente a L y como las velocidades son igual magnitud, la tangente resultante es N.

6. Conclusiones

Durante siglos, la teoría de las tangentes se mantuvo como una colección de métodos no relacionados entre sí para la construcción de tangentes a curvas especiales. Vistos de manera separada, estos procedimientos son muy interesantes y nos proveen espléndidos ejercicios de razonamiento geométrico, sin embargo, vistos en conjunto, arrojaban poca luz sobre las características de la naturaleza de las tangentes, dado que cada procedimiento se aplicaba sólo a un tipo de curva.

El primer progreso importante en la teoría de la unificación de las tangentes fue posible a principios del siglo XVII gracias al desarrollo de la Geometría Analítica realizado por René Descartes. Esencialmente, lo que la Geometría Analítica proveyó fue una manera de fusionar la geometría con el álgebra, de tal manera que los problemas en un dominio pudieran traducirse en problemas correspondientes en el otro; la base de esta identificación fue la identificación de puntos con parejas ordenadas de números y las rectas y curvas con ecuaciones algebraicas apropiadas.

Casi simultáneamente con este proceso, se tuvo el desarrollo del cálculo diferencial, el cual nos permite contar ahora, en primer lugar, con una definición precisa de recta tangente para una gran variedad de curvas (todas las que son derivables) y, además, transformar una colección de métodos para encontrar tangentes en uno sólo: la derivación.

Referencias

[1] Cervantes Gómez, L., Contreras Carreto, A., Sánchez Denicia, G. Comprendiendo mejor la derivada. Conceptos clave y aplicaciones. Primera edición en Cd-Rom, Puebla, México, 2011. Isbn 978-607-00-4183-9

[2] Cruse, Allan B. y Granberg, Millianne. *Lectures on Freshman Calculus*, Sexta edición, Addison-Wesley Publishing Company, Inc., 1971.

[3] Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, New York, Oxford University Press, 1972.

[4] López Yáñez, Alejandro. *Cálculo*, Programa de actualización y formación de profesores, Módulo III. Colegio de Bachilleres. México, 1982.

[5] Whitman, E. A., *Some Historical notes on the Cycloid*, American Mathematical Monthly, Vol. 50 (1943), págs. 309-315

[6] www.educacionplastica.net/trazbas.htm

[7] Zamorano, Rodrigo. *Los Seis Libros Primeros de la Geometria de Euclides*, Casa de Alonso de la Barrera; Sevilla, España, 1576.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

lcervant@fcfm.buap.mx

Reconocimientos: Las autoras agradecen al Dr. Agustín Contreras Carreto sus sugerencias y apoyo que contribuyeron a la realización de este trabajo. G. Sánchez y A. González agradecen las becas otorgadas en el 2010 por la Vicerrectoría de Investigación y Estudios Avanzados de la BUAP a través del proyecto: *Aprovechamiento de la Modelación Matemática para la promoción del aprendizaje significativo de las matemáticas y de su vinculación con la investigación*, las cuales les permitieron participar en este trabajo.

Capítulo 9

COTAS DE ERROR PARA LA TERCERA DERIVADA DE SPLINES CÚBICOS EMPLEANDO CÁLCULO DIFERENCIAL

LUCÍA CERVANTES GÓMEZ[1]
VALENTÍN JORNET PLÁ[2]
JOSÉ JACOBO OLIVEROS OLIVEROS[1]
[1]FCFM - BUAP
[2] DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN DE OPERACIONES - UNIVERSIDAD DE ALICANTE, ESPAÑA

Resumen. Se presenta una demostración de la convergencia de las terceras derivadas de los splines cúbicos interpolantes, para funciones cuyas terceras derivadas son absolutamente continuas en problemas de frontera fija, de esta manera puede obtenerse buen grado de generalidad al mismo tiempo que las demostraciones pueden realizarse usando cálculo diferencial e integral.

1. Introducción

Dada su importancia, los libros modernos de Análisis Numérico a nivel licenciatura incluyen el tema de la interpolación mediante splines cúbicos y algunos incluso enuncian algunos teoremas para cotas de error [2],[4], pero no incluyen las demostraciones debido a que son complicadas, por otra parte, en libros avanzados para nivel de posgrado se pueden encontrar enunciados más generales de este tipo de teoremas cuyas demostraciones requieren conocimientos avanzados, dejando una brecha muy grande entre lo que se encuentra a nivel licenciatura y después a nivel posgrado .

El propósito de este trabajo es reducir la brecha mencionada y contribuir en la profundización de la comprensión del cálculo diferencial, para ello enunciaremos dos cotas del error para la interpolación con splines cúbicos con condiciones de frontera libre, basados en el artículo de De Boor [1] cuyas demostraciones utilizan principalmente resultados de cálculo diferencial que permiten la comprensión de los teoremas a nivel licenciatura y muestran además aplicaciones interesantes de los resultados utilizados.

Una función spline está formada por varios polinomios, cada uno definido sobre un subintervalo, que se unen entre sí obedeciendo ciertas condiciones de continuidad y/o derivabilidad. La definición formal para un spline cúbico es la siguiente:

Dada una función f definida en $[a, b]$ y un conjunto de puntos de un intervalo llamados **nodos** $a = x_o < x_1 < \ldots < x_n = b$, una **función spline cúbica interpolante** s para f es una función que satisface las siguientes condiciones:

a. $s(x)$ es un polinomio cúbico, denotado $s_i(x)$, en el subintervalo $[x_i, x_{i+1}]$ para cada $i = 0, 1, \ldots, n-1$;

b. $s(x_i) = f(x_i)$ para cada $i = 0, 1, \ldots, n$;
c. $s_{i+1}(x_{i+1}) = s_i(x_{i+1})$ para cada $i = 0, 1, \ldots, n-2$;
d. $s'_{i+1}(x_{i+1}) = s'_i(x_{i+1})$ para cada $i = 0, 1, \ldots, n-2$;
e. $s''_{i+1}(x_{i+1}) = s''_i(x_{i+1})$ para cada $i = 0, 1, \ldots, n-2$;
f. Se satisface uno de los dos conjuntos de condiciones de frontera:
$i)s''(x_0) = s''(x_n) = 0$ **Condición de frontera libre o natural**;
$ii)s'(x_0) = f'(x_0)$ y $s'(x_n) = f'(x_n)$ **Condición de frontera fija**.

Aunque los splines cúbicos pueden definirse con otras condiciones de frontera, generalmente las más empleadas son las enunciadas en el inciso (f). En el caso en que se cumplen las condiciones de frontera libre, el spline se llama **natural** y su gráfica semeja la forma que una cuerda flexible tomaría cuando se le fuerza a pasar por los puntos $(x_0, f(x_0)), (x_1, f(x_1)), \ldots, (x_n, f(x_n))$.

En general, las condiciones de frontera fijas conducen a mejores aproximaciones ya que incluyen más información sobre la función, aunque para que se cumplan este tipo de condiciones, es necesario tener el valor de la derivada en sus puntos finales o una buena aproximación a esos valores, lo cual no siempre es posible, por esta razón es que algunas veces es necesario usar las condiciones de frontera libre.

2. Enunciado de las funciones involucradas

Sin pérdida de generalidad esencial podemos suponer que el intervalo en el que deseamos interpolar es $[0, 1]$, vamos a considerar además que $f'''(x)$ es absolutamente continua en $[0, 1]$. Para cualquier partición $\pi : 0 = x_0 < x_1 < \cdots < x_n = 1$ de $[0, 1]$, sea la función de interpolación tipo spline cúbico (para π) con la condiciones de frontera fijas: $s(x) \in C^2$ y

$$f(x_i) = s(x_i), \quad i = 0, \cdots, n; \quad f'(0) = s'(0) \quad f'(1) = s'. \tag{1}$$

Definimos las funciones cardinales $C_i(x)$ para una interpolación spline sobre π como las funciones spline que satisfacen

$$C_i(x_j) = \delta_{ij}, \quad C'_i(0) = C'_i(1) = 0 \quad i = 1, \cdots, n-1. \tag{2}$$

Por definición, el error en la interpolación spline es

$$e(x) = f(x) - s(x). \tag{3}$$

Si $p_f(x)$ denota el polinomio cúbico el cual satisface

$$p_f(x_k) = f(x_k) \text{ y } p'_f(x_k) = f'(x_k),$$

para $k = 0, n$, entonces el error en la interpolación spline de $f(x)$ es la misma que en la interpolación spline de $g(x) = f(x) - p_f(x)$, la cual satisface $dg'''(x) = df'''(x)$ y $g(0) = g'(0) = g(1) = g'(1) = 0$; por esta razón podemos asumir, sin pérdida de generalidad,

$$f(0) = f'(0) = f(1) = f'(1) = 0. \tag{4}$$

Para tales funciones, tenemos

$$f(x) = \int_0^1 G(x, y) df'''(y), \tag{5}$$

donde $G(x, y)$ es la función de Green para el problema de valores extremos definido por $f^{(iv)}(x) = h(x)$ y (4). Explícitamente, $G(x, y)$ está dada por

$$G(x, y) = ((x - y)_+^3)/3! - P(x, y), \tag{6}$$

donde, para y fijo, $P(x, y) = x^2(1-y)^2(x+2xy-3y)/6$ es el polinomio cúbico en x tal que $G(0, y) = G_x(0, y) = G(1, y) = G_x(1, y) = 0$. La función $(x)_+^k$ está definida por

$$(x)_+^k = \begin{cases} x^k, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Por esta razón, considerado como una función sólo de x o y, $G(x, y)$ es una función tipo spline con exactamente un nodo en $x = y$. De la misma manera, para funciones que satisfagan (4) tenemos

$$s(x) = \sum_{i=1}^{n-1} f(x_i) C_i(x), \tag{7}$$

donde, por (3),(5) y el párrafo precedente,

$$e(x) = \int_0^1 \left[G(x, y) - \sum_{i=1}^{n-1} C_i(x) G(x_i, y) \right] df'''(y). \tag{8}$$

Usando (8) y las propiedades especiales de las funciones cardinales, acotaremos la r-ésima derivada $e^{(r)}(x)$ de órdenes $r = 0, 1, 2, 3$.

3. Propiedades de las funciones cardinales

Por conveniencia, definiremos la cota M_π de la razón de la red por

$$M_\pi = |\pi|/min\Delta x_i, \;\; |\pi| = max\Delta x_i, \;\; \Delta x_i = x_{i+1} - x_i, \tag{9}$$

y escribiremos $||f|| = max|f(x)|$ en $[0,1]$. El principal resultado de esta sección será que cada función cardinal $C_i(x)$ decae exponencialmente lejos de x_i, y que $|C_i(x)|$ está acotado en $[x_{i-1}, x_{i+1}]$ por una constante K' dependiendo sólo de M_π. La prueba usará algunas propiedades cualitativas de los signos de $C_i(x)$ y sus derivadas, las cuales serán establecidas en una serie de lemas.

3.1. Lema. Si $p(x)$ es un polinomio cúbico el cual se anula en 0 y $h \neq 0$, entonces

$$p'(h) = -2p'(0) - h((p''(0))/2), \text{ y } (p''(h))/2 = -3/hp'(0) - 2((p''(0))/2). \tag{10}$$

En verdad, $p(x) \equiv p'(0)x + ((p''(0))/2)x^2 - h^{-2}[p'(0) + ((p''(0))/2)h]x^3$, de la cual se sigue (10).

3.2. Corolario. Para $i \neq j+1, j, C_i(x)$ satisface

$$C_i'(x_{j+1}) = -2C_i'(x_j) - \Delta x_j((C_i''(x_j))/2), \tag{11}$$
$$(C_i''(x_{j+1}))/2 = -(3/(\Delta x_j))C_i'(x_j) - 2((C_i''(x_j))/2).$$

El significado de las ecuaciones (11) es claro: son relaciones recursivas sobre los vectores $\{C_i'(x_j), (C_i''(x_j))/2\}$, cuyos coeficientes constituyen la matriz (con todos sus coeficientes negativos)

$$\begin{bmatrix} -2 & -\Delta x_j \\ -3/\Delta x_j & -2 \end{bmatrix}$$

3.3. Corolario. Para $i = 1, \cdots, n-1$, $C_i(x)$ satisface

$$C_i'(x_j)C_i''(x_j) \geq 0, \quad para \quad j < i, \tag{12a}$$

$$C_i'(x_j)C_i''(x_j) \leq 0, \quad para \quad j > i. \tag{12b}$$

La prueba para $j = 0, \cdots, i-1$ es por inducción sobre j. Para $j = 0$, se sigue de (2). Debido a que los coeficientes en (11) son todos negativos y la condición (12a) establece que $C_i'(x_j)$ y $C_i''(x_j)$ tienen el mismo signo, se obtiene que $C_i'(x_{j+1})$ y

$C_i'''(x_{j+1})$ tienen el mismo signo, es decir, el contrario de $C_i'(x_j)$ y $C_i'''(x_j)$. La prueba para $j > i$ se obtiene cambiando x por $-x$, lo cual cambia el signo de $C_i'(x)C_i'''(x)$.

3.4. Corolario. Para $i = 1, \cdots, n-1$, $C_i(x)$ satisface

$$|C_i'(x_j)| < 1/2|C_i'(x_{j+1})|, \quad j < i-1, \quad |C_i'(x_{j+1})| < 1/2|C_i'(x_j)|, \quad j > i. \tag{13}$$

La primera desigualdad se sigue de (12a) y (11), con la observación de que $C_i'''(x_0) \neq 0$ (de otra forma, por (11), $C_i(x) \equiv 0$), por esta razón $C_i'''(x_j) \neq 0$, $j < i$. La segunda desigualdad se sigue entonces por simetría con respecto a x_i.

El decaimiento exponencial de cada $|C_i(x)|$ lejos de x_i, se sigue del Corolario 2.4 a menos que Δx_j incremente exponencialmente lejos de x_i como función de $|j-i|$, en una tasa comparable al decremento exponencial de $|C_i'(x_j)|$.

3.5. Lema. Sea $S(x)$ cualquier función spline con nodos en x_i, la cual satisface

$$S_{i-1} = S_{i+1} = 0, \qquad S_i = h > 0, \qquad S'_{i-1}S''_{i-1} \geq 0, \qquad S'_{i+1}S''_{i+1} \leq 0, \tag{14}$$

donde $S_{i-1} = S(x_{i-1})$, $S''_{i+1} = S''(x_{i+1})$, etc.
Entonces $S''_i < 0$, $S''_{i-1} \geq 0$, $S'_{i+1} \leq 0$, y $S(x) \geq 0$ sobre $[x_{i-1}, x_{i+1}]$.

Prueba. Por cálculos directos:

$$S'_i = 3h/(\Delta x_{i-1}) - 2S'_{i-1} - 1/2S''_{i-1}\Delta x_{i-1}, \tag{15a}$$

$$S'_i = (-3h)/(\Delta x_i) - 2S'_{i+1} + 1/2S''_{i+1}\Delta x_i, \tag{15b}$$

$$1/2\Delta x_{i-1}S''_i = 3h/(\Delta x_{i-1}) - 3S'_{i-1} - S''_{i-1}\Delta x_{i-1} \tag{15c}$$

$$1/2\Delta x_i S''_i = 3h/(\Delta x_i) + 3S'_{i+1} - S''_{i+1}\Delta x_i, \tag{15d}$$

y así

$$S'_i + 1/2S''_i\Delta x_i = S'_{i+1} - 1/2S''_{i+1}\Delta x_i. \tag{16}$$

Ahora suponga que $S'_{i-1} < 0$, entonces $S''_{i-1} \leq 0$ a que por (12a) deben coincidir en signo; pero por (15a),(15c), $S'_i > 0$ y $S''_i > 0$. Si ahora $S'_{i+1} > 0$, entonces $S''_{i+1} \leq 0$, así por (15a), $S'_i < 0$, lo cual es una contradicción. De la misma manera, si $S'_{i+1} \leq 0$, entonces $S''_{i+1} \geq 0$, así $S'_i + 1/2\Delta x_i S''_i \leq 0$ por (16), lo cual es otra vez una contradicción. Por lo tanto $S'_{i-1} \geq 0$. Por simetría con respecto a x_i se sigue que $S'_{i+1} \leq 0$. Por esta razón S''_{i-1} y S''_{i+1} son no negativos. Como la segunda

diferencia dividida $S(x_{i-1}, x_i, x_{i+1})$ es negativa, se obtiene $S_i'' < 0$.

A continuación suponga que para algún $x \in [x_{i-1}, x_i]$, $S(x) < 0$. Si $S_{i-1}'' = 0$, entonces $S''(x) < 0$ en (x_{i-1}, x_i), pero $S(x_{i-1}, x_i, x_{i+1}) > 0$, lo que es una contradicción. Si, por otro lado, $S_{i-1}'' > 0$, entonces, como $S_{i-1}' \geq 0$, existe $y \in (x_{i-1}, x_i)$ tal que $S(t) > 0$ para $t \in (x_{i-1}, y)$. Pero entonces $S(x_{i-1}, y, x) < 0$, $S(y, x, x_i) > 0$, lo cual implica que la función lineal $S''(x)$ tiene dos ceros distintos en (x_{i-1}, x_i) sin ser idénticamente cero, lo que es una contradicción. Por lo tanto, $S(x) > 0$, $x \in (x_{i-1}, x_i)$. Por simetría con respecto a x_i, se obtiene que $S(x) > 0$ idénticamente en (x_i, x_{i+1}).

3.6. Lema. Sea $T(x)$ un spline con un nodo en x_i, tal que

$$T_{i-1} = T_i = T_{i+1} = 0, \quad T_{i-1}'' \leq 0, \quad T_{i+1}'' \geq 0. \tag{17}$$

Entonces $T(x) \geq 0$ sobre $[x_{i-1}, x_i]$.

Prueba. Con $h = 0$, como $T_i = 0$, (15a)-(15d) da

$$\Delta x_i T_{i-1}' + 2(\Delta x_{i-1} + \Delta x_i) T_i' + \Delta x_i T_{i+1}' = 0. \tag{18}$$

Si $T_{i-1}' < 0$, entonces se obtiene, como en el Corolario 3.3, que $T_i' > 0$, $T_i'' \geq 0$, pero $T_{i+1}' < 0$, que es una contradicción. Por lo tanto $T_{i-1}' \geq 0$. Por esta razón, si ahora $T_i' = 0$, entonces por (18), $T_{i-1}' = T_{i+1}' = 0$, y así $T(x) = 0$, lo cual completa la prueba para este caso. De otra forma, por ((18)), $T_i' < 0$, y ya que $T_i = 0$, existe un $y \in (x_{i-1}, x_i)$ tal que $T(x) > 0$, $x \in (y, x_i)$. Pero entonces la suposición de que $T(x) < 0$ para algún $x \in (x_{i-1}, y)$ implicaría que $T(x_{i-1}, x, y) > 0$, $T(x, y, x_i) < 0$, por lo que $T_{i-1}'' \leq 0$, la función lineal $T''(x)$ tendría dos ceros distintos en $[x_{i-1}, x_i]$ sin ser idénticamente ceros, lo cual es imposible.

3.7. Corolario. Sea $M = M_\pi$. Para $i = 1, \cdots, n-1$:

$$0 \leq C_i(x) \leq L \quad \text{sobre} \quad [x_{i-1}, x_{i+1}], \quad \text{donde} \quad L = 3\frac{M(M+1)^2}{3+4M}, \tag{19}$$

$$\|C_i'(x_{i-1})| \leq \frac{L}{\Delta x_{i-1}}, \quad |C_i'(x_{i+1})| \leq \frac{L}{\Delta x_i}. \tag{20}$$

Por el Corolario 3.3, $C_i(x)$ satisface la hipótesis sobre $S(x)$ en el Lema 3.5, por esta razón la primera desigualdad en (19) sigue de ese lema. Para probar la segunda desigualdad para $x \in [x_{i-1}, x_i]$, sea $U(x)$ un spline con un nodo en x_i tal

que $U_{i-1} = U_{i+1} = U''_{i-1} = U'_{i+1} = 0$, $U_i = 1$. Entonces $T(x) = U(x) - C_i(x)$ satisface la hipótesis del Lema 3.6, ya que por el Lema 3.5, como se aplicó a $C_i(x)$, $C''_{i-1} \geq 0$, $C'_{i+1} < 0$. Por esta razón $0 \leq C_i(x) \leq U(x)$ sobre $[x_{i-1}, x_i]$. Ya que $U_{i-1} = 0$, uno obtiene

$$U(x) \leq \Delta x_{i-1} max_{[x_{i-1},x_i]} U'(y). \tag{21}$$

Aplicando el Lema 3.5 a $U(x)$ nos da $U''_i < 0$. Pero $U''_{i-1} = 0$, por esta razón $U''(x) < 0$ en (x_{i-1}, x_i), y así

$$max_{[x_{i-1},x_i]} U'(y) = U''_{i-1} = \frac{3}{\Delta x_{i-1}\Delta x_i} \frac{(\Delta x_{i-1} + \Delta x_i)^2}{3\Delta x_i + 4\Delta x_{i-1}} \leq \frac{1}{\Delta x_{i-1}} 3 \frac{(M+1)^2}{(3/M) + 4}, \tag{22}$$

y (19) sigue ahora para $x \in [xi - 1, x_i]$. La primera desigualdad de (20) es una consecuencia inmediata. Las afirmaciones restantes se siguen de la simetría alrededor de x_i.

3.8. Corolario. Para $i = 1, \cdots, n-1$,

$$|C_i(x)| \leq |C'_i(x_j)|\Delta x_j \text{ sobre } [x_j, x_{j+1}], \quad j > i, \tag{23a}$$

$$|C_i(x)| \leq |C'_i(x_j)|\Delta x_{j-1} \text{ sobre } [x_{j-1}, x_j], \quad j < i - 1. \tag{23b}$$

Sea $j > i$, y suponemos sin pérdida de generalidad que $C''_i(x_j) < 0$. Entonces por el Corolario 3.3, $C'_i(x_j) \geq 0$, $C'_i(x_{j+1}) \leq 0$, y la prueba del Lema 3.6 muestra que $C_i(x) \geq 0$ sobre $[x_j, x_{j+1}]$. Además $C'_i(x_j) \geq C'_i(x)$, $x \in [x_j, x_{j+1}]$; por esta razón se obtiene (23a). Por simetría alrededor de x_i (23b) es consecuencia de (23a).

4. Primeras cotas y desigualdades fundamentales

Podemos probar ahora el primer resultado principal

4.1. Teorema. Existe una constante K que depende sólo de M_π $(K = K(M_\pi))$, tal que

$$\int_0^1 |C_i(x)|dx \leq K|\pi|. \tag{24}$$

Prueba. Para $j < i$, por (23b) y (13),

$$|C_i(x)| \leq |C_i'(x_j)| \leq \Delta x_{j-1} \leq 2^{j-i+1}|C_i'(x_{i-1})|\Delta x_(j-1), \tag{25}$$

Para $x \in [x_{j-1}, x_j]$. Por esta razón, por (22),

$$|C_i(x)| \leq 2^{j-i+1}L\Delta x_{j-1}/\Delta x_{i-1} \leq 2^{j-i+1}LM_\pi, \tag{26}$$

donde $L = L(M_\pi)$ está dado por (19). Consecuentemente

$$\int_0^{x_i} |C_i(x)|dx = \sum_{j=1}^{i} \int_{x_{j-1}}^{x_j} |C_i(x)|dx \leq \sum_{j=1}^{i-1} 2^{j-i+1}LM_\pi\Delta x_{j-1} \leq (2M_\pi + 1)L|\pi|, \tag{27}$$

ya que $\sum_0^\infty 2^{-k} = 2$ y $\Delta x_{j-1} \leq |\pi|$. Combinando la desigualdad precedente con una desigualdad como la de $j > i$, obtenemos (27) con $K = (4M_\pi + 2)L$. Como L está dada por (19), tenemos que $K \leq 3M_\pi(M_\pi + 1)^2$.
Alternativamente, podemos acotar la integral en (27) en términos del radio máximo de las longitudes de redes sucesivas. En realidad,

$$\sum_{j=1}^{i-1} 2^{j-(i-1)}\Delta x_{j-1} \leq \Delta x_{i-1} \sum_{j=1}^{i-1} 2^{j-(i-1)} \left(\frac{\Delta x_{j-1}}{\Delta x_{i-1}}\right). \tag{28}$$

Ahora elegimos R_π tal que $R_\pi^{-1} \leq \left(\frac{\Delta x_i}{\Delta x_{i-1}}\right) \leq R_\pi$, $i = 1, \cdots, n$. Si para la partición dada π, $R_\pi \leq 2\rho < 2$, entonces

$$\sum_{j=1}^{i-1} 2^{j-(i-1)} \left(\frac{\Delta x_{j-1}}{\Delta x_{i-1}}\right) \leq \sum_{j=1}^{i-1} 2^{j-(i-1)} R_\pi^{(i-1)-j} \leq \frac{1}{1-\rho}. \tag{29}$$

Además

$$\int_0^{x_i} |C_i(x)|dx \leq |\pi| \left(\frac{2}{2 - R_\pi} + 1\right) L, \tag{30}$$

donde L está dada por (19) con $M = R_\pi$.Esto prueba el

4.2. Corolario. Si $R_\pi < 2$ entonces existe una constante K que depende solo de R_π tal que (27) es verdadera.
Ahora recuerde que el error de interpolación está dado por (8) como

$$e(x) = \int_0^1 \left[G(x,y) - \sum_{i=1}^{n-1} C_i(x) G(x_i, y) \right] df'''(y). \tag{31}$$

4.3. Lema. La siguiente identidad es válida:

$$\sum_{i=1}^{n-1} C_i(x) G(x_i, y) \equiv \sum_{i=1}^{n-1} G(x, x_i) C_i(y). \tag{32}$$

Prueba. Por (3),(6),(7), obtenemos

$$\sum_{i=1}^{n-1} C_i(x) G(x_i, y) = \sum_{j=1}^{n-1} \left(\sum_{i=1}^{n-1} C_i(x) G(x_i, x_j) \right) C_j(y) = \sum_{j=1}^{n-1} G(x, x_j) C_j(y). \tag{33}$$

4.4. Corolario. La tercera derivada del error existe y satisface

$$e'''(x) = \int_0^1 G_3(x,y) df'''(y) \quad \text{donde} \tag{34a}$$

$$G_3(x,y) = \frac{\partial^3}{\partial x^3} G(x,y) - \sum_{i=1}^{n-1} \left[\frac{\partial^3}{\partial x^3} G(x, x_i) \right] C_i(y). \tag{34b}$$

La diferenciación bajo la integral está justificada ya que $G_3(x,y)$ es continua por pedazos (vea por ej. el lema 2 que garantiza la diferenciación bajo la integral en [3]). Note que aquí y en lo siguiente usamos la normalización

$$(x)_+^0 = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

4.5. Corolario. Para $x \in (0,1)$ fijo, existe $j \in [1, n-2]$ tal que

$$G_3(x,y) = g(y) - g(x_j) C_j(y) - g(x_{j+1}) C_{j+1}(y), \tag{35}$$

donde $||g|| \geq 1, g(y) = 0$ para $y \notin (x_{j-1}, x_{j+2})$.

Prueba. Se sigue de (6)que

$$\frac{\partial^3}{\partial x^3}G(x,y) = (x-y)_+^0 - (1-y)^2(1+2y). \tag{36}$$

Sea $x \in [x_{k-1}, x_k)$ y escogemos j tal que $0 \leq j-1 < k \leq j+2 \leq n$. Sea $h(y) = L(y; x_{j-1}, x_j, x_{j+1}, x_{j+2})$ la tercera diferencia dividida en z de $L(y;z) = (z-y)_+^0$ sobre $x_{j-1}, x_j, x_{j+1}, x_{j+2}$. Entonces $h(y)$ es una función spline la cual es igual a $(x-y)_+^0$ para $y \notin (x_{j-1}, x_{j+2})$ y permanece entre 0 y 1 dentro de $[x_{j-1}, x_{j+2}]$. Por esta razón con $g(y) = (x-y)_+^0 - h(y)$ se obtiene

$$G_3(x,y) = g(y) - \sum_{i=1}^{n-1} g(x_i)C_i(y), \tag{37}$$

de donde se puede concluir el Corolario 3.8.

5. Cotas de error

Dos cotas para $e'''(x)$,la tercera derivada del error (*i.e.* el error en la tercera derivada) se pueden derivar. como la función spline tiene una tercera derivada continua por pedazos, el error en la tercera derivada será, en general, acotado lejos del cero a menos que $f(x) \in C^2$. Antes de probar esta afirmación, primero estableceremos un resultado más fuerte válido para $f(x) \in C^4$.

5.1. Teorema. Sea $f(x) \in C^4$, y sea $e(x)$ el error (3) que ocurre cuando $f(x)$ es interpolada por una función spline en una partición dada

$$\pi : 0 = x_0 < x_1 < \cdots < x_n = 1.$$

Entonces existe una constante $K_1(M_\pi)$ dependiente sólo de M_π tal que

$$||e'''|| \leq ||f^{iv}||K_1(M_\pi)\pi|. \tag{38}$$

Prueba. Sea $x \in (0,1)$ fijo, por el Corolario 4.5 se tiene

$$\int_0^1 |G_3(x,y)|dy \leq \int_{x_{j-1}}^{x_{j+2}} dy + 2max_{j,j+1}\int_0^1 |C_i(y)|dy, \tag{39}$$

para alguna $j \in [1, n-2]$. Sea K la constante del Teorema 4.1, con $K_1(M_\pi) = 3+2K$, se obtiene:

$$\int_0^1 |G_3(x,y)|dy \leq K_1(M_\pi)|\pi|, \tag{40}$$

en consecuencia, con (38),

(41)
$$|e'''(x)| = \left| \int_0^1 G_3(x,y) df'''(y) \right| \leq ||f^{iv}|| \int_0^1 |G_3(x,y)| dy \leq ||f^{iv}|| K_1(M_\pi)|\pi|, \ x \in (0,1).$$

Se obtiene el Teorema 5.1 ya que $e'''(0) = e'''(0+),\ e'''(1) = e'''(1-)$. Para encontrar las cotas correspondientes a la r-ésima derivada $e^{(r)}$ cuando $r < 3$, observe que, por el teorema de Rolle, existe ξ_i^r con

$$0 = \xi_0^r \leq \xi_1^r < \xi_2^r < \cdots < \xi_{n_r-1}^r \leq \xi_{n_r}^r = 1 \tag{42}$$

tales que $e^{(r)}(\xi_i^r) = 0, i = 1, \cdots, n_r - 1$, y $max_i \Delta\xi_i^r \leq (r+1)|\pi|$. Por esta razón

$$|e^{(r)}(x)| \leq \int_{\xi_i^r}^{\xi_{i+1}^r} |e^{(r+1)}(y)| dy \leq \Delta\xi_i^r ||e^{(r+1)}||, \qquad x \in [\xi_i^r, \xi_{i+1}^r], \tag{43}$$

$$||e^{(r)}|| \leq (r+1)|\pi| ||e^{(r+1)}||, \qquad r < 3. \tag{44}$$

5.2. Corolario. Bajo la hipótesis del Teorema 5.1, existen constantes $K_r(M_\pi)$ que dependen sólo de M_π tales que

$$||e^{(r)}|| \leq ||f^{iv}|| K_r |\pi|^{4-r}, \qquad r = 0, 1, 2, 3. \tag{45}$$

Si π_n es una sucesión de particiones de $[0,1]$ tal que $|\pi_n| \to 0$ mientras que $M_{\pi_n} \leq M$ permanece acotado, entonces el Teorema 5.1 implica que para el error correspondiente $e_n(x)$ de la interpolación spline se tiene

$$|e_n'''(x)| \to 0, \qquad \text{uniformemente sobre} \qquad [0,1], \tag{46}$$

si $f(x) \in C^4$. Podemos ahora hacer la suposición más débil que $f'''(x)$ es continua y de variación acotada sobre $[0,1]$, lo cual se implica por la suposición hecha al comienzo de que $f'''(x)$ es absolutamente continua. La convergencia puede todavía ser probada bajo esta suposición por medio de un análisis más cuidadoso de la integral (22).

5.3. Teorema. Sea $f'''(x)$ absolutamente continua sobre $[0,1]$. Sea π_n una sucesión de particiones de $[0,1]$ tal que $|\pi_n| \to 0$ mientras que $M_{\pi_n} \leq M$ cuando $n \to \infty$. Sea $e_n(x)$ el error incurrido cuando $f(x)$ se interpola por la función spline sobre π_n. Entonces

$$|e_n'''(x)| \to 0 \quad \text{uniformemente sobre } [0,1], \quad \text{cuando } n \to \infty. \tag{47}$$

Prueba. Sea $\epsilon > 0$ dado. Como $f'''(x)$ es absolutamente continua existe $\delta > 0$ tal que para todo $I = [a,b] \subset [0,1]$, con $b - a < \delta$

$$\int_I |df'''(y)| < \epsilon. \tag{48}$$

Como $|\pi_n| \to 0$, existe N tal que para $n \geq N$, $|\pi_n| < \delta$. Sea ahora $n \geq N$, $\pi_n : 0 = x_0 < x_1 < \cdots < x_m = 1$, y $x \in (0,1)$. Por el Corolario 4.5

$$G_3(x,y) = g(y) - g(x_j)C_j(y) - g(x_{j+1})C_{j+1}(y), \tag{49}$$

donde $||g|| \leq 1$, $g(y) = 0$ para $y \notin (x_{j-1}, x_{j+2})$, para alguna $j \in [1, n-2]$. Por esta razón

$$\begin{aligned} |e'''(x)| = & \left| \int_0^1 G_3(x,y) df'''(y) \right| \\ & \leq \left| \int_{x_{j-1}}^{x_{j+n}} g(y) df'''(y) \right| + \left| \int_0^1 C_f(y) df'''(y) \right| + \left| \int_0^1 C_{j+1}(y) df'''(y) \right|, \end{aligned} \tag{50}$$

pero

$$\left| \int_0^1 C_i(y) df'''(y) \right| \leq \sum_{j=1}^{m} \left| \int_{x_{j-1}}^{x_j} C_i(y) df'''(y) \right| \leq \sum_{j=1}^{m} max_{[x_{j-1}, x_j]} |C_i(y)| \int_{x_{j-1}}^{x_j} |df'''(y)|. \tag{51}$$

Por esta razón, al elegir n, y por la prueba del teorema 4.1,

$$\left| \int_0^1 C_i(y) df'''(y) \right| \leq \epsilon . K(M), \tag{52}$$

donde $K(M)$ depende sólo de M. En consecuencia

$$|e'''(x)| \leq \epsilon(3 + 2K(M)), \tag{53}$$

y puede concluirse que el Teorema es válido ya que,

$$\begin{aligned} e'''(0) &= e'''(0+), \\ e'''(1) &= e'''(1-). \end{aligned}$$

Con lo cual se ha terminado la demostración.

6. Conclusiones

Como puede apreciarse, la construcción de las demostraciones requiere conocimientos de funciones de una o dos variables que incluyen los conceptos de decaimiento exponencial, convergencia de sucesiones, continuidad absoluta, así como algunos resultados básicos como el teorema de Rolle y condiciones que permitan el intercambio de la derivada y la integral; todos los cuales forman parte de los conocimientos obligatorios de cualquier licenciatura en matemáticas; por lo que cualquier estudiante puede, con un poco de paciencia, reproducir las demostraciones, desarrollando los detalles necesarios, logrando, de esta manera, practicar sus conocimientos de Cálculo Diferencial en situaciones fuera de los ejercicios del libro de texto y además probar unos teoremas de cotas de error desde la licenciatura.

Referencias

[1] Birkhoff, G. y De Boor, C., *Error Bounds for Spline Interpolation*, J. Math. and Mechanics, Vol. 13 (1964), pp. 827-835

[2] Burden, Richard L., Faires, Douglas J. *Numerical Analysis*, Ca, USA, Wadsworth Group. Brooks/Cole., Séptima ediición, 2001. Pp. 143 - 152.

[3] Flemin W. *Functions of Several Variables*, N.Y., USA. Springer Verlag, Segunda ediición, 1977. Pp. 237 - 238.

[4] Kincaid, D. y Cheney, W. *Análisis numérico. Las matemáticas del cálculo científico*, E.E. U.U., Addison-Wesley Iberoamericana, 1994. Pp. 325 - 333

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

lcervant@fcfm.buap.mx

Reconocimientos: Este trabajo pudo realizarse gracias al apoyo de la SEP a través de los proyectos de integración en redes académicas de PROMEP durante el 2010.

Capítulo 10

ESTABILIZACIÓN DE LA ORIENTACIÓN DE UN SATÉLITE POR MEDIO DE LEYES DE CONTROL NO LINEALES CON RETROALIMENTACIÓN DE SALIDA

RAFAEL CRUZ JOSÉ[1]
JOSÉ FERMI GUERRERO CASTELLANOS[2]
WUIYEVALDO FERMÍN GUERRERO SÁNCHEZ[1]
JOSÉ JACOBO OLIVEROS OLIVEROS[1]
[1]FACULTAD DE CIENCIAS DE LA ELECTRÓNICA - BUAP
[2]FCFM - BUAP

Resumen. El presente trabajo aborda el desarrollo de leyes de control no lineales para la estabilización de la orientación de un satélite. El objetivo principal consiste en alinear la antena de comunicación del satélite con la antena de la base terrestre. Para esto, dos leyes de control son propuestas; la primera considera que el vector de velocidad angular es disponible, así como un vector de dirección de la antena de la estación en tierra expresado en el sistema de referencia del satélite, el cual tiene como origen el centro de masa del mismo. La segunda ley de control solo considera al vector de dirección de la antena de la estación en tierra expresado en el sistema de referencia del satélite y el efecto de amortiguamiento introducido por el vector de velocidad angular es reemplazado por un sistema dinámico "virtual"construido por el vector antes mencionado y su derivada filtrada. Esto es lo que se conoce como un esquema de retroalimentación de salida con inyección dinámica. Desde un punto de vista práctico, la ventaja y originalidad del esquema propuesto es el de utilizar el menor número de sensores posibles, sin degradar el desempeño de la estabilización. Por medio de un análisis basado en el formalismo de Lyapunov se demuestra que las leyes de control estabilizan al sistema de manera asintótica y global. Simulaciones computacionales corroboran el desempeño del sistema en lazo cerrado y muestran su robustez con respecto a incertidumbre en el conocimiento de los parámetros y con respecto a ruido en las medidas.

1. Introducción

Desde hace varias décadas el desarrollo del control aplicado a la estabilización de la orientación de sistemas que pueden considerarse como cuerpos rígidos, ha tenido un creciente interés, especialmente en áreas como aeronáutica, aeroespacial, control y robótica. Esto se debe a que sistemas tales como aviones, helicópteros, misiles tácticos, naves espaciales, satélites, robots manipuladores e incluso vehículos subacuáticos se pueden modelar como cuerpos rígidos y todos ellos necesitan ser orientados para sus distintas aplicaciones. Varios métodos de control se han aplicado en la solución del problema de control de orientación, algunos de ellos son Linealización por retroalimentación [15], Control PD [14], Backstepping [10], [4], Control robusto [11], Métodos optimales inversos [7], [5], claramente esta lista no es exhaustiva.

Las leyes de control desarrolladas mediante los métodos mencionados anteriormente consideran que los estados del sistema son conocidos, es decir, la orientación

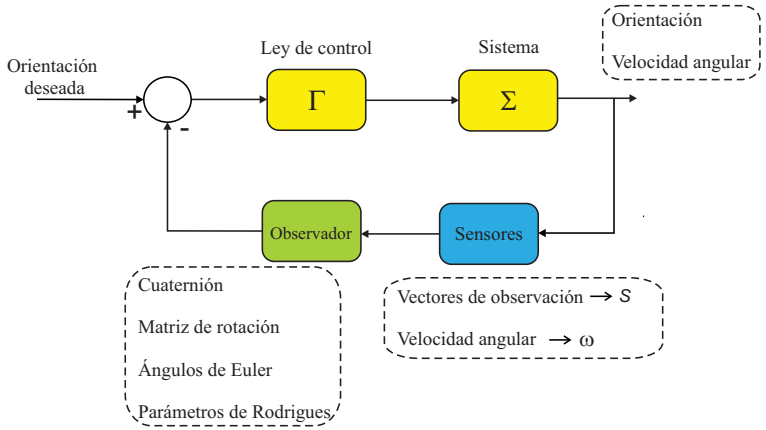


FIGURA 1. Estructura general de los sistemas de control.

del cuerpo y su velocidad angular. Puesto que la orientación no se mide directamente, se tiene que estimar usando diferentes tipos de sensores como giróscopos, magnetómetros, inclinómetros, acelerómetros, seguidores de estrellas, detectores solares, etc. Los algoritmos de estimación se basan principalmente en el filtro de Kalman y el filtro de Kalman extendido [1]. La orientación estimada se usa entonces en el control. Sin embargo, puesto que el modelo dinámico es no lineal, no es posible aplicar el principio de separación y la convergencia global del sistema no puede ser garantizada. Actualmente existen muchas unidades de orientación comerciales (estimadores de orientación) [9], [16]. De esta manera, los controladores se implementan bajo la suposición de que la estimación de la orientación corresponde a la realidad. Debido al costo de estos sistemas, se ha tratado de simplificar el problema y minimizar el número de sensores utilizados. En [8] se presenta el diseño de un observador no lineal, con la medida de torque y la orientación como entradas, para estimar la velocidad angular y estabilizar el sistema. Otros trabajos explotan la inherente pasividad de los sistemas entre el torque y la velocidad angular y entre la velocidad angular y alguna parametrización para la cinemática. La velocidad angular es reemplazada por un filtrado no lineal del cuaternión en [6] o de los parámetros de Rodrigues y los parámetros de Rodrigues modificados en [13]; el control es garantizado mediante una retroalimentación lineal. Recientemente en [12], [3] se propuso una retroalimentación de salida dinámica basada en cuaterniones para el seguimiento de orientación de una nave espacial rígida sin medida de velocidad. Los métodos mencionados han mostrado efectividad en varias aplicaciones. Sin embargo, ninguno de estos trabajos consideran solo las medidas de los vectores de referencia para el problema del control de orientación. En todos los métodos mencionados cualquier parametrización de la orientación se supone siempre como una medida de la orientación. En la Figura 1 se muestra un esquema de la estructura típica de un control de orientación en lazo cerrado.

En el presente trabajo se presentan dos leyes de control no lineales para la estabilización de la orientación de un satélite de dimensiones reducidas, el cual se considera como un cuerpo rígido en un ambiente libre de fricción con la atmósfera. Así mismo se considera que el satélite está equipado con los actuadores necesarios para modificar su orientación. En el diseño de la primera ley de control se considera

que se conocen el vector de velocidad angular y un vector de dirección de la antena de la estación en tierra expresado en el marco de referencia fijo al satélite, los cuales tienen como origen el centro de masa del mismo. Para el diseño de la segunda ley de control se omite la velocidad angular y se consideran únicamente vectores de dirección en la antena de la estación en tierra expresados en el marco de referencia fijo al satélite, en este caso, el efecto de amortiguamiento debido al vector de velocidad angular que se considera para la primera ley de control es reemplazado por un sistema dinámico "virtual"construido a partir de los vectores mencionados además de su derivada filtrada. El esquema descrito anteriormente se conoce como retroalimentación de salida con inyección dinámica. Como se mencionó anteriormente, es importante diseñar controladores que reduzcan los costos en una posible implementación práctica del sistema, en este sentido, una ventaja y originalidad del esquema propuesto es el de utilizar el menor número de sensores posibles, sin que esto implique degradar el desempeño de la estabilización. Adicionalmente, se muestra con detalle el análisis realizado a la ley de control basado en el formalismo de Lyapunov [2] donde se demuestra que estas estabilizan al sistema de manera asintótica. Como una forma de verificar la efectividad de las leyes de control se realizaron simulaciones computacionales haciendo uso de la herramienta Simulink/MATLAB, cuyos resultados se muestran en este reporte. Finalmente se construyo un prototipo de animación con el fin de tener una imagen visual del comportamiento del satélite bajo la acción de la ley de control.

El contenido adicional de este trabajo está organizado de la siguiente manera. En la sección 2 se establece la representación matemática de las herramientas a utilizar para el desarrollo de este trabajo, y se incluye una breve descripción de la forma en que se usa la información obtenida por los sensores para analizar el sistema tratado. En la sección 3 se plantea el problema a resolver en este trabajo. Los resultados principales se describen en la sección 4, se explica la naturaleza de cada uno de las dos propuestas de control y se presenta la prueba de convergencia de cada una basada en la teoría de Lyapunov. La sección 5 está dedicada a presentar los resultados en simulación de un escenario particular usando ambas propuestas. Aquí se menciona el valor de los parámetros usados con los que se obtuvieron estos resultados. Un conjunto de conclusiones se presentan en la sección 6. Finalmente se muestra la bibliografía base para la realización de este trabajo.

2. Preliminares matemáticos

Considere un satélite como el cuerpo rígido con un marco de referencia ortonormal B fijo al centro de masa, denotado por $B = [x_b, y_b, z_b]$, con ejes alineados con los ejes principales de inercia, adicionalmente un marco de referencia inercial dado por $N = [x_n, y_n, z_n]$, localizado en algún punto en el espacio. Denotando con $\vec{\omega} = [\omega_1 \;\; \omega_2 \;\; \omega_3]^T$ el vector de velocidad angular del marco del cuerpo B relativo al marco inercial N, expresado en B, el modelo matemático del satélite usado en este trabajo, conformado por la ecuación cinemática y dinámica, está dado por:

$$\dot{R} = -[\vec{\omega}]^{\times} R, \tag{1}$$

$$I\dot{\vec{\omega}} = -\vec{\omega} \times I\vec{\omega} + \vec{\Gamma}, \tag{2}$$

donde $R \in SO(3)$ es la matriz de rotación y representa la rotación del marco de referencia fijo al cuerpo B con respecto al marco de referencia inercial N, $\times$ denota

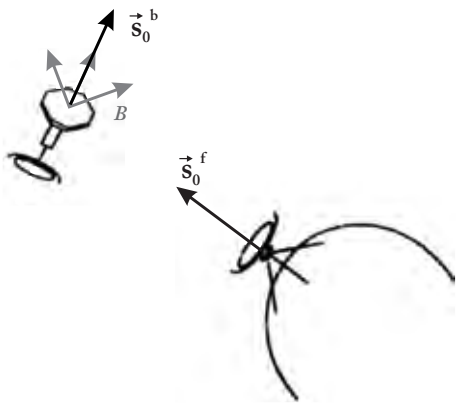


FIGURA 2. Satélite y estación en tierra.

el producto cruz, $[\xi^\times]$ es el tensor antisimétrico asociado con el vector axial ξ y está dado por

$$[\xi]^\times = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}^\times = \begin{pmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi_1 & 0 \end{pmatrix}, \tag{3}$$

$I = \text{diag}(I_1, I_2, I_3)$ representa la matriz de inercia constante definida positiva del cuerpo rígido expresada en el marco de referencia B y $\vec{\Gamma} \in \mathbb{R}^3$ es el vector de torques de control generados por los actuadores. Debe notarse que los torques también dependen de las perturbaciones del medio ambiente, pero esto no se toma en cuenta para el diseño del control. Como se puede observar, el sistema de ecuaciones (1)-(2) representa únicamente el movimiento rotacional del cuerpo rígido.

Adicionalmente, sean $\vec{s}_0^{\,b} \in \mathbb{R}^3$ un vector unitario que está en la dirección, pero en sentido opuesto, de una antena sobre el satélite y $\vec{s}_0^{\,f} \in \mathbb{R}^3$ un vector unitario fijo que esta en la dirección y el mismo sentido de la antena de recepción en la estación en tierra (ver esquema de la Figura 2).

En aplicaciones de control de orientación encontramos generalmente dos tipos de sensores:

- *Sensores de velocidad angular:* Estos sensores proporcionan la medida de la velocidad angular $\vec{\omega}$. Aquí se supone que estos sensores funcionan de forma perfecta.
- *Sensores de vectores de referencia:* Considere un vector unitario $\vec{s}_k$; con sus representaciones $\vec{s}_k^{\,f}$ en el marco de referencia inercial N y $\vec{s}_k^{\,b}$ en el marco de referencia fijo al cuerpo B. Suponiendo que $\vec{s}_k^{\,f}$ es constante, estas representaciones están relacionadas por la matriz de rotación R de la siguiente forma

$$\vec{s}_k^{\,b} = R\vec{s}_k^{\,f}. \tag{4}$$

En aplicaciones de control de orientación, los vectores $\vec{s}_k^{\,f}$ también se llaman "vectores de referencia", y en general son conocidos con bastante precisión. Los vectores

en el cuerpo $\vec{s}_k^b$ son conocidos como "vectores de observación"y se obtienen con sensores a bordo (acelerómetros, magnetómetros, detectores de sol, seguidores de estrellas, etc.). $k \in \{1, 2, ..., n\}$ representa el número de diferentes tipos de sensores de vectores de referencia. Las posibles imperfecciones como factores de escala, offset, etc. de los sensores no se toman en cuenta aquí. Los vectores de observación y la velocidad angular se relacionan a través de la ecuación cinemática

$$\dot{\vec{s}}_k^b = [\vec{s}_k^b]^\times \vec{\omega} = \vec{s}_k^b \times \vec{\omega} = -\vec{\omega} \times \vec{s}_k^b, \tag{5}$$

extendiendo a n sensores se puede hacer

$$\dot{\vec{S}} := \begin{bmatrix} \dot{\vec{s}}_1^b \\ \vdots \\ \dot{\vec{s}}_n^b \end{bmatrix} = \begin{bmatrix} (\vec{s}_1^b) \\ \vdots \\ (\vec{s}_n^b) \end{bmatrix} \vec{\omega} =: M\vec{\omega}. \tag{6}$$

Esta última expresión proporciona la idea de que es posible conocer la información de la velocidad angular $\vec{\omega}$ con solo conocer la información medida por los sensores y la derivada de esta información. Esta idea se explotará para la segunda propuesta de control.

3. Planteamiento del problema

El objetivo general consiste en alinear la antena de comunicaciones de un satélite con la antena en la base terrestre a partir de una orientación inicial cualquiera. Para esto, se asume que el satélite cuenta con sensores de velocidad angular y sensores de referencia.

Primeramente se desarrollará una ley de control considerando que la velocidad angular es disponible de manera explícita y que la dirección de la antena en la estación de tierra se obtiene por medio de un sensor de referencia, entonces la ley de control debe hacer posible que

$$\vec{s}_0^b \to \vec{s}_0^f \qquad \text{y} \qquad \vec{\omega} \to \vec{0} \qquad \text{cuando} \qquad t \to \infty, \tag{7}$$

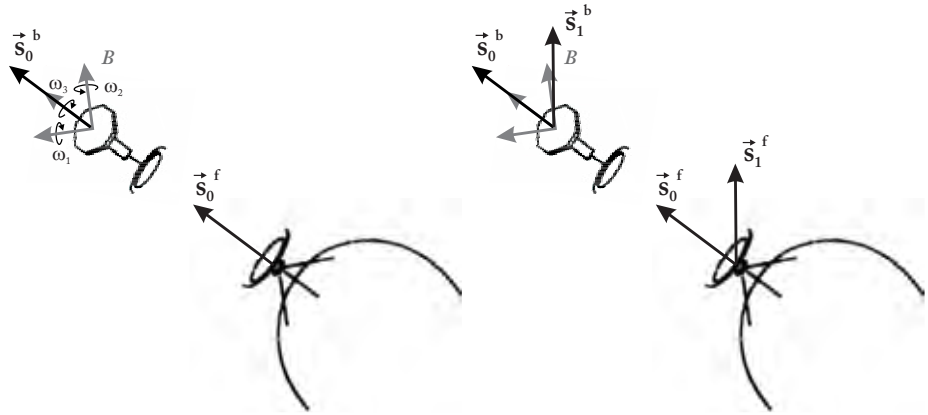
como se muestra en la Figura 3(a). La segunda propuesta considera solo la medida de dos vectores en el sistema de referencia inercial, uno que representa la dirección de la antena en la estación terrestre y otro que debe ser no colineal a este, entonces esta ley de control debe hacer posible que

$$\vec{s}_0^b \to \vec{s}_0^f \qquad \text{y} \qquad \vec{s}_1^b \to \vec{s}_1^f \qquad \text{cuando} \qquad t \to \infty, \tag{8}$$

como se muestra en al Figura 3(b).

4. Resultados principales

Para enfrentar el problema de estabilizar el satélite en una orientación predeterminada, en este trabajo se proponen dos leyes de control. La primera ley no requiere la estimación de la orientación, es decir, no se necesita el diseño de un observador (que se muestra en el esquema de la Figura 1), solo requiere sensores que midan la velocidad angular $\vec{\omega}$ del sistema y un vector de observación. La segunda ley de control además no requiere de la medida de la velocidad angular $\vec{\omega}$, pero si de por



(a) (b)

FIGURA 3. Estabilización de orientación mediante las dos leyes de control.

lo menos dos vectores de observación a partir de los cuales $\vec{\omega}$ es reconstruida. A continuación se muestra la forma de estas leyes de control.

4.1. Ley de control usando vectores de observación y velocidad angular. Supóngase que mediante un par de sensores adecuados se conocen la velocidad angular $\vec{\omega}$ y un vector de observación $\vec{s}_0^b$ sobre el sistema, el primer control propuesto para alinear las antenas del satélite y la estación en tierra es el siguiente.

4.1. PROPOSICIÓN. Considere la dinámica rotacional del satélite descrita por las ecuaciones (1)-(2) con la entrada de control $\vec{\Gamma}$ de la forma

$$\vec{\Gamma} = -\alpha\vec{\omega} - k(\vec{s}_0^f \times \vec{s}_0^b). \tag{9}$$

con $\alpha, k \in \mathbb{R}^+$.

Entonces la entrada (9) estabiliza asintóticamente el satélite al punto de equilibrio $(\vec{s}_0^b = \vec{s}_0^f, \vec{\omega} = \vec{0})$, con región de atracción $\mathbb{R}^3 \times [-1,1]^3$.

4.2. OBSERVACIÓN. De acuerdo al esquema de la Figura 2, cuando la ley de control anterior estabiliza el satélite en la orientación deseada, el vector $\vec{s}_0^b \to \vec{s}_0^f$ y por la disposición de las antenas en el satélite y la estación en tierra, estás quedan alineadas, como se puede observar en la Figura 3(a).

DEMOSTRACIÓN. Sea la función candidata de Lyapunov

$$V(\vec{\omega}, \vec{s}_0^b) = \frac{1}{2}\vec{\omega}^T I\vec{\omega} - \frac{k}{2}|\vec{s}_0^b - \vec{s}_0^f|^2 = \frac{1}{2}\vec{\omega}^T I\vec{\omega} - \frac{k}{2}(\vec{s}_0^b - \vec{s}_0^f)^T(\vec{s}_0^b - \vec{s}_0^f), \tag{10}$$

derivando lo anterior se tiene

$$\dot{V}(\vec{\omega}, \vec{s}_0^b) = \vec{\omega}^T I\dot{\vec{\omega}} - k(\vec{s}_0^b - \vec{s}_0^f)^T\dot{\vec{s}}_0^b, \tag{11}$$

sustituyendo (2) en (11) y tomando en cuenta la relación (5) se tiene

$$\dot{V}(\vec{\omega}, \vec{s}_0^b) = \vec{\omega}^T(-\vec{\omega} \times I\vec{\omega}) + \vec{\omega}^T\vec{\Gamma} - k\vec{s}_0^b(-\vec{\omega} \times \vec{s}_0^b) + k\vec{s}_0^f(-\vec{\omega} \times \vec{s}_0^b), \tag{12}$$

utilizando las propiedades del triple producto escalar, (12) se puede reescribir como

$$\dot{V}(\vec{\omega}, s_0^b) = \vec{\omega}^T \vec{\Gamma} + k\vec{\omega}(\vec{s}_0^f \times \vec{s}_0^b). \tag{13}$$

aplicando la ley de control (9) en (13) se obtiene

$$\dot{V}(\vec{\omega}, \vec{s}_0^b) = -\alpha \vec{\omega}^T \vec{\omega} \leq 0, \tag{14}$$

el teorema no permite establecer si el sistema es asintóticamente estable, pues en la ecuación (14) no está representado el vector $\vec{s}_0^b$.

Para completar la prueba, se invoca el principio de invarianza de LaSalle. Todas las trayectorias del sistema en lazo cerrado convergen al mayor conjunto invariante $\bar{\Omega}$ en $\Omega = \{(\vec{\omega}, \vec{s}_0^f) : \dot{V} = 0\} = \{(\vec{\omega}, \vec{s}_0^f) : \vec{\omega} = \vec{0}\}$. Para permanecer en este conjunto, se debe satisfacer $I\dot{\vec{\omega}} = -k[(\vec{s}_0^f \times \vec{s}_0^b)] = \vec{0}$ lo que implica que $\vec{s}_0^f \times \vec{s}_0^b = \vec{0}$. Entonces $(\vec{s}_0^b, \vec{\omega}) = (\vec{s}_0^f, \vec{0})$ es un punto de equilibrio asintóticamente estable con dominio de atracción $[-1, 1]^3 \times \mathbb{R}^3$. □

4.3. Observación. Note que una condición para la estabilidad asintótica es que $\vec{s}_0^f \times \vec{s}_0^b = 0$ y $\vec{\omega} = \vec{0}$. Esto implica que existen dos puntos de equilibrio $(\vec{s}_0^b, \vec{\omega}) = (\vec{s}_0^f, \vec{0})$ y $(\vec{s}_0^b, \vec{\omega}) = (-\vec{s}_0^f, \vec{0})$, en lazo cerrado. Estos puntos de equilibrio corresponden respectivamente a un mínimo ($V = 0$) y a un máximo ($V = 2k$) de la función de Lyapunov (10). En consecuencia $\dot{V} = 0$ en estos dos puntos de equilibrio. Si en $t_0 = 0$ el sistema en lazo cerrado se encuentra en alguno de estos puntos, entonces permanecerá ahí para todo $t > t_0$. Sin embargo, desde el punto de vista práctico y puesto que utilizamos vectores unitarios, el sistema en lazo cerrado nunca permanecerá en $(\vec{s}_0^b, \vec{\omega}) = (-\vec{s}_0^f, \vec{0})$ puesto que un pequeño ruido proporcionado por los sensores decrementará el valor de la función de Lyapunov ($V < 2k$), y debido a que $\dot{V} < 0$ fuera de estos dos puntos de equilibrio, la ley de control actúa para asegurar la convergencia al punto de equilibrio $(\vec{s}_0^b, \vec{\omega}) = (\vec{s}_0^f, \vec{0})$ para el cual $V = \dot{V} = 0$.

4.2. Ley de control usando vectores de observación. En la segunda ley de control se considerará que la velocidad angular $\vec{\omega}$ no es disponible. En su lugar se supone que se conocen las proyecciones de dos vectores fijos en el marco de referencia inercial $(\vec{s}_0^f, \vec{s}_1^f)$ sobre el marco fijo al cuerpo, es decir un par de vectores de observación.

La relación entre los vectores de observación y la velocidad angular es

$$\dot{\vec{s}}_0^b = \vec{s}_0^b \times \vec{\omega}, \tag{15}$$

$$\dot{\vec{s}}_1^b = \vec{s}_1^b \times \vec{\omega}. \tag{16}$$

que matricialmente se escribe como

$$\dot{\vec{S}} = \begin{pmatrix} \dot{\vec{s}}_0^b \\ \dot{\vec{s}}_1^b \end{pmatrix} = \begin{pmatrix} [\vec{s}_0^b]^\times \\ [\vec{s}_1^b]^\times \end{pmatrix} \vec{\omega} = M\vec{\omega} \in \mathbb{R}^6 \tag{17}$$

con $M \in \mathbb{R}^{6\times 3}$ y $\vec{\omega} \in \mathbb{R}^{3\times 1}$.

Para cuestiones prácticas, $\dot{\vec{S}}$ se obtiene por medio de la derivada filtrada que tiene la forma

$$\vec{V}_m = (pI_3 + A)^{-1} pB\vec{S}. \tag{18}$$

donde p es la variable compleja de la transformada de Laplace, $A = I_6 \cdot a \in \mathbb{R}^{6\times 6}$, $B = I_6 \cdot b \in \mathbb{R}^{6\times 6}$.

4.4. Definición. Se puede definir ahora la siguiente transformación lineal

$$\vec{\vartheta} = M^T \vec{V}_m, \tag{19}$$

la cual proporciona información sobre la velocidad angular.

4.5. Observación. El proceso de obtener el valor de la derivada filtrada $\vec{V}_m$ es una sucesión iterativa construida a partir de los vectores de observación, y multiplicarla por la matriz M^T (que también se construye a partir de estos vectores) para obtener la ley de control que da origen al sistema dinámico "virtual"que se usará en este caso, lo que se conoce como un esquema de retroalimentación de salida con inyección dinámica, con esto es posible la propuesta de control que a continuación se describe.

4.6. Proposición. Considere nuevamente la dinámica rotacional del satélite descrita por las ecuaciones (1)-(2) con la entrada de control dinámica $\vec{\Gamma}$ tal que

$$\begin{cases} \vec{\Gamma} = -\alpha_1 \vec{\vartheta} - \alpha_2 \sum_{i=0}^{1} \vec{s}_i^f \times \vec{s}_i^b, = -\alpha_1 M^T \vec{V}_m - \alpha_2 \sum_{i=0}^{1} \vec{s}_i^f \times \vec{s}_i^b, \\ \dot{\vec{V}}_m = A\vec{V}_m + B\dot{\vec{S}}. \end{cases} \tag{20}$$

con $\alpha_{1,2} \in \mathbb{R}^+$.

Entonces la entrada (20) estabiliza asintóticamente el satélite al punto de equilibrio $(\vec{s}_i^b = \vec{s}_i^f, \vec{\omega} = \vec{0})$, con $i \in \{1, 2\}$ y región de atracción $R^3 \times [-1, 1]^3$. Se puede observar que aquí solo se consideran 2 vectores de observación, sin embargo el número de vectores de observación puede ser mayor con lo que se tendría mas información y se mejoraría el control. Como antes, cuando el control estabiliza al satélite, $\vec{s}_i^b = \vec{s}_i^f$ con lo que las antenas del satélite y la estación en tierra están alineadas.A continuación se hace la prueba de convergencia de la ley de control propuesta, esta solo es para 2 vectores de observación, pero se puede extender a un número mayor.

Demostración. Sea la función candidata de Lyapunov

$$V(\vec{\omega}, \vec{s}_0^b, \vec{V}_m) = \frac{1}{2}\vec{\omega}^T I\vec{\omega} - \frac{\alpha_2}{2}\sum_{i=0}^{1} |\vec{s}_i^b - \vec{s}_i^f|^2 + \frac{k}{2}\vec{V}_m^T B^{-1}\vec{V}_m, \tag{21}$$

esto se puede reescribir como

$$V(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = \frac{1}{2}\vec{\omega}^T I\vec{\omega} - \frac{\alpha_2}{2}\sum_{i=0}^{1} (\vec{s}_i^b - \vec{s}_i^f)^T(\vec{s}_i^b - \vec{s}_i^f) + \frac{k}{2}\vec{V}_m^T B^{-1}\vec{V}_m, \tag{22}$$

derivando lo anterior se tiene

$$\dot{V}(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = \vec{\omega}^T I\dot{\vec{\omega}} - \alpha_2 \sum_{i=0}^{1} (\vec{s}_i^b - \vec{s}_i^f)^T \dot{\vec{s}}_i^b + k\vec{V}_m^T B^{-1}\dot{\vec{V}}_m. \tag{23}$$

Sustituyendo (2) y la derivada de $\vec{V}_m$ de (20) en (23), y tomando en cuenta (5) se tiene

$$\dot{V}(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = \vec{\omega}^T(-\vec{\omega} \times I\vec{\omega}) + \vec{\omega}^T\vec{\Gamma} - \alpha_2 \sum_{i=0}^{1} \vec{s}_i^b(-\vec{\omega} \times \vec{s}_i^b) + \alpha_2 \sum_{i=0}^{1} \vec{s}_i^f(-\vec{\omega} \times \vec{s}_i^b) + k\vec{V}_m^T B^{-1}(-A\vec{V}_m + B\dot{\vec{S}}), \tag{24}$$

utilizando las propiedades del triple producto escalar y tomando en cuenta que $\dot{\vec{S}} = M\vec{\omega}$, (24) se puede escribir como

$$\dot{V}(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = \vec{\omega}^T\vec{\Gamma} + \alpha_2 \sum_{i=0}^{1} (\vec{s}_i^f \times \vec{s}_i^b)\vec{\omega} + k\vec{V}_m^T B^{-1}(-A\vec{V}_m + BM\vec{\omega}). \tag{25}$$

Sustituyendo $\vec{\Gamma}$ de (20) en (25) se obtiene

$$\dot{V}(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = (k - \alpha_1)(\vec{\omega}^T M^T \vec{V}_m) - k\vec{V}_m^T B^{-1} A\vec{V}_m \tag{26}$$

haciendo $k = \alpha_1$ lo anterior se puede reescribir como

$$\dot{V}(\vec{\omega}, \vec{s}_i^b, \vec{V}_m) = -\alpha_1 \vec{V}_m^T B^{-1} A\vec{V}_m \leq 0 \tag{27}$$

Nuevamente, no es posible establecer si el sistema es asintóticamente estable, pues en la Ecuación (27) no están representados los vectores $\vec{\omega}$ y $\vec{s}_i^f$.

Finalmente, para completar la prueba, se invoca el principio de invarianza de LaSalle, del cual se tienen las siguientes implicaciones $\vec{V}_m = 0 \Rightarrow \dot{\vec{S}} = 0 \Rightarrow \vec{\omega} = 0 = \vec{\vartheta}$. Todas las trayectorias convergen al conjunto invariante mas grande $\bar{\Omega}$ en $\Omega = \{(\vec{\omega}, \vec{s}_i^f, \vec{V}_m) : \dot{V} = 0\} = \{(\vec{\omega}, \vec{s}_i^f, \vec{V}_m) : \vec{V}_m = \vec{0}\}$, donde se considera que si la velocidad angular reconstruida $\vec{V}_m = \vec{0}$, entonces la velocidad angular real $\vec{\omega} = \vec{0}$. En el conjunto invariante, $I\dot{\vec{\omega}} = [\sum_{i=1}^{2} \vec{s}_i^f \times \vec{s}_i^b] = \vec{0}$, esto es, Ω se reduce al punto de equilibrio tratado. Esto finaliza la demostración de la estabilidad asintótica del sistema en lazo cerrado. □

4.7. Observación. Como se describió en la Observación 2, una condición para la estabilidad asintótica en este caso es que $\vec{s}_0^f \times \vec{s}_0^b = 0$ y $\vec{s}_1^f \times \vec{s}_1^b = 0$. Esto implica que existe mas de un punto de equilibrio, y por razones similares a las expresadas en esa misma observación, la ley de control propuesta para este caso actúa para asegurar la convergencia al punto de equilibrio $(\vec{s}_0^b, \vec{s}_1^b) = (\vec{s}_0^f, \vec{s}_1^f)$ para el cual $V = \dot{V} = 0$.

5. Resultados en simulación

Con el fin de analizar el comportamiento del sistema en lazo cerrado, utilizando las leyes de control propuestas, se realizaron simulaciones con la herramienta Simulink/MATLAB. En esta ocasión se analiza un escenario con las condiciones iniciales para el satélite que se muestran a continuación $\vec{\omega} = (-0{,}5 \;\; 3 \;\; 1)^T$ y $(\phi \;\; \theta \;\; \phi)^T = (15° \;\; 25° \;\; -30°)^T$. Los valores utilizados para las componentes de la matriz de inercia I del satélite fueron $I_1 = I_2 = I_3 = 0.1 \; Kg\, m^2$. Los resultados obtenidos son presentados en seguida. Cabe mencionar que a pesar que en la ley de control no se utiliza ninguna parametrización de la orientación, aquí se mostrarán los ángulos de Euler con el fin de ser explícitos.

5.1. Resultados en simulación para la primera propuesta de control. En este caso, el valor utilizado para el vector de referencia es $\vec{s}_0^f = [0 \;\; 0 \;\; 1]^T$. En la Figura 4 se pueden observar gráficas con los resultados obtenidos bajo la acción de esta ley de control. La primera gráfica muestra los valores del torque aplicado al satélite, al inicio son de unos -3.5 pero rápidamente convergen a cero. La gráfica 4(b) muestra la orientación del sistema, los valor de ϕ y θ convergen a cero aproximadamente en 6 segundos y en este momento el satélite está orientado de forma que las antenas han quedado alineadas, ψ puede tomar cualquier valor ya que el movimiento alrededor del eje z del sistema no afecta la alineación de las dos antenas, incluso podría estar variando sin afectar nuestro objetivo. La gráfica 4(c) muestra la evolución de las velocidades angulares que convergen a cero en aproximadamente 6 segundos, lo que está en concordancia con la gráfica anterior, además se puede observar que los valores iniciales son los propuestos. La gráfica 4(d) expresa los valores del vector de observación $\vec{s}_0^b$ en sus tres componentes. Se puede observar como evolucionan estas componentes durante el movimiento del sistema, al estabilizarse, las componentes en x e y se anulan y solo queda la componente en z, lo que ocurre en aproximadamente 5 segundos. Con esto $\vec{s}_0^b = \vec{s}_0^f$ y las antenas quedan alineadas como se quería.

5.2. Resultados en simulación para la segunda propuesta de control. Para la simulación usando la segunda ley de control, los valores utilizados para los vectores de referencia son $\vec{s}_0^f = [0 \;\; 0 \;\; 1]^T$ y $\vec{s}_1^f = [0{,}5 \;\; 0 \;\; \sqrt{3}/2]^T$. Esta propuesta de control utiliza solo los vectores de observación para lograr la orientación del sistema, a partir de estos vectores se obtiene información de la velocidad angular a través de una derivada filtrada, lo que se describió anteriormente. Para esta simulación se utilizaron los valores $a = b = 25$ para construir las matrices diagonales A y B. Adicionalmente, los valores de los vectores de observación fueron contaminados con una perturbación de tipo ruido blanco, de amplitud 0.01 y frecuencia 1 Hz para el vector $\vec{s}_0^b$ y amplitud 0.05 con la misma frecuencia para el vector $\vec{s}_1^b$.

Los resultados obtenidos de esta simulación se muestran en las gráficas de la figura 5. En la primera gráfica del conjunto se expresan las magnitudes de los torques aplicados al sistema para su estabilización, se puede notar que la acción de los torques es una señal de alta frecuencia que al final se mantiene en un promedio de cero donde Γ_1 es el de mayor magnitud y los dos restantes son de magnitud similar. Debe mencionarse que una señal de este tipo no podría ser aplicada por un actuador tipo propulsor de gas debido a la frecuencia, pero si se pude aplicar para

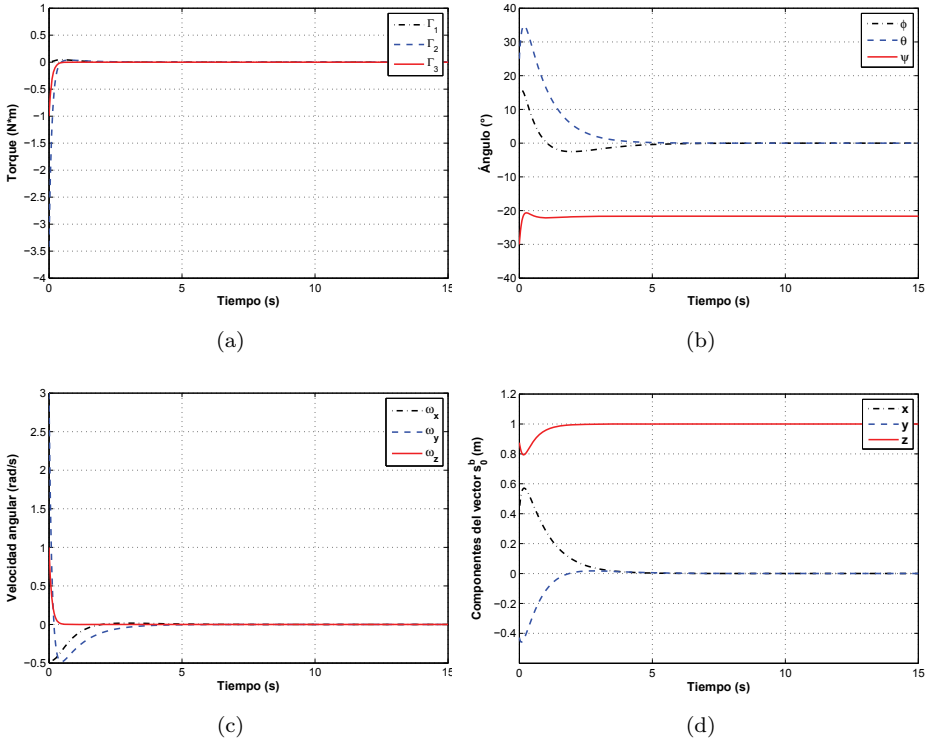
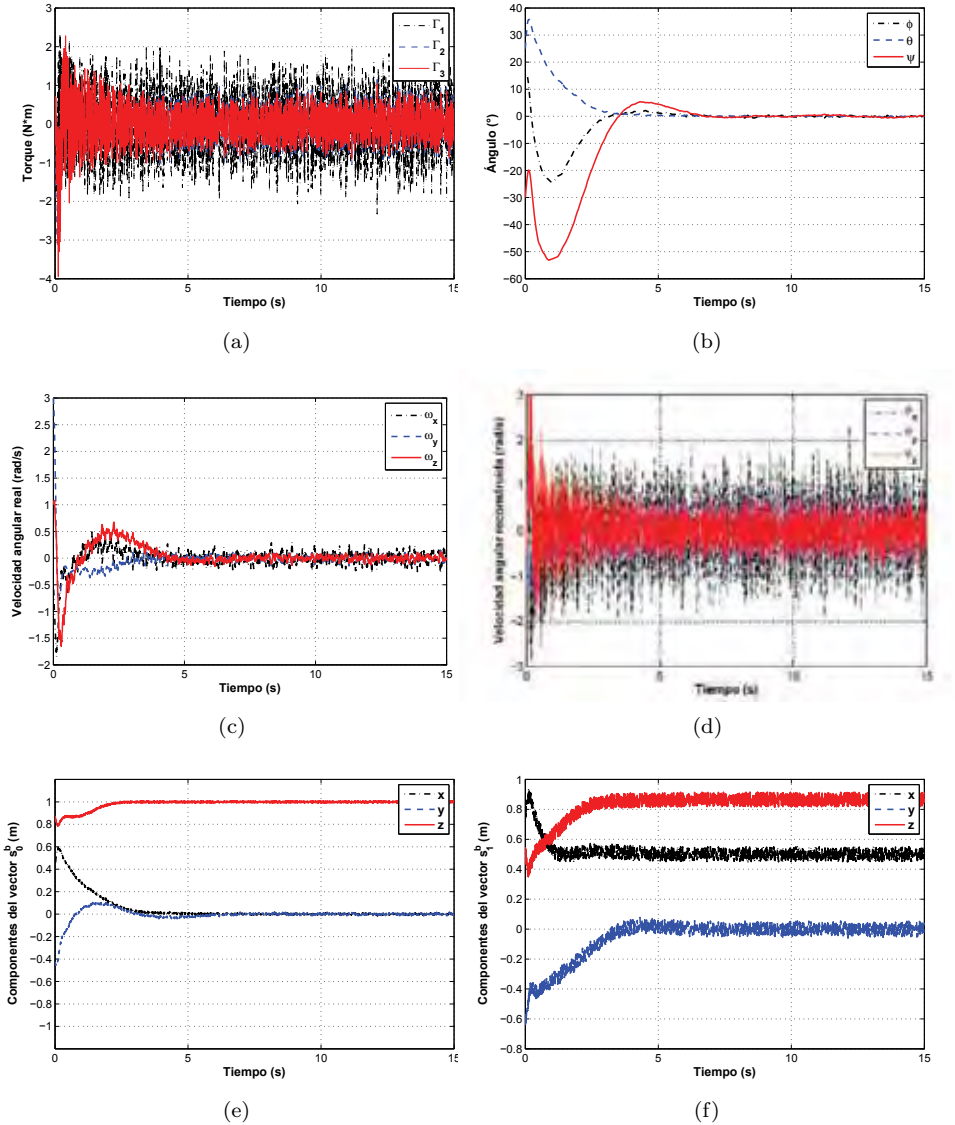


FIGURA 4. Resultados en simulación para la primera ley de control.

uno de ruedas inerciales, que funcionan en base a motores DC que en sí mismos son un filtro para este tipo de señal, con lo que lograría un mejor desempeño. La gráfica 5(b) muestra la orientación del satélite mediante los valores de los ángulos ϕ, θ y ψ, se puede observar que el tiempo de estabilización es de aproximadamente 7 segundos, llama la atención que a pesar de la forma rebuscada de la señal de control, la orientación tenga un comportamiento muy suave, lo que muestra la robustez del control. Las gráficas 5(c) y 5(d) muestran la velocidad angular real y la reconstruida mediante los vectores de observación, se observa que existe una marcada diferencia entre ambas básicamente debido al ruido registrado en los vectores base y que producen la respuesta mostrada para el control, a pesar de esto, la velocidad angular converge en un tiempo aproximado de 7 segundos y se mantiene con pequeñas variaciones. Las dos últimas gráficas 5(e) y 5(f) muestran el comportamiento de los vectores de observación $\vec{s}_0^b$ y $\vec{s}_1^b$, respectivamente, es clara la diferencia en amplitud de las perturbaciones introducidas en cada vector. Aun con esto, el control logra hacerlos coincidir, en promedio, con los vectores de referencia en un tiempo aproximado de 6 a 7 segundos en ambos casos. En este momento, $\vec{s}_0^b = \vec{s}_0^f$ y $\vec{s}_1^b = \vec{s}_1^f$, con lo que las antenas quedan alineadas en alguna vecindad debido a las perturbaciones sobre el sistema.



(a) (b) (c) (d) (e) (f)

FIGURA 5. Resultados en simulación para la segunda ley de control.

6. CONCLUSIONES

De lo expuesto anteriormente se pueden enumerar las siguientes conclusiones:

Se consiguió diseñar una ley de control que estabiliza un satélite, considerado como un cuerpo rígido, conociendo solo su orientación y velocidad angular, con un buen desempeño de este. Adicionalmente se consiguió diseñar una ley de control que estabiliza este mismo sistema conociendo solo las proyecciones sobre el sistema de referencia fijo al cuerpo de dos vectores de referencia, los cuales son usados para reconstruir la velocidad angular del cuerpo. Asimismo se demuestra que las leyes

de control estabilizan al sistema de manera asintótica mediante un análisis basado en el formalismo de Lyapunov. Por medio de simulaciones computacionales se corroboró el desempeño del sistema en lazo cerrado y se muestran su robustez con respecto a incertidumbre en el conocimiento de los parámetros y con respecto a ruido en las medidas. En términos prácticos estas leyes de control hacen posible la reducción del número de sensores necesarios para la estabilización de la orientación de sistemas tipo cuerpo rígido, lo que reduce los costos de su implementación. A sí mismo eliminan la necesidad del diseño de observadores con lo que se reduce el costo de cálculo. Finalmente, la simplicidad de estas leyes de control también permiten su uso en sistemas embarcados donde la capacidad de cómputo es reducida.

Referencias

[1] J. L. Crassidis and J. L. Junkins. *Optimal Estimation of Dynamic Systems*. Chapman and Hall/CRC, 2004.

[2] H. K. Khalil. *Nonlinear Systems, third edition*. Prentice Hall, 2002.

[3] R. Kristiansen, A. Loria, A. Chaillet, and P. J. Nicklasson. *Spacecraft relative rotation tracking whithout angular velocity measurements*. Automatica, 45(3):750-756, 2009.

[4] R. Kristiansen and P. J. Nicklasson. *Satellite attitude control by quaternion-based backstepping*. American Control Conference (ACC), 2005.

[5] M. Krstić and P. Tsiotras. *Inverse optimal stabilization of a rigid spaceraft*. IEEE Transactions on Automatic Control, 44(5):1042-1049, 1999.

[6] F. Lizarralde and J. T. Wen. *Attitude control without angular velociy measurement: A passivity approach*. IEEE Transactions on Automatic Control, 41(3):468-472, 1996.

[7] M. Osipchuck, K. D. Bharadwaj, and K. D. Mease. *Achieving good performance in global attitude stabilization*. In American Control Conference, pages 1889-1893, 1997.

[8] S. Salcudean. *A globally convergent angular velocity observer for rigid body motion*. IEEE Transactions on Automatic Control, 36(12):1493-1497, 1991.

[9] MicroStrain Microminiature Sensors. http://www.microstrain.com. October 2010.

[10] S. N. Singh and W. Yim. *Nonlinear adaptive backstepping design for spacecraft attitude control using solar radiation pressure*. In 41st IEEE conference on Decision and Control, CDC '02, 2002.

[11] C Song, S.-J. Kim, S.-H. Kim, and H. S. Nam. *Robust control of the missile attitude based on quaternion feedback*. Control Engineering Practice, 14:811-818, 2005.

[12] A. Tayebi. *Unit quaternion-based output feedback for the attitude tracking problem*. IEEE Transactions on Automatic Control, 353(6):1516-1520, 2008.

[13] P. Tsiotras. *Further passivity results for the attitude control problem*. IEEE Transactions on Automatic Control, 43(11):1597-1600, 1998.

[14] J. T. Wen and K. Kreutz-Delgado. *The attitude control problem*. IEEE Transactions on Automatic Control, 36(11):1148-1162, 1991.

[15] B. Wie, H. Weiss, and A. Arapostathis. *Quaternion feedback regulator for spacecraft eigenaxis rotations*. Journal of Guidance, Control and Dynamics, 12(3):375-380, 1989.

[16] Xsens. http://www.xsens.com. October 2010.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

r.cruzj73@gmail.com, fguerrero@ece.buap.mx, willi@fcfm.buap.mx, oliveros@fcfm.buap.mx

Capítulo 11

PROPUESTA DE ALGORITMO ESTABLE PARA LA IDENTIFICACIÓN DE FUENTES BIOELÉCTRICAS TIPO DIPOLAR

ELADIO FLORES MENA[1]
ANDRÉS FRAGUELA COLLAR[2]
JOSÉ ELIGIO MOISÉS GUTIÉRREZ ARIAS[1]
GABRIELA MORALES TIMAL[1]
MARIA MONSERRAT MORÍN CASTILLO[1]
JOSÉ JACOBO OLIVEROS OLIVEROS[2]
[1]FACULTAD DE CIENCIAS DE LA ELECTRÓNICA - BUAP
[2]FCFM - BUAP

Resumen. Los problemas de identificación son ampliamente estudiados en muchos campos de la investigación, como la medicina, donde se ha despertado un gran interés en el problema de identificación de fuentes bioeléctricas cerebrales, a partir de los datos obtenidos por medio de un electroencefalograma (EEG), entre las ventajas de este tipo de técnicas de diagnostico está el hecho de ser no invasivas, su bajo costo y su alta resolución temporal. Este problema de identificación se conoce como Problema Inverso Electroencefalográfico (PIE) y se sabe que es mal planteado en el sentido de que mediciones cercanas pueden producir grandes variaciones en la solución del problema. En este trabajo, se estudia el problema de identificar una fuente tipo dipolar con base en un modelo de medio conductor considerando una geometría esférica de la cabeza; el potencial generado en la superficie del cuero cabelludo, por efecto de la fuente, se expresa por medio del método de las funciones de Green. Por otra parte, se construyó un sistema físico que simula una fuente dipolar con la finalidad de obtener un EEG experimental. Se plantea la identificación del momento dipolar de la fuente, mediante la resolución de un problema de mínimimización, suponiendo conocida la ubicación de la misma, y a través de una técnica de regularización, se obtiene una solución estable.

1. Introducción

En diversos campos de la investigación se presentan situaciones en las cuales es necesario conocer las causas que producen cierto fenómeno a través de la información parcial que se obtiene del mismo [1]. Este tipo de problemas son llamados de identificación, y se presentan en la vida cotidiana con más frecuencia de la que imaginamos.

En particular, en el área de la medicina se ha despertado un gran interés en el estudio de los problemas de identificación, ya que permiten detectar posibles daños, anomalías o mal funcionamiento de algunos órganos del cuerpo humano. Algunos de estos problemas han sido resueltos a través las diferentes técnicas de diagnóstico, las cuales proporcionan cierta información del órgano en cuestión.

Para resolverlos, se utilizan modelos matemáticos, que por su forma se conocen como problemas inversos [2]. Por lo general, en la mayoría de los planteamientos operacionales de este tipo de problemas, se pueden presentar situaciones como la inexistencia de la solución, falta de unicidad o inestabilidad en el sentido de pequeñas perturbaciones en los datos de entrada, resultan en grandes errores en la solución. Cuando se presenta alguno de estos inconvenientes se habla de problemas mal planteados [3]. Para corregir la inexistencia o unicidad se proponen restricciones a la solución del problema, estas en general son sencillas de implementar. Sin embargo, para corregir la inestabilidad es necesario el uso de técnicas conocidas como métodos de regularización [1],[3].

Dentro del grupo de problemas inversos, uno de gran interés es el que busca identificar las fuentes bioeléctricas que son producidas por órganos como el corazón o el cerebro; cuyos efectos pueden ser observados desde el exterior del cuerpo a través de potenciales eléctricos [4].

A las fuentes que son generadas por la actividad electroquímica de estos órganos se les conoce como fuentes bioeléctricas. En este trabajo, es de especial interés recuperar aquellas que provocan los focos epilépticos. Ya que la epilepsia es un padecimiento que afecta a una parte importante de la población a nivel mundial.

2. Planteamiento del problema

El modelo matemático que se utiliza en este trabajo ha sido estudiado en [1], [3], [5], [6], en él se representa la actividad eléctrica del cerebro como un medio conductor, constituido por capas conductoras.
En este trabajo se realiza la implementación de un sistema físico que simula a una fuente dipolar inmersa en un medio conductor, con la finalidad de validar el modelo presentado, debido a las dificultades técnicas que presenta la construcción de un sistema físico completo, se construye sólo una región y se considera que dicha región representa al cerebro y la medición del potencial se realiza directamente sobre la corteza cerebral. La figura 1, muestra un esquema de esta representación, donde S denota la superficie de la corteza cerebral.

Se tiene que $\Omega = \overline{\Omega}$, donde $\overline{\Omega}$ denota a la región Ω y su frontera. En esta región se supone una conductividad constante σ_1 y $\sigma_2 = 0$, la cual corresponde a la conductividad del aire.

Se sabe que existe un potencial electrostático u en la región Ω, tal que $\mathbf{E} = \nabla u$ [1], [3] que satisface el siguiente problema de contorno:

$$
\begin{aligned}
\Delta u &= -\frac{\operatorname{div} \mathbf{J}^p}{\sigma_1} && \text{en} \quad \Omega, \\
\sigma_1 \frac{\partial u}{\partial \hat{n}} &= 0 && \text{en} \quad S,
\end{aligned}
\tag{1}
$$

donde $\mathbf{J}^p$ es la densidad de corriente primaria o impresa dentro de la región Ω. El término σ_1 denota a la conductividad del medio en la región Ω, $u = u|_\Omega$, y $\frac{\partial u}{\partial \hat{n}}$ es la

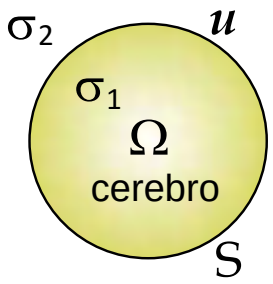


FIGURA 1. Representación de la cabeza en una sola capa conductora.

componente normal de la densidad de corriente en la superficie. En nuestro estudio, sólo se consideran densidades de corriente contenidas en el volumen cerebral, es decir, se desprecian las corrientes corticales.

Las ecuaciones (1), (2) se denomina *Problema de Contorno Electroencefalográfico* (PCE).

3. Solución del problema

La solución del Problema de Contorno Electroencefalográfico, se divide en dos partes, la solución del problema directo y la solución del problema inverso, los cuales se definen como sigue:

3.1. DEFINICIÓN. **Problema directo asociado al PCE.** Se llama Problema Directo Electroencefalográfico asociado al PCE, al problema que consiste en hallar la solución $u(x)$ del PCE cuando se conocen $f(x) = \frac{\operatorname{div} \mathbf{J}^p}{\sigma_1}$.

3.2. DEFINICIÓN. **Problema inverso asociado al PCE.** Dada una función V definida sobre S encontrar $f(x)$ de manera que para la solución $u(x)$ del problema directo correspondiente a $f(x)$ se cumpla que: $u\mid_S = V$.

El problema inverso, corresponde a la categoría de problemas mal planteados en el sentido de Hadamard [3] debido a que presenta inestabilidad numérica en su solución [1], ya que el problema es sensible a error en las mediciones, el cual se relaciona con el tipo de sistema de medición que se utilice.

Otro aspecto que se debe considerar en la solución del problema de contorno electroencefalográfico es el tipo de fuente bioeléctrica presente dentro del volumen cerebral que determina el potencial producido en el cuero cabelludo. El principal interés de este trabajo, son las fuentes bioeléctricas que generan focos epilépticos, ya que estas pueden ser modeladas por medio de dipolos.

3.1. Caso de una fuente bioeléctrica tipo dipolar.

Se sabe por [7], [8], que un foco epiléptico, usualmente se modela como una fuente tipo dipolar, ya que de esta forma se expresa a la actividad eléctrica del cerebro que se presenta en el pico del ataque, en otras palabras, la densidad de corriente se

concentra alrededor de una pequeña región.

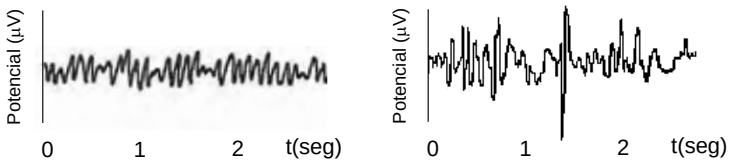


FIGURA 2. Un EEG normal y uno con presencia de un foco epiléptico.

En la figura 2, se observan dos secciones de registro de EEG, la izquierda, muestra un EEG normal, y en la derecha se observa la presencia de un exceso de potencial en aproximadamente 1.4 seg, el mayor pico en la señal se clasifica como un foco epiléptico.

Para la representación matemática de este tipo de fuentes se utilizan funciones generalizadas, las cuales no son funciones en el sentido usual, ya que su contradominio puede ser otra función. La densidad de corriente impresa $\mathbf{J}^p$ para este tipo de fuentes se expresa de la forma [8]:

$$\mathbf{J}^p = \mathbf{p}\ \delta(\mathbf{x} - a), \tag{2}$$

donde $\mathbf{p} = q\mathbf{d}$ representa el momento dipolar, q es la carga del dipolo, $\mathbf{d}$ es la distancia que separa a las cargas del dipolo y δ es la función delta de Dirac y representa la concentración de carga en el punto a.

Con base en esta expresión, para el PCE, la fuente bioeléctrica está representada por:

$$f = \frac{\text{div}\ \left[\mathbf{p}(\mathbf{y})\,\delta(\mathbf{x} - \mathbf{y})\right]}{\sigma}.$$

Una forma de hallar el potencial que satisface el problema (1), (2) es mediante el método de la función de Green, el cual permite encontrar la solución de ecuaciones diferenciales sujetas a condiciones de frontera. La función de Green es utilizada como núcleo de un operador lineal integral de modo que la solución para el problema está determinada por la expresión [9]:

$$u(\mathbf{x}) = \int_\Omega G(\mathbf{x}, \mathbf{y})\ f(\mathbf{y})\ d\mathbf{y},$$

donde $\mathbf{x}$ es el punto donde se desea conocer el potencial, $\mathbf{y}$ es el punto donde se ubica la fuente, $u(\mathbf{x})$ es la solución del problema, $G(\mathbf{x}, \mathbf{y})$ denota a la función de Green y $f(\mathbf{y})$ representa a la fuente ubicada dentro la región Ω.

Para el caso de una fuente dipolar concentrada en el punto a, la expresión anterior toma la forma [10]:

$$u(\mathbf{x}) = -\frac{q\,\mathbf{d}}{\sigma} \cdot \nabla_{\mathbf{y}} G(\mathbf{x}, \mathbf{y})\Big|_{\mathbf{y}=a}. \tag{3}$$

De modo que, el potencial en el caso de una fuente dipolar, se expresa en función de los parámetros del dipolo mismo y el gradiente de la función de Green. Por lo que se introduce la siguiente:

3.3. Definición. **Función de Green para el problema (1)-(2)** Llamaremos función de Green del problema (1)-(2) a la función $G(x, y)$ de las variables $x, y \in \Omega \subset \mathbb{R}^3$ que satisface al problema de contorno [11]

$$\Delta G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) - \frac{1}{m(\Omega)} \qquad \mathbf{x}, \mathbf{y} \in \Omega, \tag{4}$$

$$\left.\frac{\partial G}{\partial \hat{n}}\right|_S = 0 \qquad \mathbf{x} \in S = \partial\Omega, \mathbf{y} \in \Omega \subset \mathbb{R}^3, \tag{5}$$

donde $m(\Omega)$ es el volumen de la región.

Además, se sabe que la solución u del problema existe y es única salvo constantes sí [3]:

$$\int_\Omega f(\mathbf{y})\, d\Omega = 0. \tag{6}$$

3.2. Representación del potencial para el caso de una región esférica.

En el caso en que la región Ω corresponde a una esfera, para la obtención del potencial en un punto sobre la superficie de la región, se determina que dicho punto $\mathbf{x}(0, 0, z_0) \in \partial\Omega$ y la fuente se ubica en el punto $\mathbf{y}(x, y, z) \in \Omega$. Así, la función de Green se plantea de la forma [12]:

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi r} + g(\mathbf{x}, \mathbf{y}), \tag{7}$$

donde $r = |\mathbf{x} - \mathbf{y}|$, $g(\mathbf{x}, \mathbf{y}) = \alpha R^2 + g_1(\mathbf{x}, \mathbf{y})$, para $\alpha = -\frac{1}{6m(\Omega)}$, $R^2 = x^2 + y^2 + z^2$ y $\Delta g_1(\mathbf{x}, \mathbf{y}) = 0$.

Con base en lo anterior, la expresión que representa el potencial en el punto $\mathbf{x}$ para el problema (1),(2), está dado por la siguiente función de Green [12, 13]:

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi r} + \frac{1}{z_0}\frac{1}{4\pi r_1} - \frac{w}{4\pi} - \frac{R^2}{8\pi} + C, \tag{8}$$

para la cual

$$z_0' = \frac{1}{z_0}, \qquad r_1 = |\mathbf{y} - \mathbf{x}'|, \qquad w = \log_e(z_0' - z + r_1),$$

$$C = -\frac{z_0^2}{8\pi} - \frac{1}{4\pi}\log_e z_0.$$

De modo que, al calcular el gradiente de la ecuación (8) y realizar el producto interno de éste y el momento dipolar, se calcula el potencial en un punto sobre la superficie, específicamente, el polo norte de la región. Para determinar el potencial en un punto diferente, se utilizan matrices de rotación.

4. Validación del modelo matemático y su solución

Para realizar la validación del problema de contorno electroencefalográfico, se construyó un sistema físico que nos representa una fuente dipolar inmersa en un medio conductor, contenido en una región esférica.

Una forma de modelar un foco epiléptico, es a través de un dipolo de corriente equivalente [14], [15], ya que se ha observado que el campo eléctrico generado por un foco epiléptico es similar al generado por un dipolo eléctrico.

Es sabido que el comportamiento de la actividad eléctrica del cerebro es casi estático [5], [6], por lo cual la mejor forma de simular una fuente bioeléctrica dipolar es a través de un dipolo eléctrico. Se sabe, que un dipolo eléctrico está constituido por dos cargas de igual magnitud y de signo contrario separadas una distancia relativamente pequeña. Para la construcción de éste, se utilizaron generadores de Van de Graaff, éstos son máquinas electrostáticas que pueden acumular grandes cantidades de carga eléctrica [13].

El principio de funcionamiento de esta máquina se basa en el frotamiento por rotación de dos materiales distintos, en los cuales, al estar en contacto, se presenta un desprendimiento de electrones, con lo ello se genera carga eléctrica [16], ésta es transportada por una banda hacia un domo o esfera metálica hueca donde la carga es acumulada. Estos generadores, permiten la construcción del dipolo eléctrico el cual es introducido en un medio conductor cuya conductividad es de $2 \times 10^{-12} \frac{1}{\Omega \mathrm{m}}$, un valor suficientemente cercano al considerando dentro del modelo.

Los potenciales generados por esta fuente inmersa en el medio conductor artificial, se miden por medio de electrodos que son colocados de forma simétrica sobre la superficie exterior de una esfera de material no conductor, que representa al cuero cabelludo, y cuya función es convertir los potenciales iónicos en potenciales eléctricos.

Este sistema, simula la medición de potenciales sobre el cuero cabelludo y permite la obtención de valores experimentales, con la finalidad de compararlos con los obtenidos de forma teórica y realizar la identificación de la fuente.

5. Solución experimental del problema directo

Para la obtención de los potenciales teóricos en diferentes puntos sobre la frontera de la región, se desarrolló un programa en MATLAB®, por medio del cual se calcula la distribución del potencial sobre la frontera, se consideran 8 puntos, ya que éstos nos facilitarán el manejo de la información. La ubicación de dichos puntos se realiza de forma que se pueda observar la distribución del potencial en algunos puntos de la superficie, como se observa en la figura 5. El algoritmo realizado nos proporciona la solución al problema directo utilizando la ecuación (3). Este programa, considera que la región Ω está representada por una esfera y la fuente se ubica dentro de la región en una posición fija, los parámetros que conforman el momento dipolar como la carga y la distancia de separación entre ambas cargas puntuales y la conductividad del medio son valores que dependen del sistema físico que se

está utilizando.

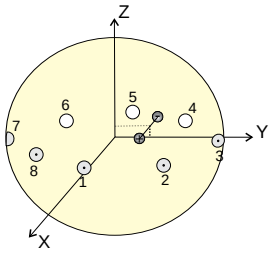


FIGURA 3. Distribución de los electrodos sobre la superficie.

Como una prueba de validación se colocó el origen de la fuente dentro de un sistema cartesiano de referencia, en la posición: (0,0.03,0.01) m y posee un momento dipolar $\mathbf{p} = 1 \times 10^{-12}(0.101,0,0)$ C m. Los valores de los potenciales obtenidos para cada uno de los electrodos se muestran en la tabla I:

No de electrodo	Potencial teórico (Volts)	Potencial experimental(Volts)
1	0,034	0,032
2	0,050	0,046
3	0,063	0,072
4	0,050	0,047
5	0,034	0,042
6	0,028	0,044
7	0,026	0,037
8	0,028	0,033

Tabla I: Comparación entre potenciales teóricos y experimentales.

Se puede observar que el valor del potencial varía en forma proporcional a la distancia entre el electrodo y la fuente, es decir, entre más cercano se ubique el electrodo de la fuente mayor será el potencial registrado en él y viceversa.
Para esta prueba, se obtuvo un error promedio de:

```
proe =

    0.0072
```

con unidades de Volts.

Estos resultados muestran, que en la mayoría de los puntos, los potenciales teóricos y experimentales son similares, de forma que se puede decir que el modelo es acorde a la situación experimental.

6. Solución del problema inverso electroencefalográfico

El problema inverso electroencefalográfico, consiste en determinar la fuente bioeléctrica que genera una determinada distribución del potencial sobre el cuero cabelludo. En este caso, ya que se estudia una fuente bioeléctrica tipo dipolar, se deben determinar los parámetros que caracterizan a dicha fuente, es decir, el momento dipolar.

Se sabe que el problema inverso electroencefalográfico es mal planteado en el sentido de que pequeñas perturbaciones en los datos resultan en una gran desviación de la solución real. Para poder obtener una solución estable del problema, se emplean las técnicas de regularización, este trabajo toma como base la técnica de regularización de Tijonov, minimizando el siguiente funcional:

$$J(X) = \|AX - Y\|^2 + \alpha\|X\|^2, \tag{9}$$

donde $A \in \mathbb{R}^{n\times p}$, para n número de datos, y p número de parámetros, $X \in \mathbb{R}^{n\times 1}$ es la solución exacta del problema, $Y \in \mathbb{R}^{p\times 1}$ denota a los datos exactos, α es el parámetro de regularización. Pero, se sabe que los datos que se tiene posee perturbaciones, de modo que la solución para esta caso, está dada por:

$$X^\delta = \underset{X}{argmin} \left\{\|AX - Y^\delta\|^2 + \alpha\|X\|^2\right\}, \tag{10}$$

donde Y^δ son los datos con error, X^δ es la solución aproximada del problema y que cumple con las ecuaciones normales modificadas:

$$A^T AX^\delta + \alpha I X^\delta = A^T Y^\delta. \tag{11}$$

Así se puede escribir que:

$$X^\delta = \left(A^T A + \alpha I\right)^{-1} A^T Y^\delta, \tag{12}$$

donde $\alpha > 0$ y se puede elegir $\alpha = \alpha(\delta)$ tal que $\alpha(\delta) \to 0$ cuando $\delta \to 0$ y X^α converja a la solución exacta cuando $\delta \to 0$ [17].

Específicamente, para el PIE, la solución del mismo se obtiene a partir de la expresión siguiente :

$$J(\mathbf{p}) = \underset{\mathbf{p}}{argmin} \left\{\sum_{k=1}^{n} \left[v(\mathbf{x}_k) - u(\mathbf{x}_k)\right]^2 + \alpha\|\mathbf{p}\|^2\right\} \tag{13}$$

donde $v(\mathbf{x}_k)$ son los datos medidos sobre la superficie de la región en los k puntos y $u(x_k) = -\mathbf{p} \cdot \nabla_{\mathbf{y}} G(\mathbf{x}_k, \mathbf{y}_k)$ son los potenciales teóricos obtenidos para cada uno de los k puntos generados por la fuente dipolar, y $\mathbf{p}$ es el momento dipolar.

Así que se establece a partir de los funcionales (9) y (13) se establecen las siguientes relaciones:

$$\begin{aligned} X &\in \mathbb{R}^{3\times 1} \qquad X = \left(\mathbf{p}_x; \mathbf{p}_y; \mathbf{p}_z\right) \\ Y &\in \mathbb{R}^{n\times 1} \qquad Y = \left(v(\mathbf{x}_1); v(\mathbf{x}_2); ...v(\mathbf{x}_n)\right) \\ A &\in \mathbb{R}^{n\times 3} \end{aligned}$$

y los elementos de la matriz A son:

$$a_{k1} = \frac{\partial G(a, \mathbf{p}_k)}{\partial x} \qquad k = 1, 2, ...n$$

$$a_{k2} = \frac{\partial G(a, \mathbf{p}_k)}{\partial y} \qquad k = 1, 2, ...n$$

$$a_{k3} = \frac{\partial G(a, \mathbf{p}_k)}{\partial z} \qquad k = 1, 2, ...n$$

donde a denota la ubicación de la fuente dentro de la región y la cual se supone conocida.

Considerando el ejemplo anterior, donde se sabe que la fuente se ubica en la posición: (0,0.03,0.01) m, de igual forma se conoce de antemano el valor de la conductividad del medio $\sigma = 2e - 12$. De modo que, con base en el gradiente de la función de Green propuesta para la solución del problema inverso, se desarrollo el algoritmo que permite la recuperación de la fuente, en este caso de los parámetros del momento dipolar; el cual se sabe que es: $\mathbf{p} = 1 \times 10^{-12}$(0.101,0,0) C m.

El algorítmo se implementó en MATLAB; primero se obtuvo una solución del problema considerando sólo la minimización del mismo, sin regularizar la solución, cuyo resultado es:

```
Y =

  1.0e-010 *
   -0.1830
    0.0000
    0.0184
```

Comparando el resultado de éste con el momento dipolar de la fuente, se observa que se tiene un error de:

```
errorp =

  1.0e-010 *
    0.1820
         0
    0.0184

errorpro =

  1.8292e-011
```

con unidades en C m.

Al aplicar diversos valores del parámetro α se obtuvieron los resultados mostrados en la tabla II; se muestran el máximo del error relativo en la primera entrada y del error absoluto, dividido entre 10^{-12} para la segunda y tercera entradas:

α	momento dipolar	error
0.001	1e-12[0.1691,0.0027,0.0269]	0.6742
0.0011	1e-12[0.1537,0.0027,0.0253]	0.5217
0.0012	1e-12[0.1409,0.0027,0.0241]	0.3950
0.0013	1e-12[0.1301,0.0027,0.0230]	0.2881
0.0014	1e-12[0.1208,0.0027,0.0221]	0.1960
0.0015	1e-12[0.1127,0.0027,0.0212]	0.1158
0.0016	1e-12[0.1056,0.0027,0.0205]	0.0455
0.0017	1e-13[0.9937,0.0268,0.1992]	0.1992
0.0018	1e-13[0.9382,0.0268,0.1936]	0.1936

Tabla II: Solución del problema para diversos valores de α

Se puede observar que el valor para el parámetro de regularización que ofrece una mejor solución al problema es $\alpha = 0{,}0016$ y para valores menores y mayores que este, la solución se aleja del valor real del momento dipolar.

La elección del parámetro de regularización, también puede ser realizada por medio del principio de discrepancia de Morozov, el cual señala que el parámetro debe elegirse de forma que [18]:

$$\|u(x^{\alpha(\delta)}) - v(x^{\alpha})\|^2 = \delta,$$

donde δ es el error que se comete en la medición. El hecho de determinar el parámetro $\alpha(\delta)$ es equivalente a encontrar el cero de la función $f(\alpha) = \|AX_{\alpha}^{\delta} - Y^{\delta}\|^2 - \delta^2$ para un delta fijo, donde X_{α}^{δ} es la solución regularizada del problema de minimización y Y^{δ} son los datos con error. La correcta elección del parámetro de regularización permite desarrollar un algoritmo estable para recuperar los parámetros del momento dipolar, es decir, realizar la identificación de la fuente bioeléctrica dipolar. Sin embargo, la elección del parámetro por este u otro método no es estudiada en este trabajo.

7. Conclusiones

En este trabajo se obtuvo la solución del problema de contorno electroencefalográfico, para una geometría esférica de la cabeza, por medio de la técnica de las funciones de Green. En particular, se consideró el caso de una fuente tipo dipolar y se obtuvo el potencial en diversos puntos sobre la superficie que representa al cuero cabelludo, es decir Con lo que se halla la solución del problema directo electroencefalográfico que a su vez sirve como base para dar solución al problema inverso. Se construyó un sistema físico, el cual consiste en una región conductora y una fuente dipolar contenida en su interior. Se observó que los resultados teóricos y experimentales para el modelo estudiado son similares; de igual forma se tomó en cuenta la presencia de errores, cuyas fuentes pueden ser muy diversas, como los errores de aproximación, los debidos a los instrumentos, al redondeo, etc, y los cuales están presentes en todas las situaciones experimentales.
Se pudieron identificar los parámetros de la fuente correspondiente al momento dipolar, mediante la propuesta de un algoritmo estable, en el cual se considera que

la fuente se localiza en una posición fija conocida, para con ello tener un problema lineal. Para realizar la identificación de la fuente en forma completa, es decir, determinar también el punto dónde se localiza el dipolo junto con el momento dipolar se requiere utilizar mínimos cuadrados no lineales, lo cual será desarrollado en trabajos futuros.

Referencias

[1] Fraguela A., Morin M., Oliveros J. 2007. *Tópicos de la Teoría de Aproximación II*. Capítulo 4. Benemérita Universidad Autónoma de Puebla. pp.73-95.

[2] Fraguela A., Morin M., Oliveros J., 1999. *Planteamiento del Problema Inverso de Localización de los Parámetros de una Fuente de Corriente Neuronal en forma de Dipolo*. Aportaciones Matemáticas, Serie Comunicaciones. Sociedad Matemática Mexicana. Vol. 25, pp 41-55.

[3] Morín M.M. 2005. *Análisis del Problema Inverso de Identificación de Fuentes a través de Planteamientos Operacionales*. Tesis Doctoral. Benemérita Universidad Autónoma de Puebla. pp. 4-45.

[4] Cromwell L., Weibell F.J., Pfeiffer E.A. 1980. *Biomedical Intrumentation and measurement*. 2a edición. Editorial Prentice-Hall U.S.A. pp. 49-62.

[5] Heller L., 1990. *Return Current in Encephalography. Variational Principles*. Biophysical Journal, Vol. 57 pp.601-607.

[6] Nunez P.L., 1981. *Electric Field of the Brain*. New York. Oxford Univ. Press pp.75-108.

[7] Grave R., González S.,Gómez C.M. 2004 *Bases biofísicas de la localización de los generadores cerebrales del electroencefalograma. Aplicación de un modelo tipo distribuido a la localización de focos epilépticos*. Revista de Neurologia, Vol 39, pp.748-756

[8] Sarvas J. 1987. *Basic Mathematical and Electromagnetic Concepts of the Biomagnetic Inverse Problem*. Phys. Med. Biol., Vol. 32, No.1 pp 11-22.

[9] Morín M.M. 1998. *Un modelo de medio conductor para el análisis de la actividad cerebral producida por fuentes puntuales*. Tesis de Licenciatura. Benemérita Universidad Autónoma de Puebla. pp. 21-33.

[10] Fraguela A., Morín M., Oliveros J. 1998 *Planteamiento del problema inverso de localización de los parámetros de una fuente neuronal en forma de dipolo*. Aportaciones matemáticas, serie comunicaciones. Vol. 25, pp. 41-55.

[11] Morín M.M. 1998. *Un modelo de medio conductor para el análisis de la actividad cerebral producida por fuentes puntuales*. Tesis de Licenciatura. Benemérita Universidad Autónoma de Puebla.

[12] Sobolev.S.L., 1964. *Partial Differential Equations of Mathematical Physics*. Addison-Wesley Publishing Company, Inc. Moscú, pp. 291-300.

[13] Rodríguez B.M. 2009. *Construcción de un Sistema Electrónico para Validar Algoritmos de Identificación de Fuentes (Puntuales) en Modelos de Medio Conductor*. Tesis de Licenciatura. Benemérita Universidad Autónoma de Puebla.

[14] Rios M.,Cabestrero, R., Maestú, F. 2007 *Neuroimagen. Técnicas y procesos cognitivos*. Editorial Elsevier-Masson. España. pp. 197.

[15] Rojo P.,Caiyoca A., Martín-Loeches M., Sola R., Pozo M. 2001 *Localización de la zona epileptógena mediante el análisis de dipolos electroencefalográficos*. Revista de Neurología Vol 32 No. 4 pp. 315-320.

[16] Serway R. 1997. *Física Tomo II*. 4ª edición, México, Mc Graw-Hill.

[17] Aquino F. 2009 *Revisión de resultados sobre el método de regularización de Tijonov* Tesis de Maestría. Benemérita Universidad Autónoma de Puebla.

[18] Kirsch A. 1996. *An Introduction to the mathematical theory of inverse problems* Springfer Verlag.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.
gmtimal@yahoo.com.mx, mmorin@ece.buap.mx, jmgutierrez@ece.buap.mx, oliveros@fcfm.buap.mx, fraguela@fcfm.buap.mx, eflores@ece.buap.mx

Capítulo 12

PROGRAMA PARA EL ANÁLISIS DEL CRECIMIENTO DE TOMATES EN AMBIENTE CONTROLADO

JOSÉ ELIGIO MOISÉS GUTIÉRREZ ARIAS[1]
IRINEO LÓPEZ CRUZ[3]
MARÍA MONSERRAT MORÍN CASTILLO[1]
RICARDO DARÍO PEÑA MORENO[2]
EDUARDO RÍOS SILVA[1]
GABRIEL ROMERO RODRÍGUEZ[1]
JUAN CARLOS TORRES-MONSIVAIS[1]
[1]FACULTAD DE CIENCIAS DE LA ELECTRÓNICA - BUAP
[2]CENTRO DE QUÍMICA DEL INSTITUTO DE CIENCIAS - BUAP
[3]ESCUELA EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA - UNIVERSIDAD AUTÓNOMA DE CHAPINGO

Resumen. El principal objetivo de este trabajo es realizar una simulación de crecimiento de cultivos partiendo de modelos matemáticos y programas por computadora capaces de describir el crecimiento aproximado del tomate (lycopersicum L.) Como resultado de una investigación y la consulta con especialistas en el campo de la Biología, se determinó que el modelo más apropiado para este trabajo es el propuesto por W. Jones y Luyten. Este modelo está descrito por ecuaciones de estado y proporciona una aproximación adecuada de las variables dependientes e independientes que participan en el crecimiento del tomate. Con este modelo se elaboró un programa por computadora utilizando la programación orientada a objetos y se creó una interfaz gráfica de usuario utilizando ventanas, botones, gráficos y tablas que nos permiten simular el crecimiento de los cultivos. Este trabajo permite predecir el crecimiento del tomate en ambiente controlado, el uso de la programación y de los métodos de optimización son utilizados para resolver las ecuaciones de crecimiento de cultivos con el fin de mejorar la calidad y aumentar la producción.

1. Introducción

Los primeros modelos de simulación datan de los años 50, donde los primeros cultivos modelados fueron el maíz, el arroz y el trigo, cerca de los años 60 apareció el concepto de sistemas dinámicos y no fue hasta los 70 que se formalizó este concepto, en los 80 y 90 se utilizaron los avances tecnológicos de la computación para estudiar y resolver diversos problemas específicos de la agricultura. Para el caso particular del tomate, en la actualidad se emplean diversos modelos de cultivos y uno de ellos es el propuesto por Jones y Luyten, esta descrito por un sistema de ecuaciones no lineal el cual describe el crecimiento potencial del tomate y haciendo uso de diversos métodos numéricos y valiéndonos de la programación orientada a objetos es posible simular y entender el crecimiento de este cultivo, y en un futuro aplicar a este modelo la teoría del Máximo de Pontryagin.

A continuación se presenta una tabla (2) que contienen las variables del modelo de Jones y Luyten y especifica las unidades y una breve definición de cada una de ellas:

Cuadro 1. Definición de Variables del modelo de Jones

Variable	**Definición**	**Unidad**
	Variables de estado	
N	Número de nodos vegetativos	no. nodos
W_s	Materia seca del follaje	$\frac{g[m.s.]}{m^2[suelo]}$
W_r	Materia seca de la raíz	$\frac{g[m.s.]}{m^2[suelo]}$
	Otras Variables	
r_m	Tasa máxima de aparición de hojas	$1/h$
$r(T)$	Función de la temperatura	$°C$
t	Tiempo transcurrido	seg
W	Materia seca total	$g[tejido]/m^2[CH_2O]$
E	Eficiencia de conversión de CH_2O	$g[m.s.]/m^2[m.s.]$
P_g	Tasa de fotosíntesis bruta	$g[CH_2O]/m^2[h]$
R_m	Tasa de respiración de mantenimiento	$g[CH_2O]/gh[tejido]$
F_c	Fracción del crecimiento total	adimensional
K	Coeficiente de extinción de luz	adimensional
τ	Conductancia de hoja	$\mu mol[CO_2]/m^2s[hojas]$
D	Coeficiente de conversión de fotosíntesis	$g[CH_2O]/m^2[h]$
C	Concentración de CO_2	$\mu mol[CO_2]/\mu mol[aire]$
α	Eficiencia de la luz	$\mu mol[CO_2]/\mu mol[fotones]$
m	Coeficiente de extinción de luz	adimensional
ϕ_1	Temperatura con fotosíntesis máxima	$°C$
ϕ_h	Temperatura con fotosíntesis cero	$°C$
T	Temperatura	$°C$
K_m	Tasa de respiración a 25 $°C$	$\mu mol[CH_2O]/\mu mol[h]$
L	Índice de área foliar	$m^2[hojas]/m^2[suelo]$
I_m	Densidad máxima de flujo de luz	$\mu mol[fotones]/m^2s[suelo]$
t_h	Longitud de día	adimensional
β	Coeficiente empírico para β	adimensional
δ	Coeficiente empírico para δ	adimensional
ρ	Densidad de población	$m^2[plantas]$
n_b	Coeficiente empírico	adimensional

2. Descripción del sistema

En este trabajo se pretende realizar un simulador de crecimiento de tomates utilizando la metodología presentada en la figura (1):

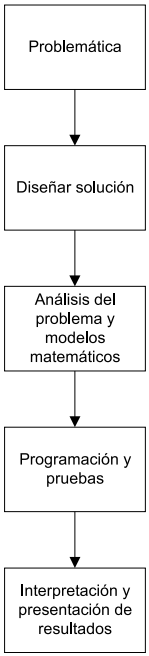


FIGURA 1. Descripción del sistema.

3. MODELO DE JONES

Las ecuaciones del modelo de Jones y Luyten (1998) [1],[3],[4] son ecuaciones que describen el crecimiento potencial de biomasa de la planta tomatera y diversos tipos de ella, para ello se considera que el sistema se encuentra a una temperatura que puede variar, además de contener parámetros ideales como son: agua, nutrientes, humedad y libre de plagas y enfermedades, entre muchas otras variables, estas ecuaciones son las siguientes:

$$\frac{dN}{dt} = r_m r(T), \tag{1}$$

$$\frac{dW_s}{dt} = E(P_g - R_m W)F_c, \tag{2}$$

$$\frac{dW_r}{dt} = E(P_g - R_m W)(1 - F_c), \tag{3}$$

donde los elementos de la ecuación (1) son:

N: Número de nodos vegetativos, este término se usa para definir el punto donde nace o crece una rama, la cual puede contener una rama solo con hojas o una rama con flores, estas flores son las que al final de la reproducción nos entregaran el fruto, dependiendo el número de flores serán el número de frutos concebidos [4].

r_m: Es la tasa máxima de aparición de hojas. La tasa de aparición de hojas tiene un comportamiento lineal. Este valor es una cantidad obtenida en cada hora dependiendo del tiempo de simulación se multiplica por el factor deseado.

t: El tiempo transcurrido. Puede darse en horas, minutos, días, semanas, pero es recomendado trabajar en lapsos de días, respetando la variable de tiempo segundo.

$r(T)$: Es una función de la temperatura $[°C]$, los rangos de temperatura son los permitidos por la planta vistos en la tabla 1;

Cuadro 2. Consideraciones de temperatura del modelo de Jones

Condición	**valor de r(T)**
$-20°C < T \leq 9°C$	$r(T) = 0$
$10°C < T \leq 15°C$	$r(T) = 0{,}55$
$16°C < T \leq 20°C$	$r(T) = 0{,}75$
$21°C < T \leq 25°C$	$r(T) = 1$
$26°C < T \leq 30°C$	$r(T) = 0{,}55$
$T > 31°C$	$r(T) = 0$

Los elementos de las ecuaciones (2) y (3) son:

W_s: Representa la biomasa del follaje o masa seca[1] del follaje. En pocas palabras, representa todo el tejido de la planta formado por encima de la tierra tomando en cuenta ramas, hojas, flores y frutos de dicha planta.

W_r: Representa la biomasa o masa seca de la raíz o comúnmente materia seca de la raíz. Esta variable representa toda aquella parte de la planta que se encuentra por debajo de la tierra.

E: Parámetro de eficiencia de conversión de CH_2O[2] a tejido de la planta. Este parámetro nos expresa la cantidad de conversión del CH_2O a tejido de la planta o materia viva de la planta.

P_g: Es la tasa de fotosíntesis bruta del follaje. Esta función es afectada por la luz, la temperatura en grados centígrados, CO_2 (dióxido de carbono) y el tamaño de la planta. Jones et al.[7], el siguiente modelo, el cual incluye la influencia de la temperatura, describe apropiadamente la fotosíntesis en el tomate. La tasa de fotosíntesis bruta P_g se expresa con la siguiente ecuación:

[1]La materia seca, ó peso seco aquí mencionada como masa seca de un cultivo se obtiene cuando una muestra es colocada en un horno a una temperatura de 105 °C durante 24 horas, el agua se evapora y el alimento seco restante es la materia seca. La materia seca de un cultivo contiene todos los nutrientes.

[2]CH_2O : En forma simplificada la fotosíntesis o reacción clorofílica, puede escribirse de la siguiente forma: $CO_2 + H_2O = CH_2O + 0_2$, en donde CH_2O representa la combinación molecular básica de un azúcar [2]. Cada molécula de CO_2 del aire queda convertida en un átomo de carbono orgánico (Corg), que pasa a formar parte de un azúcar, y en una molécula residual de oxígeno (O_2), que pasa al reservorio de la atmósfera. Por eso, de una forma más esquemática, la reacción fotosintética puede escribirse así: $CO_2 + luzsolar = Corg + O_2$.

$$P_g = D\frac{\tau C p(T)}{K}\ln\left[\frac{\alpha K I_0 + (1-m)\tau C}{\alpha K I_0 e^{-KL} + (1-m)\tau C}\right], \tag{4}$$

donde:

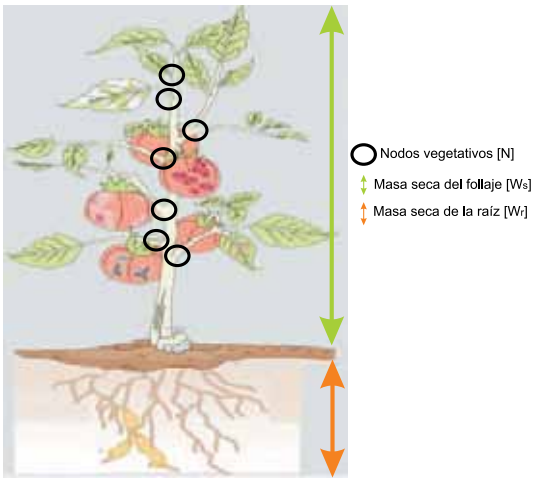


FIGURA 2. Representación de las variables de estado N, W_s y W_r.

D: Coeficiente de conversión de fotosíntesis[3] de CO_2 a CH_2O. Este proceso se produce por efecto de la fotosíntesis en las hojas y es la consecuencia de un proceso químico.

τ: Conductancia de hoja a CO_2. Esta concentración de CO_2 se encuentra en las hojas y es regulada por los estomas de las mismas.

C: Concentración de CO_2 del aire. Esta concentración se puede medir en el ambiente con aparatos especiales y se miden a partir de la cantidad de bióxido de carbono existente en aire. Las concentraciones son medidas a partir de cantidad de sustancia, y su unidad es el mol[4], ésta es una de las siete magnitudes físicas fundamentales del Sistema Internacional de Unidades.

$p(T)$: Es el factor de reducción de fotosíntesis, esta función expresa el efecto de la temperatura sobre la tasa máxima de fotosíntesis para una hoja simple.

K: Coeficiente de extinción de luz del follaje, éste coeficiente muestra la cantidad de luz que la planta pierde, ya que la luz es absorbida por la planta y debe contemplarse esta disminución en el modelo.

[3]La fotosíntesis (del griego antiguo $\varphi\omega\tau o$ [foto], "luz", y $\sigma\upsilon\nu\theta\varepsilon\sigma\iota\varsigma$ [síntesis], ünión") es la conversión de energía luminosa en energía química estable, siendo el adenosín trifosfato (ATP) la primera molécula en la que queda almacenada esa energía química.

[4]El número de unidades elementales átomos, moléculas, iones, electrones, radicales u otras partículas o grupos específicos de éstas existentes en un mol de sustancia es, por definición, una constante que no depende del material ni del tipo de partícula considerado. Esta cantidad es llamada número de Avogadro $(N_A)^2$ y equivale a: $1\ mol = 6{,}02214179 * 10^{23}$ unidades elementales.

m: Coeficiente de transmisión de luz de las hojas. Es la cantidad de luz absorbida y asimilada por la planta en cierto lapso de tiempo.

L: Índice de área foliar (IAF[5]) del follaje. Representa la cantidad de hojas por metro cuadrado.

R_m: Es la tasa de respiración de mantenimiento, la cual representa la pérdida de CO_2 debida al gasto y resíntesis[6] del tejido existente, la cual depende de la temperatura. El modelo propuesto por Gent y Enoch [5] es usado en este modelo:

$$R_m = K_m e^{0,0693(T-25)}, \tag{5}$$

donde:

T: Representa la temperatura en grados centígrados ($°C$) a la que se encuentra la planta. Se sabe que la temperatura es una de las variables de entrada más importantes y que se encuentra contenida en este modelo.

K_m: Tasa de respiración a 25 $°C$, representa la cantidad de bióxido de carbono por hora que la planta puede asimilar; para que la planta pueda respirar debe de encontrarse a una temperatura idónea, ya que si la planta se encuentra por encima o por debajo de esta temperatura, en este caso a 25 $°C$, los estomas, que son los encargados de realizar la respiración no se abrirán, esto provocará que la planta no respire y por consecuencia su muerte, es por ello la importancia de la temperatura.

W: Es la materia seca total de la planta, representa la suma total de la materia seca de la raíz y la materia seca del follaje.

f_c: Es la fracción del crecimiento total del cultivo particionado a follaje.

El cálculo de L (índice de área foliar) se realiza mediante una función exponencial la cual fue usada para ajustar el área de la hoja con el número de nodos vegetativos por Jones et al.[7] para tomates cultivados con dos niveles de CO_2 y tres niveles de temperatura. La expresión es de la siguiente forma:

$$L = \rho(\delta/\beta)\ln[1 + e^{\beta(N-n_b)}], \tag{6}$$

donde:

ρ: Densidad de población por metro cuadrado.

β,δ, n_b: Son coeficientes empíricos de la ecuación exponencial, estos coeficientes se obtiene de la experimentación al contar con datos de campo los cuales sirven para hacer un ajuste en el modelo y con esto tener resultados mas cercanos a la realidad.

$p(T)$: Es el factor de reducción de fotosíntesis, esta función expresa el efecto de la temperatura sobre la tasa máxima de fotosíntesis para una hoja simple. Se expresa con la siguiente ecuación:

[5]El término Foliar = Folio = Hoja. El área foliar de una planta se refiere a la cantidad de superficie de hoja que ella posee en determinada proporción de área de suelo.

[6]La acción o proceso de resintetizar algo en este caso conversion de la energía en tejido o materia

$$p(T) = 1 - \left[\frac{(\phi_h - T)}{(\phi_h - \phi_1)}\right]^2, \tag{7}$$

donde:

ϕ_1: Es la temperatura a la cual la fotosíntesis es máxima ($^\circ C$).

ϕ_h: Es la temperatura a la cual la fotosíntesis es cero ($^\circ C$).

α: Eficiencia de utilización de luz por la hoja.

I_0: Densidad de flujo de luz sobre el follaje, esto expresa la cantidad de fotones que hay en determinada área en un tiempo determinado. Cada hora deben ser calculados los valores de I_0, un método apropiado ha sido sugerido por Goudriaan (1986) [6], en el cual se supone una longitud de día de 12 h ($6 < t_h < 8$).

$$I_0 = I_m sen\left[2\pi\left[\frac{t_h - 6}{24}\right]\right], \tag{8}$$

donde: t_h: Es la hora del día en la que se encuentra el sistema.

I_m: Es la densidad máxima de flujo de luz por día.

4. Simulación del crecimiento

Este simulador es capaz de simular el crecimiento de una planta de tomate tomando en cuenta las variables que están involucradas en el modelo de Jones. En este software es posible cambiar parámetros y valores empíricos que en la realidad no sería posible hacerlo, además de poder intercambiar condiciones iniciales y condiciones de crecimiento para la planta como la temperatura, el flujo de luz, la respiración o la tasa de fotosíntesis en varias ocasiones y con esto es posible observar que tipo de concentraciones son las más adecuadas para el crecimiento de la planta.

Para introducir las ecuaciones de Jones al programa es necesario primero conocerlas y saber que repercusión tienen en el crecimiento de la planta, después de esto es necesario adaptar estas ecuaciones de manera correcta ya que hay algunas ecuaciones que trabajan con un lapso de días y otras que trabajan en un margen de tiempo más corto, si no se hace esto tendríamos datos incorrectos, y después es necesario resolver las ecuaciones diferenciales del modelo, esto se hace utilizando una derivada numérica para cada ecuación, ya que las variables de estado dependen entre si y sus ecuaciones involucradas. Después nosotros dentro de las ecuaciones podemos definir el tiempo de cosecha o tiempo de simulación de crecimiento de cultivo, esto depende más de nuestra experiencia del cultivo, y finalmente adecuamos los resultados obtenidos e interpretamos para saber que es lo qué se obtiene.

El diagrama de flujo general que describe al sistema se presenta a continuación, ver figura (3):

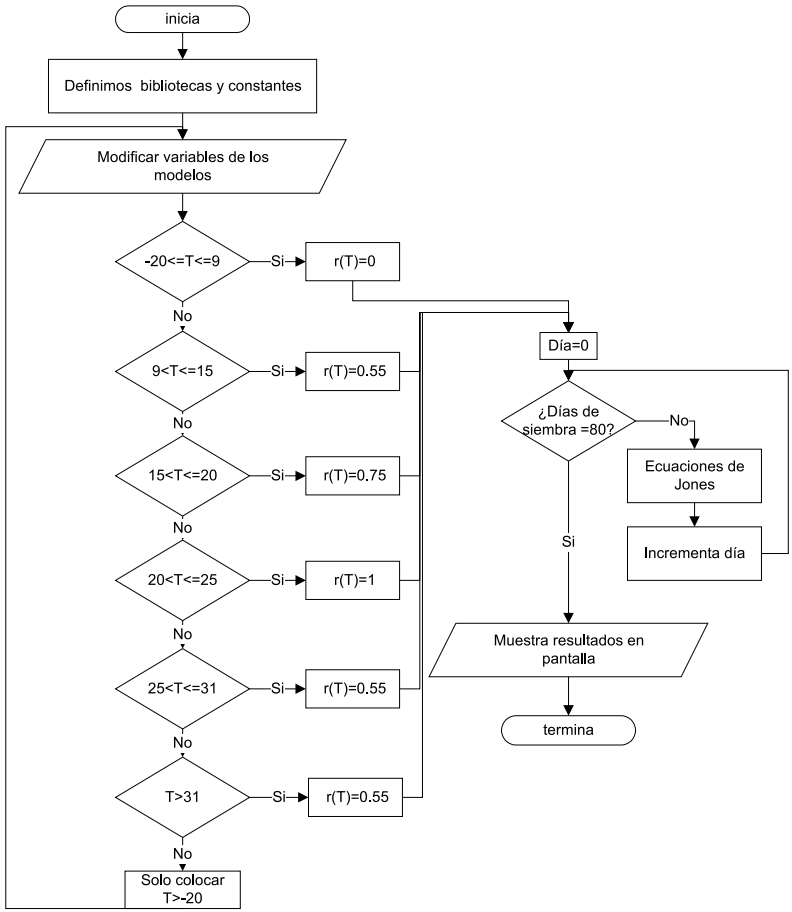


FIGURA 3. Diagrama de flujo del programa.

El programa contiene un campo para ingresar la temperatura en grados centígrados (de -20 a 50 grados centígrados), dependiendo de la temperatura que se desea iniciar será la respuesta que entregue las ecuaciones de estado. El segundo campo es para colocar las horas de luz en que el tomate trabajará, puede ser desde 0 horas hasta 24 horas; el número de horas recomendado es de 12 ya que es el periodo de horas de luz aproximado por día.

Contiene dos campos donde despliega los valores de la variable de estado, número de hojas N y la variable de estado materia seca de la raíz W_s estos usando un tipo de dato entero.

Contiene un botón llamado Ecuación 1, el cual resuelve solo la primera ecuación con respecto al tiempo, los valores de la primera variable de estado se pueden ver en la pantalla de graficado. El botón llamado Ecuación 2, resuelve solo las primeras dos ecuaciones de estado, el resultado obtenido se aprecia en la pantalla de graficado.

En la parte inferior derecha contiene 6 botones, dichos botones resuelven las tres ecuaciones de estado del sistema, podemos desplegar en la pantalla de graficado varios resultados, los cuales son:

Graficar N: entrega los valores posibles de los nodos vegetativos vs 80 días simulados (los días han sido convertidos en horas) ver gráfica (5).

Graficar W_s: muestra los valores posibles de la variable de estado materia seca del follaje vs 80 días de simulación, ver gráfica (6).

Graficar W_r: muestra los valores posibles de la variable de estado materia seca de la raíz vs 80 días de simulación, ver gráfica (7).

Graficar IAF: despliega en pantalla los valores posibles de la ecuación L denominada índice de área foliar vs 80 días de simulación, ver gráfica (8).

En la siguiente figura se presenta la pantalla principal del programa elaborado en $C\sharp$.

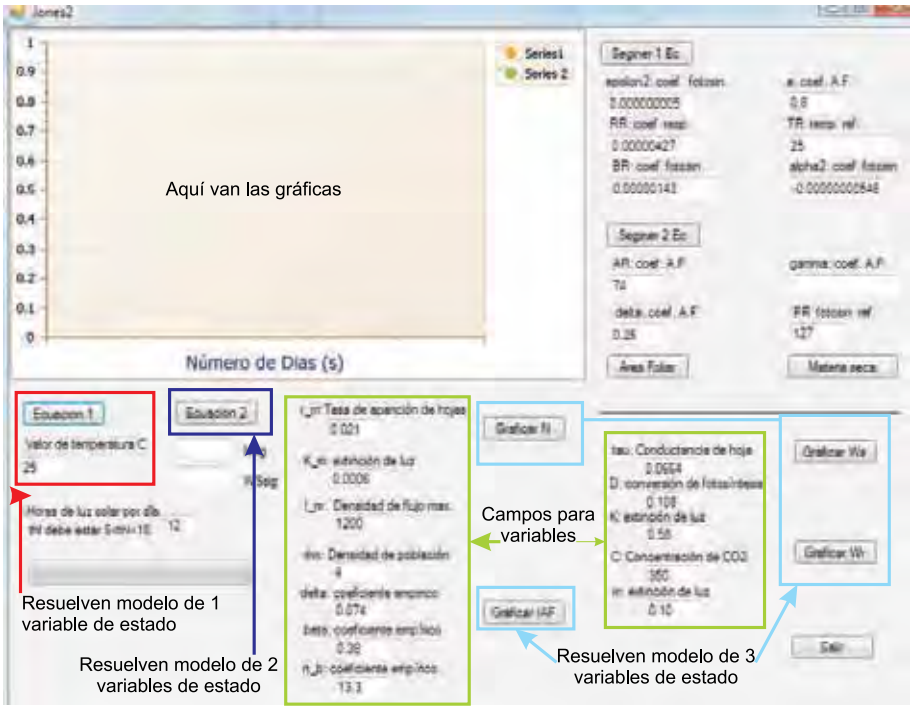


FIGURA 4. Pantalla principal del programa.

5. GRÁFICAS

La figura (5) muestra el comportamiento de la variable de estado N con diferentes temperaturas dadas en grados centígrados $°C$, la figura (5.a) muestra la variable de

estados N en el rango de $-20°C$ a $9°C$, la figura (5.b) muestra la variable N en los rangos de temperatura de $10°C$ a $15°C$, la figura (5.c) en el rango de $16°C$ a $20°C$, la gráfica (5.d) en las temperaturas de $21°C$ a $25°C$, la figura (5.e) en el rango de temperaturas de $26°C$ a $30°C$ y la figura (5.f) en rangos mayores de $31°C$. Con esto se observa que la temperatura es factor principal para el desarrollo del número de hojas con respecto al tiempo evoluciona.

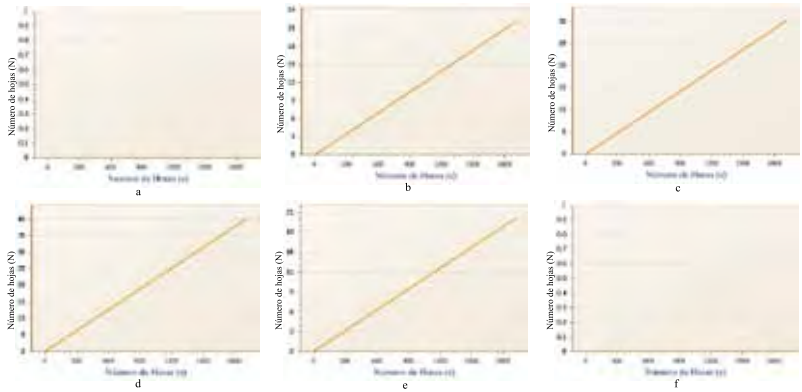


FIGURA 5. Número de nodos vegetativos vs número de horas.

La figura (6) se muestra el crecimiento de la materia seca del follaje W_s del cultivo a diferentes temperaturas, la figura (6.a) es en un rango de $-20°C$ a $9°C$, la figura (6.b) es el comportamiento debido a la temperatura de $10°C$ a $15°C$, la figura (6.c) es el crecimiento de la materia seca del follaje a un rango de $16°C$ a $20°C$, en el rango de $21°C$ se observa en la figura (6.d) el mayor crecimiento, a partir de las temperaturas superiores vistas en las figuras (6.e) y (6.f) hay un decremento de la materia seca, cabe añadir que para que las gráficas tenga un mejor resultado hay que colocar que tendrán luz durante 12 horas en el día, tiempo en que una planta real recibe la luz solar en una jornada.

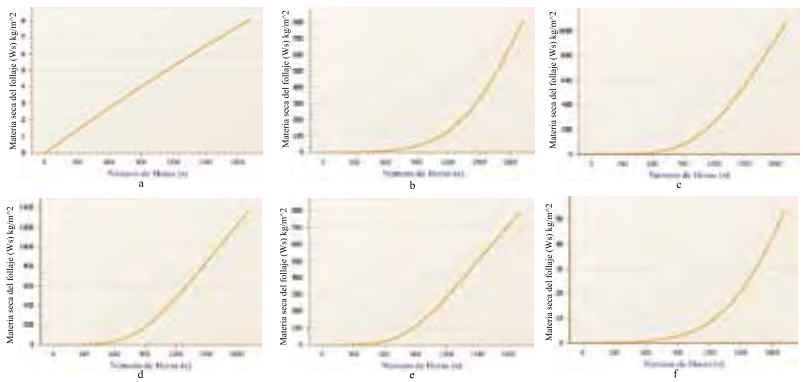


FIGURA 6. Materia seca del follaje vs número de horas.

Las siguientes figuras mostradas representan el crecimiento de la materia seca de la raíz W_r del cultivo a diferentes temperaturas, la figura (7.a) es en un rango de $-20°C$ a $9°C$, la figura (7.b) es el comportamiento debido a la temperatura de

$11°C$ a $15°C$, la figura (7.c) es el crecimiento de la materia seca del follaje a un rango de $16°C$ a $20°C$, en el rango de $21°C$ se observa en la figura (7.d) el mayor crecimiento, a partir de las temperaturas superiores vistas en las figuras (7.e) y (7.f) hay un decremento de la materia seca, cabe añadir que para que las figuras tenga un mejor resultado hay que colocar que tendrán luz durante 12 horas en el día, tiempo en que una planta real recibe la luz solar en una jornada.

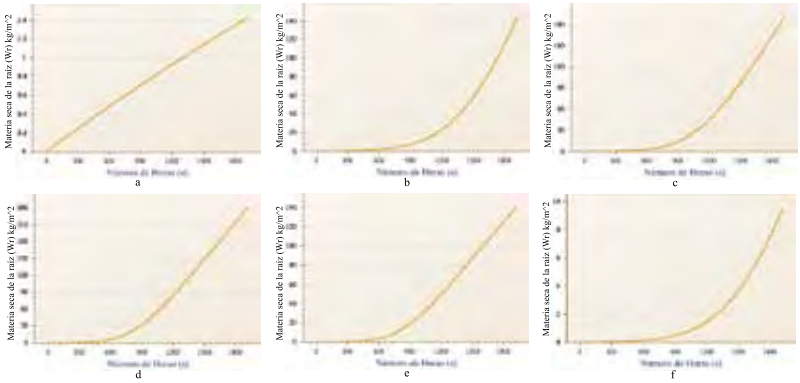


FIGURA 7. Materia seca de la raíz vs número de horas.

El índice de área foliar L se observa en la figura (8.a) diferentes temperaturas, la gráfica 8.a es en un rango de $-20°C$ a $15°C$, la figura (8.b) es el comportamiento debido a la temperatura de $16°C$ a $20°C$, la figura (8.c) es el crecimiento de la materia seca del follaje a un rango de $21°C$ a $25°C$ y por último el comportamiento debido a las temperaturas mayores a $31°C$ se observa que el mayor crecimiento se da en el rango de $25°C$ a $31°C$.

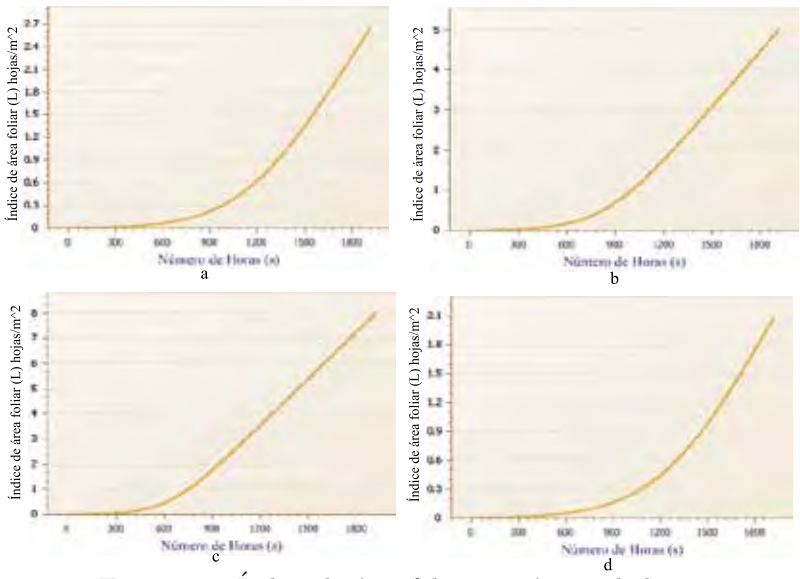


FIGURA 8. Índice de área foliar vs número de hojas.

6. Conclusiones

Utilizar el modelo de Jones para describir cultivos es apropiado para analizar el crecimiento del tomate sobre todo el de la materia seca y número de hojas, este modelo a pesar de que es un modelo simplificado del modelo denominado TOMGRO por sus siglas en inglés (TOMato GROw), y éste tiene la ventaja de mostrarnos el crecimiento potencial del cultivo a partir de la materia seca del follaje, la materia seca raíz y el número de nodos de una forma mas simple, ya que TOMGRO utiliza casi 67 variables de estado para explicar el modelo de una forma más aproximada, después en un futuro se pretende aplicar técnicas de control óptimo y con esto garantizar que el crecimiento de la planta y el producto será el óptimo.

Con este programa es posible simular el crecimiento de la planta del tomate y poder cambiar parámetros y valores que en la realidad sería imposible hacerlo y sobre todo esperar a que la planta crezca nuevamente para ver su comportamiento.

Es posible trabajar con variedades de tomates, para esto es necesario tener datos experimentales de estas variedades de cultivo y con esto realizar los cambios de los parámetros que sean necesarios en el modelo.

Referencias

[1] Irineo L. López Cruz, Introducción a la simulación de crecimiento y desarrollo de cultivos usando Fortran Simulation Translator (FST), (Universidad Autónoma de Chapingo: México).1994. pp. 31-34.

[2] Hamlyn G. Jones, Plants and microclimate: a quantitative approach to environmental plant physiology, New York, Cambridge University Press 1983. pp. 78-89.

[3] I. L. López Cruz, A. Ramírez Arias, A. Rojano Aguilar. Modelos matemáticos de hortalizas en invernadero: trascendiendo la contemplación de la dinámica de cultivos, Revista Chapingo. Serie horticultura, (Universidad Nacional de Chapingo: México). 1994. pp. 56-60.

[4] J.W. Jones, E. Dayan, L.H. Allen, H. Ankeulen, H. Challa. A dinamic tomato growth and yield model (TOMGRO). Transactions of the American Society of Agricultural Engineers, 1991. pp. 47-59.

[5] Gent, M. P. N., H. Z. Enoch. Temperature dependence of vegetative growth and dark respiration: a mathematical model. Plant physiol. 1983. 71:562-567.

[6] Gourdiaan, J. Asimple and fast numerical method for the computation of daily totals of crop photosynthesis. Agric. Forest Meteorol. 1986. 38:251-255.

[7] Jones J. W., Luyten J. C. Simulation of biological processes. University of Florida. Gainesville, Florida. 1991 pp. 47-59.

FACULTAD DE CIENCIAS DE LA ELECTRÓNICA-BUAP
Av. San Claudio y 18 Sur, Ciudad Universitaria, Col. Jardines de San Manuel
CP. 72570, Puebla, Pue. México.

jctorres_am@hotmail.com, erios@ece.buap.mx, dapena@siu.buap.mx, ilopez@correo.chapin
elezro@ece.buap.mx

Capítulo 13

DISEÑO AUTOMÁTICO DE CIRCUITOS ELECTRÓNICOS ANALÓGICOS USANDO UNA ESTRATEGIA GENERAL

JOSÉ ELIGIO MOISÉS GUTIÉRREZ ARIAS[1]
MARÍA MONSERRAT MORÍN CASTILLO[1]
RICARDO DARÍO PEÑA MORENO[3]
EDUARDO RÍOS SILVA[1]
GABRIEL ROMERO RODRÍGUEZ[1]
JUAN CARLOS TORRES MONSIVAIS[1]
ALEXANDRE ZEMLIAK[2]
[1] FACULTAD DE CIENCIAS DE LA ELECTRÓNICA - BUAP
[2] FCFM-BUAP
[3] CENTRO DE QUÍMICA DEL INSTITUTO DE CIENCIAS - BUAP

Resumen. El objetivo de este trabajo consiste en desarrollar un programa computacional para el diseño automático de circuitos electrónicos analógicos en tiempo mínimo. La técnica elegida está basada en el diseño por análisis donde se han incorporado principios de control óptimo para generar diferentes trayectorias de diseño y obtener significativas ganancias en tiempo. El modelo matemático del circuito se obtiene empleando técnicas ampliamente conocidas en el análisis de circuitos electrónicos. Se emplean diversos métodos numéricos dentro de la estrategia para solucionar el modelo y optimizar el tiempo de diseño. El programa se desarrolló bajo la plataforma de programación Visual C# usando programación orientada a objetos. Los resultados que arroja el programa permiten lograr el diseño de algunos circuitos electrónicos con base en los parámetros requeridos. Se realizaron pruebas satisfactorias con circuitos simples y se aprecia que la ganancia en tiempo se incrementa conforme aumenta su complejidad.

1. Introducción

Típicamente el proceso de diseño de circuitos electrónicos de forma automática involucra una secuencia de operaciones repetitivas de análisis que demanda grandes recursos de cómputo y tiempo de ejecución conforme los sistemas crecen en complejidad. Las tecnologías de diseño (DT[1]) para circuitos integrados florecieron en la década de 1970, y condujo a una comprensión sólida de la teoría y la práctica de modelado, análisis y síntesis de circuitos y sistemas, así como en la industria de la automatización de diseño electrónico (EDA[2]) que suministra las herramientas y los flujos [1]. Algunos de los algoritmos de la EDA evolucionaron a partir de algoritmos clásicos en ciencias de la computación (por ejemplo, la ruta más corta) y especializada para los problemas particulares de interés, mientras que otros (por ejemplo, herramientas de diseño) se inventaron para abordar las cuestiones de diseño de circuitos en la fabricación. Cuando los científicos de la década de 1970 intentaron obtener las soluciones exactas para la mayoría de los problemas de DT no las lograron debido a la complejidad intrínseca computacional, por lo que se

[1]Design Technologies por sus siglas en inglés.
[2]Electronic Design Automation por sus siglas en inglés.

emplearon métodos aproximados para la solución a problemas de diseño.

Uno de los principales retos al iniciar el proceso de diseño de un circuito electrónico, consiste en elegir la topología[3] adecuada y los valores de los componentes que se ajusten mejor para cumplir con el propósito de diseño deseado. Este problema se observa mejor en el proceso de simulación, ya que depende de los valores y condiciones que el diseñador elija, en caso de no tomar una buena elección de valores se tiende a realizar un número muy grande de simulaciones.

Para dar solución a este problema, se emplea una metodología llamada estrategia general de diseño[4] para realizar el diseño de circuitos electrónicos, la cual se basa en el modelo del circuito electrónico y en las condiciones de diseño para determinar después de un proceso de cómputo la solución de diseño en un tiempo mínimo.

Algunas ideas de la estrategia general resuelven este problema y se han presentado en diversos trabajos previos, por ejemplo en [2] se presenta una metodología de diseño óptimo en tiempo de cómputo; donde se formula una metodología para el diseño de circuitos no lineales empleando la teoría de control óptimo y un conjunto de funciones de control especiales para generalizar la metodología y producir varias estrategias de diseño diferentes dentro del mismo proceso de optimización. Otras ideas consisten en combinar dos o más de estas estrategias para obtener una ganancia en tiempo adicional además en diversos trabajos [3], [4] se evalúa la ganancia en tiempo obtenida en los resultados experimentales comparándolos con los resultados generados por dicha estrategia.

En este trabajo se presentan diversos métodos numéricos que soportan la implementación de la estrategia general y cuyo desempeño es evaluado para determinar si es necesario realizar una sustitución de métodos que ofrezcan una mayor eficiencia para lograr el objetivo de diseño en menor tiempo. Los métodos numéricos que se evalúan para la estrategia general tienen como propósito:

- a) Resolver sistemas de ecuaciones lineales y no lineales que representan al modelo del circuito.
- b) Métodos de optimización para alcanzar el punto de diseño.
- c) Métodos de integración y derivación numérica necesarios en a y b.

2. Aspectos del diseño

Ahora se presenta el planteamiento de la estrategia general de diseño de circuitos electrónicos, la cual incluye la estrategia tradicional y la estrategia tradicional modificada. La estrategia general requiere el conocimiento de tres aspectos importantes: la topología del circuito, el modelo matemático y las condiciones iniciales y de diseño.

2.1. Estrategia tradicional de diseño (ETD). El proceso de diseño de sistemas analógicos puede definirse como el problema de minimizar la función objetivo $C(\overline{x})$, donde $\overline{x} \in \mathbb{R}^{n\times 1}$ es un vector, con un conjunto de restricciones dado por el

[3]Forma de interconexión de elementos electrónicos que generan una red eléctrica.

[4]Propuesta por Zemliak, véase[6]

modelo matemático del sistema físico. Se supone que el mínimo de la función objetivo $C(\overline{x})$ lleva a cabo todos los objetivos de diseño, además la topología del circuito electrónico se conoce y se describe como:

$$g_j(\overline{x}) = 0, \quad j = 1, 2...m, \tag{1}$$

donde $g_j(\overline{x}) = 0$ representa el modelo del sistema y j es el número de ecuaciones del modelo. El vector $\overline{x}$ puede ser separado en dos partes:

$$\overline{x} = (\overline{x}', \overline{x}''), \tag{2}$$

donde $\overline{x}' \in \mathbb{R}^k$, es el vector de variables independientes, k es el número de variables independientes, $\bar{x}'' \in \mathbb{R}^m$, es el vector de variables dependientes y m es el número de variables dependientes [2],[5],[6].

Entonces el número total de variables presentes en el modelo se denota como: $n = k + m$. El sistema electrónico descrito se conoce como tradicional donde de forma natural se definen los elementos del sistema como variables independientes y los parámetros físicos (voltajes, corrientes, etcétera) como variables dependientes.

El proceso de optimización para minimizar la función objetivo $C(\bar{x})$ por un procedimiento se define en caso general como:

$$\overline{x}'^{s+1} = \overline{x}'^{s} + h^s \cdot H(\overline{x}^s), \tag{3}$$

con las restricciones del modelo (1) donde: s es el número de iteración, h^s es el paso de integración temporal, H es el vector de dirección de decremento de la función objetivo o también conocido como *método de optimización*, figura (1) [2, 3, 4].

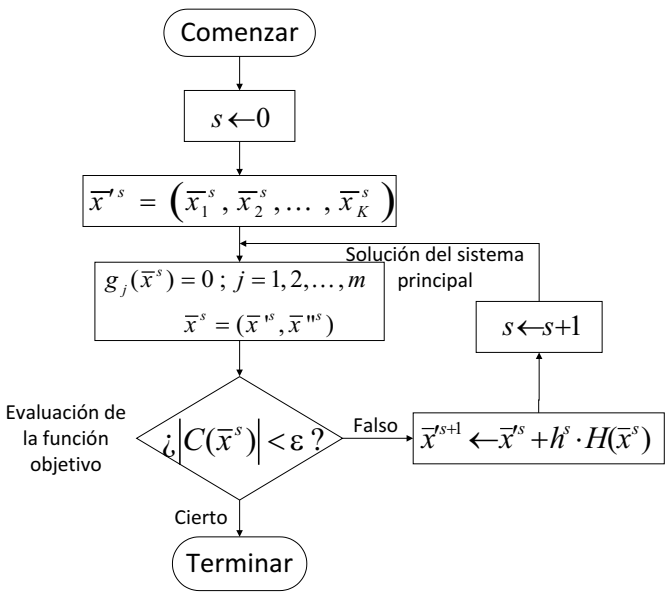


FIGURA 1. Estrategia tradicional de diseño (ETD).

El carácter específico del proceso de diseño para sistemas electrónicos consiste en el hecho de que no necesariamente satisfaga la condición del modelo (1) para todos los pasos del proceso, es suficiente satisfacer dichas condiciones en el punto final del proceso de diseño.

La figura (1) muestra el orden que sigue la metodología tradicional, donde se considera $s \leftarrow 0$ como la primera iteración en cero, después se introduce el vector de variables independientes $\overline{x}' = (x_1^s, x_2^s, ..., x_k^s)$ que representa a los elementos del circuito (resistencias, capacitores, etcétera), este vector depende de la primera iteración, pues sólo se evalua una vez.

El siguiente paso introduce y evalua el modelo del circuito $g_j(\overline{x}^s) = 0$ considerando el vector $\overline{x}^s = (\overline{x}'^s, \overline{x}''^s)$, éste realiza una separación de vectores de variables dependientes e independientes. Después se evalua la función objetivo $|C(\overline{x}^s)| < \varepsilon$ si cumple con el criterio de error ϵ termina, de lo contrario realiza un proceso de optimización para minimizar la función objetivo como: $\overline{x}'^s + h^s \cdot H(\overline{x}^s)$ y se asigna su valor al vector $\overline{x}'^{s+1}$, después vuelve a realizar el paso de la evaluación del modelo. Este proceso se realiza de forma iterativa hasta cumplir con el criterio de error.

2.2. Estrategia tradicional de diseño modificada (ETDM). Esta técnica está determinada como el procedimiento de optimización sin restricciones, es decir, sin un sistema de ecuaciones como tal y por lo tanto no existe la necesidad de resolverlo en ningún momento. Es posible utilizar este enfoque de diseño porque aparecen las funciones de penalidad que emulan al sistema, es decir, un problema con restricciones es resultado como si estas no existieran.

La función vectorial H esta definida como una funcional de la función objetivo *C(x)* y la función adicional de penalización $\varphi(x) : H^s = f(C(x), \varphi(x))$. La estructura de la función de penalización incluye todas las ecuaciones del sistema (1) y se define como:

$$\varphi(x^s) = \frac{1}{\varepsilon} \sum_{j=1}^{M} g_j^2(\overline{x}^s), \tag{4}$$

el término $\sum_{j=1}^{M} g_j^2(\overline{x}^s)$ se refiere a la sumatoria de las funciones de penalización, ε es un factor de peso, que puede ser igual a 1, cuando existe un número pequeño de funciones de penalidad. La elección de los términos $g_j^2(\overline{x}^s)$ como funciones de penalización garantiza que en el punto final $g_j^2(\overline{x}^s) = 0 \ \forall \ j$, ya que ésta es una de las condiciones de terminación del proceso de diseño.

Se define el problema de diseño como una optimización sin restricciones en el espacio $\mathbb{R}^n$, $n = m + k$ así que se define como una función aditiva:

$$F(\overline{x}) = C(\overline{x}) + \varphi(\overline{x}). \tag{5}$$

En este caso se necesita alcanzar el mínimo de la función objetivo inicial $C(\overline{x})$ y cumplir con el objetivo de diseño en el punto final del proceso de optimización. A esta aproximación se le llama método tradicional modificado. Esta metodología otorga otra estrategia de diseño así como otras trayectorias de diseño en el espacio $\mathbb{R}^n$. Desde el punto de vista computacional, el hecho de que no se tenga que resolver un sistema de ecuaciones simultáneas no lineales implica un ahorro importante en

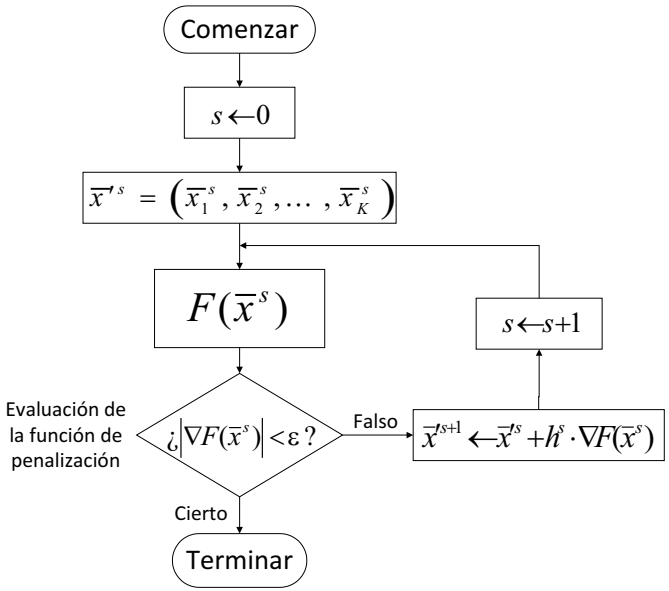


FIGURA 2. Estrategia tradicional de diseño modificado (ETDM).

el consumo de tiempo, en caso contrario un método numérico resuelve el sistema no lineal de forma iterativa hasta aproximarse a la solución.

La interpretación del diagrama de flujo (1) es similar al diagrama (2). Para procesos iterativos de diseño en la metodología modificada, el diseño inicia con la obtención del modelo $g_j(\overline{x})$, pero ahora se incluye la función de penalización $F(\overline{x})$ que incluye a la función objetivo $C(\overline{x})$ y de manera análoga se pregunta si cumple con la condición de error, en caso de cumplir el proceso termina y en caso contrario el proceso continua, ahora el vector de cambio descendente H se actualiza y se realiza el nuevo cálculo de valores y se vuelven a evaluar hasta cumplir con el criterio establecido.

2.3. Métodos de la función de penalización. Resulta conveniente plantear el problema de programación no lineal (dado como problema con restricciones (PC)); por su solución y programación dinámica, de la forma:

$$(PC) \quad min\{C^0(x) \setminus x \in D \subset \mathbb{R}^n\}, \tag{6}$$

donde $C^0 : \mathbb{R}^n \to \mathbb{R}^1$ es una función continuamente diferenciable y C es un subconjunto de $\mathbb{R}^n$.

La idea detras de los métodos de la función de penalización es la solución del problema (6), mediante la construcción de una secuencia de puntos $x_i \in \mathbb{R}^n$ los cuales son óptimos para una secuencia de problemas de minimización sin restricciones $(PS)_i$ de la forma:

$$(PS)_i \quad min\{C^0(x) + p_i(x) \setminus x \in \mathbb{R}^n\}, \quad i = 0, 1, 2, ...n, \tag{7}$$

el elemento $(PS)_i$ es construido de tal forma que $x_i \to X \in D$ cuando $i \to \infty$ y X es óptimo para (PC), el método denominado función de penalización exterior, fue propuesto por Courant [7].

2.4. Formulación de la teoría de control para el problema de diseño. La propuesta se construye a partir de la formulación del problema de diseño como el problema de control óptimo. Se considera un vector de la función especial

$$U = (u_1, u_2, ..., u_m),$$

donde $u_j \in \Omega$, $\Omega = \{0, 1\}$. Estas funciones tienen un propósito como funciones de control del proceso de diseño además de generalizar la estrategia de diseño.

El propósito de la función de control u_j es el siguiente: j representa el número de ecuaciones del sistema (1) y el término $g_j^2(\overline{x})$ es removido de la parte derecha de la expresión (4) cuando $u_j = 0$, y por otra parte el número de ecuaciones j es removido del sistema (1) y es presentado en la parte derecha de la ecuación (4) cuando $u_j = 1$.

En este caso se tienen las siguientes ecuaciones para el modelo del sistema:

$$(1 - u_j)g_j(\overline{x}) = 0, \quad j = 1, 2...m, \tag{8}$$

y para la función de penalización:

$$\varphi(\overline{x}) = \frac{1}{\varepsilon} \sum_{j=1}^{m} u_j \cdot g_j^2(\overline{x}). \tag{9}$$

Todas las variables de control u_j son las funciones de los puntos de corriente del proceso de optimización. El vector de dirección de movimiento H es la función de los vectores $\overline{x}$ y U en este caso: $H = f(\overline{x}, U)$. El número total de las trayectorias de diseño las cuales son producidas internamente en el proceso de diseño, son practicamente infinitas. Entre todas estas estrategias existe una o un número pequeño de estrategias óptimas que alcancen los objetivos de diseño para un tiempo mínimo de cómputo. Por lo tanto el problema de la búsqueda de estrategias de diseño óptimo se presentan también en la teoría de control. El problema principal de esta definición es una dependencia óptima desconocida de todas las funciones de control u_j. El problema de tiempo mínimo de la teoría de control se basa en la elección de un indicador de desempeño el cual muestre el control que cumpla con llegar a la solución en el menor tiempo.

Observación 2.1. En este trabajo los elementos del control u_j se fijan desde el inicio del proceso de optimización hasta que cumplen con el objetivo de diseño, dado que no contamos con un indicador de desempeño. Por lo tanto, la estrategia que cumpla con el menor tiempo de diseño, dentro de todas las combinaciones posibles, es la elegida como óptima.

2.5. Estrategia general de diseño (EGD). En este caso la función de penalización incluye solo el término z, $\varphi(\overline{x}^s) = \frac{1}{\varepsilon} \sum_{i=1}^{z} g_i^2(\overline{x}^s)$ donde $z \in [0, m]$ y $m - z$ ecuaciones que conforman una modificación del sistema (4):

$$g_j(\overline{x}) = 0, \quad j = z + 1, z + 2...m. \tag{10}$$

De la figura (3) resulta claro que cada nuevo valor del parametro z produce una nueva estrategia de diseño y una nueva trayectoria de diseño. Esta idea puede ser generalizada en el caso donde la función de penalización $\varphi(\overline{x})$ incluya z funciones arbitrarias. El número total de diferentes estrategias de diseño para este caso son 2^m si el parámetro z corre todos los valores de la región $[0, m]$. Todas estas estrategias existen dentro del mismo proceso de optimización.

El proceso de optimización se realiza en el espacio $\mathbb{R}^{k+z}$. El número de variables dependientes m se incrementa conforme se incrementa la complejidad del sistema y el número de estrategias de diseño crece exponencialmente. Estas estrategias tienen varios tiempos de cómputo porque tienen diferentes números de operaciones.

Resulta apropiado en este caso definir el problema de la búsqueda estrategias de diseño óptimo que cumplan con un tiempo mínimo de cómputo. En este trabajo se define la optimización de los procesos de diseño como la minimización del tiempo de cómputo.

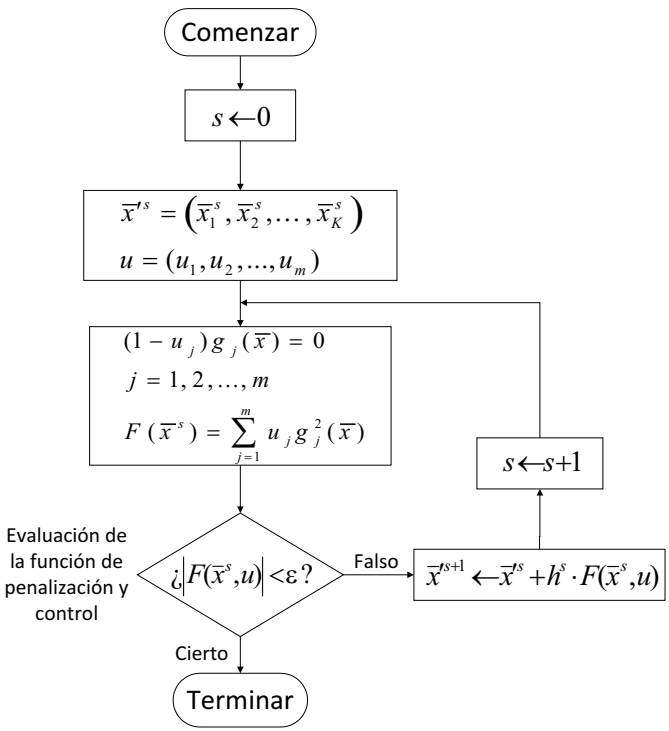


FIGURA 3. Estrategia general de diseño (EGD).

2.6. Forma continua del proceso de diseño. El proceso de diseño en la formulación de la teoría de control puede ser dispuesto en forma continua y se supone que la expresión (3) puede ser cambiada por la siguiente ecuación diferencial:

$$\frac{d(\overline{x})}{dt} = f(\overline{x}, U), \tag{11}$$

donde la función $f(\overline{x}, U)$ es la dirección del vector de movimiento H y tiene dependencia de la función objetivo generalizada $F(\overline{x}, U)$. Significa que el problema principal del proceso de diseño puede ser formulado como el problema de la integración de un sistema (11) con condiciones adicionales (8). La estructura de la función H para tres métodos se define como:

$$H \equiv f(F(\overline{x}, U)) = -F'(\overline{x}, U), \tag{12}$$

para el método del gradiente,

$$H \equiv f(F(\overline{x}, U)) = -\{F''(\overline{x}, U)\}^{-1} \cdot F'(\overline{x}, U), \tag{13}$$

para el método de Newton, donde $F''(\overline{x}, U)$ es una matriz de derivada de segundo orden,

$$H \equiv f(F(\overline{x}, U)) = -B(\overline{x}, U) \cdot F'(\overline{x}, U), \tag{14}$$

para el método de Davidon-Fletcher-Powell, donde $B(\overline{x}, U)$ es una matriz simétrica y definida positiva del algoritmo de Davidon-Fletcher-Powell. En este caso el problema de la búsqueda del proceso de diseño en tiempo óptimo ha sido formulada como el problema clásico de tiempo mínimo de la teoría de control para el sistema diferencial (11) con la parte derecha que depende concretamente del método de optimización por ejemplo en (12), (13), (14), con la función objetivo que se ha hecho para las ecuaciones (5) y (9) con condiciones adicionales (8). En este contexto el objetivo del control óptimo es poner el vector de funciones $f(\overline{x}, U)$ a cero para el tiempo final y minimizar el tiempo total de cómputo. El problema de tiempo mínimo para el sistema (11) con funciones de control no continuas puede ser resuelto de forma adecuada por medio del principio del máximo de Pontryagin [8]. Para formular el problema de control óptimo de Pontryagin es necesario definir el sistema conjugado para la función adicional ψ_i:

$$\frac{d\Psi_i}{dt} = -\sum_{l=1}^{n} \frac{\partial f_1 \overline{x}, U}{\partial x_i} \cdot \psi_l, \quad i = 1, 2, ..., n. \tag{15}$$

El Hamiltoniano del sistema esta dado como $H_0(\overline{x}, U, \Psi) = \sum_{i=1}^{n} \psi_i f_i(\overline{x}, U)$. Esta función tiene un valor supremo durante la trayectoria óptima con el principio del máximo de Pontryagin:

$$M(\overline{x}, \Psi) = supH_0(\overline{x}, U, \Psi), \qquad u \in \Omega. \tag{16}$$

Observación 2.2. El problema principal de la aplicación del principio del máximo en esta formulación es el vector desconocido Ψ_0 de los valores iniciales de la función ψ_i. Este problema tiene una solución adecuada sólo para las funciones lineales $f_i(\overline{x}, U)$.

3. Resultados

A continuación se muestran algunos de los circuitos electrónicos analógicos que se programaron para realizar el diseño electrónico empleando la metodología general y mostrar el comportamiento de las distintas trayectorias de diseño generadas.

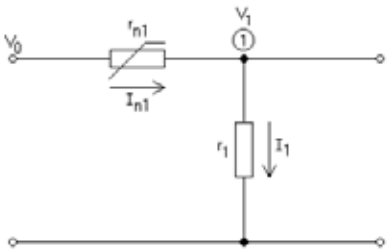


FIGURA 4. Circuito electrónico de un nodo.

EJEMPLO 3.1. 1. Se propone el diseño de un circuito muy simple, el cuál se muestra en la figura (4):

Se desea hallar el valor de r_1 para lograr que V_1 sea el 10 % de V_0, es decir $V_1 = w \cdot V_0$, con $w = 0{,}1$. Este circuito tiene una variable independiente $k = 1$ dada por la resistencia r_1 y una variable dependiente $m = 1$ representada por el voltaje nodal V_1, el número total de variables es $n = m + k = 2$.

La resistencia no lineal r_n tiene la siguiente dependencia $r_{n1} = a_n + b_n \cdot V_1$, donde a_n es una constante y b_n es un parámetro de no linealidad diferente de cero. Aplicando las leyes de Kirchhoff se tienen el modelo del circuito como:

$$V_1 = V_0 \cdot \frac{r_1}{(r_1 + r_{n1})} = V_0 \cdot \frac{r_1}{(r_1 + a_n + b_n \cdot V_1)}. \tag{17}$$

Las coordenadas del vector $\overline{x} = (x_1, x_2)$ son definidas como: $x_1^2 = r_1$ y $x_2 = V_1$. El elemento no lineal adquiere la forma $r_{n1} = a_n + b_n \cdot x_2$ entonces:

$$x_2 = V_0 \cdot \frac{(x_1^2)}{(x_1^2 + a_n + b_n \cdot x_2)}. \tag{18}$$

Así el modelo del circuito queda expresado como:

$$g(\overline{x}) = b_n \cdot x_2^2 + (x_1^2 + a_n) \cdot x_2 - V_0 \cdot x_1^2 = 0, \tag{19}$$

y sus raíces como:

$$x_2 = \frac{-(x_1^2 + a_n) \pm \sqrt{(x_1^2 + a_n)^2 + 4b_n \cdot V_0 \cdot x_1^2}}{(2b_n)}. \tag{20}$$

En este ejemplo sólo existen dos estrategias la tradicional y tradicional modificada, ambas comprenden la estrategia general. La función $C(\overline{x})$ es:

$$C(\overline{x}) = (x_2 - w \cdot V_0)^2 = 0. \tag{21}$$

Sólo una función de control u es definida en esta caso y sólo dos estrategias diferentes de diseño componen la base estructural, $u = 0$ y $u = 1$. Se emplea el método de integración de Euler y a cada estrategia:

$$\begin{aligned} \frac{dx_1}{dt} &= -b \cdot \frac{dF}{dx_1}, \\ \frac{dx_2}{dt} &= -b \cdot \frac{dF}{dx_1} \cdot u + (1 - u) \cdot \frac{\eta_2 - x_2(t - dt)}{dt}. \end{aligned} \tag{22}$$

Consecuentemente la función objetivo $F(\overline{x})$ toma las dos formas:

Caso 1.- $u = 0$ (Diseño tradicional):

$$F(\overline{x}) = C(\overline{x}) = (x_2 - w \cdot V_0)^2, \tag{23}$$

Caso 2.- $u = 1$ (Diseño tradicional modificado):

$$\begin{aligned} F(\overline{x}) &= C(\overline{x}) + \frac{1}{\varepsilon} \cdot u \cdot g_1^2(\overline{x}) \\ &= (x_2 - w \cdot V_0)^2 + \left[b_n x_2 + (x_1^2 + a_n)x_2 - V_0 x_1^2\right]^2, \end{aligned} \tag{24}$$

para todos los ejemplos $\varepsilon = 1$.

Observación 3.2. Se realizó un algoritmo y su codificación en el lenguaje de programación Visual $C\#$ para generar y graficar las trayectorias de la metodología general, el programa calcula el número de iteraciones y el tiempo de cómputo, figura (5).

Diseño n	u	Tiempo de diseño (milisegundos)	Número de iteraciones	$R1$ (ohms)	$V1$ (volts)
1	0	156	1160	0,333390747	0,100031005
2	1	202	1345	0,324910287	0,100010072

Tabla 1. Tabla de valores del circuito 1.

Se muestra la tabla (1) con los parámetros de diseño para el circuito no lineal 1, donde se observa que la estrategia 0 requiere un menor tiempo de 156 milisegundos para lograr el diseño.

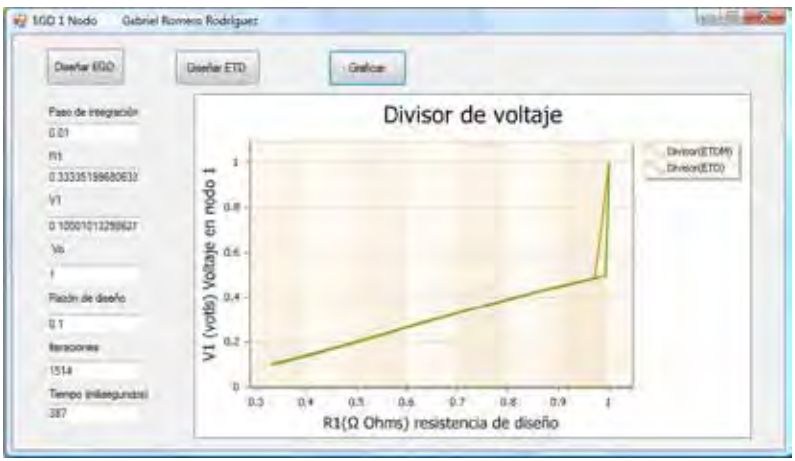


Figura 5. Diseño de un circuito de un nodo.

EJEMPLO 3.3. 2. El segundo ejemplo es un circuito no lineal pasivo de dos nodos[5] figura (6), consta de tres variables independientes $k = 3$ dadas por las admitancias y_1, y_2 y y_3 y dos variables dependientes $m = 2$ representada por los voltajes nodales V_1 y V_2.

OBSERVACIÓN 3.4. Las consideraciones siguientes son validas para ejemplos de dos y tres nodos que se presentan aquí. El número total de variables es $n = m + k$, las admitancias y_{ni}, con $i = 1 \ldots m-1$, tienen un comportamiento no lineal regido por $y_{ni} = a_n + b_n \cdot V_i$, donde $a_n \neq 0$ es constante y b_n es el parámetro de no linealidad. Se desea hallar los valores de y_j con $j = 1 \ldots k$, para lograr que V_m sea el 10 % de V_0, es decir $V_m = w \cdot V_0$, con $w = 0{,}1$.

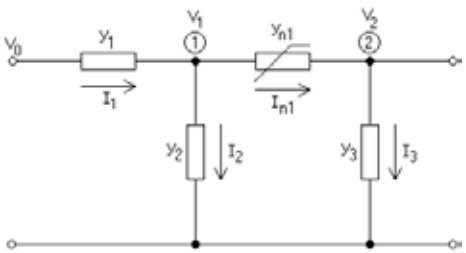


FIGURA 6. Circuito electrónico de dos nodos.

El modelo del circuito se halla aplicando el método de voltajes nodales:

$$\begin{aligned} -I_1 + I_2 + I_n &= 0, \\ -I_n + I_3 &= 0. \end{aligned} \tag{25}$$

Ahora en términos de admitancias y voltajes nodales:

$$\begin{aligned} (V_1 - V_0) \cdot y_1 + V_1 \cdot y_2 + (V_1 - V_2) \cdot y_n &= 0, \\ (V_2 - V_1) \cdot y_n + V_2 \cdot y_3 &= 0, \end{aligned} \tag{26}$$

o bien,

$$\begin{aligned} (y_1 + y_2 + y_n) \cdot V_1 + (-y_n) \cdot V_2 &= y_1 \cdot V_0, \\ (-y_n) \cdot V_1 + (y_n + y_3) \cdot V_2 &= 0. \end{aligned} \tag{27}$$

Para este caso el vector de coordenadas de fase es:

$$\overline{x} = (x_1, x_2, x_3, x_4, x_5). \tag{28}$$

Se cambian todos los valores de las admitancias y_j por x_j^2 : $x_1^2 = y_1$, $x_2^2 = y_2$, $x_3^2 = y_3$, $x_4 = V_1$, $x_5 = V_2$, así el modelo del circuito se define como:

(29)

$$\begin{aligned} g_1(\overline{x}) &= \left[(x_1^2 + x_2^2 + a_n + b_n \cdot x_4) \cdot x_4\right] - \left[(a_n + b_n \cdot x_4) \cdot x_5\right] - (x_1 \cdot V_0) = 0, \\ g_2(\overline{x}) &= \left[(a_n + b_n \cdot x_4) \cdot x_4\right] + \left[(a_n + b_n \cdot x_4 + x_3^2) \cdot x_5\right] = 0. \end{aligned}$$

Se introducen los factores mm_1 y mm_2 que limitan el proceso al caso donde se cumplan determinados valores para x_1^2 y x_2^2 y se permita variar x_3^2, así la función objetivo $C(\overline{x})$ toma la forma:

$$C(\overline{x}) = (x_5 - w \cdot V_0)^2 + (x_1^2 - mm_1)^2 + (x_2^2 - mm_2)^2 < 10^{-9}. \tag{30}$$

[5]Punto de unión de dos o más dispositivos dentro de la red eléctrica.

Para la metodología general los componentes del vector de control $u = (u_1, u_2)$ tienen diferentes combinaciones $(u_1, u_2) = (00, 01, 10, 11)$, por lo que la función objetivo $F(\overline{x})$ toma las formas:
Caso 1.- $u_1 = 0$ y $u_2 = 0$ (Diseño tradicional):

$$F(\overline{x}) = C(\overline{x}) = (x_5 - w \cdot V_0)^2 + (x_1^2 - mm1)^2 + (x_2^2 - mm2)^2. \tag{31}$$

Caso 2.- $u_1 = 0$ y $u_2 = 1$:

$$\begin{aligned} &F(\overline{x}) = C(\overline{x}) + \frac{1}{\varepsilon} \cdot u_2 \cdot g_2^2(\overline{x}), \\ &= (x_5 - w \cdot V_0)^2 + (x_1^2 - mm1)^2 + (x_2^2 - mm2)^2 \\ &+ \left[- \left[(a_n + b_n \cdot x_4) x_4 \right] + \left[(a_n + b_n \cdot x_4 + x_3^2) x_5 \right] \right]^2. \end{aligned} \tag{32}$$

Caso 3.- $u_1 = 1$ y $u_2 = 0$:

$$\begin{aligned} &F(\overline{x}) = C(\overline{x}) + \frac{1}{\varepsilon} \cdot u_1 \cdot g_1^2(\overline{x}), \\ &= (x_5 - w \cdot V_0)^2 + (x_1^2 - mm1)^2 + (x_2^2 - mm2)^2 \\ &+ \left[- \left[(x_1^2 + x_2^2 + a_n + b_n \cdot x_4) x_4 \right] - \left[(a_n + b_n \cdot x_4) x_5 \right] - \left[x_1 \cdot V_0 \right] \right]^2. \end{aligned} \tag{33}$$

Caso 4.- $u_1 = 1$ y $u_2 = 1$ (Diseño tradicional modificado):

$$\begin{aligned} &F(\overline{x}) = C(\overline{x}) + \frac{1}{\varepsilon} \cdot u_1 \cdot g_1^2(\overline{x}) + \frac{1}{\varepsilon} \cdot u_2 \cdot g_2^2(\overline{x}), \\ &= (x_5 - w \cdot V_0)^2 + (x_1^2 - mm1)^2 + (x_2^2 - mm2)^2 \\ &+ \left[- \left[(x_1^2 + x_2^2 + a_n + b_n \cdot x_4) x_4 \right] - \left[(a_n + b_n \cdot x_4) x_5 \right] - \left[x_1 \cdot V_0 \right] \right]^2 \\ &+ \left[- \left[(a_n + b_n \cdot x_4) x_4 \right] + \left[(a_n + b_n \cdot x_4 + x_3^2) x_5 \right] \right]^2, \end{aligned} \tag{34}$$

para todos los ejemplos $\varepsilon = 1$.

En este ejemplo y en el siguiente se consideran $a_n = 1$, $b_n = 10^{-5}$ y $V_0 = 1$ además se supone que los valores iniciales $k = 1 \ldots n$.

Observación 3.5. Este programa permite evaluar la estrategia tradicional y la estrategia tradicional modificada y las correspondientes combinaciones de la estrategia general. Al igual que en el programa anterior se cuenta con botones para resolver y graficar, se tienen valores iniciales con la propiedad de poder cambiar los parámetros y observar distintas trayectorias para las estrategias programadas.

A continuación se muestran la tablas (2) y (3) con los parámetros más importantes de diseño para el circuito no lineal de dos nodos, se muestran los tiempos de diseño y el número de iteraciones.

Tabla con los parámetros del circuito no lineal de dos nodos (Gradiente).

Tabla con los parámetros del circuito no lineal de dos nodos (Newton).

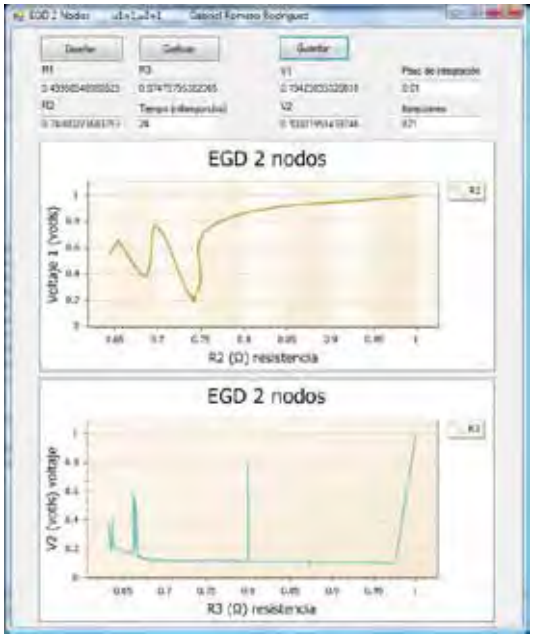


FIGURA 7. Diseño de un circuito de dos nodos.

Diseño n	Control u_1u_2	Tiempo (miliseg)	Iteración	R1 Ω	R2 Ω	R3 Ω	V1 v	V2 v
1	00	19	160	0.5	0.75	1.0	0.184	0.10
2	01	17	572	0.5	0.74	1.1	0.172	0.10
3	10	14	345	0.5	0.75	1.0	0.190	0.11
4	11	24	871	0.5	0.75	1.1	0.178	0.09

TABLA 2. Valores de los elementos del circuito 2 usando gradiente.

Diseño n	Control u_1u_2	Tiempo (miliseg)	Iteración	R1 Ω	R2 Ω	R3 Ω	V1 v	V2 v
1	00	45	1840	0.5	0.75	1.0	0.191	0.11
2	01	74	2028	0.49	0.74	0.97	0.194	0.10
3	10	79	2250	0.49	0.74	0.97	0.194	0.10
4	11	51	1899	0.49	0.73	0.96	0.194	0.10

TABLA 3. Valores de los dispositivos del circuito 2 usando Newton.

EJEMPLO 3.6. 3. El tercer ejemplo es un circuito pasivo de 3 nodos figura (8) tiene cuatro variables independientes $k = 4$ dadas por las admitancias y_1, y_2, y_3 y y_4 y tres variables dependientes $m = 3$ representada por los voltajes nodales V_1, V_2 y V_3.

Para este caso el vector de coordenadas de fase es:

$$\overline{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7), \tag{35}$$

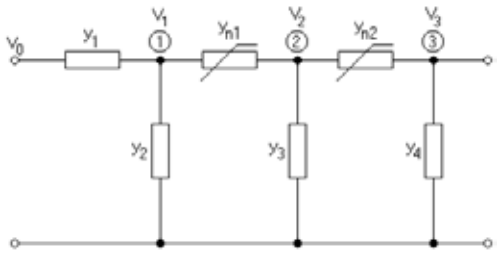


FIGURA 8. Circuito electrónico de tres nodos.

y se cambian todos los valores de las admitancias y_j por x_j^2: $x_1^2 = y_1$, $x_2^2 = y_2$, $x_3^2 = y_3$, $x_4^2 = y_4$, $x_5 = V_1$, $x_6 = V_2$, $x_7 = V_3$, así el modelo del circuito se define como:

$$
\begin{array}{ll}
 & g_1(\overline{x}) = [(x_1^2 + x_2^2 + y_0 + a \cdot x_5) \cdot x_5 - (y_0 + a \cdot x_5) \cdot x_6 - x_1^2 \cdot V_0 = 0 \\
(36) & g_2(\overline{x}) = -(y_0 + a \cdot x_5) \cdot x_5 + (2 \cdot y_0 + a \cdot x_5 + x_3^2 + a \cdot x_6) \cdot x_6 - (y_0 + a \cdot x_6) \cdot x_7 = 0 \\
 & g_3(\overline{x}) = -(y_0 + a \cdot x_6) \cdot x_6 + (y_0 + a \cdot x_6 + x_4^2) \cdot x_7 = 0
\end{array}
$$

La función objetivo $C(\overline{x})$ se define como el voltaje de salida igual a una constante, en este caso m_1, así $C(\overline{x}) = (x_7 - m_1)^2$ y con fines de comparar los tiempos de las diferentes trayectorias hacemos que todas las trayectorias lleguen al mismo punto final de diseño se agregan los términos $(x_2 - m_2)^2$ y $(x_3 - m_3)^2$, así la función objetivo final queda expresada como:

$$(37)\quad C(\overline{x}) = (x_7 - w \cdot V_0)^2 + (x_1^2 - mm_1)^2 + (x_2^2 - mm_2)^2 + (x_3^2 - mm_3)^2 < 10^{-9},$$

los factores mm_1, mm_2 y mm_3 fuerzan el proceso para que converjan en determinados valores para x_1^2, x_2^2 y x_3^2 y se permita variar x_3^2. De tal forma que las condiciones iniciales se definen como $y_0 = 1$, $a_1 = 1$, $a_2 = 1$, $V_0 = 1$, $m_1 = 0{,}1$, $m_2 = 0{,}5$ y $m_3 = 0{,}3$.

La metodología general contempla los componentes del vector de control u como: $(u_1, u_2, u_3) = (000, 001, 010, 011, 100, 101, 110, 111)$ La función objetivo $F(\overline{x})$ queda expresada como:

$$(38)\qquad F(\overline{x}) = C(\overline{x}) + \frac{1}{\varepsilon} g_1^2(\overline{x}) + \frac{1}{\varepsilon} g_2^2(\overline{x}) + \frac{1}{\varepsilon} g_3^2(\overline{x}),$$

donde se elige $\varepsilon = 1$ y considera $F(\overline{x}) < \epsilon \approx 10^{-9}$.

OBSERVACIÓN 3.7. La figura (9) muestra el programa que calcula los resultados numéricos y produce las gráficas con las trayectorias de diseño para el circuito de tres nodos, además muestra los tiempos de diseño y el número de iteraciones.

La tabla (4) muestra los parámetros de diseño obtenidos para el circuitos de tres nodos no-lineal al evaluar todas las estrategias de diseño, se observar que la estrategia 001 cumple con el objetivo de diseño en un tiempo mínimo de 768 milisegundos.

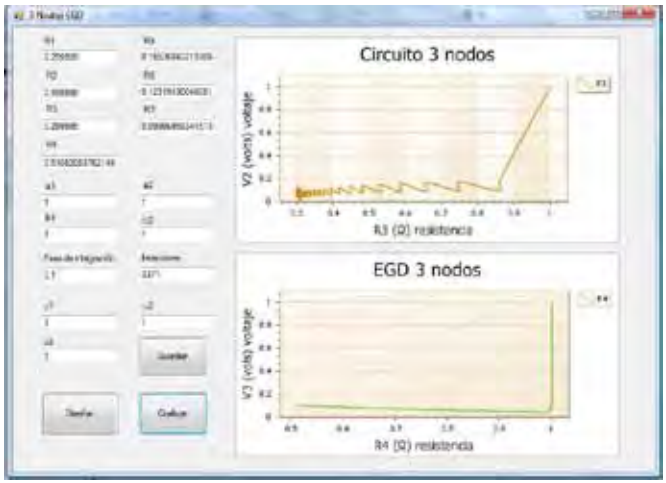


FIGURA 9. Diseño de un circuito de tres nodos.

Diseño n	Control $u_1u_2u_3$	Tiempo (miliseg)	Iteración	R1 Ω	R2 Ω	R3 Ω	R4 Ω	V1 v	V2 v	V3 v
1	000	818	27022	0.49	0.74	0.24	0.802	0.216	0.155	0.09
2	001	768	13760	0.5	0.74	0.25	0.801	0.217	0.156	0.10
3	010	867	19852	0.49	0.74	0.25	0.803	0.217	0.156	0.09
4	011	792	28645	0.5	0.75	0.25	0.801	0.218	0.157	0.10
5	100	834	18323	0.5	0.74	0.24	0.802	0.217	0.156	0.10
6	101	784	20626	0.5	0.75	0.24	0.801	0.216	0.155	0.09
7	110	806	25349	0.5	0.74	0.25	0.802	0.217	0.156	0.09
8	111	968	33711	0.5	0.74	0.24	0.802	0.217	0.156	0.10

TABLA 4. Tabla de valores del circuito 3.

OBSERVACIÓN 3.8. Todos los resultados obtenidos y mostrados en los cuadros 1-4 se programaron en un computadora personal HP $dv5-1135la$ con procesador AMD Turion^{TM} $X2$ Dual-Core con memoria $DDR2$ de 3072 MB a 800 MHz.

4. CONCLUSIONES

Después de investigar el desempeño de métodos numéricos, se programaron dentro de la metodología general, esto dió paso al diseño de circuitos no lineales donde se pudieron aplicar dichos métodos y se lograron generar distintas trayectorias de diseño.

Debido a las características que sigue la metodología en su forma matemática obliga a encontrar la solución del sistema de ecuaciones que describe el modelo en cada iteración.

El número de iteraciones durante el proceso numérico no tiene relación con el número de operaciones en cada paso, sin embargo, el conocimiento del número de operaciones incide directamente en la eficiencia del tiempo.

Referencias

[1] G. Micheli, *An Outlook on Design Technologies for Future Integrated Systems*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 28, No. 6, June (2009).

[2] E. Ríos y A. Zemliak, *Metodología de diseño óptimo en tiempo para circuitos electrónicos no lineales*, Información Tecnológica, ISSN 0716-8756. Vol. 16, No. 4, (2005), 83-90.

[3] A. Zemliak, R. Peña, E. Ríos, *Analog Network Optimization on Basis of Generalized Methodology*, 4th WSEAS Internacional Conference on Circuits, Systems, Signal and Telecomunications (CISST 2010), Harvard University, Cambridge, U.S.A., (2010).

[4] R. Peña Moreno, A. Zemliak, E. Ríos Silva, *Aplicación de la estrategia general para el diseño de circuitos analógicos con transistores* CINDET 2009, Cuernavaca, Morelos, México, (2009).

[5] A. Zemliak, *Electronic Circuit Design by General Optimization*, Nineteenth Symposium on Mathematical Programming with Data Perturbations., Book of Abstracts, pp. 5-6, Whashington D.C., USA, May (1997).

[6] A. Zemliak, *Analog system Design Problem Formulation by Optimum Control Theory*, IEICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E84-A, No. 8, pp. 2029-2041, Tokio, Japan, Aug. (2001).

[7] R. Courant, *Variational methods for the solution of problems of equilibrium and vibration*, Bull. Am. Math. Soc., Vol. 49, pp. 1-43, (1943).

[8] L. S. Pontryagin, V. G. Boltyanski, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Authorized translation from the Russian by Interscience Publishers, New York, USA, 1962.

FACULTAD DE CIENCIAS DE LA ELECTRÓNICA-BUAP
Av. San Claudio y 18 Sur, Ciudad Universitaria, Col. Jardines de San Manuel
CP. 72570, Puebla, Pue. México.

elezro@ece.buap.mx, erios@ece.buap.mx, azemliak@fcfm.buap.mx

Capítulo 14

VALIDACIÓN NUMÉRICA DE UN CONTROL ÓPTIMO DISCRETO PARA UN ROBOT MÓVIL

JOSÉ ELIGIO MOISÉS GUTIÉRREZ ARIAS
MARIA MONSERRAT MORÍN CASTILLO
GELACIO SALAS ORTEGA
FACULTAD DE CIENCIAS DE LA ELECTRÓNICA - BUAP

Resumen. En este trabajo presentamos el modelo dinámico de un robot móvil autónomo de los que también son conocidos como seguidores de línea, así como la síntesis de un algoritmo de control para la estabilización del robot en trayectorias programadas. Analizamos algunos conceptos matemáticos importantes de la teoría de optimización para el planteamiento del problema de control optimal. En la práctica, un problema de optimización es aquél en donde se desea conducir la solución del sistema a un estado objetivo y_d y para ello se minimiza la distancia entre el estado inicial y e y_d. Así, con este planteamiento, nuestro problema de control, se reduce al cálculo de puntos extremos con restricciones.

La solución al problema de optimización planteado en este trabajo se encuentra mediante el principio discreto del mínimo, esto nos lleva a resolver una ecuación algebraica matricial del tipo Riccati.

1. Introducción

La tendencia en los sistemas de ingeniería es hacia una mayor complejidad debido a los requerimientos de tareas cada vez más complicadas y de elevada precisión. En la Robótica y en el área de la inteligencia artificial por ejemplo, surgen nuevas teorías, con el fin de resolver objetivos que mejoren la capacidad de autonomía de los robots. La idea de construir un robot que pueda moverse en un entorno de manera libre (sin intervención del ser humano) es muy atractiva, estos robots son los llamados robots móviles y cuando hablamos de estos, nos referimos a un tipo particular de agentes que se construyen con un mínimo de inteligencia, necesaria para interactuar en el entorno físico que le rodea. Estos robots reciben estímulos provenientes de sensores que miden propiedades físicas, como son: Distancia, tamaño, color, sonido, intensidad luminosa, etc. del entorno físico en que se encuentren y como respuesta a estos estímulos el robot ejercerá una acción mediante movimientos oportunos y/o programados.

La autonomía de un robot móvil se basa en el sistema de navegación automática; en estos sistemas de navegación se incluyen tareas de planificación, percepción y control. En el presente manuscrito solo consideramos a la percepción y al control, cuando se realiza un recorrido determinado; en donde se requieren valores mínimos de energía y tiempo, es decir, recorrer una trayectoria definida en el menor tiempo posible. Esto nos conduce a un problema de control óptimo en el cual se considera el tipo de trayectoria que se desea que el robot realice y el modelo matemático del mismo. La optimización en la teoría de control es una técnica que tiene como objetivo aumentar o mejorar el valor de una variable que en la practica puede ser

de naturaleza muy variada: temperatura, velocidad, posición, energía, etc.

En este trabajo se hace la síntesis de un algoritmo de control óptimo discreto para un robot móvil de los denominados seguidores de línea. En trabajos previos [2] se considera este mismo problema para un sistema reducido a tres ecuaciones que describen el movimiento del robot; aquí utilizaremos cinco ecuaciones obtenidas en el modelo matemático, también se muestran resultados que se obtienen usando la misma metodología (principio discreto del mínimo) pero empleamos una forma distinta de resolver al sistema acoplado final.

En la siguiente tabla se muestran los símbolos utilizados en el desarrollo de este trabajo.

Variable	**Valor**	**Descripción**
v_o	0,352	Velocidad [m/s]
a	0,40	Distancia entre ruedas [m]
b	0,05	Distancia del centro de masa al eje de las ruedas [m]
h	0,10	Distancia del eje de las ruedas al arreglo de sensores infrarrojos [m]
m	4,5	Masa del robot [kg]
ρ	0,08	Radio de las ruedas [m]
R	0,35	Radio de inercia del carro [m]
χ	0,01	Fricción viscosa
σ	0,009	Fuerza contraelectromotriz del motor
J	0,2868	Momento de inercia
ω		Velocidad angular
α		Ángulo que forma el eje de simetría del móvil x y el eje ξ
$F_{r,l}$		Fuerzas activas
$R_{r,l}$		Fuerzas reactivas
M		Torque de los motores [kg/m]
P		Punto donde se coloca un arreglo de sensores
T		Tiempo de muestreo [milisegundos]

Tabla 1. Parámetros del robot.

2. Planteamiento del problema

Considerando el siguiente proceso controlable (1)

$$\begin{aligned} \dot{y} &= f(y,u), \\ u(\cdot) \in U &= \{u : u(t) \in \Omega \subseteq \mathbb{R}^r\}, \end{aligned} \tag{1}$$

donde y es el vector $n-$ dimensional que contiene las coordenadas de estado del sistema, u es un vector $r-$dimensional que representa los controles de entrada. El control es una función vectorial continua a trozos, la cual en cada instante de tiempo, toma sus valores en un conjunto Ω convexo, cerrado y acotado. Suponemos que dado algún movimiento $y^d(t)$ y un control $u^d(t)$ deseado, satisfacen las siguientes ecuaciones

$$\begin{cases} \dot{y}^d = f(y^d(t), u^d(t)), \\ u(\cdot) \in U, \quad t \in [t_0, t_1). \end{cases} \tag{2}$$

Se tiene un arreglo de sensores que nos dan información sobre el movimiento que realiza el móvil. Después de procesar dicha información se pueden estimar las desviaciones que ocurren $x(t) = y(t) - y^d(t)$ para así poder ejercer el control sobre los actuadores.

Dadas las siguientes notaciones:

$\Delta u = u - u^d$ control adicional,
$x = y - y^d$ desviación respecto al movimiento deseado,
$\widetilde{z} = \varphi(y) - \varphi(y^d)$ vector de la información que se recibe sobre la desviación,

las ecuaciones diferenciales que gobiernan las desviaciones $x(t) = y(t) - y^d(t)$ para algún movimiento deseado $y(t) = y^d(t)$ y un control deseado $u(t) = u^d(t)$, pueden escribirse

$$\dot{x} = \widetilde{f}(x,t), \tag{3}$$

donde $\widetilde{f}(0,t) \equiv 0$ para $t \in [t_0, t_1)$ y $x(t_0) \neq 0$. Estas ecuaciones admiten una solución trivial correspondiente al movimiento deseado y^d del sistema.

Asumimos que $u^d(t) \in \Omega$ para $t \in [t_0, t_1)$. Se considera la estrategia de control $u = u^d + \Delta u(\widetilde{z}, t)$ donde Δu es el control adicional.

Obtenemos las ecuaciones lineales en desviaciones

$$\dot{x} = A(t)x + B(t)\Delta u, \tag{4}$$

con

$$A(t) = \frac{\partial f[y^d(t), u^d(t)]}{\partial y}, \qquad B(t) = \frac{\partial f[y^d(t), u^d(t)]}{\partial u}.$$

y el modelo lineal de medición

$$\widetilde{z} = H(t)x \ , \mathbf{det}H \neq 0, \tag{5}$$

donde

$$H(t) = \frac{\partial \varphi[y^d(t)]}{\partial y}.$$

Suponemos que las perturbaciones sobre el sistema, así como los errores de medición de los sensores son nulos. Casi sin excepción, hay presentes perturbaciones

iniciales $x(t_o) \neq 0$. Posiblemente se tenga la situación que para $\Delta u \equiv 0$ se cumple $x(t) \to 0$ si $t \to \infty$; pero es mucho más probable tener el caso que para $\Delta u \equiv 0$ se cumple $x(t) \nrightarrow 0$ si $t \to \infty$.

Aparece el problema de estabilización: Mediante el empleo de la información de la trayectoria deseada, determinar el control $\Delta u(\widetilde{z}, t)$ tal que las desviaciones actuales disminuyan asintóticamente. En este trabajo asumimos que contamos con información completa y exacta de todas las coordenadas, esto es, suponemos que tenemos sensores para medir todas las coordenadas, y que además no tienen error en sus lecturas.

3. Ecuaciones dinámicas del robot

Considerando la clase de robots móviles autónomos que consisten de tres ruedas, dos activas y una pasiva, con restricciones no-holonómicas, que aparecen como consecuencia de la hipótesis de no deslizamiento [1], [10]. Las velocidades del centro de las ruedas son denotadas como v_r y v_l, ver figura (1).

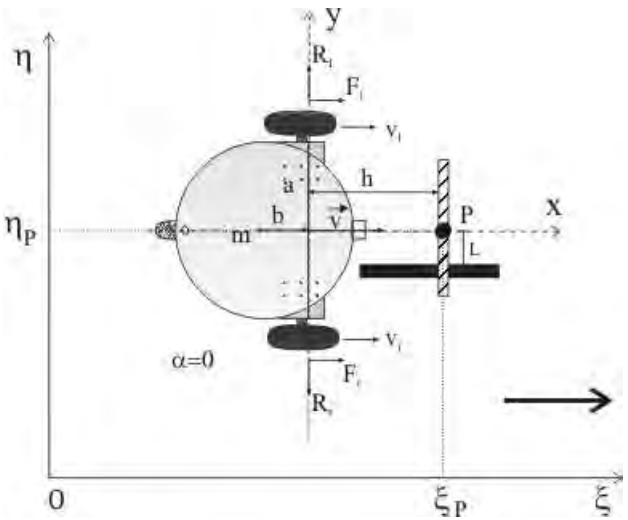


Figura 1. Sistema de coordenadas fijo (ξ, η) y sistema de referencia relativo al punto P.

La posición del robot con respecto al sistema de referencia inercial $\xi\eta$, están dadas por las siguientes relaciones

$$\begin{aligned} \dot{\alpha} &= w, \\ \dot{\xi} &= v\cos\alpha - h\omega\sin\alpha, \\ \dot{\eta} &= v\sin\alpha + h\omega\cos\alpha. \end{aligned} \tag{6}$$

Para obtener las ecuaciones dinámicas, se consideran principalmente fuerzas activas $\overline{F}_r$, $\overline{F}_l$. Se deducen las ecuaciones del sistema relativo al punto P_{xy} mediante los

teoremas principales de la mecánica, obteniendo las siguientes relaciones dinámicas

$$\begin{aligned} m(\dot{v} + b\omega^2) &= F_r + F_l, \\ J\dot{\omega} + mb\omega v &= (F_r - F_l)a. \end{aligned} \tag{7}$$

Sustituyendo las fuerzas activas por los torques de los motores y los voltajes que son aplicados a los mismos, hallamos:

$$M = F\rho, \tag{8}$$

donde M es el torque del motor, y el modelo más simple [8], [9] del motor es

$$M = \chi u - \sigma\dot{\varphi},$$

donde el miembro derecho $(\chi u - \sigma\dot{\varphi})$ es la suma de la fricción viscosa y la fuerza contraelectromotriz. Entonces para la rueda derecha

$$F_r = \frac{\chi u_r - \sigma\dot{\varphi}_r}{\rho}, \tag{9}$$

$$\varphi_r = \frac{v_r}{\rho} = \frac{v + wa}{\rho}.$$

Sustituyendo en la ecuación para cada rueda, y realizando operaciones en (7) se obtienen las ecuaciones dinámicas:

$$\begin{aligned} m\dot{v} + mbw^2 + \frac{2\sigma}{\rho^2}v &= \frac{\chi}{\rho}(u_r + u_l), \\ J\dot{w} + mbwv + \frac{2\sigma a}{\rho^2}w &= \frac{\chi}{\rho}(u_r - u_l). \end{aligned} \tag{10}$$

Finalmente se obtienen las ecuaciones que describen el movimiento del robot, cinco en total:

$$\begin{cases} \dot{\xi}_c = v\ \cos\alpha - b\ \omega\ \sin\alpha, \\ \dot{\eta}_c = v\ \sin\alpha + b\ \omega\ \cos\alpha, \\ \dot{\alpha} = \omega, \\ m\ \dot{v} = -mb\omega^2 - \ \frac{2\sigma}{\rho^2}\ v + \ \frac{\chi}{\rho}\ (u_r + u_l), \\ J\ \dot{\omega} = mb\omega v - \ \frac{2a^2\omega\sigma}{\rho^2} + \frac{\chi a}{\rho}\ (u_r - u_l), \end{cases} \tag{11}$$

donde $u_1 = (u_r + u_l)$ y $u_2 = (u_r - u_l)$.

4. Trayectorias Programadas y ecuaciones en desviaciones

Una trayectoria deseada se puede presentar considerando la configuración de líneas como se representa en la figura (2), se observa que el robot puede realizar su actividad por una línea recta horizontal o vertical en sus dos sentidos y por un semicírculo; en cuyo caso, se tienen ocho posibles configuraciones.

Ahora, las trayectorias estacionarias o programadas y^d, son determinadas a partir de la actividad o movimientos específicos del robot.

La tabla (2) muestra el conjunto de trayectorias deseadas, cuando el movimiento se realiza a lo largo de un segmento de línea paralela al eje $(0, \xi)$ o al eje $(0, \eta)$.

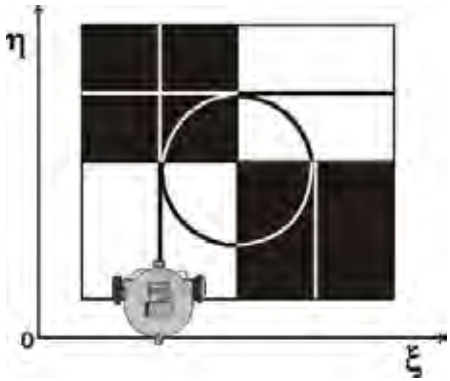


FIGURA 2. Tablero que muestra posibles trayectorias y combinaciones de las mismas.

Trayectoria	ξ^d	η^d	α^d	v^d	ω^d
1. Línea horizontal sentido positivo	$v_0 t + \xi_0$	0	0	v_0	0
2. Línea horizontal sentido negativo	$v_0 t + \xi_0$	0	π	v_0	0
3. Línea vertical sentido negativo	0	$v_0 t + \eta_0$	$\frac{\pi}{2}$	v_0	0
4. Línea vertical sentido positivo	0	$v_0 t + \eta_0$	$-\frac{\pi}{2}$	v_0	0

TABLA 2. Trayectorias programadas de las líneas horizontal y vertical

Si deseamos que el robot viaje en una línea recta horizontal o vertical, según la tabla (2), se debe suponer que no se tendrá movimiento angular ($\omega = 0$).

4.1. Ecuaciones lineales en desviaciones. Si $u^d(t)$ es una entrada nominal al sistema descrito por las ecuaciones (11) y y^d es una trayectoria nominal de dicho sistema, entonces el sistema de ecuaciones lineales en desviaciones se obtiene de la siguiente forma:

$$\begin{pmatrix} \xi^d \\ \eta^d \\ \alpha^d \\ v^d \\ \omega^d \end{pmatrix} = \begin{pmatrix} v_0 t + \xi_0 \\ 0 \\ 0 \\ v_0 \\ 0 \end{pmatrix} \tag{12}$$

considerando la línea horizontal en *sentido positivo* (*caso* 1 *de la tabla 2*) como trayectoria deseada [2], [4], vector columna (12).

Aplicamos el jacobiano a nuestro sistema de ecuaciones (11) y evaluamos, para obtener nuestras variables de estado:

$$A = \frac{\partial f(x,u)}{\partial x}|_{(x^0,u^0)}, \quad B = \frac{\partial f(x,u)}{\partial u}|_{(x^0,u^0)}.$$

Sustituyendo los valores numéricos de los parámetros del robot mostrados en la tabla 1, tenemos que las matrices A y B son:

$$A = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1{,}5 & 0 & -0{,}10 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -0{,}625 & 0 \\ 0 & 0 & 0 & 0 & -0{,}39225 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0{,}02777 & 0 \\ 0 & 0{,}17433 \end{pmatrix}.$$

Nuestro sistema lineal de *tiempo continuo* expresado en variables de estado queda de la forma:

$$\dot{x} = A\, x(t) + B\, u(t). \tag{13}$$

5. Diseño del algoritmo de control

El problema de diseño puede plantearse de la siguiente manera: Determinar el control óptimo $u^0(t)$ sobre $[0, N]$ tal que el índice de desempeño (14)

$$\mathcal{J} = \frac{1}{2}\langle x(t_f), \mathbf{S}x(t_f)\rangle + \frac{1}{2}\int_0^{t_f} [\langle x(t), \mathbf{Q}x(t)\rangle + \langle u(t), \mathbf{R}u(t)\rangle]dt, \tag{14}$$

sea mínimo, sujeto a la restricción de igualdad (4), donde $t_f = NT$, t_f es el tiempo final y T el periodo de muestreo, además

- **S** es una matriz simétrica (de dimensión $n \times n$) semidefinida positiva.
- **Q** es una matriz simétrica (de dimensión $n \times n$) semidefinida positiva.
- **R** es una matriz simétrica (de dimensión $p \times p$) definida positiva.

El término $\frac{1}{2}\langle x(t_f), \mathbf{S}x(t_f)\rangle$ que aparece en la ecuación (14) es el costo final del índice de desempeño, y se requiere como restricción final sobre la condición en el extremo sólo si $x(N)$ no es fijo.

En este trabajo se considera que $t_f = \infty$, entonces este primer término de (14) es cero y las matrices **Q** y **R** toman los siguientes valores,

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Para poder aplicar el principio del mínimo discreto [3], [5], procedemos a discretizar el sistema (13), y obtener así, al sistema:

$$x[(k+1)T] = \widetilde{A}(T)\, x(kT) + \widetilde{B}(T)\, u(kT), \tag{15}$$

donde $T = 0{,}1\ s$. Entonces, las matrices resultantes son:

$$\tilde{A} = \begin{pmatrix} 1 & 0 & 0 & -0{,}09693909 & 0 \\ 0 & 1 & -0{,}1500 & 0 & -0{,}01720930 \\ 0 & 0 & 1 & 0 & 0{,}09806414 \\ 0 & 0 & 0 & 0{,}93941306 & 0 \\ 0 & 0 & 0 & 0 & 0{,}96153433 \end{pmatrix},$$

y

$$\tilde{B} = \begin{pmatrix} -0{,}0001356 & 0 \\ 0 & -0{,}00012919 \\ 0 & 0{,}00086036 \\ 0{,}00268521 & 0 \\ 0 & 0{,}01709552 \end{pmatrix}.$$

Nuestro sistema ya discretizado es:

$$x[(k+1)] = \widetilde{A}\, x(k) + \widetilde{B}\, u(k). \tag{16}$$

Ahora, como ya hemos mencionado el objetivo de diseño es encontrar $u^o(k)$ tal que minimice al funcional $\mathcal{J}$ dado en (14), al discretizarlo queda como:

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2}x'(NT)\mathbf{S}x(NT) + \frac{1}{2}\sum_{k=0}^{N-1}[x'(kT)\hat{\mathbf{Q}}(T)x(kT) \\ &\quad + 2x'(kT)\mathbf{M}(T)u(kT) + u'(kT)\hat{\mathbf{R}}(T)u(kT)], \end{aligned} \tag{17}$$

haciendo los cálculos tenemos que,

$$\widehat{\mathbf{Q}} = \begin{pmatrix} 330 & 0 & 0 & 0 & 0 \\ 0 & 333 & 0 & 0 & 0 \\ 0 & 0 & 340 & 0 & 0 \\ 0 & 0 & 0 & 297 & 0 \\ 0 & 0 & 0 & 0 & 300 \end{pmatrix}, \ \widehat{\mathbf{R}} = 10{,}0 \times \begin{pmatrix} 0{,}2 & 0 \\ 0 & 5 \end{pmatrix}. \tag{18}$$

El siguiente paso para minimizar a la funcional $\mathcal{J}_N$, consiste en definir al Hamiltoniano, como:

$$\begin{aligned} \mathcal{H}(k) &= \mathcal{H}[x(k), p(k+1), u(k)] \\ &= \frac{1}{2}\,\langle x(k), \widehat{\mathbf{Q}}x(k)\rangle + \langle x(k), \mathbf{M}u(k)\rangle \\ &\quad + \frac{1}{2}\langle u(k), \widehat{\mathbf{R}}u(k)\rangle + \langle p(k+1), \widetilde{A}x(k) + \widetilde{B}u(k)\rangle, \end{aligned} \tag{19}$$

este funcional es la base del principio discreto del Mínimo [3], [7].

Por tanto, al resumir, la condición necesaria para que $\mathcal{J}_N$ tenga un extremo es:

$$\frac{\partial \mathcal{H}^o(k)}{\partial x^o(k)} = p^o(k) = \widehat{\mathbf{Q}}x^o(k) + \widetilde{A}'p^o(k+1) + \mathbf{M}u^o(k), \tag{20}$$

$$\frac{\partial \mathcal{H}^o(k)}{\partial p^o(k+1)} = x^o(k+1) = \widetilde{A}x^o(k) + \widetilde{B}u^o(k), \tag{21}$$

$$\frac{\partial \mathcal{H}^o(k)}{\partial u^o(k)} = \mathbf{M}'x^o(k) + \widehat{\mathbf{R}}u^o(k) + \widetilde{B}'p^o(k+1) = 0. \tag{22}$$

El control óptimo se obtiene de la ecuación (22),

$$u^o(k) = -\widehat{\mathbf{R}}^{-1}[\widetilde{B}'p^o(k+1) + \mathbf{M}'x^o(k)]. \tag{23}$$

Si sustituimos la expresión (23) en las ecuaciones (20) y (21), obtenemos las ecuaciones canónicas de estado

$$p^o(k) = (\widetilde{A}' - \mathbf{M}\widehat{\mathbf{R}}^{-1}\widetilde{B}')p^o(k+1) + (\widehat{\mathbf{Q}} - \mathbf{M}\widehat{\mathbf{R}}^{-1}\mathbf{M}')k^o(k), \tag{24}$$

$$x^o(k+1) = (\widetilde{A} - \widetilde{B}\widehat{\mathbf{R}}^{-1}\mathbf{M}')x^o(k) - \widetilde{B}\widehat{\mathbf{R}}^{-1}\widetilde{B}'p^o(k+1). \tag{25}$$

Estas expresiones representan $2n$ ecuaciones de diferencias que es necesario resolver con las condiciones de frontera $x(0)$ y $p^o(N) = \mathbf{S}x^o(N)$. Como se puede observar estas ecuaciones están acopladas en $x^o(k)$ y $p^o(k)$.

Como se plantea en trabajos anteriores [2], es complicado resolver de manera directa las ecuaciones canónicas de estado acopladas; sin embargo se demuestra en [4] que la solución es de la forma

$$p(k) = \mathbf{K}(k)x(k), \tag{26}$$

teniendo a $p(k)$ podemos resolver al sistema acoplado dado por (24) y (25), así también hallamos a $u^o(k)$ realizando las respectivas sustituciones; como podemos observar todo se reduce a encontrar a la matriz $\mathbf{K}(k)$ que se le conoce como *matriz de ganancias de Riccati*.

Entonces la solución a nuestro problema de diseño en tiempo infinito puede obtenerse al hacer que $k \to -\infty$. Cuando $N \to \infty$, la matriz de ganancia de Riccati $\mathbf{K}(K)$ se vuelve constante; esto es,

$$\lim_{k\to-\infty} \mathbf{K}(k) = \mathbf{K}.$$

La matriz de ganancias de Riccati queda,

$$\mathbf{K} = \widetilde{A}'\mathbf{K}\widetilde{A} + \widehat{\mathbf{Q}} - (\widetilde{A}'\mathbf{K}\widetilde{B} + \mathbf{M})(\widehat{\mathbf{R}} + \widetilde{B}'\mathbf{K}\widetilde{B})^{-1}(\widetilde{B}'\mathbf{K}\widetilde{A} + \mathbf{M}'), \tag{27}$$

que se conoce como *ecuación algebraica de Riccati* y que resolvemos con ayuda de MATLAB.

Ahora $u^o(k)$ podemos expresarlo como

$$u^o(k) = -(\widehat{R} + \widetilde{B}'\mathbf{K}\widetilde{B})^{-1}(\widetilde{B}'\mathbf{K}\widetilde{A} + \mathbf{M}')x^o(k), \tag{28}$$

donde el término $(\widehat{R} + \widetilde{B}'\mathbf{K}\widetilde{B})^{-1}(\widetilde{B}'\mathbf{K}\widetilde{A} + \mathbf{M}')$ se le conoce como matriz de retroalimentación y se denota por la letra $\mathbf{G}$ siendo esta matriz constante, quedando finalmente

$$u^o(k) = -\mathbf{G}x^o(k). \tag{29}$$

Ahora bien, decimos que para nuestro sistema discreto de la ecuación (16), si el índice de desempeño es $\mathcal{J}_N$, el control óptimo que minimiza a $\mathcal{J}_N$ es (29). Al sustituir $u^o(k)$ en (16) nos proporciona el sistema de lazo cerrado asintóticamente estable,

$$x^o(k+1) = [\widetilde{A} - \widetilde{B}\mathbf{G}]x^o(k), \tag{30}$$

el cual se le puede dar solución haciendo las iteraciones debidas para $k = 1, 2, \ldots, N$.

6. Simulación

Para poder medir el estado actual del robot móvil de forma experimental como se muestra en la figura (3), se pueden construir arreglos de sensores físicos que nos darían información completa sobre los movimientos del robot. Después de procesar dicha información ya se podrían estimar las desviaciones $x(t) = y(t) - y^d(t)$, donde el efecto del control se notará cuando las variables de estado $x(t) \to 0$; en nuestro caso suponemos que contamos con información completa al resolver el sistema retroalimentado (30) con condiciones iniciales distintas de cero, esto es, conocemos a todas las variables antes mencionadas. Al resolver (30), estamos resolviendo nuestro sistema en desviaciones (o diferencias) esto quiere decir que estaríamos alcanzando la trayectoria deseada $y^d(t)$, con los controles u_1 y u_2 actuando directamente sobre los motores de las ruedas.

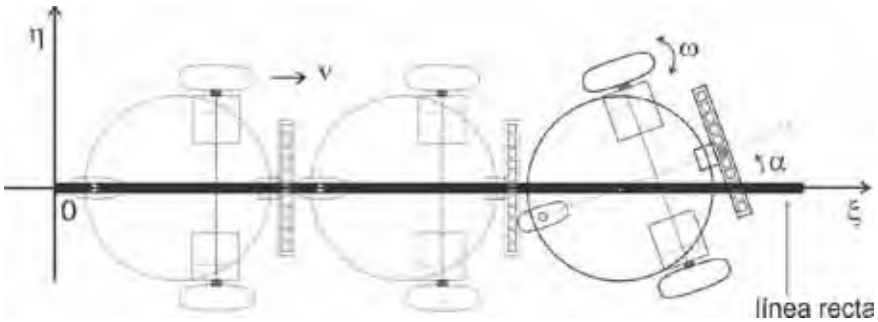


Figura 3. Movimientos al recorrer una linea horizontal en sentido positivo.

En la figura (4) se muestran las gráficas de las variables de estado y los controles; están separadas en las variables que corresponden al desplazamientos sobre la superficie del suelo (plano X, Y), los movimientos angulares (α, ω), la velocidad (ν) y los controles u_1 y u_2.

Podemos observar en las gráficas que la variable (ω) tiene un sobre tiro considerable, la interpretación física correspondiente a este fenómeno la podemos hallar cuando consideramos los movimientos que realiza el robot móvil al tratar de cubrir la tarea programada. Como el Robot parte de una posición inicial (distinta de la trayectoria deseada), este se moverá sobre la superficie del suelo buscando una línea marcada, entonces la velocidad angular del robot cambiará al realizar dicho movimiento. Su velocidad angular aumenta y una vez que encuentra dicha línea, este la seguirá. Por efecto del cambio en la velocidad angular tenemos como consecuencia también desplazamiento en X y Y, el ángulo (α) que se forma entre el eje de simetría del robot y el eje $(0, \xi)$ también cambia. Al igual que con (ω), se observan valores negativos grandes para (η) que podemos interpretar como un cabeceo que el robot haría hasta hallar la línea marcada sobre el eje $(0, \xi)$ y seguir sobre ella, ver figura (3).

En la figura (4) se muestran las gráficas de los controles u_1 y u_2 o energía necesaria para que el robot logre estabilizarse al cubrir la trayectoria programada, como describimos anteriormente al tratar de buscar la línea marcada emplea más

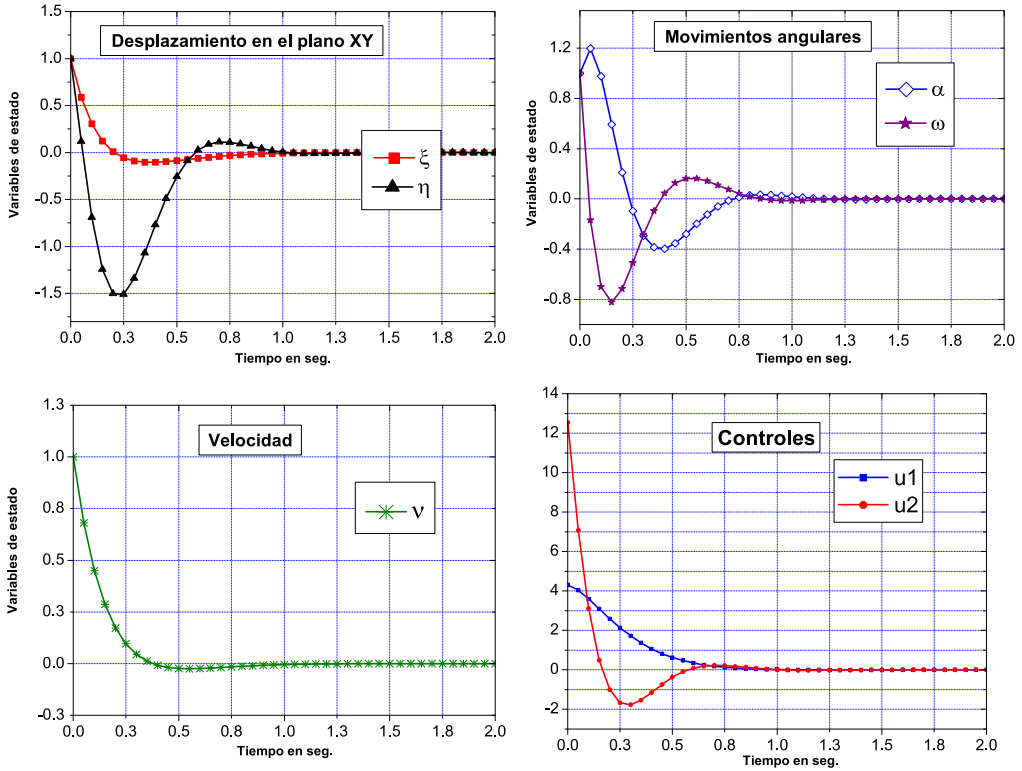


FIGURA 4. Gráficas del sistema.

energía en uno de los actuadores que en el otro (gira hacia un sentido); como observamos en esta gráfica, (u_2) toma valores grandes para alcanzar dicho tiempo de estabilización $(t = 1{,}2\ s)$. Las expresiones para $u_1 = u_r + u_l$ y $u_2 = u_r - u_l$, donde u_r y u_l son los controles para las ruedas izquierda y derecha.

7. CONCLUSIONES

La aplicación del principio del Mínimo para hallar el control óptimo que estabilizará a nuestro sistema discreto resultó menos complicada de lo se esperaba, resultando en resolver o hallar la matriz de ganancias de Riccati dicha matriz se halla con la ayuda de MATLAB, así como con MATLAB comprobamos cuestiones de controlabilidad y observabilidad del sistema y se dejan para trabajos posteriores la implementación física del algoritmo de control.

Una de las ventajas que tiene este tipo de controles es que pueden sintetizarse off-line (fuera de línea), desarrollándose previamente de forma analítica por métodos numéricos. La solución más apropiada puede consistir en aplicar una ganancia de control $\mathbf{K}$ constante, que corresponda a la solución en régimen permanente (estacionario), generalmente esta alternativa es la de implementación más simple pero requiere que el sistema cumpla con la condición necesaria de ser controlable y es suficiente que sea observable, como en nuestro caso. Independientemente de la síntesis

que se hace del algoritmo de control, podemos complementar el funcionamiento del algoritmo de control manipulando a las matrices de ponderación $\widehat{\mathbf{Q}}$ y $\widehat{\mathbf{R}}$ obviamente con el debido conocimiento del sistema a controlar.

El control sintetizado aquí, tiene una expresión relativamente simple en forma de una ecuación algebraica, cuando consideramos una trayectoria sencilla (linea recta horizontal). Otra ventaja que tienen este tipo de control es que el control es independiente del periodo de muestreo T y una extensión de este trabajo es considerar el planteamiento análogo para las trayectorias restantes.

En general creemos que este trabajo contribuye al uso de técnicas alternas para hallar solución a sistemas de control lineal y no lineal, en este caso para encontrar valores óptimos.

Referencias

[1] V.V. Alexandrov, L. Guerra, *Optimization and computer - aided testing of stabilization precision, Mathematical modeling of complex information processing systems*, Editado por la Benemérita Universidad Autónoma de Puebla en colaboración con la Moscow State University, 2001, pp. 49 - 60.
[2] 5^a Gran Semana Nacional de la Matemática (5GSNM), *Memorias, Diseño de un control óptimo digital para un robot móvil.*, Editorial de la BUAP, 2010, pp. 155 - 164.
[3] Aníbal Ollero Baturone, *Robótica Manipuladores y Robots Móviles*, Marcobombo, 2001.
[4] Benjamin C. Kuo, *Sistemas De Control Digital*, CECSA, 2000, pp. 609 - 723.
[5] Donald E. Kirk, *Optimal Control Theory An Introduction*, Prentice-Hall, 1970.
[6] Jean Jaques E. Slotine y Weiping Li., *Applied Nonlinear Control*, Pearson Education, Republic of China, 2004.
[7] Pablo Pedregal, *Introduction to Optimization*, Springer, 2004, pp. 195 - 236.
[8] Chi - Tsong Chen, *Analog and Digital Control System Design, Transfer-Function, State-Space, and Algebraic Methods*, State University of Newyork at Stony Brook, Saunders College Publishing, 2005, pp. 69 - 80.
[9] Richard C. Dorf, Robert H. Bishop, *Sistemas de control moderno, Décima Edición*, University of California, The University of Texas at Austin, Pearson Prentice Hall, 2005, pp. 58 - 60.
[10] Keith R. Symon, *Mecánica* 2da ed.,
Addison-Wesley, 1970.

FACULTAD DE CIENCIAS DE LA ELECTRÓNICA-BUAP
Av. San Claudio y 18 Sur, Ciudad Universitaria, Col. Jardines de San Manuel
CP. 72570, Puebla, Pue. México.

gelacio.salas@fce.buap.mx, mmorin@ece.buap.mx, jmgutierrez@ece.buap.mx

Capítulo 15

ALGUNAS CONSIDERACIONES SOBRE EL CONTROL DEL CAOS DETERMINISTA

EVODIO MUÑOZ AGUIRRE
FACULTAD DE MATEMÁTICAS - UNIVERSIDAD VERACRUZANA

Resumen. Al Principio se presenta una breve explicación sobre la Teoría del Control. Posteriormente, en base a un análisis del sistema de Lorenz y el modelo Logístico se discuten algunos elementos del Caos para sistemas dinámicos deterministas continuos y discretos respectivamente, con el fin de mostrar las principales definiciones del fenómeno del Caos que aparecen en la literatura. Por último, como una unión entre la Teoría del Control y la definición del Caos, se describe en qué consiste el Control del Caos, simplificando en una sola definición, describiendo sus principales objetivos; al mismo tiempo se exponen algunas aplicaciones en diferentes ramas de la ciencia de ésta importante área de la Matemática.

1. Introducción

El término control implica actuación y refleja el esfuerzo humano para intervenir en el medio que lo rodea para garantizar su supervivencia y una mejora en la calidad de vida. El área de la Teoría de Control tuvo sus orígenes en tiempos muy remotos, desde los antiguos babilonios y egipcios cuando intentaban regular la crecida de los ríos Eufrates y Nilo. Sin embargo, su auge se dio con la invención de la máquina de Vapor de James Watt en la revolución industrial. En la actualidad, se sabe que muchos de los problemas de control pueden analizarse por medio de un estudio minucioso de la ecuación de estado que modela el sistema en consideración, ya sea físico, económico, biológico, químico, etc. De hecho, "las variables de estado representan la mínima cantidad de información que nos resume todo el pasado dinámico del sistema y es todo lo que necesitamos conocer para poder predecir su evolución futura frente a cualquier señal que le apliquemos", [4].

Por otra parte, el orden lleva asociado un grado importante de predicción, al Caos le sucede lo contrario. Los sistemas lineales representan el orden, son predecibles y fáciles de manejar, de ahí nuestra tendencia a generalizarlos, y casi siempre nos va bien cuando proyectamos los datos del presente, para tratar de averiguar un comportamiento futuro. Pero existen sistemas que se resisten, pequeñas variaciones o incertidumbres en los datos iniciales desembocan en situaciones finales totalmente descontroladas e impredecibles. Se trata de los sistemas que se conocen como caóticos, [10].

La palabra Caos proviene del griego "Kaos" que significa "abertura, oquedad, insondable" y que en la Teogonía de Hesiodo designaba un espacio vacío infinito que existía antes de todas las cosas. Actualmente se usa también con el significado

de confusión y desorden.

Así, los sistemas caóticos se caracterizan por dos propiedades que los definen:

(1) Sensibilidad a la dependencia de las condiciones iniciales.
(2) Transitividad, que es equivalente a la afirmación de que existe una órbita caótica.

Además, se ha demostrado que el caos existe en una gran variedad de aplicaciones en la ciencia; como en la dinámica de los fluidos, en Química, en óptica no lineal, en circuitos electrónicos, en algunos modelos de población, en Meteorología, en oscilaciones fisiológicas, etc.[2].

Desde la década de los setenta del siglo pasado, se introdujo con fuerza el concepto del caos dentro de la ciencia. Los sistemas caóticos son una herramienta que tienen los científicos para modelar la incertidumbre que difiere de la probabilidad clásica. Los movimientos caóticos son modelados como las soluciones de Ecuaciones Diferenciales (o en Diferencias) con frecuencia y amplitud flotante, principalmente en los sistemas que presentan dificultad para construir la gráfica de las trayectorias, lo que se hace más difícil cuando se aumenta el número de variables.

Durante muchas décadas, la necesidad de científicos e ingenieros de modelar los comportamientos oscilatorios fue realizado para modelos lineales y no lineales con cíclos límites. Se creyó que ninguna otra clase de comportamiento oscilatorio podía observarse en sistemas genéricos deterministas. Es sorprendente que esto haya sido apoyado por resultados matemáticos. Por ejemplo, de acuerdo al criterio de Poincaré-Bendixon, el conjunto límite de cualquier sistema de segundo orden puede contener sólo trayectorias y equilibrio periódicos. Sin embargo, en el último cuarto del siglo pasado, esto se aclaró, descubriendo que existen fenómenos diferentes y muy complejos que pueden ocurrir, sobre todo con más frecuencia en sistemas de órdenes tercero y superiores que no se presentan en los sistemas de segundo orden para el caso de Ecuaciones Diferenciales Ordinarias. Una revolución fué inspirada por el artículo de E. Lorenz [3], quien estudió la dinámica de la turbulencia inducida por la convección térmica de fluidos en la atmósfera.

Ott, Grebogi y Yorke (OGY) publicaron en 1990 su primer artículo sobre el Control del Caos. Aunque algunos científicos ya habían realizado algunos cálculos semejantes antes que ellos, fué a partir de entonces que ha crecido el interés por el tema, especialmente por las aplicaciones: no siempre es deseable el comportamiento caótico. La estrategia OGY consiste en usar pequeñas perturbaciones de la órbita, de modo que ésta se estabilice en una de las órbitas periódicas inestables que existen en un atractor caótico. La perturbación pequeña significa que los parámetros, apenas variados, corresponden al caos que se va a controlar [5].

El artículo que se presenta está organizado como sigue.

En la primera sección se describen de manera escueta algunas características de la Teoría de control con el fin de mostrar el papel que desempeña en la tercera sección. Sin duda alguna el Caos determinista es muy importante en la ciencia actual, éste se describe mediante el desarrollo de algunos ejemplos clásicos en la

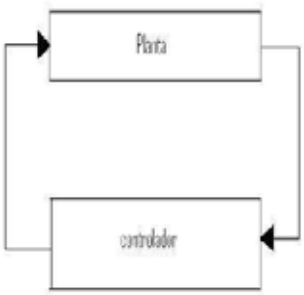


FIGURA 1. Sistema de control en retroalimentación.

tercera sección. El significado del Control del Caos y sus principales objetivos se describen en la cuarta sección, sin pasar por alto algunas aplicaciones en la ciencia.

2. CONTROL

2.1. **Significado de control.** La idea sobre el concepto de control se puede presentar familiarmente, por ejemplo cuando se toma el volante de un auto, o cuando se maneja la llave de la ducha de la regadera. El resultado puede ser bueno o malo, pero en cualquier caso el concepto de control tiene los siguientes componentes:

- El sistema a controlar (o planta) -Que es lo que se quiere controlar-
- El objetivo del control (o el comportamiento deseado del sistema a controlar)
- El conjunto de variables medibles (o salidas)
- El conjunto de variables de control (o entradas)

Otro importante componente que no se presenta en la etapa de formulación del problema de control es el controlador (o regulador), este componente siempre aparece cuando el problema ya ha sido resuelto. Por la solución del problema de control se entiende a la ley de control (o algoritmo de control). Cuando se encuentra la ley de control se utiliza el controlador para evaluar las entradas de control basadas en la medición de las salidas de la planta. Este proceso recibe un nombre especial, según se describe en la siguiente definición:

2.1. DEFINICIÓN. Al sistema compuesto de la planta y el controlador, se le llama **sistema de control en retroalimentación o sistema en lazo cerrado.**

La Figura 1 muestra esquemáticamente un sistema en lazo cerrado.

Para los casos discutidos al principio, se utilizan decisiones muy sencillas de los seres humanos para resolver esos simples problemas de control, por ejemplo:

"Si el agua está caliente, gira la llave a la izquierda".

"Si el agua está muy fría, gira la llave a la derecha".

Sin embargo, esto no se puede hacer en situaciones más generales y en casos más complicados.

2.2. **Objetivos del control.** Por lo general las plantas son modelos matemáticos descritos principalmente por ecuaciones diferenciales y algebraicas de la forma:

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \\ y(t) &= g(t, x(t), u(t)). \end{aligned}$$

A la primera ecuación se le conoce como **ecuación de estado**, que representa la dinámica de la evolución del estado del sistema, mientras que a la segunda se le conoce como **ecuación de salida**. x, y y u son vectores n-dimensionales s-dimensionales y m-dimensionales respectivamente.

De acuerdo a la tradición, se consideran dos clases de objetivos de control: Regulación y rastreo o conducción. A continuación se explican cada uno de éstos.

2.2. Definición. El objetivo de **regulación** se entiende como el objetivo de conducir el vector de las variables de estado del sistema de control a algún punto de equilibrio.

En términos matemáticos, este objetivo se describe:

$$\lim_{t \to \infty} x(t) = x_*,$$

donde x_* es un punto de equilibrio.

2.3. Definición. El objetivo de **rastreo** se entiende como el objetivo de conducir la trayectoria del sistema a una trayectoria deseada.

Es decir:

$$\lim_{t \to \infty} \mid x(t) - x_*(t) \mid = 0,$$

donde $x_*(t)$ es la trayectoria deseada.

Algunos otros objetivos son: Sincronización y la modificación del comportamiento asintótico del sistema. El primero tiene que ver con el acoplamiento de algunos sistemas, mientras que el segundo tiene que ver con la modificación de los sistemas que tienen comportamiento oscilatorio.

Un algoritmo que se utiliza con mucha frecuencia en la Teoría de Control es el principio de realimentación proporcional de la forma:

$$u(t) = K(x(t) - x_*).$$

La Teoría de Control es muy extensa y existen una gran cantidad de resultados que proveen diferentes controles, sin embargo, lo que se describió es suficiente para comprender el presente escrito. Si se desea, se puede consultar [7] para más información sobre teoría de Control.

3. Caos determinista

3.1. **Caos en Sistemas de Ecuaciones Diferenciales Ordinarias.** Un problema importante en Meteorología y en otras aplicaciones de la dinámica de fluidos hace referencia al movimiento de una capa de fluidos que está más caliente en la parte inferior que en la superior, como la atmósfera de la tierra. Si la diferencia de temperaturas vertical denotada por $\triangle T$ es pequeña, existe una variación lineal de la temperatura con la altitud, pero no se provoca ningún movimiento importante en la capa del fluido. Sin embargo, si $\triangle T$ es suficientemente grande, el aire más caliente tiende a subir, desplazando al aire más frío que está arriba y se produce un movimiento de convección estable. Si la diferencia en las temperaturas aumenta más, entonces el flujo de convección estable se rompe y sobreviene un movimiento más complejo y turbulento.

Cuando el meteorólogo estadounidense Edward Lorenz estudiaba este fenómeno, llegó al sistema de ecuaciones diferenciables ordinarias autónomo no lineal de tercer orden conocido como **Sistema de Lorenz** (Ver [3], [8] y [12]), el cual se puede escribir de la siguiente manera:

$$\begin{aligned} \dot{x} &= \sigma(x-y) \\ \dot{y} &= rx - y - xz \\ \dot{z} &= -bz + xy. \end{aligned}$$

La variable x está relacionada con la intensidad del movimiento del fluido, las variables y y z están relacionadas con las variaciones de la temperatura horizontal y vertical respectivamente. Los parámetros σ, r y b son positivos, σ y b dependen de las propiedades del material y geometrías de la capa del fluido. El parámetro r es proporcional a la diferencia de temperas $\triangle T$, y el problema es investigar de qué manera cambia la naturaleza de las ecuaciones de Lorenz con r. Para la atmósfera, $\sigma = 10$ y $b = 8/3$. Ver [3]

Para determinar los puntos críticos, es necesario resolver el siguiente sistema de ecuaciones algebraico:

$$\begin{aligned} \sigma(x-y) &= 0 \\ rx - y - xz &= 0 \\ -bz + xy &= 0. \end{aligned}$$

Realizando las operaciones respectivas para determinar la solución, se hayan los tres puntos críticos, a saber $P_1 = (0,0,0)$, $P_2 = (\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1)$, y $P_3 = (-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1)$. Si $r = 1$, claramente los tres puntos coinciden en el origen. Cuando r crece pasando por 1, el punto crítico P_1 se bifurca en el origen, surgiendo los puntos críticos P_2 y P_3. Cuando $r < 1$, los valores propios correspondientes cuando $\sigma = 10$ y $b = 8/3$ del sistema linealizado cerca del origen son $\lambda_1 = -8/3$, $\lambda_2 = -1052494$ y $\lambda_3 = -0.47506$, los tres tienen parte real negativa cuando $r = 1/2$. Por lo tanto, el sistema lineal tiene el mismo comportamiento que el sistema original

Si $r_1 \approx 1.3456$ y $r_2 \approx 24.737$, para $1 < r_1 < r_2$, P_2 y P_3 son asintóticamente estables y P_1 es inestable. Todas las soluciones cercanas tienden exponencialmente hacía

P_2 o hacia P_3. Para $r_1 < r < r_2$, los puntos críticos P_2 y P_3 son asintóticamente estables y P_1 es inestable. Todas las soluciones cercanas tienden hacia P_2 o P_3, la mayoría de ellas describen una espiral hacía el punto crítico. Para $r_2 < r$ los tres puntos críticos son inestables. La mayoría de las soluciones cercanas a P_2 o P_3 describen una espiral que se aleja del punto crítico.

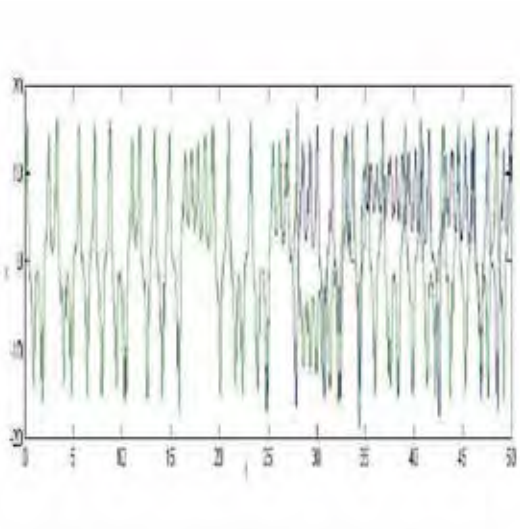


FIGURA 2. Dos soluciones del sistema de Lorenz con condiciones iniciales distintas pero muy cercanas. El eje horizontal corresponde a t, mientras que el vertical a x.

Ahora se considerarán soluciones para r algo mayores que r_2. En este caso P_1 tiene un valor propio positivo y cada uno de P_2 y P_3 tiene un par de valores propios complejos con parte real positiva. Una trayectoria puede aproximarse a cualquiera de los puntos críticos sólo sobre ciertos caminos muy restringidos. La menor desviación con respecto a estos caminos provoca que la trayectoria se aleje del punto crítico. Ya que ninguno de los puntos críticos es estable, podría esperarse que la mayor parte de las trayectorias tienden al infinito para t grande. Sin embargo, es posible demostrar que todas las soluciones permanecen acotadas cuando t tiende a infinito. Aún más, es posible demostrar que, en el infinito, todas las soluciones tienden a cierto conjunto límite de puntos cuyo volumen es cero. De hecho, esto no sólo es cierto para $r > r_2$ sino que también lo es para todos los valores positivos de r.

Las soluciones de las ecuaciones de Lorenz también son muy sensibles a las perturbaciones en las condiciones iniciales, ver la Figura 2. [1], [3]. En la figura se observa claramente el comportamiento de dos soluciones en $x(t)$ que comienzan con una condición inicial diferente. Éstas comienzan casi juntas, después de un tiempo se separan, de tal manera que esa separación es impredecible.

Fue esta propiedad la que en particular atrajo la atención de Lorenz en el estudio original de este sistema, y fue esta propiedad lo que hizo concluir que probablemente, no son posibles las predicciones climatológicas a largo plazo. Este obstáculo

a la predicción se conoce con el nombre de **Efecto Mariposa**, desde que este importante científico a nivel mundial dio una conferencia con el provocativo título: ¿Puede el batir de las alas de una mariposa en Brasil dar lugar a un tornado en Texas?

El conjunto que atrae a las soluciones en este caso, aunque de volumen cero, tiene una estructura un tanto complicada y se conoce como **atractor extraño**. En el caso que se está tratando se le llama atractor de Lorenz, ver la Figura 3.

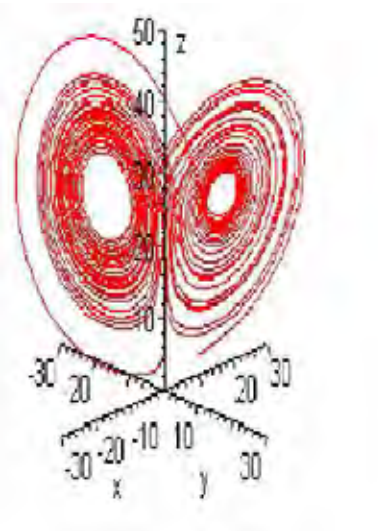


FIGURA 3. Atractor de Lorenz. Ejes horizontales x y y, eje vertical z.

El término **caótico** ya es de uso general para describir soluciones con este comportamiento. En general, se tiene la siguiente definición:

3.1. DEFINICIÓN. [3], [8]. Un **atractor** es un conjunto en el sistema espacio fase, para el cual todas las trayectorias convergen. Es decir, satisface:

- Es un conjunto invariante
- Atrae todas las trayectorias que comienzan suficientemente cerca de él.
- Es mínimo (no contiene atractores más pequeños)

En la dinámica caótica los objetos geométricos más importantes son los atractores extraños, conocidos también como atractores caóticos o atractores especiales que exhiben dependencia sensitiva a las condiciones iniciales. Estas son las principales características del **Caos**.

Es muy importante mencionar que en la presente sección y la que sigue, se trabaja con el fenómeno del caos "determinístico", es decir, no se trata de ruido, sino de soluciones aperiódicas de una o varias ecuaciones bien concretas y que determinan la evolución del sistema, principalmente gobernado por Ecuaciones Diferenciales Ordinarias y Ecuaciones en Diferencias.

3.2. **Caos Determinista mediante Ecuaciones en Diferencias.** Con el fin de describir el caos en los sistemas gobernados por Ecuaciones en Diferencias, se inicia con la presentación de modelos de poblaciones gobernados por Ecuaciones Diferenciales Ordinarias.

A continuación se tiene una ley de conservación muy sencilla que rige a una población, pero que a la vez es muy importante.

$$\dot{x} = nacimiento - muerte + migracion$$

Un modelo muy básico (Ver las referencias [3] y [8]) del crecimiento de poblaciones se debe a Thomas Malthus, quien desde 1887, sin considerar la migración, muestra que los términos de nacimiento y muerte son proporcionales a x:

$$\dot{x} = bx - cx,$$

lo que implica

$$x(t) = x_0 \exp{(b-d)t},$$

con $x(0) = x_0$.

Se ha comentado mucho en la literatura de Teoría de Ecuaciones diferenciales sobre este sistema. Sin embargo, este modelo no es muy realista porque el crecimiento de las soluciones es exponencial, aunque se utilizó para predecir de manera muy aproximada la población mundial en 1900. No obstante, para tiempos mucho más grandes, es difícil hacer predicciones. Es por esto que debía de hacerse un ajuste a este modelo. Fué así que Francois Verhulst propuso un modelo más sugestivo, conocido como **Modelo Logístico**:

$$\dot{x} = rx(1 - x/K),$$

donde r y K son constantes positivas, la constante K se le conoce como el nivel de saturación o como la capacidad cinegética del ambiente para la especie dada, mientras que a r se le denomina razón de crecimiento intrínsico.

La solución a esta ecuación está dada por:

$$x(t) = \frac{x_0 K e^{rt}}{K + x_0(e^{rt} - 1)}$$

donde $x(0) = x_0$. Se puede ver inmediatamente que el límite de esta función cuando t tiende a infinito es la constante K.

A mediados del siglo XX los ecologistas concluyeron que muchas especies tienen crecimiento regido por modelos en tiempo discreto, más que en modelos de tiempo continuo. En analogía a los modelos anteriores, propusieron modelos discretos en términos de ecuaciones en diferencias o mapeos del tipo:

$$x_{t+1} = f(x_t).$$

Así, en la década de los setenta, Robert May y Crafood Price propusieron un modelo de mapeo logístico para modelar el crecimiento de poblaciones:

$$x_{t+1} = rx_t(1 - x_t)$$

donde r es el parámetro Maltusiano que varía entre 0 y 4 y $x(0) = x_0$ pertenece a (0,1). Para ello, en el mapeo de May, la función genérica $f(x_t)$ obtiene una forma cuadrática específica:

$$x_t = rx_t(1 - x_t).$$

Para $r < 3$ hay un único punto fijo, para $3 < r < 3.4$ el punto fijo x_t oscila entre dos valores. Ver la Figura 4.

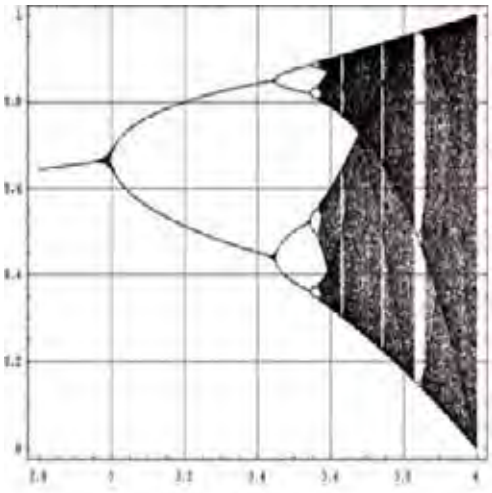


FIGURA 4. Diagrama de bifurcación del mapeo logístico. El eje horizontal representa al valor r, mientras que el vertical corresponde a x_t.

Cuando r crece, las bifurcaciones ocurren cuando el número de iteraciones se duplican. Estas bifurcaciones de doble periodo continúan a un punto límite en $r_{lim} = 3.569944$ que tiene periodo 2^∞ y la dinámica del mapeo se transforma en caótica. Este es un comportamiento que es de nuestro interés, puesto que presenta irregularidades impredecibles, es un ejemplo del comportamiento caótico; es decir, el sistema presenta el fenómeno del **Caos**.

4. Control del Caos

4.1. **Significado.** La Teoría de control es uno de los temas centrales en ciencia e ingeniería. A pesar del hecho de que ingenieros y matemáticos aplicados han estado tratando con problemas de control desde hace mucho tiempo, la idea de control del caos apenas fue introducida a principios de la década de los noventa en el siglo pasado con la publicación de un artículo de Ott, Grebogi y York ([11] donde utilizaron mapeos de poincaré para modificar comportamientos caóticos de sistemas mediante la aplicación de pequeñísimas perturbaciones, además de realimentación lineal con el mismo fin. Esta nueva idea atrajo la atención de los físicos que impulsaron una gran cantidad de trabajos sobre problemas de control, Ver [1].
El interés sobre los procesos caóticos radica en que se puede observar que su conjunto caótico sobre el cual la trayectoria del proceso vive, tiene inmerso un número de órbitas periódicas inestables conocidas. En adición, la trayectoria visita o accede a las vecindades de cada una de éstas. Algunas de estas trayectorias periódicas

pueden corresponder al funcionamiento deseado del sistema de acuerdo a algún criterio.
El segundo ingrediente es la realización que el caos, mientras signifique dependencia sensible a pequeños cambios para el estado ocurrente y a partir de este momento la representación del estado del sistema sea impredecible a largo plazo, además implique que el comportamiento del sistema puede ser alterado utilizando pequeñas perturbaciones. Entonces, la accesibilidad del sistema a varias órbitas periódicas diferentes, combinando con su sensibilidad a pequeñas perturbaciones, permite al control manipular el proceso caótico. Estas ideas estimularon al desarrollo de una gran variedad de nuevas técnicas del control del Caos [5].

En concreto, se tiene la siguiente definición, misma que se describe con base en las referencias: [1],[2], [5], [6], [7] y [11].

4.1. Definición. El término **Control del Caos**, se entiende principalmente como el área que estudia la relación entre los sistemas deterministas y el control, con el fin de analizar los métodos de control para los sistemas deterministas cuando presenten comportamiento caótico.

4.2. **Algunas aplicaciones sobre el control del Caos.** En esta sección se describen a grandes rasgos algunas aplicaciones del Control del Caos, se mencionan sólo algunas y de manera escueta, ya que no alcanzaría este espacio para enumerarlas todas y explicarlas a plenitud. Sin embargo, existen aplicaciones muy importantes en los sistemas de las comunicaciones, ya que tuvo una gran influencia en el avance de éstas. Si se está interesado en éstas y otras aplicaciones, se puede consultar [2] y [6].

- En **Mecánica** se cuenta con la clase más simple de sistemas mecánicos con características dinámicas complejas. El péndulo con fricción tiene un comportamiento caótico si se excita mediante una fuerza armónica de suficiente amplitud.
- La descripción y el control de turbulencias permanecieron como uno de los principales problemas de la **Física** en el siglo XX. Se puede dar la descripción en dimensión infinita del flujo turbulento como una solución a la ecuación de Navier-Stokes de una ecuación diferencial parcial. Si la dimensionalidad del flujo atractor en el espacio fase es relativamente pequeña, entonces el flujo turbulento se puede considerar caótico, y se pueden aplicar las técnicas del control del Caos. Un ejemplo de este tipo de flujos, es el flujo de un líquido entre dos cilindros concéntricos rotando.
- En **Química** existen oscilaciones caóticas en reacciones químicas. Se propuso una aplicación del algoritmo de control proporcional para la estabilización de la reacción química de Belousov-Zhabotinsky. El objetivo de control se formula como el intento de equilibrar el modo de la reacción. Por ejemplo, el comportamiento caótico se desea para procesar la combustión porque éste provoca aumento en la agitación de la mezcla aire-combustible y, concretamente, acelera el proceso.
- En **Biología y Ecología** se ha demostrado que las oscilaciones caóticas pueden ser transformadas a órbitas periódicas por medio de una acción periódica débil en un sistema de cuarto orden que describe la dinámica de un ecosistema acuático en donde conviven dos clases de micro algas y dos de zooplacton.

- En **Medicina** el estudio y tratamiento de arritmias cardíacas fueron una de las más apasionantes y primeras aplicaciones del control del caos. El diseño de la alta velocidad en el marcapasos parecía ser un nuevo enfoque radical en Cardiología, no obstante el comportamiento del corazón humano se tornó más complicado aún. Algunos modelos y métodos para controlar el proceso cardiaco de la actividad cardíaca han sido propuestos pr algunos científicos, uno de ellos es el que se basa en la realimentación lineal de tiempo en retraso de un paso para suprimir el ritmo patológico de periodo dos.
- Se ha demostrado que la dinámica de muchos modelos en **Economía** obedecen a sistemas no lineales, lo que provoca que suelan tener comportamientos caóticos. El problema de controlar a tales sistemas sólo es posible en niveles microeconómicos. En este caso es razonable pensar como un objetivo del control a la supresión del caos, lo cual conduce a que sean más predecibles los ciclos en los negocios. Un algoritmo de control adaptativo condujo satisfactoriamente a la supresión del caos descrito por dos compañías compitiendo con dos diferente estrategias de inversión (en el mismo sector del mercado).

CONCLUSIÓN

Mediante algunos ejemplos, se ha mostrado la fusión de la Teoría del Control con la correspondiente del Caos en una nueva área de las Matemáticas, El Control del Caos. Asimismo, se explica a grandes rasgos la importancia que tiene esta área en las aplicaciones de la ciencia en general, con base en algunos ejemplos conocidos de la ciencia en general.

Referencias

[1] B. R. Andrievskii and A.L. Fradov; *Control of Chaos: Methods and Applications I.* Automatic and Remote Control, Vol 64, N0. 5, pp 673-713, 2003.
[2] B. R. Andrievskii and A.L. Fradov; *Control of Chaos: Methods and Applications II.* Automatic and Remote Control, Vol 65, N0. 3, pp 503-533, 2004.
[3] W. E. Boyce. R. C. DiPrima; *Ecuaciones Diferenciales y Problemas con Valores en la Frontera.* Limusa Weley. México. 2005.
[4] S. Domínguez, P. Campoy, J. M. Sebastian y A. Jiménez; *Control en el espacio de estado.* Prentice Hall. España. 2006.
[5] J. I. Causabon; *Control del Caos usando la estrategia OGY* Ciencia al día, 2, Vol. 4. 2000.
[6] A.L. Fradov, R. S. Evans and B. R.; *Control of Chaos: Methods and Applications in Mechanics.* Phl. Trans. R. Soc.,pp 2279-2307. 2006.
[7] L. Fradov and Y. Pogromsky; *Introduction to Control of Oscillations and Chaos.* Uto Print, Singapore,1998.
[8] M. W. Hirsh, S. Smale and R. L. Devaney; *Differential Equations, Dynamical Systems, and an Introduction to Chaos.* Academic Press. USA. 2004.
[9] W. Just, H. Benns and E. Scholl; *Control Of Chaos By Time-Delay Feedback: A Survey of Theorical and Experimental Aspects.* Adv. in Solid State Phis. Vol. 43, pp 589-603, 2003.
[10] Z. Li; *Fuzzy Chaotic Systems* StudFuzz 199, pp 31-52. 2006.
[11] K. Pyragas; *Delayed Feedback Control of Chaos.* Phil. Trans. R. soc. A., 364, pp 2309-2334. 2006.
[12] Y. Zeng, and S. Singh; *Adaptive Control of Chaos in Lorenz System.* Dynamical and Control, 7, pp 143-154, 1997.

FACULTAD DE MATEMÁTICAS-UNIVERSIDAD VERACRUZANA
Lomas del estadio s/n, Zona Universitaria. C.P. 91000, Xalapa Veracruz México.
evmunoz@uv.mx

Enseñanza, Historia y Divulgación de las Matemáticas

Capítulo 16

ANTECEDENTES SOBRE LAS TEORÍAS PARACONSISTENTES

EDUARDO ARIZA VELÁZQUEZ
PEDRO GARCÍA JUÁREZ
ROSA GARCÍA TAMAYO
FACULTAD DE CIENCIAS DE LA COMPUTACIÓN - BUAP

Resumen. Presentamos una reseña sobre la polémica que se gestó a partir de las teorías que atentaron contra los pricipios aristotélicos, principalmente las teorías contradictorias; así como el proceso que llevó a las teorías paraconsistentes propuestas por Newton da Costa.

1. Introducción

Aunque a finales del siglo XIX la Geometría de Lobachevsky ya se daba por aceptada, en su momento fue motivo de polémica el trastocar los postulados de la Geometría Euclidiana, ampliamente respaldada por la filosofía de Kant [11]. Además, la comunidad matemática de entonces no estaba preparada para cambios tan drásticos. Esto inhibió el desarrollo de las geometrías no euclidianas, que en la actualidad son motivo de estudio y de múltiples aplicaciones [10].

Mucho más costoso resultó para quienes tuvieron el atrevimiento de poner en duda los principios de la filosofía aristotélica, que durante siglos ha dominado al menos en el mundo occidental y, por supuesto, proporcionado los fundamentos del razonamiento matemático. Retomaremos el punto de vista de Ludwig Wittgenstein, quien encabezó dicha polémica y cuya reputación resultó gravemente afectada. Sin embargo, como en la geometría no euclidiana, el tiempo es quien otorga la razón.

2. La geometría no euclidiana

Dos aspectos son motivo de inspiración para el estudio de lógicas no aristotélicas: las geometrías no euclidianas y la teoría de conjuntos de Cantor. Empezamos por dar una pequeña reseña sobre las geometrías no euclidianas.

En 1792, a la edad de 15 años, Gauss se dedicó a deducir las consecuencias de una geometría en la que pueden dibujarse más de una línea que pase por un punto dado y que cada una sea paralela a una recta dada. Compartió su teoría con su amigo, el matemático Farkas Bolyai, éste a su vez le pidió a su hijo, János, que no perdiera una sola hora en el problema del quinto postulado. Sin embargo, en 1823 János Bolyai escribió a su padre diciendo "*He descubierto cosas tan maravillosas que estoy asombrado... de la nada he creado un extraño y nuevo mundo*" [10].

Por su parte Lobachevsky publicó una obra sobre geometría no euclidiana en 1829 y ni Bolyai ni Gauss sabían del trabajo, principalmente porque fue publicado en el

idioma ruso en una publicación universitaria local. En dicho trabajo Lobachevsky remplaza el quinto postulado de Euclides por su postulado de las paralelas [10]:

Dada una línea recta, existen dos líneas paralelas a ésta y que pasan por un punto fijo que no está en la línea dada.

Gauss mantuvo en secreto su trabajo: hasta cierto punto tenía razón, pues el intento de Lobachevsky para llegar a un audiencia más amplia falló, cuando su artículo fue rechazado por Ostrogradski. La comunidad matemática no estaba lista para aceptar ideas tan revolucionarias. En esa época el pensamiento estaba dominado por Kant, quien afirmaba [10]:

La geometría euclidiana es la inevitable necesidad de pensamiento.

Años después, en 1854, Riemann dio una conferencia en la cual reformuló por completo el concepto de geometría. Discutió sobre una geometría "esférica" en la cual las rectas paralelas son imposibles. Esta plática fue publicada hasta 1868, dos años después de su muerte.

En el último tercio del siglo XIX la autoridad de Kant y la condición privilegiada que su filosofía atribuye a la geometría clásica, solian invocarse para combatir a los sistemas geométricos propuestos por Gauss-Bolyai, Lobachevsky y Riemann. Pero las nuevas geometrías acabaron por imponerse como matemáticamente legítimas.

Hoy sabemos que sustituir un axioma por su negación, en un sistema de axiomas independientes, no genera contradicción interna en el nuevo sistema [6]. De hecho, si el primer sistema es consistente, entonces el nuevo sistema sigue siendo consistente. Por lo tanto, se pone de manifiesto que ambas geometrías se encuentran en un mismo nivel de consistencia.

3. Lógica Clásica

Las geometrías no euclidianas se basan en transformar, o sustituir, uno o más axiomas propios de la geometría clásica. Nunca estuvo en duda el razonamiento matemático, fundamentado por la lógica de Aristóteles (384-322 a.C.), que dominó en el mundo occidental hasta ya iniciado el siglo XX y que esencialmente es lo que conocemos por lógica clásica. Sus principios son:

- Principio de no contradicción: No es posible que algo sea y no sea, al mismo tiempo y bajo la misma consideración.
- Principio de identidad: Algo es igual a sí mismo.
- Principio del tercero excluido: Cualquier cosa es o no es.

En el devenir de la lógica, es necesario nombrar a Friedrich Ludwig Gottlob Frege (1848-1925) que en 1879 llevó a cabo la transformación más radical de la lógica desde la época de Aristóteles, ofreciéndo por primera vez en la historia un cálculo deductivo para las lógicas de primer y segundo orden. Por ello es considerado el fundador de la Lógica Matemática, o Lógica Simbólica.

4. El programa de Hilbert

En 1874 apareció el primer trabajo de Georg Cantor (1845-1918) sobre teoría de conjuntos. Prácticamente de inmediato su teoría fue puesta en duda con la aparición de paradojas. Esto provocó una gran crisis que hizo tambalear el razonamiento matemático.

A pesar de que la teoría de conjuntos resistió la crisis, la inquietud sobre los fundamentos del razonamiento matemático creció. El problema fue tratado inicialmente de forma ortodoxa, encabezado por Hilbert. Sin embargo dicha inquietud resultó ser, en particular, un detonante para desviar la atención hacia teorías no aristotélicas. Pero nuevamente, el mundo matemático no estaba preparado para ello.

4.1. **La conferencia en Suiza.** En 1917, Hilbert dio una conferencia en Zúrich, que posteriormente se publicó con el título "Pensamiento Axiomático", en dicha plática señaló cómo diversos axiomas jugaban papeles significativos, mostrando cómo determinar la consistencia de los axiomas reduciendo el problema a un sistema más simple. Hilbert requería resolver varias cuestiones epistemológicas: el problema de la resolubilidad de cualquier cuestión matemática, la verificabilidad de cualquier resultado, un criterio para la simplicidad de las demostraciones, la relación entre contenido y formalismo en la matemática y la lógica y, finalmente, la decibilidad en un número finito de pasos (¿puede demostrarse que un enunciado dado es demostrable/refutable, o que es independiente, en una teoría dada?). La conferencia concluyó con una ilustración de la naturaleza e importancia del problema de la decisión.

Lo que empezó como un conjunto de recetas para problemas específicos, se convirtió en todo un programa para la matemática[1]. Hilbert se sentía capaz de especificar las condiciones bajo las que cualquiera, con buena voluntad, estaría de acuerdo en que se había llevado a cabo una tarea matemática. Las apuestas sobre si Hilbert lo lograría, se elevaron [7].

4.2. **El desafío de Brouwer.** La filosofía del Intuicionismo de Brouwer se basa en afirmar que: la única manera de demostrar que un conjunto tiene cierto elemento, es construyendo explícitamente un elemento semejante[2]. No basta con demostrar que: la hipótesis de que el conjunto no tiene al elemento lleva a una contradicción. En particular, Brower creía que la mente humana estaba sorprendentemente limitada en su capacidad para trabajar con conjuntos infinitos donde, según él, el principio (aristotélico) del tercero excluido ya no era aplicable [7], [9].

Para Hilbert, el Intuicionismo era motivo de excomunión, luego de una serie de existosas conferencias, por parte de Brouwer en 1927, asustado por las consecuencias que visualizaba para la matemática, Hilbert decidió proteger la revista "Mathematische Annalen", que veía como la más importante de toda la matemática y de la que él mismo era el editor en jefe. En 1928 Brouwer es despedido como editor, cargo que tenía desde 1915. Después Brouwer publicó poco sobre intuicionismo [7].

[1]Para una versión más amplia sobre el programa de Hilbert ver [1], [9].

[2]Por su carácter constructivo, la teoría intuicionista es motivo de estudio en las ciencias de la computación.

Finalmente, la filosofía de la matemática y la teoría de la demostración de Hilbert se plantan como el sistema de ideas dominante. Sin embargo, el verdadero ataque llegó desde sus partidarios.

5. El teorema de Gödel

En 1931, Kurt Gödel (1906-1978) publicó su trabajo, poniendo al descubierto dos limitaciones insuperables inherentes al método axiomático de Hilbert. De forma sucinta dice [6]:

(1) En cualquier sistema axiomático, siempre es posible construir un teorema que solamente sea demostrable cuando su negación también es demostrable. Esto es que, todo sistema de axiomas es incompleto y, más aun, que es incompletable.

(2) La consistencia de un sistema de axiomas solamente puede ser demostrada refiriéndolo a otro sistema superior, y la de este último refiriéndolo a otro sistema más superior; y sucesivamente, de manera interminable, siempre fundamentando cada sistema en otro superior.

Gödel demuestra que cualquier sistema que permita definir los números naturales es necesariamente incompleto, en el siguiente sentido: contiene afirmaciones que ni se pueden demostrar ni refutar (incompletez). El teorema dio pie a diversas interpretaciones erróneas, por ejemplo: no implica que todo sistema sea incompleto. A pesar de que la geometría euclidiana es completa, hay construcciones imposibles como la trisección del ángulo.

El teorema de incompletitud fue presentado en Königsberg, ciudad natal de Hilbert. Acabada la exposición, Gödel dijo: *deseo indicar expresamente que este teorema no contradice el punto de vista formalista de Hilbert, pues este punto de vista presupone sólo la existencia de una demostración de consistencia en la que no se utiliza otra cosa que medios finitos de demostración, y es concebible que existan demostraciones finitas que no pueden expresarse en el formalismo de* P.

Para muchos, el ataque al trabajo de Hilbert era evidente, a pesar de que el mismo Gödel afirmara lo contrario. Las reacciones iniciaron con una serie de interpretaciones tanto a favor como en contra, seguido de un escrutinio minucioso a la demostración. En 1939 se publicó el libro "*Los fundamentos de las Matemáticas*" de Hilbert y Bernays, el cual calmó la seria oposición al trabajo de Gödel.

6. La filosofía de Wittgenstein

Durante su estancia en Manchester, Ludwig Wittgenstein (1889-1951) descubrió su interés tanto en la matemática pura como en sus fundamentos, inspirado por *Principles of Mathematics*, de Russell. Por consejo de Frege se dirige a estudiar lógica con Russell en Cambridge. En 1923 mantuvo una serie de conversaciones con F. P. Ramsey. Posteriormente, el profesor Moritz Schlick, fundador del Círculo de Viena, así como Friedrich Waismann, uno de sus miembros, entraron en contacto con Wittgenstein a raíz de la publicación del Tractatus [15].

A continuación presentamos textualmente algunas observaciones y puntos de vista de Wittgenstein sobre los fundamentos de la matemática tomados de las pláticas y conversaciones que sostiene con Schlick y Waismann, entre 1929 y 1931, taquigrafiadas por Waismann [14][3].

> *Luego de leer un trabajo de Hilbert sobre consistencia, parece que la cuestión ha sido planteada equívocamente. Surge una pregunta:*
>
> *¿Pueden las matemáticas contener contradicciones?*
>
> *Es decir, si surgieran contradicciones en las reglas del juego de la matemática, lo más fácil del mundo sería remediarlo. Lo único que tenemos que hacer es una nueva estipulación que cubra el caso en el que las reglas entran en conflicto y asunto arreglado. Lo que Hilbert hace es matemática y no metamatemática. Es una vez más un cálculo, tan bueno como cualquier otro.*
>
> *Debe decirse que las antinomias nada tienen que ver con las contradicciones de la matemática, que no hay aquí ninguna contradicción. Porque las antinomias no surgen en el cálculo, sino en el lenguaje usual y ello porque se usaban las palabras de manera ambigua. De manera que la resolución de las antinomias consiste en reemplazar el modo vago de hablar por uno preciso. Las antinomias se disuelven gracias a un análisis, no por medio de una prueba.*
>
> *Por medio de autorizaciones y prohibiciones se puede siempre determinar un juego, pero nunca el juego. Lo que Hilbert quiere mostrar con su prueba es que los axiomas de la aritmética tienen las propiedades del juego, y eso es imposible. Es como si quisiera probar que la contradicción es inadmisible. Cuando Hilbert dice: "$0 \neq 0$ no debe aparecer como una fórmula demostrable", él determina un cálculo por medio de permisos y prohibiciones.*
>
> *No puede haber ninguna cuestión como la de si eventualmente se caerá en una contradicción por proceder en concordancia con las reglas. Yo creo que éste es el punto esencial, del cual depende todo en la cuestión de la consistencia.*
>
> *Si veo esto candidamente, lo que nos tiene que llamar la atención es que los matemáticos siempre tengan miedo sólo de una cosa, que es para ellos como una especie de pesadilla: la contradicción.*
>
> *Creo que dar una prueba de consistencia sólo puede significar una cosa: examinar cuidadosamente las reglas. No hay otra cosa que se pueda hacer.*
>
> *Pero ¿qué pasa si examino sistemáticamente las reglas del juego? A partir del momento en que me muevo dentro de un sistema, tengo una vez más un cálculo; pero de nuevo se plantea la cuestión de la*

[3]Publicado post mortem.

consistencia. De hecho, no puedo hacer otra cosa que inspeccionar una regla tras otra.

¿Qué significaría que un cálculo produjera el resultado $0 \neq 0$*?*

Claro está, no estaríamos lidiando con una aritmética modificada, sino con una aritmética completamente diferente, con una aritmética que no tiene el más ligero parecido con la aritmética cardinal. Que pueda aplicar un cálculo así es otra cuestión. Desde este punto de vista, una fórmula como $0 \neq 0$ *es absolutamente inentendible, porque obviamente significaría que* 0 *no es sustituible por* 0*.*

¡Qué interesante sería si se produjera una contradicción! En verdad, yo predigo desde ahora que:

> *Habrá investigaciones matemáticas sobre cálculos que contengan contradicciones y hasta se estará orgulloso de haberse emancipado de la consistencia.*

¿Qué pasaría si quisiera aplicar un cálculo así? ¿Lo aplicaría con mala conciencia mientras no hubiera probado que no es inconsistente?. "Si pude aplicar un cálculo, entonces simplemente pude aplicarlo".

6.1. **Septiembre de 1931.** *. . . No tiene ningún sentido hablar de una inconsistencia oculta. Porque ¿qué sería una inconsistencia oculta? Se puede decir, por ejemplo: la divisibilidad del número 357567 por 7 está oculta, esto es, mientras no haya aplicado el criterio, de la división. Para transformar en abierta la divisibilidad oculta, lo único que debo hacer es aplicar el criterio. Todo este discurso acerca de la inconsistencia oculta no tiene ningún sentido, y el peligro del cual hablan los matemáticos es pura fantasía.*

Si alguién describiera la introducción de los números irracionales afirmando que descubrió que entre los puntos racionales de una línea hay todavía más puntos, nosotros le responderíamos: "es que no descubriste nuevos puntos entre los anteriores sino que construíste nuevos puntos. Por consiguiente, lo que tienes delante de ti es un nuevo cálculo". Esto es lo que se le debe decir a Hilbert cuando afirma que es un descubrimiento el que las matemáticas son consistentes. En realidad Hilbert no comprueba nada, sólo estipula.

Por último, consideremos el caso en que no tenemos ningún método para determinar si surge una contradicción. ¿Surge ahora aquí algún peligro?, ¡ni el más mínimo!, ¿de qué deberíamos estar asustados?, ¿de una contradicción? Pero una contradicción es dada sólo a través del método para encontrarla. Mientras no surja no nos incumbe. Podemos estar totalmente tranquilos y hacer más cálculos.

¿Acaso porque se encontrara una contradicción en la matemática cesaría súbitamente todo lo que han logrado los matemáticos a lo largo de centurias?

¿Diríamos que no se trataba de cálculos? En absoluto. Si surgiera una contradicción, sencillamente lidiariamos con ella. Pero no necesitamos tener ahora ninguna preocupación.

7. Las reacciones en contra de Wittgenstein

En 1956 se editó la primera versión de: *Observaciones sobre los fundamentos de la matemática*. La publicación produjo una fuerte reacción adversa tanto entre filósofos como entre matemáticos. Las primeras interpretaciones encontraron que las referencias a los resultados de Gödel no sólo constituían una especie de profanación, de aquello que los matemáticos ya habían canonizado, sino que también era fácil poner en evidencia una clara ingenuidad de parte del autor.

Los lógicos considerados los más grandes de todos los tiempos son: Aristóteles (384 a.C.-322 a.C.), Frege (1848-1925), Gödel (1906-1978) y Tarski (1902-1983). Sólo Frege tuvo contacto directo con Wittgenstein, como ya mencionamos, él mismo lo recomendo con Russell. Por su parte Gödel y Tarski mostraron una fuerte reacción hacia el punto de vista de Wittgenstein.

7.1. **Wittgenstein y Gödel.** En 1972, en respuesta a una petición de Karl Menger, Gödel escribió:

> *Por lo que se refiere a mi teorema acerca de las proposiciones indecidibles, se desprende con toda claridad de los pasajes que usted cita que Wittgenstein no lo entendió (o fingió no entenderlo). Lo interpreta como si fuera una especie de paradoja lógica, cuando de hecho es precisamente lo contrario, a saber, un teorema matemático* [13], [12].

Otros comentarios de Gödel recogidos por su amigo Hao Wang [12], [13] también en respuesta a Karl Menger, son:

- *El pasaje entero que usted cita me parece, dicho sea de paso, un sinsentido. Repárese, por ejemplo, en:*

  "El supersticioso temor de los matemáticos a las contradicciones".

- *¿Ha perdido Wittgenstein su cabeza? ¿Opina él esto seriamente? Lo que él dice acerca del conjunto de todos los números cardinales revela una visión perfectamente ingenua. Toma posición cuando él realmente no tiene nada que hacer aquí. Por ejemplo:*

  "*Usted no puede derivar cualquier cosa a partir de una contradicción*".

  Es sorprendente que Turing consiguiera sacar algo de sus discusiones con alguien como Wittgenstein.

7.2. **Tarski y Wittgenstein.** Alfred Tarski contribuyó a la madurez de la lógica estándar, de primer orden, fundando una metodología conjuntista de las teorías deductivas. Es de Tarski una de las primeras demostraciones del teorema de Deducción. Sus métodos semánticos, que culminaron en la teoría de modelos (desarrollada en los años 50 y 60), transformaron radicalmente la metamatemática consolidándola como ciencia estricta. En su artículo de 1936 "*On the Concept of Logical consecuence*" defendió que la explicación de la consecuencia lógica depende de la teoría semántica de la verdad.

Aunque poco, y poco documentado, rescatamos el punto de vista que Tarski tenía sobre Wittgenstein, por tratarse de uno de los más grandes en la historia de la lógica.

En el libro *Alfred Tarski Life and Logic* [2], se comenta uno de lo últimos eventos a los que Tarski asistió, en Calgary:

> Después de enterarse de que el interés hacía el trabajo Wittgenstein era grande, al día siguiente le dio a Verena Huber Dyson[4] una seria conferencia moral acerca de su deber de proteger a los estudiantes de malas influencias, como la de Wittgenstein (pág. 373).

La postura de Wittgenstein sobre temas asociados con los problemas de incompletitud y consistencia se hallaba definitivamente desacreditada entre personajes que contaban con un reputación intachable. Sin embargo, al final, el tiempo le ha dado la razón a Wittgenstein.

8. Teorías paraconsistentes

De forma aislada surgió la inquietud de estudiar teorías que rechazaran total o parcialmente el principio de no contradicción. Jan Łukasiewicz y Nicolai Vasiliev son considerados los pioneros en el estudio de las teorías paraconsistentes [4].

1. En su trabajo "On the Principle of Contradiction in Aristotle", en 1910, Łukasiewicz presenta tres propuestas del principio de no contradicción, en cada una argumenta que tal principio no puede ser tan básico como generalmente se pensaba, creando un precedente para el inicio de las lógicas no clásicas.

2. Inspirado por los métodos de Lobachevsky (que inicalmente se llamó geometría imaginaria), Vasiliev es considerado un precursor debido a sus ideas sobre lógicas imaginarias en 1911. Además tenía la creencia de que, como en la geometría de Lobachevsky, sus sistemas también podrían presentar alguna interpretación clásica.

[4]Profersora emérita del Departamento de Filosofía de la Universidad de Calgary.

8.1. **El trabajo de Jaśkowski.** El inicio formal al estudio de lógicas contradictorias no triviales se atribuye a Stanisław Jaśkowski. En 1948 propone el siguiente problema [8]:

Trabajar con sistemas que incluyan contradicciones y todavía contar con deducciones razonables.

Él mismo da respuesta: con la influencia de Łukasiewicz, propone un cálculo proposicional en el que la presencia de contradicciones no implica la trivialización del sistema, dando paso a la primera formulación de teorías no triviales presentando inconsistencia. A pesar de que su respuesta no es completa, el trabajo de Jaśkowski resultó ser de gran importancia y todavía motivo de estudio.

8.2. **Newton da Costa.** El crédito al origen de la Lógica Paraconsistente, como se conoce hoy a los Sistemas Formales de Inconsistencia, es para Newton Caneiro Affonso da Costa. Desde 1954 ha creado de forma independiente, muchos *Sistemas Formales de Inconsistencia* del cálculo proposicional, extendiéndose al cálculo de predicados y teoría de conjuntos.

En 1959 da Costa propone el principio de Tolerancia en Matemática. Al respecto, las primeras referencias de su pensamiento son:

> *De manera inmediata se sigue que, sintáctica o semánticamente, un lenguaje objeto en el que aparezcan contradicciones no puede ser excluido a priori. En este ámbito, claro, no sería conveniente utilizar, en la estructuración del lenguaje en consideración, el cálculo lógico tradicional pues como ya sabemos, esto lo transforma en una banalidad, algo desprovisto de toda importancia matemática. Sin embargo, si cambiamos de manera apropiada las reglas "lógicas" a utilizarse, nada lo diferenciaría -en esencia- de las teorías consistentes* [5].

En su primera publicación [3], en 1963, da Costa presenta el sistema C_1 y la jerarquía de lógicas C_n , $1 \leq n \leq \omega$, proponiendo a C_ω como el sistema límite de la familia, que herede las propiedades comunes. A ésta le siguen publicaciones que conducen a establecer las propiedades más importantes de la jerarquía. La primera teoría paraconsistente de conjuntos es propuesta con base en C_1.

Durante el "*Third Latin-American Symposium of Mathematical Logic*" en 1976, en la ciudad de Campinas, Brasil, F. Miró Quesada, junto con da Costa, propusieron dejar atrás la expresión "Sistemas Formales de Inconsistencia" para sustituirla con el nombre de Lógicas Paraconsistentes.

La contribución de da Costa no sólo está en crear y desarrollar la lógica paraconsistente como un campo autónomo de investigación matemática, también ha organizado y respaldado el tema, mejorándolo. Junto con sus colaboradores, da Costa ha introducido el estudio de diversas lógicas y teorías paraconsistentes, tales como:

Teoría paraconsistente de conjuntos.
Semánticas apropiadas y álgebras asociadas a estos sistemas.
Procedimientos de decidibilidad.
Teoría de modelos paraconsistentes.
Un cálculo diferencial paraconsistente.
También cuenta con análisis de aplicaciones del análisis funcional y teorías físicas.

Además de las aplicaciones en los fundamentos de la ciencia y su análisis filosófico, la Paraconsistencia es actualmente un campo del conocimiento, con aplicaciones en informática y tecnología.

Por último, cabe señalar que las teorías paraconsistentes no tienen como objetivo sustituir a las teorías desarrollasdas antes, como las de Brouwer y Hilbert. Retomando palabras del propio Wittgenstein: no se trata de eliminar todo lo que los matemáticos han desarrollado a lo largo de la historia. La Paraconsistencia no excluye a priori una teoría en la que aparecen contradicciones [4].

Referencias

[1] Ariza E. y Otros, Memorias de la 5^a Gran Semana Nacional de Matemáticas, *El Razonamiento Matemático*, pp. 41-60, Primera edición 2010.
[2] Burdman Feferman Anita y Solomon Feferman, *Alfred Tarski Life and Logic*, Cambridge University Press, 2004.
[3] Newton C. A. da Costa, Présenté par M. René Garnier, *Calculs Propositionnels pour les Systèmes Formels Inconsistants*, 1963.
[4] Newton C. A. da Costa, D. Krause and O. Bueno, *Paraconsistent Logics and Paraconsistency*, Handbook of the Philosophy of Science. Volume 5: Philosophy of Logic Volume editor Dale Jacquette. Handbook editors; Dov M. Gabbay, Paul Thagard and and John Woods, 2006.
[5] N. C. da Costa, *Nota sobre o conceito de contradicão*, Anuário da Soc. Paranaense de Matemática, 1, NS: 6-8, 1958.
[6] De Gortari E., *Elementos de Lógica Matemática*, Océano, 1983.
[7] Gray J., *El reto de Hilbert: Los 23 problemas que desafiaron a la matemática*, Crítica Drakontos Barcelona 2005.
[8] Jaśkoswki S., *Propositional Calculus for Contradictory Deductive Systems* (in Polish), Studia Societatis Scientiarum Torunensis, 1948, Translated into English: Studia Logica, 1967.
[9] Kleene S. C., *Introduction to Metamathematics*, Nort-Holland, 1952.
[10] O'Connor J. J. and E. F. Robertson, *Non-Euclidean Geometry*, 1996.
[11] Torreti Roberto, *La Geometría en el pensamiento de Kant*, Universidad de Puerto Rico.
[12] Wang Hao, *Reflections on Kurt Gödel*, Massachusetts Institute of Technology, Sexta edición 2002.
[13] Wang Hao, *from Gödel to Philosophy*, A logical Journey, Massachusetts Institute of Technology, 1996.
[14] Wittgenstein L., *Tractatus lógico filosófico*, Traducción de Luis M. Valdés Villanueva, tercera edición, Ed. Tecnos (grupo anaya) 2008.
[15] Wittgenstein L., *Observaciones filosóficas*, traducido por Alejandro Tomasini Bassols, Instituto de investigaciones filosóficas UNAM, 2008.

Facultad de Ciencias de la Computación - BUAP.
Av. San Claudio y 14 Sur, Ciudad Universitaria.
avеariza@hotmail.com, pgarciajz-2@hotmail.com, tamayorg@hotmail.com

Capítulo 17

UN BOSQUEJO HISTÓRICO DE ALGUNAS RELACIONES ENTRE LAS CIENCIAS Y LA MILICIA

JUAN FRANCISCO ESTRADA GARCÍA
FCFM - BUAP

Resumen. En este artículo se hará un recuento histórico de algunas relaciones entre los científicos y ciertas aplicaciones militares de sus conocimientos.

1. ¿Qué es la Guerra?

La guerra es un acto de violencia para imponer nuestra voluntad al adversario y privarlo de toda resistencia. La violencia, para enfrentarse a la violencia, recurre a las creaciones de las artes y de las ciencias. La guerra es una serie de actos irracionales que ha existido sin importar el nivel de civilización de los pueblos. Sin embargo, estos actos son conducidos racionalmente, utilizando conocimientos sobre estrategia y táctica. Así, la invención de la pólvora y el constante perfeccionamiento de las armas para la guerra, muestran por sí mismas con suficiente claridad, que la necesidad inherente al concepto teórico de la guerra, la de destruir al enemigo, no ha sido en modo alguno debilitada o desviada por el avance de la civilización (vale decir, de sus conocimientos en las artes y las ciencias). Por el contrario, con el avance de la civilización, el armamento bélico se perfecciona, se hace más sofisticado.

La guerra de una comunidad (guerra entre naciones, y particularmente de naciones civilizadas) surge casi siempre de una circunstancia política, y se pone de manifiesto por un motivo político. La guerra, es la mera continuación de la política por otros medios. La guerra no es simplemente un acto político, sino un verdadero instrumento político. Aquí, se considera la política como lo relativo a la organización del poder de un Estado, a su ejercicio.

La milicia es la encargada oficial del Estado de ejecutar la orden de guerra. Sin embargo, el país en guerra, con su superficie y población, con su infraestructura económica-política, con su industria y todos sus recursos, es no sólo la fuente de las fuerzas militares propiamente dichas, sino que, en sí mismo, es también una parte integral de los factores que actúan en la guerra.

Observación: De acuerdo a lo anterior, puede decirse que para llevar a cabo el cometido del encabezado, es necesario un recorrido más amplio, es decir, al menos se requiere de la historia de las relaciones entre política, economía, comercio, ciencias, tecnologías, industrias, armamentos y milicia, lo cual es una tarea que será minimizada en lo que sigue.

2. Del Renacimiento a la Revolución Francesa

Antes del Renacimiento, las armas de fuego y la artillería trastocaron al armamento, pero los científicos no tomaron parte; principalmente ligada a la medicina o a la alquimia, la química, como ciencia, participó casi nada en la confección de la pólvora para los cañones o para la metalurgia, al menos hasta 1780; todo el progreso se debía a artesanos con cierta técnica. Algo parecido puede decirse en la construcción naval, pero no *en el arte de la navegación*, la cual, a partir de 1550 aproximadamente, suscitó en Londres el interés de los matemáticos, astrónomos y cartógrafos (que con frecuencia son los mismos) como Robert Recorde, John Dee, Thomas Digges y, por 1620, Henry Briggs, quien popularizó los logaritmos de Neper; todas esas gentes estaban en contacto directo con los navegantes y los exploradores y, difundiendo las matemáticas en un público relativamente vasto, contribuyeron en la formación del movimiento que condujo a la obra de Newton (1643-1727).

Los científicos del Renacimiento, que son tanto artistas como ingenieros, imaginaron tanques, máquinas que vuelan, submarinos, etc. El Barón escocés Neper, inventó una bomba para sacar el agua de las minas de carbón; antipapista fanático y propagandista popular del Apocalipsis, propuso a la Reina Elizabeth un submarino mucho más imaginario, para luchar contra la Armada española; al final de su vida, él pretendió haber inventado una máquina

> "capaz de deshacerse en un campo de cuatro millas de radio, de toda criatura viva con más de un pie de alto".

Tentador programa que el ingeniero estadounidense nacionalizado británico, Hiram Maxim realizó con su ametralladora automática de 600 tiros por minuto en 1884. Muy cristiano, el padre de los logaritmos había rehusado divulgar los planos de su máquina porque alegó:

> "Se ha dado ya a los hombres tantas armas para matarse entre ellos, que si eso dependiera de mí, haría todo para reducir su nḿero. Pero viendo que la maldad enraizada en el corazón del los hombres no lo permitirá jamás, quiero al menos evitar contribuir a su aumento".

Los siglos XVII y XVIII son más calmos, las naciones europeas controlaban la mayor parte de África, América y gran parte de Asia. Los científicos comenzaron a organizarse, particularmente en Academias, y soñaron con la ventaja de establecerse en "comunidad internacional" y, hacer que se reconociera el valor intelectual de sus trabajos, sin tener que desarrollar sus aplicaciones, si bien se habla de ellas con frecuencia. En esta época, se puede decir que los problemas de mayor interés fueron la cartografía, la geodesia, así como los problemas técnicos de la navegación marítima, sobre todo la determinación de la longitud en el mar, lo cual permitiría evitar errores catastróficos de navegación. Para la solución de este problema, las potencias marítimas (Portugal e Inglaterra), ofrecieron recompensas extraordinarias, lo cual inspiró a Galileo, Huygens, Hooke y otros más. El problema fue resuelto en Inglaterra en el siglo XVIII por un modesto artesano, relojero autodidacta. La relojería provoca el desarrollo de maquinaria y de herramienta de tamaño reducido, incorporando ciertos principios que florecieron en el siglo siguiente. Todos estos avances fueron utilizados posteriormente por los marinos militares, en particular, la fabricación de resortes suscitó la invención en 1740, por el inglés Benjamin Huntsman, de la técnica del acero acrisolado, la cual fue aplicada hacia 1860, por

la artillería en la confección de placas de blindaje.

Utilizando la mecánica Newtoniana, la balística comienzó a parecer una ciencia matemática y experimental hacia 1740; el matemático inglés Robins, inventó instrumentos para medir la velocidad inicial de una bola y la resistencia al aire (proporcional al cuadrado de la velocidad abajo de 300 m/s, y curiosamente al cubo por encima de esa velocidad), lo cual le permitió corregir errores de cálculo hechos por Galileo; se comienzaron también a calcular tablas de tiro "con el método de trapecios". El libro de Robins (1742) fue traducido por Euler, quien lo adornó con comentarios. Se estudiaron también las reacciones químicas producidas por la pólvora y se descubrieron nuevos explosivos, como el clorato de potasio(Berthollet 1793) y el fulminante de mercurio (Howard 1800), el cual se utilizó en la dinamita de Nobel.

En un dominio cercano, están los trabajos de Monge sobre geometría descriptiva, aplicada al arte de las fortificaciones, lo cual era enseñado en las escuelas militares y de marina francesas.

La gran innovación científico-militar de la revolución francesa fue la Escuela Politécnica, establecimiento público de enseñanza e investigación, fundada en 1794 y militarizada por Napoleón en 1805. Se le llama "la X", porque en su símbolo de armas se pueden ver dos cañones cruzados que forman una X . La X preparó la liga entre la terna ciencias-milicia-industria, si bien esta institución fue acompañada de la creación de otras instituciones como el Museo, La Escuela Normal, El Sistema Métrico, La Oficina de las Longitudes, etc., las cuales en conjunto se interesaron en "el progreso de las ciencias y de las artes" y, por supuesto, en la educación. Mientras que en EEUU, la academia militar más antigua (USMA), conocida como West Point, se fundó en 1802.

3. La Revolución Industrial

La Revolución Industrial se realizó inicialmente de 1739 a 1869, en la que Gran Bretaña en primer lugar y el resto de Europa continental después, sufrieron el mayor conjunto de transformaciones socio-económicas, tecnológicas y culturales de la Historia de la humanidad. La sociedad feudal estamental dejó de existir y en su lugar se erigió la sociedad de clases capitalista, articulada en torno a dos grupos sociales: la burguesía y el proletariado.
Los elementos mś importantes que permitieron las transformaciones que llamamos Revolución Industrial son:

(1) La aplicaión de las ciencias y las tecnologías en la invención de máquinas que aceleraban los procesos productivos.

(2) El uso de nuevas fuentes energéticas como el carbón y el vapor.

(3) La revolución en el transporte: ferrocarriles y barcos de vapor.

(4) La despersonalización de las relaciones de trabajo.

(5) El surgimiento del proletariado urbano.

Lo cual trajo consecuencias como:

(1) Económicas: producción en serie, desarrollo del capitalismo, intercambios desiguales.

(2) Demográficas: traspaso de la población del campo a la ciudad, migraciones internacionales (crecimeiento sostenido de la población).

(3) Sociales: con el proleteriado nace la cuestión social.

(4) Ambientales: deterioro del ambiente y degradación del paisaje, explotación irracional de los recursos naturales.

La Revolución Industrial se clasifica en las siguientes tres etapas: la primera, desde los primeros usos del carbón en 1732 hasta la producción de electricidad en 1869; la segunda, desde 1869 hasta la Primera Guerra Mundial (1914); la tercera, desde el fin de la Segunda Guerra Mundial (1945) hasta la actualidad. Su base técnico-científica es revolucionaria, generando así el problema de la obsolescencia tecnológica en períodos cada vez más breves. Esta característica de obsolescencia e innovación, se extiende a toda la estructura socio-económica de las sociedades modernas. En este contexto, la innovación es por definición, negación, destrucción, cambio, la transformación; es la esencia permanente de la modernidad. Nunca se consideran los procesos de producción como definitivos o acabados. El desarrollo de nuevas tecnologías, como ciencias aplicadas en un receptivo clima social, lleva a un proceso acumulativo de tecnología, que genera bienes y servicios, mejorando el nivel y calidad de vida. Todo esto, exige un sistema educativo adecuado y espíritu emprendedor. Los procesos de industrialización producen estados sociales muy inestables, en la práctica inevitables, lo cual produce abundancia de problemas de equilibrio y justicia social.

La revolución industrial, en particular, los inicios de la industria química y la siderúrgica, tuvieron efectos sobre el armamento y uniformes militares, en lo concerniente a la fabricación de cañones y de recámaras de plomo, para la fabricación del ácido sulfúrico, necesario en la industria textil. La siderúrgica y la química permitieron la industrialización agraria, la cual también repercute eventualmente en la manutención del ejército. Entre las múltiples innovaciones que produjo, podemos resaltar: la mecanización textil, la industria química (ácido sulfúrico, cloro, sosa, fertilizantes, colorantes, fotografía, plásticos, fibras artificiales, medicamentos, electro-química, petróleo, etc.), ferrocarriles, navegación con vapor, máquinas agrícolas, telégrafo eléctrico, teléfono, luz eléctrica, energía y tracción eléctrica, turbinas hidráulicas, a vapor y gas, motores de explosión (gas, gasolina, diesel), automóvil, aeronáutica, refrigeración (alimentación, licuefacción de gas), imprenta rotativa y linotipos, máquinas de escribir o contables, cine, radio, etc. Cabe señalar, que no todas las innovaciones fueron hechas por gente con carreras universitarias o

técnicas, sino por los llamados "hombres de la práctica".

A nivel mundial, aparecieeon entonces los países ricos, es decir industrializados, y los países pobres, los no indutrializados, lo cual aunado a la ambición de poder comercial-económico-político de parte de los países ricos, fortaleció sus ambiciones coloniales y preparó el terreno para la guerra entre estos países.

4. El siglo XIX

Instituciones educativas y de investigación

Paralelamente a los desarrollos industriales, en el siglo XIX apareció la enseñanza científica y técnica, la cual varía mucho de un país a otro, especialmente en función del papel que jugaban los poderes públicos en Alemania y Francia, y que en Inglaterra y EEUU casi no aparecieron. En el mundo germánico, que sirvió de modelo y desafío al resto del mundo a partir de 1870, la X inspiró una docena de imitaciones en Praga (1806), en Viena (1815), y después en Karlsruhe, Múnich, Dresde, etc., y en Zúrich en 1855, agregando buena cantidad de escuelas profesionales de nivel menos elevado, fortalecido esto con el establecimiento obligatorio de la escuela primaria un siglo antes por Prusia, así como un excelente sistema de enseñanza secundaria, inaugurado en 1819. En principio, la pedagogía y programas de estudio, no estaban encaminados a conectar las teorías con las prácticas industriales. El objetivo pedagógico de los fundadores fue sobre todo proveer a las futuras élites de una formación cultural, fuertemente influenciada por las ideas de filósofos, filólogos e historiadores que deseaban terminar con el obscurantismo clerical de las universidades tradicionales. Es en este contexto que se desarrolló la ideología de una "ciencia pura", opuesta a las concepciones utilitarias y comerciales. Es así como en Königsberg en 1835, el joven Jacobi fundó junto con Franz Neumann, el primer seminario de matemáticas y de física, inspirado en el modelo de un seminario de filología que él frecuentaba en Berlín, y donde él preconizaba la práctica de la ciencia "por el honor del espírtu humano".

Sin embargo, desde 1826, intentando contrarrestar la invasión de manufacturas inglesas, J. Liebig en la universidad de Giessen en Alemania, instauró un sistema de aprendizaje metódico del análisis químico, con presencia constante en el laboratorio. A pesar de múltiples problemas financieros, estas ideas lograron conducir a la creación y al desarrollo de laboratorios universitarios en Múnich, dirigidos por los mejores químicos de la época. A partir de 1870, Francia retomó el ejemplo alemán a instancias de Pasteur y Berthelot. A fines de siglo XIX, Alemania produjo aproximadamente el doble de ingenieros y de técnicos que Francia. Esta característica se mantiene hasta nuestros días.

Hasta la Guerra de Secesión o Guerra Civil Estadounidense (1861-1865), EEUU era un país pequeño, tanto por su población como por su producción intelectual. En la enseñanza técnica, se hicieron tentativas aisladas. West Point (1805), se reorganizó más tarde y produjo un Corps of Engineers, para la construcción de las vías del ferrocarril, el Polytechnic Institute fundado por un filántropo en 1826, y algunos servicios gubernamentales que contribuyeron a la actividad científica, especialmente el Geological Survey, al cual el Congreso rehusó financiar sus eventuales

actividades científicas (son las minas lo que le interesa), y la Marina.
El primer "gran salto hacia adelante" de este país es dado por la ley federal de 1862, que le permite crear colegios de agricultura y de mecánica, dotándolos de un capital que proviene de la venta de terrenos federales. La mayor parte de los colegios se convirtieron en las State Universities actuales por extensiones sucesivas de su vocación inicial. El MIT fue fundado por el geólogo W. B. Rogers en 1861. En las universidades se instauró la organización en departamentos, independientes unos de otros, correspondientes a las diversas disciplinas de interés, en donde se proclamó como prioridad la investigación. Los ejemplos más célebres son: Johns Hopkins en Baltimore, fundada en 1876, gracias a un capitalista filántropo quien donó 3.5 millones de dólares, en donde nació en seguida el American Journal of Mathematics; la Universidad de Chicago, fundada en 1892 por John D. Rockefeller; las universidades de Stanford y de Cornell, fundadas antes de 1900 también por filaństropos; el rey del acero, Andrew Carnegie, quien no tenía heredero, vendió en 1901 su negocio en cerca de 400 millones, dejando casi todo a fundaciones, entre las cuales dejó 20 millones a un Carnegie Institution (Washinton), dedicada principalmente a la investigación aplicada; George Eastman (Kodak) donó, durante el transcurso de su vida, aproximadamente 35 millones de dólares a la Universidad de Rochester y 20 al MIT.

Esas fundaciones y filántropos ejercieron hasta 1940 una gran influencia sobre las univerisdades y la investigación científica, no sólo en razón de la casi ausencia de todo financiamiento federal, sino también en razón del hecho de que ellas ayudaron en general sólo a establecimientos administrados "racionalmente" y donde el nivel intelectual era suficientemente elevado. Otros capitalistas siguieron estos ejemplos hasta la actualidad. El control de la universidades pasó entonces a los generosos donadores y a sus amigos industriales, banqueros o abogados de negocios, que terminaron por constituir el 90% de sus consejos de administración. A partir del final del siglo XIX, fueron principalmente las nuevas empresas capitalistas (General Electric, AT&T, Westinghouse, Stanford Oil, DuPont, etc.) las que, de más en más, emplearon a los ingenieros e investigadores, dotándose antes de la Primera Guerra Mundial de los servicios de investigación y desarrollo. Así encontramos en EEUU una simbiosis más o menos completa entre la industria y los departamentos técnicos o científicos de las universidades o institutos de tecnología. El ingeniero en jefe de la AT&T expresa en 1916 una idea con gran futuro:

> *En último análisis, la distinción entre la investigación científica pura y la investigación industrial, es simplemente una cuestión de motivos.*

Nobel hace notar que la investigación desinteresada en busca de la verdad y la que busca la utilidad y el provecho, no son incompatibles, en resumen:

> *El científico podrá continuar inmerso en los misterios del universo, sin ocuparse de cuestiones prácticas o financieras, tanto como quiera, ya que sus descubrimientos pueden fácilmente ser traducidas por otros en provecho del desarrollo industrial capitalista. No se le pide estar de acuerdo con los motivos. Lo que se le pide, es una coordinación satisfactoria entre los medios y los fines y, para tal efecto, una organización adecuada.*

Otros usuarios de la investigación fundamental, que no es la industria capitalista, tomarán ese comentario por su cuenta. La investigación fundamental consistirá entonces en elaborar sin motivos prácticos un catálogo de herramientas a la disposición de cualquier usuario. Las dos Guerrras Mundiales, la Guerra Fría y otras guerras, con este punto de vista, pusieron en problemas éticos a ciertos científicos.

Estos desarrollos de EEUU, no pasaron desapercibidos por el resto de los países industrializados, lo cual condujo a la búsqueda constante, por tales países, de la supremacía en las ciencias y tecnologías, o al menos a evitar rezagarse. Estos enfoques, al intentar ser realizados por los distintos países, acarrearon problemas de índole diversa, en particular económicos (presupuestos), políticos y sociales, los cuales no serán analizados aquí. Es conveniente señalar un hecho que liga una vez más la industria con la milicia: en Alemania, donde se desarrolló una gran industrria textil, tales empresas se transformaron repentinamente, durante la Primer Guerra mundial, en fábricas de explosivos y de gas tóxico; además proveyeron al régimen nazi de grandes cantidades de petróleo y de hule sintético, indispensables en la guerra, sin contar otras cuestiones.

Los nuevos explosivos

Las propiedades explosivas de la nitroglicerina y de la nitrocelulosa, que dieron lugar a una revolución militar de primer nivel, fueron descubrimientos accidentales y sin el mínimo objetivo militar, conseguidos por los químicos Sobrero y Schönbein, en 1846-47, lo cual es poco sorprendente, ya que una de las distracciones favoritas de los químicos de esa época, era ligar radicales de NO_2 a todos los compuestos orgánicos que los aceptaran, lo cual convierte la química de los colorantes muy próxima a la de los explosivos. Pero su utilización en la artilllería así como a su fabricación y almacenamiento, dio lugar a accidentes espectaculares (aun antes de 1914, naves de guerra explotaron espontáneamente) y se enfrentó a dificultades mayores que no fueron resueltas antes de 1880-90.

Nobel estabilizó en 1867 la nitroglicerina y la transformó en dinamita. Más importante aun, inventó los detonadores con fulminantes de mercurio accionados eléctricamente. Él mejoró a continuación su dinamita, agregándole pasta de nitrato de amonio. El nuevo explosivo revolucionó no solamente el armamento, sino también los trabajos públicos y las minas, las cuales hicieron un consumo prodigioso. Después de 15 años de trabajo, Nobel produjo para la artillería, en 1888, una "pólvora", mezclando nitroglicerina y colodión; él esperó, uno más, hacer la guerra suficientemente horrible y hacer que los hombres la rechazaran. Muchos otros explosivos, militares (TNT sobre todo, conocido mucho tiempo antes en la industria de los colorantes y que la armada alemana utilizara a partir de 1904) o civiles, fueron descubiertos a continuación en los laboratorios, especializados o no. Un químico alemán publica en 1908, en *Berichte*, una historia del tema, donde hace notar que en la química orgánica, cada químico tiene un explosivo en sus utensilios; con los deto-
nadores a la Nobel, permite hacer explotar casi cada uno de ellos.

Los comerciantes de la muerte

A partir de 1860-70, los progresos técnicos del armamento fueron favorecidos por la competencia entre Gran Bretaña, Francia, Alemania, Rusia, EEUU, Japón e Italia. El sector naval era el más técnico. Pero las fabricaciones resultaban muy costosas. El ritmo de cambio técnico resultó tan rápido, que los arsenales gubernamentales fueron obligados a recurrir, al menos en parte, a las empresas privadas, las cuales les permitieron amortizar las enormes inversiones necesarias, y compensar las altas y las bajas de pedidos nacionales, gracias a los productos civiles, por ejemplo, naves de pasajeros y sobre todo a las exportaciones de armas, donde la mejor manera de obtener pedidos, es hacer el material de guerra cada vez más sofisticado. El progreso de los armamentos tiene, naturalmente, como principal objetivo, tener mejores armas que los potenciales enemigos. Todo mundo se dio cuenta que la carrera armamentista necesita de sumas astronómicas, sin otro resultado previsible que transformar cada uno de los participantes en una amenaza para los otros.

5. El siglo XX

En la Conferencia de la Paz de 1906, que no llegó a casi nada, todo el mundo reconoció que la mejor manera de evitar la guerra, es estar preparado para ella y que a la postre hizo la Primer Guerra Mundial inevitable (1914-1918).

La energía nuclear

En 1896 H. Becquerel descubrió que algunos elementos químicos emitían radiaciones. Se descubrió que estas radiaciones eran diferentes a los ya conocidos Rayos X y que poseían propiedades distintas. En 1911 Rutherford describió estas radiaciones (alfa, beta, gama), que provenían del núcleo atómico. En 1930 Pauli describió teóricamente una partícula llamada *neutrino*. En 1932 J. Chadwick descubrió la existencia del neutrón predicho por W. Pauli, e inmediatamente después, E. Fermi descubrió que ciertas radiaciones emitidas en ciertos fenómenos de desintegración, eran en realidad neutrones. Fermi y sus colaboradores bombardearon con neutrones más de 60 elementos, entre ellos *uranio*, produciendo las primeras fisiones nucleares artificiales. En 1938, en Alemania, L. Meitner, O. Hahn y F. Strassmann verificarón los experimentos de Fermi, y en 1939 demostraron que parte de los productos que aparecían al llevar a cabo estos experimentos con uranio, eran núcleos de *bario*. Muy pronto, llegaron a la conclusión de que eran resultado de la división de los núcleos del uranio, fenómeno llamado *fisión*, el cual produce energía. En Francia, J. Curie descubrió que, además del bario, se emitían neutrones secundarios en esa reacción, haciendo factible la reacción en cadena, las energías liberadas por cada una de ésas fisiones, se suman, pudiendo alcanzar proporciones gigantezcas. De ahí *la idea de una bomba*, la cual se intuía posible desde 1935. Un kilo de uranio libera más calor que mil toneladas de dinamita

Durante la Segunda Guerra Mundial(1939-1945), el Departamento de Desarrollo de Armamento de la Alemania nazi desarrolló un proyecto de enrgía nuclear, con vistas a la producción de un artefacto explosivo nuclear. A. Einstein, en 1939, firmó una carta al presidente F. D. Roosevelt de EEUU, escrita por L.Szilard (húngaro

judío), en la que prevenía sobre este hecho. Ellos no fueron muy lejos en tal dirección, ya que no recibió el visto bueno de Hitler. En menos de diez años, la bomba atómica pasó del estado de especulación pura, a la de relidad terrorífica, engendrada por una de las más prodigiosas concentraciones de materia gris de la humanidad. En 1942, en Los Álamos, pueblo perdido en los desiertos de Nuevo México, se reúnen un grupo de los mejores científicos del planeta: físicos, matemáticos y químicos. Algunos de los participantes en este proyecto (Manhattan), fueron: Hans Bethe, director de la sección teórica, Robert Wilson, el hombre de las máquinas para acelerar las partículas, Philip Morrison, Richard Feynman, Robert Oppenheimer, Stanisław M. Ulam, John von Neumann, Edward Teller y Enrico Fermi, varios de ellos de ascendencia judía.

La armada de EEUU instaló un súper-laboratorio donde todos los medios son puestos a disposición de los investigadores, utilizando a fondo su formidable infraestructura industrial y técnica. Se trabaja día y noche sin tomar vacaciones, estimulados por las victorias alemanas y los campos de exterminación de judíos. El parto es largo y difícil. La bomba se manifiesta por primera vez en julio de 1945, en Alamogordo, Nuevo México. Poco después, el 6 y 9 de agosto de 1945, ella muestra su verdadera cara. Dos ciudades japonesas son aniquilidas: Hiroshima y Nagasaki. En algunos segundos, decenas de miles de personas, pasan literalmente a estado gaseoso. En total, se estima que las bombas mataron 140 000 personas en Hiroshima y 80 000 en Nagasaki.

Después de EEUU, la Unión Soviética, Francia, Inglaterra, China e India hicieron explotar artefactos termonucleares. La bomba ganó poder. En las nieves Siberianas de la Nueva-Zembla, ella alcanzará el equivalente a *decenas de millones de toneladas de dinamita*. Además proliferó. Más de 30 000 fueron diseminadas en los arsenales del planeta.

Referencias

[1] Clausewitz K. V., *De la Guerra*, Ed. Diógenes S. A. México
[2] Godement R., *Science et Défense,* Gazette des Mathématiciens, n. 61, julio 1994 pp. 3-60
[3] Manegold K. H., *Universität, Technische Hochschulen, und Industrie*, Ed. Duncker & Humblot, 1970
[4] Reeves H., *L'heure de s'enivrer*, Éditions du Seuil, 1992

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

festrada@fcfm.buap.mx

Capítulo 18

ESTRATEGIAS PARA RESOLVER PROBLEMAS DE MATEMÁTICAS DE NIVEL PREUNIVERSITARIO

LIDIA AURORA HERNÁNDEZ REBOLLAR
MARÍA ARACELI JUÁREZ RAMÍREZ
FRANCISCO JAVIER RODRÍGUEZ MARTÍNEZ
FCFM - BUAP

Resumen. El objetivo de este artículo es presentar algunas herramientas que ayuden a desarrollar la habilidad general para resolver problemas. Nos centramos particularmente en el estudio de la metodología de Pólya presentada en su libro "Cómo plantear y resolver problemas", extrayendo de su literatura el modelo general que él propone como método para resolver cualquier problema de matemáticas, y además, estrategias aplicables a la resolución de problemas de matemáticas en el ámbito de pruebas de selección por el hecho de que éstas son utilizadas como vehículos al servicio de otros objetivos curriculares tales como probar un dominio del conocimiento, manejo de conceptos, actitudes y procedimientos.

1. Introducción

La resolución de problemas que impliquen el uso de un razonamiento lógico y matemático ha sido centro de atención de todos aquellos que de una forma u otra tienen la tarea de educar, especialmente de aquellos que lo hacen en el campo de la matemática. Esto se debe a que la resolución de problemas tiene un carácter desarrollador del pensamiento lógico en los estudiantes, donde la independencia hará que el educando sea más organizado en un trabajo dirigido a las exigencias educativas de la actualidad.

Luego, la resolución de problemas de matemáticas se visualiza como una disciplina de la matemática, caracterizada por resultados precisos pero con procedimientos flexibles e infalibles cuyos elementos básicos son las operaciones aritméticas, los procedimientos algebraicos, manejo de conceptos geométricos y teoremas elementales; saber y dominar esta matemática básica es equivalente a ser hábil en desarrollar procedimientos e identificar vías de solución rápidas y eficaces.

Por lo tanto, se propone la resolución de problemas matemáticos no rutinarios como una de las habilidades más importantes a ser desarrolladas en el currículo escolar, caracterizándola como una habilidad necesaria en el proceso de formación de los estudiantes que están en transición del nivel medio superior al superior, ya que, independientemente de la carrera que deseen estudiar se encontrarán con un examen de admisión que pone a prueba sus conocimientos y dominio de conceptos, definiciones y teoremas más importantes y significativos, mediante una variada colección de problemas de matemáticas.

En este trabajo, se busca ayudar al lector a desarrollar la habilidad de solución de problemas en el ámbito de exámenes de admisión de nivel universitario. Para esto, se presenta una sugerencia de solución para cada modelo de consigna y se proponen algunas estrategias que pueden ser aplicadas, después, se presenta la solución de los problemas con la aplicación de dichas estrategias, todo esto con el afán de promover en el estudiante el uso deliberado de estrategias, principalmente las mencionadas en la Metodología de Pólya.

Finalmente, se busca contribuir al mejoramiento del proceso docente mediante el diseño de una guía para la preparación de estudiantes en la solución de problemas de matemáticas del estilo de los exámenes de admisión para nivel superior. El trabajo completo se puede consultar en la referencia [1], donde se presentan 60 problemas resueltos con la aplicación. Al mismo tiempo se cumple con lo que las tendencias de enseñanza y aprendizaje exigen hoy en día, en el hecho de hacer hincapié en el desarrollo o promoción de los procesos de pensamiento propios de la matemática más que en la mera transferencia de contenidos.

2. Metodología de la investigación

Se investigó acerca de estrategias para la solución de problemas de matemáticas expuestos en la literatura, y se seleccionaron las expuestas por G. Pólya en su publicación "Cómo plantear y resolver problemasçon la finalidad de aplicarlas en los problemas tipo examen de admisión.

Se llevó a cabo una selección de problemas tipo examen de admisión a partir de diversas guías de estudio y material relacionado, para mostrar su solución mediante la aplicación de las estrategias específicas.

3. Marco Teórico: Metodología de Pólya

Pólya plantea que en la resolución de problemas (Cómo plantear y resolver problemas, 1965), los aspectos matemáticos deben ser primero imaginados y luego resueltos. Para esto, propone una metodología compuesta por cuatro etapas, misma que nos servirá de guía para el desarrollo de la propuesta plasmada en este trabajo, y que ha sido enfocada a la resolución de problemas de matemáticas del estilo examen de admisión de nivel superior.

Las reflexiones que Pólya establece, están dirigidas a afrontar un problema mediante una estrategia general constituida por cuatro etapas. Esta estrategia general es la que el profesor debe enfatizar en el aula, mostrándola a lo estudiantes como método general de solución.

Etapa 1: Entender el Problema. El alumno debe ser capaz de identificar las partes principales del problema, la incógnita, los datos, la condición. Se deben considerar estas partes atentamente y pensarlas desde diferentes ángulos a fin de extraer todos los datos implícitos que ésta contenga. Si hay una figura relacionada al problema, debe *dibujarla* y destacar en ella la incógnita y los datos. Es necesario dar nombres a dichos elementos y por consiguiente *introducir una notación* adecuada.

Etapa 2: Configurar un Plan. Cuando ya se está seguro de haber entendido bien el problema y se tiene toda la información necesaria, entonces será el momento de elegir una estrategia para resolverlo. Existe una gran variedad de estrategias que conviene conocer, practicar para mejorar la capacidad de resolver problemas, e interiorizar para hacer un manejo consiente y automático de ellas.

Etapa 3: Ejecutar el Plan. Ya elegida la o las estrategias adecuadas, el estudiante debe trabajarlas con decisión y no abandonar el proceso a la primera dificultad. Pero si el procedimiento se complica sin frutos significativos, se debe volver al paso anterior y probar con una estrategia diferente. Por lo general hay varias formas de llegar a la solución y no podemos esperar acertar siempre con la más apropiada al primer intento.

Etapa 4: Una Visión Retrospectiva. Esta etapa es muchas veces omitida. Pólya insiste mucho en su importancia (y nosotros también), no sólo porque consideremos que comprobar los pasos realizados y verificar su corrección, realización efectiva y verdadera nos puedan ahorrar muchas respuestas erróneas, sino porque la visión retrospectiva nos puede conducir a nuevos resultados que generalicen, amplíen o fortalezcan el que acabamos de hallar. Entonces, en esta etapa conviene ***examinar, revisar, familiarizarse,*** y finalmente ***reflexionar*** sobre el procedimiento realizado.

Estrategias Específicas

Ahora bien, para determinar la solución de un problema, además de una estrategia general se sugiere contar también con unas *estrategias específicas* que se

aplicarán de acuerdo al tipo de problema que se presente. Disponer de un buen repertorio de estrategias, y de la experiencia en el uso de éstas, es de gran ayuda para el que soluciona problemas matemáticos. Sin embargo, el éxito en su aplicación depende mucho de la experiencia, juicio y buen sentido de quien las use. Pólya sugiere las siguientes estrategias específicas:

1. El conocimiento de los recursos matemáticos. El papel de un estudiante ante una situación matemática (ya sea de interpretación o de resolución de problemas), parte de tener claro cuáles son las herramientas matemáticas que tiene a su disposición, qué de la información que se proporciona es relevante para afrontar la situación matemática o problema, cómo acceder a esa información y cómo utilizarla.

2. Diversas posturas ante un problema matemático. Al tratar de encontrar la solución a un problema podemos cambiar repetidamente nuestro punto de vista, nuestra forma de considerar un problema cuantas veces sea necesario.

3. Plantearse preguntas. Considérense las preguntas: ¿Cuál es la incógnita?, ¿cuáles son los datos?, ¿cuál es la condición? Éstas son preguntas aplicables en toda clase de problemas, ya sea un problema aritmético, algebraico o geométrico, etc. Estas preguntas ayudan a esclarecer el problema.

4. Imitación y práctica. El estudiante debe adquirir en su trabajo personal la más amplia experiencia posible. El resolver problemas es una cuestión de habilidad práctica, habilidad que se adquiere mediante la imitación y la práctica.

5. Buscar semejanzas con otros problemas. ¿A qué le recuerda al estudiante la situación planteada en el enunciado del problema? ¿No se intuye que tal vez sea como alguna otra resuelta con anterioridad?

6. Marcar las palabras clave. Encerrar o subrayar las palabras clave es una técnica muy efectiva, pues muchas veces por la tensión y la presión del tiempo se puede entender mal lo que se pide en el examen y seleccionar por esto una respuesta equivocada.

7. Extraer información útil. Poner mayor atención para obtener información implícita al leer un problema escrito puede ayudar a entenderlo y resolverlo rápidamente. Identificar los datos que dan y decidir cuáles servirán para resolver el problema.

8. Reducir lo complicado a lo simple. Normalmente el camino correcto para la resolución de un problema complicado es la división de éste en otros más sencillos.

9. Hacer un dibujo. A veces, una imagen aporta más que toda la información implícita en un enunciado. En el dibujo o esquema que se haga, se deben incorporar los datos realmente importantes y prescindir de lo demás.

10. Estudiar todos los casos posibles. Se trata de ver todas y cada una de las posibilidades, analizar si se pueden aceptar o descartar y por qué. El enunciado

de muchos problemas suele impactar por su formalidad y generalidad de lo que plantean y tipo de solución que solicitan.

11. Elegir una buena notación. Eligiendo una buena notación, un problema se puede simplificar notablemente. El objetivo es relacionar los datos con las variables desconocidas y tratar de hacer los cálculos de la mejor manera posible.

12. Incorporar algo adicional. A veces, al incorporar un elemento nuevo, por ejemplo, una línea, una incógnita, una fórmula conocida (aritmética, algebraica, etc.), se ponen de manifiesto relaciones que de otra forma pueden pasar desapercibidas.

13. Sustituyendo números. Cuando un problema involucra variables (letras o cantidades desconocidas) puede parecer difícil y confuso, simplemente reemplazar las variables por números.

14. Trabajando desde las respuestas. En ocasiones, la solución a un problema será obvia. En otras será más útil trabajar usando las respuestas que se proveen. Se recomienda tener muy presente esta estrategia cuando se trabaja con problemas de opción múltiple.

15. Determinar un patrón. Principalmente cuando se trata de problemas donde la hipótesis está dada por una serie de números o condiciones que determinan una situación reiterativa se recomienda analizar elemento a elemento y descubrir el comportamiento de éstos.

16. Hacer estimaciones. Esta estrategia será recomendable aplicarla cuando sólo baste hacer cálculos sencillos y hasta mentales.

17. Construir una ecuación. La modelación de problemas que conducen a la solución de ecuaciones lineales o a un sistema de dos ecuaciones lineales con dos variables es una estrategia que contribuye a la solución exitosa de un problema.

4. Aplicación de la Teoría de Pólya

En esta sección se muestra la aplicación de las estrategias sugeridas por Pólya en problemas representativos de las tres áreas elementales: aritmética, álgebra y geometría; atendiendo a los diferentes modelos que generalmente aparecen en una prueba de admisión. Cada problema cuenta con su respectiva sugerencia y solución, así como la o las estrategias que recomendamos utilizar. Se menciona el número de la lista dada en la sección anterior y el nombre de la estrategia. En la solución del problema se puede apreciar la aplicación de la o de las estretegias sugeridas. En la referencia [17] se pueden consultar 60 problemas como los que aquí se enlistan y 60 más que se proponen pero sólo con una sugerencia.

1. Si Carlos ahorra \$1 el día 1, \$2 el día 2, \$3 el día 3, y así sucesivamente, ¿cuánto habrá ahorrado en total para el día 31?

Sugerencia. El ahorro del día corresponde al número de día, por lo que para el día 31 Carlos habrá ahorrado : $1 + 2 + 3 + 4 + \cdots + 29 + 30 + 31$. La fórmula de Gauss. *Estrategia 12: incorporar algo adicional.*

Solución. Para calcular el ahorro del día 1 al 31 podemos *incorporar algo adicional,* aplicar la fórmula de Gauss, la cual sirve para hallar sumas de números naturales consecutivos hasta un determinado número n, estableciéndose que:

$$1 + 2 + 3 + \cdots + n - 1 + n = \frac{n\,(n+1)}{2}.$$

Luego, la suma de los ahorros diarios hasta el día 31 es:

$$\frac{31\,(31+1)}{2} = \frac{992}{2} = 496.$$

2. Una compañía de 720 hombres está dispuesta en filas. El número de soldados de cada fila es 6 más que el número de filas que hay. ¿Cuántas filas y cuántos soldados hay en cada una?

Sugerencia. Para resolver este problema, debemos generar una ecuación de la forma $ax^2 + bx + c = 0$, y posteriormente determinar alguno de los métodos de solución para dicha ecuación. *Estrategias 12, 17, 7 y 1: incorporar algo adicional, construir una ecuación, extraer información útil y conocimiento de recursos matemáticos.*

Solución. Definamos a $\boldsymbol{S}$ como el número de soldados de cada fila, y a $\boldsymbol{F}$ como el número de filas en que está dispuesta la compañía de 720 hombres. Como "El número de soldados de cada fila es 6 más que el número de filas que hay", entonces, $\boldsymbol{F} = \boldsymbol{S} - 6$. Esta ecuación será nuestra ecuación 1.

Por otra parte, si se multiplican las filas ($\boldsymbol{F}$) por el número de soldados que hay en cada una ($\boldsymbol{S}$) obtendremos el total de soldados, por lo que $\boldsymbol{F} \cdot \boldsymbol{S} = 720$. Esta ecuación será nuestra ecuación 2.

Entonces, sustituyendo la ecuación 1 en la ecuación 2 obtendremos:

$$\boldsymbol{F} \cdot \boldsymbol{S} = \boldsymbol{S} - 6\boldsymbol{S} = 720.$$

De donde $\boldsymbol{S}^2 - 6\boldsymbol{S} - 720 = 0$, será nuestra ecuación cuadrática que cumple la forma $ax^2 + bx + c = 0$.

Ahora bien, resolveremos esta ecuación por medio de a factorización. Luego:

$$\boldsymbol{S}^2 - 6\boldsymbol{S} - 720 = (\boldsymbol{S} + 24)\,(\boldsymbol{S} - 30)$$

donde claramente el conjunto solución es $\boldsymbol{S}_1 = -24$ y $\boldsymbol{S}_2 = 30$.

Elegimos el número de soldados que tiene cada fila como $\boldsymbol{S}_2 = 30$, pues -24 soldados en una fila no es una cantidad lógica. Entonces, volviendo a nuestra ecuación 1, $\boldsymbol{F} = \boldsymbol{S} - 6$, se tiene que $\boldsymbol{F} = 30 - 6 = 24$.

Por lo tanto, el número de **filas es 24**, habiendo **30 soldados en cada una.**

3. El dividendo de una división es 112. El divisor sumado con el triple del cociente da 37. Hallar el divisor.

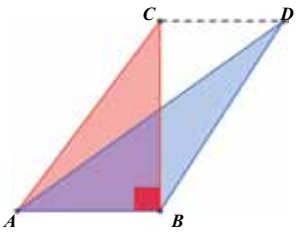
a) 7 **b)** 58 **c)** $\frac{1}{2}$ **d)** 16 **e)** Indeterminable

Sugerencia. Debemos determinar el divisor (considerando que el dividendo es 112), por lo que partiendo desde las respuestas podemos realizar las operaciones correspondientes donde se evidenciará el cociente. *Estrategias 14 y 16: trabajando desde las respuestas y hacer estimaciones.*

Solución. Atendiendo a la condición: *"El divisor sumado con el triple del cociente da 37".*

Realizando divisiones se nota que con la opción **a)** obtenemos: $\frac{112}{16} = 7$, de donde se cumple que: $divisor + 3(cociente) = 37$.

4. Si AB y DC son paralelas y de igual magnitud, $CD = 3$ cm y $BD = 5$ cm. ¿Cuál es el área del triángulo ADB?



Sugerencia. La base se conoce y la altura es la longitud de uno de los catetos del triángulo BCD. *Estrategias 1, 12 y 16: conocimiento de recursos matemáticos, incorporar algo adicional y hacer estimaciones.*

Solución. Para determinar el área del triángulo ADB necesitamos conocer su base y de su altura. La base se conoce pero la altura no, sin embargo, es posible descubrirla si observamos que tal altura es uno de los catetos del triángulo CBD. Incorporado algo adicional, como lo es el Teorema de Pitágoras sabemos que: $BD^2 = CD^2 + CB^2$.

Luego $5^2 = 3^2 + CB^2$, de donde obtenemos que $CB = \sqrt{25-9} = \sqrt{16} = 4$ cm.

Por lo que la altura del triángulo CBD (misma altura para el triángulo ADB) es: 4 y su base $AB = CD = 3$. Entonces su área será:

$$\frac{\text{base} \cdot \text{altura}}{2} = \frac{3 \cdot 4}{2} = \underline{\mathbf{6\ cm^2}}.$$

5. En un grupo de 15 personas, seis son mujeres. Si la mitad de ellas tiene ojos de color café, ¿cuál es la probabilidad de que al elegir una persona del grupo sea una mujer con ojos de color café?

a) $\frac{1}{15}$ **b)** $\frac{3}{15}$ **c)** $\frac{3}{9}$ **d)** $\frac{6}{15}$ **e)** $\frac{3}{6}$

Sugerencia. Únicamente hay que identificar al conjunto que cumple la condición (mujer con ojos de color café) respecto al conjunto total de elementos. *Estrategia 1: Conocimiento de recursos matemáticos.*

Solución. De un total de 15, seis son mujeres y 3, entonces, la probabilidad de elegir a una mujer que tenga ojos de color café es **b)** $\frac{3}{15}$.

Instrucciones. **El problema siguiente consiste en comparar dos cantidades o informaciones. La opción correcta se marcará de la siguiente manera:**

a) Si la cantidad de la Columna A es mayor.
b) Si la cantidad de la Columna B es mayor.
c) Si las dos informaciones expresan la misma cantidad.
d) Si la información sea insuficiente.

Columna A	**Columna** B

6. Sabiendo que x y y son números enteros.

$4x^2+4y^2$	$(2x+ 2y)^2$

Sugerencia. La Columna B aporta $(2x+ 2y)^2$ el cual es el cuadrado de un binomio y se puede desarrollar algebraicamente. El enunciado sólo menciona que x y y son enteros, pero no especifica su signo. *Estrategia 10: Estudiar todos los casos posibles.*

Solución. La Columna B aporta $(2x+ 2y)^2$ el cual es el cuadrado de un binomio, que sabemos que se desarrolla como $(2x+ 2y)^2 = 4x^2+ 8xy+ 4y^2$.

Ahora bien, podríamos pensar que la Columna A es menor a la Columna B. Sin embargo se sugiere no tomar decisiones precipitadas, ya que si bien es cierto que en la Columna A hay un término algebraico más que en la Columna B, no conocemos exactamente los valores de las literales x y y.

Lo anterior sugiere *estudiar todos los casos posibles,* se refiere a lo siguiente:

Si x y y son ambos positivos, entonces: $4x^2+4y^2 < 4x^2+ 8xy+ 4y^2$.

Si x y y son ambos negativos, entonces: $4x^2+4y^2 < 4x^2+ 8xy+ 4y^2$.

Si x es positivo y y negativo (o viceversa), entonces: $4x^2+4y^2 > 4x^2+ 8xy+ 4y^2$.

Por lo tanto, la respuesta es ***d).***

5. Conclusiones

El presente artículo es el resultado de una investigación acerca de las estrategias de resolución de problemas y su aplicación a problemas matemáticos del tipo exámenes de selección para acceder a la educación de nivel superior. Para esto, se hizo una revisión bibliográfica de diversos autores que de alguna manera se han involucrado en el diseño y análisis de estrategias para abordar problemas de matemáticas.

Además, se hizo una revisión de diversas guías de estudio enfocadas a la acreditación de evaluaciones de las principales instituciones educativas de nivel superior, esto con el fin de extraer los distintos modelos de problemas.

A los problemas extraídos le han sido aplicadas algunas de las principales estrategias definidas en su gran mayoría por uno de los principales autores de este ámbito, como lo es Pólya. Además, se dan algunas sugerencias y recomendaciones, así como soluciones para los problemas propuestos. Es pertinente comentar que por lo general estos problemas analizados no requieren de conocimientos especializados, y las mismas técnicas y estrategias que ejemplificamos con problemas elementales se pueden combinar para atender otros más avanzados.

En suma, se pretende que este trabajo sirva como apoyo a profesores que deseen preparar estudiantes en la solución de problemas, en particular problemas del estilo de examen de admisión de nivel profesional. Nos enfocamos a la resolución de problemas en el ámbito de pruebas de selección por el hecho de que estas son utilizadas como vehículos al servicio de otros objetivos curriculares tales como probar un dominio de conocimiento, manejo de conceptos, actitudes y procedimientos. El trabajo completo se puede revisar en la tesis de licenciatura que tiene el mismo título.

Referencias

[1] L. Galdós, *Consultor Matemático Álgebra, Aritmética y Geometría* CULTURAL S.A., Madrid, España.

[2] Baldor, *Álgebra y Aritmética* Publicaciones Cultural S.A. de C.V., México, 2006.

[3] *Guía EXANI II ingreso a la licenciatura.* CENEVAL México, 2005.

[4] *Guía de estudios para el examen de selección I.P.N.* México, 2005.

[5] *Centro Nacional de Evaluación para la Educación Superior A.C.* (CENEVAL) Guía de examen. México. Última Edición.

[6] Jesús Antonio Márquez Gallardo, *Guía de examen desarrollada para el ingreso a la licenciatura de la UNAM.* Guía parte I y parte II.

[7] José Alberto Romero Ascencio, Sergio Carrasco Romo, *Guía para aprobar tu examen de ingreso a la Universidad,* Mc Graw Hill. México, 2006.

[8] *Evaluación del ingreso a la educación superior tecnológica Consejo del Sistema Nacional de Educación Superior,* México, 2004.

[9] *Orientación para el examen de admisión para carreras profesionales Guía del estudiantes - Razonamiento matemático,* Dirección de Desarrollo e Integración Estudiantil, (DDIE-BUAP-THE COLLEGE BOARD), México, Puebla, 2009.

[10] G. Pólya, *Cómo plantear y resolver problemas,* (Traducción de: "How to solve it"), Pinceton University Press, U.S.A. Editorial Trillas. Ed. 1965.

[11] Romo, H., Delgado, V. y Terrazas, J. B., *Estrategias de Estudio UNAM,* Área Química 3, Educación secundaria, México: Ediciones Castillo, 2008.

[12] *Guía para preparar el examen de selección para ingresar a la licenciatura UNAM-2006,* Estrategias de estudio y Sugerencias para responder el examen de selección.
[13] Carmen Batanero, *Significados de la probabilidad en la educación secundaria,* http://www.uv.mx/eib/curso_pre/videoconferencia/51SignificadosProbabilidad.pdf. Revista Latinoamericana de Investigación en Matemática Educativa, noviembre, 2005. D.F., México.
[14] *Le Educación Matemática, El papel de la resolución de problemas en el aprendizaje,* Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales de la Universidad Nacional del Mar del Plata, Argentina. OEI-Revista Iberoamericana de Educación.
[15] George Pólya, *Estrategias para la solución de problemas,* I.E.S. Rosa Chacel, Departamento de Matemáticas. Http://ficus.pntic.mec.es/fheb0005/hojas_varias/material_de_apoyo/estrategias %20de %20polya.pdf
[16] Fernando García Fresneda, *A pie de aula − El papel del profesor en la resolución de problemas,* http://www.profes.net/newweb/mat/apieaula2.asp?id_contenido=33334.
[17] Francisco Javier Rodríguez Martínez, *Estrategias para Resolver Problemas de Matemáticas de Nivel Preuniversitario,* Tesis de Licenciatura, Facultad de Ciencias Físico Matemáticas, BUAP, 2010.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

lhernan@fcfm.buap.mx, arjuarez@fcfm.buap.mx, fj.roma@yahoo.com.mx

Capítulo 19

ACERCA DEL ABUSO DE LA PROPORCIONALIDAD POR ESTUDIANTES DE NIVEL MEDIO SUPERIOR

LIDIA AURORA HERNÁNDEZ REBOLLAR
ARACELI JUÁREZ RAMÍREZ
JOSIP SLIŠKO IGNJATOV
JOSUÉ VÁZQUEZ RODRÍGUEZ
FCFM - BUAP

Resumen. La proporcionalidad es un tema importante que recibe mucha atención en educación básica, media y media superior de las matemáticas; los estudiantes se enfrentan con frecuencia a problemas proporcionales del valor faltante, comúnmente llamados problemas de *Regla de Tres*, caracterizados por una igualdad entre razones. Los estudios han indicado que los estudiantes tienden fuertemente a aplicar métodos proporcionales a la solución de problemas del valor faltante, incluso a problemas donde el esquema es cuestionable o aun claramente inadecuado. En la presente investigación se utiliza un instrumento diseñado especialmente para detectar el abuso de la proporcionalidad, el cual fue aplicado a 100 estudiantes del tercer grado de preparatoria, con la intención además de describir y exponer el desenvolvimiento de las capacidades cognitivas, presentadas por dichos jóvenes, cuando se confrontan a un problema con dichas características.

1. Introducción

El abuso de métodos proporcionales en la solución de problemas ha sido documentado y discutido por varios eruditos en el campo, tanto el concepto de linealidad como sus características y representaciones [1].

El razonamiento lineal es una herramienta importante que se utiliza para interpretar los fenómenos del mundo real, incluso cuando los fenómenos son, en realidad, no lineales. Es esencial para entender, solucionar e incluso formular una amplia gama de los problemas matemáticos y científicos (elementales y avanzados).

La familiaridad cada vez mayor y la experiencia de los estudiantes con ideas lineales es importante, pero puede llevar a una tendencia a aplicar ciertos aspectos de linealidades dondequiera, incluyendo en situaciones donde no son aplicables en absoluto, los estudiantes se muestran seducidos frecuentemente por la idea de ver cada relación numérica como si fuese lineal, debido a la *Ilusión de linealidad*.

En resumen, utilizan métodos lineales más allá de su gama de aplicabilidad.

2. Antecedentes

El razonamiento lineal desempeña un papel primordial en la educación de las matemáticas. Observamos que los estudiantes viven a todo tiempo experiencias que se pueden caracterizar como lineales y que pueden establecer los fundamentos

para el razonamiento lineal. Wim Van Dooren en [2] menciona 3 momentos donde los estudiantes experimentan lo anterior:

- Desde los 6 ó 7 años hacen juicios lógicos y eficaces en situaciones lineales, por ejemplo se le dice que 1 coche tiene 4 ruedas, 2 coches tienen 8 ruedas, ellos, desde esa edad, deducen que para 3 coches se tendrán 12 ruedas.
- Los alumnos de segundo y tercer grado de primaria aprende a multiplicar y dividir y reconocer cuando aplicar estas operaciones en problemas del valor faltante, por ejemplo se le dice al estudiante si una caja de 12 lápices de colores costó 48 pesos, responder el costo de cada color.
- La noción de la proporcionalidad se introduce desde el cuarto, quinto grado de primaria, después se hace más formalmente. Desde entonces, enfrentan a los estudiantes con frecuencia a los problemas de proporcionalidad en un formato de los llamados problemas de *regla de 3*.

Posteriormente, con los fundamentos para un razonamiento lineal, los estudiantes tienen muchas ocasiones para practicar habilidades proporcionales en problemas de razonamiento más complejos:

- Durante la enseñanza secundaria la noción más abstracta de la función lineal se introduce, que es ahora en un plano cartesiano y aprendiendo a manipular la relación entre problemas de proporcionalidad directa y su representación en un plano.
- Luego los estudiantes encuentran varias relaciones proporcionales en las ciencias: entre el diámetro y el perímetro de un círculo, entre el tiempo y la distancia a velocidad constante, entre la masa y el volumen de una sustancia uniforme, o entre el voltaje y el amperaje en un circuito eléctrico.

2.1. Abuso de proporcionalidades en los problemas aritméticos pseudo-proporcionales del valor faltante y en patrones numéricos.

Existen problemas aritméticos a los que llamamos *Pseudo-proporcionales del valor faltante* para los cuales es claramente inadecuado un método proporcional. Es importante aclarar que los problemas de una pseudo-proporcionalidad son claramente solucionables, haciendo notar que las habilidades matemáticas de los estudiantes enfrentados a dichos problemas son las suficientes para resolverlo.

La carencia de resultados satisfactorios en este tipo de problemas por parte de los alumnos, pende de la presentación del problema, se llega pues a pensar que porque la mayoría de las tareas proporcionales del razonamiento encontradas en la escuela fueron indicadas en un formato de valor faltante, y al mismo tiempo los problemas no-proporcionales raramente fueron indicados en este formato, el problema carece de una solución fácil. Ello indica una comprensión superficial de dicho concepto, así como de las circunstancias bajo las cuales es aplicable. En lugar, aprendieron la asociación superficial, entre una formulación lingüística del problema y un procedimiento de la solución. Se concluye que parte del problema radica en que los libros de texto no acentúan suficientemente la capacidad de discriminar entre las situaciones proporcionales de las que no lo son.

En investigaciones previas, por ejemplo en [1], se ha demostrado que la tendencia a dar respuestas proporcionales a los problemas no proporcionales en una formulación de valor faltante *o de regla de 3* está presente antes de la instrucción formal en el razonamiento proporcional, asimismo el abuso de métodos proporcionales llega a ser más prominente cuando los problemas de la proporcionalidad del *valor faltante* son centrales en el curso, siendo que disminuye cuando estos problemas dejan de ser el tema a aprender.

Así, por ejemplo, en un primer curso de cálculo al solucionar problemas tales como *Si una planta mide 30 cm al principio de un experimento y su altura aumenta un 50 porciento mensual, ¿Cuánto medirá después de 3 meses?* En este problema el 62 por ciento de los estudiantes respondió de manera lineal con respecto al aumento de la altura en función del tiempo, en vez de considerar el carácter exponencial de este proceso de crecimiento [2].

La más frecuente de las respuestas erróneas observada en los problemas de patrones numéricos es debido a una asunción de proporción en vez de una determinación del razonamiento correcto. Esto se debe a que los patrones de respuesta confirman que los estudiantes están intuitivamente familiarizados con las relaciones que aplican proporcionalidad. Estas relaciones generalizadas se deben quizás a un malentendido anterior al estudio formal de los asuntos relacionados (cociente, ley distributiva).

3. METODOLOGÍA

Se llevó a cabo una investigación con 100 estudiantes de nivel medio superior en vísperas de presentar su examen de admisión a nivel licenciatura de la Benemérita Universidad Autónoma de Puebla (BUAP), en la cual el objetivo era percatarnos que aun a nivel medio superior existe entre algunos jóvenes la incapacidad de discernir entre ocupar un razonamiento proporcional o no. Esta investigación se basa en un instrumento que posee la caractetística de ser un problema aritmético pseudoproporcional del valor faltante.

3.1. Instrumento de investigación. Se diseñó un instrumento especial para esta investigación, el cual fue ideado para detectar el abuso de proporcionalidad con la intención además de describir y exponer el desenvolvimiento de las capacidades cognitivas presentadas por dichos jóvenes, cuando se confrontan a un problema con dicha característica.

El instrumento es el siguiente:

Juntando las mesas.

Para una fiesta se tienen disponibles mesas rectangulares. Alrededor de una mesa pueden sentarse 6 personas. Al juntar dos mesas, pueden sentarse 10 personas.

- *Al juntar 4 mesas, ¿Cuántas personas podrán sentarse?*
- *Al juntar 10 mesas, ¿Cuántas personas podrán sentarse?*
- *Al juntar 100 mesas, ¿Cuántas personas podrán sentarse?*
- *Al juntar N mesas, ¿Cuántas personas podrán sentarse?*

Se le solicita justificar sus respuestas con palabras, además de que en las primeras dos preguntas disponen de un espacio para realizar un dibujo para llegar a la respuesta, si lo creen necesario.

3.2. Análisis de resultados. En esta sección nos dedicaremos a analizar e interpretar los resultados obtenidos del instrumento de investigación.
Observamos en primera instancia que el patrón numérico que llega a resolver el problema de manera satisfactoria está dado por la función $\mathbf{F(n) = 4n + 2}$ (además de todo aquel patrón que sea equivalente al dado) donde $\mathbf{n}$ representa el número de mesas y $\mathbf{F(n)}$ el número de personas, y es a partir de dicha función que se llegan a escrutar las características y capacidades cognitivas de los estudiantes.

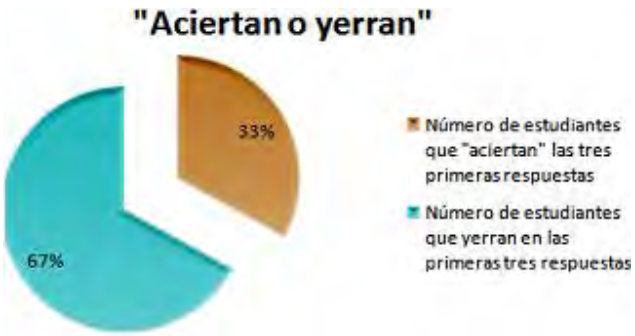


FIGURA 1. *Clasificación de los estudiantes de acuerdo a sus 3 primeras respuestas*

El primer gráfico es para distinguir aquellos estudiantes que logran responder hasta la pregunta 3 correctamente, con lo cual se desea mostrar que dos terceras partes de los estudiantes investigados tienen dificultades en aspectos cualitativos del problema. Deducimos de este hecho que los estudiantes tienen conflictos al momento de abordar ejercicios en donde no está presente una proporcionalidad directa, aunque el enunciado indicase posiblemente que sí hubiere (*La llamada pseudo-proporcionalidad*) y se presentan dichos conflictos desde la parte numérica y al momento de discriminar entre proporcionalidad o no.

Conforme a los que aciertan:

Este segundo gráfico corresponde a los 33 estudiantes que acertaron las tres primeras cuestiones, pero ahora apreciamos si fueron capaces de llegar a plantear un modelo para un número general de mesas. Nos percatamos que solamente 17 de los 100 estudiantes que elaboraron la prueba llegaron a solucionar por completo el problema de una manera correcta, además de que responder en la parte numérica de manera correcta no brinda la garantía de que estos estudiantes sean capaces de transportar sus respuestas a lo general.

Un rasgo primordial para evaluar una respuesta como correcta es la justificación de ésta, es ahí donde analizamos los tipos de argumentación que sostuvieron los estudiantes. Se decidió clasificar estas argumentaciones como válidas e inválidas, las primeras son aquellas que tienen una base matemática acertada, por ejemplo: *en cada mesa habrá cuatro personas a los lados y se sumarán solamente 2 que*

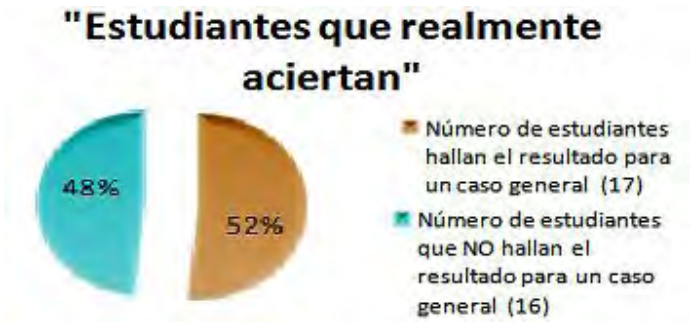


FIGURA 2. *Clasificación de los estudiantes que acertaron al hallar el caso de* **N** *mesas*

FIGURA 3. *Clasificación de los 33 estudiantes que acertaron de acuerdo a su tipo de justificación.*

corresponden a las personas de las orillas e inválidas como aquellas que no le brindan ningún aporte para llegar a la respuesta o que descansan en una tesis matemática inválida, por ejemplo: *pues depende de cómo vayan llegando a sentarse* o *el producto de mesas por 6 disminuye de acuerdo al número de mesas, es decir si hay dos mesas se tendrá 6 por 2 y disminuye 2, si hay 10 mesas entonces se tendrá las 10 mesas por 6 personas menos 10, que es el número de mesas y el resultado obtenido será el número de personas que andamos buscando*(figura 3).

La esquematización del problema es uno de los aspectos a analizar de los estudiantes investigados. Percibimos todo tipo de dibujos, van desde aquellos completamente realistas hasta los más abstractos donde, para representar a las personas, emplean números, aunque cabe decir que solamente 7 de los alumnos fueron los que recurrieron a este tipo de representación. Resaltamos el hecho que encontramos una relación entre dibujos abstractos y buenos resultados en la solución,(figura 4).

Ahora analizando los 67 personas que erraron en las 3 primeras respuestas:

Este quinto gráfico (figura 5) está presente debido a que deseamos manifestar el número de estudiantes que emplean un razonamiento lineal proporcional, entendido como aquel de regla de 3, a partir de su respuesta y justificación, por ejemplo, al

FIGURA 4. *Clasificación de los estudiantes que acertaron de acuerdo a la elaboración de dibujos*

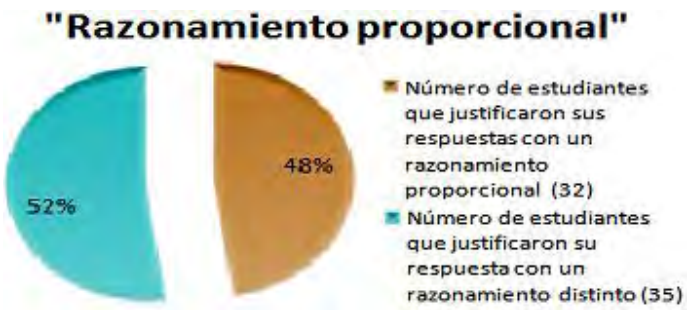


FIGURA 5. *Clasificación, de acuerdo a su justificación, de los estudiantes que se equivocaron*

decir *si en dos mesas caben 10, en cuatro cabrán 20*. Observamos que aproximadamente de 1 de cada 3, a nivel medio superior, aún no son capaces de discernir entre problemas proporcionales y no proporcionales.

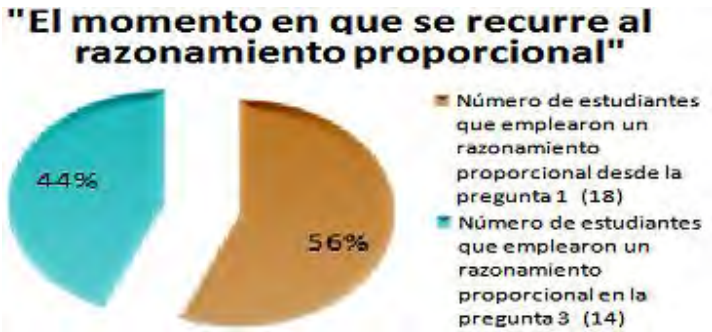


FIGURA 6. *Clasificación de los estudiantes que emplearon un razonamiento proporcional*

Observamos que al momento de ocupar un razonamiento proporcional hay una distinción entre aquellos que desde el inicio tienen dudas de cómo plantear el problema y es por ello que recurren a un razonamiento proporcional y aquellos estudiantes

que al no poder disponer de los dibujos, en el caso de la pregunta de las 100 mesas, recurren a un razonamiento de regla de 3. Esto lo observamos al momento de su justificación en la tercera pregunta: *Al caber 42 personas en 10 mesas, cabrán 420 en 100 mesas*, (figura 6).

FIGURA 7. *Clasificación de los estudiantes, de acuerdo a la elaboración de sus dibujos, que utilizaron un razonamiento proporcional desde la pregunta 1*

La figura 7 nos refleja la clara confusión de los estudiantes, que recurren a un razonamiento proporcional desde la pregunta 1, debido a que no son capaces de concebir un dibujo; sabemos que existen muchos problemas que no pueden respaldarse en un dibujo, pero este ejercicio es uno en los cuales un esquema es de gran ayuda para deducir la respuesta.

Analizando a los 14 estudiantes restantes, nos percatamos que elaboran sus dibujos ante las primeras preguntas, y que es un hecho imprescindible el apoyarse en un esquema para poder responder de manera acertada, y al no disponer de un dibujo en la tercera cuestión, titubean en su razonamiento y recurren al que intuitivamente está ahí presente: **un razonamiento de regla de 3**. Además diremos que dichos dibujos tienen características menos abstractas que los de aquellos estudiantes que acertaron en la respuesta. Consideramos que una característica para lograr acertar la respuesta es el hecho de no apoyarse de manera total en un dibujo, y por ende, el dibujar de manera que solamente se abstraiga la información necesaria para hallar la respuesta(Figura 8).

Por último en este análisis consideraremos aquellos que se equivocaron por una razón distinta de un razonamiento proporcional. Los clasificamos de igual manera que en el caso de los que acertaron:

- Una justificación válida como por ejemplo: *La relación que se va obteniendo es la tabla de 4 pero recorrida dos números* e
- Inválidos como por ejemplo *pues dependen qué tan amontonados estén.*

Observamos que es muy reducido el número de estudiantes que a pesar de haber justificado de manera idónea no logran dar una respuesta correcta (solamente 7) (*ver Figura 9*).

Analizando aquellos que optan por argumentos distintos a un razonamiento proporcional está presente una tendencia a basarse en ideas tales como *depende qué tan delgadas estén las personas* o *¿son niños o adultos?*; justificando de esta manera está claro que se presentarán mayores dificultades para concebir una respuesta para

FIGURA 8. *Clasificación de los estudiantes, de acuerdo a la elaboración de sus dibujos, que utilizaron un razonamiento proporcional a partir de la pregunta 3*

FIGURA 9. *Clasificación de los estudiantes, de acuerdo su tipo de argumentación, que erraron por una razón distinta a un razonamiento proporcional.*

nuestro problema.

Se ha observado que los estudiantes seleccionan y utilizan un método proporcional de una manera intuitiva. Los estudiantes optaron inmediatamente por éste, y encontraron muy difícil justificar su opción (*alrededor de una quinta parte de los estudiantes no justifican sus respuestas*).

Muchos demostraron deficiencias particulares con su conocimiento geométrico (*al momento de idear las mesas*) y tienen actitudes y creencias inadecuadas para solucionar los problemas matemáticos del valor faltante (*por ejemplo la creencia que la primera solución es siempre la mejor*) y un tedio para utilizar demás recursos (*abordando un problema no trivial primero haciendo un bosquejo o un dibujo*).

4. CONCLUSIONES

Existe una problemática en la resolución de ejercicios del valor faltante: Alrededor de una tercera parte de estudiantes emplea un razonamiento proporcional aún cuando este tipo de razonamiento no es aplicable a nuestro problema, provocando que no acierten la solución.

Las justificaciones con palabras de las respuestas nos indican que los estudiantes están altamente familiarizados en el empleo de razonamiento de *Regla de Tres* para resolver este tipo de problemas y es por ello que yerran, pero debemos hacer hincapié que la resolución del problema no rebasa las capacidades de los estudiantes, es decir, los jóvenes a este nivel de estudios poseen las herramientas cognitivas suficientes para resolver este problema de manera satisfactoria.

Debido a ciertas situaciones (sabemos que en ciertos momentos en el plan de estudios de matemáticas existe una clara extensión para el asunto referente a la proporcionalidad; que los problemas no proporcionales, indicados en un formato de valor faltante, son escasos en los libros de texto y el hecho de que las proporcionalidades se encuentran en las experiencias más comunes y frecuentes de los estudiantes) está provocando la incapacidad de emplear el razonamiento idóneo para problemas del valor faltante, incapacidad que perdura a través de los años, conllevando a un abuso del razonamiento proporcional aun a nivel medio superior.

Referencias

[1] VAN DOREN, Wim. *Not everything is proportional: Effects of age and problem type on propensities for Overgeneralization*, Cognition and instruction, 23(1), 57-86, 2005.

[2] VAN DOOREN, Wim. DE BOCK, Dirk. EVERS, Marleen y VERSCHAFFEL, Lieven. *Students´ overuse of proportionality on missing-value problems: How numbers may change solution.* Journal for research in Mathematics education, 40 (2), 2008.

[3] VAN DOOREN, Wim. DE BOCK, Dirk. JANSSENS, Dirk y VERSCHAFFEL, Lieven. *The linear imperative: An inventory and conceptual analysis of students´ overuse of linearity*, Journal for research in Mathematics education, 39 (0), 2008.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

lhernan@fcfm.buap.mx, arjuarez@fcfm.buap.mx, jslisko@fcfm.buap.mx,
200820897@alumnos.fcfm.buap.mx

Capítulo 20

FÍSICA Y MATEMÁTICA DESDE ARQUÍMEDES

RAÚL LINARES GRACIA
JUAN ARMANDO REYES FLORES
FCFM - BUAP

Resumen. Presentamos un ejemplo de cómo trabajaba Arquímedes y de los conceptos físicos involucrados al resolver un problema matemático como es el de la cuadratura de la parábola.

1. Introducción

La genialidad de la mente humana es maravillosa sin duda alguna, y para muestra de ello hay que dar un vistazo a la historia, voltear a ver a esos grandes personajes que aportaron grandes conocimientos al mundo. Fijemos nuestra atención en la antigua Grecia, siendo aun más precisos, en la ciudad de Siracusa. Ahí vivió Arquímedes cuya genialidad ha sorprendio ha sus conteporáneos y generaciones siguientes.

La manera en que Arquímedes fue desarrollando sus ideas para concluirlas en conocimientos concretos es digna de mencionarse y analizase, un ejemplo es aquel pasaje cuando en un momento de relajación en un baño público descubre la manera de como comprobar un posible fraude que le estarían haciendo a su amigo el rey Hierón y sale gritando del lugar aquella frase famosa: *¡Eureka, Eureka!* Arquímedes también inventó numerosas e ingeniosas máquinas para mantener a distancia a los romanos que trataban de invadir la ciudad como catapultas de alcance variable, una máquina que incendiaba los barcos romanos, pértigas móviles que proyectaban masas considerables sobre los barcos situados demasiado cerca de los muros de la ciudad. Y no nada más en aplicaciones muestra su talento sino también en la teória, como muestran sus principales obras. En las que da una contribución original a las matemáticas de su época siendo otra prueba de su genialidad.

En este trabajo analizaremos su obra *El Método* en la cuál explica a Eratóstenes como llegaba a tales resultados dando un ejemplo de como utilizó conceptos físicos para llegar a esos resultados pero antes daremos un breve esbozo de su vida y sus obras principales.

2. La vida de Arquímedes

Desafortunadamente no se cuenta en especifico con una biografía de Arquímedes ya que Heracleides escribió su biografía pero se perdió y por eso hay que recurir a varias fuentes antiguas de veracidad irregular que nos dan pasajes de la vida de Arquímedes. Tzetzes (gramático bizantino) refiere que Arquímedes murió a la edad de setenta y cinco años; puesto que murio en la caída de Siracusa en el año 212 a.C. entonces se deduce que nació hace el 287 a.C. Del historiador griego Diodoro sabemos que estudió en la universidad de Alejandría, posiblemente con los sucesores de Euclides.

Tzetzes escribió en su *Libro de historias* (El bizantino del siglo XII, Juan Tzetzes, conservó en su libro un gran tesoro de detalles literarios, históricos, teológicos y científicos, pero se ha de usar con precaución) sobre Arquímedes lo siguiente:

"Arquímedes el sabio, el famoso constructor de máquinas, fue de raza siracusana, y trabajó con la geometría hasta la vejez, llegando a los setenta y cinco años. Dominó muchas fuerzas mecánicas y, con un aparato de tres poleas arrastró una nave de cincuenta mil medimnos de peso, con sólo la mano izquierda. Cuando el general romano Marcelo asaltaba Siracusa por tierra y por mar, levantó con sus aparatos algunos buques de carga, los suspendió sobre la muralla de Siracusa y los lanzó amontonados a la profundidad, con su tripulación. Después de retirar Marcelo sus barcos a distancia, el anciano proporcionó a los siracusanos la ma-nera de levantar piedras tan grandes como un carro y, disparándolas continuamente, hundir los navíos. Al retirarlos Marcelo a un tiro de arco, el anciano construyó una especie de espejo hexagonal y puso espejos más pequeños, tetragonales, a distancias proporcionadas al tamaño del espejo, los movió con cuerdas y bisagras, constituyendolo en centro de los rayos del Sol, los rayos de mediodía, tanto en verano como en invierno. Después, al reflejarse los rayos en el espejo, provocó en los barcos un terrible incendio, convirtiéndolos en ceniza a un tiro de arco de distancia. De esta manera superó el anciano a Marcelo con sus armas. Y en su dialecto dórico de Siracusa, dijo: Si tuviera sitio para apoyarme, movería con mi caristión toda la Tierra."

Plutarco en su obra *Marcelo* cuenta la dramática historia de la muerte de Arquímedes tras la caida de Siracusa:

Dicen, que él, viendo desde arriba la ciudad bella y espaciosa, derramó muchas lágrimas, al pensar con dolor en lo que iba pasar, cuando en poco tiempo cambiaría de aspecto y de forma, saqueada por la soldadesca. Marcelo permitió que se hiciera presa del dinero y de los esclavos. En cuanto a las personas libres, dio órdenes de que no fueran molestadas, ni se matara o ultrajara, o se esclavizara a ningún siracusano. Pero lo que apenó especialmente a Marcelo fue la desgracia de Arquímedes. Estaba solo, examinando una figura de geometría, y había entregado de tal forma su pensamiento y sus ojos a la contemplación, que no percibió la irrupción de los romanos ni la toma de la ciudad. De pronto, se le presenta un soldado y le ordena que le siga hacia Marcelo. Arquímedes no quiere hacerlo antes de acabar el problema y establecer la demostración. El soldado se irrita y, desenvainando la espada, lo mata. Otros dicen que el soldado se presentó directamente con la espada para matarlo y que Arquímedes, al verlo, le rogó y suplicó que esperara un poco para no dejar la investigación sin terminar y sin demostrar, pero que el soldado, sin preocuparse, lo mató. Se está de acuerdo, sin embargo, en que Marcelo se apartó del asesino con horror, como de un sacrílegio, y que buscó a los parientes para honrarlos.

3. Las principales obras

Los tratados de Arquímedes, que versan a la vez sobre geometría plana y geometría en el espacio, aritmética, mecánica, hidrostática y astronomía, estan siempre orientados hacia el descubriemiento de nuevos conocimientos y contiene material nuevo, resultado de aproximaciones originales y personales.

Arquímedes escribió más de diez obras que se clasifican en el orden siguiente:

1. *Primer libro de los equilibrios*, que trata de los centros de gravedad, los paralelogramos y los triángilos.
2. *Cuadratura de la parábola*, que incluye un prefacio dirigido a Dositeo y trata de la cuadratura de cualquier segmento parabólico, problema para el que ofrece dos soluciones: una mecánica y otra geometríca.
3. *Segundo libro de los equilibrios*, que trata de los centros de gravedad de los segmentos de parábola.
4. *Sobre la esfera y el cilindro I y II.* Los principales resultados contenidos en estas dos obras son:
   a) la superficie de una esfera es cuatro veces la del gran círculo;
   b) la superficie de un segmento de esfera es igual a un círculo de radio igual a un segmento de recta trazado desde el vértice del segmento al punto sobre la circunferencia de la base;
   c) si un cilindro está circunscrito a una esfera y su altura es igual al diámetro de la esfera, entonces, 1) el volumen y 2) la superficie (con las bases) del cilindro son una vez y media el volumen y la superficie respectiva de la esfera.

   Este resultado debió gustar de manera especial a Arquímedes, ya que pidió a los suyos que sobre su tumba representaran la figura de la esfera inscrita en el cilindro. Gracias a esta inscripción, en el año 75 a. C., Marco Tulio Cicerón, cuestor por entonces en Sicilia, pudo identificar la tumba de Arquímedes a pesar del estado ruinoso en que se encontraba.
5. *Sobre las espirales*, que concierne a la espiral de Arquímedes, la figura engendrada por un punto que se mueve con velocidad constante sobre una semirrecta, radio vector, que a su vez gira con velocidad angular constante alrededor de su origen ($f = r\theta$). El estudio de las tangentes y de las áreas barridas por el radio vector.
6. *Sobre los conoides y los esferoides*, que trata de los volúmenes barridos por las elipses y parábolas que giran alrededor de un eje de simetría y por las hipérbolas que giran alrededor de un eje transversal.
7. *Medida del círculo*, breve tratado que está compuesto por tres proposiciones; la primera demuestra la equivalencia de los problemas de la cuadratura y la rectificación; la segunda prueba que el círculo es los $\frac{11}{14}$ del cuadrado circunscrito si la longitud de la circunferencia es 3 veces el diámetro más un séptimo; por último, la última proposición afirma que el perímetro del círculo es menor que los $3\frac{1}{7}$ del diámetro, puesto que es superior a los $3\frac{10}{71}$ de este diámetro.
8. *Arenario*, que contiene un sistema de números grandes, concebido en un principio para que Arquímedes pudiese escribir un número superior al número de granos de arena necesarios para llenar todo el universo.
9. *los cuerpos flotantes, libros I y II*, que trata del equilibrio de un segmento de paraboloide de revolución que flota en un líquido, del principio hidrostático de Arquímedes, etc.
10. *Tratado del método*, en donde Arquímedes revela a Eratóstenes algunos de sus métodos de investigación, utilizados para descubrir varios de sus teoremas.

4. El método de Arquímedes

El manuscrito del *Método* de Arquímedes se encontró en 1906 por Heiberg, tuvo noticias del hallazgo en el convento del Santo Sepulcro de Constantinopla de un palimpsesto de contenido matemático. Un palimpsesto es un pergamino, en él, el primer texto escrito fue lavado para poder volver a escribir una nueva obra, en este caso un libro de oraciones de la iglesia ortodoxa. Examinando el texto con técnicas fotográficas, Heiberg descubrió que en el pergamino había escritas obras de Arquímedes que habían sido copiadas alrededor del siglo X. En sus 185 páginas estaban *Sobre la esfera y el cilindro*, *Sobre las espirales*, *La medida del círculo*, *Sobre el equilibrio de los planos* y *Sobre los cuerpos flotantes* además de la única copia de *El método*. Durante la primera Guerra Mundial el texto volvió a desaparecer, reapareciendo en 1998, en una de las célebres subastas de la Galería Christie y un coleccionista anónimo lo adquirió por dos millones de dólares y lo donó, para su cuidado, al Museo Walters de Baltimore.

La particularidad de este libro radica en el uso de la experimentación previa a la hora de resolver los problemas. Arqumedes en una carta a Eratóstenes lo expresa de la siguiente manera:

"Reconociendo, como digo, tu celo y tu excelente dominio en materia de filosofía, amén de que sabes apreciar, llegado el caso, la investigación de cuestiones matemáticas, he creído oportuno confiarte por escrito, y explicar en este mismo libro, las características propias de un método según el cual te será posible abordar la investigación de ciertas cuestiones matemáticas por medio de la mecánica.

Algo que por lo demás, estoy convencido, no es en absoluto menos útil en orden a la demostración de los teoremas mismos. Pues algunos de los que primero se me hicieron patentes por la mecánica, recibieron luego demostración por geometría, habida cuenta de que la investigación por ese método queda lejos de una demostración; como que es más fácil construir la demostración después de haber adquirido por ese método cierto conocimiento de los problemas, que buscarla sin la menor idea al respecto"

En otros términos, los procedimientos mecánicos sirven para proponer y explorar los teoremas y no para probarlos. Así, el método mecánico empleado para descubrir teoremas no proporciona las pruebas de los mismos, pruebas en las que se utiliza con frecuencia el método exhausivo. Las características importantes de este método exhausivo son esencialemente las siguientes:

Llamemos X a una figura plana o sólido cuya área o volumen se desconozcen. El método [3] consiste en pesar elementos muy pequeños (infinitesimales) de X comparándolos con los elementos correspondientes de una figura Y de la que se conocen el área, o el volumen y el centro de gravedad. Para conseguir el equilibrio mecánico, las figuras se disponen de un eje de tal manera que las figuras se encuentren en una misma recta. Entonces, si los elementos infinitesimales son seccionados de figuras obtenidas mediante planos paralelos, perpendiculares al eje común, que cortan a las figuras, los centros de gravedad de todos los elementos están en algún punto de este eje. Eje que se convierte en el brazo de una balanza. Estos elementos se ven entonces como líneas rectas (figuras planas) o áreas planas (sólidos).

El propósito de Arquímedes consiste en balancear los elementos de X, aplicándolos todos en un único punto de la palanca, mientras los de Y permanecen en su sitio. Y, como el centro de gravedad de Y, así como el volumen o su área son conocidos, se imagina Y como una masa que actúa sobre su centro de gravedad y por tanto,

si X e Y están situados en sus puntos respectivos, conocemos las distancias de los dos centros de gravedad al punto de aplicación de la palanca y tambien el área y el volumen de Y. Así, se calcula el área o volumen de X.

Ejemplo de su método, consideramos la proposición 1 del *Tratado del Método* [3]

PROPOSICIÓN. Sea ABC segmento de parábola acotado por el segmento de recta AC y BD el eje de simetría de la parábola (Figura 1). Entonces el área del segmento de parábola ABC es $\frac{4}{3}$ el área del triángulo ABC.

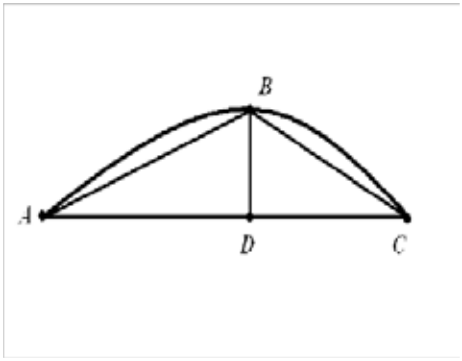


FIGURA 1. Segmento ABC

DEMOSTRACIÓN.

Trazamos la tangente del segmento parabólico en el punto C, la llamamos línea L_1. Trazamos una línea L_2 paralela a BD que pasa por A, la intersección de L_1 y L_2 es Z, consideramos a la línea AZ y ahora el segmento parabólico ABC se encuentra dentro del triángulo ACZ (Figura 2). Extendemos la línea DB hasta intersecar a la línea CZ en E y la línea CB hasta el punto K,

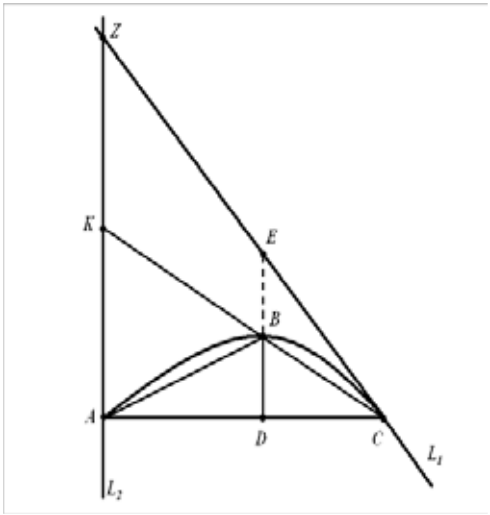


FIGURA 2. Triángulo ACZ

Prolongamos CK hasta T de tal manera que $CK = TK$. El segmento TC será el brazo de la palanca y K su punto medio. Sea MX una recta paralela a ED que corta en M, N, O, X, a la tangente, el segmento CK, la parábola y base respectivamente (Figura 3). $EB = DB$ y $AK = KZ$ (por ser la tangente y la semiordenada, esto se demuestra según Arquímedes en los elementos de las cónicas de Aristeo y Euclides).

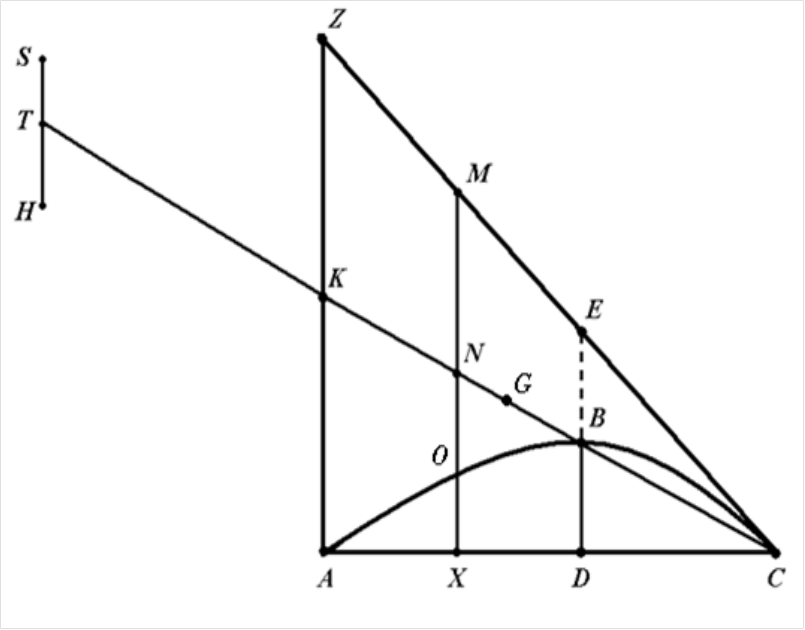


Figura 3

Ahora por la propiedad de la parábola probada en la proposición 5 de su libro cuadratura de la parábola

$$\frac{MX}{XO} = \frac{CA}{XA}$$

de donde

$$\frac{MX}{XO} = \frac{CK}{KN} = \frac{TK}{KN} \qquad \text{(Euclides VI, 2)}$$

Medir SH igual a XO y colocarla en su centro de gravedad en T, de manera que $ST = TH$, entonces puesto que N es el centro de gravedad de MX, se tiene $\frac{MX}{SH} = \frac{TK}{KN}$; se desprende que SH en T y MX en N están en equilibrio alrededor de K. (proposiciones 6 y 7 del primer libro de los equilibrios). Además K es el centro de gravedad de todo el sistema.

Puesto que el triángulo ACZ está constituido por todas las paralelas como MX, y el segmento CBA está constituido por todas las paralelas como OX bajo la curva limitada por AC, se desprende que el triángulo ABC está en equilibrio alrededor de K con el segmento CBA situado con su centro de gravedad en T.

Dividir KC en G de manera que $CK = 3KG$, entonces G es el centro de gravedad del triángulo ACZ.

Además

$$\frac{\Delta ACZ}{segmento \quad ABC} = \frac{TK}{KG} = 3$$

Además

$$segmento \quad ABC = \tfrac{1}{3}\Delta ACZ$$

Pero

$$AZC = 4\Delta ABC$$

Luego

$$segmento \quad ABC = \tfrac{4}{3}\Delta ABC$$

□

"Ahora bien, lo enunciado no está realmente demostrado por el argumento utilizado; pero este argumento nos ha proporcionado un indicio de que la conclusión es verdadera. Entonces viendo que el teorema no está demostrado, pero sospechando sin embargo que la conclusión es verdadera, debemos recurrir a la demostración geométrica que yo mismo he descubierto y que ya he publicado. [3]"

Los conceptos que Arquímedes usa para su exploración son los siguientes:

1. La teoría de las razones y las proporciones.
2. La teoría de los centros de gravedad.
3. La teoría de los equilibrios.
4. Métodos de comprensión y aproximación.

La demostración a la que se refiere Arquímedes de la cuadratura de un segmento de parábola se encuentra en la *Cuadratura de la parábola*, proposiciones 16 y 17. Sin embargo, en las siete últimas proposiciones de esta misma obra, Arquímedes realiza una demostración diferente, basada sobre todo en la existencia de una serie infinita de áreas, cada una de ellas cuatro veces mayor que la siguiente, y cuya suma es inferior al área del segmento de parábola. El método utilizado en esta demostración es el exhausivo y no el de la suma infinita de áreas de la forma

$$A + \tfrac{A}{4} + \tfrac{A}{4^2} + ... + \tfrac{A}{4^n} + ...$$

que es evidentemente igual a $\frac{4}{3}A$

Con este ensayo podemos ver que Arquímedes es uno de los matemáticos más importantes de todos los tiempos, fue un hombre práctico de sentido común, podemos decir que fue el Newton de su época, que poseía la habilidad imaginativa y la perspicacia para tratar la geometría, la mecánica y que sento las bases del cálculo integral. Y es facil ver porque se considera a Arquímedes el padre de la Física Matemática.

"*Quien comprenda a Arquímedes y Apolonio admirará menos los logros de hombres posteriores.*"
G. W. LEIBNIZ

Referencias

[1] Jean-Paul Collette, *HISTORIA DE LAS MATEMATICAS I*. Segunda Edición, Siglo Veintiuno Editores, México 1986.
[2] Santiago Gutiérrez, *El Método: una carta reveladora de Arquímedes a Eratóstenes*. Revista Suma, Vol. 53, Noviembre 2006, pp. 69-73
[3] T. L. Heath, *THE WORKS OF ARCHIMEDES WHIT THE METHOD OF ARCHIMEDES*. Editorial Dover, 1953.
[4] Josehp E. Hofmanm, *HISTORIA DE LAS MATEMATICAS I Tomo I*, UTEHA, México 1978.
[5] James R. Newman, *EL MUNDO DE LAS MATEMATICAS 1*. Novena Edición, Editores Grijalbo, Barcelona 1983.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

rlinares@fcfm.buap.mx, j.arm.rf@gmail.com

Capítulo 21

EL TRATAMIENTO DE LOS INFINITESIMALES SEGÚN L'HOSPITAL

RAÚL LINARES GRACIA
JOSUÉ VÁZQUEZ RODRÍGUEZ
FCFM - BUAP

Resumen. Presentaremos un ejemplo de cómo hallar la recta tangente a una curva, comparándolo con la forma de trabajar de Arquímedes 21 siglos antes, viendo así la influencia de este último.

1. Introducción

Nuestro cálculo integral y diferencial se ha desarrollado mediante los trabajos y aportaciones realizadas por grandes matemáticos de distintas épocas y lugares conforme al paso de los años.

Al hacer un análisis de cómo se dio la formación de esta rama de las matemáticas nos percatamos que las bases fueron ideadas por los griegos (resaltando la figura de Arquímedes), y dichas bases están presentes en el primer libro de cálculo de la humanidad: *Analyse des Infinitiment petits, pour líntelligence des lignes courbes* (Análisis de los infinitamente pequeños para el estudio de las líneas curvas) de Guillaume Francois Antoine de L'Hospital, mejor conocido como Marqués de L'Hospital.

Este primer libro apareció tan sólo 12 años después de la primera publicación del cálculo como tal, compara las diferencias infinitamente pequeñas de las magnitudes finitas y descubre las razones de estas diferencias. Aborda uno de los problemas más trascendentes para la geometría: El cálculo de tangentes a cualquier línea curva, dicha tesis se desarrolla a través del libro logrando resultados que hoy en día aún tienen vigencia, es por ello que se mostrará la manera en que el Marqués de L'Hospital trabajó con los infinitesimales considerando unas respectivas reglas y condiciones.

2. Antecedentes

Se remonta a la época helénica los fundamentos para el cálculo de áreas y volúmenes que dieron dirección a lo que hoy conocemos y aprendemos como cálculo integral y diferencial, dichos fundamentos están inscritos en el estricto dominio de la geometría:

1. Eudoxio al indagar el volumen del cono y la pirámide, brindó los primeros modelos para el cálculo de éstos (constatado por Euclides en sus *Elementos* -libro XII, prop. 7,10-).
2. Arquímedes, que desarrolla de manera rigurosa el llamado método de exhaución, en particular halla el área de la llamada espiral de Arquímedes, donde da una definición cinemática de la espiral (es decir, está describiendo una curva mediante el movimiento, como si tuviéramos una partícula en movimiento) permitiéndole calcular la tangente a la espiral, siendo éste el

único caso que se conoce como fuente del cálculo diferencial, además de la determinación relativamente fácil de las tangentes a las cónicas y de algunos problemas de máximos y mínimos entendidos en aquellos tiempos.

Posteriormente, los matemáticos del siglo XVII ante la multitud de problemas nuevos que se les planteaban buscaron en el estudio de los escritos de Arquímedes los medios de superarlo. Son en Galileo y Kepler, donde, primeramente, se manifiesta dicha influencia; a partir de ese momento y más o menos hasta 1670, no hay ningún nombre que aparezca más a menudo que el de Arquímedes en los escritos de los precursores a los fundadores del Cálculo infinitesimal: Descartes, Fermat Pascal, Roberval, Torricelli, J. Wallis.

Una herencia más para dicha época es que adquirieron el rigor en la forma de trabajo de Arquímedes: Fermat incluso se limita a no adelantar nada que no pueda ser justificado, condenándose a no enunciar ningún resultado general más que mediante alusiones o en forma de método.

Subsiguientemente, durante el siglo XVI se contaba con:

1. Un álgebra literal (escuela italiana).
2. Tablas de "funciones" trigonométricas (en esta época aún no se contaba con un concepto formal de función pero se tenía presente una idea intuitiva).
3. Elementos necesarios para el despeje de la teoría de ecuaciones algebraicas.
4. Un cálculo logarítmico
5. El método de los indivisibles de Cavalieri.

Es en el segundo tercio del siglo XVII cuando Francia se convierte en el principal centro de la actividad matemática encabezada por Descartes, Fermat, Desargues, Roberval y Pascal. Desarrollando la geometría analítica y las bases del cálculo diferencial e integral entre otros, resolvieron los problemas de diferenciación que se les presentaron para dicha época, como el de poder hallar un método para la pendiente de la recta tangente a la curva descrita por un circunferencia, dicho método emplea argumentos plenamente algebraicos.

René Descartes (1596-1650) publica en 1637 el *Discurso del Método* . El primer libro está dedicado a la Geometría, contiene sus ideas sobre la geometría de coordenadas y el álgebra (resolución de ecuaciones algebraicas, su método consistía en utilizar la representación gráfica de las ecuaciones), proporcionando una base geométrica al álgebra; no sólo a propósito de las tangentes, sino también a propósito de las velocidades, Galileo busca la ley de las velocidades en la caída de los cuerpos, Descartes hace lo mismo dentro del lenguaje de los indivisibles. En ambos el papel principal corresponde a la gráfica de la velocidad. En el libro II de la Geometría aparece el método de la tangente, siendo éste un estudio completamente algebraico y no recurre a conceptos de límite o infinitésimo. Sin embargo, queriendo corregir la regla de máximos y mínimos de Fermat, utiliza un procedimiento que es prácticamente equivalente a definir la tangente como límite de una secante.

P. Fermat (1601-1665) se propone someter la teoría de los lugares geométricos a un análisis que indique el camino hacia un estudio general de los problemas de lugares. *Cuando una ecuación contiene dos cantidades desconocidas, hay un lugar correspondiente, y el punto extremo de una de estas cantidades describe una línea recta o una línea curva*. Fermat está introduciendo no sólo la geometría analítica,

sino la idea de variable algebraica. Además contribuyó al cálculo diferencial e integral (estudio de los máximos y mínimos), a la geometría analítica, teoría de números y teoría de probabilidades.

Tenemos que Descartes y Fermat se dedicaron al estudio del mismo problema con enfoques distintos. Mientras que Descartes estudia el lugar geométrico a partir del cual obtiene una ecuación del lugar, Fermat parte de la ecuación para deducir las propiedades de su curva.

Gilles Personne de Roberval (1602-1675) consideraba que la materia podía dividirse infinitamente en partículas cada vez más pequeñas, sin poder afirmar un fin a esta descomposición. Esta concepción, transportada al campo de las matemáticas, caracteriza su método de los indivisibles. Con su método logró integrar funciones de una potencia entera, cuadrar todas las parábolas en el infinito, la integral de la función seno. Además atacó el problema de la tangente a una curva. El método de los indivisibles de Roberval difiere del de Cavalieri ya que este último consideraba que la descomposición de la materia era limitada, estando compuesta de partículas indivisibles.

E. Torricelli (1608-1647) conocía bien los trabajos de Arquímedes y el método de los indivisibles de Cavalieri y no satisfecho de las demostraciones de éste, elaboró otras pruebas a la manera de Arquímedes (método de exhaución) como suplemento a las demostraciones. Contribuyó de manera original al desarrollo del cálculo integral demostrando que la rotación de sólidos de longitud infinita engendraba un volumen finito (aunque Oresme, Fermat y Roberval habían previsto resultados semejantes). También se interesó por las tangentes, su método se basa esencialmente en una concepción dinámica de la tangente, la que recuerda a la tangente a la espiral de Arquímedes (consideró las curvas generadas, por un punto que se mueve a lo largo de una línea en notación uniforme).

J. Wallis (1616-1703): Su contribución fue la aritmetización de los indivisibles asignándole valores numéricos, convirtiendo el cálculo de áreas en cálculo aritmético (algo más que un primitivo paso al límite). Propone una genealogía del cálculo:

1. Método de exhaución (Arquímedes).
2. Método de los indivisibles (Cavalieri).
3. Aritmética de los infinitos (Wallis).
4. Método de series infinitas (Newton).

Además de los infinitésimos, cada vez se usan más fórmulas y menos dibujos, y así fueron creados de forma geométrica y aritmética los principios del cálculo diferencial e integral.

Es en el último tercio del siglo XVII que se produce el descubrimiento del cálculo diferencial e integral en el sentido propio de la palabra.

Para I. Newton (1642-1727) el cálculo de fluxiones era simplemente el reflejo de la idea de que las leyes elementales de la naturaleza se expresaban por ecuaciones diferenciales, y la obtención de resultados que describen estos procesos mediante las ecuaciones requiere de su integración.

Mientras que para G.N. Leibniz (1646-1716) el cálculo se forma bajo las siguientes premisas:

1. Problema de sumatoria de series y la utilización de los sistemas de diferencias finitas.

2. Resolución de problemas sobre tangentes, el triángulo de Pascal y el paso gradual de las relaciones entre elementos finitos a arbitrarios y después infinitesimales.
3. Problemas inversos de tangentes, sumatoria de diferencias infinitamente pequeñas, descubrimiento de la inversibilidad mutua entre los problemas diferenciales e integrales.

3. L'Hospital y el primer libro de texto del cálculo diferencial

Las primeras publicaciones donde Leibniz presenta su cálculo, ideado nueve años antes, son dos breves artículos en el *Acta Eruditorum* -revista científica y literaria de Alemania- que datan de 1684 y 1686, los cuales llegan a ser del conocimiento de los hermanos Jacques y Jean Bernoulli, son dichos hermanos los primeros en contribuir al desarrollo del cálculo creado por Leibniz (alrededor del año 1700, junto con Leibniz, ya habían creado, casi en su totalidad, lo que hoy se conoce como cálculo diferencial e integral elemental).

Es en el año de 1691 cuando se conocen Jean Bernoulli y Guillaume Francois Antoine de L'Hospital, matemático descendiente de una familia aristocrática, se dice que mostró talento matemático desde muy temprana edad.

A partir de finales de 1691, L'Hospital contrató a Bernoulli para que le explicara el nuevo cálculo: El compromiso incluía tener que entregarle por escrito una lección en cada sesión a cambio de un buen salario, además de que todo aquel nuevo descubrimiento de Jean sólo debía ser comunicado al Marqués. Literalmente se puede decir que el cálculo infinitesimal pensado por Leibniz, y a cuyo desarrollo contribuyeron prominentemente los hermanos Bernoulli, fue vendido al Marqués de L'Hospital por Jean Bernoulli, incluyendo los nuevos logros de éste.

Al dominar ya el nuevo cálculo, L'Hospital publica el primer libro de texto sobre esta materia: *Analyse des Infinitiment petits, pour líntelligence des lignes courbes* (Análisis de los infinitamente pequeños para el estudio de las líneas curvas) en el año de 1696, tan sólo 12 años después de la primera publicación de Leibniz.

El libro es una introducción a la Geometría de los infinitamente pequeños. No contiene ejercicios o problemas para que el lector los resuelva: Todos los ejemplos están resueltos. Sigue el estilo clásico de la estructura axiomática de Euclides (y Arquímedes), -este hecho se puede interpretar como un intento por formalizar el concepto intuitivo de cantidades infinitesimales y las operaciones que regían su uso-.

Es en los años posteriores cuando se presentó un debate en *La Academia de Ciencias de París* sobre la admisibilidad lógica del nuevo cálculo, se dice que incluso para Leibniz el problema era asegurar el uso confiable de las cantidades infinitesimales y no su existencia. Fueron los resultados obtenidos por el cálculo lo que impulsaron su avance y aceptabilidad, algunos presentados en *Analyse des inifiment petits* y a partir de él se continuaron las investigaciones.

4. Análisis de los infinitamente pequeños para el estudio de líneas curvas

El análisis presentado en el libro compara las diferencias infinitamente pequeños de las magnitudes finitas, descubre las razones de estas diferencias, se puede decir que este análisis se extiende más allá de lo infinito, pues no se limita a las diferencias infinitamente pequeãs, sino que descubre las razones de las diferencias de sus diferencias; más aún aquéllas de las diferencias terceras, cuartas y así sucesivamente, de

tal manera que no sólo abarca el infinito, sino el infinito del infinito, o una infinidad de infinitos.

Se retoman ideas de los matemáticos de la era helénica, principalmente de Arquímedes:

"Hicieron lo que nuestras mejores mentes hubieran hecho en su lugar, y si hubieran estado en nuestra época, se podría creer que tendrían las mismas concepciones que nosotros. Todo eso es una continuidad de la igualdad natural de las mentes y de la sucesión necesaria de los descubrimientos."(Marqués de L'Hospital)

En el prefacio El Marqués elabora una cronología acerca de las matemáticas donde resaltan algunos puntos:

- Se sorprende que grandes hombres (los llega a igualar con los antiguos) hayan permanecido tanto tiempo sin avanzar más y que, por una admiración casi supersticiosa por sus obras, se hayan contentado con leerlas y comentarlas, sin permitirse otra utilización de su genio que la que se necesitaba para comprenderlos, sin atreverse a cometer el crimen de pensar alguna vez por sí mismos, y de llevar su concepción más allá de la que los antiguos habían descubierto.
- Comenta que se continuó pensando y actuando de ésa manera hasta las aportaciones de René Descartes, un "rebelde"capaz de abandonar las ideas de sus tiempos y comenzar donde los antiguos griegos habían terminado.
- Abarca su opinión acerca de Pascal, Fermat, Barrow, Leibniz y los hermanos Bernoulli, así como una interpretación propia de las obras de éstos personajes.

.

Describe además al libro como tal, donde específica que está dividido en 10 secciones de acuerdo a la evolución que va obteniendo el desarrollo de los infinitesimales.

4.1. SECCIÓN 1: Donde se dan las reglas del cálculo de las diferencias.

El Marqués decide partir de dos definiciones:

4.1. Definición. Se llaman cantidades variables aquéllas que aumentan o disminuyen constantemente, y por lo contrario, cantidades *constantes* las que permanecen siendo las mismas mientras otras cambian.

4.2. Definición. La parte infinitamente pequeña en la que la cantidad variable aumenta o disminuye continuamente se llama *Diferencia*.

Ambas definiciones las ejemplifica para que se pueda obtener un mayor entendimiento, consideremos el ejemplo que brinda para su segunda definición:

Sea AMB, por ejemplo, una línea curva cualquiera que tiene por eje o diámetro a la línea AC y como una de sus ordenadas a la recta PM (Fig. 1), y sea pm otra ordenada infinitamente cercana a la primera. Admitido eso, si se trazan MR paralela a AC, y las cuerdas AM y Am, y luego se describe, con centro en A y radio AM, el pequeño arco de círculo MS, entonces:

Pp es la diferencia de AP, Rm la diferencia de PM, Sm la diferencia de AM, $\widehat{\text{M}}\text{m}$ la diferencia del arco AM, el triángulo Amm que tiene por base el arco Mm *será* la diferencia del segmento AM, y el espacio $Mppm$ será la diferencia del espacio comprendido por las rectas AP y PM y el arco AM.

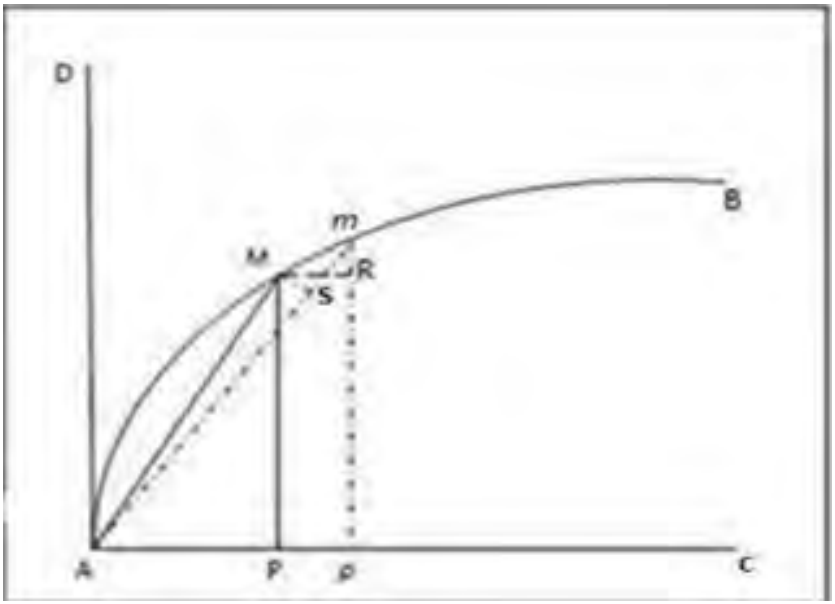


FIGURA 1

4.3. COROLARIO. Es evidente que la diferencia de una cantidad constante es nula o cero, (lo cual es lo mismo) que las cantidades constantes no tienen diferencia.

Advierte que hará uso de la notación ∂ para denotar la diferencia de una cantidad variable, denotará por las primeras letras del alfabeto a, b, c, etc. a las cantidades constantes y por las últimas letras del alfabeto x, y, z, etc. a las cantidades variables. Y también considera los siguientes postulados:

1. POSTULADO: Se pide tomar indistintamente una por la otra a dos cantidades que no difieran entre sí más que por una cantidad infinitamente pequeña.
2. POSTULADO: Se pide que una línea curva pueda ser considerada como el ensamblaje de una infinidad de líneas rectas, cada una infinitamente pequeña, o (lo cual es lo mismo) como una poligonal de un número infinito de lados, cada uno infinitamente pequeño, los cuales determinan, por los ángulos que forman entre sí, la curvatura de la línea.

Posteriormente el Marqués plantea 4 proposiciones que son problemas resueltos y a partir de las cuales define las reglas para el cálculo de las diferencias:

- REGLA 1: *Para la adición o sustracción de cantidades.* Se tomará la diferencia de cada término de la cantidad propuesta y conservando los mismos signos, se compondrá otra cantidad que será la diferencia buscada. (Así por ejemplo la diferencia de $a + x + y - z$ será $\partial x + \partial y - \partial z$)
- REGLA 2: *Para las cantidades multiplicadas.* La diferencia del producto de varias cantidades multiplicadas entre sí es igual a la suma de los productos de la diferencia de cada una de estas cantidades por el producto de las otras.(Así por ejemplo la diferencia de $(a+x)(b-y)$ será $b\partial x - y\partial x - a\partial y - x\partial y$)

- REGLA 3: *Para las cantidades divididas, o para las fracciones.* La diferencia de una fracción cualquiera es igual al producto de la diferencia del numerador por el denominador, menos el producto de la diferencia del denominador por el numerador, el total dividido entre el cuadrado del denominador. (Así la diferencia de $\frac{a}{x}$ será $\frac{-a\partial x}{x^2}$)
- REGLA 4: *Para las potencias perfectas o imperfectas.* La diferencia de una potencia cualquiera, perfecta o imperfecta, de una cantidad variable es igual al producto del exponente de esta potencia por esta misma cantidad elevada a una potencia menor en una unidad, y multiplicada por su diferencia. (Así la diferencia de x^m será $mx^{m-1}\partial x$)

Cabe decir que para ir deduciendo las reglas del cálculo de las diferencias es a partir de las reglas anteriores, por ejemplo de la regla del producto logra deducir la regla del cociente y la regla de la potencia, solamente ajustándola a ciertos detalles, los cuales se describen en los problemas resueltos que anteceden a las reglas enunciadas.

Así se concluyen las reglas para el cálculo de las diferencias y con ello la primera sección de su libro, posteriormente ejemplificará que a través de dichas reglas se pueda hallar la tangente a cualquier curva.

4.2. SECCIÓN 2: Uso del cálculo de las diferencias para encontrar las tangentes de todos los tipos de líneas curvas. El Marqués considera imprescindible el hecho de definir qué se entenderá por tangente debido a que es un concepto que se ocupa en el transcuso de su libro:

4.4. Definición. Si se prolonga uno de los infinitésimos lados Aa de la poligonal que compone cualquier curva, este infinitésimo lado, así prolongado, será llamado la tangente de la curva en el punto A o a.

Posteriormente se propone el problema que dará el desarrollo del resto del libro:

4.5. Problema. Sea AM una línea curva tal que la relación de la abscisa AP con la ordenada PM esté expresada por una ecuación cualquiera, se requiere trazar la tangente MT por el punto M dado sobre está curva. (Fig. 2)

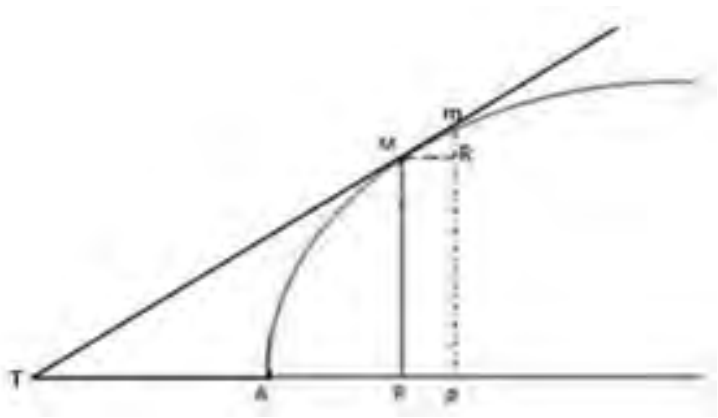


Figura 2

Daremos la solución del problema tal como la expresó el Marqués de L'Hopital:

Habiendo trazado la ordenada MP, y suponiendo que la recta MT que intersecta el eje en el punto T sea la tangente buscada, se concebirá otra ordenada mp infinitamente cercana a la primera, con una pequeña recta MR paralela a AP. Y al denominar a AP, x, y a PM, y, (y que están dadas luego $Pp = MR = \partial x$ y $RM = \partial y$), los triángulos semejantes mRM y MPT darán:

$$\frac{mR}{RM} = \frac{MP}{PT} \qquad \text{o} \qquad \frac{\partial y}{\partial x} = \frac{y}{PT} \qquad \text{o} \qquad PT = \frac{y\partial x}{\partial y}$$

Luego, por la diferencia de la ecuación dada, se encontrará un valor de ∂x en términos que estarán afectados por ∂y, el cual al ser multiplicado por y y dividido entre ∂y dará un valor de la sub-tangente PT, en términos completamente conocidos y libres de diferencias, el cual servirá para trazar la tangente buscada MT.

Observamos que el Marqués de L'Hopital realiza su demostración a partir de las ideas de Arquímedes de tomar la curva como una poligonal de una infinidad de lados que difieren entre ellos sólo por la diferencia de los ángulos que estos lados infinitamente pequeños forman entre sí; y del uso de los infinitésimos de tal manera como se empleaban en las matemáticas griegas clásicas, además de retomar ideas euclidianas como las razones y proporciones que brindan los triángulos semejantes.

Ahora analizaremos que a partir de la unión de sus reglas de cálculo de diferencias y con la solución del problema anterior logra hallar la tangente en cualquier punto de cualquier curva.

4.6. Ejemplo. Si se quiere que $ax = y^2$ (fig. 3) exprese la relación entre AP a PM, la curva AM será una parábola, al tomar las diferencias de una y otra parte se tendrá:

$$a\partial x = 2y\partial y \qquad \partial x = \frac{2y\partial y}{a} \qquad \text{y} \qquad PT = \frac{y\partial x}{\partial y} = \frac{2y^2}{a}$$

o, al sustituir a y^2 por su valor ax , $PT = 2x$, de donde se sigue que si se toma PT igual al doble de AP, y se traza la recta MT, ésta será tangente en el punto M. Lo cual se había propuesto.

4.7. Ejemplo. Sea $a^2 = xy$ (fig. 4) *la ecuación que expresa la naturaleza de la hipérbola* se tendrá al tomar las diferencias de ambos lados:

$$x\partial y + y\partial x = 0 \qquad \partial x = -\frac{x\partial y}{y} \qquad y \qquad PT = \frac{y\partial x}{\partial y} = -x$$

de donde se sigue que si se toma $PT = PA$ del lado opuesto del punto A, y se traza la recta MT, ésta será la tangente en M.

5. Conclusión

En esta primera elaboración del análisis infinitesimal está presente ideas de la geometría clásica griega, donde la más sobresaliente es el uso de los infinitésimos de manera tal como Arquímedes los empleó en su método de exhausción.

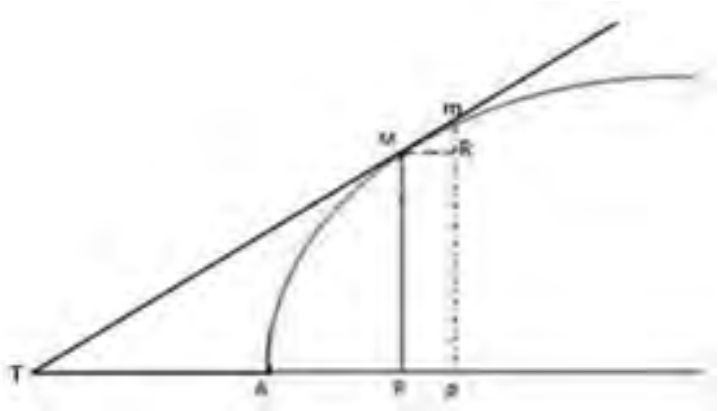


Figura 3

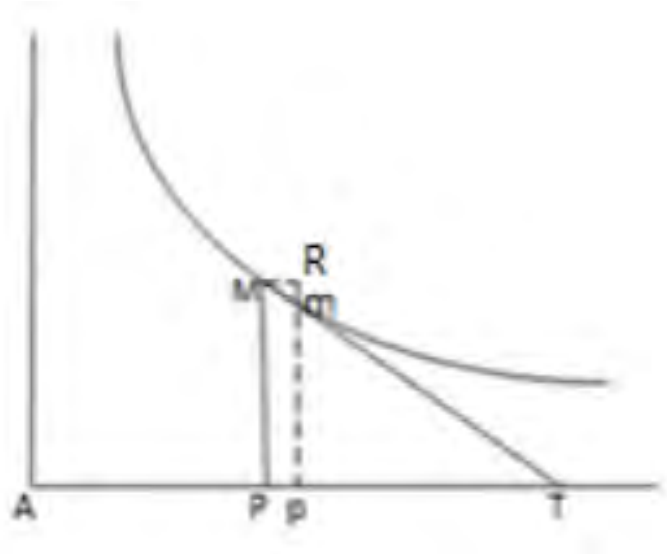


Figura 4

El *Análisis de los infinitamente pequeños para el estudio de las líneas curvas* es la combinación ideal de los avances que se lograron a través de los trabajos de Descartes, Fermat, Leibniz, los hermanos Bernoulli y toda la base estructurada matemática ideada por los griegos de la época helénica.

Referencias

[1] L'Hospital, Guillaume, *ANÁLISIS DE LOS INFINITAMENTE PEQUEÑOS PARA EL ESTUDIO DE LAS LÍNEAS CURVAS*. Vigésima primera edición, Mathema Editores, México 1998.

[2] Jean-Paul Collette, *HISTORIA DE LAS MATEMATICAS I*. Segunda Edición, Siglo Veintiuno Editores, México 1986.

[3] James R. Newman, *EL MUNDO DE LAS MATEMATICAS 1.* Novena Edición, Editores Grijalbo, Barcelona 1983.
[4] Linares, Raúl, *El cálculo (Memorias de la* 5^o *Gran Semana Nacional de las Matemática)*, El Errante Editor, Puebla 2010.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

rlinares@fcfm.buap.mx, 200820897@fcfm.buap.com

Lógica Matemática

Capítulo 22

BREVE RESEÑA DE MODEL CHECKING

JOSÉ ARRAZOLA RAMÍREZ
IVÁN CORTÉS
JESÚS LAVALLE MARTÍNEZ*
FCFM - BUAP
*FACULTAD DE CIENCIAS DE LA COMPUTACIÓN - BUAP

1. Introducción

En las pasadas dos décadas surge un planteamiento muy atractivo hacia la correción de sistemas de control basados en computadora, esto es lo que ahora se conoce como "*model checking*".

"*Model checking*" es una técnica de verificación formal, la cual permite verificar el comportamiento de ciertas propiedades de un sistema dado, esto en base a un adecuado modelo del sistema y medio una sistemática inspección de todos los estados de dicho modelo. Lo atractivo de Model Checking viene del hecho que es completamente automático, es de rápido aprendizaje y ofrece contraejemplos en caso de que el modelo falle al satisfacer alguna propiedad, esto sirve como información indispensable para la depuración. Además de esto, el rendimiento de las herramientas del Model Checking se ha demostrado con el tiempo, gracias a un gran número de aplicaciones a la industria. Por ejemplo: imaginemos que nuestro teléfono móvil falle al digitar ciertos comandos o que nuestra videograbadora reacciones de una manera inesperada; es claro que estos errores no atentan contra nuestra vida, pero es claro que tendrá sustanciales consecuencias financieras para el manufacturador de dichos dispositivos. Como un ejemplo dramático tenemos el caso del error del procesador Intel Pentium I en su unidad de punto flotante que causó una perdida de 475 millones de dólares para remplazar las unidades defectuosas además del daño a la reputación de esta compañía. Mas aún, los errores también pueden ser catastróficos. El software es usado para el proceso de control de sistemas de seguridad crítica tales como plantas químicas, reactores nucleares, control de tráfico aéreo y sistemas de alerta. Claramente los errores en el software pueden traer problemas catastróficos. Por ejemplo, una falla en el software de control de la máquina Therac-25 causó la muerte de 6 pacientes con cáncer entre 1985 y 1987 ya que fueron sometidos a una sobre exposición.

La creciente dependencia en el software de aplicación crítica nos lleva a decir que la fiabilidad de los sistemas de información y comunicación es un punto clave en el proceso de diseño del sistema.

La magnitud de los sistemas de información y comunicación, crece tanto en su complejidad, como en su espacio; dado que estos sistemas no son aplicaciones de escritorio sino software embebido por todas partes, así, al conectar e interactuar con otros sistemas los vuelve más vulnerable a errores y el número de defectos crece de una manera exponencial. De manera particular, un fenómenos tal como

TRABAJO APOYADO POR EL PROYECTO VIEP-BUAP, PROGRAMACIÓN LÓGICA POSIBILISTA Y TEORÍA DE LA ARGUMENTACIÓN.

la concurrencia es lo que vuelve muy dificil de manejar a un modelo con técnicas estandar.

Podemos decir de manera informal que los Métodos formales son la matemática aplicada al modelado y análisis de los sistemas tecnológicos de información y comunicación (TIC). Sin embargo, al parecer hasta ahora una técnica en desarrollo, existen instituciones que actualmente la emplean tal es el caso de:

- IEC - International Electrotechnical Commission, China.
- FAA - Federal Aviation Administration, EUA.
- NASA - EUA.

2. Un poco de teoría

Hasta ahora, hemos descrito algunas de las virtudes de Model Checking, que al ser una técnica formal de verificación, está sustentada en fuertes teorías, por ejemplo, la Lógica Matemática y la Teoría de Automatas entre algunas otras, al ser así cuenta también con algunas definiciones propias; en la sección anterior habíamos comentado acerca de la realización de un modelo del sistema en cuestión para luego realizar la técnica mencionada. Este modelo suele representarse con un tipo de grafo, que para nuestros fines lo denominaremos *Sistema de transición*:

Definición 1. Un sistema de transición TS es una tupla $(S, Act, \rightarrow, I, AP, L)$ donde

- S es el conjunto de estados,
- Act es el conjunto de acciones,
- $\rightarrow \subseteq S \times Act \times S$ es la relación de transición,
- $I \subseteq S$ es el conjunto de estados iniciales,
- AP es el conjunto de proposiciones atómicas,
- $L : S \rightarrow 2^{AP}$ es la función codificadora.

Decimos que un ST es finito si S, Act y AP son finitos.

Notas 1.

1. Escribiremos $s \xrightarrow{\alpha} s'$ en lugar de $(s, \alpha, s') \in \longrightarrow$
2. Cuando $|I| > 1$, el estado inicial se selecciona de manera no determinista.
3. Note que puede que $I = \emptyset$, en tal caso el sistema de transición no tiene comportamiento.

Vamos a describir de manera intuitiva y algorítmica el comportamiento de un sistema de transición

1. El ST se inicia seleccionando al azár algún $s_0 \in S$.
2. $s_0 \in S$ evoluciona de acuerdo a la relación de transición $\longrightarrow$ $(s_0 \xrightarrow{\alpha} s')$
   a) Se selecciona de manera no determinista s_0.
   b) Se realiza α.
   c) Evoluciona a s'.
3. El procedimiento de selección es similar para s'.
4. Finaliza el proceso cuando se encuntra un estado que no tiene transiciones de salida.

Aunque hemos presentado una la definición de un sistema de transición y algunas notas acerca de él, a continuación presentaremos un ejemplo particular.

Ejemplo 1. El sistema de transición de la siguiente figura modela una máquina despachadora de bebidas. Su descripción es: *La máquina puede entregar cualquiera de las dos bebidas, soda o cerveza.*

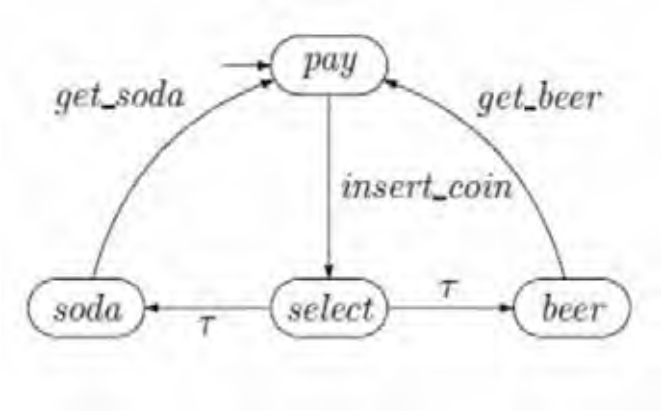


FIGURA 1. La máquina vendedora

La figura 1 es la representación gráfica del sistema de transición que modela la descripción de la máquina vendedora, de éste podemos notar que los óvalos son los estados, los arcos, son las transiciones y las etiquetas de los arcos son las acciones, el estado inicial es aque que está siendo apuntado por una flecha que no viene de algún lado. Aplicando la definición a nuestro ejemplo:

1. El espacio de estados $S = \{pay,\ select,\ soda,\ beer\}$.
2. El conjunto de estados iniciales $I = \{pay\}$.
3. El conjunto de acciones Act= {get_soda, get_beer, insert_coin, τ}.
4. Las proposiciones atómicas dependen de las propiedades bajo consideración. Una elección simple es dejar que los nombres de los estados actuen como proposiciones atómicas.
5. Del punto anterior se deduce que $L(s) = \{s\}$.

Notemos que en el punto 3 anterior, $\tau \in Act$, lo cual nos dice que existe una transición del estado *select* a los estados *beer* y *soda*, el motivo de denotar de esta manera a esta transición, es solo para recalcar que en este grado de abstracción esa transición no la modelaremos, de igual manera puede pensarse como "no habrá problemas en la transición de *select* a *beer* o a *soda*".

Es natural preguntarse que al estar en un estado s a qué otros estados podremos llegar; ésto nos lo dice la siguiente definición:

Definición 2 (Sucesor y predecesor)**.** Sea $TS = (S, Act, \rightarrow, I, AP, L)$ un sistema de transición. Para $s \in S$ y $\alpha \in Act$, el conjunto de sucesores directos α de s está definido como:

$$Post(s, \alpha) = \{s' \in S : s \xrightarrow{\alpha} s'\} \qquad Post(s) = \bigcup_{\alpha \in Act} Post(s, \alpha)$$

El conjunto de los predecesores α de s está definido por:

$$Pre(s, \alpha) = \{s' \in S : s' \xrightarrow{\alpha} s\} \qquad Pre(s) = \bigcup_{\alpha \in Act} Pre(s, \alpha)$$

De manera coloquial podemos decir que el conjunto de los sucesores directos α de s, se llena con todos los estados a los cuales podemos llegar por medio de la

realización de la acción α. Por el contrario, el conjunto de los predecesores α de s, es construído con todos aquellos estados s' que llegan a s por medio de α.

De manera intuitiva los estados terminales son los estados que no tienen transiciones de salida.

Definición 3 (Estado terminal)**.** Un estado s en un sistema de transiciones TS es llamado *terminal* si y sólo si $Post(s) = \emptyset$.

Ejemplo 2. Para un sistema de transición que modele un programa secuencial, los estados terminales ocurren como un fenómeno natural que representa el fin de ejecución del programa.

Definición 4 (Fragmentos de ejecución)**.** Sea un $TS = (S, Act, \rightarrow, I, AP, L)$ un sistema de transiciones. Un fragmento de ejecución *finito* ϱ de TS es una secuencia alternante de estados y acciones que finalizan con un estado

$$\varrho = s_0\alpha_1 s_1\alpha_2 s_2 \ldots \alpha_n s_n \text{ tal que } s_i \xrightarrow{\alpha_{i+1}} s_{i+1} \text{ para todo } 0 \leq i \leq n$$

donde $n \geq 0$. Nos referimos a n como la longitud del fragmento de ejecución ϱ. Un fragmento de ejecución *infinito* ρ de TS, es una secuencia alternante infinita de estados y acciones

$$\rho = s_0\alpha_1 s_1\alpha_2 s_2\alpha_3 \ldots \text{ tal que } s_i \xrightarrow{\alpha_{i+1}} s_{i+1} \text{ para todo } 0 \leq i$$

De la definición anterior surge lo siguiente:

Notas 2.

1. Observe que la secuencia $s_0 \in S$ es un fragmento de ejecución de longitud $n = 0$.
2. Cada *sub* fragmento de ejecución infinito de longitud impar, es un fragmento de ejecución finito.
3. De ahora en adelante, el término *fragmento de ejecución* será usado sin distinción para denotar a un fragmento de ejecución finito o infinito.

Notación 1. Los fragementos de ejecución $\varrho = s_0\alpha_1 \ldots \alpha_n s_n$ y $\rho = s_0\alpha_1 s_1\alpha_2 \ldots$ se escribiran respectivamente como:

$$\varrho = s_0 \xrightarrow{\alpha_1} \cdots \xrightarrow{\alpha_n} s_n \qquad \text{y} \qquad \rho = s_0 \xrightarrow{\alpha_1} s_1 \xrightarrow{\alpha_2} \cdots$$

Definición 5 (Inicial y máximo)**.** Un fragmento de ejecución *máximo* es o un fragmento finito de ejecución que culmina en un estado terminal, o un fragmento de ejecución infinito.

Un fragmento de ejecución es llamado *inicial* si comienza en un estado inicial, es decir, si $s_0 \in I$.

Definición 6 (Ejecución)**.** Una *ejecución* del sistema de transición TS es un fragmento de ejecución inicial y máximo.

Quizas nos parezca redundante la última definición, sin embargo, con ésta queremos enfatizar que para nuestro fin, una ejecución debe comenzar en un estado inicial y culminar en un estado terminal o bien no terminar, decimos esto ya que puede ocurrir que no se finalice en un estado terminal, esto puede ser porque se ha encontrado un error y que se llegó a éste por la violación de alguna propiedad que estamos modelando. Es también por esto que queremos saber qué estados son alcanzable y cuáles no.

Definición 7 (Estado alcanzable)**.** Sea $TS = (S, Act, \rightarrow, I, AP, L)$ un sistema de transición. Un estado $s \in S$ es llamado *alcanzable* en TS, si existe un fragmento finito de ejecución inicial

$$s_0 \xrightarrow{\alpha_1} s_1 \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_n} s_n = s.$$

$Reach(TS)$ denota el conjunto de todos los estados alcanzables en TS.

Ahora que contamos con un conjunto de definiciones "más desarrollado" es momento de revisitar nuestro modelo de máquina vendedora y agragar algunos detalles a la definición de ésta.

Ejemplo 3. *La máquina vendedora cuenta el número de latas de soda y cerveza y regresa la moneda si está vacía.* Introduciremos de manera informal un concepto, a saber, *transición condicional*.

$$start \overset{true:coin}{\hookrightarrow} select \qquad \text{y} \qquad start \overset{true:refill}{\hookrightarrow} start$$

Notas 3.

1. Las etiquetas de las transiciones condicionales son de la forma $g : \alpha$ donde g es una *condición* booleana (llamada guardia).
2. α es la acción que se realiza una vez que g se cumple.

Las transiciones condicionales:

$$select \overset{nsoda>0:get\_s}{\hookrightarrow} start \quad \text{y} \quad select \overset{nbeer>0:get\_b}{\hookrightarrow} start$$

modelan que una soda o una cerveza pueden ser obtenidas si por lo menos hay alguna de las dos.

Las variables *nsoda* y *nbeer* almacenan el número de latas de soda o cerveza respectivamente.

Finalmente la máquina regresa a la localización inicial, *start*, mientras regresa la moneda insertada cuando ya no hay latas. El modelo es:

$$select \overset{nsoda=0 \,\wedge\, nbeer=0:ret\_coin}{\hookrightarrow} start$$

Notas 4.

1. Denotaremos con *max* a la máxima capacidad de latas de la máquina.
2. La inserción de una moneda, deja sin cambio el número de latas.
3. Lo mismo aplica cuando la moneda es regresada.

Los efectos de las otras acciones son los siguientes

Acción	**Efecto**
refill	`nsoda = max; nbeer = max`
sget	`nsoda = nsoda - 1`
bget	`nbeer = nbeer - 1`

El siguiente ejemplo muestra el caso para `max = 2`.

Como puede observarse en la figura anterior, al realizar pocas modificaciones a la descripción de la máquina vendedora, crece el número de estados, ya que tenemos que realizar una inspección sistemática de todas las combinaciones posibles, esto nos lleva a decir que el problema que enfrentamos es el crecimiento exponencial del espacio de estados.

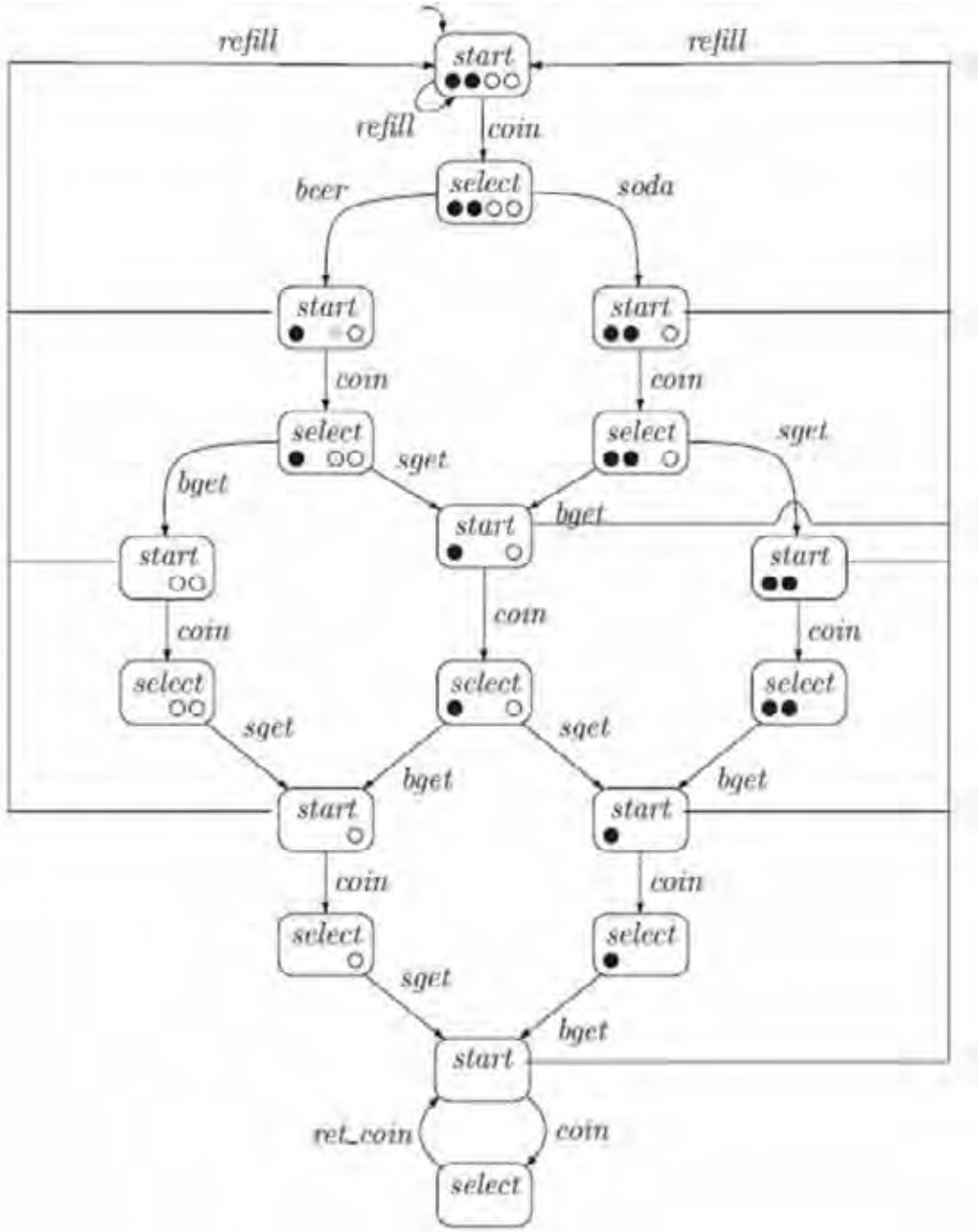


FIGURA 2. La máquina vendedora revisitada

De esta manera podemos llegar a decir que los sistemas con estado infinitos son intratables. Sin embargo, con la modificación de las estructuras de datos y los algoritmos empleados para el problema que estemos tratando, podemos notar una reducción de nuestro problema, ya que podemos tratar una mayor cantidad de estados.

3. Estado del arte

Algunas tendencias actuales:

1. Integrar diferentes lenguajes de especificación. Cada una capaz de manejar diferentes aspectos de un sistema.
2. Manejar aspectos diferentes al comportamiento del sistema, como
   - Cuestiones de rendimiento.
   - Limitaciones de tiempo-real.
   - Políticas de seguridad.
   - Diseño de la arquitectura.

4. Trabajo a futuro

Model Checking y la Demostración automática de Teoremas son las más prometedoras direcciones en la integración de métodos; su uso:

- Emplear Model Checking como procedimiento de decisión sin un motor deductivo.
- Combinar el marco de trabajo deductivo y Model Checking para obtener una abstracción de estados finitos, luego implementar usando Model Checking.

Referencias

[1] Principles of Model Checking, Christel Baier and Joost-Pieter Katoen, The MIT Press, 2009.
[2] Formal Methods: State of the art and future directions, Edmund M. Clarke, Carnegie Mellon University, 2008.
[3] The Beginning of Model Checking: A Personal Perspective, E. Allen Emerson, Springer, 2008.

Capítulo 23

ALGUNOS SISTEMAS LÓGICOS

JOSÉ ARRAZOLA RAMÍREZ
OSCAR ESTRADA ESTRADA
JOSÉ DE JESÚS LAVALLE MARTÍNEZ*
FCFM - BUAP
*FACULTAD DE CIENCIAS DE LA COMPUTACIÓN - BUAP

1. Introducción

La Lógica ha evolucionado desde los tiempos del griego Aristóteles hasta la moderna Lógica Matemática creada por David Hilbert, entre otros. La Lógica comprende entre otras temáticas: la construcción de lenguajes formales, sistemas deductivos, semánticas formales y los procesos para obtener conclusiones a partir de información incompleta o contradictoria.

2. Las Tres Leyes del Pensamiento

Luego de algunos siglos de estudio se llegó a la conclusión de que para un "correcto" razonamiento es *obligatorio* seguir tres reglas, a las que les han llamado "Los Tres Principios Básicos del Pensamiento" [1]:

Identidad: Todo enunciado verdadero es verdadero.
Contradicción: Ningún enunciado puede ser verdadero y falso a la vez.
Tercero Excluido: Cualquier enunciado es verdadero o falso.

Otra notación es:

Identidad: Es verdadero $\mathcal{B} \rightarrow \mathcal{B}$
Contradicción: Es falso $\mathcal{B} \wedge \neg\mathcal{B}$
Tercero Excluido: Es verdadero $\mathcal{B} \vee \neg\mathcal{B}$

Estos tres principios, se supone, bastan para que no existan "problemas" en la lógica. Estos principios fueron tomados como "evidentes" durante muchos siglos, de hecho, los tres principios fueron planteados de manera explícita por Aristóteles, $(384 - 322$ A.C.), y hasta principios de los años veinte se comenzó a poner en tela de juicio su caracter de "evidente".

3. Omisión de la Tercera Ley: Intuicionismo

El primero de los principios en ponerse en tela de juicio fue el Principio del Tercero Excluido: $\mathcal{B}\vee\neg\mathcal{B}$. Este principio permite suponer la veracidad de una expresión, una vez que se ha demostrado que su negación es falsa o conduce a contradicciones. Por ejemplo, este principio aplicado en el proceso de hacer demostraciones nos permite afirmar que la siguiente demostración es correcta:

TRABAJO APOYADO POR EL PROYECTO VIEP-BUAP, PROGRAMACIÓN LÓGICA POSIBILISTA Y TEORÍA DE LA ARGUMENTACIÓN.

> Queremos demostrar que existen dos números irracionales a y b tales que a^b es un número racional. Sea $c = \sqrt{2}^{\sqrt{2}}$. Si c es racional, entonces hacemos $a = b = \sqrt{2}$ y terminamos. Si c es irracional, es decir, no es racional, entonces observando que $c^{\sqrt{2}} = 2$ podemos hacer $a = c$ y $b = \sqrt{2}$. En cualquier caso, tenemos dos números irracionales a y b tales que a^b es un número racional.

Pero queda la pregunta: ¿Cuáles son esos dos números?. Durante el siglo pasado se dió un gran debate entorno a la pregunta de como tratar las demostraciones *no-constructivas* en la matemática. ¿Es válido decir que se ha demostrado la existencia de un objeto matemático con cierta propiedad sin realmente poder encontrarlo o "producirlo"?

A principios de los años veinte, algunos lógicos como Emil Post y los polacos Jan Lukasiewicz y Alfred Tarski, pusieron en duda la necesidad del Principio del Tercero Excluido y mostraron que era posible construir sistemas lógicos trivalentes perfectamente consistentes. En estos sistemas, a los valores"verdadero" y "falso", se añade un tercer valor de verdad al que se denomina "indeterminado", "dudoso" o "incierto". A partir de entonces se han venido desarrollando sistemas lógicos con más de tres valores de verdad, e incluso se han construido sistemas con infinitos valores de verdad.

Después de los trabajos de Brower(1976), Heyting(1956), Kleene(1952), entre otros, se creo un sistema axiomático para describir las consecuencias de omitir el Principio del Tercero Excluido como un principio básico. Tal sistema se llama Lógica Intuicionista; Una formalización para el Cálculo Proposicional Intuicionista, Int, es la siguiente [4]:

1. Los símbolos que se utilizarán en Int son: $\neg, \rightarrow, \wedge, \vee, (,)$ y las letras A_i, donde i es un entero positivo; es decir: $A_1, A_2, A_3, \ldots$ Los símbolos $\neg, \rightarrow$ se llaman *conectivos primitivos*, y las letras A_i *letras proposicionales*.
2. Las fórmulas de intuicionismo se construyen de la misma forma que en lógica clásica.
3. Si $\mathcal{B}, \mathcal{C}$ y $\mathcal{D}$ son fórmulas de C, entonces los siguientes son axiomas:
   A1) $\mathcal{B} \rightarrow (\mathcal{C} \rightarrow \mathcal{B})$
   A2) $(\mathcal{B} \rightarrow (\mathcal{C} \rightarrow \mathcal{D})) \rightarrow ((\mathcal{B} \rightarrow \mathcal{C}) \rightarrow (\mathcal{B} \rightarrow \mathcal{D}))$
   A3) $\mathcal{B} \wedge \mathcal{C} \rightarrow \mathcal{B}$
   A4) $\mathcal{B} \wedge \mathcal{C} \rightarrow \mathcal{C}$
   A5) $\mathcal{B} \rightarrow (\mathcal{C} \rightarrow (\mathcal{B} \wedge \mathcal{C}))$
   A6) $\mathcal{B} \rightarrow (\mathcal{B} \vee \mathcal{C})$
   A7) $\mathcal{C} \rightarrow (\mathcal{B} \vee \mathcal{C})$
   A8) $(\mathcal{B} \rightarrow \mathcal{D}) \rightarrow ((\mathcal{C} \rightarrow \mathcal{D}) \rightarrow (\mathcal{B} \vee \mathcal{C}) \rightarrow \mathcal{D})$
   A9) $(\mathcal{B} \rightarrow \mathcal{C}) \rightarrow ((\mathcal{B} \rightarrow \neg\mathcal{C}) \rightarrow \neg\mathcal{B})$
   A10) $\neg\mathcal{B} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})$
4. La única regla de inferencia de Int es *Modus Ponens*: $\mathcal{C}$ es una consecuencia directa de $\mathcal{B}$ y $\mathcal{B} \rightarrow \mathcal{C}$.

Observación El conjunto de axiomas del Cálculo Proposicional Clásico C, consta de todos los axiomas de Int excepto que el axioma (A_{10}) es substituido por $\neg\neg\mathcal{B} \rightarrow \mathcal{B}$ axioma de C, como se puede ver en [4], además con los axiomas de C se prueban los de Int, por ende, el conjunto de teoremas del Int está contenido en el

conjunto de teoremas de C

Algunas consecuencias son que las fórmulas $\mathcal{A} \vee \neg\mathcal{A}$ y $\neg\neg\mathcal{A} \to \mathcal{A}$ no son teoremas de Int. Más aún, una diferencia entre C e Int es que para cualquier $n \in \mathbb{N}$, la fórmula

$$(A_1 \leftrightarrow A_2) \vee (A_1 \leftrightarrow A_3) \vee \cdots \vee (A_1 \leftrightarrow A_n) \vee (A_2 \leftrightarrow A_3) \vee \cdots \vee (A_{n-1} \leftrightarrow A_n)$$

resulta ser un teorema en clásica, pero no en intuicionismo [4].

Las siguientes son algunas fórmulas que en C son equivalencias y que en Int sólo conservan una de sus implicaciones [3]:

1. $\mathcal{B} \to \neg\neg\mathcal{B}$
2. $\neg\mathcal{B} \vee \neg\mathcal{C} \to \neg(\mathcal{B} \wedge \mathcal{C})$
3. $(\neg\mathcal{B} \vee \mathcal{C}) \to (\mathcal{B} \to \mathcal{C})$
4. $(\mathcal{B} \to \mathcal{C}) \to (\neg\mathcal{C} \to \neg\mathcal{B})$

Observe que la implicación $(\mathcal{B} \to \mathcal{C}) \to (\neg\mathcal{B} \vee \mathcal{C})$ no es teorema de Int, podemos pensar en la siguiente *justificación* intuitiva: Defina $b_n :=$ "existe una cadena de n números 7 consecutivos en la expansión decimal de π". Es claro que la implicación $b_{100} \to b_{99}$ es válida, pero la validez de la fórmula $\neg b_{100} \vee b_{99}$ no es segura.

La no aceptación del Principio del Tercero Excluido tiene consecuencias importantes, debido a que sugiere que la forma en que *hacemos matemáticas* no necesariamente debe seguir y utilizar procedimientos clásicos; ¿Porque no utilizar la Lógica Intuicionista como fundamento de los procedimientos y técnicas de la Matemática? Con esta visión de la Matemática es que se crea lo que se conoce como *Matemática Intuicionista*. En esta matemática, el concepto central es el de "demostración". ¿qué significa demostrar? ¿Sigue siendo válida la demostración presentada anteriormente para la existencia de dos números irracionales con las propiedades mencionadas? En la matemática intuicionista, **NO**. En esta matemática, el concepto de demostración está relacionado con el de *construcción*, y por tanto demostraciones de *existencia* como la anterior no son admitidas. En intuicionismo, demostrar quiere decir *exhibir un procedimiento finito para obtener* el objeto matemático del que se habla o presentar el objeto mismo. Por ejemplo, la demostración de Euclides ($\approx 325 - 265$ a.C) de que hay un número infinito de números primos; Para que fuera válida en Matemática Intuicionista, se debería proporcionar un método para que dado un número primo, se pudiera obtener el siguiente, lo que no se hace en la demostracion conocida por todos.

El hecho de que algunas de las equivalencias entre fórmulas de C no sean válidas en Int, da lugar a preguntas acerca del "verdadero" significado de los conectivos en Int ¿ El significado de los mismos puede definirse por medio de tablas de verdad como en el caso de la C? La respuesta a esta pregunta es *negativa* y fué dada por Gödel (1933). A partir de esa fecha, el problema de asignar una semántica a los conectivos de la Int ha sido abordado desde diferentes perspectivas; Algunas de éstas han sido la semántica por medio de álgebras de Heyting, la semántica de Árboles de Beth(1956), la semántica por medio de espacios Topológios dada por Stone(1937) y Tarski(1938) [y mas tarde desarrollada por McKinsey(1941)], la semántica relacional dada por Kripke (1965) y la semántica probabilista dada por Morgan-LeBlanc (1983). En este trabajo se platicará acerca de la semántica relacional de Kripke.

3.1. Semántica de Kripke. La Semántica de Kripke está basada en la idea intuitiva de "conocimiento", a partir de relaciones entre elementos de un conjunto ordenado y fórmulas lógicas. Veamos lo que significa que una fórmula $\mathcal{A}$ sea "satisfecha" en el nodo k, lo cual denotaremos por $k \Vdash \mathcal{A}$ y diremos que en k *se fuerza* a $\mathcal{A}$.

- $k \Vdash A$ si la variable proposicional A esta asignada al nodo k
- $k \Vdash \mathcal{B} \wedge C$ si $k \Vdash \mathcal{B}$ y $k \Vdash \mathcal{C}$
- $k \Vdash \mathcal{B} \vee C$ si $k \Vdash \mathcal{B}$ o $k \Vdash \mathcal{C}$
- $k \Vdash \mathcal{B} \to C$ si para toda $l \geq k$, $l \Vdash \mathcal{B}$ implica que $l \Vdash \mathcal{C}$.
- $\perp$ no es forzado en algún nodo.

En el caso de C, el asignar valores de verdadero o falso a las proposiciones, refleja el estado de "las cosas", el *estado actual del mundo*: algunas proposiciones son falsas o son verdaderas. La semántica para Int descrita por Kripke refleja un aspecto dinámico: Nuestro conocimiento actual del estado de verdad de las proposiciones puede mejorar. Algunas proposiciones cuyo valor de verdad era indeterminado previamente, pueder ser establecido como verdadero. El valor "verdadero" corresponde a una verdad firmemente establecida, la cual se conserva durante el avance en el conocimiento; el valor "falso" corresponde a "no verdadero todavia"[6].
Tenemos el siguiente teorema:

Theorem 3.1. [3] *Sea* $\Gamma \vdash \mathcal{A}$. *Si todas las fórmulas de* Γ *son forzadas en algún nodo de un modelo de Kripke, entonces también* $\mathcal{A}$ *es forzada en ese nodo.*

Corollary 3.2 (Robustez)**.** [3] *Si* $\vdash \mathcal{A}$, *entonces* $\mathcal{A}$ *es forzada en todos los modelos de Kripe.*

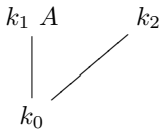
La proposición contrarrecíproca del Teorema de Robustez permite probar que una fórmula no es derivable en Int, esto es:

Corollary 3.3. *Si existe un modelo de Kripke en que* $\mathcal{A}$ *no es forzada, entonces* $\nvdash \mathcal{A}$.

Así, para demostrar que una fórmula no es un teorema en el Cálculo Proposicional Intuicionista basta con encontrar un modelo de Kripke en el que esa fórmula no sea forzada.
Ejemplo 1: Considere la fórmula $\neg\neg A \to A$; Veamos que ésta fórmula no es derivable, para esto considere el siguiente modelo de Kripke:

recordando que $\neg A := A \to \perp$, se puede ver que $k_0 \nVdash \neg A$ y también que $k_0 \Vdash \neg\neg A$, pero $k_0 \nVdash \neg\neg A \to A$.
Ejemplo 2: Veamos que $\nvdash (A \to B) \to (\neg A \vee B)$. Para esto, considere el siguiente modelo de Kripke:

En este modelo, se puede ver que $k_0 \Vdash A \to B$, pero $k_0 \nVdash \neg A \vee B$.
Ejemplo 3: Veamos que $\nvdash (\neg\neg A \to A) \to (\neg A \vee A)$. Para ello, considere el siguiente modelo de Kripke:



Mostraremos que $k_0 \Vdash \neg\neg A \to A$, pero $k_0 \nVdash \neg A \vee A$.
Primero, directamente del modelo de Kripe se obtiene la información:

$$k_0 \nVdash A \quad k_1 \Vdash A$$
$$k_2 \nVdash A$$

Con base en lo anterior, se tiene:

$$k_0 \nVdash \neg A \qquad k_1 \nVdash \neg A \qquad k_2 \Vdash \neg a$$
$$k_0 \nVdash \neg\neg A \quad k_1 \nVdash \neg\neg A \quad k_2 \nVdash \neg\neg A$$

Por lo tanto, se tiene $k_0 \Vdash \neg\neg A \to A$, pero $k_0 \nVdash \neg A \vee A$.
La semántica presentada por Kripke es Robusta y Completa, es decir, los teoremas y fórmulas válidas son *las mismas*; Se cumple el siguiente teorema:

Theorem 3.4. [3] *Sea $\mathcal{A}$ una fórmula del Cálculo Proposicional Intuicionista, entonces:*

$$\vdash_{Int} \mathcal{A} \text{ si y sólo si } \mathcal{A} \text{ es forzada en todos los modelos de Kripke.}$$

3.2. Lógicas Intermedias. Una *lógica intermedia L* es un conjunto de fórmulas el cual es cerrado bajo Modus Ponens y bajo Substitución [2], tal que Th(Int) $\subseteq$ Th(L) $\subset$ Th(C), donde Th(X) abrevia conjunto de teoremas del sistema X. Observe que el Cálculo Proposicional Intuicionista es considerado una lógica intermedia, mientras que el Cálculo Proposicional Clásico no lo es.
Existe un *continuo* de lógicas intermedias [9], las cuales se construyen agregando uno o más axiomas a los axiomas del Cálculo Proposicional Intuicionista. Las lógicas intermedias que presentaremos a continuación están relacionadas de la siguiente manera:

$$I \subseteq J_v \subseteq \ldots \subseteq \mathcal{D} = \cap_{n=2}^{\infty} G_n \subseteq \ldots \subseteq G_3 \subset C.$$

Lógica de Jankov (J_v)
Esta lógica se forma añadiendo a Int el axioma $\neg\mathcal{A} \vee \neg\neg\mathcal{A}$, llamado el principio débil del tercero excluido.

Lógica de Gödel-Dummett ($\mathcal{D}$)
Esta lógica se forma añadiendo a Int el axioma $(\mathcal{A} \to \mathcal{B}) \vee (\mathcal{B} \to \mathcal{A})$.
Otras lógicas intermedias son:

Lógica de Kreisel-Putnam (**KP**)
Esta lógica se forma añadiendo a Int el axioma $(\neg\mathcal{A} \to (\mathcal{B} \vee \mathcal{C})) \to ((\neg\mathcal{A} \to \mathcal{B}) \vee (\neg\mathcal{A} \to \mathcal{C}))$.

Lógica de Scott (SL)
Esta lógica se forma añadiendo a Int el axioma $((\neg\neg\mathcal{A} \to \mathcal{A}) \to (\mathcal{A} \vee \neg\mathcal{A})) \to (\neg\neg\mathcal{A} \vee \neg\mathcal{A})$.

Lógicas de Cardinalidad Acotada ($\mathbf{BC}_n$)
Cada lógica $\mathbf{BC}_n$ se forma añadiendo a Int el axioma $\bigvee_{i=0}^{n}(\bigwedge_{j<i}\mathcal{A}_j \to \mathcal{A}_i)$.

Lógicas de Profundidad Acotada ($\mathbf{BD}_n$)
Cada lógica $\mathbf{BD}_n$ se forma añadiendo a Int el axioma:

$$\mathcal{A}_n \vee (\mathcal{A}_n \to (\mathcal{A}_{n-1} \vee (\mathcal{A}_{n-1} \to \cdots \to (\mathcal{A}_2 \vee (\mathcal{A}_2 \to (\mathcal{A}_1 \vee \neg\mathcal{A}_1)))\cdots))).$$

Lógica G_3 (o Lógica de Smetanich)
Esta lógica se forma añadiendo a Int el axioma:

$$(\neg\mathcal{B} \to \mathcal{A}) \to (((\mathcal{A} \to \mathcal{B}) \to \mathcal{A}) \to \mathcal{A}).$$

Lógicas $n-$valuada de Gödel (G_n)
Cada lógica se forma añadiendo a la Lógica de Gödel-Dummett el axioma $\mathbf{BC}_{n-1}$ o el axioma $\mathbf{BD}_{n-1}$, es decir, $G_n = \mathcal{D}{+}\mathbf{BC}_{n-1} = \mathcal{D}{+}\mathbf{BD}_{n-1}$.

Como pudimos ver la eliminación del Trecero excluido aporta una gran variedad de sistemas lógicos cuyos teoremas son un subconjunto del conjunto de los teoremas del Cálculo Proposicional Clásico. Ahora, podemos preguntarnos ¿Qué sucedería si el principio que se elimina es el de contradicción? La respuesta es que de manera análoga se genera un conjunto de sistemas lógicos denominados Paraconsistentes [7], estos fueron desarrollados por Newton da Costa y actualmente tienen amplia aplicación en Ciencias de la computación, veamos lo siguiente:

4. Omisión de la Segunda Ley: Paraconsistencia

Otra ley del pensamiento puesta en duda fué el Principio de No-Contradicción: *Ningún enunciado puede ser verdadero y falso a la vez.* Dicho principio resulta por demás evidente para nuestra noción intuitiva de verdad. De hecho, definimos una expresión como falsa, cuando su negación es verdadera. Una teoría (conjunto de fórmula de alguna lógica) de la que se puede deducir una fórmula $\mathcal{B}$ y a su negación $\neg\mathcal{B}$ se le denomina *teoría contradictoria*. En el Cálculo Proposicional Clásico, el hecho de que podamos deducir a una fórmula y a su negación lleva, irremediablemente, a que podemos deducir cualquier fórmula; Por ejemplo, considere la siguiente demostración en la que se han deducido, de alguna forma, las fórmulas $\mathcal{B}$ y $\neg\mathcal{B}$:

1.	$\mathcal{B}$	se dedujo de alguna forma
2.	$\neg\mathcal{B}$	se dedujo de alguna forma
3.	$\mathcal{B} \to (\neg\mathcal{C} \to \mathcal{B})$	Instancia Axioma 1
4.	$\neg\mathcal{C} \to \mathcal{B}$	$1, 3$ y M.P.
5.	$\neg\mathcal{B} \to (\neg\mathcal{C} \to \neg\mathcal{B})$	Instancia Axioma 1
6.	$\neg\mathcal{C} \to \neg\mathcal{B}$	$2, 5$ y M.P.
7.	$(\neg\mathcal{C} \to \mathcal{B}) \to ((\neg\mathcal{C} \to \neg\mathcal{B}) \to \mathcal{C})$	Instancia de Axioma 3
8.	$(\neg\mathcal{C} \to \neg\mathcal{B}) \to \mathcal{C}$	$4, 7$ y M.P.
9.	$\mathcal{C}$	$6, 8$ y M.P.

observe que el simple hecho de haber deducido a una fórmula $\mathcal{B}$ y a su negación $\neg\mathcal{B}$, nos lleva que podemos deducir la fórmula $\mathcal{C}$, independientemente de que tan compleja sea, o de si tiene relación con deducciones anteriores. Este mismo *fenómeno*

sucede en Int, donde el axioma A10, a saber: $\neg\mathcal{B} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})$, nos da más claridad al respecto.

Decimos que una teoría, o una lógica, es *trivial* si se puede deducir cualquier fórmula en ella. Como acabamos de ver, una teoría contradictoria es una teoría trivial. Observe que el recíproco también es válido, tanto en C como en Int: si una teoría es trivial, entonces es una teoría contradictoria; Esto es así, debido a que en ella se pueden deducir una fórmula $\mathcal{B}$ y su negación $\neg\mathcal{B}$. De esta forma, acabamos de mostrar que los conceptos de Teoría Trivial y de Teoría Contradictoria coinciden en C y en Int.

Un tercer concepto que tiene relación con los de *Teoría Trivial* y *Teoría Contradictoria* es el de *Teoría Explosiva*; decimos que una teoría, o una lógica, es explosiva si al agregarle una contradicción ($\mathcal{B}$ y $\neg\mathcal{B}$) se vuelve trivial. Así, tanto C como Int son lógicas explosivas. Por mucho tiempo, el caracter de *explosividad* de una lógica o de una teoría fue algo indeseable, y se creía que desde el momento mismo en que se encontraban contradicciones en una teoría ésta debía ser desechada o modificada. Se pensaba que una teoría contradictoria, al ser trivial, no podía aportar información útil. Pero en la vida cotidiana es común obtener información útil aun en situaciones donde se ha encontrado una contradicción; Considere la siguiente situación:

> Usted va manejando su auto por una carretera rumbo a la ciudad X y llega a una bifurcación. No existe alguna señal que le indique cual es el camino hacia X. Al lado del camino se encuentran dos lugareños y usted les pregunta señalando unos de los caminos: ¿este camino me lleva a la ciudad X? Sucederá alguna de las siguientes tres situaciones:
>
> 1. Los dos lugareños le dicen “SI”
> 2. Los dos lugareños le dicen “NO”
> 3. Un lugareño le dice “SI” y el otro lugareño de dice “NO”

Resulta que en los primeros dos casos no se puede tener la certeza de que los lugareños estén mintiendo o no. Es sólo en el tercer caso, en donde existe información contradictoria, donde uno puede estar seguro de que uno de los dos está mintiendo (o dando información equivocada). Luego, aún al encontrarnos con información contradictoria podemos obtener información útil.

Otra situación en donde se tiene información contradictoria entre sí es en las llamadas “paradojas de la autoreferencia” tales como la *paradoja del mentiroso*:

$$\mathcal{B} : \text{Esta oración es falsa}$$

Si aceptamos que $\mathcal{B}$ es verdadera, entonces se debe tener que la oración es falsa, es decir, también tenemos que $\neg\mathcal{B}$ es verdadera. Si, por otro lado, aceptamos que $\mathcal{B}$ es falsa (y por tanto que $\neg\mathcal{B}$ es verdadera) entonces tenemos que la oración $\mathcal{B}$ es verdadera. En cualquiera de los dos casos, aceptar la veracidad o la falsedad de la oración $\mathcal{B}$ nos lleva irremediablemente a que tanto $\mathcal{B}$ como $\neg\mathcal{B}$ son ambas verdaderas. Los ejemplos anteriores muestran la necesidad de una lógica que sea capaz de manejar inconsistencias sin que toda la teoría se vuelva trivial. Una lógica adecuada para modelar estos ejemplos debería ser capaz de manejar el conocimiento inconsistente de la misma forma en que se maneja el conocimiento consistente.

La observación de que en el devenir del *mundo real* existen contradicciones no es una observación nueva; Desde los tiempos de Heráclito (535 a.C. - 484 a.C.) ya se proclamaba la existencia de tales contradicciones. Pero como es sabido, el trabajo de Aristóteles (384 a.C. - 322 a.C.) obscureció por mas de 2,000 años dichas observaciones. Por supuesto que hubo muchos intentos, entre los que destacan los

de Hegel y Marx, pero la mayoria no fueron lo suficientemente fructiferos y contundentes como para acabar con la tradición Aristotélica. Los primeros sistemas formales de lógica paraconsistente fueron creados después de la Segunda Guerra Mundial, en lugares muy distintos y de manera independiente unos de otros; El primero fue creado por el lógico polaco Staniław Jaśkowski (1949), y los restantes se deben a F. G. Asenjo (Argentina, 1949), Newton C. A. da Costa (Brasil, 1958) y T. J. Smiley (Reino Unido, 1959)[5].

4.1. Cálculo Proposicional Positivo (CPP). Primero, comenzamos presentando el Cálculo Proposicional Positivo, el cual es construido por un lenguaje proposicional usual, que consta de: Una formalización para CPP:

1. Los símbolos que se utilizarán en CPP son: $\neg, \rightarrow, \wedge, \vee, (,)$ y las letras A_i, donde i es un entero positivo; es decir: $A_1, A_2, A_3, \ldots$ Los símbolos $\neg, \rightarrow$ se llaman *conectivos primitivos* y las letras A_i *letras proposicionales.*
2. Las fórmulas de CPP se construyen de la misma forma que en C.
3. Si $\mathcal{B}, \mathcal{C}$ y $\mathcal{D}$ son fórmulas de C. El conjunto axiomático del Cálculo Proposicional Positivo , está definido por los axiomas A1)-A8), inciso (3) de la formalización de Int.
4. Como única regla de inferencia, Modus Ponens(MP).

4.2. Lógicas Paraconsistentes. Una *Lógica Paraconsistente* es una lógica que "tolera" inconsistencias, es decir, que no se trivializa en presencia de contradicciones. Algunas lógicas paraconsistentes pueden ser construidas a partir del Cálculo Proposicional Positivo, agregando uno o más axiomas. Alternativamente, se pueden definir de manera semántica.

El Sistema C_ω.

El sistema C_ω se define como el Cálculo Proposicional Positivo, más los siguientes dos axiomas:

$C_\omega 1)$ $\quad \mathcal{B} \vee \neg\mathcal{B}$
$C_\omega 2)$ $\quad \neg\neg\mathcal{B} \rightarrow \mathcal{B}$

Se puede demostrar que C_ω es la lógica paraconsistente "mas pequeña" (en el sentido de la contención de teoremas).

La Lógica G_3'. Para definir la lógica G_3', debemos utilizar las siguientes abreviaciones:

$\sim \mathcal{B} := (\mathcal{B} \rightarrow (\neg\mathcal{B} \wedge \neg\neg\mathcal{B}))$
$\mathcal{B} \vee \mathcal{C} := ((\mathcal{B} \rightarrow \mathcal{C}) \rightarrow \mathcal{C}) \wedge ((\mathcal{C} \rightarrow \mathcal{B}) \rightarrow \mathcal{B})$
$\mathcal{B} \leftrightarrow \mathcal{C} := (\mathcal{B} \rightarrow \mathcal{C}) \wedge (\mathcal{C} \rightarrow \mathcal{B})$
$\bigtriangledown\mathcal{B} := (\sim\sim \mathcal{B}) \wedge (\neg\mathcal{B})$

Los axiomas de la Lógica G_3' son los de la lógica C_ω más los siguientes:

$G_3'1)$ $\quad (\neg\mathcal{B} \rightarrow \neg\mathcal{C}) \leftrightarrow (\neg\neg\mathcal{C} \rightarrow \neg\neg\mathcal{B})$
$G_3'2)$ $\quad \neg\neg(\mathcal{B} \rightarrow \mathcal{C}) \leftrightarrow ((\mathcal{B} \rightarrow \mathcal{C}) \wedge (\neg\neg\mathcal{B} \rightarrow \neg\neg\mathcal{C}))$
$G_3'3)$ $\quad \neg\neg(\mathcal{B} \wedge \mathcal{C}) \leftrightarrow (\neg\neg\mathcal{B} \wedge \neg\neg\mathcal{C}))$
$G_3'4)$ $\quad ((\mathcal{B} \wedge \neg\mathcal{B}) \wedge \bigtriangledown\mathcal{C}) \rightarrow \mathcal{C}$

La Lógica Pac.

La lógica paraconsistente Pac se obtiene agregando a los axiomas de la lógica C_ω los siguientes axiomas:

$Pac1)$ $\quad ((\mathcal{B} \rightarrow \mathcal{C}) \rightarrow \mathcal{B}) \rightarrow \mathcal{B}$

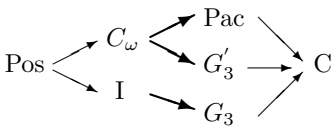


FIGURA 1. Comparación de algunas lógicas

$Pac2)$ $\quad \mathcal{B} \to \neg\neg\mathcal{B}$.
$Pac3)$ $\quad (\neg(\mathcal{B} \vee \mathcal{C}) \to (\neg\mathcal{B} \wedge \neg\mathcal{C})) \wedge ((\neg\mathcal{B} \wedge \neg\mathcal{C}) \to \neg(\mathcal{B} \vee \mathcal{C}))$
$Pac4)$ $\quad (\neg(\mathcal{B} \wedge \mathcal{C}) \to (\neg\mathcal{B} \vee \neg\mathcal{C})) \wedge ((\neg\mathcal{B} \vee \neg\mathcal{C}) \to \neg(\mathcal{B} \wedge \mathcal{C}))$
$Pac5)$ $\quad (\neg(\mathcal{B} \to \mathcal{C}) \to (\mathcal{B} \wedge \neg\mathcal{C})) \wedge ((\mathcal{B} \wedge \neg\mathcal{C}) \to \neg(\mathcal{B} \to \mathcal{C}))$

Esta lógica introduce las leyes de De Morgan como axiomas, lo cual nos permite cancelar dobles negaciones y expresar la implicación en términos de la disyunción. Estas propiedades no las posee la lógica C_ω.

Algunas de las lógicas anteriores también pueden presentarse de manera semántica, es decir, mediante tablas de verdad. La valuación de los conectivos $\neg_{Pac}$, $\to_{Pac}$, $\neg_{G_3}$, $\neg_{G'_3}$ y $\to_{G_3}$ se encuentra en los Cuadros 1 y 2, mientras que: $a \vee_{Pac} b = \max\{a, b\}$ y $a \wedge_{Pac} b = \min\{a, b\}$. En la Figura 1 se ilustran las relaciones entre las diferentes lógicas expuestas [8].

a	$\neg_{Pac}a$
0	2
1	1
2	0

$\to_{Pac}$	0	1	2
0	2	2	2
1	0	1	2
2	0	1	2

CUADRO 1. Valuación de los conectivos $\neg_{Pac}$ y $\to_{Pac}$.

a	$\neg_{G_3}a$
0	2
1	0
2	0

a	$\neg_{G'_3}a$
0	2
1	2
2	0

$\to_{G_3}$	0	1	2
0	2	2	2
1	0	2	2
2	0	1	2

CUADRO 2. Valuación de los conectivos $\neg_{G_3}$, $\neg_{G'_3}$ y $\to_{G_3}$.

Las lógicas que acabamos de describir en este trabajo resultan útiles, no sólo desde el punto de vista filosófico, sino también desde el punto de vista técnico, ya que en Programación Lógica permiten establecer un enlace entre la teoría de modelos y la parte sintáctica (de demostración) de alguna lógica.

Referencias

[1] Irvin Copi, Carl Cohen *Introducción a la Lógica*. Limusa 2007.
[2] van Dalen Dirk, *Intuitionistic Logic*. Algorithms and Decision Problems: *a crash course in recursion theory*. Eds. Gabbay, Guenthner, Handbook of Philosophical Logic, vol 1, Kluwer, Dordrecht. 2001.
[3] van Dalen Dirk *Logic and structure*, Springer, Universitext, 2008.
[4] Elliot Mendelson, *Introduction to Mathematical Logic*. Chapman & Hall, 2004.
[5] Gladys Palau, *Introducción Filosófica a las Lógicas No Clásicas*. Editorial Gedisa, 2002.
[6] Grigori Mints, *A Short Introduction to Intuitionistic Logic*. Eds. Sylvain E. Cappell and Joseph J. Kohn, Kluwer Academic Publishers, 2002.
[7] M. Osorio, J.Arrazola, and J. L. Carballido, *Logical weak completions of paraconsistent logics*. Journal of Logic and Computation, 2008.
[8] M. Osorio, J. A. Navarro, J. Arrazola, V. Borja, *Logics with common weak completions*. Journal of Logic and Computation, 16(6), 2006.
[9] *Intermediate Logic*. Wikipedia, 2010.

arrazola@fcfm.buap.mx, oestrada2005@gmail.com, jlavalle@aleteya.cs.buap.mx

Capítulo 24

LÓGICA POSIBILISTA ESTÁNDAR

JOSÉ ARRAZOLA RAMÍREZ
OSCAR ESTRADA ESTRADA
JOSÉ DE JESÚS LAVALLE MARTÍNEZ*
FELIPE MAZÓN CAMBRÓN
FCFM - BUAP
*FACULTAD DE CIENCIAS DE LA COMPUTACIÓN - BUAP

Resumen. Se presenta una breve reseña de la Lógica Posibilista Estándar, se demuestran teoremas de deducción y Refutación. Culminando el trabajo con un teorema de robustez y completitud para la Lógica Posibilista Estándar.

1. Introducción

El presente trabajo está fuertemente apoyado en [5]. La Lógica Posibilista fué creada por Didier Dubois y Henri Prade, la cual emerge de la Teoría de la Posibilidad de Zadeh, ofreciendo un marco para representar el estado parcial de ignorancia, debido al uso de un par de funciones una denominada medida de necesidad y la otra medida de posibilidad.

La Lógica Posibilista es una lógica que utiliza medidas de incertidumbre para razonamiento con evidencia incompleta y conocimiento parcialmente inconsistente. En un nivel sintáctico se trabaja con fórmulas de una lógica proposicional o de una lógica de primer orden, a las que se les asigna un número en el intervalo $[0, 1]$, en general podría asignarse cualquier elemento de un conjunto totalmente ordenado. Estas cotas inferiores son llamadas grados de necesidad o grados de posibilidad de las fórmulas correspondientes.

En [10] Zadeh introduce la medida de posibilidad como un índice escalar que evalúa la consistencia de una proposición difusa respecto al estado de conocimiento expresado por medio de una restricción difusa. Una restricción difusa es un conjunto difuso de valores *posibles* y su función membresía es llamada una distribución de posibilidad, para ésta existe una la noción dual denominada distribución de certeza [9] o distribución de necesidad [2].

La Lógica Posibilista puede ser empleada para modelar partes de conocimiento impregnado de incertidumbre, cuando la incertidumbre puede ser representada en el contexto de la teoría de la posibilidad.

2. Teoría de la Posibilidad

El objeto básico de la teoría de la posibilidad es la *distribución de posibilidad*, la cual es una función que le asigna a cada elemento de un conjunto U de *alternativas* un grado de posibilidad $\pi(u) \in [0, 1]$. La distribución de posibilidad es la representación de lo que un *agente* sabe acerca del valor de alguna cantidad x que

TRABAJO APOYADO POR EL PROYECTO VIEP-BUAP, PROGRAMACIÓN LÓGICA POSIBILISTA Y TEORÍA DE LA ARGUMENTACIÓN.

toma valores en el conjunto U. La función π_x representa los valores más o menos plausibles para la cantidad desconocida x. Se asume que x puede tomar un sólo valor.

Adoptamos las siguientes convenciones

$\pi_x(u) = 0$ significa que $x = u$ es imposible.
$\pi_x(u) = 1$ significa que $x = u$ es completamente permitido.
$\pi_x(u) > \pi_x(u')$ significa que $x = u$ es preferido a $x = u'$.

2.1. Ejemplo. La distribución de posibilidad más simple es la función característica de un subconjunto E de U. Considere la función $\pi_x : U \longrightarrow [0, 1]$, definida por

$$\pi_x(u) = \begin{cases} 1 & \text{si } u \in E \\ 0 & \text{otro caso} \end{cases}$$

Esta distribución representa la situación de que todo lo que se sabe sobre el valor de la variable x es que no puede pertenecer al complemento de E. Este tipo de distribución surge de manera natural cuando se quiere representar, por ejemplo, que "estamos en el mes de Febrero". En este caso se tiene que $U = \{$ 1 de Enero, 2 de Enero, ..., 31 de Diciembre$\}$, x es la variable que representa el dia actual y $E = \{$ 1 de Febrero, 2 de Febrero, ..., 28 de Febrero$\}$[1].

2.2. Definición. Sean π_x y π'_x dos distribuciones de posibilidad. Se dice que π_x es más específica que π'_x si y sólo si $\pi_x < \pi'_x$, es decir si y sólo si $\forall u \in U,\ \pi_x(u) < \pi'_x(u)$ [11].

La especificidad se refiere al nivel de precisión de una distribución de posibilidad. Es decir, si $\pi_x < \pi'_x$, entonces π_x proporciona más información que π'_x. Si existe $u_0 \in U$ tal que $\pi_x(u_0) = 1$ mientras que $\pi_x(u) = 0$ para $u \neq u_0$, entonces decimos que el *estado de conocimiento* es completo (sabemos que $x = u_0$). Similarmente, se tiene un estado de total ignorancia cuando $\pi_x(u) = 1$ para todo $u \in U$.

Un principio importante en la teoría de la posibilidad es el Principio de Mínima Especificidad, el cual dice que, dado un conjunto de restricciones que acotan el valor de x, se debería escoger π_x de tal forma que le asigne a cada $u \in U$ el máximo grado de posibilidad acorde con las restricciones.

2.3. Definición. Una medida de posibilidad es una función conjunto-valuada Π que asocia a cada subconjunto $A \subseteq U$ un número $\Pi(A) \in [0, 1]$. Los axiomas básicos de la medida de posibilidad son

$$\Pi(\emptyset) = 0$$
$$\Pi(U) = 1$$
$$\Pi(\cup_{i\in I} A_i) = \sup_{i\in I} \Pi(A_i)$$

para un conjunto de índices I.

Una medida de posibilidad se puede obtener de una distribución de posibilidad π_x, la cual verifica que $\forall u \in U,\ \pi_x(u) = \Pi(\{u\})$. En particular, tenemos

$$\Pi(A) = \sup_{u\in A} \pi_x(u).$$

[1]No se están considerando años biciestos.

$\Pi(A)$ expresa en qué medida existe un valor $u \in A$ que pueda presentarse como valor de x. La función dual de la función conjunto-valuada Π es llamada una medida de necesidad, denotada por N y se define como [2]:

$$N(A) = 1 - \Pi(\overline{A}) = \inf_{u \notin A} \{1 - \pi_x(u)\},$$

donde $\overline{A}$ es el complemento de A. $N(A)$ evalua en qué medida todos los posibles valores de x pertenecen a A, es decir, en qué medida uno está seguro de que x pertenece a A.

2.4. Proposición. Sean $A, B \subseteq U$, entonces tenemos que

1. $\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$
2. $N(A \cap B) = \min\{N(A), N(B)\}$
3. $\Pi(A \cap B) \leq \min\{\Pi(A), \Pi(B)\}$
4. $N(A \cup B) \geq \max\{N(A), N(B)\}$
5. $\min\left\{N(A), N(\overline{A})\right\} = 0$
6. $\max\left\{\Pi(A), \Pi(\overline{A})\right\} = 1$

Demostración.

1) En esta parte de la prueba utilizaremos la propiedad de que $\sup\{A \cup B\} = \max\{\sup A, \sup B\}$. Así,
$\Pi(A \cup B) = \sup_{u \in A \cup B} \pi_x(u)$
$= \sup\{\pi_x(u) \mid u \in A \cup B\}$
$= \sup\{\pi_x(u) \mid u \in A \text{ ó } B\}$
$= \sup\{\{\pi_x(u) \mid u \in A\} \cup \{\pi_x(u) \mid u \in B\}\}$
$= \max\{\sup\{\pi_x(u) \mid u \in A\}, \sup\{\pi_x(u) \mid u \in B\}\}$
$= \max\{\Pi(A), \Pi(B)\}$.

2) $N(A \cap B) = 1 - \Pi(\overline{A} \cup \overline{B})$
$= 1 - \max\left\{\Pi(\overline{A}), \Pi(\overline{B})\right\}$
$= \min\left\{1 - \Pi(\overline{A}), 1 - \Pi(\overline{B})\right\}$
$= \min\{N(A), N(B)\}$.

3) $\Pi(A \cap B) = \sup_{u \in A \cap B} \pi_x(u)$. Supongamos que el supremo se alcanza en u^*, es decir, $\Pi(A \cap B) = \pi_x(u^*)$. Además, $\Pi(A) = \sup_{u \in A} \pi_x(u) \geq \pi(u^*)$ y $\Pi(B) = \sup_{u \in A} \pi_x(u) \geq \pi(u^*)$ lo cual implica que $\min\{\Pi(A), \Pi(B)\} \geq \pi(u^*) = \Pi(A \cap B)$.

4) $N(A \cup B) = 1 - \Pi(\overline{A \cup B})$
$= 1 - \Pi(\overline{A} \cap \overline{B}) \geq 1 - \min\left\{\Pi(\overline{A}), \Pi(\overline{B})\right\}$
$= \max\left\{1 - \Pi(\overline{A}), 1 - \Pi(\overline{B})\right\}$
$= \max\{N(A), N(B)\}$.

5) $\min\left\{N(A), N(\overline{A})\right\}$
$= \min\left\{1 - \Pi(\overline{A}), 1 - \Pi(A)\right\}$
$= 1 - \max\left\{\Pi(\overline{A}), \Pi(A)\right\}$
$= 1 - \Pi(A \cup \overline{A}) = 1 - \Pi(U) = 0$.

6) $\max\left\{\Pi(A), \Pi(\overline{A})\right\} = \Pi(A \cup \overline{A}) = \Pi(U) = 1$.

□

3. Lógica Posibilista Estándar. Fórmulas de Necesidad Valuadas

La incertidumbre es un atributo de la información. El trabajo pionero de Claude Shannon en la teoría de la información condujo a la aceptación universal de que la información es estadística en su naturaleza. Así, el manejo de la incertidumbre, sin importar su forma, ha sido delegado a la Teoría de la Probabilidad.

Entre las aplicaciones de la Lógica Posibilista se encuentran:

1. Razonamiento No-Monótono [3, 4], razonamiento con *default rules* [12].
2. Revisión de creencias [1].
3. Se puede usar la programación lógica para crear la *Programación Lógica Posibilista*, la cual es particularmente útil cuando se trata con incertidumbre o con optimización min-max. Los detalles formales sobre la semántica declarativa y procedural de la programas lógicos posibilistas se pueden encontrar en [6] y algunas extensiones que incorporan la negación por falla se pueden encontrar en los trabajos de Wagner [8].

3.1. **Lenguaje.**

3.1. Definición. Una fórmula *de necesidad valuada* (también llamada una fórmula posibilista estándar) es un par $(\varphi\ \alpha)$, donde φ es una fórmula proposicional clásica y $\alpha \in (0, 1]$. $(\varphi\ \alpha)$ expresa que φ es cierta al menos desde un grado α, esto es, $N(\varphi) \geq \alpha$, donde N es una *medida de necesidad* que modela el estado de conocimiento. A la constante α se le conoce como la *valuación* (o peso) de la fórmula y se denota como $val(\varphi)$.

Una base de conocimiento posibilista estándar $\mathcal{F}$ se define como un conjunto finito de fórmulas de necesidad valuadas. Denotamos por $\mathcal{F}^*$ al conjunto de fórmulas clásicas que se obtiene de $\mathcal{F}$ de la siguiente manera: si $\mathcal{F} = \{(\varphi_i\ \alpha_i) \mid i = 1, \ldots, n\}$ entonces, $\mathcal{F}^* = \{\varphi_i \mid i = 1, \ldots, n\}$, tal $\mathcal{F}^*$ se denomina la *proyección clásica* de $\mathcal{F}$.

Una base de conocimiento posibilista estándar también puede ser vista como una colección anidada de fórmulas clásicas: si α es cualquier valuación de [0,1], definimos el $\alpha-$corte $\mathcal{F}_\alpha$ y el $\alpha-$corte estricto $\mathcal{F}_{\overline{\alpha}}$ como:

$$\begin{aligned} \mathcal{F}_\alpha &= \{(\varphi\ \beta) \in \mathcal{F} \mid \beta \geq \alpha\} \\ \mathcal{F}_{\overline{\alpha}} &= \{(\varphi\ \beta) \in \mathcal{F} \mid \beta > \alpha\} \end{aligned}$$

sus proyecciones clásicas son $\mathcal{F}^*_\alpha$ y $\mathcal{F}^*_{\overline{\alpha}}$, es decir

$$\begin{aligned} \mathcal{F}^*_\alpha &= \{\varphi \mid (\varphi\ \beta) \in \mathcal{F},\ \beta \geq \alpha\} \\ \mathcal{F}^*_{\overline{\alpha}} &= \{\varphi \mid (\varphi\ \beta) \in \mathcal{F},\ \beta > \alpha\} \end{aligned}$$

Sea $\mathcal{L}$ un lenguaje clásico asociado con el conjunto $\mathcal{F}^*$ de fórmulas clásicas obtenido del conjunto $\mathcal{F}$ de fórmulas posibilistas y sea Ω el conjunto de interpretaciones clásicas para $\mathcal{L}$. Sea $\mathcal{L}'$ el conjunto de fórmulas cerradas de $\mathcal{L}$, [7].

La semántica de un conjunto de fórmulas clásicas $\mathcal{F}^*$ está definida por medio de un conjunto de interpretaciones que satisfacen todas las fórmulas en $\mathcal{F}^*$. Cada una de tales interpretacines es denominada un *modelo*.

En la Lógica Posibilista Estándar los conceptos de *satisfacción* y *consecuencia lógica* están definidos en términos de distribuciones de posibilidad sobre el conjunto de "mundos clásicos". Una distribución de posibilidad π es una función de Ω (el

conjunto de todos los mundos posibles) a $[0,1]$. $\pi(\omega)$ refleja cuan posible es que ω sea el "mundo real". Cuando $\pi(\omega) = 1$ (respectivamente, $\pi(\omega) = 0$) entonces es completamente posible (respectivamente, imposible) que ω sea el mundo real. Una distribución de posibilidad está *normalizada* si y sólo si $\exists\omega$ tal que $\pi(\omega) = 1$.

La *medida de posibilidad* Π inducida por π es una función de $\mathfrak{L}$ (un lenguaje lógico de primer orden o proposicional) a $[0,1]$ definida por

$$\Pi(\varphi) = \sup\{\pi(\omega) \ : \ \omega \models \varphi\}$$

La *medida de necesidad* N inducida por π se define como

$$N(\varphi) = 1 - \Pi(\neg\varphi) = \inf\{1 - \pi(\omega) \ : \ \omega \models \neg\varphi\}$$

Se cumplen las siguientes propiedades:

$$\begin{array}{rcl} N(\top) & = & 1 \\ N(\varphi \wedge \psi) & = & \min\{N(\varphi), N(\psi)\} \\ N(\varphi \vee \psi) & \geq & \max\{N(\varphi), N(\psi)\} \\ \text{Si } \varphi \models \psi & \text{entonces} & N(\psi) \geq N(\varphi) \end{array}$$

Demostración.

1) $N(\top) = 1 - \Pi(\neg\top) = 1 - \Pi(\bot)$
$= 1 - \sup\{\pi(\omega) : \omega \models \bot\}$
$= \inf\{1 - \pi(\omega) : \omega \models \bot\}$
$= \inf \emptyset = 1.$

2) Para esta parte de la prueba utilizaremos la siguiente propiedad $\sup\{A \cup B\} = \sup\{\sup A, \sup B\}$. Así,
$N(\varphi \wedge \psi) = 1 - \Pi(\neg(\varphi \wedge \psi))$
$= 1 - \Pi(\neg\varphi \vee \neg\psi)$
$= 1 - \sup\{\pi(\omega) : \omega \models \neg\varphi \vee \neg\psi\}$
$= 1 - \sup\{\{\pi(\omega) : \omega \models \neg\varphi\} \cup \{\pi(\omega) : \omega \models \neg\psi\}\}$
$= 1 - \sup\{\sup\{\pi(\omega) : \omega \models \neg\varphi\}, \sup\{\pi(\omega) : \omega \models \neg\psi\}\}$
$= \inf\{1 - \sup\{\pi(\omega) : \omega \models \neg\varphi\}, 1 - \sup\{\pi(\omega) : \omega \models \neg\psi\}\}$
$= \inf\{N(\varphi), N(\psi)\} = \min\{N(\varphi), N(\psi)\}.$

3) En esta parte de la prueba utilizaremos la siguiente propiedad $\sup\{A \cap B\} \leq \min\{\sup A, \sup B\}$, entonces $1 - \sup\{A \cap B\} \geq 1 - \min\{\sup A, \sup B\}$. Así,
$N(\varphi \vee \psi) = 1 - \Pi(\neg(\varphi \vee \psi))$
$= 1 - \sup\{\pi(\omega) : \omega \models \neg\varphi \wedge \neg\psi\}$
$= 1 - \sup\{\{\pi(\omega) : \omega \models \neg\varphi\} \cap \{\pi(\omega) : \omega \models \neg\psi\}\}$
$\geq 1 - \min\{\sup\{\pi(\omega) : \omega \models \neg\varphi\}, \sup\{\pi(\omega) : \omega \models \neg\psi\}\}$
$= \max\{1 - \sup\{\pi(\omega) : \omega \models \neg\varphi\}, 1 - \sup\{\pi(\omega) : \omega \models \neg\psi\}\}$
$= \max\{N(\varphi), N(\psi)\}.$

4) Por hipótesis tenemos que $\varphi \models \psi$, entonces toda ω tal que $\omega \models \varphi$ satisface que $\omega \models \psi$; Así, tenemos que, $\omega \not\models \psi$ implica $\omega \not\models \varphi$ es decir, que $\omega \models \neg\psi$ implica $\omega \models \neg\varphi$. Por lo tanto $\{\omega : \omega \models \neg\psi\} \subseteq \{\omega : \omega \models \neg\varphi\}$. Así, $\{\pi(\omega) : \omega \models \neg\psi\} \subseteq \{\pi(\omega) : \omega \models \neg\varphi\}$. Por lo tanto, $\sup\{\pi(\omega) : \omega \models \neg\psi\}$

$\leq \sup\{\pi(\omega) : \omega \models \neg\varphi\}$. Así,

$$1 - \sup\{\pi(\omega) : \omega \models \neg\psi\} \geq 1 - \sup\{\pi(\omega) : \omega \models \neg\varphi\}.$$

Por lo tanto,

$$N(\psi) \geq N(\varphi).$$

□

Decimos que una distribución de posibilidad π satisface a la fórmula posibilista estándar $(\varphi\ \alpha)$ si y sólo si $N(\varphi) \geq \alpha$, donde N es la medida de necesidad inducida por π. Usaremos la notación $\pi \models (\varphi\ \alpha)$. Una distribución de posibilidad π satisface una base de conocimiento posibilista estándar $\mathcal{F} = \{(\varphi_i\ \alpha_i) \mid i = 1, \ldots, n\}$ si y sólo si $\forall_i\ \pi \models (\varphi_i\ \alpha_i)$. Esto lo denotaremos por $\pi \models \mathcal{F}$. Si $\pi \models (\varphi\ \alpha)$ es verdadera para toda π, lo denotaremos por $\models (\varphi\ \alpha)$ y diremos que $(\varphi\ \alpha)$ es válida.

3.2. Definición. Una fórmula posibilista estándar Φ es una *consecuencia lógica* de una base de conocimiento posibilista estándar $\mathcal{F}$ si y sólo si para cualquier π que satisfaga a $\mathcal{F}$, se tiene que también π satisface a Φ, esto es, para todo π se tiene que si $(\pi \models \mathcal{F})$ entonces $(\pi \models \Phi)$. Esto se denota como $\mathcal{F} \models \Phi$.

3.3. Proposición.

(1) $(\varphi\ \alpha) \models (\varphi\ \beta)$ para todo $\alpha \geq \beta$.
(2) $\forall\alpha > 0, \models (\varphi\ \alpha)$ si y sólo si φ es una tautología[2].

Demostración. En [5] se presenta este resultado sin demostración, a continuación damos la siguiente prueba:

(1) Sea π una distribución de posibilidad tal que $\pi \models (\varphi\ \alpha)$. Por lo tanto, por definición, $N(\varphi) \geq \alpha$. Si $\alpha \geq \beta$, entonces $N(\varphi) \geq \beta$, así $\pi \models (\varphi\ \beta)$. Luego $(\varphi\ \alpha) \models (\varphi\ \beta)$.

2. ($\Longrightarrow$) Supongamos que $\models (\varphi\ \alpha)$ por lo tanto, para toda π y para toda α, se tiene $\pi \models (\varphi\ \alpha)$, es decir, $N(\varphi) \geq \alpha$. Así, en particular, si $\alpha = 1$ se debe tener que para todo π se cumple $N(\varphi) = 1$. De aqui que para todo π se cumple $\max\{\pi(\omega) \mid \omega \models \neg\varphi\} = 0$. Esto implica que es imposible que alguna interpretación cumpla con $\omega \models \neg\varphi$ entonces para toda $\omega \in \Omega$ se tiene $\omega \models \varphi$. Por lo tanto φ es una tautología.

($\Longleftarrow$) Supongamos que φ es una tautología. Así, para toda interpretación $\omega \in \Omega$ se tiene $\omega \models \varphi$, y como el cálculo proposicional clásico es consistente, se tiene que para ningún $\omega \in \Omega$ se cumple $\omega \models \neg\varphi$. Así para toda π, el conjunto $\{\pi(\omega) \mid \omega \models \neg\varphi\}$ es el conjunto vacío. Así, se tiene que $\max\{\pi(\omega) \mid \omega \models \neg\varphi\} = 0$ y por lo tanto $N(\varphi) = 1 - \max\{\pi(\omega) \mid \omega \models \neg\varphi\} = 1 \geq \alpha$ para toda $\alpha \in [0, 1]$.

□

3.4. Definición. Definimos la valuación de una fórmula posibilista estándar de la siguiente manera:

$$Val(\varphi, \mathcal{F}) = \sup\{\alpha \in (0, 1] \mid \mathcal{F} \models (\varphi\ \alpha)\}.$$

[2]En el sentido clásico.

Un resultado fundamental de la deducción a partir de bases de conocimiento posibilistas estándar es que siempre existe una distribución de posibilidad menos *específica* que satisface una base de conocimiento posibilista estándar $\mathcal{F}$. A saber, si $\mathcal{F} = \{(\varphi_i\ \alpha_i) \mid i = 1, \ldots, n\}$ entonces la distribución de posibilidad menos específica $\pi_{\mathcal{F}}$ que satisface $\mathcal{F}$ se define como

$$\pi_{\mathcal{F}}(\omega) = \begin{cases} 1 & \text{si } \omega \models \varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_n \\ 1 - \max\{\alpha_i \mid \omega \models \neg\varphi,\ i = 1, \ldots, n\} & \text{otro caso} \end{cases}$$

3.5. Proposición. [5] Para cualquier distribución de posibilidad π, π satisface a $\mathcal{F}$ si y sólo si $\pi \leq \pi_{\mathcal{F}}$.

Demostración.

$$\begin{array}{lll} \pi \models \mathcal{F} & \text{sii} & (\forall i = 1, 2, \ldots, n)\ \pi \models (\varphi_i\ \alpha_i) \\ & \text{sii} & (\forall i = 1, 2, \ldots, n)\ N(\varphi_i) \geq \alpha_i \text{ (donde } N \text{ es la medida de necesidad inducida por } \pi) \\ & \text{sii} & (\forall i = 1, 2, \ldots, n)\ \inf\{1 - \pi(\omega) \mid \omega \models \neg\varphi_i\} \geq \alpha_i \\ & \text{sii} & (\forall i = 1, 2, \ldots, n)\ (\forall \omega \models \neg\varphi_i)\ \pi(\omega) \leq 1 - \alpha_i \\ & \text{sii} & \pi(\omega) \leq \inf\{1 - \alpha_i \mid \omega \models \neg\varphi_i,\ i = 1, 2, \ldots, n\} \\ & \text{sii} & \pi(\omega) \leq \pi_{\mathcal{F}}(\omega) \end{array}$$

□

3.6. Corolario. [5] Sea $\mathcal{F}$ una base de conocimiento posibilista y sea $(\varphi\ \alpha)$ una fórmula de necesidad valuada. Entonces,

$$\mathcal{F} \models (\varphi\ \alpha) \text{ si y sólo si } \pi_{\mathcal{F}} \models (\varphi\ \alpha).$$

En otros términos, $Val(\varphi, \mathcal{F}) = N_{\mathcal{F}}(\varphi)$, donde $N_{\mathcal{F}}$ es la medida de necesidad inducida por $\pi_{\mathcal{F}}$.

Demostración. En [5], se presenta este resultado sin demostración, a continuación damos la siguiente prueba:

Supongamos que $\mathcal{F} \models (\varphi\ \alpha)$. Por la Proposición 3.5 se tiene que $\pi_{\mathcal{F}}$ satisface a $\mathcal{F}$, por lo tanto, $\pi_{\mathcal{F}} \models (\varphi\ \alpha)$.

Inversamente, supongamos que $\pi_{\mathcal{F}} \models (\varphi\ \alpha)$. Así $N_{\pi_{\mathcal{F}}}(\varphi) \geq \alpha$. Sea π una distribución de posibilidad tal que $\pi \models \mathcal{F}$. Entonces, por la Proposición 3.5 $\forall\omega$, $\pi(\omega) \leq \pi_{\mathcal{F}}(\omega)$. Así, $\inf\{1 - \pi(\omega) \mid \omega \models \neg\varphi\} \geq \inf\{1 - \pi_{\mathcal{F}}(\omega) \mid \omega \models \neg\varphi\}$, luego, $N_{\pi}(\omega) \geq N_{\pi_{\mathcal{F}}}(\varphi) \geq \alpha$. Por lo tanto, $\pi \models (\varphi\ \alpha)$. □

3.2. **Inconsistencia Parcial.**

3.7. Definición. Una base de conocimiento $\mathcal{F}$ cuya distribución de posibilidad asociada $\pi_{\mathcal{F}}$ es tal que $0 < \sup \pi_{\mathcal{F}} < 1$ se llama *parcialmente inconsistente*. La cantidad

$$\text{Cons}(\mathcal{F}) = \sup_{\pi \models \mathcal{F}} \sup_{\omega \in \Omega} \pi(\omega) = \sup_{\omega \in \Omega} \pi_{\mathcal{F}}(\omega)$$

se llamará *grado de consistencia* de $\mathcal{F}$ y la cantidad $\text{Incons}(\mathcal{F}) = 1 - \text{Cons}(\mathcal{F})$ se llamará *grado de inconsistencia* de $\mathcal{F}$.

La inconsistencia parcial extiende la inconsistencia clásica en la siguiente forma: Sea $F = \{\varphi_i \mid i = 1, 2, \ldots, n\}$ un conjunto de fórmulas proposicionales y asociemos con F el conjunto de fórmulas de necesidad valuadas totalmente certeras $\mathcal{F} =$

$\{(\varphi_i\ 1) \mid i = 1, 2, \ldots, n\}$, entonces se puede demostrar inmediatamente que si F es consistente entonces $\text{Incons}(\mathcal{F}) = 0$ y si F es inconsistente entonces $\text{Incons}(\mathcal{F}) = 1$.

3.8. PROPOSICIÓN. Sea $\mathcal{F} = \{(\varphi_1\ \alpha_1), (\varphi_2\ \alpha_2), \ldots, (\varphi_n\ \alpha_n)\}$. Entonces, $\text{Incons}(\mathcal{F}) = 0$ si y sólo si la proyección clásica $\mathcal{F}^*$ es consistente en el sentido clásico.

DEMOSTRACIÓN. Damos la siguiente demostración:

$$\begin{array}{lll}
\text{Incons}(\mathcal{F}) = 0 & \text{si y sólo si} & 1 - \sup_{\omega\in\Omega} \pi_{\mathcal{F}}(\omega) = 0 \\
& \text{si y sólo si} & 1 = \sup_{\omega\in\Omega} \pi_{\mathcal{F}}(\omega) \\
& \text{si y sólo si} & \exists\omega_0 \in \Omega, \pi_{\mathcal{F}}(\omega_0) = 1 \\
& \text{si y sólo si} & \exists\omega_0 \in \Omega, \omega_0 \models \varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_n \\
& \text{si y sólo si} & \exists\omega_0 \in \Omega, \omega_0 \text{ es un modelo clásico de } \mathcal{F}^* \\
& \text{si y sólo si} & \mathcal{F}^* \text{ es consistente.}
\end{array}$$

□

3.9. DEFINICIÓN. Sea $\mathcal{F}$ una base de conocimiento de necesidad valuada, decimos que $\mathcal{F}$ es *parcialmente inconsistente*, si

$$\mathcal{F} \models (\bot\ \text{Incons}(\mathcal{F})) \text{ con } \text{Incons}(\mathcal{F}) > 0$$

Así, puesto que para cualquier fórmula φ tenemos $N(\varphi) \geq N(\bot)$, cualquier fórmula φ es deducible a partir de $\mathcal{F}$ con una valuación mayor o igual a $\text{Incons}(\mathcal{F})$. Esto significa que cualquier deducción tal que $\mathcal{F} \models (\varphi\ \alpha)$ con $\alpha = \text{Incons}(\mathcal{F})$ puede deberse sólo a la inconsistencia parcial de $\mathcal{F}$ y tal vez no tiene nada que ver con φ. Esas deducciones son llamadas *deducciones triviales*. Por lo tanto, las deducciones de fórmulas de necesidad valuadas $\mathcal{F} \models (\varphi\ \alpha)$ con $\alpha > \text{Incons}(\mathcal{F})$ que no son causadas por la inconsistencia parcial, son llamadas *deducciones no triviales*.

3.10. PROPOSICIÓN. [5] Sea $\mathcal{F}$ un conjunto de fórmulas posibilistas y sea $\text{Incons}(\mathcal{F}) = inc$, entonces

(1) $\mathcal{F}$ es semánticamente equivalente a $\mathcal{F}_{inc}$ y a $\mathcal{F}_{\overline{inc}} \cup \{(\bot\ inc)\}$
(2) $\mathcal{F}_{\overline{inc}}$ es consistente
(3) Si $\mathcal{F} \models (\psi\ \alpha)$ no trivialmente (*i.e.*, con $\alpha > inc$) entonces $\mathcal{F}_{\overline{inc}} \models (\psi\ \alpha)$.

DEMOSTRACIÓN. Ver [5]. □

3.11. PROPOSICIÓN (Inconsistencia parcial y $\alpha - corte$). [5]

1. Sea $\mathcal{F}$ un conjunto de fórmulas de necesidad valuadas, entonces $\text{Incons}(\mathcal{F}) = 0$ si y sólo si $\mathcal{F}^*$ es consistente en el sentido clásico.
2. $$\begin{array}{rcl}
\text{Incons}(\mathcal{F}) & = & \sup\{\alpha \mid \mathcal{F}^*_\alpha \text{ es inconsistente}\} \\
& = & \inf\{\alpha \mid \mathcal{F}^*_{\overline{\alpha}} \text{ es consistente}\}.
\end{array}$$
y estas cotas se alcanzan.

DEMOSTRACIÓN.

1. $\Rightarrow$) Sea $\mathcal{F} = \{(\varphi_i\ \alpha_i) \mid i = 1, \ldots, n\}$. De acuerdo con su definición, $\text{Incons}(\mathcal{F}) = 0$ si y sólo si $\pi_{\mathcal{F}}$ es normalizada, es decir, si y sólo si $\exists\ \omega^* \in \Omega$ tal que $\pi_{\mathcal{F}}(\omega^*) = 1$. Esto implica que $\omega^* \models \varphi_i$, para todo i. Por lo tanto, $\mathcal{F}^*$ es consistente.

$\Leftarrow$) Si $\mathcal{F}^*$ es consistente, entonces tiene un modelo $\overline{\omega}$; así,

$$\pi_{\mathcal{F}}(\overline{\omega}) = \inf\{1 - \alpha_i \mid \overline{\omega} \models \neg\varphi_i\} = 1$$

ya que, para todo i, $\overline{\omega} \models \varphi_i$. Así, $\pi_{\mathcal{F}}$ es normalizada e Incons $(\mathcal{F}) = 0$.

2. Se sigue de 1) y de los puntos 1) y 2) de la Proposición 3.10.

□

Los siguientes resultados [5], generalizan las versiones semánticas de los teoremas clásicos de la Deducción y de la Refutación:

3.12. Proposición (Teorema de la Deducción).

$$\mathcal{F} \cup \{(\varphi\ 1)\} \models (\psi\ \alpha) \text{ si y sólo si } \mathcal{F} \models (\varphi \to \psi\ \alpha).$$

Demostración.

($\Longrightarrow$) $\mathcal{F} \cup \{(\varphi\ 1)\} \models (\psi\ \alpha)$ implica que $N_{\mathcal{F}\cup\{(\varphi\ 1)\}}(\psi) \geq \alpha$, por el Corolario 3.6, se tiene que $\inf\left\{1 - \pi_{\mathcal{F}\cup\{(\varphi\ 1)\}}(\omega) \mid \omega \models \neg\psi\right\} \geq \alpha$. Así, para todo ω tal que $\omega \models \varphi \wedge \neg\psi$, se tiene que, $\pi_{\mathcal{F}}(\omega) \leq 1 - \alpha$, ya que $\pi_{\mathcal{F}\cup\{(\varphi\ 1)\}}(\omega) = \pi_{\mathcal{F}}(\omega)$ para cualesquier $\omega \models \varphi$. Y $N_{\mathcal{F}}(\varphi \to \psi) \geq \alpha$, ya que $\varphi \to \psi$ es equivalente a $\neg(\varphi \wedge \neg\psi)$. Por lo tanto, $\mathcal{F} \models (\varphi \to \psi\ \alpha)$, nuevamente, por el Corolario 3.6.

($\Longleftarrow$) $\mathcal{F} \models (\varphi \to \psi\ \alpha)$ implica que $\forall\pi \models \mathcal{F},\ N(\varphi \to \psi) \geq \alpha$. Luego, $\forall\pi \models \mathcal{F},\ N(\varphi) = 1$ implica $N(\psi) \geq \alpha$, dado que $N(\psi) \geq \min\{N(\varphi), N(\varphi \to \psi)\}$. Así, $\forall\pi \models \mathcal{F} \cup \{(\varphi\ 1)\},\ N(\psi) \geq \alpha$. Por lo tanto, $\mathcal{F} \cup \{(\varphi\ 1)\} \models (\psi\ \alpha)$.

□

3.13. Proposición (Teorema de la Refutación).

$$\mathcal{F} \models (\varphi\ \alpha) \text{ si y sólo si } \mathcal{F} \cup \{(\neg\varphi\ 1)\} \models (\bot\ \alpha)$$

o, equivalentemente,

$$Val(\varphi, \mathcal{F}) = \text{Incons}(\mathcal{F} \cup \{(\neg\varphi\ 1)\}).$$

Demostración. Apliquemos el Teorema de la Deducción, reemplazando φ por $\neg\varphi$ y ψ por $\bot$. Así, tenemos que $\mathcal{F} \cup \{(\neg\varphi\ 1)\} \models (\bot\ \alpha)$ si y sólo si $\mathcal{F} \models (\neg\varphi \to \bot\ \alpha)$; es decir, $\mathcal{F} \cup \{(\neg\varphi\ 1)\} \models (\bot\ \alpha)$ si y sólo si $\mathcal{F} \models (\varphi\ \alpha)$. □

La equivalencia se debe al hecho de que

$$\begin{aligned} Val(\varphi\ \mathcal{F}) &= \sup\{\alpha \in (0,1] \mid \mathcal{F} \models (\varphi\ \alpha)\} \\ &= \sup\{\alpha \in (0,1] \mid \mathcal{F} \cup \{(\neg\varphi\ 1)\} \models (\bot\ \alpha)\} \\ &= \text{Incons}(\mathcal{F} \cup \{(\neg\varphi\ 1)\}). \end{aligned}$$

Como complemento a uno de los teoremas anteriores, tenemos el siguiente resultado, el cual establece que en el proceso de deducción de una fórmula posibilista $(\varphi\ \alpha)$ sólo son "útiles" las fórmulas que tienen un grado mayor o igual que α.

3.14. Proposición. Sean $\mathcal{F}$ una base de conocimiento posibilista y $(\varphi\ \alpha)$ una fórmula de necesidad valuada. Entonces

$$\mathcal{F} \models (\varphi\ \alpha) \text{ si y sólo si } \mathcal{F}_\alpha \models (\varphi\ \alpha).$$

Demostración. Por el Teorema de Refutación, $\mathcal{F} \models (\varphi\ \alpha)$ es equivalente a Incons$(\mathcal{F} \cup \{(\neg\varphi\ 1)\}) \geq \alpha$; entonces, por la proposición 3.10, tenemos que Incons$(\mathcal{F}_\alpha \cup \{(\neg\varphi\ 1)\}) \geq \alpha$, es decir, $\mathcal{F}_\alpha \models (\varphi\ \alpha)$ por otra aplicación del Teorema de Refutación. El inverso se obtiene debido a que, $\mathcal{F}_\alpha \subseteq \mathcal{F}$. □

3.3. **Aspectos No-monótonos de la Lógica Posibilista Estándar.** Es posible definir [5] un operador deductivo no-monótono como sigue: Definimos el operador de deducción no trivial[3] $\mid\!\approx$ como

$$\mathcal{F} \mid\!\approx (\varphi\ \alpha) \text{ si y sólo si } \mathcal{F} \models (\varphi\ \alpha) \text{ y } \alpha > \text{Incons}(\mathcal{F}).$$

Con este operador puede suceder que $\mathcal{F} \mid\!\approx (\varphi\ \alpha)$ y $\mathcal{F} \cup \mathcal{F}' \not\mid\!\approx (\varphi\ \alpha)$.

3.15. Ejemplo. [5] Consideremos la siguiente base de conocimiento $\mathcal{F} = \{\Phi_1, \ldots, \Phi_6\}$, en relación a una elección cuyos candidatos son Mary y Pedro.

$\Phi_1 :=$ ((Electo(Pedro) $\vee$ Electo(Mary)) $\wedge$ ($\neg$Electo(Pedro) $\vee$ $\neg$Electo(Mary)) 1)
$\Phi_2 :=$ ($\forall x$ $\neg$Presidente-actual(x) $\vee$ Electo(x) 0.5)
$\Phi_3 :=$ (Presidente-actual(Mary) 1)
$\Phi_4 :=$ ($\forall x$ $\neg$Apoya(Juan, x) $\vee$ Electo(x) 0.6)
$\Phi_5 :=$ (Apoya(Juan, Mary) 0.2)
$\Phi_6 :=$ ($\forall x$ Víctima-de-un-incidente(x) $\vee$ $\neg$Electo(x) 0.7)

Se puede establecer que Incon($\mathcal{F}$) = 0, es decir, $\mathcal{F}$ es consistente. Estamos interesados en saber quién será electo y con qué grado de certeza maximal. Puede probarse que

$$\mathcal{F} \models (\text{Electo(Mary) } 0.5)$$
$$\mathcal{F} \models (\neg\text{Electo(Mary) } 0)$$
$$\mathcal{F} \models (\text{Electo(Pedro) } 0)$$
$$\mathcal{F} \models (\neg\text{Electo(Pedro) } 0.5)$$

es decir, es moderadamente cierto que Mary sea electa (o equivalentemente, que Pedro no lo sea); el grado de 0.5 es maximal, es decir, $Val(\mathcal{F}, \text{Electo (Mary)}) = 0.5$. Ya que Incons($\mathcal{F}$) = 0 también podemos escribir que

$$\mathcal{F} \mid\!\approx (\text{Electo(Mary) } 0.5)$$
$$\mathcal{F} \mid\!\approx (\neg\text{Electo(Pedro) } 0.5)$$

Entonces nos enteramos de que Mary es la víctima de un incidente (lo cual es información completamente cierta). Esto nos lleva a actualizar la base de conocimiento adicionándole a $\mathcal{F}$ la fórmula posibilista

$$\Phi_7 := (\text{Víctima-de-un-incidente(Mary) } 1)$$

Sea $\mathcal{F}_1$ la nueva base de conocimiento tal que $\mathcal{F}_1 = \mathcal{F} \cup \{\Phi_7\}$. Se puede probar que $\mathcal{F}_1$ es parcialmente inconsistente, con Incons($\mathcal{F}_1$) = 0.5. En efecto, la nueva información nos permite inferir que Mary no será electa, mientras que la base de conocimiento previa $\mathcal{F}$ nos permite inferir que Mary será electa. Se puede probar que

$$\mathcal{F}_1 \models (\text{Electo(Mary) } 0.5)$$
$$\mathcal{F}_1 \models (\neg\ \text{Electo(Pedro) } 0.5)$$

pero esas deducciones ahora son invalidadas por el umbral de inconsistencia y por lo tanto trivial. Usando el operador de deducción no trivial, todo esto se reduce a escribir que la deducción no trivial previa $\mathcal{F}|\!\approx$ (Electo (Mary) 0.5) y $\mathcal{F}|\!\approx$ ($\neg$Electo (Pedro) 0.5) ya no puede hacerse con $\mathcal{F}_1$. En este caso vemos un comportamiento no-monótono. Además de que tenemos

[3]Recordemos que la deducción $\mathcal{F} \models (\varphi\ \alpha)$ es *no trivial* si y sólo si $\alpha >$Incons($\mathcal{F}$)

$$\mathcal{F}_1 \models (\neg\text{Electo(Mary)}\ 0.7)$$
$$\mathcal{F}_1 \models (\ \text{Electo(Pedro)}\ 0.7)$$

y estas deducciones son no triviales ya que $0.7 > \text{Incons}(\mathcal{F}_1)$, es decir

$$\mathcal{F}_1 \not\approx (\neg\text{Electo(Mary)}\ 0.7)$$
$$\mathcal{F}_1 \not\approx (\ \text{Electo(Pedro)}\ 0.7)$$

lo cual significa ahora que Mary no es electa y Pedro sí. Por lo tanto, el actualizar la base de conocimiento nos lleva a una conclusión opuesta.

3.4. **Axiomatización.** En [5] se presenta un sistema axiomático para la Lógica Posibilista. Donde se proponen los siguientes tres axiomas,

(A1) $(\varphi \to (\psi \to \varphi)\ 1)$
(A2) $((\varphi \to (\psi \to \xi)) \to ((\varphi \to \psi) \to (\varphi \to \xi))\ 1)$
(A3) $((\neg\varphi \to \neg\psi) \to ((\neg\varphi \to \psi) \to \varphi)\ 1)$

junto con las reglas de inferencia

(GMP) $(\varphi\ \alpha),\ (\varphi \to \psi\ \beta) \vdash (\psi\ \min\{\alpha, \beta\})$
(S) $(\varphi\ \alpha) \vdash (\varphi\ \beta)$ si $\alpha \geq \beta$

Como puede observarse estos axiomas son los axiomas del Cálculo Proposicional Clásico ponderados con 1. Además, observe que la regla de inferencia (GMP) conserva fórmulas válidas, donde GMP es llamado Modus Ponens Graduado. Este sistema formal hace a la Lógica Posibilista *robusta* y *completa* con respecto a la semántica tolerante a la inconsistencia. Se tiene el siguiente resultado [5].

3.16. Proposición. Para cualquier conjunto de fórmulas posibilistas $\mathcal{F}$ tenemos

$$\mathcal{F} \models (\varphi\ \alpha) \text{ si y sólo si } \mathcal{F} \vdash (\varphi\ \alpha),$$

así, la Lógica Posibilista es *axiomatizable*.

Para la demostración de esta proposición, necesitamos el siguiente lema:

3.17. Lema. Sean $\mathcal{F}$ un conjunto de fórmulas de necesidad valuadas y $(\varphi\ \alpha)$ una fórmula de necesidad valuada. Entonces

$$\mathcal{F} \models (\varphi\ \alpha) \text{ si y sólo si } \mathcal{F}^*_\alpha \models \varphi \text{ en el sentido clásico.}$$

Demostración. (Del Lema 3.17):

$\mathcal{F} \models (\varphi\ \alpha)$	si y sólo si	$\mathcal{F}_\alpha \models (\varphi\ \alpha)$ (Proposición 3.14)
	si y sólo si	$\text{Incons}(\mathcal{F}_\alpha \cup \{(\neg\varphi\ 1)\}) \geq \alpha$ (Teorema de Refutación)
	si y sólo si	$\mathcal{F}^*_\alpha \cup \{\neg\varphi\}$ es inconsistente en el sentido clásico. (Proposición 3.11 (1))
	si y sólo si	$\mathcal{F}^*_\alpha \models \varphi$ (propiedad de deducción clásica).

□

Demostración. (De proposición 3.16):

⇒) Usando el Lema 3.17, $\mathcal{F} \models (\psi\ \alpha)$ es equivalente a $\mathcal{F}^*_\alpha \models \psi$. Entonces, ya que el sistema formal constituido por la parte sin peso del esquema de axiomas y de las reglas de inferencia (excepto (S) cuya parte no valuada es trivial), es bien conocido por ser un sistema formal tipo Hilbert completo y robusto para lógicas de primer orden, entonces existe una prueba de ψ a partir de $\mathcal{F}^*_\alpha$ en este sistema formal clásico. Entonces, considerando las valuaciones, la prueba obtenida por lo anterior es una prueba de $(\psi\ \gamma)$ a partir de $\mathcal{F}_\alpha$ por el sistema formal dado, con $\gamma \geq \alpha$. Por último, usando (S) obtenemos una prueba de $(\psi\ \alpha)$ a partir de $\mathcal{F}_\alpha$ y *a fortiori* de $\mathcal{F}$.

⇐) Por inducción sobre la longitud de la deducción.
Caso base: Si $\mathcal{F}$ deduce a $(\varphi\ \alpha)$ en un sólo paso, tenemos que es un axioma, es decir, $(\varphi\ 1)$ o que $(\varphi\ \alpha) \in \mathcal{F}$.
Para toda π, $\pi \models (\varphi\ 1)$, en particular $\forall \pi'$ tal que $\pi' \models \mathcal{F}'$ implica que $\pi' \models (\varphi\ 1)$. Ahora, si $(\varphi\ \alpha) \in \mathcal{F}$, para toda π, $\pi \models \mathcal{F}$ implica que $\pi \models (\varphi\ \alpha)$.
Supongamos que $\mathcal{F} \vdash (\psi\ \beta)$ y que la prueba requirío n pasos, entonces $\mathcal{F} \models (\psi\ \beta)$.
Supongamos que $\mathcal{F} \vdash (\varphi\ \alpha)$ se probo en $n+1$ pasos. Analicemos el último paso de la prueba de $(\varphi\ \alpha)$ supuesto $\mathcal{F}$.
Supongamos que $(\varphi\ \alpha)$ se obtuvo por GMP, es decir, que existen $(\psi \rightarrow \varphi\ \beta)$ y $(\psi\ \gamma)$ tal que $\mathcal{F} \vdash (\psi \rightarrow \varphi\ \beta)$ y $\mathcal{F} \vdash (\psi\ \gamma)$ y donde $\alpha \leq \min\{\beta, \gamma\}$. Por hipótesis inductiva, se tiene que $\mathcal{F} \models (\psi \rightarrow \varphi\ \beta)$ y $\mathcal{F} \models (\psi\ \gamma)$. Por el Lema anterior tenemos que $\mathcal{F}^*_\beta \models \psi \rightarrow \varphi$ y $\mathcal{F}^*_\gamma \models \psi$ en el sentido clásico. Así, obtenemos por Modus Ponens que $\mathcal{F}^*_\alpha \models \varphi$ si y sólo si $\mathcal{F} \models (\varphi\ \alpha)$.
Ahora, supongamos que $(\varphi\ \alpha)$ es aplicación de la regla (S) y sea $\mathcal{F} \vdash (\varphi\ \beta)$ con $\beta > \alpha$. Por hipótesis inductiva $\mathcal{F} \models (\varphi\ \beta)$, además sabemos que $(\varphi\ \beta) \models (\varphi\ \alpha)$, es decir, para todo π, $\pi \models \mathcal{F}$ implica que $\pi \models (\varphi\ \beta)$ y para todo π, $\pi \models (\varphi\ \beta)$ implica que $\pi \models (\varphi\ \alpha)$. Así, $\forall \pi$, $\pi \models \mathcal{F}$ implica que $\pi \models (\varphi\ \alpha)$ y por lo tanto $\mathcal{F} \models (\varphi\ \alpha)$.

□

Así, la Proposición 3.16 establece una semántica adecuada para la Lógica Posibilista Estándar, es decir, una semántica para la cual la Lógica Posibilista Estándar es Robusta y Completa.

Referencias

[1] D. Dubois and H. Prade. Belief Change and Possibility Theory. In P. Gärdenfors, ed., Belief Revision, 142-182, Cambridge University Press, 1992.

[2] D. Dubois and H. Prade. Fuzzy Sets and Systems: Theory and Applications. Mathematics in Sciences and Engineering Series, Vol. 144. Academic Press, New York, 1980.

[3] D. Dubois and H. Prade. Possibilistic logic, preferential models, non-motonicity and related issues. In Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91), Sydney, Australia, Aug. 24-30, pages 419-424, 1991.

[4] D. Dubois, J. Lang and H. Prade. Automated Reasoning Using Possibilisitic Logic: Semantics, Belief Revision, and Variable Certainty Weights. IEEE Trans.on Data and Knowledge Engineering, 1994.

[5] D. Dubois, J. Lang and H. Prade. Possibilistic Logic, Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, 439-513, Oxford University Press.

[6] D. Dubois, J. Lang and H. Prade. Towards Possibilistic Logic Programming. Proc. of ICLP91, 581-595.

[7] E. Mendelson. Introduction to Mathematical Logic. Fourth Edition, CHAPMAN-HALL-CRC, 2000.

[8] G. Wagner. Negation in Fuzzy and Possibilistic Logic Programs.

[9] L. A. Zadeh. Fuzzy sets and information granularity. In M.M. Gupta, R.K. Ragade, and R.R. Yager, editors, Advances in Fuzzy Set Theory and Applications, pages 3-18. North-Holland, Amsterdam, 1979.
[10] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1(1):3-28, 1978.
[11] R. R. Yager. An introduction to applications of possibility theory. Human Systems Management, 3:246-269, 1983.
[12] S. Benferhat, D. Dubois and H. Prade. Default Rules and Possibilistic Logic. Proceedings of KR' 92, 673-684.

arrazola@fcfm.buap.mx, oestrada2005@hotmail.com, jlavalleneitor@gmail.com, cambron99@hotmail.com

Capítulo 25

DEMOSTRACIÓN AUTOMÁTICA DE TEOREMAS

JOSÉ ARRAZOLA RAMÍREZ
JOSÉ DE JESÚS LAVALLE MARTÍNEZ*
JUAN PABLO MUÑOZ TORIZ
FCFM - BUAP
* FACULTAD DE CIENCIAS DE LA COMPUTACIÓN - BUAP

Resumen. En este trabajo se desarrolla un demostrador automático de teoremas en ML basado en el sistema Gentzen. Para esto se da una breve introducción a ML. Esto es: qué es ML, cuál es la sintaxis usada en este lenguaje de programación, así como algunos ejemplos de como programar en ML. Luego se explica en que consiste formalmente Gentzen y como codificar en ML las definiciones dadas. Finalmente se dan algunos ejemplos de ejecución del programa.

1. Introducción

ML[1] es un lenguaje de programación de propósito general de la familia de los lenguajes de programación funcional, desarrollado por Robin Milner y sus colaboradores a finales de la década de 1970 en la Universidad de Edimburgo. ML es un acrónimo de *Meta Lenguaje* dado que fue concebido como el lenguaje para desarrollar tácticas de demostración en el sistema LCF.

Entre las características de ML se incluyen: alto orden, evaluación por valor, álgebra de funciones, manejo automático de memoria por medio de recolección de basura, polimorfismo parametrizado, análisis estático de tipos, inferencia de tipos, tipos de datos algebráicos, llamada por patrones y manejo de excepciones. Esta combinación particular de conceptos hace que sea posible producir uno de los mejores lenguajes actualmente disponibles. Es un lenguaje funcional fuertemente tipificado, esto es que toda expresión en ML tiene un único tipo, es un lenguaje interpretado no compilado.

Una función en ML se define poniendo la palabra reservada **fun**, seguida por los parámetros de la función, el símbolo = y finalmente el cuerpo de la función.

1.1. Ejemplo. Definir en ML la función $sucesor(x) = x + 1$ se puede hacer de las siguientes maneras:

```
fun sucesor  x = x+1;
fun sucesor1 x:int = x+1;
fun sucesor2 x = x+1:int;
fun sucesor3 x:int = x+1:int;
```

Trabajo apoyado por el proyecto VIEP-BUAP, Programación Lógica Posibilista y Teoría de la Argumentación.

1.2. Ejemplo. La función factorial definida mediante

$$factorial(n) = \begin{cases} 1 & \text{si } n = 0 \\ n * factorial(n-1) & \text{si } n > 0 \end{cases}$$

se puede codificar en ML por:

```
fun factorial 0 = 1
|   factorial n = n * factorial(n-1);
```

Por otro lado, como ya se dijo anteriormente ML es un lenguaje de programación fuertemente tipificado, así que debemos caracterizar mediante tipos el sistema que queremos construir. Para lo cual utilizamos el constructor de tipo **datatype**.

1.3. Ejemplo. Supongamos que deseamos construir una lista de cadenas, la cual se define recursivamente como:

- Una lista de cadenas está vacía o
- Tiene una cadena como primer elemento y un resto que es una lista de cadenas.

Expresado en ML:

```
datatype strlist = nul | ht of string * strlist;
```

Con lo anterior no sólo hemos definido el tipo de datos strlist, también hemos creado dos constructores para dicho tipo, a saber nul de aridad 0 y ht de aridad 2. Así para construir una lista de cadenas tenemos que basarnos en tales constructores, como en

```
ht("cadena1",ht("cadena2",ht("cadena3",nul)));
```

Al introducir la línea anterior, ML nos dirá que el tipo al que pertenece es a strlist.

ML también cuenta con una forma de crear alias para tipos usando **type**, pero con ésta no se define constructor alguno. Como ejemplo podemos escribir

```
type float = real;
```

En este caso float se puede usar posteriormente como un sinónimo del tipo real. Una primera forma de ocultar código en ML es usando **let** e **in** como a continuación se muestra:

```
fun invierteLista(Lista) =
  let
    fun invierteLPA(nul,pa) = pa
    |   invierteLPA(ht(h,t),pa) = invierteLPA(t,ht(h,pa))
  in
    invierteLPA(Lista,nul)
  end;
```

En este caso la función invierteLista delega su trabajo a una segunda función llamada invierteLPA. La función invierteLPA se encuentra oculta entre las palabras reservadas **let** e **in**. La palabra reservada **in** indica el fin del **let** y después de ella comienza el cuerpo de la función principal.

Los lenguajes de la familia ML se usan para el diseño y manipulación de cualquier lenguaje, de ahí su nombre *Meta Language* (compiladores, analizadores, demostradores de teoremas, etc.).

2. Implementación

Primero definiremos formalmente el lenguaje de la lógica proposicional [2][3]. Para ello debemos definir su alfabeto y dar sus reglas sintácticas de formación.

2.1. Definición. El alfabeto del lenguaje de la lógica proposicional está conformado por:

1. Paréntesis: (,);
2. Conectivos lógicos: $\neg$, $\wedge$, $\vee$, $\rightarrow$;
3. Letras proposicionales: Cualquier letra con o sin subíndices.

2.2. Definición. Los elementos del lenguaje de la lógica proposicional, que llamaremos proposiciones, se definitionnen mediante:

1. Toda letra proposicional es una proposición,
2. Si $\mathcal{A}$ y $\mathcal{B}$ son proposiciones entonces $(\neg\mathcal{A})$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$ y $(\mathcal{A} \rightarrow \mathcal{B})$ también son proposiciones.

Lo cual podemos escribir, prácticamente tal cual, en ML de la siguiente manera:

```
datatype Prop = lp of string |
              nega of Prop |
                  conj of Prop * Prop |
                  disy of Prop * Prop |
                  impl of Prop * Prop;
```

Así tenemos cinco constructores para el tipo Prop:

1. lp de aridad uno y cadenas como argumento,
2. nega de aridad uno y proposiciones como argumento,
3. conj de aridad dos y proposiciones como argumentos,
4. disy de aridad dos y proposiciones como argumentos,
5. impl de aridad dos y proposiciones como argumentos.

Definimos Sec para crear listas de proposiciones, ya que en ML está interconstruido el tipo lista y es polimorfo utilizamos **type** como alias para una lista de proposicones. Para fines prácticos pensaremos en Sec como un conjunto de proposiciones.

```
type Sec = Prop list;
```

El sistema de Gentzen [2][3] nos permite saber si una proposición es una tautología, en caso de que no sea una tautología también nos permite saber bajo que asignaciones, de valores de verdad a las letras proposicionales, la proposición es falsificable, se basa en el concepto de secuente. Un secuente es una pareja (Γ, Δ) de secuencias (o conjuntos, posiblemente vacíos) de proposiciones: $\Gamma =< A_1, ..., A_n >$ y $\Delta =< B_1, ..., B_m >$, decimos que Γ y Δ son el antecedente y el consecuente, respectivamente. Por brevedad escribiremos los secuentes como $\Gamma \rightsquigarrow \Delta$ y en las secuencias no usaremos paréntesis angulares.

Por lo tanto el constructor de tipo Secuente se define de la siguiente manera:

```
datatype Secuente = lr of Sec * Sec;
```

Las reglas de inferencia están formadas como un conjunto de secuentes premisas sobre una raya horizontal y un único secuente conclusión bajo la raya.

$$\frac{premisa_1, premisa_2, \ldots, premisa_n}{conclusion}\ NombreDeLaRegla$$

Las premisas resultan de la aplicación de la definición de la regla a la conclusión. Las reglas de inferencia se clasifican en dos categorías: las que operan sobre el antecedente ($* \rightsquigarrow$) y las que operan sobre el consecuente ($\rightsquigarrow *$): cada regla descompone a la proposición principal en subproposiciones que son colocadas ya sea en el antecedente o consecuente de las premisas, pudiendo incluso dividir un secuente (conclusión) en dos (premisas). Las premisas obtenidas después de esta operación pueden contener o no operadores lógicos, en caso de contenerlos las reglas de inferencia vuelven a ser aplicadas. Ello naturalmente permite representar a las pruebas como árboles.

2.3. Definición. El sistema de Gentzen se define a continuación:

1. Γ, Δ representan secuencias arbitrarias de fórmulas (conjuntos),
2. A, B representan proposiciones arbitrarias,
3. El único axioma es:
$$\frac{}{\Gamma, A \rightsquigarrow A, \Delta}\ Ax$$
4. Las reglas de inferencia de Gentzen son:

$$\frac{\Gamma \rightsquigarrow A, \Delta}{\Gamma, (\neg A) \rightsquigarrow \Delta}\ \neg \rightsquigarrow \qquad \frac{\Gamma, A \rightsquigarrow \Delta}{\Gamma \rightsquigarrow (\neg A), \Delta}\ \rightsquigarrow \neg$$

$$\frac{\Gamma, A, B \rightsquigarrow \Delta}{\Gamma, (A \wedge B) \rightsquigarrow \Delta}\ \wedge \rightsquigarrow \qquad \frac{\Gamma \rightsquigarrow A, \Delta \quad \Gamma \rightsquigarrow B, \Delta}{\Gamma \rightsquigarrow (A \wedge B), \Delta}\ \rightsquigarrow \wedge$$

$$\frac{\Gamma, A \rightsquigarrow \Delta \quad \Gamma, B \rightsquigarrow \Delta}{\Gamma, (A \vee B) \rightsquigarrow \Delta}\ \vee \rightsquigarrow \qquad \frac{\Gamma \rightsquigarrow A, B, \Delta}{\Gamma \rightsquigarrow (A \vee B), \Delta}\ \rightsquigarrow \vee$$

$$\frac{\Gamma \rightsquigarrow A, \Delta \quad \Gamma, B \rightsquigarrow \Delta}{\Gamma, (A \to B) \rightsquigarrow \Delta}\ \to \rightsquigarrow \qquad \frac{\Gamma, A \rightsquigarrow B, \Delta}{\Gamma \rightsquigarrow (A \to B), \Delta}\ \rightsquigarrow \to$$

Note que en el sistema de Gentzen tenemos un axioma (Ax) y dos formas de reglas de inferencia, aquellas que dado un secuente conclusión obtenemos sólo un secuente premisa llamémosles InRuUno (coloquialmente las que no bifurcan) y aquellas que dado un secuente conclusión obtenemos dos secuentes premisas llamémosles InRuDos (coloquialmente las que bifurcan), lo cual podemos definir en ML mediante:

```
datatype SistemaG = Ax of Secuente |
                    InRuUno of Secuente |
                    InRuDos of Secuente * Secuente;
```

Hasta aquí se ha caracterizado el sistema de Gentzen, es decir, hemos creado un tipo SistemaG cuyos elementos son la base para poder hacer demostraciones en lógica proposicional. A continuación vamos a definir la función reduce = fun: Secuente → SistemaG, ésta se encarga de aplicar las reglas de inferencia, en caso de no poder aplicar una regla de inferencia, debido a que no encontró operadores lógicos a la cabeza de la lista, se llama a la función decide. El código es el siguiente:

```
fun reduce(lr(nega(a)::gama,delta)) =
    InRuUno(lr(gama,a::delta)) |
    reduce(lr(conj(a,b)::gama,delta)) =
    InRuUno(lr(a::b::gama,delta)) |
    reduce(lr(gama,nega(a)::delta))   =
    InRuUno(lr(a::gama,delta)) |
    reduce(lr(gama,disy(a,b)::delta)) =
    InRuUno(lr(gama,a::b::delta)) |
    reduce(lr(gama,impl(a,b)::delta)) =
    InRuUno(lr(a::gama,b::delta)) |
    reduce(lr(disy(a,b)::gama,delta)) =
    InRuDos(lr(a::gama,delta),lr(b::gama,delta)) |
    reduce(lr(impl(a,b)::gama,delta)) =
    InRuDos(lr(b::gama,delta),lr(gama,a::delta)) |
    reduce(lr(gama,conj(a,b)::delta)) =
    InRuDos(lr(gama,a::delta),lr(gama,b::delta)) |
    reduce(sec)                       = decide(sec);
```

2.4. Ejemplo. La línea

```
reduce (lr(conj(a,b)::gama, delta)) =
InRuUno(lr(a::b::gama, delta))
```

es la codificación de la regla

$$\frac{\Gamma, A, B \rightsquigarrow \Delta}{\Gamma, (A \wedge B) \rightsquigarrow \Delta} \wedge \rightsquigarrow$$

y la línea

```
reduce(lr(impl(a,b)::gama,delta)) =
InRuDos(lr(gama,a::delta),lr(b::gama,delta))
```

es la codificación de la regla

$$\frac{\Gamma \rightsquigarrow A, \Delta \quad \Gamma, B \rightsquigarrow \Delta}{\Gamma, (A \rightarrow B) \rightsquigarrow \Delta} \rightarrow\rightsquigarrow$$

Como se puede observar una instancia de nuestro único axioma es cualquier secuente $\Gamma \rightsquigarrow \Delta$ tal que $\Gamma \cap \Delta \neq \varnothing$, esto es, contienen alguna proposición en común.

En ML la forma de determinar si una proposición es un teorema será por medio de la función prueba: SistemaG → bool. Esta función se llama recursivamente mientras reciba elementos de tipo InRuUno o InRuDos, es decir las premisas aún contienen conectivos lógicos, con los cuales llama a la función reduce que es la función que aplica las reglas de inferencia. Cuando prueba recibe un elemento de tipo Ax (premisa sin operadores lógicos) entonces llamará a la función intersecta la cual devuelve true si los conjuntos no son disjuntos, es decir la conclusión es una axioma, y false en otro caso.

```
fun prueba(Ax(lr(left ,right))) =
    intersecta(left ,right) |
    prueba(InRuUno(up))         =
```

```
    prueba(printVal(reduce(up))) |
    prueba(InRuDos(up1,up2))     =
    prueba(printVal(reduce(up1))) andalso
    prueba(printVal(reduce(up2)));
```

3. Funciones auxiliares

En esta sección presentaremos funciones auxiliares que son llamadas por las discutidas anteriormente.

3.1. Función intersecta: Secuente → bool. Recordemos que Secuente es de tipo Sec ∗ Sec, es decir tiene una lista del lado izquierdo y otra del lado derecho, así intersecta busca si la parte derecha tiene elementos en común con la parte izquierda, si este es el caso devuelve true, en caso contrario devuelve false. Esto es porque si se tiene una o mas proposiciones iguales del lado derecho y del lado izquierdo se llega a una contradicción, lo cual quiere decir que no hay valores de verdad que hagan falsa a la proposición inicial.

```
fun intersecta([],_)    = false
|   intersecta(h::t,r) = let
                            fun esta(_,[]) = false
                            |   esta(x,h::t) = x = h
                        orelse esta(x,t);
                        in
                          esta(h,r)
                    orelse intersecta(t,r)
                        end;
```

3.2. Función decide: Secuente → SistemaG. La función decide se llama cuando no se ha encontrado un conectivo a la cabeza de gama ni a la cabeza de delta, pero podría haber alguno todavía en el resto de alguna de las dos secuencias, por lo tanto esta función tiene como propósito buscar un conectivo y de hallarlo dejarlo a la cabeza de la secuencia en la que lo halló. Si no lo encuentra en ninguna de las dos secuencias se sospecha de que se trata de una instancia del único axioma.

```
fun decide(lr(gama,delta)) =
        let
          fun hayoperador(nil,pa) = (false,pa)
          |   hayoperador(lp(a)::t,pa) =
              hayoperador(t,lp(a)::pa)
          |   hayoperador(x,pa) = (true,x@pa)
          val (valor,gama1) =
              hayoperador(gama,nil)
        in
          if valor
          then InRuUno(lr(gama1,delta))
          else
            let
              val (valor1,delta1) =
```

```
        hayoperador(delta, nil)
      in
        if valor1
        then InRuUno(lr(gama, delta1))
        else Ax(lr(gama, delta))
      end
end;
```

Con lo cual tenemos el demostrador completo del sistema de Gentzen.

4. EJEMPLOS

A continuación damos dos ejemplos de como funciona el demostrador y sus correspondientes árboles de demostración.

4.1. EJEMPLO. Para saber si la proposición $((p \to q) \to ((\neg q) \to (\neg p)))$ es siempre verdadera, usando el sistema de Gentzen, tenemos que construir un árbol de deducción para el secuente $\rightsquigarrow ((p \to q) \to ((\neg q) \to (\neg p)))$ o usar el demostrador automático que hemos desarrollado de la siguiente manera:

```
prueba(InRuUno(lr([],[impl(impl(lp("p"),lp("q")),
impl(nega(lp("q")),nega(lp("p"))))])));
```

Obteniendo

```
InRuUno(lr([impl(lp "p", lp "q")],
           [impl(nega(lp "q"), nega(lp "p"))]))
InRuUno(lr([nega(lp "q"), impl(lp "p", lp "q")],
           [nega(lp "p")]))
InRuUno(lr([impl(lp "p", lp "q")],
           [lp "q", nega(lp "p")]))
InRuDos(lr([lp "q"], [lp "q", nega(lp "p")]),
        lr([], [lp "p", lp "q", nega(lp "p")]))
InRuUno(lr([lp "q"],
           [nega(lp "p"), lp "q"]))
InRuUno(lr([lp "p", lp "q"],
           [lp "q"]))
Ax(lr([lp "p", lp "q"],
      [lp "q"]))
InRuUno(lr([],
           [nega(lp "p"), lp "q", lp "p"]))
InRuUno(lr([lp "p"],
           [lp "q", lp "p"]))
Ax(lr([lp "p"],
      [lp "q", lp "p"]))
> val it = true : bool
```

Es decir, la proposición $((p \to q) \to ((\neg q) \to (\neg p)))$ es una tautología. Su árbol de deducción como lo hariamos las personas es el siguiente:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overline{p, q \rightsquigarrow q}\ Ax}{q \rightsquigarrow q, (\neg p)}\ \rightsquigarrow\neg \quad \dfrac{\overline{p \rightsquigarrow p, q}\ Ax}{\rightsquigarrow p, q, (\neg p)}\ \rightsquigarrow\neg}{(p \to q) \rightsquigarrow q, (\neg p)}\ \to\rightsquigarrow}{(\neg q), (p \to q) \rightsquigarrow (\neg p)}\ \neg\rightsquigarrow}{(p \to q) \rightsquigarrow ((\neg q) \to (\neg p))}\ \rightsquigarrow\to}{\rightsquigarrow ((p \to q) \to ((\neg q) \to (\neg p)))}\ \rightsquigarrow\to$$

4.2. Ejemplo. De la misma forma si queremos saber si la proposición $((\neg p) \wedge p)$ es siempre verdadera, podemos construir un árbol de deducción para el secuente $\rightsquigarrow ((\neg p) \wedge p)$ o introducirla al programa de la siguiente manera:

```
prueba(InRuUno(lr([],[conj(nega(lp("p")),lp("p"))]))));
```

Obteniendo

```
InRuDos(lr([], [nega(lp "p")]),
        lr([], [lp "p"])))
InRuUno(lr([lp "p"], []))
Ax(lr([lp "p"], []))>
val it = false : bool
```

Es decir, la proposición $((\neg p) \wedge p)$ no es una tautología y se puede falsificar cuando p toma el valor veritativo verdadero (aunque también cuando p es falso, ya que en realidad es una contradicción). Su árbol de deducción es el siguiente:

$$\dfrac{\dfrac{\dfrac{p \rightsquigarrow}{\rightsquigarrow (\neg p)}\ \rightsquigarrow\neg \qquad \rightsquigarrow p}{\rightsquigarrow ((\neg p) \wedge p)}\ \rightsquigarrow\wedge$$

Conclusión

Como se ha mostrado ML es muy útil para programar objetos definidos matemáticamente, ya que nos permite modelar fácilmente el alfabeto, la sintaxis y la semántica que se le da a dichos objetos. ML permite tener código elegante y simple, también facilita su entendimiento, interpretación y mantenimiento.

Referencias

[1] Robert Harper; Programming in Standard ML; Carnegie Mellon University, Spring 2005, `http://www.cs.cmu.edu/~rwh/smlbook/offline.pdf`, last visited September 2010.

[2] Jean Gallier; Logic for Computer Science: Foundations of Automatic Theorem Proving; 2003, `http://www.cis.upenn.edu/~jean/gbooks/logic.html`, last visited September 2010.

[3] Steve Reeves and Mike Clarke; Logic for Computer Science, Addison-Wesley Publishers Ltd. 1990, `http://www.cs.waikato.ac.nz/~stever/LCS.pdf`, last visited September 2010.

Facultad de Ciencias Físico Matemáticas, BUAP.
Avenida San Claudio y 18 Sur, Colonia San Manuel,
Puebla, Pue. C.P. 72570.
Facultad de Ciencias de la Computación, BUAP.
Av. 14 Sur y Av. San Claudio, CU, San Manuel
Puebla, Puebla, México. `arrazola@fcfm.buap.mx`, `jlavalle@cs.buap.mx`, `jp_190999@hotmail.`

Topología

Capítulo 26

LA TOPOLOGÍA DE LOS HIPERESPACIOS

VIANEY CÓRDOVA SALAZAR
DAVID HERRERA CARRASCO
FERNANDO MACÍAS ROMERO
FCFM - BUAP

Resumen. El material que contiene este trabajo pertenece a la rama de la topología denominada Teoría de los Continuos e Hiperespacios. Un continuo es un espacio métrico, no vacío, compacto y conexo. Para un continuo X se define el hiperespacio $2^X = \{A \subset X : A \text{ es cerrado en } X \text{ y no vacío}\}$ con la topología generada por la métrica de Hausdorff. En este artículo analizamos las propiedades de dicha métrica y vemos que la Topología generada por ésta coincide con la Topología de Vietoris.

1. Introducción

En términos generales, la Teoría de los Continuos se encarga de estudiar las propiedades de los espacios métricos, no vacíos, compactos y conexos; precisamente, un espacio con estas cuatro características es un continuo; la Teoría de los Hiperespacios se dedica a estudiar ciertos subconjuntos del conjunto potencia de un continuo, a los cuales se les asigna una métrica (conocida como la métrica de Hausdorff) y a dichos subconjuntos distinguidos se les denomina hiperespacios. los primeros trabajos se deben a los topólogos alemanes L. Vietoris y F. Hausdorff. Vietoris determinó las propiedades básicas de los hiperespacios y Hausdorff definió una métrica para éstos. En este artículo exponemos los conceptos y demostramos resultados de los hiperespacios de continuos, necesarios para dotar de una métrica al conjunto de todos los subconjuntos cerrados y no vacíos de un continuo dado, es decir, dado un continuo X, dotamos al conjunto $2^X = \{A \subset X : A \text{ es cerrado y no vacío}\}$ de una métrica que se conoce como la métrica de Hausdorff. También dotamos de una topología a este conjunto, llamada topología de Vietoris. El objetivo principal de este artículo es probar que la topología de Vietoris coincide con la topología generada por la métrica de Hausdorff.

2. Métrica de Hausdorff

Un *continuo* es un espacio métrico, no vacío, compacto y conexo. A lo largo de todo este trabajo, la letra d representará la métrica para un continuo X. Los *hiperespacios* son ciertas familias de subconjuntos de X, con alguna característica particular. Dado un continuo X, consideraremos los siguientes hiperespacios de X.

$2^X = \{A \subset X : A \text{ es cerrado en } X \text{ y no vacío}\}$,
$C(X) = \{A \in 2^X : A \text{ es conexo}\}$.

Como es usual, los siguientes símbolos $\mathbb{R}$, $\mathbb{R}^+$ y $\mathbb{N}$ denotarán el conjunto de los números reales, números reales positivos y números naturales, respectivamente. Las nociones generales de topología fueron tomadas de la referencia [1]. El contenido de este material fue inspirado de los ejercicios 2.2, 2.3, 2.6, 2.7 y 2.8 del Capítulo

2 de la referencia [2]. El libro indicado por [3] al final, es el libro básico del buen continuero, en el cual el lector encontrará más información alrededor de la Teoría de los Continuos.

Sea X un continuo, definimos la *bola abierta* en X con centro en un punto $x \in X$ y de radio $\varepsilon > 0$, como sigue:

$$B(\varepsilon, x) = \{y \in X : d(x, y) < \varepsilon\}.$$

Dados un continuo X y subconjuntos A y B de X, denotamos por $d(A, B)$ la distancia de A a B que se define como:

$$d(A, B) = ínf\{d(a, b) : a \in A \text{ y } b \in B\}$$

y la distancia de un punto p a un conjunto C es

$$d(p, C) = d(\{p\}, C).$$

Por otra parte, dado un subconjunto cerrado A de un continuo X, definimos la *nube en X con centro en A y de radio $\varepsilon > 0$*, denotada por $N(\varepsilon, A)$, como sigue:

$$N(\varepsilon, A) = \{x \in X : \textit{existe } a \in A \textit{ tal que } d(a, x) < \varepsilon\}.$$

2.1. Lema. Sea X un continuo. Entonces se cumple lo siguiente:

1. Si $\varepsilon > 0$ y $A \in 2^X$, entonces
   (a) $A \subset N(\varepsilon, A)$,
   (b) $N(\varepsilon, A) = \bigcup_{a \in A} B(\varepsilon, a)$. Así, $N(\varepsilon, A)$ es un abierto en X,
   (c) $N(\delta, A) \subset N(\varepsilon, A)$ para cada $\delta > 0$ tal que $\delta < \varepsilon$, y
   (d) $N(\varepsilon, A) = \bigcup\{N(\delta, A) : \delta > 0, \delta < \varepsilon\}$.
2. Si $A \in 2^X$ y U es un subconjunto abierto en X tal que $A \subset U$, entonces existe $\varepsilon > 0$ tal que
$$A \subset N(\varepsilon, A) \text{ y } N(\varepsilon, A) \subset U.$$
3. Sean A, $B \in 2^X$. Si $0 < \delta \leq \varepsilon$ y $A \subset B$, entonces $N(\delta, A) \subset N(\varepsilon, B)$.
4. Sea $\varepsilon > 0$. Entonces para cualesquiera A, $B \in 2^X$, se tiene que
$$N(\varepsilon, A) \cup N(\varepsilon, B) = N(\varepsilon, A \cup B).$$
5. Si A, $B \in 2^X$ tal que $A \cap B = \emptyset$, entonces existe $\varepsilon > 0$ tal que
$$N(\varepsilon, A) \cap N(\varepsilon, B) = \emptyset.$$

Demostración.

Veamos 1.

(a) Es claro que se cumple.

(b) Sea $x \in N(\varepsilon, A)$. Entonces existe $a \in A$ tal que $d(a, x) < \varepsilon$, luego $x \in B(\varepsilon, a)$, así, $x \in \bigcup_{a \in A} B(\varepsilon, a)$. Por lo tanto

$$N(\varepsilon, A) \subset \bigcup_{a \in A} B(\varepsilon, a). \tag{2.1}$$

Ahora bien, sea $x \in \bigcup_{a \in A} B(\varepsilon, a)$. Entonces existe $a \in A$ tal que $x \in B(\varepsilon, a)$, así, $d(a, x) < \varepsilon$, es decir, $x \in N(\varepsilon, A)$. Por lo tanto

$$\bigcup_{a \in A} B(\varepsilon, a) \subset N(\varepsilon, A). \tag{2.2}$$

De (2.1) y (2.2), se tiene la igualdad.

(c) Sean $\delta > 0$ tal que $\delta < \varepsilon$ y $x \in N(\delta, A)$. Entonces existe $a \in A$ tal que $d(a, x) < \delta$. Como $\delta < \varepsilon$, se tiene que $d(a, x) < \varepsilon$, es decir, $x \in N(\varepsilon, A)$. Por lo tanto, $N(\delta, A) \subset N(\varepsilon, A)$.

(d) Sea $x \in N(\varepsilon, A)$. Entonces existe $a \in A$ tal que $d(a, x) < \varepsilon$. Sea $\delta > 0$ tal que $d(a, x) < \delta < \varepsilon$. Así, $d(a, x) < \delta$, de manera que $x \in N(\delta, A)$. Luego, $x \in \bigcup\{N(\delta, A) : \delta > 0,\ \delta < \varepsilon\}$. Por lo tanto

$$N(\varepsilon, A) \subset \bigcup\{N(\delta, A) : \delta > 0,\ \delta < \varepsilon\}. \tag{2.3}$$

Por (c), tenemos que

$$\bigcup_{\delta<\varepsilon} N(\delta, A) \subset N(\varepsilon, A). \tag{2.4}$$

De (2.3) y (2.4), concluimos (d).

2. Como $A \subset U$, se sigue que $A \cap (X \setminus U) = \emptyset$. Luego, A y $X \setminus U$ son cerrados en X y, así, compactos. De manera que $d(A, X \setminus U) > 0$. Sea $\varepsilon = \frac{d(A, X \setminus U)}{2}$, notemos que $\varepsilon > 0$. Se sigue que $N(\varepsilon, A) \subset U$. En efecto, si $x \in N(\varepsilon, A)$, entonces existe $a \in A$ tal que $d(a, x) < \varepsilon$. Así, $x \in B(\varepsilon, a)$. De manera que $x \in U$. En caso contrario, es decir, si $x \in X \setminus U$, entonces $B(\varepsilon, a) \cap (X \setminus U) \neq \emptyset$, de modo que existe $z \in B(\varepsilon, a) \cap (X \setminus U)$. Así, $d(a, z) < \varepsilon$ y $\varepsilon < d(A, X \setminus U)$, lo cual es una contradicción. Por lo tanto, $x \in U$. En consecuencia, $N(\varepsilon, A) \subset U$. Ahora, por 1.(a), se tiene que $A \subset N(\varepsilon, A)$ y $N(\varepsilon, A) \subset U$.
3. Sea $x \in N(\delta, A)$. Entonces existe $a \in A$ tal que $d(a, x) < \delta$. Como $\delta \leq \varepsilon$, se sigue que $d(a, x) < \varepsilon$. Puesto que $A \subset B$, implicamos que $a \in B$, de donde $x \in N(\varepsilon, B)$. Por lo tanto, $N(\delta, A) \subset N(\varepsilon, B)$.
4. Por 3, inferimos que $N(\varepsilon, A) \subset N(\varepsilon, A \cup B)$ y $N(\varepsilon, B) \subset N(\varepsilon, A \cup B)$. Así,

$$N(\varepsilon, A) \cup N(\varepsilon, B) \subset N(\varepsilon, A \cup B). \tag{2.5}$$

Ahora, sea $z \in N(\varepsilon, A \cup B)$. Luego, existe $b \in A \cup B$ tal que $d(b, z) < \varepsilon$. Tenemos dos casos:

(i) Si $b \in A$, entonces $z \in N(\varepsilon, A)$.

(ii) Si $b \in B$, entonces $z \in N(\varepsilon, B)$.

En ambos casos, se tiene que $z \in N(\varepsilon, A) \cup N(\varepsilon, B)$. Por lo tanto

$$N(\varepsilon, A \cup B) \subset N(\varepsilon, A) \cup N(\varepsilon, B). \tag{2.6}$$

De (2.5) y (2.6), concluimos que

$$N(\varepsilon, A) \cup N(\varepsilon, B) = N(\varepsilon, A \cup B).$$

5. Supongamos que para cada $\varepsilon > 0$, tenemos que $N(\varepsilon, A) \cap N(\varepsilon, B) \neq \emptyset$. Puesto que $A \cap B = \emptyset$ y A, B son compactos en X, se tiene que $d(A, B) > 0$. Sea $\varepsilon = \frac{d(A,B)}{2}$, notemos que $\varepsilon > 0$. Por lo supuesto, deducimos que $N(\varepsilon, A) \cap N(\varepsilon, B) \neq \emptyset$. Luego, existe $z \in N(\varepsilon, A) \cap N(\varepsilon, B)$. Así, existen $a \in A$ y $b \in B$ tales que $d(a, z) < \varepsilon$ y $d(b, z) < \varepsilon$. Luego, aplicando la desigualdad del triángulo, $d(a, b) \leq d(a, z) + d(z, b) < 2\varepsilon = d(A, B)$, de manera que $d(a, b) < d(A, B)$, lo cual es una contradicción. Con esto demostramos 5. □

2.2. Definición. Sea A un subconjunto no vacío de un continuo X. El *diámetro* de A, denotado por $diám(A)$, es:

$$sup\{d(a, b) : a, b \in A\}.$$

2.3. Teorema. Sea X un continuo. La función diám : $2^X \longrightarrow [0,\infty)$, dada por diám$(A) = sup\{d(a,b) : a,b \in A\}$ es una función continua.

Demostración. Sean $\varepsilon > 0$, A, $B \in 2^X$ y $\delta = \frac{\varepsilon}{2} > 0$. Si $H(A,B) < \delta$, por el Lema 2.1, tenemos que $A \subset N(\delta, B)$ y $B \subset N(\delta, A)$. Como A es compacto, existen a_1, $a_2 \in A$ tales que $diám(A) = d(a_1, a_2)$. Luego, existen b_1, $b_2 \in B$ tales que $d(a_1, b_1) < \delta$ y $d(a_2, b_2) < \delta$ (dado que $A \subset N(\delta, B)$). Por la desigualdad del triángulo:

$$diám(A) = d(a_1,a_2) \leq d(a_1,b_1) + d(a_2,b_2) + d(b_1,b_2)$$
$$< 2\delta + d(b_1,b_2) = \varepsilon + d(b_1,b_2).$$

Como $d(b_1,b_2) < diám(B)$, tenemos que

$$\varepsilon + d(b_1,b_2) < \varepsilon + diám(B).$$

Por lo tanto,

$$diám(A) - diám(B) < \varepsilon. \tag{2.7}$$

De manera análoga, como B es compacto, existen b_3, $b_4 \in B$ tales que $diám(B) = d(b_3, b_4)$. Luego, existen a_3, $a_4 \in A$ tales que $d(b_3, a_3) < \delta$ y $d(b_4, a_4) < \delta$ (como $B \subset N(\delta, A)$). Por la desigualdad del triángulo:

$$diám(B) = d(b_3,b_4) \leq d(b_3,a_3) + d(b_4,a_4) + d(a_3,a_4)$$
$$< 2\delta + d(a_3,a_4) = \varepsilon + d(a_3,a_4).$$

Como $d(a_3,a_4) < diám(A)$, tenemos que

$$\varepsilon + d(a_3,a_4) < \varepsilon + diám(A).$$

Por lo tanto,

$$diám(B) - diám(A) < \varepsilon.$$

Así,

$$-\varepsilon < diám(A) - diám(B). \tag{2.8}$$

Por (2.7) y (2.8), concluimos que:

$$| \, diám(A) - diám(B) \, | < \varepsilon.$$

Por tanto, la función *diám* es uniformemente continua, y así continua. □

2.4. Notación. Para cada A, $B \in 2^X$ sean

$$E(A,B) = \{\varepsilon > 0 : A \subset N(\varepsilon, B) \ y \ B \subset N(\varepsilon, A)\}$$

y

$$E(A,B) + E(B,C) = \{\varepsilon + \delta : \varepsilon \in E(A,B) \ y \ \delta \in E(B,C)\}.$$

2.5. Teorema. Sea X un continuo. La función $H : 2^X \times 2^X \longrightarrow \mathbb{R}^+ \cup \{0\}$ definida, para cada A, $B \in 2^X$, por $H(A,B) = ínf E(A,B)$ es una métrica para 2^X (conocida como la *métrica de Hausdorff*).

Demostración. Sean A, B, $C \in 2^X$ y

$$E(A,B) + E(B,C) = \{\varepsilon + \delta : \varepsilon \in E(A,B) \ y \ \delta \in E(B,C)\}.$$

(a) Veamos que H está bien definida. Para esto, tenemos que probar que el conjunto $E(A,B)$ es no vacío y está acotado inferiormente. Observemos que $d(x,y) < diám(X)+1$, para cada x, $y \in X$. Así, $A \subset N(diám(X)+1, B)$ y $B \subset N(diám(X)+1, A)$. De manera que $diám(X)+1 \in E(A,B)$. Por tanto, $E(A,B) \neq \emptyset$. Claramente, $E(A,B)$ está acotado inferiormente por el cero.

(b) Ahora, mostremos que para cada A, $B \in 2^X$, tenemos que $H(A,B) \geq 0$. Como $E(A,B)$ está acotado inferiormente por el cero, por definición de ínfimo, se sigue que $H(A,B) \geq 0$, para cada A, $B \in 2^X$.

(c) A continuación, probamos que para cada A, $B \in 2^X$, tenemos que $H(A,B) = H(B,A)$. Por definición de $E(A,B)$, deducimos que $E(A,B) = E(B,A)$, de esto obtenemos (c).

(d) Probaremos para cada A, $B \in 2^X$ que $H(A,B) = 0$ si y sólo si $A = B$. Sean A, $B \in 2^X$. Supongamos que $H(A,B) = 0$. Mostraremos que $A = B$. Para esto, sean $\varepsilon > 0$ y $x \in A$. Como $H(A,B) = 0$, existe $\delta \in E(A,B)$ tal que $\delta < \varepsilon$. Luego, $A \subset N(\delta, B)$ y como $x \in A$, se sigue que $x \in N(\delta, B)$. Entonces existe $y \in B$ tal que $d(x,y) < \delta < \varepsilon$. Así, $y \in B(x,\varepsilon) \cap B$, de donde $B(x,\varepsilon) \cap B \neq \emptyset$ y como ε fue arbitrario, se tiene que $x \in \overline{B}$. Puesto que B es cerrado en X, se sigue que $x \in B$. Por lo tanto, $A \subset B$. Análogamente, se tiene que $B \subset A$. Así, $A = B$.

Ahora supongamos que $A = B$. Luego, para todo $\varepsilon > 0$, tenemos que $\varepsilon \in E(A,B)$, así, $H(A,B) = 0$.

(e) Finalmente veamos que para cada A, $B \in 2^X$, tenemos que $H(A,C) \leq H(A,B) + H(B,C)$.

Para esto, demostremos que $E(A,B) + E(B,C) \subset E(A,C)$. Sea $\beta \in E(A,B) + E(B,C)$, así existen $\varepsilon \in E(A,B)$ y $\delta \in E(B,C)$ tales que $\beta = \varepsilon + \delta$. Luego, $A \subset N(\varepsilon, B)$ y $B \subset N(\delta, C)$. Notemos que $A \subset N(\beta, C)$, dado que si $x \in A$, existe $y \in B$ tal que $d(x,y) < \varepsilon$. Luego, existe $z \in C$ tal que $d(y,z) < \delta$. Así, $d(x,z) \leq d(x,y) + d(y,z) < \varepsilon + \delta = \beta$. Por lo tanto, $A \subset N(\beta, C)$. Similarmente, $C \subset N(\beta, A)$. De esto, deducimos que $\beta \in E(A,C)$. Por lo tanto, $E(A,B) + E(B,C) \subset E(A,C)$. De lo anterior, es inmediato que $H(A,C) \leq H(A,B) + H(B,C)$. □

Para cada continuo X, tenemos que $(2^X, H)$ es un espacio métrico. Como $C(X)$ está contenido en 2^X, observemos que H también es una métrica para $C(X)$. La idea intuitiva de esta métrica es que dos conjuntos están cercanos si ellos casi se empalman uno en el otro. Esta idea geométrica es buena pero tenemos que notar que, por ejemplo, si A es un disco en el plano, se pueden dar conjuntos finitos tan cercanos a A como se quiera, simplemente se toma una cuadrícula muy fina dentro del disco y se toma como conjunto finito al conjunto de los cruces de la cuadrícula.

2.6. Observación. Sea X un continuo no degenerado. Aunque H se puede definir en la familia de todos los subconjuntos no vacíos de X, en este caso H no es una métrica para dicha familia.

2.7. Notación. Sean $A \in 2^X$ y $\varepsilon > 0$. Por $\mathbf{B}(\varepsilon, A)$ entendemos *la bola abierta en* 2^X con centro en A y de radio ε, es decir,

$$\mathbf{B}(\varepsilon, A) = \{B \in 2^X : H(A,B) < \varepsilon\}.$$

2.8. Lema. Sean X un continuo, A, $B \in 2^X$ y $\varepsilon > 0$. Entonces $H(A,B) < \varepsilon$ si y sólo si $A \subset N(\varepsilon, B)$ y $B \subset N(\varepsilon, A)$.

Demostración. Supongamos que $H(A,B) < \varepsilon$. Luego, existe $\delta' \in E(A,B)$ tal que $\delta' < \varepsilon$, $A \subset N(\delta', B)$ y $B \subset N(\delta', A)$. Además, por el Lema 2.1.1.(c), tenemos que $N(\delta', B) \subset N(\varepsilon, B)$ y $N(\delta', A) \subset N(\varepsilon, A)$. Por tanto, $A \subset N(\varepsilon, B)$ y $B \subset N(\varepsilon, A)$.

Recíprocamente, supongamos que $A \subset N(\varepsilon, B)$ y $B \subset N(\varepsilon, A)$. Así, por el Lema 2.1.1.(d), se tiene que $A \subset \bigcup_{\delta<\varepsilon} N(\delta, B)$ para cada $\delta > 0$ tal que $\delta < \varepsilon$. Dado que A es compacto, existen números positivos, $\delta_1, \delta_2, \ldots, \delta_n$, con $n \in \mathbb{N}$, tales que $\delta_i < \varepsilon$ y $A \subset \bigcup_{i=1}^{n} N(\delta_i, B)$. Sea $\alpha = máx\{\delta_1, \delta_2, \ldots, \delta_n\}$. Luego, para cada $i \in \{1, 2, \ldots, n\}$, tenemos que $N(\delta_i, B) \subset N(\alpha, B)$. Así, $\bigcup_{i=1}^{n} N(\delta_i, B) \subset N(\alpha, B)$. De manera que $A \subset N(\alpha, B)$. De manera análoga, como $B \subset N(\varepsilon, A)$, existe $\gamma > 0$ tal que $\gamma < \varepsilon$ y $B \subset N(\gamma, A)$. Sea $\beta = máx\{\alpha, \gamma\}$. Se tiene que $\beta < \varepsilon$, $A \subset N(\beta, B)$ y $B \subset N(\beta, A)$. Así, $\beta \in E(A,B)$. En consecuencia $H(A,B) \leq \beta < \varepsilon$. Por tanto, $H(A,B) < \varepsilon$. □

2.9. Definición. Dado un continuo X, definimos la función $D : 2^X \times 2^X \longrightarrow \mathbb{R}^+ \cup \{0\}$, para cada A, $B \in 2^X$, por $D(A,B) = máx\{sup\{d(a,B) : a \in A\}, sup\{d(A,b) : b \in B\}\}$.

2.10. Teorema. Sea X un continuo. Si A,$B \in 2^X$, entonces $D(A,B) = H(A,B)$.

Demostración. Sean $\varepsilon > 0$ y $r = H(A,B) + \varepsilon$, se sigue que $H(A,B) < r$. Luego, por el Lema 2.8, tenemos que $A \subset N(\varepsilon, B)$ y $B \subset N(\varepsilon, A)$. Así, para cada $a \in A$, existe un $b \in B$ tal que $d(a,b) < r$. De esta forma, para cada $a \in A$, deducimos que $d(a,B) < r$. De modo que $sup\{d(a,B) : a \in A\} < r$. De manera análoga, como $B \subset N(\varepsilon, A)$, se sigue que $sup\{d(b,A) : b \in B\} < r$. Por lo tanto, $D(A,B) \leq r$, es decir, $D(A,B) \leq H(A,B) + \varepsilon$. Como $\varepsilon > 0$ fue arbitrario, inferimos que $D(A,B) \leq H(A,B)$.

Veamos que $H(A,B) \leq D(A,B)$. Sean $\varepsilon > 0$ y $r = D(A,B) + \varepsilon$. Probemos que $A \subset N(r, B)$. Tomemos $a_1 \in A$, así, $d(a_1, B) \leq sup\{d(a,B) : a \in A\}$. Como $sup\{d(a,B) : a \in A\} \leq D(A,B)$, se sigue que $d(a_1, B) \leq D(A,B)$. Así, $d(a_1, B) < r$. De manera que $a_1 \in N(r, B)$. Por lo tanto, $A \subset N(r, B)$.
De manera análoga, se prueba que $B \subset N(r, A)$. Luego, por el Lema 2.8, tenemos que $H(A,B) < r$, es decir, $H(A,B) < D(A,B) + \varepsilon$. Dado que $\varepsilon > 0$ fue arbitrario, se sigue que $H(A,B) \leq D(A,B)$. Por lo tanto, $H(A,B) = D(A,B)$. □

Con el resultado anterior conluimos que D es una métrica para el hiperespacio 2^X, que coincide con la métrica de Hausdorff. De manera que los hiperespacios 2^X y $C(X)$ pueden ser considerados con cualquiera de estas dos métricas, según nos convenga.

2.11. Definición. Dado un subconjunto A de un continuo X. Consideremos las siguientes subcolecciones del hiperespacio 2^X :

$\Gamma(A) = \{B \in 2^X : B \subset A\}$,
$\Lambda(A) = \{B \in 2^X : B \cap A \neq \emptyset\}$ y
$\Phi(A) = \{B \in 2^X : A \subset B\}$.

2.12. Teorema. Sean X un continuo y A un subconjunto de X. Entonces se tiene lo siguiente:

1. Si A es un abierto en X, entonces $\Gamma(A)$ y $\Lambda(A)$ son abiertos en 2^X.
2. Si A es cerrado en X, entonces $\Gamma(A)$, $\Lambda(A)$ y $\Phi(A)$ son cerrados en 2^X.

Demostración.

1. Sea A un subconjunto abierto de X. Probemos que $\Gamma(A)$ es abierto en 2^X. Sea $B \in \Gamma(A)$, luego, $B \in 2^X$ y $B \subset A$. Como A es abierto en X, por el Lema 2.1.2, tenemos que existe $\varepsilon > 0$ tal que $B \subset N(\varepsilon, B) \subset A$. Veamos que $\mathbf{B}(\varepsilon, B) \subset \Gamma(A)$. Sea $C \in \mathbf{B}(\varepsilon, B)$, se sigue que $H(B, C) < \varepsilon$. Por el Lema 2.8, tenemos que $C \subset N(\varepsilon, B)$ y como $N(\varepsilon, B) \subset A$, se sigue que $C \subset A$. Así, $C \in \Gamma(A)$. Luego, $\mathbf{B}(\varepsilon, B) \subset \Gamma(A)$. En resumen, para cada $B \in \Gamma(A)$, existe $\varepsilon > 0$ tal que $\mathbf{B}(\varepsilon, B) \subset \Gamma(A)$, es decir, $\Gamma(A)$ es abierto en 2^X.

   Ahora, demostremos que $\Lambda(A)$ es abierto en 2^X. Sea $B \in \Lambda(A)$, luego $B \in 2^X$ y $B \cap A \neq \emptyset$. Sea $x \in B \cap A$. Como A es abierto en X, existe $\varepsilon > 0$ tal que $B(\varepsilon, x) \subset A$. Probemos que $\mathbf{B}(\varepsilon, B) \subset \Lambda(A)$. Sea $C \in \mathbf{B}(\varepsilon, B)$, luego $H(B, C) < \varepsilon$, por el Lema 2.8, inferimos que $B \subset N(\varepsilon, C)$. Como $x \in B$, existe $y \in C$ tal que $d(x, y) < \varepsilon$, es decir, $y \in B(\varepsilon, x)$. Así, $y \in A$, luego $y \in C \cap A$, de manera que $C \cap A \neq \emptyset$. Por lo tanto, $C \in \Lambda(A)$. Así, $\mathbf{B}(\varepsilon, B) \subset \Lambda(A)$. Esto prueba que $\Lambda(A)$ es abierto en 2^X.
2. Sea A un subconjunto cerrado de X. Para probar que $\Gamma(A)$ es cerrado en 2^X, basta ver que $\overline{\Gamma(A)} \subset \Gamma(A)$. Sea $B \in \overline{\Gamma(A)}$ y supongamos que $B \notin \Gamma(A)$, es decir, $B \not\subset A$. Tomemos $b \in B \setminus A$, como A es compacto en X y $b \notin A$, tenemos que $d(A, b) > 0$. Sea $\varepsilon = d(A, b)$. Dado que $B \in \overline{\Gamma(A)}$, se sigue que $\mathbf{B}(\varepsilon, B) \cap \Gamma(A) \neq \emptyset$. Tomemos $E \in \mathbf{B}(\varepsilon, B) \cap \Gamma(A)$, de manera que $H(B, E) < \varepsilon$ y $E \subset A$. Por el Lema 2.8, implicamos que $B \subset N(\varepsilon, E)$. Como $b \in B$, existe $e \in E$ tal que $d(b, e) < \varepsilon$. Dado que $E \subset A$ y $e \in A$, tenemos que $d(b, A) \leq d(b, e)$. De manera que $d(b, A) < \varepsilon$, lo cual es una contradicción. Así, $B \in \Gamma(A)$. Por lo que $\overline{\Gamma(A)} \subset \Gamma(A)$. Concluimos que $\Gamma(A)$ es cerrado en 2^X.

   Por otro lado, si A es cerrado en X, entonces $X \setminus A$ es abierto en X. Por 1. de este lema, se sigue que $\Gamma(X \setminus A)$ es abierto en 2^X, así, $2^X \setminus \Gamma(X \setminus A)$ es cerrado en 2^X. Notemos que $\Lambda(A) = 2^X \setminus \Gamma(X \setminus A)$ (como $2^X = \Lambda(A) \cup \Gamma(X \setminus A)$ y $\Lambda(A) \cup \Gamma(X \setminus A) = \emptyset$). Por lo tanto, $\Lambda(A)$ es cerrado en 2^X.

   Ahora, veamos que $\Phi(A)$ es cerrado en 2^X. Sea $B \in \overline{\Phi(A)}$ y supongamos que $B \notin \Phi(A)$. Luego, $A \not\subset B$, tomemos $a \in A \setminus B$. Notemos que $d(B, a) > 0$. Sea $\varepsilon = d(B, a)$. Como $B \in \overline{\Phi(A)}$, tenemos que $\Phi(A) \cap \mathbf{B}(\varepsilon, B) \neq \emptyset$. Tomemos $E \in \Phi(A)$ tal que $H(B, E) < \varepsilon$. Por el Lema 2.8, se sigue que $E \subset N(\varepsilon, B)$. Como $a \in A$ y $E \in \Phi(A)$, existe $b \in B$ tal que $d(a, b) < \varepsilon$. Como $d(a, B) \leq d(a, b)$, tenemos que $\varepsilon < \varepsilon$, lo cual no puede ser. Por lo tanto, $B \in \Phi(A)$. Así, $\overline{\Phi(A)} \subset \Phi(A)$. En consecuencia, hemos demostrado que $\Phi(A)$ es cerrado en 2^X. □

3. Topología de Vietoris.

En esta sección veremos que todos los hiperespacios de un continuo, los podemos considerar ya sea con la topología de Vietoris o con la topología inducida por la métrica de Hausdorff, indistintamente.

3.1. Definición. Sean X un continuo, $n \in \mathbb{N}$ y $U_1, U_2, \ldots, U_n$ subconjuntos no vacíos de X.
Diremos que el *vietórico* de $U_1, U_2, \ldots, U_n$, denotado por $\langle U_1, U_2, \ldots, U_n \rangle$, es el

conjunto

$$\{A \in 2^X : A \subset \bigcup_{i=1}^{n} U_i \text{ y } A \cap U_i \neq \emptyset, \text{ para cada } i \in \{1, 2, \ldots, n\}\}.$$

3.2. Lema. Sean X un continuo, $n \in \mathbb{N}$ y $U_1, U_2, \ldots, U_n$ subconjuntos no vacíos de X. Las siguientes afirmaciones se cumplen

1. $\langle U_1, U_2, \ldots, U_n\rangle = \Gamma(\bigcup_{i=1}^{n} U_i) \cap [\bigcap_{i=1}^{n} \Lambda(U_i)]$,
2. Para cada $A \subset X$, tenemos que $\Gamma(A) = \langle A\rangle$,
3. Para cada $A \subset X$, tenemos que $\Lambda(A) = \langle X, A\rangle$.

Demostración. Para ver que se cumple (1) notemos que $\langle U_1, U_2, \ldots, U_n\rangle = \{A \in 2^X : A \subset \cup_{i=1}^{n} U_i\} \cap \{A \in 2^X : A \cap U_i \neq \emptyset, \text{ para cada } i \in \{1, 2, \ldots, n\}\} = \Gamma(\cup_{i=1}^{n} U_i) \cap [\cap_{i=1}^{n} \Lambda(U_i)]$. Por lo tanto, $\langle U_1, U_2, \ldots, U_n\rangle = \Gamma(\cup_{i=1}^{n} U_i) \cap [\cap_{i=1}^{n} \Lambda(U_i)]$. El resto de la demostración de este lema es similar. □

3.3. Teorema. Sean $m, n \in \mathbb{N}$, $U_1, U_2, \ldots, U_n$ y $V_1, V_2, \ldots, V_m$ subconjuntos de un continuo X. Si $U = \bigcup_{i=1}^{n} U_i$ y $V = \bigcup_{i=1}^{m} V_i$, entonces

$$\langle U_1, U_2, \ldots, U_n\rangle \cap \langle V_1, V_2, \ldots, V_m\rangle =$$

$$\langle V \cap U_1, V \cap U_2, \ldots, V \cap U_n, U \cap V_1, U \cap V_2, \ldots, U \cap V_m\rangle.$$

Demostración. Sea

$$A \in \langle U_1, U_2, \ldots, U_n\rangle \cap \langle V_1, V_2, \ldots, V_m\rangle$$

$$= \Gamma(U) \cap \Big[\bigcap_{i=1}^{n} \Lambda(U_i)\Big] \cap \Gamma(V) \cap \Big[\bigcap_{i=1}^{m} \Lambda(V_i)\Big].$$

Así,

$$A \subset U \cap V = (U \cap V) \cup (V \cap U)$$

$$= \Big[U \cap \big(\bigcup_{i=1}^{m} V_i\big)\Big] \cup \Big[V \cap \big(\bigcup_{i=1}^{n} U_i\big)\Big]$$

$$= \Big[\bigcup_{i=1}^{m} \big(U \cap V_i\big)\Big] \cup \Big[\bigcup_{i=1}^{n} \big(V \cap U_i\big)\Big].$$

Por otro lado, como $A \cap U_i \neq \emptyset$, para cada $i \in \{1, 2, \ldots, n\}$ y $A \subset V$, se tiene que $A \cap (V \cap U_i) \neq \emptyset$, para cada $i \in \{1, 2, \ldots, n\}$. Similarmente, $A \cap (U \cap V_i) \neq \emptyset$, para cada $i \in \{1, 2, \ldots, m\}$. De manera que

$$A \in \langle V \cap U_1, V \cap U_2, \ldots, V \cap U_n, U \cap V_1, U \cap V_2, \ldots, U \cap V_m\rangle.$$

Para probar la otra contención, sea

$$A \in \langle V \cap U_1, V \cap U_2, \ldots, V \cap U_n, U \cap V_1, U \cap V_2, \ldots, U \cap V_m\rangle.$$

Entonces, $A \subset U \cap V$. Es decir, $A \subset U$ y $A \subset V$. Así, $A \in \Gamma(U)$ y $A \in \Gamma(V)$. Por otra parte, como $A \cap (U \cap V_i) = A \cap V_i \neq \emptyset$, para cada $i \in \{1, 2, \ldots, m\}$, se tiene que $A \in \cap_{i=1}^{m}(V_i)$. Similarmente, $A \in \cap_{i=1}^{n}(U_i)$. Por tanto,

$$A \in \langle U_1, U_2, \ldots, U_n\rangle \cap \langle V_1, V_2, \ldots, V_m\rangle.$$

□

El siguiente resultado dota de una topología al hiperespacio 2^X de un continuo X dado.

3.4. Teorema. Sean X un continuo y $\mathcal{B} = \{\langle U_1, U_2, \ldots, U_n\rangle : U_1, U_2, \ldots, U_n \textit{ son abiertos en } X \textit{ y } n \in \mathbb{N}\}$. Entonces $\mathcal{B}$ es una base para una topología del hiperespacio 2^X (la topología generada por $\mathcal{B}$), denotada por τ_V y conocida como la *Topología de Vietoris*.

Demostración. . Primero veamos que $2^X = \bigcup \mathcal{B}$. Notemos que $\langle X\rangle = \{A \in 2^X : A \subset X\} = 2^X$. Así, $2^X \in \mathcal{B}$. De manera que $2^X \subset \bigcup \mathcal{B}$, luego, $2^X = \bigcup \mathcal{B}$.
La demostración de la segunda condición, es decir, que para cada $\mathcal{U}$, $\mathcal{V} \in \mathcal{B}$ con $A \in \mathcal{U} \cap \mathcal{V}$, existe $\mathcal{W} \in \mathcal{B}$ tal que $A \in \mathcal{W} \subset \mathcal{U} \cap \mathcal{V}$, se tiene del Lema 3.3. Por lo tanto, $\mathcal{B}$ es una base para la topología de Vietoris. □

3.5. Teorema. Sea X un continuo. El conjunto $\mathcal{S} = \{\Gamma(U) : U \textit{ es abierto en } X\} \cup \{\Lambda(U) : U \textit{ es abierto en } X\}$ es una subbase para la Topología de Vietoris.

Demostración. Pongamos $\mathcal{S}' = \{\bigcap \mathcal{W} : \mathcal{W} \text{ es un subconjunto finito de } \mathcal{S}\}$. Para ver que $\mathcal{S}$ es subbase para la topología de Vietoris, basta probar que $\mathcal{S}' = \mathcal{B}$.

Sea $\mathcal{U} \in \mathcal{B}$. Luego, sean $U_1, U_2, \ldots, U_n$ conjuntos abiertos en X tales que $\mathcal{U} = \langle U_1, U_2, \ldots, U_n\rangle$ y pongamos $U = \bigcup_{i=1}^n U_i$. Notemos que por el Lema 3.2, $\mathcal{U} = \langle U_1, U_2, \ldots, U_n\rangle = \Gamma(U) \cap \Lambda(U_1), \ldots, \Lambda(U_n)$. Es decir, $\mathcal{U}$ es una intersección finita de elementos de $\mathcal{S}$. Así, $\mathcal{U} \in \mathcal{S}'$. De manera que $\mathcal{B} \subset \mathcal{S}'$.

Por otra parte, observemos que $\mathcal{S} \subset \mathcal{B}$, pues si $\mathcal{V} \in \mathcal{S}$, entonces $\mathcal{V} = \Gamma(U)$ o $\mathcal{V} = \Lambda(U)$, para algún conjunto U abierto de X, es decir, $\mathcal{V} = \langle U\rangle$ o $\mathcal{V} = \langle X, U\rangle$, de cualquier forma $\mathcal{V} \in \mathcal{B}$. Además, por el Lema 3.3, sabemos que $\mathcal{B}$ es cerrado bajo intersecciones finitas, de manera que $\mathcal{S}' \subset \mathcal{B}$. Por lo tanto, $\mathcal{S}' = \mathcal{B}$. Lo que demuestra que $\mathcal{S}$ es subbase para la topología de Vietoris. □

3.6. Teorema. Sea X un continuo. La Topología de Vietoris τ_V y la topología inducida por la métrica de Hausdorff τ_H en 2^X son iguales.

Demostración. Sean $\mathcal{U} \in \tau_V$ y $A \in \mathcal{U}$. Por el Teorema 3.4, tenemos que $\mathcal{B}$ es una base de τ_V. Así, existen $n \in \mathbb{N}$ y conjuntos abiertos $U_1, U_2, \ldots, U_n$ de X tales que $A \subset \langle U_1, U_2, \ldots, U_n\rangle \subset \mathcal{U}$. Luego, $\bigcup_{i=1}^n U_i$ es abierto en X. Por el Teorema 2.12, se sigue que $\Gamma(\bigcup_{i=1}^n U_i) \in \tau_H$ y $\Lambda(U_i) \in \tau_H$, para cada $i \in \{1, 2, \ldots, n\}$. Así,

$$\Gamma(\bigcup_{i=1}^{n} U_i) \cap [\bigcap_{i=1}^{n} \Lambda(U_i)] \in \tau_H.$$

Por el Lema 3.2.1, inferimos que:

$$\langle U_1, U_2, \ldots, U_n\rangle = \Gamma(\bigcup_{i=1}^{n} U_i) \cap [\bigcap_{i=1}^{n} \Lambda(U_i)].$$

Luego, $\mathcal{U} \in \tau_H$. De manera que $\tau_V \subset \tau_H$.

Ahora, sean $\mathcal{V} \in \tau_H$ y $A \in \mathcal{V}$. Probemos que existe $\mathcal{W} \in \mathcal{B}$ tal que $A \in \mathcal{W} \subset \mathcal{V}$. Recordemos que una base para τ_H está dada por $\gamma_H = \{\mathbf{B}(\delta, C) : C \in 2^X \textit{ y } \delta > 0\}$. De manera que existen $F \in 2^X$ y $\varepsilon > 0$ tales que $A \in \mathbf{B}(\varepsilon, F) \subset \mathcal{V}$.

Por otro lado, observemos que la colección $\{B(\frac{\varepsilon}{2}, b) : b \in F\}$ es una cubierta abierta para F. Como F es compacto, existen $n \in \mathbb{N}$ y $\{b_1, b_2, \ldots, b_n\} \subset F$ tales que $F \subset \bigcup_{i=1}^n B(\frac{\varepsilon}{2}, b_i)$.

Sea $U_i = B(\frac{\varepsilon}{2}, b_i)$, para cada $i \in \{1, 2, \ldots, n\}$. Consideremos

$$\mathcal{W} = \langle U_1, U_2, \ldots, U_n\rangle = \Gamma(\bigcup_{i=1}^{n} U_i) \cap [\bigcap_{i=1}^{n} \Lambda(U_i)]$$

(por el Lema 3.2.1). Notemos que $\mathcal{W} \in \mathcal{B}$.

Ahora, probemos que $\mathcal{W} \subset \mathbf{B}(\varepsilon, F)$. Sea $D \in \mathcal{W}$, luego,

$$D \in \Gamma(\bigcup_{i=1}^{n} U_i) \cap [\bigcap_{i=1}^{n} \Lambda(U_i)].$$

Así, $D \subset \bigcup_{i=1}^{n} U_i$ y $D \cap U_i \neq \emptyset$, para cada $i \in \{1, 2, \ldots, n\}$. Afirmamos que:

$$D \subset N(\varepsilon, F). \tag{3.1}$$

En efecto, si $e \in D$, entonces existe $j \in \{1, 2, \ldots, n\}$ tal que $e \in U_j = B(\frac{\varepsilon}{2}, b_j)$. Así, $d(e, b_j) < \frac{\varepsilon}{2}$, además $b_j \in F$. En resumen, para cada $e \in D$, existe $b_j \in F$ tal que $d(e, b_j) < \varepsilon$. De manera que $D \subset N(\varepsilon, F)$.

Veamos que:

$$F \subset N(\varepsilon, D). \tag{3.2}$$

Si $b \in F$, entonces existe $k \in \{1, 2, \ldots, n\}$ tal que $b \in U_k = B(\frac{\varepsilon}{2}, b_k)$. Así, $d(b, b_k) < \frac{\varepsilon}{2}$. Dado que $D \cap U_i \neq \emptyset$, para cada $i \in \{1, 2, \ldots, n\}$, inferimos que $D \cap U_k \neq \emptyset$, es decir, $D \cap B(\frac{\varepsilon}{2}, b_k) \neq \emptyset$, se sigue que existe $z \in D \cap B(\frac{\varepsilon}{2}, b_k)$. Por la desigualdad del triángulo, tenemos que $d(b, z) \leq d(b, b_k) + d(b_k, z) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, es decir, $d(b, z) < \varepsilon$. Por lo tanto, para cada $b \in F$, existe $z \in D$ tal que $d(b, z) < \varepsilon$. En consecuencia, $F \subset B(\varepsilon, D)$.

De (3.1), (3.2) y por el Lema 2.8, implicamos que $H(F, D) < \varepsilon$. Así, $D \in \mathbf{B}(\varepsilon, F)$. Por lo tanto, $\mathcal{W} \subset \mathbf{B}(\varepsilon, F)$. Dado que $\mathbf{B}(\varepsilon, F) \subset \mathcal{V}$, deducimos que $\mathcal{W} \subset \mathcal{V}$.

En resumen, para cada $\mathcal{V} \in \tau_H$ tal que $A \in \mathcal{V}$, existe $\mathcal{W} \in \mathcal{B}$ tal que $A \subset \mathcal{W} \subset \mathcal{V}$. Esto demuestra que $\tau_H \subset \tau_V$. Por tanto, $\tau_H = \tau_V$. □

Referencias

[1] C. O. Christenson and W. L. Voxman, *Aspects of Topology*, Monographs and Textbooks in Pure and applied Math., Vol. 39, Marcel Dekker, Inc., New York, 1977.

[2] A. Illanes, *Hiperespacios de continuos*, Aportaciones Matemáticas, Serie Textos N. 28, Sociedad Matemática Mexicana, ISBN: 968-36-3594-6, 2004.

[3] S. B. Nadler, Jr., *Continuum Theory. An introduction.* Monographs and Texbooks in Pure and Applied Mathematics, Vol. 158, Marcel Dekker, New York, ISBN:0-8247-8659-9, 1992.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

cosvi_07@hotmail.com, dherrera@fcfm.buap.mx, fmacias@fcfm.buap.mx

Capítulo 27

DENDRITAS LOCALES

LUIS ALBERTO GUERRERO MÉNDEZ
DAVID HERRERA CARRASCO
FERNANDO MACÍAS ROMERO
FCFM-BUAP

Resumen. Un *continuo* es un espacio métrico no vacío, compacto y conexo. Un *subcontinuo* es un continuo que es subespacio de un espacio topológico. Una *dendrita* es un continuo localmente conexo que no contiene curvas cerradas simples. Una *dendrita local* es un continuo tal que cada punto tiene una vecindad que es una dendrita. En este trabajo retomamos varias propiedades de las dendritas locales.

1. Introducción

Este trabajo pertenece a la rama de la Topología conocida como Teoría de los Continuos, la cual trata del estudio de las propiedades topológicas de los espacios métricos no vacíos, compactos y conexos. De hecho a un espacio topológico con estas características se le llama **continuo.**

Casi todos los resultados que se mencionan en este artículo están demostrados en las referencias de la bibliografía señalada al final de este artículo, sin embargo, nuestra aportación consiste en la presentación y la demostración detallada de los resultados que se prueban.

Todos los conceptos no definidos en este trabajo los consideramos igual que en la referencia [4].

En este trabajo demostramos, vea el Teorema 3.7, que toda dendrita local es un continuo regular, basándonos en el hecho de que toda dendrita es un continuo regular. Exhibimos un ejemplo de un continuo regular que no es dendrita local (vea el Ejemplo 3.8), por lo que la colección de las dendritas locales está contenida propiamente en la colección de los continuos regulares.

Además de esto, entre los resultados sobre dendritas locales que revisamos en este trabajo están los siguientes:

(a) Si X es una dendrita local, entonces existen dendritas $D_1, \ldots, D_n$ tales que $X = \bigcup_{i=1}^{n} \operatorname{Int}_X(D_i)$ (vea el Teorema 3.10).
(b) Si X es una dendrita local, entonces existe un número $\varepsilon > 0$ tal que todo subcontinuo C de diámetro menor que ε es una dendrita (vea el Teorema 3.14).
(c) Sean X un continuo localmente conexo y $\varepsilon = \text{ínf}\,\{\text{diám}(C)\colon\ C \text{ es una curva cerrada simple contenida en } X\}$. Si $\varepsilon > 0$, entonces X es una dendrita local (vea el Teorema 3.15).

(d) Sea X un continuo. Entonces las siguientes condiciones son equivalentes (vea el Teorema 3.17):
  (1) X es una dendrita local.
  (2) X es un continuo localmente conexo que contiene a lo más un número finito de curvas cerradas simples.
  (3) X es la unión de una colección finita de dendritas $\{D_1, \ldots, D_n\}$ tal que para $i, j \in \{1, \ldots, n\}$ con $i \neq j$, tenemos que $|D_i \cap D_j| < \infty$.

2. Preliminares

En la presente sección enunciaremos la notación, las definiciones y los resultados esenciales que usaremos a lo largo de este trabajo. Para los resultados que presentamos sin demostración, damos en cada caso una referencia adecuada.

El símbolo $\mathbb{N}$ representará al conjunto de los números naturales y $\mathbb{R}$ al de los números reales. Dado $n \in \mathbb{N}$ y $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, supondremos que la norma de $(x_1, x_2, \ldots, x_n)$ es $\| (x_1, x_2, \ldots, x_n) \| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ y que $\mathbb{R}^n$ posee la topología usual, inducida por la norma. Cuando escribamos sobre un subconjunto de $\mathbb{R}^n$, lo consideraremos como un espacio topológico y supondremos que su topología es la usual. La cardinalidad de un conjunto Z la representaremos por $|Z|$.

Sean X un espacio topológico y $p \in X$, un subconjunto A de X es una **vecindad** de p si existe un abierto U en X tal que $p \in U \subset A$.

Sean X un espacio topológico y $A \subset X$, denotaremos al interior, la cerradura y la frontera de A en X como $\mathrm{Int}_X(A)$, $\mathrm{Cl}_X(A)$ y $\mathrm{Fr}_X(A)$, respectivamente.

Sean X un espacio métrico con métrica d; $p \in X$ y $\varepsilon > 0$, la **bola abierta en X con centro en p y radio ε**, que denotaremos por $B_\varepsilon(p)$, es el conjunto $B_\varepsilon(p) = \{x \in X : d(x, p) < \varepsilon\}$.

Veamos las siguientes definiciones básicas para el desarrollo de nuestro trabajo.

2.1. Definición. Un **continuo** es un espacio métrico no vacío, compacto y conexo. Un **subcontinuo** es un continuo que es subespacio de un espacio topológico.

Sea $S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Una **curva cerrada simple** es un espacio que es homeomorfo a S^1.

2.2. Definición. Un espacio topológico X es **localmente conexo en** $p \in X$ si para cada vencindad V de p, existe una vecindad U de p tal que U es conexo, U es abierto en X y $U \subset V$. Si X es localmente conexo en cada uno de sus puntos decimos que X es **localmente conexo.**

La noción de conexidad local está dada en términos de vecindades. También puede darse en términos de abiertos, de hecho en [1, Teorema 1.40, pág. 15] se prueba una caracterización de esta propiedad.

2.3. Definición. Una **dendrita** es un continuo localmente conexo que no contiene curvas cerradas simples.

Recordemos el concepto de base local de un punto.

2.4. Definición. Si X es un espacio topológico y $p \in X$, entonces una **base local de p** en X es una colección $\mathcal{B}_p$ de abiertos en X tal que

(1) Si U es abierto en X con $p \in U$, entonces existe $V \in \mathcal{B}_p$ tal que $p \in V \subset U$.
(2) $p \in \bigcap_{V \in \mathcal{B}_p} V$.

2.5. Ejemplo. Para todo $p \in \mathbb{R}^n$, tenemos que $\mathcal{B}_p = \left\{ B_{\frac{1}{n}}(p) \colon n \in \mathbb{N} \right\}$ es una base local de p en $\mathbb{R}^n$.

El siguiente resultado nos proporciona una base local de un punto para un subespacio, a partir de una base local del mismo punto en el espacio topológico.

2.6. Teorema. Sean X un espacio topológico y $p \in X$. Si $\mathcal{B}_p$ es una base local de p en X y Y es un subconjunto de X tal que $p \in Y$, entonces $\{B \cap Y \colon B \in \mathcal{B}_p\}$ es una base local de p en Y.

Demostración. Sean $\mathcal{B}_p$ una base local de p en X y Y un subconjunto de X tal que $p \in Y$. Sea $\mathcal{B}_p^* = \{B \cap Y \colon B \in \mathcal{B}_p\}$.

Como cada $B \in \mathcal{B}_p$ es abierto en X, tenemos que $\mathcal{B}_p^*$ es una colección de abiertos en Y.

Sea U un abierto en Y con $p \in U$. Veamos que existe $V \in \mathcal{B}_p^*$ tal que $p \in V \subset U$. Como U es abierto en Y existe W abierto en X tal que $U = W \cap Y$. Luego, $p \in W \cap Y$. Como $p \in W$ y W es abierto en X, existe $B \in \mathcal{B}_p$ tal que $p \in B \subset W$. Así, $p \in B \cap Y$. Notemos que $B \cap Y \subset W \cap Y$. Luego, $p \in B \cap Y \subset U$.

Ahora, veamos que $p \in \bigcap_{V \in \mathcal{B}_p^*} V$.

Como $p \in \bigcap_{B \in \mathcal{B}_p} B$ y $p \in Y$, tenemos que

$$p \in \left(\bigcap_{B \in \mathcal{B}_p} B \right) \cap Y = \bigcap_{B \in \mathcal{B}_p} (B \cap Y) = \bigcap_{V \in \mathcal{B}_p^*} V.$$

Así, $\mathcal{B}_p^*$ es una base local de p en Y. □

2.7. Ejemplo. Sean $p = (1,\ sen\ (1)) \in \mathbb{R}^2$ y

$$Y = \left\{ \left(x, sen\left(\frac{1}{x}\right) \right) \in \mathbb{R}^2 \colon 0 < x \leq 1 \right\}.$$

La colección $\mathcal{B}_p = \left\{ B_{\frac{1}{n}}(p) \colon n \in \mathbb{N} \right\}$ es una base local de p en $\mathbb{R}^2$. Notemos que $p \in Y$. Luego, por el Teorema 2.6, tenemos que

$$\mathcal{B}_p^* = \left\{ B_{\frac{1}{n}}(p) \cap Y \colon n \in \mathbb{N} \right\}$$

es una base local de p en Y.

Estamos listos para anotar el concepto de continuo regular.

2.8. Definición. Si X es un continuo y $p \in X$, entonces X es un **continuo regular en** p si existe una base local de p en X, $\mathcal{B}_p$, tal que la frontera de todo elemento de $\mathcal{B}_p$ es de cardinalidad finita. Un continuo X es un **continuo regular** si X es un continuo regular en cada uno de sus puntos.

2.9. Ejemplo. Consideremos el continuo $X = [0,1]$. Sean $p \in X$ y $\mathcal{B}_p = \{(p - \varepsilon, p + \varepsilon) \cap [0,1] \colon \varepsilon > 0\}$. Notemos que $\mathcal{B}_p$ es una base local de p en X tal que para todo $B \in \mathcal{B}_p$, tenemos que $|\mathrm{Fr}_X(B)| < \infty$. Luego, X es un continuo regular.

2.10. Ejemplo. Sean $Y = [0,1] \times [0,1]$ y $p \in Y$. Notemos que cualquier abierto U en Y con $p \in U$ tiene frontera de cardinalidad no finita. Luego, para cualquier base local $\mathcal{B}_p$ de p en Y, tenemos que todo $B \in \mathcal{B}_p$ tiene frontera de cardinalidad no finita. Así, Y no es un continuo regular.

En particular, las dendritas tienen esta propiedad, como lo dice el siguiente resultado que se prueba en [4, 10.20, pág. 173].

2.11. Teorema. Toda dendrita es un continuo regular.

En seguida un teorema relativo a fronteras que nos ayudará a probar una propiedad de los continuos regulares.

2.12. Teorema. Si A y B son subconjuntos de un espacio topológico Z, entonces $\mathrm{Fr}_B(A \cap B) \subset \mathrm{Fr}_Z(A) \cap B$.

Demostración. Supongamos que Z, A y B son como en la hipótesis del teorema. Observemos que

$$\begin{aligned} \mathrm{Fr}_B(A \cap B) &= \mathrm{Cl}_B(A \cap B) \cap \mathrm{Cl}_B(B - (A \cap B)) \\ &= \mathrm{Cl}_Z(A \cap B) \cap \mathrm{Cl}_Z(B - A) \cap B \subset \mathrm{Cl}_Z(A) \cap \mathrm{Cl}_Z(Z - A) \cap B \\ &= \mathrm{Fr}_Z(A) \cap B. \end{aligned}$$

□

La propiedad de ser un continuo regular es una propiedad hereditaria.

2.13. Teorema. Todo subcontinuo de un continuo regular es un continuo regular.

Demostración. Supongamos que X es un continuo regular y que Y es un subcontinuo de X. Veamos que Y es un continuo regular. Para esto sea $y \in Y$. Como X es un continuo regular existe una base local $\mathcal{B}_y$ de y en X tal que todo elemento de $\mathcal{B}_y$ tiene frontera de cardinalidad finita. Por el Teorema 2.6, tenemos que $\mathcal{B}_y^* = \{B \cap Y : B \in \mathcal{B}_y\}$ es una base local de y en Y. Además, por el Teorema 2.12, para todo $B \in \mathcal{B}_y$, tenemos que

$$\mathrm{Fr}_Y(B \cap Y) \subset \mathrm{Fr}_X(B) \cap Y,$$

y como $\mathrm{Fr}_X(B) \cap Y \subset \mathrm{Fr}_X(B)$, obtenemos que $\mathrm{Fr}_Y(B \cap Y) \subset \mathrm{Fr}_X(B)$.

Luego, para todo $B \in \mathcal{B}_y$, tenemos que

$$|\mathrm{Fr}_Y(B \cap Y)| \leq |\mathrm{Fr}_X(B)| < \infty.$$

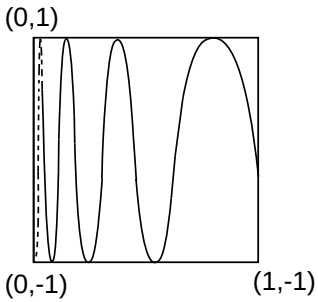
Así, Y es un continuo regular. □

Ahora, veremos un concepto que nos será de utilidad en resultados posteriores.

2.14. Definición. Un continuo X es **hereditariamente localmente conexo** si todo subcontinuo de X es localmente conexo.

Obviamente todo continuo hereditariamente localmente conexo es localmente conexo, sin embargo no todo continuo localmente conexo es hereditariamente localmente conexo, como se muestra a continuación.

2.15. Ejemplo. Sean $W = \left\{\left(x, sen\left(\frac{1}{x}\right)\right) \in \mathbb{R}^2 : 0 < x \leq 1\right\}$ y $J = \{0\} \times [-1,1]$. Tenemos que $\mathrm{Cl}_{\mathbb{R}^2}(W) = W \cup J$ es un continuo, conocido como el **continuo *sen*** $\left(\frac{1}{x}\right)$, el cual no es localmente conexo. Consideremos $Y = [0,1] \times [-1,1]$. Notemos que Y es un continuo localmente conexo y el continuo $sen\left(\frac{1}{x}\right)$ está contenido en Y. Luego, Y no es hereditariamente localmente conexo.



$\mathrm{Cl}_{\mathbb{R}^2}(W) \subset Y$

El siguiente resultado se prueba en [4, 10.16, pág. 171].

2.16. Teorema. Todo continuo regular es hereditariamente localmente conexo.

El recíproco del Teorema 2.16 no es en general cierto, como se ve a continuación.

2.17. Ejemplo. [2, Observación (i) del Teorema 2, págs. 283 y 284] Sea

$$A = \{(x, 0) \in \mathbb{R}^2 : 0 \leq x \leq 1\};$$

para cada $n \in \mathbb{N}$ y cada $k \in \{1, 2, \ldots, 2^{n-1}\}$, sea

$$B_{n,k} = \left\{(x, y) \in \mathbb{R}^2 \colon \left(x - \frac{2k-1}{2^n}\right)^2 + y^2 = 4^{-n},\ y \geq 0\right\};$$

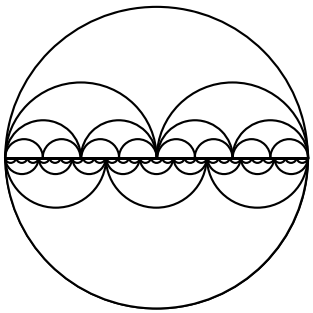
y para cada $n \in \mathbb{N} \cup \{0\}$ y cada $k \in \{1, 2, \ldots, 3^n\}$, sea

$$C_{n,k} = \left\{(x, y) \in \mathbb{R}^2 \colon \left(x - \frac{2k-1}{2 \cdot 3^n}\right)^2 + y^2 = \frac{1}{4} \cdot 9^{-n},\ y \leq 0\right\}.$$

Si

$$X = A \cup \left[\bigcup_{n=1}^{\infty} \left(\bigcup_{k=1}^{2^{n-1}} B_{n,k}\right)\right] \cup \left[\bigcup_{n=0}^{\infty} \left(\bigcup_{k=1}^{3^n} C_{n,k}\right)\right],$$

entonces X es un continuo hereditariamente localmente conexo y no es un continuo regular. Una aproximación del continuo X es la siguiente:



2.18. Teorema. Toda dendrita es hereditariamente localmente conexo.

Demostración. Supongamos que X es una dendrita. Por el Teorema 2.11, tenemos que X es un continuo regular. Por el Teorema 2.16, concluimos que X es hereditariamente localmente conexo. □

La propiedad de ser una dendrita es una propiedad hereditaria.

2.19. Corolario. Todo subcontinuo de una dendrita es una dendrita.

Demostración. Supongamos que X es una dendrita y Y es un subcontinuo de X. Por el Teorema 2.18, tenemos que X es hereditariamente localmente conexo, luego, Y es localmente conexo. Como $Y \subset X$ y X no contiene curvas cerradas simples, entonces Y no contiene curvas cerradas simples, concluimos que Y es una dendrita. □

El teorema que sigue, nos da condiciones suficientes para que la cerradura (frontera) de un conjunto A en un espacio topológico Z sea igual a la cerradura (frontera) del mismo conjunto A en algún subespacio B de Z.

2.20. Teorema. Sean A y B subconjuntos de un espacio topológico Z tales que $A \subset B$. Las siguientes condiciones son verdaderas.

(1) Si B es cerrado en Z, entonces $\mathrm{Cl}_B(A) = \mathrm{Cl}_Z(A)$.
(2) Si A es abierto en Z y B es cerrado en Z, entonces $\mathrm{Fr}_B(A) = \mathrm{Fr}_Z(A)$.

Demostración. Supongamos que Z, A y B son como en las hipótesis del teorema.

Veamos que se cumple (1). Supongamos que B es cerrado en Z. Como $A \subset B$ y B es cerrado en Z, tenemos que $\mathrm{Cl}_Z(A) \subset \mathrm{Cl}_Z(B) = B$. Luego, $\mathrm{Cl}_Z(A) \cap B = \mathrm{Cl}_Z(A)$. Notemos que $\mathrm{Cl}_B(A) = \mathrm{Cl}_Z(A) \cap B$. Así, $\mathrm{Cl}_B(A) = \mathrm{Cl}_Z(A)$.

Probemos (2). Supongamos que A es abierto en Z y B es cerrado en Z. Como A es abierto en Z y $A \subset B$, se sigue que A es abierto en B. Recordemos que una propiedad básica de la frontera es que $\mathrm{Fr}_B(A) = \mathrm{Cl}_B(A) - \mathrm{Int}_B(A)$. Como A es abierto en B, tenemos que $A = \mathrm{Int}_B(A)$.

Luego,

$$\mathrm{Fr}_B(A) = \mathrm{Cl}_B(A) - A.$$

Ahora, como B es cerrado en Z, por (1), tenemos que $\mathrm{Cl}_B(A) = \mathrm{Cl}_Z(A)$. Como A es abierto en Z, tenemos que $A = \mathrm{Int}_Z(A)$. Luego, $\mathrm{Cl}_B(A) - A = \mathrm{Cl}_Z(A) - \mathrm{Int}_Z(A)$. Por lo tanto, $\mathrm{Fr}_B(A) = \mathrm{Cl}_Z(A) - \mathrm{Int}_Z(A) = \mathrm{Fr}_Z(A)$. □

Concluimos esta sección con otro resultado, demostrado en [1, Teorema 2.25, pág. 41], que nos será de utilidad.

2.21. Teorema. Si X es un continuo localmente conexo, $p \in X$ y $\varepsilon > 0$, entonces existe $C \subset X$ tal que C es un continuo localmente conexo, C es una vecindad de p y $\mathrm{diám}(C) < \varepsilon$.

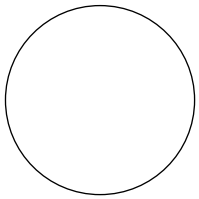
3. Dendritas locales

En la presente sección definiremos lo que es una dendrita local, daremos ejemplos y probaremos algunas propiedades de las dendritas locales, así como una caracterización interesante de éstas.

3.1. Definición. Una **dendrita local** es un continuo tal que cada punto tiene una vecindad que es una dendrita.

3.2. Ejemplo. Toda dendrita es una dendrita local.

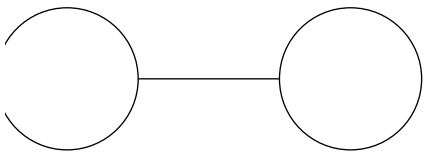
3.3. Ejemplo. La circunferencia unitaria S^1 no es una dendrita y sí es una dendrita local.



Circunferencia unitaria S^1.

3.4. Definición. Una **gráfica finita** es un continuo que puede ser escrito como la unión finita de arcos tales que cada dos de ellos se intersectan en un conjunto finito.

3.5. Ejemplo. Consideremos en $\mathbb{R}^2$ los siguientes puntos $p_1 = (-2, 0)$ y $p_2 = (2, 0)$. Sean C_1 y C_2 las circunferencias de radio 1 y centradas en p_1 y p_2, respectivamente. Sea $P = C_1 \cup ([-1, 1] \times \{0\}) \cup C_2$. El espacio P, **la pesa,** es un continuo tal que no es una dendrita y sí es una gráfica finita.



La pesa.

3.6. Ejemplo. Toda gráfica finita es una dendrita local.

Veamos una primera propiedad de las dendritas locales.

3.7. Teorema. Toda dendrita local es un continuo regular.

Demostración. Supongamos que X es una dendrita local y que $p \in X$. Existe una dendrita Y que es una vecindad de p, luego, existe U abierto en X tal que $p \in U \subset Y$. Notemos que $U = \operatorname{Int}_X(U) \subset \operatorname{Int}_X(Y)$. Así, $p \in \operatorname{Int}_X(Y)$.

Por el Teorema 2.11, existe una base local de p en Y, $\mathcal{B}_p$, tal que para todo $B \in \mathcal{B}_p$, tenemos que $|\operatorname{Fr}_Y(B)| < \infty$.

Sea $\mathcal{B}_p^* = \{B \in \mathcal{B}_p \colon B \subset \operatorname{Int}_X(Y)\}$. Afirmamos que $\mathcal{B}_p^*$ es una base local de p en X.

En primer lugar veamos que $\mathcal{B}_p^* \neq \emptyset$. Notemos que $\operatorname{Int}_X(Y)$ es abierto en Y. Luego, como $p \in \operatorname{Int}_X(Y)$, existe $B \in \mathcal{B}_p$ tal que $p \in B \subset \operatorname{Int}_X(Y)$. Así, $B \in \mathcal{B}_p^*$. Por lo tanto, $\mathcal{B}_p^* \neq \emptyset$.

Veamos que todo $B \in \mathcal{B}_p^*$ es abierto en X. Para esto supongamos que $B \in \mathcal{B}_p^*$. Como B es abierto Y, existe V abierto en X tal que

$$B = V \cap Y \subset \operatorname{Int}_X(Y).$$

Luego,

$$B = V \cap Y = (V \cap Y) \cap \operatorname{Int}_X(Y) = V \cap (Y \cap \operatorname{Int}_X(Y))$$

$$= V \cap \operatorname{Int}_X(Y) = \operatorname{Int}_X(V) \cap \operatorname{Int}_X(Y) = \operatorname{Int}_X(V \cap Y) = \operatorname{Int}_X(B).$$

Así, B es abierto en X.

Ahora, supongamos que W es abierto en X tal que $p \in W$. Veamos que existe $B \in \mathcal{B}_p^*$ tal que $p \in B \subset W$. Notemos que $p \in U \cap W \subset U \subset Y$. Como $W \cap U$ es abierto en X y $(W \cap U) \cap Y = W \cap U$, tenemos que $W \cap U$ es abierto en Y. Por lo tanto, existe $B \in \mathcal{B}_p$ tal que $p \in B \subset W \cap U \subset U \subset \operatorname{Int}_X(Y)$. Así, $B \in \mathcal{B}_p^*$. Además, $p \in B \subset W \cap U \subset W$.

Como $p \in \bigcap_{B \in \mathcal{B}_p} B$ y $\mathcal{B}_p^* \subset \mathcal{B}_p$, tenemos que $p \in \bigcap_{B \in \mathcal{B}_p^*} B$. Así, tenemos que $\mathcal{B}_p^*$ es una base local de p en X.

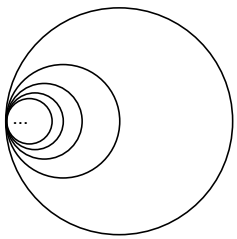
Veamos que todo elemento de $\mathcal{B}_p^*$ tiene frontera de cardinalidad finita. Supongamos que $B \in \mathcal{B}_p^*$. Por el Teorema 2.20, tenemos que $\operatorname{Fr}_Y(B) = \operatorname{Fr}_X(B)$. Luego, $|\operatorname{Fr}_X(B)| = |\operatorname{Fr}_Y(B)| < \infty$. Por lo tanto, X es un continuo regular. □

El siguiente ejemplo nos muestra que no todo continuo regular es una dendrita local.

3.8. Ejemplo. Para todo $n \in \mathbb{N}$, sea

$$C_n = \left\{ (x, y) \in \mathbb{R}^2 \colon \left(x - \frac{1}{2n} \right)^2 + y^2 = \left(\frac{1}{2n} \right)^2 \right\}.$$

Ahora, sea $A = \bigcup_{n \in \mathbb{N}} C_n$. El continuo A es conocido como **el arete hawaiano armónico,** el cual es un continuo regular y no es una dendrita local, porque para el punto $p = (0, 0) \in \mathbb{R}^2$, no existe una vecindad que sea una dendrita.



Arete hawaiano armónico

A continuación una propiedad de las dendritas locales, la cual es consecuencia del Teorema 3.7.

3.9. Corolario. Toda dendrita local es hereditariamente localmente conexo.

Demostración. Supongamos que X es un dendrita local. Por el Teorema 3.7, tenemos que X es un continuo regular. Luego, por el Teorema 2.16, concluimos que X es hereditariamente localmente conexo. □

Ahora, veamos una propiedad de dendritas locales, que nos asegura que toda dendrita local es unión finita de ciertos conjuntos abiertos.

3.10. Teorema. Si X es una dendrita local, entonces existen dendritas $D_1, \ldots, D_n$ tales que $X = \bigcup_{i=1}^n \operatorname{Int}_X(D_i)$.

Demostración. Sean X una dendrita local y $p \in X$. Existe una vecindad D_p de p que es una dendrita, luego, existe U_p abierto en X tal que $p \in U_p \subset D_p$.

Notemos que $X = \bigcup_{p\in X} U_p$. Luego, la colección $\mathcal{C} = \{U_p \colon p \in X\}$ es una cubierta abierta para X. Como X es compacto, existe una subcolección finita $\{U_{p_1}, \ldots, U_{p_n}\}$ de $\mathcal{C}$, tal que $X = \bigcup_{i=1}^{n} U_{p_i}$.

Así,

$$X = U_{p_1} \cup \cdots \cup U_{p_n} = \operatorname{Int}_X(U_{p_1}) \cup \cdots \cup \operatorname{Int}_X(U_{p_n}) \subset \operatorname{Int}_X(D_{p_1}) \cup \cdots \cup \operatorname{Int}_X(D_{p_n}).$$

Como $\operatorname{Int}_X(D_{p_1}) \cup \cdots \cup \operatorname{Int}_X(D_{p_n}) \subset X$, concluimos que

$$X = \operatorname{Int}_X(D_{p_1}) \cup \cdots \cup \operatorname{Int}_X(D_{p_n}). \qquad \square$$

Más aún, toda dendrita local es unión finita de dendritas.

3.11. Corolario. Si X es una dendrita local, entonces existen dendritas $D_1, \ldots, D_n$ tales que $X = \bigcup_{i=1}^{n} D_i$.

Demostración. Supongamos que X es una dendrita local. Por el Teorema 3.10, existen dendritas $D_1, \ldots, D_n$ contenidas en X tales que $X = \bigcup_{i=1}^{n} \operatorname{Int}_X(D_i)$. Notemos que

$$\operatorname{Int}_X(D_1) \cup \cdots \cup \operatorname{Int}_X(D_n) \subset D_1 \cup \cdots \cup D_n.$$

Luego, $X \subset D_1 \cup \cdots \cup D_n$. Además, claramente $D_1 \cup \cdots \cup D_n \subset X$. Así, $X = D_1 \cup \cdots \cup D_n$. $\square$

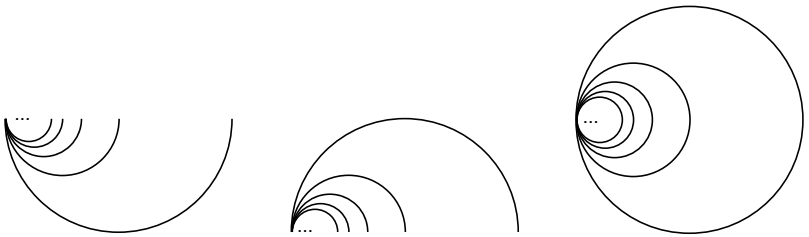
El siguiente ejemplo nos muestra que la unión finita de dendritas no es en general una dendrita local.

3.12. Ejemplo. Para todo $n \in \mathbb{N}$, sean

$$S_n = \left\{(x, y) \in \mathbb{R}^2 \colon \left(x - \frac{1}{2n}\right)^2 + y^2 = \left(\frac{1}{2n}\right)^2,\ y \geq 0\right\},$$

$$I_n = \left\{(x, y) \in \mathbb{R}^2 \colon \left(x - \frac{1}{2n}\right)^2 + y^2 = \left(\frac{1}{2n}\right)^2,\ y \leq 0\right\}.$$

Ahora, sean $D_1 = \bigcup_{n\in\mathbb{N}} S_n$ y $D_2 = \bigcup_{n\in\mathbb{N}} I_n$. Los espacios D_1 y D_2 son dendritas y $D_1 \cup D_2$ no es una dendrita local porque precisamente $D_1 \cup D_2$ coincide con el arete hawaiano armónico A (vea el Ejemplo 3.8).



Antes de ver una propiedad más de las dendritas locales, veamos el siguiente lema que nos ayudará a demostrar el Teorema 3.14.

3.13. Lema. [3, Lema 27.5, pág. 199] **Lema del número de Lebesgue.** Si X es un espacio métrico compacto y $\mathcal{C}$ es una cubierta abierta de X, entonces existe un número real $\delta > 0$ tal que para todo subconjunto de X con diámetro menor que δ, existe un elemento de $\mathcal{C}$ conteniéndolo.

El número δ se denomina **número de Lebesgue** para la cubierta $\mathcal{C}$.

Ahora sí, veamos otra propiedad de las dendritas locales.

3.14. Teorema. Si X es una dendrita local, entonces existe un número $\varepsilon > 0$ tal que todo subcontinuo C de diámetro menor que ε es una dendrita.

Demostración. Supongamos que X es una dendrita local. Por el Teorema 3.10, existen dendritas, $D_1, \ldots, D_n$, tales que $X = \bigcup_{i=1}^{n} \operatorname{Int}_X(D_i)$.

Por el Lema 3.13, existe un número $\varepsilon > 0$ tal que todo subconjunto de X con diámetro menor que ε, está contenido en alguno de los conjuntos $\operatorname{Int}_X(D_i)$. Por lo tanto, si C es un subcontinuo de X con diámetro menor que ε, existe $j \in \{1, \ldots, n\}$ tal que $C \subset \operatorname{Int}_X(D_j) \subset D_j$. Luego, por el Corolario 2.19, tenemos que C es una dendrita. □

En seguida, un resultado que nos da una condición suficiente para que un continuo localmente conexo sea una dendrital local.

3.15. Teorema. Sean X un continuo localmente conexo y $\varepsilon = \text{ínf}\,\{\text{diám}(C)\colon\ C$ es una curva cerrada simple contenida en $X\}$. Si $\varepsilon > 0$, entonces X es una dendrita local.

Demostración. Sea $p \in X$. Por el Teorema 2.21, existe $C \subset X$ tal que C es un continuo localmente conexo, C es una vecindad de p y $\text{diám}(C) < \varepsilon$.

Como $\text{diám}(C) < \varepsilon$, tenemos que C no contiene curvas cerradas simples. Luego, C es una dendrita.

Así, hemos probado que p tiene una vecindad que es una dendrita y por lo tanto X es una dendrita local. □

El teorema que sigue nos ayuda a probar una interesante caracterización de las dendritas locales.

3.16. Teorema. [2, Observación 2 del Teorema 11, pág. 288] Si X es un continuo regular, entonces para todo $\varepsilon > 0$, existe una colección finita de continuos regulares, $\{F_1, \ldots, F_n\}$, tal que

(1) $X = \bigcup_{i=1}^{n} F_i$.
(2) Para todo $i \in \{1, \ldots, n\}$, tenemos que $\text{diám}(F_i) < \varepsilon$.
(3) Para todo $i, j \in \{1, \ldots, n\}$ con $i \neq j$, tenemos que $|F_i \cap F_j| < \infty$.
(4) Para todo $i, j, k \in \{1, \ldots, n\}$ con $i \neq j, i \neq k, j \neq k$, tenemos que $F_i \cap F_j \cap F_k = \emptyset$.

3.17. Teorema. Sea X un continuo. Entonces las siguientes condiciones son equivalentes:

(1) X es una dendrita local.
(2) X es un continuo localmente conexo que contiene a lo más un número finito de curvas cerradas simples.
(3) X es la unión de una colección finita de dendritas $\{D_1, \ldots, D_n\}$ tal que para $i, j \in \{1, \ldots, n\}$ con $i \neq j$, tenemos que $|D_i \cap D_j| < \infty$.

DEMOSTRACIÓN. La prueba de (3) implica (2) se puede ver en [2, Teorema 4, págs. 303 y 304].

Veamos que (2) implica (1). Si X no contiene curvas cerradas simples, entonces X es una dendrita. Luego, X es una dendrita local. Si X sí contiene curvas cerradas simples, entonces sean $C_1, \ldots, C_n$ las curvas cerradas simples de X y $\varepsilon = \text{mín}\{\text{diám}(C_i)\colon i \in \{1, \ldots, n\}\}$. Como toda curva cerrada simple tiene diámetro positivo, se sigue que $\varepsilon > 0$. Por el Teorema 3.15, tenemos que X es una dendrita local.

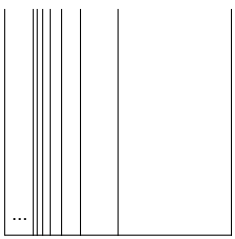
Veamos que (1) implica (3). Por el Teorema 3.14, existe un número $\varepsilon > 0$ tal que todo subcontinuo C de X con diámetro menor que ε es una dendrita. Por el Teorema 3.7, tenemos que X es un continuo regular. Por el Teorema 3.16, existe una colección de continuos regulares $\{D_1, \ldots, D_n\}$ tal que

(a) $X = \bigcup_{i=1}^{n} D_i$.
(b) Para todo $i \in \{1, \ldots, n\}$, tenemos que $\text{diám}(D_i) < \varepsilon$ (Notemos que cada D_i es una dendrita).
(c) Para todo $i, j \in \{1, \ldots, n\}$ con $i \neq j$, tenemos que $|D_i \cap D_j| < \infty$. □

Del Teorema 3.17, tenemos que si X es una continuo tal que no es localmente conexo o si X contiene una cantidad no finita de curvas cerradas simples, entonces X no es una dendrita local.

Veamos algunos ejemplos.

3.18. EJEMPLO. Sea $P = ([0,1] \times \{0\}) \cup (\{\frac{1}{n}\colon n \in \mathbb{N}\} \times [0,1]) \cup (\{0\} \times [0,1])$. El conjunto P, el **espacio peine,** es un continuo que no es localmente conexo. Luego, por el Teorema 3.17, tenemos que P no es una dendrita local.



El espacio peine.

3.19. EJEMPLO. Consideremos el arete hawaiano armónico A (vea el Ejemplo 3.8). El hecho de que A no sea una dendrita local se desprende más fácilmente del Teorema 3.17, porque A contiene una cantidad no finita de curvas cerradas simples.

Para concluir este trabajo, consideremos las siguientes notaciones:

$\mathcal{C}_1 = \{X\colon X \text{ es dendrita}\}$
$\mathcal{C}_2 = \{X\colon X \text{ es dendrita local}\}$
$\mathcal{C}_3 = \{X\colon X \text{ es continuo regular}\}$
$\mathcal{C}_4 = \{X\colon X \text{ es continuo hereditariamente localmente conexo}\}$
$\mathcal{C}_5 = \{X\colon X \text{ es continuo localmente conexo}\}$,

de acuerdo a algunos teoremas y ejemplos aquí revisados tenemos que estas clases de continuos guardan la siguiente relación:

$$\mathcal{C}_1 \subsetneq \mathcal{C}_2 \subsetneq \mathcal{C}_3 \subsetneq \mathcal{C}_4 \subsetneq \mathcal{C}_5.$$

Referencias

[1] L. A. Guerrero Méndez, *Clases de Continuos Localmente Conexos,* Tesis de Licenciatura, Facultad de Ciencias Físico Matemáticas, BUAP, 2009.
[2] K. Kuratowski, *Topology, Vol. II,* Academic Press, New York, N. Y., 1968.
[3] J. R. Munkres, *Topología, Segunda Edición,* Prentice Hall. Madrid. España, 2002.
[4] S. B. Nadler, Jr., *Continuum Theory: An Introduction,* Monographs and Textbooks in Pure and Applied Math. Vol. 158, Marcel Dekker, New York, Basel, Hong Kong, 1992.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

luisalberto_gm4@hotmail.com, dherrera@fcfm.buap.mx, fmacias@fcfm.buap.mx

Capítulo 28

¿TIENEN LAS DENDRITAS LOCALES PRODUCTO SIMÉTRICO ÚNICO?

DAVID HERRERA CARRASCO
FERNANDO MACÍAS ROMERO
FRANCISCO VÁZQUEZ JUÁREZ
FCFM - BUAP

Resumen. Sean X una dendrita y $\mathfrak{D}$ la clase de todas las dendritas cuyo conjunto de puntos extremos es cerrado. Sean X una dendrita local y $\mathfrak{LD}$ la clase de las dendritas locales tal que cada uno de sus puntos tiene una vecindad que pertenece a $\mathfrak{D}$. En este artículo estudiamos diversas propiedades de las clases $\mathfrak{D}$ y $\mathfrak{LD}$.

1. Introducción

Un **continuo** es un espacio métrico, compacto, conexo y no degenerado. Un continuo X es **localmente conexo** en $x \in X$ si para cada vecindad V de x existe un subconjunto abierto y conexo U de X tal que $x \in U \subset V$. El espacio X es **localmente conexo** cuando X es localmente conexo en cada uno de sus puntos. Una **dendrita** es un continuo localmente conexo sin curvas cerradas simples.

Sean X un continuo y $p \in X$. Sea $\mathfrak{n} \in \mathbb{N} \cup \{\aleph_0, \omega, \mathfrak{c}\}$, donde $\aleph_0$, ω, y $\mathfrak{c}$ denotan la cardinalidad de los números naturales, el primer ordinal numerable y la cardinalidad de los números reales, respectivamente. Sea $\mathfrak{n} \in \mathbb{N} \cup \{\aleph_0, \mathfrak{c}\}$. Se dice que p es de orden menor o igual que $\mathfrak{n}$ en X, se denota por $ord(p, X) \leq \mathfrak{n}$, si para cada abierto V en X tal que $p \in V$, existe U abierto en X tal que $p \in U \subset V$ y $|Fr(U)| \leq \mathfrak{n}$ Se dice que p es de **orden** $\mathfrak{n}$ en X, denotado por $ord(p, X) = \mathfrak{n}$, si $ord(p, X) \leq \mathfrak{n}$ y $ord(p, X) \nleq \alpha$ para cualquier $\alpha < \mathfrak{n}$. Se dice que p es de orden ω en X, se denota por $ord(p, X) = \omega$ si para cada abierto V en X tal que $p \in V$, existe U abierto en X tal que $p \in U \subset V$ y $|Fr(U)|$ es finito, pero $ord(p, X) \notin \mathbb{N}$ (por conveniencia, para cada $n \in \mathbb{N}$, suponemos que $n < \omega < \aleph_0$). Los puntos de orden 1, 2 y mayor o igual que 3 son los puntos extremos, puntos ordinarios y puntos de ramificación de X, respectivamente. Los conjuntos de puntos extremos, puntos ordinarios y puntos de ramificación de X son denotados por $E(X)$, $O(X)$ y $R(X)$, respectivamente.

En este artículo investigamos dendritas X cuyo conjunto de puntos extremos, $E(X)$, es cerrado. Sea

$$\mathfrak{D} = \{X \text{ dendrita} : E(X) \text{ es cerrado}\}.$$

Los resultados principales sobre la clase $\mathfrak{D}$ que estudiamos en este trabajo son los siguientes:

1. Si $X \in \mathfrak{D}$ y Y es un subcontinuo de X, entonces $Y \in \mathfrak{D}$.
2. Establecemos una caracterización de los elementos que pertenecen a la clase $\mathfrak{D}$.

Para conocer más resultados de la clase $\mathfrak{D}$, el lector puede consultar [10].

Una **dendrita local** es un continuo tal que cada uno de sus puntos tiene una vecindad que es una dendrita. También, en este artículo, investigamos algunas propiedades de las dendritas locales X pero con la condición que cada uno de sus puntos tiene una vecindad que pertenece a la clase $\mathfrak{D}$. Sea

$$\mathfrak{LD}= \{X \text{ dendrita local : cada punto de } X \text{ tiene una vecindad que está en } \mathfrak{D}\}.$$

Las clases $\mathfrak{D}$ y $\mathfrak{LD}$ son importantes en el tema de Hiperespacios de Continuos, para precisar en qué sentido son substanciales necesitamos lo siguiente. Dado un continuo X, un **hiperespacio** de X es una colección de subconjuntos de X con ciertas propiedades. Sea $n \in \mathbb{N}$, algunos hiperespacios son los siguientes.

$$C_n(X) = \{A \subset X : A \text{ es cerrado no vacío y tiene a lo más } n \text{ componentes}\} \text{ y}$$

$$F_n(X) = \{A \subset X : A \text{ tiene a lo más } n \text{ puntos}\}.$$

Los hiperespacios $C_n(X)$ y $F_n(X)$ tienen la topología inducida por la métrica de Hausdorff y son conocidos como el n-ésimo hiperespacio de X y el n-ésimo producto simétrico de X, respectivamente. Cuando $n = 1$, escribimos $C(X)$ en lugar de $C_1(X)$ y $C(X)$ es conocido como el hiperespacio de los subcontinuos de X.

Si dos continuos X y Y son homeomorfos, entonces, para $n \in \mathbb{N}$, es claro que $C_n(X)$ es homeomorfo a $C_n(Y)$ y $F_n(X)$ es homeomorfo a $F_n(Y)$. Sin embargo, puede ocurrir que dos espacios no homeomorfos tengan el mismo hiperespacio; por ejemplo, los hiperespacios de los subcontinuos de la circunferencia unitaria, $C(S^1)$, y del intervalo unitario $[0, 1]$, $C([0, 1])$, respectivamente, son homeomorfos, vea [14, Ejemplos 3.1 y 3.2].

1.1. Definición. Para un continuo X, sea $\mathcal{H}(X)$ alguno de los hiperespacios $C_n(X)$ o $F_n(X)$. Un continuo X tiene **hiperespacio único** $\mathcal{H}(X)$, si para cualquier continuo Y tal que $\mathcal{H}(X)$ es homeomorfo a $\mathcal{H}(Y)$, entonces X es homeomorfo a Y.

Surge de manera natural la siguiente pregunta:

1.2. Problema. ¿Bajo qué condiciones el continuo X tiene hiperespacio único $\mathcal{H}(X)$?

Para la clase $\mathfrak{D}$ hasta el momento sabemos lo siguiente:

(1) Si $X \in \mathfrak{D}$ y no es un arco, entonces X tiene hiperespacio único $C(X)$, vea [6, Teorema 10].
(2) Si X es una dendrita y $X \notin \mathfrak{D}$, entonces X no tiene hiperespacio único $C(X)$, vea [1, Teorema 5.2].
(3) Si $X \in \mathfrak{D}$, entonces X tiene hiperespacio único $C_2(X)$, vea [12] y [15, Teorema 3.1].
(4) Si $X \in \mathfrak{D}$, entonces X tiene hiperespacio único $C_n(X)$ para cada $n \in \mathbb{N} - \{1, 2\}$, vea [8, Teorema 5.7].
(5) Si $X \in \mathfrak{D}$, entonces X tiene hiperespacio único $F_2(X)$, vea [16, Teorema 8].

(6) Si $X \in \mathfrak{D}$, entonces X tiene hiperespacio único $F_n(X)$ para cada $n \in \mathbb{N}-\{2\}$, vea [11, Teorema 3.7].

Para la clase $\mathfrak{LD}$ hasta el momento sabemos lo siguiente:

(7) Si $X \in \mathfrak{LD}$ diferente de un arco y una curva cerrada simple, entonces X tiene hiperespacio único $C(X)$, vea [2, Corolario 5.2].
(8) Si $X \in \mathfrak{LD}$ y $n \in \mathbb{N} - \{1, 2\}$, entonces X tiene hiperespacio único $C_n(X)$, vea [9, Teorema 5.4].

Los autores de este artículo actualmente están investigando las siguientes cuestiones.

1.3. CONJETURA. Si $X \in \mathfrak{LD}$, entonces X tiene hiperespacio único $F_2(X)$.

1.4. CONJETURA. Si $X \in \mathfrak{LD}$, entonces X tiene hiperespacio único $F_n(X)$ para cada $n \in \mathbb{N} - \{1, 2\}$.

2. Preliminares

Sean X un espacio topológico y A un subconjunto de X, los símbolos $\overline{A}$, $Fr(A)$, $Int(A)$ y A' denotan la cerradura de A, la frontera de A, el interior de A y el derivado de A en X, respectivamente. Si $A \subset Y \subset X$, entonces $\overline{A}_Y$, $Fr_Y(A)$ y $Int_Y(A)$ denotan la cerradura de A, la frontera de A y el interior de A en el subespacio Y de X, respectivamente. La cardinalidad de un conjunto A se representa por $|A|$. Como es usual, los símbolos $\emptyset$, $\mathbb{N}$, y $\mathbb{R}$, representan el conjunto vacío, los números naturales y los números reales, respectivamente. De hecho, todos los conceptos no definidos aquí serán tomados como en [19].

En esta sección presentamos los resultados necesarios para el desarrollo de este artículo; en cada uno de ellos damos una referencia adecuada para el lector interesado en sus demostraciones.

2.1. Teorema. [21, Corolario 2.2, pág. 90] En un continuo hereditariamente localmente conexo, cuando hay una cantidad infinita de componentes en un conjunto abierto, ellas forman una sucesión nula.

2.2. Teorema. [19, Corolario 5.9] Sean X un continuo y A un subcontinuo propio de X. Si C es una componente de $X - A$, entonces $C \cup A$ es un continuo.

El siguiente resultado es útil en la construcción de funciones continuas a partir de otras funciones continuas, esta técnica se puede observar en las pruebas de los Teoremas 3.1 y 3.2.

2.3. Teorema. [5, Teorema 9.4, pág. 83] Sean X un espacio topológico y $\{A_\alpha : \alpha \in \Lambda\}$ una cubierta de X tal que:

(1) para cada $\alpha \in \Lambda$, tenemos que A_α es abierto en X o;
(2) para cada $\alpha \in \Lambda$, tenemos que A_α es cerrado en X, y tal cubierta forma una familia de vecindades finita.

Para cada $(\alpha, \beta) \in \Lambda \times \Lambda$, sea $f_\alpha : A_\alpha \to Y$ continua tal que $f_\alpha|_{A_\alpha \cap A_\beta} = f_\beta|_{A_\alpha \cap A_\beta}$. Entonces existe una única función continua $f : X \to Y$ extensión de f_α, es decir, $f|_{A_\alpha} = f_\alpha$.

2.4. Definición. Un **árbol** es un continuo que se puede escribir como una unión finita de arcos tales que la intersección de cualesquiera dos de ellos es finita y no contiene curvas cerradas simples.

A continuación damos dos resultados relacionados con el concepto de árbol; uno de ellos hace notar que la propiedad de árbol es una propiedad hereditaria.

2.5. Teorema. [19, Teorema 9.28] Sea X un continuo. Entonces X es un árbol si y sólo si el conjunto de puntos de no corte de X es finito.

2.6. Teorema. [19, Corolario 9.10.1] Cada subcontinuo de un árbol es un árbol.

2.7. Definición. Sean X un continuo y $A \subset X$. Se dice que X es un **continuo irreducible** respecto a A si ningún subcontinuo propio de X contiene a A.

2.8. Teorema. [21, 11.2] Si X es un continuo y A es subconjunto cerrado de X, entonces X contiene un subcontinuo irreducible respecto a A.

2.9. Teorema. [20, Teorema 2.16] Sean X un continuo localmente conexo, p un punto de corte de X y C una componente de $X-\{p\}$. Si $q \in E(\overline{C})$ y $q \neq p$, entonces $q \in E(X)$.

A continuación exponemos algunos ejemplos de dendritas, éstas juegan un papel importante en las siguientes dos secciones.

2.10. Ejemplo. Para cada $n \in \mathbb{N}$, sean $a_n = (\frac{1}{n}, \frac{1}{n^2})$ puntos de $\mathbb{R}^2$. Definimos $F_\omega = \bigcup_{n\in\mathbb{N}} [a, a_n]$, donde $a = (0, 0)$, vea la Figura 2.1. El continuo F_ω es una dendrita.

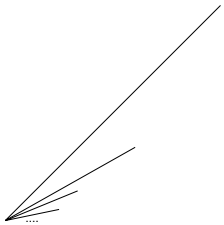


Figura 2.1: Dendrita F_ω.

2.11. Ejemplo. Para cada $n \in \mathbb{N}$, sean $a_n = (\frac{1}{n}, \frac{1}{n})$, $b_n = (\frac{1}{n}, 0)$ puntos de $\mathbb{R}^2$; sean $a = (0, 0)$ y $c = (-1, 0)$. Definimos $W_R = [a, b_1] \cup \big(\bigcup_{n\in\mathbb{N}} [a_n, b_n] \big)$ y $W = [c, a] \cup W_R$, vea las Figuras 2.2 y 2.3, respectivamente. Los continuos W_R y W son dendritas.

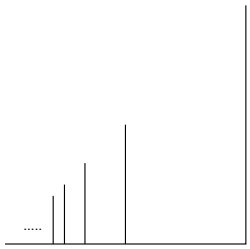


Figura 2.2: Dendrita W_R.

Existen muchas formas de caracterizar a las dendritas, vea [3], [13] y [19], enseguida enunciamos dos de ellas y algunas de sus propiedades.

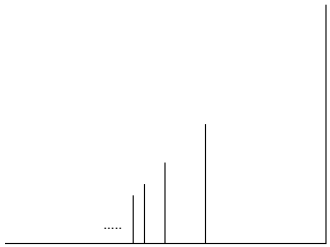


Figura 2.3: Dendrita W.

2.12. Teorema. [19, Teorema 10.7] Sea X un continuo. Entonces X es una dendrita si y sólo si cada punto de X es un punto de corte de X o un punto extremo de X.

2.13. Teorema. [19, Teorema 10.13] Sea X un continuo. Entonces X es una dendrita si y sólo si para cada $p \in X$ cuando $c(p, X)$ o $ord(p, X)$ es finito, tenemos que $c(p, X) = ord(p, X)$.

2.14. Teorema. [19, Corolario 10.5] Toda dendrita es hereditariamente localmente conexo.

2.15. Teorema. [19, Teorema 10.9] Cada subconjunto conexo no degenerado de una dendrita es arco conexo.

2.16. Teorema. [19, Corolario 10.20.1] Si X es una dendrita y $x \in X$, entonces $ord(x, X) \in \mathbb{N}$ o $ord(x, X) = \omega$.

Para concluir nuestra sección damos el concepto de dendroide y una propiedad de dicho concepto.

2.17. Definición. Un **dendroide** es un continuo hereditariamente unicoherente y arco conexo.

2.18. Definición. Un dendroide X es **suave** en $p \in X$ si para toda sucesión $\{x_n\}_{n=1}^{\infty}$ en X tal que $\lim_{n\to\infty} x_n = x$, tenemos que $\lim_{n\to\infty} [p, x_n] = [p, x]$.

2.19. Teorema. [18, Teorema 8] Sea X un dendroide. Entonces X es una dendrita si y sólo si X es suave en cada uno de sus puntos.

3. La clase $\mathfrak{D}$

En esta sección exponemos diversas propiedades de la clase $\mathfrak{D}$, para ello, empezamos probando dos resultados, el primero involucra a la dendrita F_ω, y el segundo resultado es una caracterización de árbol en la clase de las dendritas. Posteriormente mostramos que si $X \in \mathfrak{D}$, entonces todo subcontinuo Y de X es una dendrita y $E(Y)$ es cerrado, además, establecemos una caracterización de los elementos que están en $\mathfrak{D}$.

En seguida un resultado que nos da una condición suficiente para que una dendrita tenga un subcontinuo homeomorfo a F_ω.

3.1. Teorema. Sean X una dendrita y $x \in X$. Si $ord(x, X) = \omega$, entonces X contiene un subcontinuo homeomorfo a F_ω.

Demostración. Supongamos que $x \in X$ es tal que $ord(x, X) = \omega$. Si $c(x, X)$ es finito, por el Teorema 2.13, concluimos que $ord(x, X)$ es finito, esto contradice la hipótesis. De modo que $c(x, X) = \aleph_0$. Sean $C_1, C_2, \ldots$ las componentes de $X - \{x\}$. Por el Teorema 2.14, implicamos que X es hereditariamente localmente conexo. Aplicando el Teorema 2.1, tenemos que $\lim_{n\to\infty} diám(C_n) = 0$. Por el Teorema 2.2, para cada $n \in \mathbb{N}$, obtenemos que $C_n \cup \{x\}$ es un continuo. Para cada $n \in \mathbb{N}$, usando el Teorema 2.15, deducimos que $C_n \cup \{x\}$ es arco conexo. Para cada $n \in \mathbb{N}$, sean $y_n \in C_n$ y $A_n = [y_n, x] \subset C_n \cup \{x\}$. Así, $\lim_{n\to\infty} diám(A_n) = 0$ y $A_n \cap A_m = \{x\}$, si $n \neq m$.

Sea $Y = \bigcup_{n=1}^{\infty} A_n$. Veamos que Y es homeomorfo a F_ω. Para cada $n \in \mathbb{N}$, sea $B_n = [a_n, a]$ (donde a_n y a son como en el Ejemplo 2.10). Para cada $n \in \mathbb{N}$, existe $f_n : A_n \to B_n$ un homeomorfismo con $f_n(x) = a$. En particular, para cada $n \in \mathbb{N}$, deducimos que $f_n|_{A_n - \{x\}} : A_n - \{x\} \to F_\omega$ es continua. Notemos que la colección $\{A_n - \{x\}\}_{n=1}^{\infty}$ es una cubierta abierta de $Y - \{x\}$, luego por el Teorema 2.3, existe una función $f : Y - \{x\} \to F_\omega$ continua tal que $f|_{A_n - \{x\}} = f_n$ para cada $n \in \mathbb{N}$.

Sea $g : Y \to F_\omega$ definida, para cada $y \in Y$, por:

$$g(y) = \begin{cases} f(y), & \text{si } y \neq x; \\ a, & \text{si } y = x. \end{cases}$$

Veamos que g es continua en x. Sea U abierto en F_ω tal que $a \in U$. Como $\{B_n\}_{n=1}^{\infty}$ es una sucesión nula, existe $N \in \mathbb{N}$ tal que si $n \geq N$, entonces $B_n \subset U$. Esto implica que:

$$(i)\ \bigcup_{n=N}^{\infty} B_n \subset U.$$

Notemos que $U \cap B_1, \ldots, U \cap B_{N-1}$ son abiertos en $B_1, \ldots, B_{N-1}$, respectivamente. Dado que $U \cap B_1, \ldots, U \cap B_{N-1}$ contienen al punto a y que las funciones f_n son continuas en x, existen $V_1, \ldots, V_{N-1}$ abiertos en $A_1, \ldots, A_{N-1}$, respectivamente, que contienen al punto x tales que:

$$(ii)\ f_1(V_1) \subset U \cap B_1, \ldots, f_{N-1}(V_{N-1}) \subset U \cap B_{N-1}.$$

Para cada $n \in \{1, \ldots, N-1\}$, tenemos que $V_n = A_n \cap W_n$ con W_n abierto en Y. Sea $W = \bigcap_{n=1}^{N-1} W_n$. Notemos que W es abierto en Y y $x \in W$.

Veamos que $g(W) \subset U$, para esto, empezamos describiendo a W de otra manera.

$$W = W \cap \bigcup_{n=1}^{\infty} A_n = \big[W \cap \big(\bigcup_{n=1}^{N-1} A_n\big)\big] \bigcup \big[W \cap \big(\bigcup_{n=N}^{\infty} A_n\big)\big].$$

Por tanto,

$$g(W) = g\big(W \cap \big(\bigcup_{n=1}^{N-1} A_n\big)\big) \bigcup g\big(W \cap \big(\bigcup_{n=N}^{\infty} A_n\big)\big).$$

Dado que

$$g\big(W \cap \big(\bigcup_{n<N} A_n\big)\big) = g\big(\bigcup_{n=1}^{N-1} (W \cap A_n)\big)$$

$$\subset g(\bigcup_{n=1}^{N-1}(W_n \cap A_n)) = g(\bigcup_{n=1}^{N-1} V_n)$$

$$= \bigcup_{n=1}^{N-1} g(V_n) = \bigcup_{n=1}^{N-1} f_n(V_n),$$

aplicando (ii), deducimos que $g(W \cap (\bigcup_{n=1}^{N-1} A_n)) \subset \bigcup_{n=1}^{N-1}(U \cap B_n)$. Por lo que $g(W \cap (\bigcup_{n=1}^{N-1} A_n)) \subset U$.

Por otra parte,

$$g(W \cap (\bigcup_{n=N}^{\infty} A_n)) \subset g(\bigcup_{n=N}^{\infty} A_n)$$

$$= \bigcup_{n=N}^{\infty} g(A_n) = \bigcup_{n=N}^{\infty} f_n(A_n)$$

$$= \bigcup_{n=N}^{\infty} B_n,$$

aplicando (i), deducimos que $g(W \cap (\bigcup_{n=N}^{\infty} A_n)) \subset U$.

Con todo esto, $g(W) \subset U$. Por lo que g es continua en x. Luego, g es continua en Y.

Como,

$$g(Y) = g(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} g(A_n) = \bigcup_{n=1}^{\infty} f_n(A_n) = \bigcup_{n=1}^{\infty} B_n = F_\omega.$$

Por lo tanto, g es suprayectiva.

Veamos que g es inyectiva. Sean $p, q \in Y$ con $p \neq q$. Consideremos los dos casos posibles:

(a) Existe $n \in \mathbb{N}$ tal que $p, q \in A_n$. Dado que g es una extensión de f_n, tenemos que $g(p) = f_n(p)$ y $g(q) = f_n(q)$. Como f_n es inyectiva, $f_n(p) \neq f_n(q)$. Así, $g(p) \neq g(q)$.

(b) Existen $n, m \in \mathbb{N}$ con $n \neq m$ tales que $p \in A_n$ y $q \in A_m$. Dado que g es una extensión de f_n y f_m, implicamos que $f_n(p) = g(p)$ y $f_m(q) = g(q)$. Tenemos que $f_n(p) \in B_n$ y $f_m(q) \in B_m$. Así, $g(p) \neq g(q)$.

Así, g es inyectiva. Por lo que g es una función continua y biyectiva entre continuos. De aquí, concluimos que g es un homeomorfismo. Es decir, Y es un subcontinuo de X homeomorfo a F_ω. □

Dado que todo árbol es una dendrita, a continuación vemos cuándo se cumple el recíproco.

3.2. Teorema. Sea X una dendrita. Entonces X es un árbol si y sólo si X no contiene subcontinuos homeomorfos a F_ω ni a W_R.

Demostración. Supongamos que X es un árbol. Por el Teorema 2.5, implicamos que F_ω y W_R no son árboles. Aplicando el Corolario 2.6, deducimos que X no contiene subcontinuos homeomorfos a F_ω ni a W_R.

Recíprocamente, supongamos que X no contiene subcontinuos homeomorfos a F_ω ni a W_R. Por el Teorema 3.1 y Teorema 2.16, para cada $z \in X$, tenemos que $ord(z, X)$ es finito.

Veamos que $E(X)$ es finito. Para esto supongamos, por el contrario que $E(X)$ es infinito. Como X es compacto, existe una sucesión convergente $\{x_n\}_{n=1}^{\infty}$ de puntos distintos de $E(X)$. Sea $x = \lim_{n\to\infty} x_n$. Como $ord(x, X)$ es finito, por el Teorema 2.13, deducimos que $c(x, X)$ es finito y así podemos suponer que la sucesión $\{x_n\}_{n=1}^{\infty}$ está en una misma componente de $X - \{x\}$.

Para cada $n \in \mathbb{N}$, sea $p_n = \text{ínf}_{\leq_x}\{x_1, \ldots, x_n\}$. Dado que $p_n \leq_x x_i$ para cada $i \in \{1, \ldots, n\}$ y $p_{n+1} \leq_x x_i$ para cada $i \in \{1, \ldots, n+1\}$, tenemos que $p_{n+1} \leq_x p_n$. Es decir, con respecto al orden $\leq_x$, la sucesión $\{p_n\}_{n=1}^{\infty}$ es una sucesión decreciente. Además, para cada $n \in \mathbb{N} - \{1\}$, tenemos que $p_n \in R(X)$.

Veamos que $\lim_{n\to\infty} p_n = x$. Por el Teorema 2.19, tenemos que X es suave en x. Por lo que $[x, x_n] \to \{x\}$. Sea $\epsilon > 0$. Existe $N \in \mathbb{N}$ tal que si $n \geq N$, entonces $H_d([x, x_n], \{x\}) < \epsilon$. De aquí, tenemos que $[x, x_n] \subset N(\epsilon, \{x\})$. Dado que $p_n \in [x, x_n]$, implicamos que $p_n \in N(\epsilon, \{x\})$; es decir, $d(x, p_n) < \epsilon$. Por lo tanto, $\lim_{n\to\infty} p_n = x$.

Sin perdida de generalidad, supongamos que $p_n \neq p_m$ si $n \neq m$. Por el Teorema 2.8, existe un continuo irreducible Y respecto a $\{x_n : n \in \mathbb{N}\} \cup \{x\}$.

Veamos que Y es homeomorfo a W_R. Para cada $n \in \mathbb{N}-\{1\}$, sean $A_n = [x_{n-1}, p_n]$ y $B_n = [a_{n-1}, b_n]$, (donde a_{n-1} y b_n son como se definieron en el Ejemplo 2.11); notemos que $A_n \subset Y$.

Dado que A_n y B_n son homeomorfos, para cada $n \in \mathbb{N} - \{1\}$, existen homeomorfismos $g_n : A_n \to B_n$ tales que $g_n(x_n) = a_n$, $g_n(p_{n-1}) = b_{n-1}$ y $g_n(p_n) = b_n$. Por otra parte, como la colección $\{A_n : n \in \mathbb{N}-\{1\}\}$ es una cubierta de $Y-\{x\}$, son conjuntos cerrados, es una familia de vecindades finita y $g_n|_{A_n\cap A_m} = g_m|_{A_n\cap A_m}$ para cada $n, m \in \mathbb{N} - \{1\}$, así por el Teorema 2.3, existe una función $f : Y - \{x\} \to W_R$ continua tal que para cada $n \in \mathbb{N} - \{1\}$, tenemos que $f|_{A_n} = g_n$.

Sea $g : Y \to W_R$ definida, para cada $y \in Y$, por:

$$g(y) = \begin{cases} f(y), & \text{si } y \neq x; \\ a, & \text{si } y = x. \end{cases}$$

Veamos que g es continua en x. Para ello, por el Teorema 2.3, notemos que para cada $n \in \mathbb{N} - \{1\}$, las funciones $f_n : A_n \cup \{x\} \to B_n \cup \{a\}$ definidas, para cada $y \in A_n \cup \{x\}$, por:

$$f_n(y) = \begin{cases} g_n(y), & \text{si } y \neq x; \\ a, & \text{si } y = x. \end{cases}$$

son continuas.

Sea U abierto en W_R tal que $g(x) = a \in U$. Como tenemos que $\lim_{n\to\infty} diám(B_n \cup \{a\}) = 0$, existe $N \in \mathbb{N}$ tal que si $n \geq N$, entonces $B_n \cup \{a\} \subset U$. Así,

$$(i) \bigcup_{n=N}^{\infty} B_n \cup \{a\} \subset U.$$

Como $U \cap (B_1 \cup \{a\}), \ldots, U \cap (B_{N-1} \cup \{a\})$ son abiertos en $B_1 \cup \{a\}, \ldots, B_{N-1} \cup \{a\}$, respectivamente, tal que contienen al punto a y dado que las funciones f_n son continuas en x, existen $V_1, \ldots, V_{N-1}$ abiertos en $A_1 \cup \{x\}, \ldots, A_{N-1} \cup \{x\}$, respectivamente, que contienen al punto x tales que:

$$(ii)\ f_1(V_1) \subset U \cap (B_1 \cup \{a\}), \ldots, f_{N-1}(V_{N-1}) \subset U \cap (B_{N-1} \cup \{a\}).$$

Ahora para cada $n \in \{1, \ldots, N-1\}$, tenemos que $V_n = (A_n \cup \{x\}) \cap W_n$ con W_n abierto en Y. Sea $W = \bigcap_{n=1}^{N-1} W_n$. Notemos que W es abierto en Y y $x \in W$.

Veamos que $g(W) \subset U$, para esto, observemos que:

$$W = W \cap Y = W \cap \bigcup_{n=1}^{\infty} (A_n \cup \{x\})$$
$$= \big[W \cap (\bigcup_{n=1}^{N-1} A_n \cup \{x\})\big] \bigcup \big[W \cap (\bigcup_{n=N}^{\infty} A_n \cup \{x\})\big].$$

Por lo que,

$$g(W) = g(W \cap (\bigcup_{n=1}^{N-1} A_n \cup \{x\})) \bigcup g(W \cap (\bigcup_{n=N}^{\infty} A_n \cup \{x\})).$$

Como,

$$g(W \cap (\bigcup_{n=1}^{N-1} A_n \cup \{x\})) = g(\bigcup_{n=1}^{N-1} (W \cap (A_n \cup \{x\})))$$

$$\subset g(\bigcup_{n=1}^{N-1} (W_n \cap (A_n \cup \{x\}))) = g(\bigcup_{n=1}^{N-1} V_n)$$

$$= \bigcup_{n=1}^{N-1} g(V_n) = \bigcup_{n=1}^{N-1} f_n(V_n),$$

por (ii), implicamos que $g(W \cap (\bigcup_{n=1}^{N-1} A_n \cup \{x\})) \subset \bigcup_{n=1}^{N-1} (U \cap (B_n \cup \{a\}))$, de aquí, $g(W \cap (\bigcup_{n=1}^{N-1} A_n \cup \{x\})) \subset U$.

Por otra parte,

$$g(W \cap (\bigcup_{n=N}^{\infty} A_n \cup \{x\})) \subset g(\bigcup_{n=N}^{\infty} A_n \cup \{x\})$$

$$= \bigcup_{n=N}^{\infty} g(A_n \cup \{x\}) = \bigcup_{n=N}^{\infty} f_n(A_n \cup \{x\})$$

$$= \bigcup_{n=N}^{\infty} B_n \cup \{a\},$$

por (i), inferimos que $g(W \cap (\bigcup_{n=N}^{\infty} A_n \cup \{x\})) \subset U$.

Así, $g(W) \subset U$. Por lo que g es continua en x. Concluimos que g es continua en Y.

Para probar que g es biyectiva hacemos un procedimiento similar al que se hace para probar que la g, de la demostración del Teorema 3.1, lo es.

Como $g : Y \to W_R$ es una función continua y biyectiva entre continuos, tenemos que g es un homeomorfismo, este hecho contradice la hipótesis. Por lo tanto, $E(X)$ es finito.

Aplicando el Teorema 2.5, concluimos que X es un árbol. □

A continuación damos algunos ejemplos de dendritas que pertenecen a la clase $\mathfrak{D}$.

3.3. Ejemplo. Todo árbol pertenece a la clase $\mathfrak{D}$.

3.4. Ejemplo. La dendrita W_R pertenece a la clase $\mathfrak{D}$.

3.5. Ejemplo. Las dendritas F_ω y W no pertenecen a la clase $\mathfrak{D}$.

3.6. Ejemplo. Sea L un subconjunto de $\mathbb{N} - \{1, 2\}$ o bien $L = \{\omega\}$. En [4], se prueba que existe una dendrita D_L tal que:

(1) el $ord(p, D_L) \in L$ para cada $p \in R(D_L)$;
(2) para cada arco $[x, y]$ en D_L y cada $m \in L$, existe $p \in [x, y]$ tal que $ord(p, D_L) = m$.

En la Figura 3.1, aparece una aproximación de la dendrita D_3.

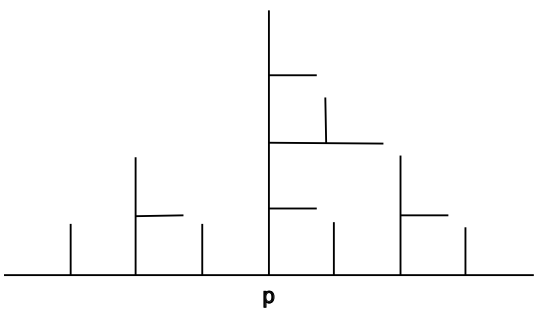


Figura 3.1: Dendrita D_3.

Observemos que p es un punto de acumulación de $E(D_3)$, pero $p \notin E(D_3)$. Es decir, el conjunto $E(D_3)$ no es cerrado. Por lo que D_3 no pertenece a la clase $\mathfrak{D}$. De hecho, para cada L subconjunto de $\mathbb{N} - \{1, 2\}$, tenemos que D_L no pertenece a la clase $\mathfrak{D}$.

El siguiente resultado muestra que el ser un elemento de la clase $\mathfrak{D}$ es una propiedad hereditaria.

3.7. Teorema. [20, Teorema 3.7] Sea $X \in \mathfrak{D}$. Si Y es un subcontinuo de X, entonces $Y \in \mathfrak{D}$.

Una manera sencilla para determinar si una dendrita pertenece a la familia $\mathfrak{D}$ es la siguiente.

3.8. Teorema. Sea X una dendrita. Entonces $X \in \mathfrak{D}$ si y sólo si X no contiene subcontinuos homeomorfos a F_ω ni a W.

Demostración. Supongamos que $X \in \mathfrak{D}$. Dado que $E(F_\omega)$ y $E(W)$ no son conjuntos cerrados, por el Teorema 3.7, concluimos que X no contiene subcontinuos homeomorfos a F_ω ni a W.

Recíprocamente, supongamos que X no contiene subcontinuos homeomorfos a F_ω ni a W. Veamos que $X \in \mathfrak{D}$. Supongamos lo contrario, es decir, existe $x \in X$ y una sucesión $\{x_n\}_{n=1}^{\infty}$ de puntos distintos de $E(X)$ tal que $\lim_{n\to\infty} x_n = x$, pero $x \notin E(X)$. Procediendo como en la demostración del Teorema 3.2, tenemos que X contiene un subcontinuo Y homeomorfo a W_R, donde x es la imagen de a. Por el Teorema 2.12, deducimos que x es un punto de corte de X, de aquí, $c(x, X) \geq 2$. Así, existe C una componente de $X - \{x\}$ tal que $Y \cap C = \emptyset$. Observemos que $C \cup \{x\}$ es arco conexo. Sean $y \in C$ y $[x, y] \subset C \cup \{x\}$. De modo que $Y \cup [x, y]$ es homeomorfo a W, esto contradice la hipótesis. Concluimos que $X \in \mathfrak{D}$. □

3.9. Teorema. Si $X \in \mathfrak{D}$, entonces $ord(x, X)$ es finito para cada $x \in X$.

Demostración. Supongamos que existe $x \in X$ tal que $ord(x, X) = \omega$. Por el Teorema 3.1, existe un subcontinuo de X homeomorfo a F_ω, lo cual no es posible por el Teorema 3.8. □

Si tenemos una sucesión convergente de puntos distintos de ramificación en una dendrita X tal que $E(X)$ es cerrado, entonces la sucesión converge a un punto extremo de X; esto lo establecemos en el resultado que sigue.

3.10. Teorema. [20, Teorema 3.9] Si $X \in \mathfrak{D}$ y $\{r_n\}_{n=1}^{\infty}$ es una sucesión de puntos distintos dos a dos en $R(X)$ convergente, entonces dicha sucesión converge a un punto extremo de X. Es decir, $R(X)' \subset E(X)$.

Demostración. Sea $r = \lim_{n\to\infty} r_n$. Por el Teorema 3.8, deducimos que X no contiene subcontinuos homeomorfos a F_ω, así por el Teorema 3.1 y el Teorema 2.16, cada punto de X es de orden finito, por lo que podemos suponer que todos los puntos r_n están en una componente de $X - \{r\}$. Sea $p_n = \text{ínf}_{\leq_r}\{r_1, \ldots, r_n\}$. Observemos que, cada p_n es un punto de ramificación de X. Además, todos los puntos p_n estań en el arco $[r, p_1]$, porque $p_{n+1} \leq_r p_n$ y la relación $\leq_r$ es transitiva.

Como $\lim_{n\to\infty} r_n = r$, $p_n \in [r, r_n]$ y dado que X es suave en r, tenemos que $\lim_{n\to\infty} p_n = r$. Además, podemos suponer que $p_n \neq p_m$, si $n \neq m$. Por el Teorema 2.9 y dado que $p_n \in R(X)$, para cada $n \in \mathbb{N}$, podemos elegir un punto extremo e_n de X en una componente C_n de $X - \{p_n\}$ tal que $C_n \cap [r, p_1] = \emptyset$. De modo que por el Teorema 2.1, la sucesión $\{C_n\}_{n=1}^{\infty}$ es una sucesión nula. Así, $\lim_{n\to\infty} diám(\overline{C_n}) = \lim_{n\to\infty} diám\big(C_n \cup \{p\}\big) = 0$. Esto implica que $\lim_{n\to\infty} d(p_n, e_n) = 0$ (donde d es la métrica de X), y como $\lim_{n\to\infty} p_n = r$, tenemos que $\lim_{n\to\infty} e_n = r$. Como $E(X)$ es un conjunto cerrado, se concluye que $r \in E(X)$. □

Terminamos esta sección con el siguiente resultado.

3.11. Teorema. [7, Teorema 2.31] Si $X \in \mathfrak{D}$ y $\{e_n\}_{n=1}^{\infty}$ es una sucesión en $E(X)$ convergente tal que $e_n \neq e_m$, si $n \neq m$, y $\lim_{n\to\infty} e_n = e \neq e_1$, entonces existe una sucesión $\{r_n\}_{n=1}^{\infty}$ de puntos distintos en $R(X) \cap [e, e_1]$ tal que $\lim_{n\to\infty} r_n = e$.

4. La clase $\mathfrak{LD}$

En esta sección anotamos, con su referencia adecuada, propiedades de la clase $\mathfrak{LD}$, recordemos que $\mathfrak{LD} = \{X$ dendrita local : cada punto de X tiene una vecindad

que está en $\mathfrak{D}$; después de leer dichas propiedades, el lector puede notar la similitud con los resultados establecidos en la Sección 3.

4.1. Observación. Si $X \in \mathfrak{D}$, entonces $X \in \mathfrak{LD}$.

4.2. Teorema. [2, Teorema 3.4] Sea X una dendrita local. Entonces $X \in \mathfrak{LD}$ si y sólo si X no contiene subcontinuos homeomorfos a F_ω ni a W.

4.3. Teorema. [2, Corolario 3.6] Si $X \in \mathfrak{LD}$ y Y es un subcontinuo de X, entonces $Y \in \mathfrak{LD}$.

4.4. Teorema. [2, Teorema 3.11] Si $X \in \mathfrak{LD}$, entonces $E(X)$ es cerrado en X.

En lo que sigue, ocuparemos el siguiente subconjunto de X:

$$E_a(X)= \{p \in X : \text{existe una sucesión en } E(X) - \{p\} \text{ que converge a } p\}.$$

4.5. Teorema. [2, Teorema 3.12] Sean $X \in \mathfrak{LD}$ y $x \in X$. Entonces las siguientes condiciones son equivalentes:

(1) $x \in E_a(X)$;
(2) x es el límite de una sucesión de puntos de ramificación de X distintos, todos están en un arco de X que contiene a x.

4.6. Teorema. [2, Corolario 3.13] Si $X \in \mathfrak{LD}$, entonces $O(X)$ es abierto en X.

4.7. Teorema. [2, Corolario 3.14] Si $X \in \mathfrak{LD}$ y A es un subcontinuo de X. Entonces se cumple lo siguiente:

(1) $E_a(A) \subset E_a(X)$;
(2) Si $A \cap E_a(X) = \emptyset$, entonces A es una gráfica finita.

Los resultados que hemos expuesto en este artículo acerca de la clase $\mathfrak{LD}$ son algunos de los que conocemos hasta el momento, creemos que dichos resultados forman una parte esencial para probar las Conjeturas 1.3 y 1.4.

Referencias

[1] G. Acosta y D. Herrera-Carrasco, *Dendrites without unique hyperspace,* Houston J. Math. 35 (2009) 451–467.

[2] G. Acosta, D. Herrera-Carrasco, F. Macías-Romero *Local dendrites with unique hiperespace $C(X)$*, Topology Appl. 157 (2010) 2069–2085.

[3] J. J. Charatonik y W. J. Charatonik, *Dendrites*, Aportaciones Matemáticas, Sociedad Matemática Mexicana, 22 (1998), 227-253.

[4] W. J. Charatonik y A. Dilks, *On self-homeomorphic spaces*, Topology Appl., 55 (1994), 215-238.

[5] J. Dugundji, *Topology*, Allyn and Bacon, Inc., Boston, 1966.

[6] D. Herrera-Carrasco, *Dendrites with unique hyperspace*, Houston J. Math., 33 (2007), 795-805.

[7] D. Herrera-Carrasco, *Hiperespacios de dendritas*, Tesis de Doctorado, Facultad de Ciencias UNAM, 2005.

[8] D. Herrera-Carrasco y F. Macías-Romero, *Dendrites with unique $n-$ fold hyperspace*, Topology Proc., 32 (2008), 321-337.

[9] D. Herrera-Carrasco y F. Macías-Romero *Local dendrites with unique $n-$fold hyperspace,* Topology Appl., doi:10.1016/j.topol.2010.11.004.

[10] D. Herrera-Carrasco, F. Macías-Romero y F. Vázquez-Juárez, *Dendritas Cuyo Conjunto de Puntos Extremos es Cerrado*, Memorias, 5^a Gran Semana Nacional de la Matemática, 5GSNM, 279-290, ISBN: 978-607-487-133-3, 2010.

[11] D. Herrera-Carrasco, M. de J. López y F. Macías-Romero *Dendrites with Unique Symmetric Products,* Topology Proc., 34 (2009), 175-190.

[12] D. Herrera-Carrasco, A. Illanes, M. de J. López y F. Macías-Romero *Dendrites with unique hyperspace $C_2(X)$,* Topology Appl., 156 (2009), 549–557.
[13] R. J. Hernández Gutiérrez, *Dendritas*, Tesis de Licenciatura, Facultad de Ciencias de la UNAM, 2007.
[14] A. Illanes, *Hiperespacios de continuos*, Aportaciones Matemáticas, Serie Textos N. 28, Sociedad Matemática Mexicana, ISBN: 968-36-3594-6, 2004.
[15] A. Illanes, *Dendrites with unique hyperspace $C_2(X)$, II*, aceptado Topology Proc., 34 (2009), 77-96.
[16] A. Illanes, *Dendrites with unique hyperspace $F_2(X)$*, JP J. Geom. Topol., 2 (2002), 75-96.
[17] J. L. Kelley, *Hyperspaces of a continuum*, Trans. Amer. Math. Soc., 52 (1942), 22-36.
[18] L. Lum, *Weakly smooth dendroids*, Fund. Math. 83 (1974), 111-120.
[19] S. B. Nadler, Jr., *Continuum Theory. An introduction.* Monographs and Texbooks in Pure and Applied Mathematics, Vol. 158, Marcel Dekker, New York, ISBN:0-8247-8659-9, 1992.
[20] F. Vázquez-Juárez, *Dendritas Cuyo Conjunto de Puntos Extremos es Cerrado*, Tesis de Maestría, Facultad de Ciencias Físico Matemáticas BUAP, 2010.
[21] G. T. Whyburn, *Analytic Topology.* American Mathematical Society Colloquium Publications, v. 28. New York: American Mathematical Society, 1942. Reprinted with corrections, 1971.

Facultad de Ciencias Físico Matemáticas, BUAP.
Av. San Claudio y 18 Sur, Col. San Manuel,
Puebla, Pue., C.P. 72570.

dherrera@fcfm.buap.mx, fmacias@fcfm.buap.mx, paco2013@hotmail.com

Capítulo 29

SUBESPACIOS EN ESPACIOS ORDENADOS

MANUEL IBARRA CONTRERAS
ARMANDO MARTÍNEZ GARCÍA
FCFM-BUAP

Resumen. Es conocido que si X es un espacio topológico linealmente ordenado y $Y \subseteq X$ entonces la topología en Y, inducida por el orden de X restringido a Y, no coincide con la topología relativa en Y. En este capítulo se mostrará que si Y es denso en el sentido del orden, o Y es compacto o Y es convexo, entonces estas dos topologías sobre Y, coinciden.

1. Introducción

Una relación sobre un conjunto X, $<$, es un orden lineal si se satisface que para todo $a, b, c \in X$:

(1) $a < b$ y $b < c$ implica que $a < c$,
(2) $a < b$ o $b < a$,
(3) Si $a < b$, es falso que $b < a$.

A la pareja $(X, <)$ se le llama conjunto linealmente ordenado. Ahora, para cualquier conjunto linealmente ordenado $(X, <)$, se dirá que $I \subset X$ es un intervalo abierto de X si existen $a, b \in X$ tales que:

$$I = (a, b) = \{x \in X : a < x < b\} \text{ o}$$
$$I = (\leftarrow, b) = \{x \in X : x < b\} \text{ o}$$
$$I = (a, \rightarrow) = \{x \in X : a < x\} \text{ o}$$
$$I = (\leftarrow, \rightarrow) = X.$$

Un espacio topológico (X, τ) *linealmente ordenado* si existe un orden lineal $<$ sobre X tal que la topología inducida por este orden

$$\tau_< = \{\emptyset\} \cup \{A \subset X \ : A \textit{ es unión de intervalos abiertos}\},$$

coincide con τ. En la siguiente sección describiremos esta situación con detalle, veremos que en un subespacio Y, de un espacio topológico linealmente ordenado, las topologías relativa y la inducida por el orden en el espacio restringido al conjunto Y, no coinciden y, finalmente, daremos algunas condiciones sobre Y que provocan la coincidencia de estas dos topologías. Si el lector está interesado en estudiar más acerca de esta problemática y otras más, relacionadas con los espacios topológicos linealmente ordenados, puede consultar [1] (Problemas 1.7.4, 2.7.5 y 3.12.3), [2] (capítulos 1 y 3) y [3].

2. Espacios ordenados y algunos de sus subespacios

En esta sección mostramos que en un subconjunto Y de un espacio ordenado X, las topologías relativa y la del orden restringido a Y no coinciden. También probaremos que si Y es denso en el sentido del orden, convexo o compacto entonces las dos topologías coinciden en Y.

2.1. Lema. Sea $(X, <)$ un conjunto linealmente ordenado. Entonces la familia

$$\beta = \{I \subset X : I \textit{ es un intervalo abierto de } X\}$$

es base para una topología en X, la cual será denotada como $\tau_<$.

Demostración. Para esto será suficiente ver que:

1) Para cualesquiera $I_1, I_2 \in \beta$ y $x \in I_1 \cap I_2$ existe $I \in \beta$ tal que $x \in I \subseteq I_1 \cap I_2$.

2) Para todo $x \in X$, existe $I \in \beta$ tal que $x \in I$.

Para ver (1) analicemos los posibles casos:

i) $I_1 = (a, b)$ con $a < b$; $I_2 = (c, d)$ con $c < d$.

Como $x \in I_1 \cap I_2$. Sea $r_1 = max\{a, c\}$ y $r_2 = min\{b, d\}$, eligiendo $I = (r_1, r_2)$ se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

ii) $I_1 = (a, b)$ con $a < b$; $I_2 = (\leftarrow, d)$.

Como $x \in I_1 \cap I_2$ entonces, $a < d < b$ o $b < d$, eligiendo $I = (a, d)$ o $I = I_1$, respectivamente, se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

iii) $I_1 = (a, b)$ con $a < b$; $I_2 = (c, \rightarrow)$.

Como $x \in I_1 \cap I_2$ entonces, $a < c < b$ o $c < a$, eligiendo $I = (c, b)$ o $I = I_1$, respectivamente, se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

iv) $I_1 = (\leftarrow, b)$; $I_2 = (\leftarrow, d)$.

Como $x \in I_1 \cap I_2$ entonces, $b < d$ o $d < b$, eligiendo $I = I_1$ o $I = I_2$, respectivamente, se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

v) $I_1 = (a, \rightarrow)$; $I_2 = (c, \rightarrow)$.

Como $x \in I_1 \cap I_2$ entonces, $a < c$ o $c < a$, eligiendo $I = I_1$ o $I = I_2$, respectivamente, se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

vi) $I_1 = (a, \rightarrow)$; $I_2 = (\leftarrow, b)$.

Como $x \in I_1 \cap I_2$ entonces, $a < b$, eligiendo $I = (a, b)$ se sigue que,

$$x \in I,\ I \in \beta \text{ y } I \subseteq I_1 \cap I_2.$$

vii) Si uno de los dos intervalos es $(\leftarrow, \rightarrow) = X$, entonces la intersección es igual al otro intervalo.

(2) sigue de lo que $X = (\leftarrow, \rightarrow)$ es un intervalo abierto. □

El Lema 2.1 justifica la siguiente definición.

2.2. Definición. Un espacio topológico X es un espacio topológico linealmente ordenado si su topología es generada por un orden lineal, $<$, en X. En este caso, denotaremos a este espacio como $(X, \tau_<)$ donde

$$\tau_< = \{\emptyset\} \cup \{A \subset X \ : A \textit{ es unión de intervalos abiertos}\}.$$

Por el Lema 2.1 tenemos que la familia

$$\beta = \{I \subset X : I \textit{ es un intervalo abierto de } X\}$$

es una base para esta topología.

Sean $(X, \tau_<)$ un espacio linealmente ordenado y $Y \subseteq X$.

La topología relativa en Y, inducida por $\tau_<$ la denotaremos como $\tau_<^Y$ y la topología en Y, generada por el orden lineal en X restringido a Y, la denotaremos como $\tau_{<_Y}$.

Es claro que una base para el espacio $(Y, \tau_<^Y)$ es la familia

$$\beta_1 = \{I \cap Y : I \textit{ es un intervalo abierto de } X\}$$

y una base para el espacio $(Y, \tau_{<_Y})$ es la familia

$$\beta_2 = \{J \subseteq Y : J \text{ es un intervalo abierto de } Y\}.$$

Observemos que si $J \in \beta_2$ entonces existen $a, b \in Y$ tal que,

$$J = (a, b) \text{ o } J = (a, \rightarrow) \text{ o } J = (\leftarrow, b) \text{ o } J = Y.$$

En caso necesario, a los intervalos abiertos en X los denotaremos como $(a, b)_X$ con $a, b \in X$ y de manera análoga en los otros casos y a los intervalos abiertos en Y los denotaremos como $(a, b)_Y$ con $a, b \in Y$ y de forma análoga en los otros casos.

El siguiente resultado, nos hace ver que siempre se satisface la siguiente contención:

$$\tau_{<_Y} \subseteq \tau_<^Y.$$

2.3. Lema. Sea $(X, \tau_<)$ un espacio topológico linealmente ordenado y $Y \subseteq X$. Entonces

$$\tau_{<_Y} \subseteq \tau_<^Y.$$

Demostración. Sea $U \in \tau_{<_Y}$ y $x \in U$, entonces existe un intervalo abierto $J \in \beta_2$ tal que $x \in J \subset U$, es decir, existen $a, b \in Y$ tal que,

$$J = (a, b)_Y \text{ o } J = (a, \rightarrow)_Y \text{ o } J = (\leftarrow, b)_Y \text{ o } J = Y.$$

En cualquiera de los casos como $Y \subseteq X$, y $a, b \in Y$ se sigue que $a, b \in X$ y, por lo tanto, eligiendo $I \in \beta$ con

$$I = (a, b)_X \text{ o } I = (a, \rightarrow)_X \text{ o } I = (\leftarrow, b)_X \text{ o } I = Y.$$

se sigue que $J \subseteq I$ y, de aquí, eligiendo $V = I \cap Y$, tenemos que

$$V \in \tau_<^Y \text{ y } x \in V \subset U.$$

Por lo tanto,

$$\tau_{<_Y} \subseteq \tau_<^Y.$$

□

El siguiente ejemplo nos permite ver que, en general, estas dos topologías son diferentes.

2.4. Ejemplo. Sean $X = \mathbb{R}$ con la topología inducida por el orden usual y

$$Y = \{x \in \mathbb{R} : x < 0\} \cup \{x \in \mathbb{R} : 1 \leq x\}.$$

Dado $z \in Y$ con $z > 1$ tenemos que

$$[1, z) = (0, z)_X \cap Y \in \tau_<^Y;$$

sin embargo, para todo $J \in \beta_2$ tal que $1 \in J$ se tiene que

$$J \cap \{x \in \mathbb{R} : x < 0\} \neq \emptyset \text{ y,}$$

por lo tanto,

$$[1, z) \notin \tau_{<_Y},$$

es decir, $\tau_<^Y$ es estrictamente mas fina que $\tau_{<_Y}$.

En estos momentos uno podría plantearse la siguiente pregunta:

Si ya sabemos que $\tau_<^Y$ es mas fina que $\tau_{<_Y}$, ¿existirá algún orden sobre Y de tal manera que $\tau_<^Y$ coincida con la topología inducida por ese orden?

Para darse cuenta de la magnitud de este problema (no es fácil) invitamos al lector a consultar [4], páginas 247-252.

La siguiente definición nos permitirá enunciar la primera condición sobre Y para que se dé la igualdad en el Lema 2.1.

2.5. Definición. Sean $(X, \tau_<)$ un espacio topológico linealmente ordenado y $Y \subseteq X$. Diremos que Y es *denso* en X en el sentido del orden si

$$\text{para todo } x, y \in X \text{ con } x < y \text{ existe } z \in Y \text{ tal que } x < z < y.$$

2.6. Teorema. Sean $(X, \tau_<)$ un espacio topológico linealmente ordenado y $Y \subseteq X$ denso en X en el sentido del orden. Entonces

$$\tau_<^Y = \tau_{<_Y}.$$

Demostración. Por el Lema 2.1 es suficiente ver que $\tau_<^Y \subseteq \tau_{<_Y}$.

Sean $U \in \tau_<^Y$ y $x \in U$, entonces existe $I \in \beta$ tal que,

$$x \in I \cap Y \subset U.$$

(a) Si x no es extremo de X, entonces existen $a, b \in X$ tal que $I = (a, b)_X$, es decir, $a < x < b$ y, como $a, x, b \in X$ y Y es denso en el sentido del orden, existen $s, t \in Y$ tales que $a < s < x < t < b$; por lo tanto, si $J = (s, t)_Y$ se sigue que

$$J \subseteq I,\ J \in \beta_2 \text{ y } x \in J \subseteq U.$$

(b) Si x es máximo y no es mínimo de X entonces

$$I = (\leftarrow, x]_X \text{ o } I = (a, x]_X$$

para algún $a \in X$. En el primer caso existe $a \in I$ tal que $a < x$, y entonces $(a, x] \subset I$. Como Y es denso en el sentido del orden existe $s \in Y$ tal que $a < s < x$ Por lo tanto, si $J = (s, x]_Y$ se sigue que

$$J \subseteq I,\ J \in \beta_2 \text{ y } x \in J \subseteq U.$$

(c) El caso cuando x es mínimo y no es máximo de X es ismilar al caso (b).

(d) Si x es el mínimo y el máximo de X, entonces X tiene sólo un punto, y $X = Y$.

□

El siguiente ejemplo nos muestra que no es suficiente pedir que $cl_X Y = X$ para que se dé la conclusión del Teorema 2.6.

2.7. Ejemplo. Sea

$$X = \{x \in \mathbb{R} : x \leq 0\} \cup \{x \in \mathbb{R} : 1 \leq x \leq 2\} \cup \{x \in \mathbb{R} : x \geq 3\},$$

con la topología inducida por el orden usual en $\mathbb{R}$ restringido a X y consideremos

$$Y = \{x \in \mathbb{R} : x < 0\} \cup \{x \in \mathbb{R} : 1 \leq x \leq 2\} \cup \{x \in \mathbb{R} : x > 3\}.$$

Observemos que $cl_X Y = X$. Es claro que

$$U = [1, 2) \in \tau_<^Y$$

ya que para $I = (1, 2)_X$ tenemos que $I \in \beta$ y $U = I \cap Y$. Sin embargo

$$U \notin \tau_{<_Y}$$

ya que para todo $J \in \beta_2$ tal que, $1 \in J$ tenemos que,

$$J \cap \{x \in \mathbb{R} : x < 0\} \neq \emptyset.$$

De donde se sigue que $\tau_<^Y$ y $\tau_{<_Y}$ no coinciden. Observemos que Y no es denso en el sentido del orden ya que para $0, 1 \in X$ no existe $y \in Y$ tal que $0 < y < 1$.

La siguiente definición nos permitirá dar la segunda condición bajo la cual se da la igualdad del Lema 2.1.

2.8. Definición. Sea $(X, \tau_<)$ un espacio topológico linealmente ordenado. $Y \subseteq X$ es convexo si

$$\text{para todo } x, y \in Y \ (x, y)_X \subseteq Y.$$

2.9. Teorema. Sean $(X, \tau_<)$ un espacio topológico linealmente ordenado y $Y \subseteq X$ convexo. Entonces

$$\tau_<^Y = \tau_{<_Y}.$$

Demostración. Por el Lema 2.1 es suficiente ver que $\tau_<^Y \subseteq \tau_{<_Y}$. Notemos que si Y tiene menos de dos puntos, entonces Y tiene sólo una topología, y la igualdad de las dos topologías es trivial. Entonces suponemos que Y tiene al ménos dos puntos.

Sea $U \in \tau_<^Y$ y $x \in U$.

Existen $a, b \in X \cup \{\leftarrow, \rightarrow\}$ tal que

$$x \in I = (a, b)_X \text{ y } x \in I \cap Y \subset U.$$

Si existe $y \in Y$ tal que $y < x$, ponemos $a = y$, en el caso contrario ponemos $c = \leftarrow$.

Si existe $z \in Y$ tal que $x < z$, ponemos $d = z$, en el caso contrario ponemos $d = \leftarrow$.

Si $(x, b)_X \cap Y \neq \emptyset$, sea $z \in (x, b)_X \cap Y$. Eligiendo $J = (y, z)_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

Si $(x, b)_X \cap Y = \emptyset$ y $(b, \rightarrow)_X \cap Y \neq \emptyset$ al considerar $z \in (b, \rightarrow)_X \cap Y$ se tiene que $(y, z) \subset Y$ y, por lo tanto, $b \in Y$. Así, eligiendo $J = (y, b)_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

Si $(x, \rightarrow)_X \cap Y = \emptyset$. Eligiendo $J = (y, \rightarrow)_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

Análogamente si $x < y$ y $y \in I \cap Y$.

Ahora supongamos que $y < x$ y $y \notin I$. Como $y < x$ y $y \notin I$ se sigue que $y < a$ y $(y, x)_X \subseteq Y$, es decir, $a \in Y$.

Si $(x, b)_X \cap Y \neq \emptyset$, sea $z \in (x, b)_X \cap Y$. Eligiendo $J = (a, z)_Y$ se sigue que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

Si $(x, b)_X \cap Y = \emptyset$ y $(b, \rightarrow)_X \cap Y \neq \emptyset$ sea $z \in (b, \rightarrow)_X \cap Y$; entonces $(x, z)_X \subseteq Y$ y, por lo tanto, $b \in Y$. Eligiendo $J = (a, b)_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

Si $(x, b)_X \cap Y = \emptyset$ y $(x, \rightarrow)_X \cap Y \neq \emptyset$. Eligiendo $J = (a, \rightarrow)_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

.

El caso $x < y$ y $y \notin I \cap Y$ se resuelve en forma similar al caso anterior.

En el caso que x sea extremo de X entonces

$$I = (a, x]_X \text{ o } I = [x, b)_X \text{ o } I = [x, \rightarrow)_X \text{ o } I = (\leftarrow, x]_X$$

para algún $a, b \in X$.

Si $I = (a, x]_X$ como Y tiene al menos dos puntos existe $y \in Y$ con $y \neq x$. Si $y \in I$, eligiendo $J = (y, x]_Y$ tenemos que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

. Si $y \notin I$ entonces $(y, x]_X \subseteq Y$, de donde se sigue que $a \in Y$. Eligiendo $J = (a, x]_Y$ se sigue que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

El caso en que $I = [x, b)_X$ se resuelve en forma similar.

Si $I = [x, \rightarrow)_X$, como Y tiene al menos dos puntos, existe $y \in Y$ con $y \neq x$. Eligiendo $J = [x, y)_Y$ se tiene que

$$J \in \beta_2,\ x \in J \text{ y } J \subseteq U.$$

El caso en que $I = (\leftarrow, x]_X$ se resuelve en forma similar. □

2.10. Corolario. Sean $(X, \tau_<)$ es un espacio topológico linealmente ordenado y $Y \subseteq X$ componente convexa de X. Entonces

$$\tau_<^Y = \tau_{<_Y}.$$

2.11. Ejemplo. Tomando X y Y como en el Ejemplo 2.7 podemos ver que Y no es convexo y que no se da la igualdad del Lema 2.1.

Antes de enunciar y probar el último resultado anunciado en este capítulo conviene recordar algunos resultados referentes a la compacidad en espacios linealmente ordenados.

2.12. Lema. Sean $(X, \tau_<)$ es un espacio topológico linealmente ordenado y $A \subseteq X$. Si A no es acotado superiormente, entonces X no es compacto.

Demostración. Como $A \subseteq X$ no es acotado superiormente para cada $x \in X$ existe $p(x) \in A$ tal que $x < p(x)$.

Para $a_0 \in A$ fijo consideremos la siguiente cubierta abierta de X:

$$\mathcal{U} = \{(a_0, x) : x \in X \text{ y } a_0 < x\} \cup \{(\leftarrow, p(a_0))\}$$

que no tiene ninguna subcubierta finita. En efecto, para cada cada familia finita

$$\{(a_0, x_i)\}_{i=1}^{i=n} \cup \{(\leftarrow, p(a_0))\}$$

consideremos a $x = max\,\{x_i : 1 \leq i \leq n\}$, entonces existe $p(x) \in A$ tal que

$$p(x) \notin \cup_{i=1}^{i=n}(a_0, x_i) \cup (\leftarrow, p(a_0))$$

de donde se sigue el resultado deseado. □

En forma similar se puede demostrar el siguiente resultado.

2.13. Corolario. Sean $(X, \tau_<)$ es un espacio topológico linealmente ordenado y $A \subseteq X$. Si A no es acotado inferiormente, entonces X no es compacto.

2.14. Teorema. Sea $(X, \tau_<)$ un espacio topológico linealmente ordenado. X es compacto si y sólo si para todo $A \subseteq X$ con $A \neq \emptyset$, A tiene supremo e ínfimo.

Demostración. La necesidad de sigue del Lema 2.12 y Corolario 2.13.

Suficiencia. Sea $\mathcal{U}$ cubierta abierta y y_0 el primer elemento de X.

Consideremos el conjunto:

$$S = \{y \in X : [y_0, y) \textit{ se puede cubrir por una familia finita de } \mathcal{U}\}.$$

Por hipótesis podemos considerar $\alpha = supS$. Entonces existe $U \in \mathcal{U}$ tal que $\alpha \in U$, por lo tanto existe $I \in \beta$ tal que $\alpha \in I \subseteq U$.

Supongamos que existen $a, b \in X$ con $a < b$ tal que $I = (a, b)$, entonces $a < \alpha < b$ lo cual implica que $a \in S$ y que $(\alpha, b) = \emptyset$ ya que si $z \in (\alpha, b)$ entonces $z \in S$ contradiciendo la definición de α. De aquí se sigue que $b \in S$ lo cual también es imposible, por lo tanto $I = (a, \rightarrow)$. Entonces $[a_0, a)$ puede ser cubierto por un número finito de elementos de $\mathcal{U}$ y $(a, \rightarrow) \subseteq U$. Por lo tanto X es cubierto por un número finito de elementos de $\mathcal{U}$, es decir, X es compacto. □

Ahora sí, podemos enunciar y probar el siguiente teorema

2.15. Teorema. Sean $(X, \tau_<)$ un espacio topológico linealmente ordenado y $Y \subseteq X$ compacto. Entonces

$$\tau_<^Y = \tau_{<_Y}.$$

Demostración. Por el Lema 2.1 es suficiente ver que $\tau_<^Y \subseteq \tau_{<_Y}$.

Sea $U \in \tau_<^Y$, entonces existe $V \in \tau_<$ tal que $U = V \cap Y$. Para cada $z \in U$ existe $I \in \beta$ tal que $z \in I \cap Y \subset U$.

Aplicando el Teorema 2.14 podemos considerar

$$z_1 = inf\,\{y \in I : y < z, y \in Y\}$$
$$z_2 = sup\,\{y \in I : z < y, y \in Y\}.$$

Por lo tanto para $J = (z_1, z_2)_Y$ tenemos que

$$J \in \beta_2,\ z \in J \text{ y } J \subseteq U.$$

Si $(\leftarrow, z) \cap Y = \emptyset$, entonces z es el primer elemento de Y y $z \in (\leftarrow, z_2)_Y \subset I \cap Y \subset U$.

Si $(z, \rightarrow) \cap Y = \emptyset$, entonces z es el último elemento de Y y $z \in (z_1, \rightarrow)_Y \subset I \cap Y \subset U$.

de donde se sigue que

$$\tau_<^Y \subseteq \tau_{<_Y}.$$

□

2.16. Ejemplo. Tomando X y Y como en el Ejemplo 2.7 podemos ver que Y no tiene supremo y que no se da la igualdad del Lema 2.1.

3. colofón

A sugerencia del árbitro incluimos la siguiente demostración alternativa del Teorema 2.15:

- Todo espacio linealmente ordenado X tiene la propiedad de Hausdorff. En efecto, si $x, y \in X$ con $x \neq y$, entonces, sin pérdida de generalidad, podemos suponer que $x < y$. Si existe $z \in (x, y)$, entonces $(\leftarrow, z)$ y $(z, \rightarrow)$ son abiertos ajenos que separan a x y a y. Si $(x, y) = \emptyset$, entonces $(\leftarrow, y)$ y $(x, \rightarrow)$ separan a x y a y.
- Con las hipótesis del Teorema 2.15, la función identidad $Id : (Y, \tau_<^Y) \rightarrow (Y, \tau_{<_Y})$, es un homeomorfismo, pues $(Y, \tau_<^Y)$ es un espacio compacto, la función Id es una biyección continua y $(Y, \tau_{<_Y})$ es Hausdorff.

La conclusión de los puntos anteriores nos da la conclusión deseada: $\tau_<^Y = \tau_{<_Y}$.

Referencias

[1] R. Engelking, *General Topology, Revised and completed edition*, Sigma Series in Pure Mathematics 6, Heldermann Verlag, Berlin.
[2] M. Ibarra, A. Martínez, *Espacios topológicos linealmente ordenados*, Tesis Profesional, UNAM, 1986.
[3] D.J. Lutzer, *On generalized ordered spaces*, Dissertationes Math. 89 1971.
[4] D.J. Lutzer, *Ordered topological spaces*, en: G. M. Reed (ed.), Surveys in General Topology, Academic Press, New York, 1980, 247-295.

Facultad de Ciencias Físico Matemáticas, BUAP.
Avenida San Claudio y 18 Sur, Colonia San Manuel,
Puebla, Pue. C.P. 72570.

mibarra@fcfm.buap.mx, maga@fcfm.buap.mx

Matemáticas y sus Aplicaciones I,
de Miguel Ángel García Ariza, Fernando Macías Romero
y José Jacobo Oliveros Oliveros, se terminó de imprimir
el 30 de septiembre de 2011, en los talleres de
El Errante editor, S.A. de C.V., sito en
Privada Emiliano Zapata 5947,
Col. San Baltasar Campeche, Puebla, Pue.

La composición tipográfica y el cuidado
de edición son de los autores y la coordinación,
de Fernando Macías Romero.

El tiraje consta de 283 ejemplares.