

**ESTUDIO TEÓRICO Y COMPUTACIONAL DE MÉTODOS
NUMÉRICOS PARA ECUACIONES DIFERENCIALES ORDINARIAS**

CHRISTIAM FERNANDO PISTALA CEBALLOS

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
UNIVERSIDAD DE NARIÑO
SAN JUAN DE PASTO
2017**

**ESTUDIO TEÓRICO Y COMPUTACIONAL DE MÉTODOS
NUMÉRICOS PARA ECUACIONES DIFERENCIALES ORDINARIAS**

CHRISTIAM FERNANDO PISTALA CEBALLOS

**Trabajo presentado como requisito parcial para optar al título de
Licenciado en Matemáticas**

**Asesora
Catalina María Rúa Álvarez
Doctora en Matemática Aplicada**

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
UNIVERSIDAD DE NARIÑO
SAN JUAN DE PASTO
2017**

Nota de Responsabilidad

Todas las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva de los autores.

Artículo 1^{ro} del Acuerdo No. 324 de octubre 11 de 1996 emanado por el Honorable Consejo Directivo de la Universidad de Nariño.

San Juan de Pasto, Junio 1 de 2017

Nota de Aceptación

Catalina María Rúa Álvarez
Presidenta de Tesis

Luis Eduardo López Montenegro
Jurado

Luis Andrés Santacruz Almeida
Jurado



ACUERDO NÚMERO 057 DE 2017
(24 de Mayo)

Por la cual se otorga una distinción de **LAUREADO** al trabajo de Grado del estudiante **CHRISTIAM FERNANDO PISTALA CEBALLOS**.

EL CONSEJO DE LA FACULTAD DE CIENCIAS EXACTAS Y NATURALES DE LA UNIVERSIDAD DE NARIÑO, en uso de sus atribuciones reglamentarias y estatutarias y,

CONSIDERANDO:

Que mediante Proposición No.029 de Mayo 18 de 2017, el Comité Curricular y de Investigaciones del Departamento de Matemáticas y Estadística, solicitó se otorgue la distinción de **LAUREADO** al trabajo de grado titulado: "ESTUDIO TEÓRICO Y COMPUTACIONAL DE MÉTODOS NUMÉRICOS PARA ECUACIONES DIFERENCIALES ORDINARIAS", presentado por el estudiante **CHRISTIAM FERNANDO PISTALA CEBALLOS**, como requisito parcial para optar al título de Licenciado en Matemáticas, bajo la Asesoría de la doctora CATALINA MARÍA RÚZ ALVAREZ, LUIS ANDRÉS SANTACRUZ ALMEIDA y LUIS EDUARDO LÓPEZ MONTENEGRO como integrantes del Jurado Evaluador;

Que mediante Acuerdo 030 de mayo 10 de 2016, el Comité Curricular y de Investigaciones del Departamento de Matemáticas y Estadística aprobó la realización del Proyecto de Grado denominado "ESTUDIO TEÓRICO Y COMPUTACIONAL DE MÉTODOS NUMÉRICOS PARA ECUACIONES DIFERENCIALES ORDINARIAS" bajo la responsabilidad del estudiante CHRISTIAM FERNANDO PISTALA CEBALLOS como requisito parcial para optar al título de Licenciado en Matemáticas, bajo la Asesoría de la doctora CATALINA MARÍA RÚZ ALVAREZ, LUIS ANDRÉS SANTACRUZ ALMEIDA y LUIS EDUARDO LÓPEZ MONTENEGRO como integrantes del Jurado Evaluador,

Que mediante Acuerdo 021 de mayo 2 de 2017 y previo el cumplimiento de los requisitos exigidos por la Institución, se autorizó la sesión de sustentación del citado Trabajo de Grado,

Que en sesión celebrada el 8 de mayo de 2017 los integrantes del Jurado Evaluador otorgaron una calificación conjunta de 100/100 puntos, hecho consignado en el Acta de Sustentación No. 004 de 2017,

Que según comunicaciones emitidas por los integrantes del Jurado Evaluador, se sustenta las razones por las cuales el trabajo antes citado es acreedor a la distinción de **LAUREADO**,

Que los integrantes del Jurado Evaluador destacan la alta calidad, evidenciando un buen trabajo investigativo, su profundidad en el análisis y el aporte para los estudiosos de las ecuaciones diferenciales tanto de la Universidad como fuera de ella, observando igualmente la excelente sesión de sustentación fundamentada en el dominio del tema y la difusión que ha



tenido el trabajo incluso antes de estar culminado, ya que el prenombrado estudiante Pístala Ceballos lo ha dado a conocer en eventos nacionales e internacionales de orden académico,

Que el Comité Curricular y de Investigaciones del Departamento de Matemáticas y Estadística, una vez evaluados los conceptos, considera que la distinción de LAUREADO se ajusta a las normas,

Que este organismo considero viable la petición y;

ACUERDA:

PRIMERO: Otorgar la distinción de **LAUREADO** (100/100 puntos) al Trabajo de Grado denominado "**ESTUDIO TEÓRICO Y COMPUTACIONAL DE MÉTODOS NUMÉRICOS PARA ECUACIONES DIFERENCIALES ORDINARIAS**", presentado por el estudiante **CHRISTIAM FERNANDO PISTALA CEBALLOS** como requisito parcial para optar al título de Licenciado en Matemáticas, desarrollado bajo la Asesoría de la doctora **CATALINA MARÍA RÚA ALVAREZ** y los profesores **LUIS ANDRÉS SANTACRUZ ALMEIDA** y **LUIS EDUARDO LÓPEZ MONTENEGRO** como integrantes del Jurado Evaluador, según la parte motiva del presente acuerdo.

SEGUNDO: El Departamento de Matemáticas y Estadística, la oficina de Registro Académico y la Facultad de Ciencias Exactas y Naturales, anotarán lo de su cargo

San Juan de Pasto, 24 de Mayo de 2017.

HERNÁN ABDÓN GARCÍA
Presidente

DUVI CASTILLO MENESSES
Secretaria

Elaboró: Duvi C.

*Este trabajo está dedicado a:
Mi hija Emmily Karolina Pistala.*

Agradecimientos

En primer lugar quisiera dar gracias a Dios por permitir la culminación de este proyecto. A mis padres Jorge Pistala y Dora Ceballos, quienes han sido el pilar fundamental en mi vida, hoy que culmino esta etapa quiero decirles muchas gracias, por su apoyo incondicional, por sus consejos, su amor, su esfuerzo y sobre todo porque hoy más que nunca sé que el estudio y educación recibida es la mejor herencia que han podido concederme, estoy seguro que no me alcanzará la vida para agradecerles todo lo que han hecho por mí.

A mi asesora de tesis Catalina Rúa, por su orientación y dedicación durante la elaboración de este trabajo, por estar siempre presta a resolver dudas e impulsarme a asumir retos que me han hecho crecer como persona. Agradezco a todos los profesores del programa de Licenciatura en Matemáticas de los cuales aprendí no solo conocimiento teórico sino también la gran responsabilidad de asumir mi profesión con seriedad y pensamiento crítico.

A mis hermanos que han sido mis consejeros, mis ejemplos a seguir. A la madre de mi hija por sus consejos, motivación, apoyo y sobre todo su comprensión en aquellos momentos difíciles de sobrellevar en el proceso de este trabajo de grado. A mi hija que es mi mayor fuente de inspiración, ese motor que me impulsa a seguir adelante con entusiasmo, luchando por alcanzar mis metas.

Finalmente agradezco a todas aquellas personas que de alguna forma han contribuido en la realización y logro de este trabajo de grado, el cual ha sido de gran satisfacción y motivo de orgullo para mí, pues durante el desarrollo de este trabajo me superarme como persona de forma intelectual y moral.

Resumen

La modelación matemática de una gran variedad de fenómenos en la naturaleza se relacionada con ecuaciones diferenciales ordinarias, EDO. A pesar de que en la teoría se presentan algunos métodos analíticos que resuelven diferentes clases de EDO, la cantidad de familias de EDO que pueden ser resueltas analíticamente es bastante restringida, por tanto los métodos numéricos son una excelente opción para encontrar aproximaciones discretas a la solución.

En este trabajo se presenta la base teórica de algunos métodos numéricos para resolver una EDO o un sistema de EDO sujeto a una condición inicial, conocido como problemas de Cauchy. Se distingue entre métodos de paso único y métodos de paso múltiple, entre los que se destacan respectivamente, la familia de métodos de Runge Kutta y la de Adams Bashford. Se analiza la deducción de estos y otros métodos, además se presenta un estudio de propiedades teóricas como son la consistencia, convergencia y estabilidad, entre otras. El estudio de estas propiedades es muy importante ya que permiten determinar si la solución numérica obtenida es de buena precisión.

Con base a la teoría y a diferentes simulaciones numéricas realizadas, se listan ventajas y desventajas al usar los diferentes métodos y se comparan soluciones exactas con aproximaciones en diferentes problemas de Cauchy. Finalmente, se aplican algunos de los métodos estudiados para determinar aproximaciones de un modelo matemático que describe la dinámica de infección del VIH/SIDA presentado en [17]. En este modelo la aproximación numérica es indispensable dado que no posee una solución analítica.

Abstract

The mathematical modeling of a variety of phenomena in the nature are related with Ordinary Differential Equations, ODE. Although, the theory presents analytics methods that solve some ODE the quantity of families of ODE that could be solved analytically is strongly restricted. Due to it numerical methods are an excellent option to obtain discrete approximations to a solution.

In this work we present the theory of some numerical methods that solve an ODE or a system of ODE joint with an initial condition, also known as the Cauchy's problem. We highlight one-step and multi-step methods, such as Runge Kutta and Adams Bashforth methods. We analyse their deduction and also we present a study of theoretical properties, furthermore here we distinguish the consistency, convergence and stability of them. The study of this properties allow us to accept or reject the approximation when it is compared with the precision.

We list the advantages and the disadvantages when are used some numerical methods, based on the theory and different numerical simulations done during this work. Therefore, we compare some approximations with their exact solution in several Cauchy's problems. Finally, we apply the numerical methods studied to approximate a mathematical model related to the dynamic of HIV/AIDS infection in [17]. That model do not have an exact solution and the relevance of the numerical approximation is evident.

Índice general

Lista de Tablas	xI
Lista de Figuras	xII
Abreviaciones	xIII
Introducción	xIV
1. Teoría general de ecuaciones diferenciales	1
1.1. Ecuaciones diferenciales ordinarias	1
1.2. Sistemas de ecuaciones diferenciales ordinarias de primer orden	2
1.2.1. Existencia y unicidad	4
1.2.2. Sistema lineal de EDO de primer orden	6
1.2.3. Solución de SEDOP con coeficientes constantes	8
1.2.4. Método matriz exponencial	11
1.3. Introducción a la solución numérica de EDO	13
1.3.1. Método de Euler	14
1.3.2. Método de Euler para sistemas de EDO	16
1.3.3. Una cota para el error	19
1.3.4. Métodos implícitos	22
2. Métodos de paso único	23
2.1. Forma general de los métodos de paso único	23
2.2. Consistencia, estabilidad y convergencia	24
2.3. Métodos de Taylor	28
2.4. Métodos de Runge Kutta	29
2.5. Estabilidad absoluta	32
3. Métodos de paso múltiple	37
3.1. Forma general de los métodos de paso múltiple lineales	37
3.2. Deducción por series de Taylor	38
3.3. Integración numérica	40
3.4. Métodos de Adams	42
3.5. Consistencia, convergencia y estabilidad	45
3.6. Estabilidad absoluta	50

4. Resultados Numéricos	53
4.1. Validación de implementaciones	53
4.1.1. Métodos de paso único	54
4.1.2. Métodos de paso múltiple	67
4.1.3. Comparación entre los métodos de paso único y paso múltiple	71
4.2. Aplicación VIH/SIDA.	72
4.2.1. Generalidades VIH/SIDA	72
4.2.2. Formulación del modelo	73
4.2.3. Modelo con estrategia de control.	75
4.2.4. Simulaciones numéricas para el modelo VIH.	75
5. Conclusiones y trabajos futuros	82
5.1. Conclusiones	82
5.2. Trabajos futuros	83
Apéndice	84
A.1. Interpolación	84
A.2. Adaptatividad y estimación para el error	86
A.3. Ecuaciones de diferencias	90
A.4. Implementaciones	92
Referencias	95
Índice alfabético	97

Índice de tablas

1.1. Solución Ejemplo 1.4, método de Euler.	17
1.2. Solución del Ejemplo 1.5, método de Euler para diferentes tamaños de paso.	18
1.3. Cota para el error método de Euler.	21
2.1. Intervalos de estabilidad absoluta para los métodos de Runge-Kutta	35
4.1. Convergencia MPU, Problema 4.1.	56
4.2. Convergencia MPU, Problema 4.2.	56
4.3. Convergencia MPU, Problema 4.6.	57
4.4. Solución numérica Problema 4.4, método RKF.	68
4.5. Convergencia MPM, Problema 4.1.	69
4.6. Convergencia MPM, Problema 4.2.	69
4.7. Convergencia MPM, Problema 4.6.	70
4.8. Valores iniciales AB4 con Euler y RK44, Problema 4.2.	71
4.9. Conteo de células T sanas y diagnóstico.	74
4.10. Parámetros del modelo VIH/SIDA.	76
4.11. Comparación numérica para $T(t)$	77
4.12. Comparación numérica para $I(t)$	77
4.13. Comparación numérica para $V(t)$	77
4.14. Solución numérica modelo VIH/SIDA.	79
4.15. Solución numérica modelo VIH/SIDA, falla Euler.	80
4.16. Solución numérica para modelo VIH/SIDA, eficiencia tratamiento, método Rk44. . .	81

Índice de figuras

1.1. Interpretación geométrica método de Euler.	15
2.1. Región de estabilidad absoluta para los métodos de Runge Kutta.	35
3.1. Región de estabilidad absoluta Adams Bashforth 2.	52
4.1. Convergencia MPU, Problema 4.1.	57
4.2. Estabilidad absoluta Euler, Problema 4.5.	58
4.3. Estabilidad absoluta método de Euler, valores de h críticos, Problema 4.3.	59
4.4. Estabilidad absoluta método de Euler, h en la región de estabilidad, Problema 4.3. .	60
4.5. Estabilidad absoluta método RK33, valores de h críticos, Problema 4.3.	61
4.6. Estabilidad absoluta método RK33, h en la región de estabilidad, Problema 4.3. . .	61
4.7. Estabilidad absoluta método de Euler, Problema 4.7.	63
4.8. Estabilidad absoluta método RK33, Problema 4.7.	64
4.9. Estabilidad absoluta método de Euler implícito, Problema 4.3.	65
4.10. Prueba de tiempos MPU, Problema 4.1.	66
4.11. Solución exacta Problema 4.4.	67
4.12. Convergencia MPM, Problema 4.1.	68
4.13. Prueba de tiempos MPM, Problema 4.1.	70
4.14. Ciclo de vida del VIH	73
4.15. Comparación en la solución numérica del método Rk44 y [10].	78
4.16. Comparación en la solución numérica del método Rk44 y [19].	78
4.17. Solución numérica para modelo VIH/SIDA, método Rk44.	79
4.18. Variación del tamaño de paso modelo VIH/SIDA, método RKF.	80
4.19. Solución numérica para modelo VIH/SIDA, eficiencia tratamiento, método Rk44. .	81

Abreviaciones

- EDO: Ecuación diferencial ordinaria.
- SEDO: Sistema de ecuaciones diferenciales ordinarias.
- SEDOP: Sistema de ecuaciones diferenciales ordinarias de prime orden.
- MPU: Métodos de paso único.
- MPM: Métodos de paso múltiple.
- MPML: Métodos de paso múltiple lineales.
- MNEDO: Métodos numéricos para ecuaciones diferenciales ordinarias.
- RK22: Runge Kutta de orden dos con dos estados.
- RK33: Runge Kutta de orden tres con tres estados.
- RK44: Runge Kutta de orden cuatro con cuatro estados.
- RK56: Runge Kutta de orden cinco con seis estados.
- RKF: Runge Kutta Fehlberg.
- AB2: Adams Bashforth de dos pasos.
- AB3: Adams Bashforth de tres pasos.
- AB4: Adams Bashforth de cuatro pasos.

Introducción

Las ecuaciones diferenciales ordinarias, EDO, son usadas frecuentemente en la descripción de diferentes fenómenos de la naturaleza desde sus orígenes en problemas de física y recientemente como herramienta indispensable en la formulación matemática de modelos biológicos, químicos, de la ingeniería y la economía. A pesar de que en la teoría se presentan algunos métodos analíticos que resuelven diferentes clases de EDO, como puede verse en [4] y [23], la cantidad de familias de EDO que pueden ser resueltas analíticamente es muy restringida, por tanto los métodos numéricos son una excelente opción para encontrar al menos una solución aproximada a problemas de EDO con valor inicial.

Históricamente uno de los primeros métodos numéricos para la solución de ecuaciones diferenciales ordinarias, MNEDO, fue el **método de Euler**. Este método fue desarrollado por Leonard Euler en 1768 y por su facilidad de deducción y comprensión, a pesar de ser un método que aproxima con poca precisión, este método hoy en día continua siendo muy usado en los primeros cursos de análisis numérico como base para comprender métodos de orden superior y en la solución numérica de modelos matemáticos básicos.

Hacia 1855 aparecen los **métodos de Adams-Bashford**, los cuales fueron creados por John C. Adams y solo hasta 1883 fueron publicados por Francis Bashford. Estos métodos resultan del uso de interpolación e integración numérica y calculan una aproximación aplicando en el método numérico el valor de por lo menos dos aproximaciones en instantes anteriores al deseado, propiedad por la cual se caracterizaron como métodos de paso múltiple. Siguiendo estas ideas, en 1926, Forest R. Moulton propuso los **métodos de Adams-Moulton** que a diferencia de los anteriores tienen la propiedad de ser implícitos¹, lo cual permitió ampliar el número de problemas a ser resueltos numéricamente.

Uno de los mayores avances en la historia de los MNEDO fue la obra de 1895 del matemático alemán Carl Runge. A diferencia de los métodos de Adams, los métodos que propone Runge son de paso único, siendo así que estos pueden usar para obtener la aproximación final, a ella misma y a la aproximación en el instante inmediatamente anterior. Los métodos definidos en ese momento fueron el **método del punto medio** y el **método del trapecio**. A partir de los aportes de Adams y de Runge, los matemáticos Karl Heun y Martin Kutta desarrollaron alrededor de 1900 los famosos **métodos de Runge-Kutta**. En particular, el aporte de Heun fue elevar el orden de estos métodos al desarrollar un método de cuarto orden, esto se traduce en obtener una aproximación con la mis-

¹Usan el valor desconocido que se desea aproximar en el método numérico, dependiendo así de la solución de una ecuación usualmente no lineal.

ma precisión a la obtenida con los otros métodos pero de forma más rápida, con tamaños de paso mayores. Seguidamente, de las manos de Nystrom y Moulton, se produjeron mayores avances en el análisis numérico con el desarrollo de métodos de Rungge-Kutta de quinto orden y los **métodos de Predictor Corrector**, lo cual puede ser visto con mayor detalle en [6].

A pesar de que en la antigüedad el uso de estos métodos conllevaba a realizar una gran cantidad de cálculos manualmente, en la actualidad se hicieron grandes avances teóricos y numéricos que se han visto impulsados con el uso de computadores. Actualmente existen diferentes software matemáticos utilizados en las ciencias aplicadas y la ingeniería, que incorporan MNEDO programados con diferentes estructuras de datos y de forma eficiente. El uso de estos programas en algunas aplicaciones no retorna el resultado esperado por el investigador, lo cual puede suceder al no tener en cuenta las propiedades particulares de los métodos numéricos implementados en el software. Por tal motivo, conocer la región de estabilidad, orden de convergencia y otras propiedades de los métodos es importante en la interpretación final de resultados y con esto garantizar desde el punto de vista del análisis numérico que las aproximaciones obtenidas son correctas. Esta situación justifica en gran medida la importancia de realizar un estudio teórico para MNEDO donde nos enfoquemos en el estudio de propiedades teóricas y computacionales.

En este trabajo se estudia la base teórica de los MNEDO clásicos y de algunos de los métodos de orden superior más usados con el fin de tener un referente teórico en el cual se discutan las propiedades más relevantes inherentes a la constitución de los MNEDO, en base a esto realizar la implementación computacional para algunos de los métodos estudiados y aplicarlos así a un modelo que describe la dinámica de infección del VIH/SIDA presentado en [17]. Este modelo consiste de un sistema no lineal de EDO que no cuenta con una solución analítica. También se realizan algunas modificaciones al modelo inicial a fin de simular numéricamente el efecto de un tratamiento antiviral en la dinámica de la enfermedad.

En el Capítulo 1 se presentan nociones básicas acerca de EDO, así como también una introducción a la solución numérica de EDO, particularmente se estudia el método de Euler. La deducción del método de Euler mediante series de Taylor sugiere que se puede tomar una mayor cantidad de términos de la serie para generar métodos de mayor precisión, es así que surgen los métodos de Taylor. Con estos métodos se puede alcanzar altos ordenes, sin embargo se podrá ver que en la práctica esto requerirá del cálculo de derivadas de alta orden lo cual resulta costoso computacionalmente. Como una mejora de los métodos de Taylor, se presenta los métodos de Runge Kutta, los cuales alcanzan la misma precisión que los métodos de Taylor y evitan el cálculo de las derivadas. Los MNEDO mencionados hasta el momento son métodos de paso único, es así que el Capítulo 2 está dedicado al estudio de las propiedades más relevantes inherentes a los métodos de paso único como son consistencia, convergencia y estabilidad. Análogamente, en el Capítulo 3 se desarrolla un estudio para los métodos de paso múltiple, algunas técnicas empleadas para deducirlos y las principales propiedades teóricas de ellos; particularmente se estudia la familia de métodos de Adams Bashford. Para finalizar, en el Capítulo 4 se valida las implementaciones realizadas en el software MATLAB, de algunos métodos. Además en este capítulo, para resaltar la importancia que cobran los métodos estudiados en la práctica, se estudia numéricamente un modelo que describe la dinámica del VIH/SIDA, el cual consiste de un sistema de EDO no lineal que no posee solución analítica.

Capítulo 1

Teoría general de ecuaciones diferenciales

A lo largo de este capítulo se presentan aspectos básicos en la teoría general de ecuaciones diferenciales, indispensables para el desarrollo de este trabajo tales como: la definición de EDO, sistemas de EDO, SEDO, el problema de Cauchy y sus propiedades básicas, condición de Lipschitz, unicidad de la solución, entre otros. Adicional a esto, se presentan dos métodos analíticos para resolver SEDO y una introducción a la solución numérica de EDO.

1.1. Ecuaciones diferenciales ordinarias

Muchos de los principios que rigen el comportamiento de diferentes fenómenos en la naturaleza relacionan las razones de cambio según las cuales un fenómeno acontece. En lenguaje matemático, las relaciones son ecuaciones y las razones de cambio tienen que ver con derivadas. Las ecuaciones que involucran razones de cambio se conocen como **Ecuaciones Diferenciales**.

Definición 1.1. Se llama **Ecuación Diferencial Ordinaria** a una ecuación que relaciona la variable independiente x , la función incógnita $y = y(x)$ y sus derivadas $y', y'', \dots, y^{(n)}$; es decir, una ecuación de la forma

$$F(x, y, y', y'', \dots, y^{(n)}) = 0, \quad (1.1.1)$$

donde F es una función real con $n + 2$ variables, $x, y, y', \dots, y^{(n)}$, y $y^{(n)} = \frac{d^n y}{dx^n}$.

El orden de una EDO se define como el orden de la derivada más alta de la ecuación, así la EDO (1.1.1) es de orden n . Si la función incógnita y depende de dos o más variables independientes, la ecuación diferencial se llama **Ecuación Diferencial Parcial**.

La solución de una EDO es una función $y = \phi(x)$ definida en un intervalo $[a, b]$, $a, b \in \mathbb{R}$, para la cual existen sus derivadas sucesivas, hasta el n -ésimo orden, tal que al hacer la sustitución $y = \phi(x)$ en la EDO, esta se convierte en una identidad con respecto a x en el intervalo $[a, b]$.

En algunos casos la solución $y = \phi(x)$ puede encontrarse usando métodos analíticos, los cuales son producto de la integración, algunos de estos métodos son: separación de variables, variación del parámetro, sustitución, factor integrante, entre otros. La descripción detallada de estos métodos se puede ver en [4, 13, 18] y [23]. Sin embargo, no siempre es posible determinar la solución analítica de una EDO usando estos métodos. Por tal motivo en la práctica se suele usar métodos numéricos, que son menos restrictivos que los métodos analíticos y permiten obtener una aproximación discreta a $y = \phi(x)$.

Dado que sin importar el orden, cualquier EDO puede ser escrita como un SEDO de primer orden, los métodos analíticos y numéricos que se estudiarán posteriormente estarán enfocados a resolver sistemas.

1.2. Sistemas de ecuaciones diferenciales ordinarias de primer orden

Definición 1.2. Un sistema de EDO de primer orden, **SEDOP**, es

$$\begin{cases} \frac{dy_1}{dx} = f_1(x, y_1, y_2, \dots, y_n) \\ \frac{dy_2}{dx} = f_2(x, y_1, y_2, \dots, y_n) \\ \vdots \\ \frac{dy_n}{dx} = f_n(x, y_1, y_2, \dots, y_n), \end{cases} \quad (1.2.1)$$

donde cada $f_i(x, y_1, y_2, \dots, y_n)$ es una función escalar, $i = 1 \dots n$.

Para reducir una EDO de orden superior a un SEDOP se realiza un cambio de variable, tal como se describe a seguir.

Consideremos la EDO de n -ésimo orden

$$y^{(n)} = f(x, y, y', y'', \dots, y^{(n-1)}). \quad (1.2.2)$$

Para reducir (1.2.2) a un SEDOP, se introducen las variables dependientes y_1, y_2, \dots, y_n , tales que

$$y_1 = y, \quad y_2 = y', \quad y_3 = y'', \quad \dots, \quad y_n = y^{(n-1)}. \quad (1.2.3)$$

Luego $y'_1 = y' = y_2$, $y'_2 = y'' = y_3$, y así sucesivamente $y'_{n-1} = y^{(n-1)} = y_n$. Por tanto, la sustitución de (1.2.3) en la ecuación (1.2.2) proporciona el SEDOP

$$\begin{cases} y'_1 = y_2 \\ y'_2 = y_3 \\ \vdots \\ y'_{n-1} = y_n \\ y'_n = f(x, y_1, y_2, \dots, y_n). \end{cases} \quad (1.2.4)$$

Ejemplo 1.1. Reescribir como un SEDOP la EDO de segundo orden

$$y''(x) + 2y'(x) - 3y(x) = 0. \quad (1.2.5)$$

Sean $y_1(x)$ y $y_2(x)$, tales que

$$y_1(x) = y(x) \text{ y } y_2(x) = y'(x),$$

entonces

$$y'_1(x) = y'(x) = y_2(x) \text{ y } y'_2(x) = y''(x). \quad (1.2.6)$$

Remplazando (1.2.6) en (1.2.5) tenemos el SEDOP

$$\begin{cases} y'_1(x) = y_2(x) \\ y'_2(x) = -2y_2(x) + 3y_1(x). \end{cases} \quad (1.2.7)$$

□

Al trabajar con SEDOP es conveniente utilizar una notación vectorial por ser más manejable y compacta. Sean

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ y } \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} f_1(x, \mathbf{y}) \\ f_2(x, \mathbf{y}) \\ \vdots \\ f_n(x, \mathbf{y}) \end{pmatrix},$$

entonces el sistema (1.2.1) se puede escribir como

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}). \quad (1.2.8)$$

Resaltamos, que se tomó como convención marcar en negrita las variables que denotan vectores.

Al igual que para una EDO, es posible determinar una solución particular para un SEDOP de n ecuaciones por medio de condiciones iniciales de la forma

$$y(x_0) = y_0^0, \quad y'(x_0) = y_0^1, \quad \dots, \quad y^{(n-1)}(x_0) = y_0^{n-1}, \quad (1.2.9)$$

donde $x_0, y_0^0, \dots, y_0^{n-1}$ son variables reales. Notemos que si en (1.2.9) hacemos el cambio de variable (1.2.3), tenemos que

$$y_1(x_0) = y_0^0, \quad y_2(x_0) = y_0^1, \quad \dots, \quad y_n(x_0) = y_0^{n-1},$$

lo cual en notación vectorial se expresará por

$$\mathbf{y}(x_0) = \mathbf{y}_0. \quad (1.2.10)$$

Esta condición es muy importante en nuestro estudio, pues un SEDOP como (1.2.8) sujeto a la condición inicial (1.2.10) define un **problema de Cauchy** y estudiar métodos para solucionar numéricamente este tipo de problemas es nuestro objetivo en este trabajo.

Definición 1.3. Un SEDOP sujeto a una condición inicial, de la forma

$$\begin{cases} \mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x)) \\ \mathbf{y}(x_0) = \mathbf{y}_0, \end{cases} \quad (1.2.11)$$

donde $\mathbf{f} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ y $x \in [a, b]$, se conoce como un **Problema de Cauchy**.

1.2.1. Existencia y unicidad

Las siguientes definiciones y teoremas nos permiten enunciar y demostrar el teorema que establece la existencia y unicidad de la solución al problema (1.2.11).

Definición 1.4. La función $f : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ se dice que satisface la **Condición de Lipschitz** en su segunda variable, si existe una constante $L > 0$, conocida como la constante de Lipschitz, tal que para algún $x \in [a, b]$ y $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, se satisface que

$$\|f(x, \mathbf{y}) - f(x, \mathbf{z})\| \leq L \|\mathbf{y} - \mathbf{z}\|,$$

donde $\|\cdot\|$ es una norma vectorial.

Definición 1.5. Una sucesión $\{x_n\}$ de un espacio métrico (E, d) se llama **sucesión de Cauchy** si para todo $\epsilon > 0$ existe un entero N tal que

$$d(x_n, x_m) < \epsilon \quad \text{siempre que } n \geq N \text{ y } m \geq N.$$

Teorema 1.1. En el espacio Euclídeo \mathbb{R}^n toda sucesión de Cauchy es convergente.

Definición 1.6. Un espacio métrico (E, d) se llama **completo** si toda sucesión de Cauchy de E converge en E .

El siguiente teorema será importante para garantizar la unicidad en el problema de Cauchy

Teorema 1.2. *Sea (E, d) un espacio métrico completo y $F : E \rightarrow E$ una contracción, es decir*

$$d(F(x), F(y)) \leq \alpha, \quad 0 \leq \alpha < 1.$$

Entonces existe un único punto fijo $p \in E$ para F , es decir $F(p) = p$.

Las demostraciones de los anteriores teoremas se pueden encontrar en [1]. A continuación se enuncia el teorema de existencia y unicidad para el problema de Cauchy, la demostración que se presenta de este teorema fue tomada de [7] y por su importancia se extiende con más detalle aquí.

Teorema 1.3. *Si en el problema de Cauchy (1.2.11), $f : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ es continua en su primera variable y satisface la condición de Lipschitz en su segunda variable con constante L , entonces existe una única solución para este problema.*

Demostración. Sea E el conjunto de todas las funciones continuas $\mathbf{y} : [a, b] \rightarrow \mathbb{R}^n$, tal que $\mathbf{y}(a) = \mathbf{y}_0$. Dadas las funciones $\mathbf{y}, \mathbf{z} \in E$, se define la métrica

$$d(\mathbf{y}, \mathbf{z}) = \sup_{x \in [a, b]} \exp(-K(x - a)) \|\mathbf{y}(x) - \mathbf{z}(x)\|, \quad (1.2.12)$$

donde $K > L$. El conjunto E junto con la métrica d definen un espacio métrico completo. Sea $\phi(\mathbf{y})$ la solución exacta del problema de Cauchy (1.2.11), la cual se obtiene de aplicar el teorema fundamental del cálculo, es decir

$$\phi(\mathbf{y}) = \mathbf{y}_0 + \int_a^x \mathbf{f}(s, \mathbf{y}(s)) ds.$$

Esta función es una contracción, dado que para $\mathbf{y}, \mathbf{z} \in E$, se tiene

$$d(\phi(\mathbf{y}), \phi(\mathbf{z})) = \sup_{x \in [a, b]} \exp(-K(x - a)) \left\| \int_a^x (\mathbf{f}(s, \mathbf{y}(s)) - \mathbf{f}(s, \mathbf{z}(s))) ds \right\|.$$

Aplicando la desigualdad triangular se obtiene

$$d(\phi(\mathbf{y}), \phi(\mathbf{z})) \leq \sup_{x \in [a, b]} \exp(-K(x - a)) \int_a^x \|\mathbf{f}(s, \mathbf{y}(s)) - \mathbf{f}(s, \mathbf{z}(s))\| ds.$$

Por hipótesis \mathbf{f} es satisface la condición de Lipschitz en su segunda variable, por lo tanto

$$d(\phi(\mathbf{y}), \phi(\mathbf{z})) \leq L \sup_{x \in [a, b]} \exp(-K(x - a)) \int_a^x \|\mathbf{y}(s) - \mathbf{z}(s)\| ds. \quad (1.2.13)$$

Aplicando la métrica (1.2.12) en la expresión (1.2.13), se tiene

$$d(\phi(\mathbf{y}), \phi(\mathbf{z})) \leq L d(\mathbf{y}, \mathbf{z}) \sup_{x \in [a, b]} \exp(-K(x - a)) \int_a^x \exp(K(s - a)) ds,$$

de donde

$$d(\phi(\mathbf{y}), \phi(\mathbf{z})) \leq \frac{L}{K} d(\mathbf{y}, \mathbf{z}). \quad (1.2.14)$$

Dado que $K > L$, de la expresión (1.2.14), se concluye que ϕ es una contracción. Por lo tanto, por el Teorema 1.2, existe una única función $\mathbf{y} \in E$ que satisface que $\phi(\mathbf{y}) = \mathbf{y}$, y en consecuencia es la única solución al problema de Cauchy. \square

A continuación estudiaremos características generales de la solución de SEDOP lineales y presentaremos dos métodos analíticos para resolver sistemas lineales de coeficientes constantes descritos detalladamente en [11, 21] y [23]. Los métodos estudiados permitirán observar como la dificultad para obtener una solución analítica aumenta conforme crece la dimensión del sistema. Esta situación justifica en gran medida la importancia de usar métodos numéricos.

1.2.2. Sistema lineal de EDO de primer orden

Si cada una de las funciones f_1, \dots, f_n en el sistema (1.2.1) son lineales, entonces se dice que el SEDOP es lineal. La forma general de un SEDOP lineal es

$$\begin{cases} y'_1 = a_{11}(x)y_1 + a_{12}(x)y_2 + \dots + a_{1n}(x)y_n + b_1(x) \\ y'_2 = a_{21}(x)y_1 + a_{22}(x)y_2 + \dots + a_{2n}(x)y_n + b_2(x) \\ \vdots \\ y'_n = a_{n1}(x)y_1 + a_{n2}(x)y_2 + \dots + a_{nn}(x)y_n + b_n(x), \end{cases} \quad (1.2.15)$$

donde las funciones a_{ij} y b_i , con $i, j = 1 \dots n$, son reales y dependientes únicamente de la variable independiente x . Si cada una de las funciones b_i , $i = 1 \dots n$, son idénticamente cero, entonces se dice que el sistema es **homogéneo** y en caso contrario es **no homogéneo**. Los sistemas lineales son los más simples entre todos los sistemas de primer orden, pero incluso estos son difíciles de resolver analíticamente.

Usaremos notación matricial para el sistema (1.2.15) a fin de simplificar cálculos y destacar las propiedades matriciales de estos sistemas.

Sean

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{A}(x) = \begin{pmatrix} a_{11}(x) & a_{12}(x) & \dots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \dots & a_{2n}(x) \\ \vdots & \vdots & \dots & \vdots \\ a_{n1}(x) & a_{n2}(x) & \dots & a_{nn}(x) \end{pmatrix} \quad \text{y} \quad \mathbf{b}(x) = \begin{pmatrix} b_1(x) \\ b_2(x) \\ \vdots \\ b_n(x) \end{pmatrix},$$

entonces el sistema (1.2.15) se puede escribir como

$$\mathbf{y}'(x) = \mathbf{A}(x)\mathbf{y}(x) + \mathbf{b}(x). \quad (1.2.16)$$

Cuando el sistema es homogéneo $b(x)=0$ y en este caso se tiene

$$\mathbf{y}'(x) = \mathbf{A}(x)\mathbf{y}(x). \quad (1.2.17)$$

La solución de una EDO de la forma $y' = \lambda y$ con $\lambda \in \mathbb{R}$ es $y = ce^{\lambda x}$ con c constante. Sin embargo, aunque al comparar esta EDO con la que se presenta en la ecuación (1.2.17) se observa una similitud se debe recordar que \mathbf{A} es una matriz y por tanto se debe tener un cuidado especial en la solución de este sistema.

El siguiente teorema conoce como **principio de superposición** para sistemas lineales, este resultado será usado para encontrar una solución general de (1.2.17).

Teorema 1.4. *Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$ un conjunto de vectores solución del sistema homogéneo (1.2.17) en un intervalo $[a, b]$. Entonces la combinación lineal*

$$\mathbf{y} = c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \dots + c_k\mathbf{y}_k,$$

donde $c_i, i = 1 \dots k$, son constantes escalares arbitrarias, es también una solución en el intervalo.

Definición 1.7. Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$ un conjunto de vectores solución del sistema homogéneo (1.2.17) en un intervalo $[a, b]$. Se dice que el conjunto es **linealmente dependiente** en el intervalo si existen constantes c_1, c_2, \dots, c_k , no todas cero, tales que

$$c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \dots + c_k\mathbf{y}_k = 0,$$

para todo x en el intervalo. Si el conjunto de vectores no es linealmente dependiente en el intervalo, se dice que es **linealmente independiente**.

Para determinar si un conjunto de vectores es linealmente dependiente o independiente, se puede usar la Definición 1.7. Sin embargo en la práctica el **Wronskiano** define un criterio que permite determinar fácilmente la dependencia lineal de un conjunto de vectores.

Definición 1.8. Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$, un conjunto de vectores en \mathbb{R}^n definidos como en (1.7). Entonces el **Wronskiano** de este conjunto es el determinante de la matriz obtenida a partir de estos vectores y se denota por

$$W(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = \det(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n).$$

Es importante señalar que la Definición 1.8 de Wronskiano tomada de [23], difiere de la definición habitual, pues esta definición no involucra derivación, pero dado que la idea es aplicarlo como criterio para verificar independencia lineal en funciones vectoriales, se acostumbra mantener el mismo nombre.

Teorema 1.5. Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$ un conjunto de vectores solución del sistema homogéneo (1.2.17) en un intervalo $[a, b]$. Este conjunto es linealmente independiente si y sólo si

$$W(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) \neq 0.$$

La importancia de establecer la independencia lineal de un conjunto de vectores está en que al encontrar un conjunto de n vectores solución de (1.2.17) linealmente independientes, se puede encontrar una solución general para el sistema lineal homogéneo (1.2.17) como se muestra a seguir.

Definición 1.9. Cualquier conjunto $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ de n vectores solución linealmente independientes del sistema homogéneo (1.2.17) en un intervalo $[a, b]$ es un **conjunto fundamental de soluciones** en el intervalo.

Teorema 1.6. Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ un conjunto fundamental de soluciones del sistema homogéneo (1.2.17) en un intervalo $[a, b]$. Entonces la **solución general** del sistema (1.2.16) en el intervalo $[a, b]$ es

$$\mathbf{y} = c_1 \mathbf{y}_1 + c_2 \mathbf{y}_2 + \dots + c_n \mathbf{y}_n,$$

donde c_i , $i = 1 \dots, n$, son constantes escalares arbitrarias.

1.2.3. Solución de SEDOP con coeficientes constantes

Los teoremas y definiciones que se presentan a continuación permiten obtener un método para encontrar la solución analítica para sistemas lineales con coeficientes constantes; es decir, aquellos en los que las funciones a_{ij} , $i, j = 1 \dots n$, del sistema (1.2.15) son funciones constantes.

Definición 1.10. Sea \mathbf{A} una matriz cuadrada con $x \times n$ elementos reales. El número $\lambda_i \in R$ es un **valor propio** de \mathbf{A} , si es solución de la ecuación

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = 0, \quad (1.2.18)$$

esta expresión se conoce como **ecuación característica** de la matriz \mathbf{A} .

Un **vector propio** asociado con el valor propio λ_i es un vector no cero \mathbf{v}_i tal que $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, es decir que

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v}_i = 0. \quad (1.2.19)$$

Teorema 1.7. Cualquier conjunto de k vectores propios $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ de \mathbf{A} con valores propios diferentes $\lambda_1, \lambda_2, \dots, \lambda_k$, respectivamente, es linealmente independiente.

Teorema 1.8. Sean $\lambda_1, \lambda_2, \dots, \lambda_n$ n valores propios reales y distintos de la matriz de coeficientes \mathbf{A} del SEDOP homogéneo (1.2.17) y sean $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ los vectores propios correspondientes asociados

a ellos. Entonces la **solución general** del SEDOP homogéneo (1.2.17), en el intervalo $(-\infty, \infty)$ es

$$\mathbf{y} = c_1 \mathbf{v}_1 e^{\lambda_1 x} + c_2 \mathbf{v}_2 e^{\lambda_2 x} + \dots + c_n \mathbf{v}_n e^{\lambda_n x},$$

donde $c_i, i = 1, \dots, n$ son constantes escalares arbitrarias.

Observación 1.1. Los n valores propios de la matriz de coeficientes \mathbf{A} no necesariamente son todos distintos, es decir algunos de los valores propios pueden repetirse. En general, si m es un entero positivo y $(\lambda - \lambda_j)^m$ es un factor de la ecuación característica, mientras que $(\lambda - \lambda_j)^{m+1}$ no es un factor, entonces se dice que λ_j es un valor propio de multiplicidad m . Se tienen los siguientes casos:

1. Para algunas matrices \mathbf{A} de $n \times n$, relacionadas con el SEDOP homogéneo (1.2.17), es posible encontrar m vectores propios linealmente independientes $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$, correspondientes a un valor propio λ_j de multiplicidad $m \leq n$. En este caso la solución general del sistema contiene la combinación lineal

$$c_1 \mathbf{v}_1 e^{\lambda_j x} + c_2 \mathbf{v}_2 e^{\lambda_j x} + \dots + c_n \mathbf{v}_m e^{\lambda_j x}.$$

2. Si solo hay un vector propio \mathbf{v}_j que corresponde al valor propio λ_j de multiplicidad m , se pueden encontrar m soluciones linealmente independientes de la forma

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{v}_{11} e^{\lambda_j x} \\ \mathbf{y}_2 &= \mathbf{v}_{21} x e^{\lambda_j x} + \mathbf{v}_{22} e^{\lambda_j x} \\ &\vdots \\ \mathbf{y}_m &= \mathbf{v}_{m1} \frac{x^{m-1}}{(m-1)!} e^{\lambda_j x} + \mathbf{v}_{m2} \frac{x^{m-2}}{(m-2)!} e^{\lambda_j x} + \dots + \mathbf{v}_{mm} e^{\lambda_j x},\end{aligned}$$

donde los \mathbf{v}_{ij} son vectores.

Además de lo mencionado arriba, puede suceder el caso en que los valores propios sean complejos, tal como sucede en la solución de EDO de orden superior con coeficientes constantes. Esto puede ser visto con mayor detalle en [23].

Ejemplo 1.2. Encontrar la solución general del SEDOP homogéneo

$$\begin{cases} y'_1(x) = y_1(x) - y_2(x) \\ y'_2(x) = -4y_1(x) + y_2(x) \\ y_1(0) = 0, \quad y_2(0) = 1. \end{cases} \quad (1.2.20)$$

La matriz de coeficientes del SEDOP homogéneo es

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix}.$$

Para encontrar la solución general del sistema se deben encontrar los valores y vectores propios de esta matriz. Así, dado que los valores propios vienen dados por la expresión (1.2.18) se tiene

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 1 - \lambda & -1 \\ -4 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1) = 0.$$

Por lo tanto, los valores propios de \mathbf{A} son $\lambda = 3$ y $\lambda = -1$.

Ahora se debe encontrar los vectores propios asociados a estos valores propios.

Para $\lambda = 3$ el vector propio asociado debe satisfacer que $|\mathbf{A} - 3\mathbf{I}|\mathbf{v}_1 = 0$. De donde al solucionar el sistema

$$\begin{bmatrix} -2 & -1 \\ -4 & -2 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

se obtiene $m_1=1$ y $m_2=-2$. Por lo tanto

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

De forma análoga se determina el vector propio \mathbf{v}_2 asociado al valor propio $\lambda = -1$, donde

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

De esta forma se obtienen dos soluciones particulares linealmente independientes

$$\mathbf{y}_1(x) = e^{3x} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{y} \quad \mathbf{y}_2(x) = e^{-x} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Así la solución general del sistema es

$$\mathbf{y}(x) = c_1 e^{3x} \begin{bmatrix} 1 \\ -2 \end{bmatrix} + c_2 e^{-x} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} c_1 e^{3x} + c_2 e^{-x} \\ -2c_1 e^{3x} + 2c_2 e^{-x} \end{bmatrix}. \quad (1.2.21)$$

Finalmente, usando las condiciones iniciales para $x = 0$ tenemos que

$$\mathbf{y}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 e^0 + c_2 e^0 \\ -2c_1 e^0 + 2c_2 e^0 \end{bmatrix},$$

de donde $c_1 = -\frac{1}{4}$ y $c_2 = \frac{1}{4}$. Por lo tanto remplazando el valor de estas constantes en (1.2.21) tenemos la solución del problema de Cauchy (1.2.20)

$$\mathbf{y}(x) = \frac{1}{4} \begin{bmatrix} -e^{3x} + e^{-x} \\ 2e^{3x} + 2e^{-x} \end{bmatrix}.$$

□

1.2.4. Método matriz exponencial

Una importante aplicación de la exponencial de una matriz es encontrar soluciones a un SEDO lineales de coeficientes constantes. Mostraremos como la forma de la solución de una EDO de primer orden con coeficientes constantes tiene características similares a la de un SEDOP con coeficientes constantes.

Para una EDO de la forma $y'(x) = \lambda y(x)$, la solución general es $y(x) = ce^{\lambda x}$. Podría suponerse que para el SEDO

$$\mathbf{y}'(x) = \mathbf{A}\mathbf{y}(x), \quad (1.2.22)$$

la solución general tiene la forma $\mathbf{y} = e^{\mathbf{Ax}}\mathbf{c}$, donde \mathbf{A} es una matriz constante de $n \times n$ y \mathbf{c} es un vector columna de n componentes. Por tal motivo, es importante definir la matriz exponencial $e^{\mathbf{Ax}}$, en la cual usaremos el desarrollo de la función exponencial por series.

Definición 1.11. Sea \mathbf{A} una matriz $n \times n$ con elementos reales o complejos. Entonces la matriz $e^{\mathbf{Ax}}$ es una matriz de $n \times n$ elementos que se conoce como **Matriz Exponencial**, y está definida por

$$e^{\mathbf{Ax}} = \mathbf{I} + \mathbf{Ax} + \mathbf{A}^2 \frac{x^2}{2!} + \dots + \mathbf{A}^k \frac{x^k}{k!} + \dots = \sum_{k=0}^{\infty} \mathbf{A}^k \frac{x^k}{k!}.$$

Teorema 1.9. Para cualquier vector constante \mathbf{c} , $\mathbf{y}(x) = e^{\mathbf{Ax}}\mathbf{c}$ es una solución al SEDOP (1.2.22). Más aún, la solución dada por $\mathbf{y}(x) = e^{\mathbf{Ax}}\mathbf{y}_0$ satisface $\mathbf{y}(0) = \mathbf{y}_0$.

Demostración. Si $\mathbf{y}(x) = e^{\mathbf{Ax}}\mathbf{c}$ es solución del SEDO (1.2.22), al sustituirlo en este tendremos una identidad. Primero notamos que dado que $e^{\mathbf{Ax}}$ es una matriz el producto por un vector debe ser a su derecha y así $e^{\mathbf{Ax}}\mathbf{c}$ está bien definido. Ahora, usado de matriz exponencial tenemos

$$\mathbf{y}(x) = e^{\mathbf{Ax}}\mathbf{c} = \left[\mathbf{I} + \mathbf{Ax} + \mathbf{A}^2 \frac{x^2}{2!} + \dots + \mathbf{A}^k \frac{x^k}{k!} + \dots \right] \mathbf{c}. \quad (1.2.23)$$

Dado que las entradas \mathbf{A} son constantes, se puede verificar que

$$\frac{d}{dx} \mathbf{A}^k \frac{x^k}{k!} = \mathbf{A}^k \frac{d}{dx} \frac{x^k}{k!} = \mathbf{A}^k \frac{kx^{k-1}}{k!} = \frac{\mathbf{A}^k x^{k-1}}{(k-1)!} = \mathbf{A} \left[\frac{\mathbf{A}^{k-1} x^{k-1}}{(k-1)!} \right]. \quad (1.2.24)$$

Usando las ecuaciones (1.2.23) y (1.2.24), calculamos $\mathbf{y}'(x)$. De donde,

$$\mathbf{y}'(x) = \frac{d}{dx} [e^{\mathbf{Ax}}\mathbf{c}] = \mathbf{A} \left[\mathbf{I} + \mathbf{Ax} + \mathbf{A}^2 \frac{x^2}{2!} + \dots + \mathbf{A}^k \frac{x^k}{k!} + \dots \right] \mathbf{c} = \mathbf{A} e^{\mathbf{Ax}} \mathbf{c} = \mathbf{A} \mathbf{y}(x).$$

Por lo tanto, $\mathbf{y}(x) = e^{\mathbf{Ax}}\mathbf{c}$ es solución de (1.2.22).

Finalmente, como $e^{\mathbf{A}0} = e^0 = \mathbf{I}$, se tiene

$$\mathbf{y}(0) = e^{\mathbf{A}0} \mathbf{y}_0 = \mathbf{I} \mathbf{y}_0 = \mathbf{y}_0.$$

□

Usualmente la matriz exponencial se conoce como la matriz solución principal del sistema o matriz fundamental, pues el cálculo de esta matriz resuelve el sistema. A pesar de que la definición ofrece una forma para calcular esta matriz, en la práctica se aplican técnicas alternativas más simples para calcular $e^{\mathbf{A}x}$, como son la transformada de Laplace, la forma canónica de Jordan, entre otras descritas en [21] y [23].

El caso más trivial se presenta cuando \mathbf{A} es una matriz diagonal, pues para una matriz diagonal

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

la matriz exponencial es

$$e^{\mathbf{D}x} = \begin{bmatrix} e^{\lambda_1 x} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 x} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n x} \end{bmatrix}.$$

La simplicidad de calcular la exponencial para una matriz diagonalizable será aprovechada usando la forma canónica de Jordan para el cálculo de $e^{\mathbf{A}x}$.

Teorema 1.10. *Sea \mathbf{J} la forma canónica de Jordan de una matriz \mathbf{A} , $\mathbf{J} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$. Entonces $\mathbf{A} = \mathbf{C}\mathbf{J}\mathbf{C}^{-1}$ y*

$$e^{\mathbf{A}x} = \mathbf{C}e^{\mathbf{J}x}\mathbf{C}^{-1}.$$

Demostración. Observemos que

$$\begin{aligned} \mathbf{A}^n &= (\mathbf{C}\mathbf{J}\mathbf{C}^{-1})^n = (\mathbf{C}\mathbf{J}\mathbf{C}^{-1})(\mathbf{C}\mathbf{J}\mathbf{C}^{-1})\dots(\mathbf{C}\mathbf{J}\mathbf{C}^{-1}) \\ &= \mathbf{C}\mathbf{J}(\mathbf{C}^{-1}\mathbf{C})\mathbf{J}(\mathbf{C}^{-1}\mathbf{C})\mathbf{J}(\mathbf{C}^{-1}\mathbf{C})\dots(\mathbf{C}^{-1}\mathbf{C})\mathbf{J}\mathbf{C}^{-1} \\ &= \mathbf{C}\mathbf{J}^n\mathbf{C}^{-1}. \end{aligned}$$

Así

$$\begin{aligned} e^{\mathbf{A}x} &= \mathbf{I} + (\mathbf{A}x) + \frac{(\mathbf{A}x)^2}{2!} + \dots = \mathbf{C}\mathbf{I}\mathbf{C}^{-1} + \mathbf{C}(\mathbf{J}x)\mathbf{C}^{-1} + \mathbf{C}\frac{(\mathbf{J}x)^2}{2!}\mathbf{C}^{-1} + \dots \\ &= \mathbf{C} \left[\mathbf{I} + (\mathbf{J}x) + \frac{(\mathbf{J}x)^2}{2!} + \dots \right] \mathbf{C}^{-1} = \mathbf{C}e^{\mathbf{J}x}\mathbf{C}^{-1}. \end{aligned}$$

□

El Teorema 1.10 dice que para calcular $e^{\mathbf{A}x}$ en realidad solo se necesita calcular $e^{\mathbf{J}x}$. Calcular $e^{\mathbf{J}x}$ es más sencillo, pues la matriz \mathbf{J} es diagonal o diagonal por bloques de Jordan.

Ejemplo 1.3. Resolver el problema (1.2.20) usando la matriz exponencial.

En el Ejemplo 1.2 se obtuvo que los valores propios de la matriz de coeficientes \mathbf{A} son $\lambda_1 = 3$ y $\lambda_2 = -1$ con vectores propios asociados $\mathbf{v}_1 = [1, -2]^T$ y $\mathbf{v}_2 = [1, 2]^T$. De aquí se tiene

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ -2 & 2 \end{bmatrix}, \quad \mathbf{C}^{-1} = \frac{1}{4} \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, \quad e^{\mathbf{J}x} = \begin{bmatrix} e^{3x} & 0 \\ 0 & e^{-x} \end{bmatrix}. \quad (1.2.25)$$

Por lo tanto

$$\begin{aligned} e^{\mathbf{A}x} &= \mathbf{C}e^{\mathbf{J}x}\mathbf{C}^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} e^{3x} & 0 \\ 0 & e^{-x} \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} 2e^{3x} + 2e^{-x} & -e^{-x} + e^{-x} \\ -4e^{3x} + 4e^{-x} & 2e^{3x} + 2e^{-x} \end{bmatrix}. \end{aligned}$$

Usando las condiciones iniciales en $x = 0$, tal como se obtuvo en el Ejemplo 1.2, la solución del problema de Cauchy es

$$\begin{aligned} \mathbf{y}(x) &= e^{\mathbf{A}x}\mathbf{y}_0 = \frac{1}{4} \begin{bmatrix} 2e^{3x} + 2e^{-x} & -e^{-x} + e^{-x} \\ -4e^{3x} + 4e^{-x} & 2e^{3x} + 2e^{-x} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} -e^{3x} + e^{-x} \\ 2e^{3x} + 2e^{-x} \end{bmatrix}. \end{aligned}$$

□

1.3. Introducción a la solución numérica de EDO

Dado que los métodos para la solución de SEDO presentados son para sistemas con coeficientes constantes, lo cual no sucede en la mayoría de modelos matemáticos y además, el problema de encontrar valores propios aumenta con el orden de la matriz. Se evidencia la necesidad de encontrar aproximaciones para el problema de Cauchy. En esta sección se presenta una introducción a la solución numérica de EDO. Primero presentaremos métodos numéricos para solucionar el problema de Cauchy en los reales y luego en \mathbb{R}^n . Estos métodos se presentarán de una forma más profunda en los siguientes capítulos.

Los métodos numéricos que estudiamos son métodos iterativos aritméticos que aproximan la solución al problema de Cauchy

$$\begin{cases} y'(x) = f(x, y(x)) \\ y(x_0) = y_0. \end{cases} \quad (1.3.1)$$

Con estos métodos no se obtiene una aproximación continua a la solución $y(x)$, se obtienen aproximaciones en un conjunto discreto de puntos $a = x_0, x_1, \dots, x_n = b$ del intervalo $[a, b]$, este proceso

se conoce como **Discretización**. Las discretizaciones numéricas se aplican a menudo en la aproximación de soluciones de ecuaciones diferenciales en las que no es posible encontrar una solución analítica. Con estas discretizaciones se espera que la aproximación encontrada sea lo más cercana posible a la solución exacta y a un bajo costo computacional.

En esta sección estudiaremos el método de Euler, este método es el más básico de las técnicas de aproximación para resolver problemas de Cauchy, por la simplicidad de la deducción de este método puede ser utilizado para ilustrar las técnicas involucradas en la construcción de algunos de los métodos más complejos.

1.3.1. Método de Euler

El método de Euler, también llamado **método de la recta recta tangente**, es el método más simple y más antiguo de los métodos usados para aproximar soluciones al problema de Cauchy. Fue creado por Leonhard Euler y fue publicado en su obra de cálculo integral entre los años 1768 y 1770, tal y como se presenta en [5]. Existen diferentes formas para deducir el método de Euler, para aproximar la solución al problema de Cauchy (1.3.1), entre las cuales están: series de Taylor, la idea geométrica de aproximación por rectas tangentes, aproximaciones a una integral y la aplicación del teorema de valor medio. En esta sección se presentará la deducción mediante series de Taylor, la aproximación de una integral y geométricamente mediante rectas tangentes.

- Deducción del método de Euler mediante series de Taylor.

En primer lugar consideramos un intervalo $[a, b]$, donde $a = x_0$ es el valor inicial para x en el cual se conoce el valor de y_0 por la condición inicial y $b = x_f$ es el extremo final del intervalo, en el cual se quiere aproximar el valor de $y(x_f)$ desconocido. Dividimos dicho intervalo en n subintervalos de longitud h , por lo tanto $h = (b - a)/n$, de lo cual se obtiene el conjunto discreto de $n+1$ puntos $\{x_0, x_1, x_2, \dots, x_n\}$, llamados **puntos de malla**, donde $x_{k+1} = x_k + h$ para $0 \leq k \leq n - 1$, es decir que la distancia entre cada par de puntos sucesivos es h , esta distancia se conoce como **tamaño de paso**.

Supongamos que la solución exacta de (1.3.1), $y(x)$, tiene dos derivadas continuas sobre $[a, b]$, de modo que para cada $0 \leq k \leq n - 1$. La expansión de Taylor en torno a x_k es

$$y(x_{k+1}) = y(x_k) + (x_{k+1} - x_k)y'(x_k) + \frac{(x_{k+1} - x_k)^2}{2!}y''(\varsigma_k),$$

para algún número $\varsigma_k \in (x_k, x_{k+1})$. Como $h = x_{k+1} - x_k$, tenemos

$$y(x_{k+1}) = y(x_k) + hy'(x_k) + \frac{h^2}{2!}y''(\varsigma_k). \quad (1.3.2)$$

Eliminando el último término de (1.3.2), se deduce el método de Euler

$$\begin{cases} y_0 = y(x_0) \\ y_{k+1} = y_k + hf(x_k, y_k), \end{cases} \quad (1.3.3)$$

donde $x_{k+1} = x_k + h$, $0 \leq k \leq n - 1$, $h = \frac{b-a}{n}$, $f : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$.

El término que se ha truncado en la serie de Taylor para deducir el método de Euler está relacionado con el **error Local de discretización**, α_k , que es el error cometido en el cálculo $y(x_{k+1})$ entre x_k y x_{k+1} . Es decir que para el método de Euler

$$\alpha_{k+1} = \frac{1}{2}hy''(\varsigma_k),$$

para ς_k en el intervalo (x_k, x_{k+1}) .

Si M es una cota superior para $y''(x)$, con $x \in [a, b]$, entonces el error local de discretización satisface que $\alpha_k \leq Mh/2$ y se dice que el método de Euler es de orden uno. En el siguiente capítulo se presentarán métodos de mayor orden, los **Métodos de Taylor**, los cuales se obtienen truncando la serie de Taylor con mayor número de términos.

- Deducción geométrica del método de Euler.

Geométricamente, el método de Euler calcula la aproximación y_{k+1} por medio de la recta tangente a la curva que define la solución exacta $y(x)$. La ecuación de la recta tangente a la curva de $y(x)$ en el punto (x_0, y_0) está dada por

$$r(x) = y'(x_0)(x - x_0) + y_0 = (x - x_0)f(x_0, y_0) + y_0.$$

Tomando $x = x_0 + h = x_1$, se tiene que $r(x_0 + h) = y_0 + hf(x_0, y_0) = y_1$, es decir que se encontró la aproximación y_1 a partir del valor de y_0 . En general, si se supone conocido el valor de y_k , el valor de y_{k+1} puede encontrarse usando la expresión $y_{k+1} = y_k + hf(x_k, y_k)$, que es la fórmula de integración del método de Euler (1.3.3). La Figura 1.1 ilustra esta interpretación.

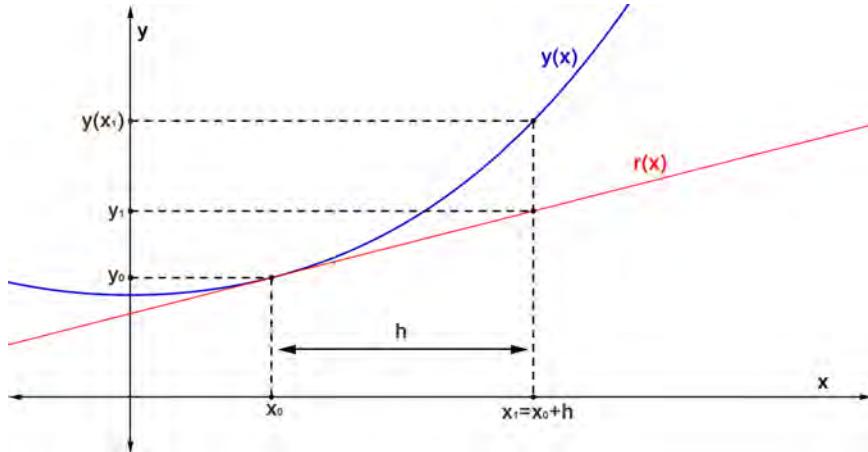


Figura 1.1: Interpretación geométrica método de Euler.

- Deducción integral del método de Euler.

Integrando cada lado de (1.3.1) entre dos puntos de malla consecutivos, por ejemplo entre x_k y x_{k+1} , se obtiene la expresión

$$y(x_{k+1}) = y(x_k) + \int_{x_k}^{x_{k+1}} f(x, y(x)) dx.$$

Si la función $f(x, y(x))$ se considera constante en el intervalo $[x_k, x_{k+1}]$, puede salir de la integral y se obtiene

$$y(x_{k+1}) = y(x_k) + f(x_k, y(x_k)) \int_{x_k}^{x_{k+1}} dx = y(x_k) + f(x_k, y(x_k))(x_{k+1} - x_k),$$

de lo cual se deduce el método de Euler $y_{k+1} = y_k + hf(x_k, y_k)$.

Más adelante usaremos este mismo enfoque para obtener métodos de mayor precisión, dando mejores aproximaciones a la integral de $f(x, y(x))$.

Ejemplo 1.4. Usar el método de Euler con $h = 0.02$ para aproximar el valor de $y(0.1)$, si

$$\begin{cases} y'(x) = -4y + e^{-2x} \\ y(0) = 1, \end{cases} \quad (1.3.4)$$

con solución exacta $y(x) = \frac{1}{2}(e^{-4x} + e^{-2x})$ y $y(0.1) = 0.74452$.

Usando la fórmula de integración del método de Euler (1.3.3) y la condición inicial $y(0) = 1$ se tiene el siguiente proceso iterativo

$$\begin{aligned} x_0 &= 0.00, & y_0 &= 1 \\ x_1 &= 0.02, & y_1 &= y_0 + h(-4y_0 + e^{-2x_0}) = 1 + 0.02(-4(1) + e^{-2(0.00)}) = 0.94 \\ x_2 &= 0.04, & y_2 &= y_1 + h(-4y_1 + e^{-2x_1}) = 0.94 + 0.02(-4(0.94) + e^{-2(0.02)}) = 0.88401 \\ x_3 &= 0.06, & y_3 &= y_2 + h(-4y_2 + e^{-2x_2}) = 0.88401 + 0.02(-4(0.88401) + e^{-2(0.04)}) = 0.83175. \end{aligned}$$

Continuamos con este proceso hasta llegar al extremo final del intervalo de integración y se obtiene la aproximación $y(0.1) \approx 0.73736$.

Para comparar las aproximaciones obtenidas con el método de Euler, dado que se conoce la solución analítica de este problema, usamos el **error global de discretización** $e_k = |y(x_k) - y_k|$, que es el valor absoluto de la diferencia entre la solución exacta y la solución aproximada del problema en x_k . Los resultados obtenidos para este problema se resumen en la Tabla 1.1, donde se observa que a medida que x_k se aleja de la condición inicial, el error global aumenta. \square

1.3.2. Método de Euler para sistemas de EDO

Los métodos numéricos para resolver problemas de Cauchy que involucran una sola ecuación diferencial de primer orden y una condición inicial escalar pueden ser generalizados a fin de resolver SEDOP. La idea básica es reescribir el sistema de ecuaciones diferenciales como el problema de

x_k	y_k	$y(x_k)$	e_k
0.00	1.00000	1.00000	0
0.02	0.94000	0.94195	1.952×10^{-3}
0.04	0.88401	0.88763	3.617×10^{-3}
0.06	0.83175	0.83677	5.013×10^{-3}
0.08	0.78295	0.78914	6.198×10^{-3}
0.10	0.73736	0.74452	7.164×10^{-3}

Tabla 1.1: Discretización obtenida por el método de Euler en la solución numérica de (1.3.4).

Cauchy (1.3.1) y aplicar el método teniendo en cuenta que las funciones involucradas son vectoriales.

El método de Euler para resolver SEDOP es

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h \mathbf{f}(x_k, \mathbf{y}_k), \end{cases} \quad (1.3.5)$$

donde $x_{k+1} = x_k + h$, $0 \leq k \leq n - 1$, $h = \frac{b-a}{n}$, $\mathbf{f} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Ejemplo 1.5. Usar el método de Euler para aproximar el valor de $u(1)$ y $v(1)$, si

$$\begin{cases} u'(x) = v(x) \\ v'(x) = 4v(x) - 3u(x) \\ u(0) = 1 \quad v(0) = 0, \end{cases} \quad (1.3.6)$$

con solución exacta $u(x) = \frac{1}{2}(3e^x - e^{3x})$, $v(x) = \frac{1}{2}(3e^x - 3e^{3x})$ y $u(1) = -5.96534$, $v(1) = -26.05088$.

Si hacemos $\mathbf{y} = [u(x) \ v(x)]^T$, $\mathbf{f}(x, \mathbf{y}) = [v(x) \ 4v(x) - 3u(x)]^T$ y $\mathbf{y}_0 = [u(0) \ v(0)]^T$, el sistema (1.3.6) puede escribirse como

$$\begin{cases} \mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x)) \\ \mathbf{y}(\mathbf{x}_0) = \mathbf{y}_0, \end{cases} \quad (1.3.7)$$

donde $\mathbf{f} : [0, 1] \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ y $x \in [0, 1]$. Es decir que describe un problema de Cauchy como en (1.2.11).

Aplicando el método de Euler (1.3.5) a este problema con $h = 1/16$ se tiene

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{y}_0 + h\mathbf{f}(x, \mathbf{y}_0) = \begin{bmatrix} u(x_0) \\ v(x_0) \end{bmatrix} + (1/16) \begin{bmatrix} v(x_0) \\ 4v(x_0) - 3u(x_0) \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (0.0625) \begin{bmatrix} 0 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.1875 \end{bmatrix} \approx \mathbf{y}(0.0625), \\ \mathbf{y}_2 &= \mathbf{y}_1 + h\mathbf{f}(x, \mathbf{y}_1) = \begin{bmatrix} u(x_1) \\ v(x_1) \end{bmatrix} + (1/16) \begin{bmatrix} v(x_1) \\ 4v(x_1) - 3u(x_1) \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -0.1875 \end{bmatrix} + (0.0625) \begin{bmatrix} -0.1875 \\ -3.75 \end{bmatrix} = \begin{bmatrix} 0.98828 \\ -0.42187 \end{bmatrix} \approx \mathbf{y}(0.125).\end{aligned}$$

Este proceso se repite hasta llegar al extremo final del intervalo de integración. Los resultados de las aproximaciones numéricas a $y(1)$ con diferentes tamaños de paso se resumen en la Tabla 1.2, donde se observa que al reducir el tamaño de paso h a la mitad, el error también se reduce a la mitad, por lo tanto el error es proporcional al tamaño de paso h .

Tamaño de paso h	1/16	1/32	1/64	1/128
Aproximación $u(1)$	-3.86132	-4.78178	-5.33472	-5.63946
Aproximación $v(1)$	-19.49776	-22.37633	-24.09632	-25.04160
Error global	6.88261	3.86046	2.05376	1.06059

Tabla 1.2: Discretización obtenida por el método de Euler en la solución numérica de 1.3.4 para diferentes tamaños de paso.

□

Tanto los métodos de paso único como los métodos de paso múltiple que se presentarán más adelante se pueden generalizar para resolver SEDOP de manera similar a como se ha hecho con el método de Euler. Al trabajar con SEDO, tanto las aproximaciones numéricas como los diferentes tipos de errores obtenidos son vectores, para el análisis del error y en general el comportamiento de la soluciones numéricas es conveniente conocer un valor numérico del vector por este motivo se debe usar una norma vectorial. Según se presenta en [2] y [9], para los MEDO es conveniente usar la **norma infinito**

$$\mathbf{v}_\infty = \max_{0 \leq i \leq m} |v_i|, \quad (1.3.8)$$

donde v_i , para $0 < i \leq m$, son las entradas del vector \mathbf{v} de dimensión m . En este trabajo se usará la norma infinito.

1.3.3. Una cota para el error

En la Tabla 1.2 se observa que a medida que el tamaño de paso disminuye, el error global también disminuye. A continuación, tal y como se muestra en [5], se calcula una cota para el error la cual dependerá linealmente del tamaño de paso h . Para simplificar los cálculos en la búsqueda de la cota se considera los lemas a seguir.

Lema 1.1. *Para todo $z \geq -1$ y para cualquier m positivo, se cumple*

$$0 \leq (1+z)^m \leq e^{mz}.$$

Demostración. Desarrollando en serie de Taylor la función $f(z) = e^z$ en torno al punto $z_0 = 0$ hasta $n = 1$ tenemos

$$e^z = 1 + z + \frac{1}{2}z^2 e^\xi,$$

donde $\xi \in [z, 0]$. Por lo tanto,

$$0 \leq 1 + z \leq 1 + z + \frac{1}{2}z^2 e^\xi = e^z$$

y como $1 + z \geq 0$ se tiene que

$$0 \leq (1+z)^m \leq e^{mz}.$$

□

Lema 1.2. *Si s y x son números reales positivos, $\{a_i\}_{i=0}^k$ es una sucesión que satisface $a_0 \geq -x/s$ y*

$$a_{i+1} \leq (1+s)a_i + x, \quad \text{para } i = 0, 1, \dots, k, \quad (1.3.9)$$

entonces

$$a_{i+1} \leq e^{(i+1)s} \left(a_0 + \frac{x}{s} \right) - \frac{x}{s}.$$

Demostración. Para un centro fijo i , la desigualdad (1.3.9) implica que

$$\begin{aligned} a_{i+1} &\leq (1+s) a_i + x \\ &\leq (1+s) [(1+s) a_{i-1} + x] + x \\ &\leq (1+s) \{(1+s) [(1+s) a_{i-2} + x] + x\} + x \\ &\vdots \\ &\leq (1+s)^{i+1} a_0 + \left[1 + (1+s) + (1+s)^2 + \dots + (1+s)^i \right] x. \end{aligned}$$

Pero

$$1 + (1+s) + (1+s)^2 + \dots + (1+s)^i = \sum_{j=0}^i (1+s)^j$$

es una serie geométrica de razón $(1 + s)$ y su suma es

$$\frac{1 - (1 + s)i + 1}{1 - (1 + s)} = \frac{1}{s} \left[(1 + s)^{i+1} - 1 \right].$$

Por tanto,

$$a_{i+1} \leq (1 + s)^{i+1} a_0 + \frac{(1 + s)^{i+1} - 1}{s} x = (1 + s)^{i+1} \left(a_0 + \frac{x}{s} \right) - \frac{x}{s}$$

y de acuerdo con el Lema 1.1, con $z = 1 + s$, dada

$$a_{i+1} \leq e^{(i+1)s} \left(a_0 + \frac{x}{s} \right) - \frac{x}{s}.$$

□

Teorema 1.11. *Supongamos que la función f , del problema de Cauchy (1.3.1), es continua y satisface la condición de Lipschitz con la constante L en $D = \{(x, y) | a \leq x \leq b, -\infty < y < \infty\}$, y que existe una constante M con la propiedad de que $|y''| \leq M$ para todo $x \in [a, b]$.*

Sea $y(x_k)$ la solución exacta del problema de Cauchy (1.3.1) y sean y_0, y_1, \dots, y_n las aproximaciones generadas con el método de Euler para algún entero positivo n . Entonces para cada $k = 0, 1, \dots, n$,

$$e_k \leq \frac{hM}{2L} \left[e^{L(x_k-a)} - 1 \right]. \quad (1.3.10)$$

Demostración. Desarrollando la serie de Taylor para $y(x_{i+1})$ en torno a x_i tenemos

$$y(x_{k+1}) = y(x_k) + hf(x_k, y(x_k)) + \frac{h^2}{2} y''(\xi_k).$$

Por otro lado del método de Euler se tiene que

$$y_{k+1} = y_k + hf(x_k, y_k).$$

Restando estas ecuaciones tenemos

$$y(x_{k+1}) - y_{k+1} = y(x_k) - y_k + h [f(x_k, y(x_k)) - f(x_k, y_k)] + \frac{h^2}{2} y''(\xi_k).$$

De donde

$$e_{k+1} = |y(x_{k+1}) - y_{k+1}| \leq |y(x_k) - y_k| + h |f(x_k, y(x_k)) - f(x_k, y_k)| + \frac{h^2}{2} |y''(\xi_k)|.$$

Puesto que f satisface una condición de Lipschitz en la segunda variable con la constante L y como $|y''(x)| \leq M$, se sigue que

$$e_{k+1} \leq (1 + hL) |y(x_k) - y_k| + \frac{h^2 M}{2}.$$

Usando el Lema (1.3.9) con $a_j = e_k$ para cada $j = 0, 1, \dots, n$, y con $s = hL$ y $x = h^2M/2$, tenemos

$$e_k \leq e^{(k+1)hL} \left(|y(x_0) - y_0| + \frac{h^2M}{2hL} \right) - \frac{h^2M}{2hL}.$$

Puesto que $e_0 = |y(x_0) - y_0| = 0$ e $(k+1)h = x_{k+1} - x_0 = x_{k+1} - a$, tenemos

$$e_k \leq \frac{hM}{2L} (e^{(x_{k+1}-a)L} - 1),$$

para cada $k = 0, 1, \dots, n-1$.

□

La dificultad en este teorema está en conocer una cota para la segunda derivada de la función incógnita y , sin embargo por la regla de la cadena es posible en algunos casos calcular la segunda derivada de la función $y(x)$ sin necesidad de conocerla explícitamente.

Ejemplo 1.6. Para el problema de Cauchy

$$\begin{cases} y'(x) = y \\ y(0) = 1, \end{cases}$$

aplicar el método de Euler para aproximar $y(1)$. Además, con $h=0.2$, encontrar el error y la cota del error para cada x_k .

La solución exacta es $y(x) = e^x$, por lo tanto la constante de Lipschitz es $L = 1$ y una cota para la segunda derivada es $M = e \approx 2.71828$. Usando estos valores en la desigualdad (1.3.10), se obtiene la cota de error

$$e_k \leq \frac{0.2(2.71828)}{2} [e^{x_k} - 1].$$

En la Tabla 1.3 se muestra el error junto con la cota del error para cada x_k , se observa que la cota es mucho más grande que el error real. Estas cotas dependen del problema de Cauchy y del método numérico, muchas veces no es fácil determinar una cota para el error y si se encuentra en la práctica son mucho mayores que el error real.

x_i	0.2	0.4	0.6	0.8	1.0
Error	0.02140	0.05182	0.09411	0.15194	0.22996
Cota de error	0.06018	0.13369	0.22347	0.33313	0.46707

Tabla 1.3: Cota para el error método de Euler.

□

1.3.4. Métodos implícitos

Algunos de los métodos numéricos que se presentarán más adelante poseen fórmulas de integración implícitas, es decir que el valor y_{k+1} aparece en el lado izquierdo y en el lado derecho de la ecuación. Estos métodos se conocen como **métodos implícitos** y la principal dificultad en su aplicación está en que si la función $f(x, y)$ es trascendente puede resultar una ecuación implícita no lineal donde no sea posible despejar explícitamente el valor y_{k+1} . En ese caso se debe recurrir a métodos numéricos de resolución de ecuaciones no lineales como el método de Newton. Algunos ejemplos de métodos implícitos para solucionar numéricamente el problema de Cauchy (1.2.11) se presentan a seguir.

Método de Euler implícito

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{f}(x_k + h, \mathbf{y}_{k+1}). \end{cases} \quad (1.3.11)$$

Método del Trapecio implícito

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{2} [\mathbf{f}(x_k, \mathbf{y}_k) + \mathbf{f}(x_k + h, \mathbf{y}_{k+1})]. \end{cases} \quad (1.3.12)$$

Método de Simpson implícito

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), \quad \mathbf{y}_1 \text{ conocido} \\ \mathbf{y}_{k+2} = \mathbf{y}_{k+1} + \frac{h}{3} [\mathbf{f}(x_{k+2}, \mathbf{y}_{k+2}) + 4\mathbf{f}(x_{k+1}, \mathbf{y}_{k+1}) + \mathbf{f}(x_k, \mathbf{y}_k)]. \end{cases} \quad (1.3.13)$$

En los anteriores métodos y en cualquier otro MNEDO que se presenta en este trabajo $x_{k+1} = x_k + h$, $0 \leq k \leq n - 1$, $h = \frac{b-a}{n}$, $\mathbf{f} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Los métodos (1.3.11), (1.3.12) y 1.3.13 son implícitos. La diferencia está en que mientras el método de Euler implícito y el método del trapecio implícito son de paso único, el método de Simpson es de paso múltiple y para generar una aproximación necesitan dos aproximaciones en instantes de tiempos anteriores.

Capítulo 2

Métodos de paso único

En el Capítulo 1 se presentó el método de Euler (1.3.3), el cual permite obtener una aproximación y_{k+1} únicamente a partir de una aproximación y_k de un paso anterior. Los métodos numéricos que cumplen esta propiedad se conocen como métodos de paso único, MPU. Los MPU son de frecuente uso en la práctica y existen diferentes software que incorporan estos métodos. Sin embargo, cuando se utilizan estos algoritmos podría no tenerse indicios del proceso que se está realizando internamente. Por este motivo es importante conocer las propiedades generales de los MPU, analizar de donde surgen estos métodos y obtener criterios que permitan identificar cuando los resultados generados mediante la aplicación de uno de estos métodos ofrece resultados de buena precisión. Entre los MPU se destacan las familias de métodos de Taylor y de Runge-Kutta, por lo cual en este capítulo se presenta el desarrollo de estos métodos, además de un estudio de las propiedades de consistencia, convergencia y estabilidad para los MPU.

2.1. Forma general de los métodos de paso único

Un MPU es un método que calcula una aproximación \mathbf{y}_{k+1} únicamente a partir del valor de una aproximación \mathbf{y}_k del paso anterior. En forma general un MPU se escribe como

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h_k \phi(x_k, \mathbf{y}_k, \mathbf{y}_{k+1}, h_k), \end{cases} \quad (2.1.1)$$

donde ϕ se denomina **función incremento** y h_k es el tamaño de paso adoptado en el subintervalo $[x_k, x_{k+1}]$. A fin de simplificar el análisis se considerará fijo h_k , es decir $h_k = h$ para todo k . Si la variable \mathbf{y}_{k+1} aparece en los dos lados de la expresión (2.1.1), se dice que el método es implícito, caso contrario se dice que es un método explícito. Las definiciones y resultados que se presentan en este capítulo son válidas tanto para métodos implícitos como para métodos explícitos, por lo tanto por simplicidad enunciaremos estos resultados en el contexto de métodos explícitos.

La forma general de los MPU explícitos con h fijo que se considera es

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h\phi(x_k, \mathbf{y}_k, h). \end{cases} \quad (2.1.2)$$

Observación 2.1. El método de Euler (1.3.3) es un MPU, con

$$\phi(x_k, \mathbf{y}_k, h) = \mathbf{f}(x_k, \mathbf{y}_k).$$

Otro MPU, es el **método de Euler mejorado** que se estudiará más adelante y cumple (2.1.2) con

$$\phi(x_k, \mathbf{y}_k, h) = \frac{1}{2} [f(x_k, y_k) + f(x_k + h, y_k + hf(x_k, y_k))].$$

2.2. Consistencia, estabilidad y convergencia

Cuando se usan métodos numéricos para obtener una aproximación a la solución, es importante conocer que tanto se acerca la solución numérica a la solución exacta. En el caso de los métodos numéricos para EDO vamos a considerar dos tipos de error: el **error local de discretización** y el **error global de discretización**.

Definición 2.1. Dado el MPU (2.1.2), el **error local de discretización** para el método en x_k se define por

$$\tau_k = \frac{\mathbf{y}(x_{k+1}) - \mathbf{y}(x_k)}{h} - \phi(x_k, \mathbf{y}(x_k), h). \quad (2.2.1)$$

Definición 2.2. Dado el MPU (2.1.2), el **error global de discretización** del método en x_k es la diferencia entre la solución exacta y la solución numérica del problema de Cauchy en x_k , es decir

$$e_k = \mathbf{y}(x_k) - \mathbf{y}_k. \quad (2.2.2)$$

Observación 2.2. Multiplicando (2.2.1) por el tamaño de paso h se tiene

$$h\tau_k = \mathbf{y}(x_{k+1}) - [\mathbf{y}(x_k) + h\phi(x_k, \mathbf{y}(x_k), h)] = \mathbf{y}(x_{k+1}) - \mathbf{y}_{k+1}. \quad (2.2.3)$$

Lo cual puede ser visto como la diferencia entre la solución exacta y la solución aproximada, donde la solución numérica se ha calculado en una sola aplicación del método numérico, es decir que ningún error fue cometido anteriormente, esto le confiere el carácter local. La ecuación (2.2.2), también puede interpretarse como la diferencia entre la solución exacta y la solución aproximada, con la diferencia que en este caso la solución numérica se ha calculado a partir de valores obtenidos por la aplicación del método numérico en instantes de tiempo anteriores, esto le confiere el carácter global.

Para tener buenas aproximaciones numéricas se debe mantener controlados estos errores, es decir procurar que la magnitud de estos errores sea lo más baja posible y que de hecho tienda a cero a medida que h tienda a cero, o equivalentemente a medida que n tienda a infinito. En este sentido se presenta las siguientes definiciones.

Definición 2.3. Se dice que el MPU (2.1.2) es **consistente** con el problema de Cauchy (1.2.11) si y solo si la función de incremento, $\phi(x, \mathbf{y}, h)$, satisface la siguiente relación

$$\phi(x, \mathbf{y}, 0) = \mathbf{f}(x, \mathbf{y}).$$

En otras palabras, un MPU es consistente si y solo si

$$\lim_{h \rightarrow 0} \|\boldsymbol{\tau}_k\| = 0, \quad \forall x \in [a, b], \quad \mathbf{y} \in \mathbb{R}^n.$$

Definición 2.4. Si existieran constantes positivas C , h_0 y q , independiente del paso de integración h y del subíndice k , con $0 < h \leq h_0$, tales que el error local de discretización satisface

$$\max_k \|\boldsymbol{\tau}_k\| \leq Ch^q,$$

entonces se dice que el método numérico tiene **orden de consistencia** q y se denota $O(h^q)$.

Observación 2.3. La consistencia de un MPU asegura que el método es al menos de orden de consistencia uno. El orden de consistencia indica que tan rápido el error local de discretización se acerca a cero cuando h disminuye. Análogamente el **orden de convergencia** indica que tan rápido el error global de discretización disminuye. Dado que para las definiciones de error local y error global de discretización que se han adoptan en este trabajo los ordenes de consistencia y convergencia coinciden, en adelante simplemente se dirá que un método es de orden q .

Definición 2.5. Se dice que el MPU (2.1.2) es **convergente** en el punto x_k si y solo si

$$\lim_{h \rightarrow 0} \|\mathbf{e}_k\| = 0.$$

Es decir, que la convergencia se tiene en un MPU si y solamente si el error global de discretización tiende a cero cuando h tiende a cero. El método numérico es convergente si fuera convergente para todo $x \in [a, b]$ y para cualquier problema de Cauchy.

La condición inicial y_0 es dada de forma exacta por el problema de Cauchy, sin embargo, en la práctica la condición inicial puede estar sujeta a un tipo de error. Es decir que $y(0) = y_0 + \delta_0$, donde δ_0 es una perturbación a la condición inicial. Por lo tanto en la práctica se trabaja con problemas de Cauchy de la forma

$$\begin{cases} \mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x)) \\ \mathbf{y}(\mathbf{x}_0) = \mathbf{y}_0 + \delta_0. \end{cases}$$

Este tipo de problemas se presenta cuando la condición inicial $y(0)$ es calculada de forma experimental e incluso la precisión finita de un computador impide almacenar el valor de $y(0)$ de forma exacta. Por lo tanto, es importante garantizar que los métodos usados sean **estables numéricamente** en el sentido de que pequeñas perturbaciones en las condiciones iniciales no produzcan grandes perturbaciones en los resultados finales.

Definición 2.6. Se dice que el MPU es **estable** si existe una constante $K > 0$ tal que, para cualquier par de soluciones numéricas, \mathbf{y}_{k+1} y $\hat{\mathbf{y}}_{k+1}$, obtenidas de aplicar el MPU al mismo problema de Cauchy pero con diferentes condiciones iniciales, \mathbf{y}_0 y $\hat{\mathbf{y}}_0$, respectivamente, se tiene que

$$\|\mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1}\| \leq K \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|,$$

para $x_{k+1} \leq b$, y para h que tiende a 0.

Como se verá más adelante, en la Observación 3.3 se concluye, para los MPU la consistencia del método garantiza la estabilidad del mismo, es decir que para verificar que un MPU es estable basta verificar que el método sea consistente.

Garantizar la convergencia de un MPU mediante la Definición 2.5 resulta una tarea compleja ya que en la práctica normalmente no se conoce el error global de discretización y además la definición de convergencia de un MPU exige la verificación de que para todo problema de Cauchy, el error global tiende a cero cuando h tiende a cero, lo cual no es posible. Por tal motivo, se presenta el siguiente resultado que ofrece una forma alternativa de garantizar la convergencia de un MPU a partir de la consistencia como se muestra en [2] y [7].

Teorema 2.1. *Considere el MPU (2.1.2), donde la función incremento $\phi(x, \mathbf{y}, h)$ es Lipschitziana en \mathbf{y} y continua en sus argumentos. Si el MPU es consistente entonces es convergente.*

Demostración. Para determinar la convergencia del MPU se debe obtener una expresión que describa el comportamiento del error global de discretización, de la forma

$$\mathbf{e}_{k+1} = \mathbf{y}(x_{k+1}) - \mathbf{y}_{k+1}. \quad (2.2.4)$$

Considerando la expresión (2.2.3) del producto entre el error local de discretización τ_k y el paso de integración h se tiene que

$$\mathbf{y}(x_{k+1}) = \mathbf{y}(x_k) + h\phi(x_k, \mathbf{y}(x_k), h) + h\tau_k. \quad (2.2.5)$$

Por otro lado considérese \mathbf{y}_{k+1} como una aproximación numérica obtenida por la aplicación del MPU (2.1.2), por tanto

$$\mathbf{y}_{k+1} = \mathbf{y}_k + h\phi(x_k, \mathbf{y}_k, h). \quad (2.2.6)$$

Sustituyendo (2.2.5) y (2.2.6) en (2.2.4) y agrupando términos se tiene

$$\mathbf{e}_{k+1} = \mathbf{y}(x_k) - \mathbf{y}_k + h [\phi(x_k, \mathbf{y}(x_k), h) - \phi(x_k, \mathbf{y}_k, h)] + h\boldsymbol{\tau}_k \quad (2.2.7)$$

$$= \mathbf{e}_k + h [\phi(x_k, \mathbf{y}(x_k), h) - \phi(x_k, \mathbf{y}_k, h)] + h\boldsymbol{\tau}_k. \quad (2.2.8)$$

Por hipótesis la función $\phi(x, y, h)$ satisface la condición de Lipschitz en la variable \mathbf{y} , es decir

$$\|\phi(x, \mathbf{y}_1, h) - \phi(x, \mathbf{y}_2, h)\| \leq L \|\mathbf{y}_1 - \mathbf{y}_2\|,$$

para cualquier x, h y para una constante positiva L . Además ϕ es continua en todas sus variables, por tanto aplicando estas hipótesis en (2.2.8) se tiene que

$$\begin{aligned} \|\mathbf{e}_{k+1}\| &\leq \|\mathbf{e}_k\| + h \|\phi(x_k, \mathbf{y}(x_k), h) - \phi(x_k, \mathbf{y}_k, h)\| + h \|\boldsymbol{\tau}_k\| \\ &\leq \|\mathbf{e}_k\| + hL \|\mathbf{y}(x_k) - \mathbf{y}_k\| + h \|\boldsymbol{\tau}_k\| \\ &\leq \|\mathbf{e}_k\| + hL \|\mathbf{e}_k\| + h \|\boldsymbol{\tau}_k\|, \end{aligned}$$

esto es

$$\|\mathbf{e}_{k+1}\| \leq (1 + hL) \|\mathbf{e}_k\| + h \|\boldsymbol{\tau}_k\|, \text{ para todo } k. \quad (2.2.9)$$

A partir de (2.2.9), de manera recursiva se puede obtener una expresión que dependa de e_0 así

$$\begin{aligned} \|\mathbf{e}_1\| &\leq (1 + hL) \|\mathbf{e}_0\| + h \|\boldsymbol{\tau}_k\| \\ \|\mathbf{e}_2\| &\leq (1 + hL)^2 \|\mathbf{e}_0\| + [(1 + hL) + 1] h \|\boldsymbol{\tau}_k\| \\ \|\mathbf{e}_3\| &\leq (1 + hL)^3 \|\mathbf{e}_0\| + [(1 + hL)^2 + (1 + hL) + 1] h \|\boldsymbol{\tau}_k\| \\ &\vdots \\ \|\mathbf{e}_k\| &\leq (1 + hL)^k \|\mathbf{e}_0\| + \left[(1 + hL)^{k-1} + \dots + (1 + hL) + 1 \right] h \|\boldsymbol{\tau}_k\|. \end{aligned}$$

El segundo término del lado derecho en la última desigualdad es una suma de los k términos de una progresión geométrica de término inicial 1 y razón $(1 + hL)$. Por tanto

$$\|\mathbf{e}_k\| \leq (1 + hL)^k \|\mathbf{e}_0\| + \frac{(1 + hL)^k - 1}{L} h \|\boldsymbol{\tau}_k\|. \quad (2.2.10)$$

Aplicando el Lema 1.1, se tiene que

$$(e^{hL})^k \geq (1 + hL)^k.$$

Luego usando esta desigualdad en (2.2.10) se tiene que

$$\|\mathbf{e}_k\| \leq e^{khL} \|\mathbf{e}_0\| + \frac{e^{khL} - 1}{L} h \|\boldsymbol{\tau}_k\|. \quad (2.2.11)$$

Por hipótesis el método es consistente, por lo tanto es consistente al menos de orden 1, es decir que

$$\max_k \|\boldsymbol{\tau}_k\| \leq Ch^q \text{ donde } q \geq 1,$$

entonces se tiene que

$$\|\mathbf{e}_k\| \leq e^{khL} \|\mathbf{e}_0\| + \frac{e^{khL} - 1}{L} h^q. \quad (2.2.12)$$

Si $\|\mathbf{e}_0\| = 0$ el lado derecho de la desigualdad (2.2.12) tiende a cero cuando h tiende a cero. De esta manera el límite del error global tiende a cero cuando el tamaño de paso h tiende a cero, es decir el MPU es convergente. \square

2.3. Métodos de Taylor

En el Capítulo 1 para deducir el método de Euler se usó la expansión de Taylor truncando hasta el segundo término. En este capítulo, a fin de alcanzar una mayor precisión que la obtenida con el método de Euler, se considera una mayor cantidad de términos en el desarrollo de Taylor. Esto da origen a una familia de MPU conocida como métodos de Taylor.

Suponga que la solución $\mathbf{y}(x)$ es $q + 1$ veces continuamente diferenciable. El objetivo es usar la serie de Taylor para expresar $\mathbf{y}(x + h)$ en términos de $\mathbf{y}(x)$ para algún tamaño de paso h , utilizando información acerca de la ecuación diferencial del problema de Cauchy. Expandiendo en series de Taylor la función $\mathbf{y}(x + h)$ con centro en $x = x_k$, se tiene

$$\mathbf{y}(x_k + h) = \mathbf{y}(x_{k+1}) = \mathbf{y}(x_k) + h\mathbf{y}'(x_k) + \frac{h^2}{2!}\mathbf{y}''(x_k) + \dots + \frac{h^{q+1}}{(q+1)!}\mathbf{y}^{(q+1)}(\xi),$$

donde $x_k < \xi < x_k + h$. Sustituyendo $\mathbf{y}'(x_k) = \mathbf{f}(x_k, \mathbf{y}(x_k))$ del problema de Cauchy, se obtiene el siguiente método.

Método de Taylor de Taylor de orden q

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{f}(x_k, \mathbf{y}_k) + \frac{h^2}{2!} \frac{d}{dt} \mathbf{f}(x_k, \mathbf{y}_k) + \dots + \frac{h^q}{q!} \frac{d^q}{dt^q} \mathbf{f}(x_k, \mathbf{y}_k). \end{cases} \quad (2.3.1)$$

El error local de discretización del método de Taylor es

$$\tau_k(h) = \frac{h^q}{(q+1)!} \mathbf{y}^{(q+1)}(\xi).$$

Por lo tanto el método de Taylor de orden q tiene un error local de discretización $O(h^q)$.

Para calcular las derivadas de \mathbf{f} en la expresión (2.3.1) se debe calcular las derivadas totales de $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$ teniendo en cuenta que \mathbf{f} es una función implícita de \mathbf{y} . Para simplificar el cálculo de las derivadas de \mathbf{f} se usa la siguiente notación

$$\mathbf{f} = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{f}_x = \frac{\partial \mathbf{f}(x, \mathbf{y})}{\partial x}, \quad \mathbf{f}_{xx} = \frac{\partial^2 \mathbf{f}(x, \mathbf{y})}{\partial x^2}, \quad \mathbf{f}_{yy} = \frac{\partial^2 \mathbf{f}(x, \mathbf{y})}{\partial y^2}, \quad \mathbf{f}_{xy} = \frac{\partial^2 \mathbf{f}(x, \mathbf{y})}{\partial x \partial y}.$$

Aparentemente con los métodos de Taylor se puede obtener métodos numéricos de ordenes arbitrariamente altos. Para desarrollar un método de Taylor de orden 2 es necesario el cálculo de las 2 derivadas de \mathbf{y} ,

$$\begin{aligned}\mathbf{y}'(x) &= \mathbf{f} \\ \mathbf{y}''(x) &= \mathbf{f}' = \mathbf{f}_x + \mathbf{f}_y \mathbf{f},\end{aligned}\tag{2.3.2}$$

de lo cual sustituyendo en (2.3.1) se obtiene el correspondiente método de Taylor.

Método de Taylor de orden 2 (Taylor 2)

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{f}(x_k, \mathbf{y}_k) + \frac{1}{2}h^2(\mathbf{f}_x(x_k, \mathbf{y}_k) + \mathbf{f}_y(x_k, \mathbf{y}_k)\mathbf{f}(x_k, \mathbf{y}_k)). \end{cases}\tag{2.3.3}$$

Sin embargo, el cálculo de las derivadas (2.3.2) es tedioso, a menos que la función \mathbf{f} sea lo suficientemente simple para que muchas de las derivadas parciales en (2.3.2) desaparezcan. Una alternativa en la práctica sería usar software de cálculo simbólico para calcular tales derivadas, aunque se debe tener en cuenta que esto demanda mayores costos computacionales. Por lo tanto los métodos de Taylor no son de gran utilidad práctica.

2.4. Métodos de Runge Kutta

Los métodos de Runge Kutta forman una importante familia de MPU. La principal característica de estos métodos, al compararlos con los de Taylor, es que ayudan a evitar el cálculo y la evaluación de la derivada de $\mathbf{f}(x, \mathbf{y})$ incluso con alta orden. Un ejemplo de estos métodos es el método Euler (1.3.3). Estos métodos surgen en 1895 por la idea de Runge de generalizar el método de Euler, teniendo en cuenta una serie de evaluaciones en la función $\mathbf{f}(x, \mathbf{y})$ para aproximar sus sucesivas derivadas en un solo paso. Otras contribuciones fueron hechas por Heun y Kutta al rededor de 1900. Este último caracterizó por completo el conjunto de métodos de Runge Kutta de orden 4, y propuso los primeros métodos de orden 5, más información histórica se presenta en [6].

Definición 2.7. Un **método de Runge Kutta explícito de R estados**, es un MPU en el cual el número de estados hace referencia al número de evaluaciones de $\mathbf{f}(x, \mathbf{y})$ que deben hacerse en cada paso. Estos métodos tienen la forma

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h\phi(x_k, \mathbf{y}_k, h), \end{cases}$$

donde

$$\phi(x, \mathbf{y}, h) = \sum_{r=1}^R c_r \mathbf{k}_r,\tag{2.4.1}$$

con

$$\begin{aligned} \mathbf{k}_1(x, \mathbf{y}) &= \mathbf{f}(x, \mathbf{y}) \\ \mathbf{k}_2(x, \mathbf{y}) &= \mathbf{f}(x + ha_2, \mathbf{y} + hb_{21}\mathbf{k}_1) \\ \mathbf{k}_3(x, \mathbf{y}) &= \mathbf{f}(x + ha_3, \mathbf{y} + hb_{31}\mathbf{k}_1 + hb_{32}\mathbf{k}_2) \\ &\vdots \\ \mathbf{k}_r(x, \mathbf{y}) &= \mathbf{f}\left(x + ha_r, \mathbf{y} + h \sum_{s=1}^{r-1} b_{rs} \mathbf{k}_s\right), \quad 2 \leq r \leq R, \end{aligned} \quad (2.4.2)$$

y los parámetros a_r , b_{rs} y c_r satisfacen las relaciones

$$\sum_{r=1}^R c_r = 1 \quad y \quad a_r = \sum_{s=1}^{r-1} b_{rs}, \quad 2 \leq r \leq R. \quad (2.4.3)$$

Para deducir métodos de tipo Runge Kutta, se debe considerar la expresión general de la Definición 2.7 y determinar las constantes c_r , a_r y b_{rs} en (2.4.3). Una técnica para determinar las constantes es desarrollar en series de Taylor las funciones \mathbf{k}_r en (2.4.2) y comparar este desarrollo con los coeficientes de un método de Taylor de orden q . A seguir se ejemplifica este proceso.

Ejemplo 2.1. Para deducir un método de tipo Runge Kutta de dos estados, para aproximar el problema de Cauchy (1.3.1), se considera la Definición 2.7 con $R = 2$ y se obtiene

$$\begin{cases} y_{k+1} = y_k + h(c_1 k_1 + c_2 k_2) \\ k_1 = f(x_k, y_k) \\ k_2 = f(x_k + a_2 h, y_k + hb_{21} k_1). \end{cases} \quad (2.4.4)$$

Como se deben de cumplir las condiciones (2.4.3), se tiene que $a_2 = b_{21}$. Por lo tanto, la expresión (2.4.4) se puede escribir como

$$\begin{cases} y_{k+1} = y_k + h(c_1 k_1 + c_2 k_2) \\ k_1 = f(x_k, y_k) \\ k_2 = f(x_k + a_2 h, y_k + ha_2 k_1). \end{cases} \quad (2.4.5)$$

Expandiendo k_2 en series de Taylor con centro en (x_k, y_k) hasta el orden 2, se tiene

$$k_2 = f + a_2 h [f_x + k_1 f_y] + \frac{a_2^2 h^2}{2!} [f_{xx} + 2f_x f_{xy}]. \quad (2.4.6)$$

Remplazando la ecuación (2.4.6) en la ecuación (2.4.5) se tiene

$$\begin{aligned} y_{k+1} &= y_k + (c_1 + c_2)hf + a_2 c_2 h^2 [f_x + ff_y] \\ &\quad + \frac{a_2^2 c_2 h^3}{2!} [f_{xx} + 2ff_{xy} + f^2 f_{yy}]. \end{aligned} \quad (2.4.7)$$

Comparando los coeficientes de la ecuación (2.4.7) con los coeficientes de la ecuación del método de Taylor de orden 3 (2.3.3), se concluye que

$$c_1 + c_2 = 1 \quad y \quad a_2 c_2 = \frac{1}{2},$$

de donde se puede elegir estas constantes de diferentes formas. Si elegimos $c_1 = c_2 = \frac{1}{2}$, se tiene que $a_2 = 1$, con esta elección se obtiene el método de Euler mejorado.

Método de Euler mejorado

$$\begin{cases} y_0 = y(x_0) \\ y_{k+1} = y_k + \frac{h}{2} [f(x_k, y_k) + f(x_k + h, y_k + hf(x_k, y_k))] . \end{cases} \quad (2.4.8)$$

Si se toma $c_1 = 0$ y $c_2 = 1$, se tiene que $a_2 = \frac{1}{2}$ que corresponden al método de Euler modificado.

Método de Euler modificado

$$\begin{cases} y_0 = y(x_0) \\ y_{k+1} = y_k + hf \left(x_k + \frac{h}{2}, y_k + \frac{h}{2} f(x_k, y_k) \right) . \end{cases} \quad (2.4.9)$$

□

Los métodos de Euler mejorado (2.4.8) y Euler modificado (2.4.9) también pueden ser aplicados para resolver SEDOP.

Usando la técnica presentada en el Ejemplo 2.1 se pueden obtener métodos de alto orden, para lo cual se debe de considerar un mayor número de estados. Se presentan a seguir algunos de los métodos de Runge Kutta clásicos.

Método de Runge Kutta de orden tres con tres estados (RK33)

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{6} (\mathbf{k}_1 + 4\mathbf{k}_2 + \mathbf{k}_3) \\ \mathbf{k}_1 = \mathbf{f}(x_k, \mathbf{y}_k) \\ \mathbf{k}_2 = \mathbf{f}(x_k + \frac{h}{2}, \mathbf{y}_k + \frac{h}{2}\mathbf{k}_1) \\ \mathbf{k}_3 = \mathbf{f}(x_k + h, \mathbf{y}_k - h\mathbf{k}_1 + 2h\mathbf{k}_2) . \end{cases} \quad (2.4.10)$$

Método de Runge Kutta de orden cuatro con cuatro estados (RK44)

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{6} (\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4) \\ \mathbf{k}_1 = \mathbf{f}(x_k, \mathbf{y}_k) \\ \mathbf{k}_2 = \mathbf{f}(x_k + \frac{h}{2}, \mathbf{y}_k + \frac{h}{2}\mathbf{k}_1) \\ \mathbf{k}_3 = \mathbf{f}(x_k + \frac{h}{2}, \mathbf{y}_k + \frac{h}{2}\mathbf{k}_2) \\ \mathbf{k}_4 = \mathbf{f}(x_k + h, \mathbf{y}_k + h\mathbf{k}_3) . \end{cases} \quad (2.4.11)$$

Método de Runge Kutta de orden cinco con seis estados (RK56)

$$\left\{ \begin{array}{l} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + h(\frac{16}{135}\mathbf{k}_1 + \frac{6656}{12825}\mathbf{k}_3 + \frac{28561}{56430}\mathbf{k}_4 - \frac{9}{50}\mathbf{k}_5 + \frac{2}{55}\mathbf{k}_6) \\ \mathbf{k}_1 = \mathbf{f}(x_k, \mathbf{y}_k) \\ \mathbf{k}_2 = \mathbf{f}(x_k + \frac{h}{4}, \mathbf{y}_k + \frac{h}{4}\mathbf{k}_1) \\ \mathbf{k}_3 = \mathbf{f}(x_k + \frac{3h}{8}, \mathbf{y}_k + \frac{3h}{32}\mathbf{k}_1 + \frac{9h}{32}\mathbf{k}_2) \\ \mathbf{k}_4 = \mathbf{f}(x_k + \frac{12h}{13}, \mathbf{y}_k + \frac{1932h}{2197}\mathbf{k}_1 - \frac{7200h}{2197}\mathbf{k}_2 + \frac{7296h}{2197}\mathbf{k}_3) \\ \mathbf{k}_5 = \mathbf{f}(x_k + h, \mathbf{y}_k + \frac{439h}{216}\mathbf{k}_1 - 8h\mathbf{k}_2 + \frac{3680h}{513}\mathbf{k}_3 - \frac{845h}{4104}\mathbf{k}_4) \\ \mathbf{k}_6 = \mathbf{f}(x_k + \frac{h}{2}, \mathbf{y}_k - \frac{8h}{27}\mathbf{k}_1 + 2h\mathbf{k}_2 - \frac{3544h}{2565}\mathbf{k}_3 + \frac{1859h}{4104}\mathbf{k}_4 - \frac{11h}{40}\mathbf{k}_5). \end{array} \right. \quad (2.4.12)$$

Si se observan los métodos (2.4.10), (2.4.11) y los deducidos en el Ejemplo 2.1, podría suponerse que si se considera un método de R etapas siempre se puede obtener un método que alcance orden R , sin embargo, en general esto es falso. Butcher demostró que no existe un método de cinco etapas de orden cinco y demostró el siguiente resultado, el cual se presenta en [7] y [12].

Teorema 2.2. *Sea $p^*(R)$ el mayor orden que se puede alcanzar mediante un método de Runge Kutta de R estados, entonces*

$$\begin{aligned} p^*(R) &= R, \quad R = 1, 2, 3, 4, \\ p^*(5) &= 4, \\ p^*(6) &= 5, \\ p^*(7) &= 6, \\ p^*(8) &= 6, \\ p^*(9) &= 7, \\ p^*(R) &= R - 2 \quad R = 10, 11, \dots \end{aligned} \quad (2.4.13)$$

El Teorema 2.2 justifica el hecho de que en la práctica se utilice con mayor frecuencia los métodos de Runge Kutta de orden 4, ya que usar métodos de más de 4 estados demanda un número mayor de evaluaciones en la función f , lo cual genera grandes costos computacionales, con un orden de convergencia menor a la cantidad de estados del método.

2.5. Estabilidad absoluta

En el estudio de las propiedades de consistencia, estabilidad y convergencia para los MPU se consideraba que el tamaño de paso h tiende a cero. Sin embargo, en la práctica tomar tamaños de paso muy pequeños demanda una mayor cantidad de cálculos, lo cual se traduce en un alto costo computacional y de hecho por la aritmética finita del computador, al trabajar con h muy pequeño

se pueden generar mayores errores de redondeo. Por lo tanto, se debe elegir el valor de h lo más grande posible para el cual la solución aproximada se comporte por lo menos de forma similar con la solución exacta. El estudio de la elección del tamaño de paso adecuado se denomina **estabilidad absoluta**, donde ahora la palabra estabilidad se refiere al comportamiento de la solución numérica para un h fijo.

Para estudiar la estabilidad absoluta de los MPU se debe realizar un análisis cualitativo del conjunto de puntos en el plano complejo en el que la solución numérica tenga el mismo comportamiento de la solución analítica, este conjunto de puntos se conoce como la **región de estabilidad**. Para simplificar el análisis cualitativo se trabaja con problemas lineales con coeficientes constantes. Iniciamos el estudio de estabilidad absoluta a partir del método de Euler.

Estabilidad método de Euler

Considérese el SEDOP

$$\begin{cases} \mathbf{y}'(x) = M\mathbf{y}(x), \\ \mathbf{y}(0) = \mathbf{y}_0, \end{cases} \quad (2.5.1)$$

donde M es una matriz cuadrada de entradas constantes. La solución exacta del SEDOP en $x_k = kh$ es

$$\mathbf{y}(x_k) = e^{khM}\mathbf{y}_0.$$

Al hacer un cambio de base, de modo que $\mathbf{y}(x) = S\hat{\mathbf{y}}(x)$, el sistema (2.5.1) se puede reescribir en la forma

$$\hat{\mathbf{y}}'(x) = \hat{M}\hat{\mathbf{y}}(x), \quad (2.5.2)$$

donde $\hat{M} = S^{-1}MS$ es la forma canónica de Jordan de M . Aplicando el método de Euler al problema (2.5.2) se obtiene la solución numérica

$$\hat{\mathbf{y}}_k = (I + h\hat{M})^k\hat{\mathbf{y}}_0.$$

Para el caso en el que todos los valores propios λ de M sean diferentes, una de las ecuaciones del sistema tiene la forma simple

$$y'(x) = \lambda y(x). \quad (2.5.3)$$

Por lo tanto, para obtener información aceptable del comportamiento de la solución de (2.5.1) se puede estudiar el comportamiento de (2.5.3). La solución numérica por el método de Euler para (2.5.3) es

$$y_k = (1 + \lambda h)^k y_0. \quad (2.5.4)$$

Se desea que esta solución presente el mismo comportamiento que la solución analítica $y = e^{(nh\lambda)}$, donde λ puede ser real o complejo.

En el caso de que $\lambda \in \mathbb{R}$ se presentan dos casos, $\lambda < 0$ o $\lambda \geq 0$.

- Si $\lambda < 0$, entonces $\lim_{k \rightarrow \infty} y(x_k) = \lim_{k \rightarrow \infty} y_0 e^{\lambda kh} = 0$.

Para $k \rightarrow \infty$, la solución numérica (2.5.4), tendrá el mismo comportamiento que la solución analítica si

$$|1 + \lambda h| < 1 \Rightarrow -2 < \lambda h < 0.$$

Por lo tanto, la solución numérica tiene el mismo comportamiento de la solución analítica, cuando $k \rightarrow \infty$, si $\lambda h \in (-2, 0)$. En este caso existe una restricción sobre h a fin de que el método de Euler sea estable.

- Si $\lambda \geq 0$, se tiene que

$$\lim_{k \rightarrow \infty} y(x_k) = \lim_{x \rightarrow \infty} y_0 e^{\lambda kh} = \begin{cases} y_0, & \text{si } \lambda = 0 \\ \infty, & \text{si } \lambda > 0. \end{cases}$$

En este caso la solución numérica (2.5.4) tiene el mismo comportamiento de la solución analítica cuando $k \rightarrow \infty$, por tanto en este caso no existen restricciones sobre h para que el método de Euler sea estable.

De lo anterior se concluye que el método de Euler es absolutamente estable para $\lambda h \in (-2, 0)$.

Si se considera que $\lambda \in \mathbb{C}$, $\lambda h \in \mathbb{C}$ y se tiene

$$\begin{aligned} z &= \lambda h = a + ib, \quad a, b \in \mathbb{R} \\ |1 + \lambda h| &< 1 \\ |z + 1| &< 1 \\ |a + ib + 1| &< 1, \end{aligned}$$

es decir

$$[a - (-1)]^2 + b^2 < 1. \quad (2.5.5)$$

Por lo tanto, en el plano complejo la región de estabilidad para el método de Euler son los puntos interiores de un disco de centro en el punto $(-1, 0)$ y de radio 1.

Análogamente al estudio de estabilidad absoluta para el método de Euler, se puede considerar cualquier MPU y al aplicarlo al problema (2.5.1) obtener una expresión de la forma

$$y_{k+1} = \psi(\lambda h)y_k,$$

donde la expresión $\psi(\lambda h)$ se denomina **factor de amplificación** y el conjunto

$$\Omega = \{\mu \in C; |\psi(\mu)| < 1\},$$

se conoce como **región de estabilidad Absoluta**. La intercepción de la región Ω con la recta real determinan el **intervalo de estabilidad absoluta** para el MPU.

Los principales resultados de estabilidad absoluta para los métodos de Runge Kutta explícitos se resumen en la Figura 2.1 y en la Tabla 2.1. De estos notamos que entre mayor es el orden del método es más complicado determinar la región de estabilidad absoluta asociada, pero esta es mayor al compararla con los de orden menor.

Método	Factor de amplificación	Intervalo de EA
Euler	$1 + \lambda h$	(-2, 0)
RK22	$1 + \lambda h + \frac{(\lambda h)^2}{2}$	(-2, 0)
RK33	$1 + \lambda h + \frac{(\lambda h)^2}{2} + \frac{(\lambda h)^3}{6}$	(-2.51, 0)
RK44	$1 + \lambda h + \frac{(\lambda h)^2}{2} + \frac{(\lambda h)^3}{6} + \frac{(\lambda h)^4}{24}$	(-2.78, 0)
RK56	$1 + \lambda h + \frac{(\lambda h)^2}{2} + \frac{(\lambda h)^3}{6} + \frac{(\lambda h)^4}{24} + \frac{(\lambda h)^5}{120} + \frac{(\lambda h)^6}{1280}$	(-2.78, 0)

Tabla 2.1: Intervalos de estabilidad absoluta para los métodos de Runge-Kutta [12].

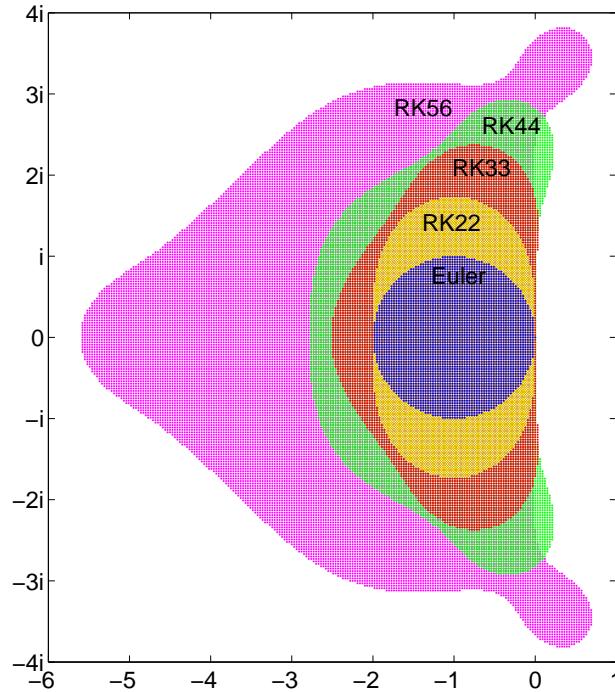


Figura 2.1: Región de estabilidad absoluta para los métodos de Runge Kutta.

Observación 2.4. Para diseñar la Figura 2.1 en Matlab, se consideró un conjunto de puntos λh en las vecindades del origen del plano complejo. De estos puntos se representó gráficamente en el plano complejo los puntos que cumplían la condición $|\psi(\mu)| < 1$, es decir los puntos que estaban en las regiones de estabilidad para cada uno de los métodos de Runge Kutta.

Capítulo 3

Métodos de paso múltiple

En la sección anterior se estudiaron los MPU, particularmente los métodos de Taylor y Runge Kutta, mediante los cuales es posible obtener una solución aproximada, \mathbf{y}_{k+1} , del problema de Cauchy (1.2.11) a partir de una única solución aproximada \mathbf{y}_k del paso anterior. Los métodos de paso múltiple, MPM, necesitan conocer varios valores de la discretización en pasos anteriores al de interés, es decir que para producir una solución aproximada, \mathbf{y}_{k+N} , es necesario conocer las aproximaciones $\mathbf{y}_k, \mathbf{y}_{k+1}, \dots, \mathbf{y}_{k+N-1}$, de pasos anteriores, en este caso el MPM se conoce como **Método de N-pasos**. Los MPU pueden ser considerados un caso particular de los MPM, cuando $N = 1$.

La idea de John C. Adams y Bashford de mejorar las aproximaciones obtenidas mediante el método de Euler dio origen en 1833 a los primeros métodos de paso múltiple, conocidos como métodos de Adams-Bashford. Sin embargo, la teoría moderna de los métodos de paso múltiple lineales, MPML, fue desarrollada solo hasta 1956 por Dahlquist, y se dio a conocer por los trabajos de Henrici entre 1962 y 1963, tal y como se muestra en [7].

En este trabajo, únicamente se estudiarán propiedades generales para los MPML, que son métodos en los cuales la expresión que los define es una combinación lineal de $\mathbf{y}_k, \mathbf{y}_{k+1}, \dots, \mathbf{y}_{k+N}$ y de la derivada \mathbf{f} evaluada en $(x_k, \mathbf{y}_k), (x_{k+1}, \mathbf{y}_{k+1}), \dots, (x_{k+N}, \mathbf{y}_{k+N})$. Se omite el estudio de métodos de paso múltiple no lineales ya que estos se usan especialmente en problemas de Cauchy que son rígidos o cuyas soluciones poseen singularidades, problemas sobre los cuales no se profundiza en este trabajo, algunos MPM no lineales se presentan en [8].

3.1. Forma general de los métodos de paso múltiple lineales

El método de Simpson (1.3.13) del Capítulo 1 es un método de 3 pasos lineal, ya que la expresión que define el método es una combinación lineal de $\mathbf{y}_{k-1}, \mathbf{y}_k$ y \mathbf{y}_{k+1} y de la derivada \mathbf{f} evaluada en

(x_k, \mathbf{y}_k) , $(x_{k+1}, \mathbf{y}_{k+1})$ y $(x_{k+2}, \mathbf{y}_{k+2})$. Note que si se aplica este método para resolver el problema de Cauchy (1.2.11) solo disponemos de una condición inicial \mathbf{y}_0 , pero el método exige que además de la condición inicial se debe conocer el valor de \mathbf{y}_1 . Por lo tanto, es necesario usar diferentes técnicas para aproximar los valores iniciales necesarios para utilizar un MPML.

En forma general un MPML de N pasos se escribe como

$$\sum_{j=0}^N \alpha_j \mathbf{y}_{k+j} = h \sum_{j=0}^N \beta_j \mathbf{f}_{k+j}, \quad (3.1.1)$$

donde α_j y β_j son reales, $\alpha_N \neq 0$, $\mathbf{f} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, $h = (b - a)/n$, $x_{k+j} = x_k + jh$ y $\mathbf{f}_{k+j} = \mathbf{f}(x_{k+j}, \mathbf{y}_{k+j})$ para $0 \leq j \leq N$.

La expresión (3.1.1) se conoce como **ecuación de diferencias lineal**, cuya solución es una sucesión \mathbf{y}_n , más información de este tipo de ecuaciones se presenta en el Apéndice A.3. El método definido por esta ecuación es implícito cuando $\beta_N \neq 0$ y explícito cuando $\beta_N = 0$. Sin perdida de generalidad, se asume que $\alpha_N = 1$ dado que en caso contrario se podría despejar. Es importante señalar que para aplicar el MPML (3.1.1) los valores iniciales \mathbf{y}_{k+j} para $0 \leq j \leq N - 1$ deben ser conocidos.

Existen diferentes formas para obtener MPML. A continuación se muestran algunas técnicas para deducir estos métodos.

3.2. Deducción por series de Taylor

Para obtener un MPML a partir de series de Taylor se debe suponer que se conoce la solución exacta del problema de Cauchy (1.2.11) y trabajar con la expresión obtenida al desarrollar dicha solución en series de Taylor. Por lo tanto, se debe suponer que \mathbf{f} es continua y diferenciable en sus argumentos x e \mathbf{y} . La deducción del método de Euler (1.3.3) es un ejemplo de esta técnica, pues el método de Euler es un MPML de un paso.

Para deducir un MPML de dos pasos consideremos la expansión de Taylor para $\mathbf{y}(x_k + h)$ y $\mathbf{y}(x_k - h)$ en torno al punto x_k dadas respectivamente por

$$\mathbf{y}(x_k + h) = \mathbf{y}(x_k) + h\mathbf{y}'(x_k) + \frac{h^2}{2!}\mathbf{y}''(x_k) + \frac{h^3}{3!}\mathbf{y}'''(\xi_1), \quad (3.2.1)$$

$$\mathbf{y}(x_k - h) = \mathbf{y}(x_k) - h\mathbf{y}'(x_k) + \frac{h^2}{2!}\mathbf{y}''(x_k) - \frac{h^3}{3!}\mathbf{y}'''(\xi_2), \quad (3.2.2)$$

donde $x_k < \xi_1 < x_{k+1}$ y $x_{k-1} < \xi_2 < x_k$.

Restando las expresiones (3.2.1) y (3.2.2), se tiene

$$\mathbf{y}(x_k + h) - \mathbf{y}(x_k - h) = 2h\mathbf{y}'(x_k) + \frac{h^3}{3} [\mathbf{y}'''(\xi_1) + \mathbf{y}'''(\xi_2)].$$

Si $\mathbf{y}'''(x)$ está acotada por M para todo $x \in [a, b]$, se tiene

$$\mathbf{y}(x_k + h) - \mathbf{y}(x_k - h) \leq 2h\mathbf{y}'(x_k) + 2M \frac{h^3}{3}. \quad (3.2.3)$$

Para valores de h pequeños, el término que contiene h^3 en (3.2.3) que se refiere al error local, se puede omitir de lo cual se obtiene

$$\mathbf{y}_{k+1} - \mathbf{y}_{k-1} = 2h\mathbf{f}_k.$$

Este método se conoce como **método del punto medio** y puede ser escrito en la forma general (3.1.1) mediante la sustitución de k por $k + 1$, de lo cual se obtiene

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_1 \text{ conocido} \\ \mathbf{y}_{k+2} = \mathbf{y}_k + 2h\mathbf{f}_{k+1}. \end{cases}$$

En la expresión (3.2.3) se puede observar que el error local de discretización para el método del punto medio es

$$\tau_k = 2M \frac{h^3}{3}.$$

Otra técnica que permite obtener MPML a partir de las series de Taylor consiste en considerar la forma general (3.1.1) y mediante series de Taylor obtener expresiones que permitan calcular los coeficientes α_j y β_j para $j = 0, 1, \dots, N$. Por ejemplo, para deducir un MPML de un paso se considera la forma general

$$\mathbf{y}_{k+1} + \alpha_0 \mathbf{y}_k = h(\beta_1 \mathbf{f}_{k+1} + \beta_0 \mathbf{f}_k).$$

Suponiendo que se conoce la solución exacta del problema de Cauchy (1.2.11), $\mathbf{y}(x)$, y dado que $\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y})$ se cumple que

$$\mathbf{y}(x_k + h) + \alpha_0 \mathbf{y}(x_k) \approx h [\beta_1 \mathbf{y}'(x_k + h) + \beta_0 \mathbf{y}'(x_k)]. \quad (3.2.4)$$

Expandiendo en series de Taylor $\mathbf{y}(x_k + h)$ y $\mathbf{y}'(x_k + h)$ en torno a x_k se tiene

$$\mathbf{y}(x_k + h) = \mathbf{y}(x_k) + h\mathbf{y}'(x_k) + \frac{h^2}{2!}\mathbf{y}''(x_k) + \frac{h^3}{3!}\mathbf{y}'''(x_k) + \frac{h^4}{3!}\mathbf{y}^{(4)}(\xi_1), \quad (3.2.5)$$

$$\mathbf{y}'(x_k + h) = \mathbf{y}'(x_k) + h\mathbf{y}''(x_k) + \frac{h^2}{2!}\mathbf{y}'''(x_k) + \frac{h^3}{3!}\mathbf{y}^{(4)}(\xi_2), \quad (3.2.6)$$

donde $x_k < \xi_1 < x_{k+1}$ y $x_k < \xi_2 < x_{k+1}$.

Cuando la cuarta derivada de y esta acotada en $[a, b]$ y h es pequeño, el error local se puede omitir. Luego sustituyendo (3.2.5) y (3.2.6) en (3.2.4) y agrupando términos se obtiene la expresión

$$C_0\mathbf{y}(x_k) + C_1h\mathbf{y}'(x_k) + C_2h^2\mathbf{y}''(x_k) + C_3h^3y'''(x_k) \approx 0, \quad (3.2.7)$$

donde

$$C_0 = 1 + \alpha_0, \quad C_1 = 1 - \beta_1 - \beta_0, \quad C_2 = \frac{1}{2} - \beta_1, \quad C_3 = \frac{1}{6} - \frac{1}{2}\beta_1. \quad (3.2.8)$$

Para que se cumpla la expresión (3.2.7) se debe hacer que los coeficientes de esta expresión se anulen, es decir que $C_0 = 0$, $C_1 = 0$, $C_2 = 0$ y $C_3 = 0$. Por lo tanto, de las tres primeras expresiones en (3.2.8) se concluye que $\alpha_0 = -1$ y $\beta_0 = \beta_1 = \frac{1}{2}$. De donde C_3 toma el valor de $\frac{1}{12}$. De lo anterior se obtiene el MPML conocido como el método del trapecio implícito. Este método se presentó en la ecuación (1.3.12) del Capítulo 1 como un ejemplo de métodos implícitos.

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0) \\ \mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{2}(\mathbf{f}_{k+1} + \mathbf{f}_k). \end{cases} \quad (3.2.9)$$

Observe que el coeficiente C_3 no se ha anulado. De aquí se concluye que el error local para este método depende del término

$$\frac{1}{12}h^3y'''(x_k).$$

A continuación se presenta una técnica para deducir MPM usando integración numérica. Esta técnica da origen a una importante familia de métodos conocidos como los métodos de Adams.

3.3. Integración numérica

Integrando el problema de Cauchy $\mathbf{y}' = \mathbf{f}(x, \mathbf{y}(x))$ entre x_k y x_{k+j} para $j \geq 1$ se obtiene la expresión

$$\mathbf{y}(x_{k+j}) - \mathbf{y}(x_k) = \int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x))dx. \quad (3.3.1)$$

El objetivo es usar método de integración numérica para encontrar aproximaciones a la integral en (3.3.1) para diferentes valores de j y de esto derivar MPML.

Considerando la integral en (3.3.1)

$$\int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x))dx, \quad (3.3.2)$$

la idea básica de los métodos de integración numérica está en aproximar la función $\mathbf{f}(x, \mathbf{y}(x))$ de la integral (3.3.2) por un polinomio interpolador $\mathbf{P}(x)$ y determinar analíticamente la integral de ese

polinomio en el intervalo $[x_k, x_{k+j}]$. Las fórmulas de integración numérica se conocen como cuadraturas numéricas y son sumatorias cuyos términos son valores de la función $\mathbf{f}(x, \mathbf{y}(x))$ multiplicados por pesos w convenientemente elegidos, es decir

$$\int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x)) dx \approx \int_{x_k}^{x_{k+j}} \mathbf{P}(x) dx \approx \sum_{i=k}^j w_{k+i} \mathbf{f}(x_{k+i}, \mathbf{y}(x_{k+i})).$$

Usando para calcular $\mathbf{P}(x)$ el método de interpolación de Lagrange, presentado en el Apéndice A.1, se tiene

$$\int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x)) dx = \int_{x_k}^{x_{k+j}} \sum_{i=k}^j \mathbf{f}(x_{k+i}, \mathbf{y}(x_{k+i})) L_{k+i}(x) dx + \int_{x_k}^{x_{k+j}} \prod_{i=k}^j (x - x_{k+i}) \frac{\mathbf{f}^{(j+1)}(\varsigma(x))}{(n+1)!} dx, \quad (3.3.3)$$

es decir

$$\int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x)) dx = \sum_{i=k}^j w_{k+i} \mathbf{f}(x_{k+i}, \mathbf{y}(x_{k+i})) + \frac{1}{(n+1)!} \int_{x_k}^{x_{k+j}} \prod_{i=k}^j (x - x_{k+i}) \mathbf{f}^{(j+1)}(\varsigma(x)) dx,$$

donde $\varsigma(x)$ y x están en $[x_k, x_{k+j}]$ y para cada x y

$$w_{k+i} = \int_{x_k}^{x_{k+j}} L_{k+i}(x) dx, \text{ para } 0 \leq i \leq j.$$

Por lo tanto, la fórmula de cuadratura es

$$\int_{x_k}^{x_{k+j}} \mathbf{f}(x, \mathbf{y}(x)) dx \approx \sum_{i=k}^j w_{k+i} \mathbf{f}(x_{k+i}, \mathbf{y}(x_{k+i})),$$

con un error dado por

$$\frac{1}{(n+1)!} \int_{x_k}^{x_{k+j}} \prod_{i=k}^j (x - x_{k+i}) \mathbf{f}^{(j+1)}(\varsigma(x)) dx.$$

Ejemplo 3.1. Si tomamos $j = 1$ de la integral (3.3.3) se tiene

$$\begin{aligned} \int_{x_k}^{x_{k+1}} \mathbf{f}(x, \mathbf{y}(x)) dx &= \int_{x_k}^{x_{k+1}} \frac{(x - x_{k+1})}{(x_k - x_{k+1})} \mathbf{f}(x_k) dx + \int_{x_k}^{x_{k+1}} \frac{(x - x_k)}{(x_{k+1} - x_k)} \mathbf{f}(x_{k+1}) dx \\ &\quad + \frac{1}{2} \int_{x_k}^{x_{k+1}} \mathbf{f}''(\varsigma(x))(x - x_k)(x - x_{k+1}) dx. \end{aligned}$$

Donde aplicando el teorema del valor medio al término del error, se tiene que para un $\varsigma \in (x_k, x_{k+1})$

$$\begin{aligned} \int_{x_k}^{x_{k+1}} \mathbf{f}''(\varsigma(x))(x - x_k)(x - x_{k+1}) dx &= \mathbf{f}''(\varsigma) \int_{x_k}^{x_{k+1}} (x - x_k)(x - x_{k+1}) dx \\ &= \mathbf{f}''(\varsigma) \left[\frac{x^3}{3} - \frac{(x_{k+1} + x_k)}{2} x^2 + x_k x_{k+1} x \right]_{x_k}^{x_{k+1}} \\ &= -\frac{h^3}{6} \mathbf{f}''(\varsigma). \end{aligned}$$

Por lo tanto

$$\begin{aligned}\int_{x_k}^{x_{k+1}} \mathbf{f}(x, \mathbf{y}(x)) dx &= \left[\frac{(x - x_{k+1})^2}{2(x_k - x_{k+1})} \mathbf{f}(x_k) + \frac{(x - x_k)^2}{2(x_{k+1} - x_k)} \mathbf{f}(x_{k+1}) \right]_{x_k}^{x_{k+1}} \\ &= \frac{(x_{k+1} - x_k)}{2} [\mathbf{f}(x_k, \mathbf{y}(x_k)) + \mathbf{f}(x_{k+1}, \mathbf{y}(x_{k+1}))].\end{aligned}$$

Dado que $h = x_{k+1} - x_k$, se tiene

$$\int_{x_k}^{x_{k+1}} \mathbf{f}(x, \mathbf{y}(x)) dx = \frac{h}{2} [\mathbf{f}(x_k, \mathbf{y}(x_k)) + \mathbf{f}(x_{k+1}, \mathbf{y}(x_{k+1}))] - \frac{h^3}{12} \mathbf{f}''(\varsigma). \quad (3.3.4)$$

La expresión (3.3.4) se conoce como la **regla del trapecio** para aproximar integrales.

De manera similar se puede obtener otros métodos de integración numérica, si se toma $j = 2$ en (3.3.3) y se aplica el polinomio de Lagrange que interpola los puntos $(x_k, f(x_k))$, $(x_{k+1}, f(x_{k+1}))$ y $(x_{k+2}, f(x_{k+2}))$, se obtiene la **regla de Simpson** para aproximar integrales

$$\int_{x_k}^{x_{k+2}} \mathbf{f}(x, \mathbf{y}(x)) dx = \frac{h}{3} [\mathbf{f}(x_k, \mathbf{y}(x_k)) + 4\mathbf{f}(x_{k+1}, \mathbf{y}(x_{k+1})) + \mathbf{f}(x_{k+2}, \mathbf{y}(x_{k+2}))] - \frac{h^5}{90} \mathbf{f}^{(4)}(\varsigma). \quad (3.3.5)$$

□

La idea es usar métodos de integración numérica para aproximar la integral en (3.3.1). Usando la regla del trapecio se obtiene nuevamente el método del trapecio implícito (3.2.9). Por otro lado aplicando la regla de Simpson se tiene

$$\mathbf{y}(x_{k+2}) - \mathbf{y}(x_k) = \frac{h}{3} [\mathbf{f}(x_k, \mathbf{y}(x_k)) + 4\mathbf{f}(x_{k+1}, \mathbf{y}(x_{k+1})) + \mathbf{f}(x_{k+2}, \mathbf{y}(x_{k+2}))] - \frac{h^5}{90} \mathbf{f}^{(4)}(\varsigma),$$

de donde se obtiene el MPML de dos paso

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_1 \text{ conocido} \\ \mathbf{y}_{k+2} = \mathbf{y}_k + \frac{h}{3} [\mathbf{f}_k + 4\mathbf{f}_{k+1} + \mathbf{f}_{k+2}], \end{cases}$$

que es comúnmente conocido como el método de Simpson implícito. Este método se presentó en la ecuación (1.3.13) del Capítulo 1 como un ejemplo de métodos implícitos.

3.4. Métodos de Adams

Entre los métodos deducidos por integración numérica se destaca la familia de métodos de Adams, los cuales se deducen de la expresión

$$\mathbf{y}_{k+1} - \mathbf{y}_k = \int_{x_k}^{x_{k+1}} \mathbf{P}(x) dx, \quad (3.4.1)$$

donde $\mathbf{P}(x)$ es un polinomio que interpola a $\mathbf{f}(x, \mathbf{y}(x))$.

Si los métodos obtenidos de (3.4.1) son explícitos se conocen como métodos de Adams Bashforth, y si son implícitos, se conocen como métodos de Adams Moulton.

Para obtener el método de Adams Bashforth de 2 pasos se debe aproximar la función $\mathbf{f}(x, \mathbf{y}(x))$ por el polinomio de Lagrange de grado 1 que pasa por los puntos $(x_{k-1}, \mathbf{y}_{k-1})$ y (x_k, \mathbf{y}_k) , es decir el polinomio

$$\mathbf{P}_1(x) = \frac{x - x_k}{x_{k-1} - x_k} \mathbf{f}_{k-1} + \frac{x - x_{k-1}}{x_k - x_{k-1}} \mathbf{f}_k. \quad (3.4.2)$$

Tomando

$$u = \frac{x - x_{k-1}}{h} \text{ y } dx = hdu, \quad (3.4.3)$$

se sigue que $x - x_{k-1} = hu$ y $x - x_k = x - x_{k-1} - h = h(u - 1)$. Sustituyendo estos valores en (3.4.2), se tiene

$$\mathbf{P}_1(x) = u\mathbf{f}_k + (1 - u)\mathbf{f}_{k-1}.$$

Por lo tanto, de la ecuación (3.4.1) se obtiene

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \int_{x_k}^{x_{k+1}} \mathbf{P}_1(x) dx.$$

Así

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{y}_k + \int_1^2 [u\mathbf{f}_k + (1 - u)\mathbf{f}_{k-1}] hdu \\ &= \mathbf{y}_k + h \left[\left(\frac{u^2}{2} \right) \mathbf{f}_k + \left(u - \frac{u^2}{2} \right) \mathbf{f}_{k-1} \right]_1^2 \\ &= \mathbf{y}_k + h \left(\frac{3}{2}\mathbf{f}_k - \frac{1}{2}\mathbf{f}_{k-1} \right). \end{aligned}$$

De aquí se obtiene la fórmula explícita de Adams Bashforth de 2 pasos (AB2)

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{2}(3\mathbf{f}_k - \mathbf{f}_{k-1}).$$

Este método puede ser escrito en la forma general (3.1.1) mediante la sustitución de k por $k + 1$, de lo cual se obtiene

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_1 \text{ conocido} \\ \mathbf{y}_{k+2} = \mathbf{y}_{k+1} + \frac{h}{2}(3\mathbf{f}_{k+1} - \mathbf{f}_k). \end{cases} \quad (3.4.4)$$

El método (3.4.4) fue obtenido de la integración del polinomio $\mathbf{P}_1(x)$ en el intervalo $[x_k, x_{k+1}]$, donde $\mathbf{P}_1(x)$ interpola los puntos $(x_{k-1}, \mathbf{f}_{k-1})$ y (x_k, \mathbf{f}_k) . A fin de obtener un método que ofrezca

mejores resultados se repite el proceso anterior usando el polinomio $\mathbf{P}_2(x)$ que interpola los puntos $(x_{k-1}, \mathbf{f}_{k-1})$, (x_k, \mathbf{f}_k) y $(x_{k+1}, \mathbf{f}_{k+1})$, por tanto

$$\mathbf{P}_2(x) = \frac{(x-x_k)(x-x_{k+1})}{(x_{k-1}-x_k)(x_{k-1}-x_{k+1})} \mathbf{f}_{k-1} + \frac{(x-x_{k-1})(x-x_{k+1})}{(x_k-x_{k-1})(x_k-x_{k+1})} \mathbf{f}_k + \frac{(x-x_{k-1})(x-x_k)}{(x_{k+1}-x_{k-1})(x_{k+1}-x_k)} \mathbf{f}_{k+1}. \quad (3.4.5)$$

Nuevamente considerando el cambio de variable (3.4.3) y sus consideraciones, al sustituir en (3.4.5) se concluye que

$$\mathbf{P}_2(x) = \frac{u(u-1)}{2} \mathbf{f}_{k+1} - u(u-2) \mathbf{f}_k + \frac{(u-1)(u-2)}{2} \mathbf{f}_{(k-1)}.$$

Por lo tanto de la expresión (3.4.1) se tiene

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \int_{x_k}^{x_{k+1}} \mathbf{P}_2(x) dx.$$

Así

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{y}_k + \int_1^2 \left[\frac{u(u-1)}{2} \mathbf{f}_{k+1} - u(u-2) \mathbf{f}_k + \frac{(u-1)(u-2)}{2} \mathbf{f}_{(k-1)} \right] h du \\ &= \mathbf{y}_k + h \left[\left(\frac{u^3}{6} - \frac{u^2}{4} \right) \mathbf{f}_{k+1} - \left(\frac{u^3}{3} - u^2 \right) \mathbf{f}_k + \left(\frac{u^3}{6} - \frac{3u^2}{4} + u \right) \mathbf{f}_{(k-1)} \right]_1^2 \\ &= \mathbf{y}_k + \frac{h}{12} (5\mathbf{f}_{k+1} + 8\mathbf{f}_k - \mathbf{f}_{k-1}). \end{aligned}$$

Por lo tanto se obtiene la fórmula implícita del método de Adams Moulton de 2 pasos

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \frac{h}{12} (5\mathbf{f}_{k+1} + 8\mathbf{f}_k - \mathbf{f}_{k-1}).$$

Este método puede ser escrito en la forma general (3.1.1) mediante la sustitución de k por $k+1$, de lo cual se obtiene

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_1 \text{ conocido} \\ \mathbf{y}_{k+2} = \mathbf{y}_{k+1} + \frac{h}{12} (5\mathbf{f}_{k+2} + 8\mathbf{f}_{k+1} - \mathbf{f}_k). \end{cases}$$

De forma similar se obtienen el método de Adams Bashforth de 3 pasos y los métodos de Adams Bashforth y Moulton de 4 pasos que son de frecuente uso en la práctica.

Método de Adams Bashforth 3 pasos (AB3)

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_p \text{ conocido para } 1 \leq p \leq 2 \\ \mathbf{y}_{k+3} = \mathbf{y}_{k+2} + \frac{h}{12} [23\mathbf{f}_{k+2} - 16\mathbf{f}_{k+1} + 5\mathbf{f}_k]. \end{cases} \quad (3.4.6)$$

Método de Adams Bashforth 4 pasos (AB4)

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_p \text{ conocido para } 1 \leq p \leq 3 \\ \mathbf{y}_{k+4} = \mathbf{y}_{k+3} + \frac{h}{24} [55\mathbf{f}_{k+3} + 59\mathbf{f}_{k+2} + 37\mathbf{f}_{k+1} - 9\mathbf{f}_k]. \end{cases} \quad (3.4.7)$$

Método de Adams Moulton 4 pasos

$$\begin{cases} \mathbf{y}_0 = \mathbf{y}(x_0), & \mathbf{y}_p \text{ conocido para } 1 \leq p \leq 3 \\ \mathbf{y}_{k+4} = \mathbf{y}_{k+3} + \frac{h}{720} [251\mathbf{f}_{k+4} + 646\mathbf{f}_{k+3} - 264\mathbf{f}_{k+2} + 106\mathbf{f}_{k+1} - 19\mathbf{f}_k]. \end{cases}$$

3.5. Consistencia, convergencia y estabilidad

En esta sección se describe las propiedades de consistencia, convergencia y estabilidad para los MPML, análogamente como se hizo en el Capítulo 2 para los MPU.

Definición 3.1. Para el MPML (3.1.1) de N pasos, el **error local de discretización** τ_k está dado por

$$\tau_k = \frac{\sum_{j=0}^N \alpha_j \mathbf{y}(x_k + jh)}{h} - \sum_{j=0}^N \beta_j \mathbf{f}(x_k + jh, \mathbf{y}(x_k + jh)). \quad (3.5.1)$$

Observación 3.1. La expresión (3.5.1) se puede escribir como

$$h\tau_k = \sum_{j=0}^N \alpha_j \mathbf{y}(x_k + jh) - h \sum_{j=0}^N \beta_j \mathbf{y}'(x_k + jh). \quad (3.5.2)$$

Si se expande $\mathbf{y}(x_k + jh)$ y su derivada $\mathbf{y}'(x_k + jh)$ en series de Taylor al rededor de x_k , se tiene

$$\mathbf{y}(x_k + jh) = \mathbf{y}(x_k) + jh\mathbf{y}'(x_k) + \frac{j^2h^2}{2!}\mathbf{y}''(x_k) + \dots + \frac{j^ph^p}{p!}\mathbf{y}^{(p)}(x_k) + \dots \quad (3.5.3)$$

$$\mathbf{y}'(x_k + jh) = \mathbf{y}'(x_k) + jh\mathbf{y}''(x_k) + \frac{j^2h^2}{2!}\mathbf{y}'''(x_k) + \dots + \frac{j^{p-1}h^{p-1}}{(p-1)!}\mathbf{y}^{(p)}(x_k) + \dots \quad (3.5.4)$$

Remplazando (3.5.3) y (3.5.4) en (3.5.2), y agrupando términos se tiene la expresión

$$h\tau_k = C_0 \mathbf{y}(x_k) + C_1 h \mathbf{y}'(x_k) + C_2 h^2 \mathbf{y}''(x_k) + \dots + C_p h^p \mathbf{y}^{(p)}(x_k) + \dots, \quad (3.5.5)$$

donde los coeficientes constantes C_i son

$$\begin{aligned} C_0 &= \sum_{j=0}^N \alpha_j, \\ C_1 &= \sum_{j=0}^N j\alpha_j - \sum_{j=0}^n \beta_j, \\ C_2 &= \sum_{j=0}^N \frac{j^2\alpha_j}{2!} - \sum_{j=0}^n j\beta_j, \\ &\vdots \\ C_p &= \sum_{j=0}^N \frac{j^p\alpha_j}{p!} - \sum_{j=0}^N \frac{j^{p-1}\beta_j}{(p-1)!}. \end{aligned}$$

Definición 3.2. El MPML (3.1.1) de N pasos es **consistente** con el problema de Cauchy (1.2.11) si y solo si

$$\lim_{h \rightarrow 0} \|\boldsymbol{\tau}_k\| = 0.$$

Usando la Observación 3.1 se puede decir que el MPML es consistente si y solo si $C_0 = 0$ y $C_1 = 0$, ya que para la consistencia necesitamos que $\|\boldsymbol{\tau}_k\| \rightarrow 0$ cuando $h \rightarrow 0$.

Definición 3.3. El MPML (3.1.1) de N pasos tiene **orden** q si y solo si existen constantes positivas Q , h_0 y q , independientes del paso de integración h y del subíndice k , con $0 < h \leq h_0$, tales que el error local de discretización satisface

$$\max_k \|\boldsymbol{\tau}_k\| \leq Qh^q.$$

Usando la Observación 3.1, el MPML (3.1.1) tiene orden q si y solo si

$$C_0 = C_1 = \dots = C_q = 0 \text{ y } C_{q+1} \neq 0.$$

Observe que para asegurar que el MPML (3.1.1) es consistente basta probar que el método tiene orden de consistencia $q \geq 1$, pues al ser el método de al menos orden 1 se verifica que $C_0 = C_1 = 0$, lo cual es equivalente a demostrar que el método es consistente.

Definición 3.4. El MPML (3.1.1) de N pasos es **convergente** si para todo problema de Cauchy (1.2.11), se tiene que

$$\lim_{h \rightarrow 0} \|\mathbf{y}_{k+N}\| = \|\mathbf{y}(x_{k+N})\|,$$

para todo $x \in [a, b]$, y para toda solución $\{\mathbf{y}_{k+N}\}$ de la ecuación de diferencias, que satisfacen las condiciones iniciales $\|\mathbf{y}_i\| = \|\boldsymbol{\eta}_i(h)\|$, donde $\lim_{h \rightarrow 0} \|\boldsymbol{\eta}_i(h)\| = \|\mathbf{y}_0\|$ para $i = 0, 1, \dots, N - 1$.

En el capítulo anterior se observó que la consistencia era una condición necesaria y suficiente para asegurar la convergencia, la estabilidad solo tenía lugar en el caso en el que se consideraba que la condición inicial \mathbf{y}_0 del problema de Cauchy (1.2.11) estuviera perturbada y aun así la consistencia garantizaba la estabilidad. En este capítulo se mostrará que para asegurar la convergencia de un MPML es necesario garantizar consistencia y estabilidad.

Para los MPML el estudio de estabilidad cobra mayor relevancia que para los MPU, pues para un MPML de N pasos (3.1.1), \mathbf{y}_0 está dada por la condición inicial del problema de Cauchy (1.2.11), pero los demás valores de inicio $\mathbf{y}_1, \dots, \mathbf{y}_{N-1}$, usualmente son calculados aplicando MPU, usando herramientas estadísticas e incluso de datos experimentales. En cualquiera caso, generalmente los valores iniciales están sujetos a grandes errores numéricos, en este sentido es importante saber cómo estos errores afectarán a las aproximaciones \mathbf{y}_k para $N \leq k$, obtenidas por un MPML de N pasos (3.1.1). Por lo tanto en este caso se estudiará la estabilidad de un MPML con respecto a pequeñas perturbaciones en las condiciones de partida.

Definición 3.5. Se dice que el MPML (3.1.1) es **estable** si existe una constante K tal que, para cualquier par de sucesiones $\{\mathbf{y}_{k+N}\}$ y $\{\hat{\mathbf{y}}_{k+N}\}$ las cuales han sido generadas por la aplicación de (3.1.1) pero con diferentes valores de inicio $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{N-1}$ y $\hat{\mathbf{y}}_0, \hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_{N-1}$, respectivamente, se tiene que

$$\|\mathbf{y}_{k+N} - \hat{\mathbf{y}}_{k+N}\| \leq K \max \{\|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|, \|\mathbf{y}_1 - \hat{\mathbf{y}}_1\|, \dots, \|\mathbf{y}_{N-1} - \hat{\mathbf{y}}_{N-1}\|\},$$

para h que tiende a 0.

A pesar de que la Definición (3.5) permite entender intuitivamente la idea de estabilidad en el sentido en que pequeñas perturbaciones en las condiciones iniciales dan lugar a pequeñas perturbaciones en los datos de salida, es muy tedioso usar esta definición para verificar la estabilidad de un MPML. Por esta razón estudiaremos técnicas alternativas más sencillas que permiten verificar la estabilidad de un MPML. Antes de empezar es necesario definir algunos conceptos, los cuales se enuncian para el caso escalar.

Definición 3.6. Dado el MPML,

$$\sum_{j=0}^N \alpha_j y_{k+j} = h \sum_{j=0}^N \beta_j f_{k+j},$$

se conocen como el **primer y segundo polinomio característico** del MPML a los polinomios $\rho(z)$ y $\sigma(z)$ respectivamente, donde

$$\begin{aligned} \rho(z) &= \sum_{j=0}^N \alpha_j z^j, \\ \sigma(z) &= \sum_{j=0}^n \beta_j z^j, \end{aligned}$$

con $\alpha_N = 1$.

Usando la Definición 3.6, se presenta a seguir algunos teoremas que permiten verificar fácilmente la consistencia y estabilidad para un MPML.

Teorema 3.1. *El MPML (3.1.1) es consistente si y solo si*

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1).$$

Demostración. El Teorema 3.1 resulta evidente, pues

$$0 = \rho(1) = \alpha_0 + \alpha_1 + \dots + \alpha_n = C_0,$$

es decir $C_0 = 0$.

Por otro lado

$$\rho'(1) = \sum_{j=0}^N j \alpha_j \quad y \quad \sigma(1) = \sum_{j=0}^N \beta_j.$$

Por lo tanto usando la condición del Teorema 3.1, $\rho'(1) = \sigma(1)$, se tiene

$$0 = \rho'(1) - \sigma(1) = \sum_{j=0}^N j\alpha_j - \sum_{j=0}^N \beta_j = C_1.$$

Se concluye que si se cumplen las hipótesis del Teorema 3.1, se tiene que $C_0 = 0$ y $C_1 = 1$, por lo tanto el MPML es consistente. \square

Ejemplo 3.2. Considere el problema de Cauchy

$$\begin{cases} y'(x) = 0 \\ y(x_0) = 0, \end{cases} \quad (3.5.6)$$

cuya solución exacta es $y(x) = 0$. Aplicando el MPML (3.1.1) al problema (3.5.6), se tiene la ecuación en diferencias homogénea

$$\sum_{j=0}^N \alpha_j y_{n+j} = 0. \quad (3.5.7)$$

La expresión (3.5.7) es una ecuación de diferencias homogénea (ver Apéndice A.3), por lo tanto para encontrar la solución de esta ecuación se deben considerar las raíces del polinomio

$$\rho_n(z) = \sum_{j=0}^N \alpha_j z^j,$$

el cual coincide con el primer polinomio característico de la Definición 3.6. Si se considera que todas las raíces de este polinomio son distintas, entonces la solución de (3.5.7) es de la forma

$$y_k = d_1 r_1^k + d_2 r_2^k + \dots + d_N r_N^k,$$

donde d_i , $1 \leq i \leq N$ son constantes y recordamos que se conoce una raíz del polinomio, la raíz principal $z_1 = 1$. Si se supone que el MPML que se aplicó es convergente, se cumple que

$$y_i \rightarrow y(0) = 0 \text{ cuando } h \rightarrow 0, \text{ para } i = 0, 1, \dots, N-1.$$

Por lo tanto se considera y_k de la forma

$$y_k = h(d_1 r_1^k + d_2 r_2^k + \dots + d_N r_N^k).$$

Dado que el MPML es convergente se cumple que $y_k \rightarrow 0$ cuando $h \rightarrow 0$ o cuando $n \rightarrow \infty$, siendo $x = nh$. Sin embargo dado que

$$\lim_{h \rightarrow 0} h r_j^k = \lim_{k \rightarrow \infty} \frac{x}{k} r_j^k = x \lim_{k \rightarrow \infty} \frac{r_j^k}{k} = 0 \text{ si y solo si } |r_j| \leq 1,$$

es claro que el método no será convergente si alguna de las raíces z_1 tiene modulo mayor que uno. Si se considera el caso en el que las raíces del polinomio $\rho(z)$ tienen multiplicidad $m > 1$ de manera similar al caso anterior, se concluye del Apéndice A.3, la solución de (3.5.7) tiene la forma

$$\begin{aligned} y_k = & [d_{1,1} + d_{1,2}k + \dots + d_{1,m_1}k(k-1)(k-2)\dots(k-m_1+2)] hr_1^k + \\ & + [d_{2,1} + d_{2,2}k + \dots + d_{2,m_2}k(k-1)(k-2)\dots(k-m_2+2)] hr_2^k + \\ & \vdots \\ & + [d_{j,1} + d_{j,2}k + \dots + d_{j,m_j}k(k-1)(k-2)\dots(k-m_j+2)] hr_j^k, \end{aligned}$$

donde $\sum_{x=1}^j x_x = N$ y

$$\lim_{h \rightarrow 0} hk^j r_j^k = \lim_{k \rightarrow \infty} \frac{x}{k} k^j r_j^k = x \lim_{k \rightarrow \infty} k^{j-1} r_j^k. \quad (3.5.8)$$

El método será convergente si y solo si el límite (3.5.8) es cero, esto ocurre cuando $|r_j| < 1$. \square

El ejemplo anterior motiva la siguiente definición de estabilidad

Definición 3.7. Se dice que el MPML (3.1.1) es **estable** si ninguna raíz del primer polinomio característico

$$\rho(z) = \sum_{j=0}^N \alpha_j z^j,$$

tiene modulo mayor que 1 y toda raíz con modulo 1 tiene multiplicidad 1.

La Definición 3.7 usualmente se conoce como **condición de la raíz** y ofrece una técnica sencilla para verificar la estabilidad de un MPML.

Observación 3.2. El Teorema 3.1 establece que para que un MPML sea consistente, se debe cumplir que el primer polinomio característico del método, $\rho(z)$, siempre tiene que tener la raíz $z = 1$. Esta raíz se denomina **raíz principal** y será de suma importancia en el estudio de la estabilidad absoluta para MPML, en adelante esta raíz será denotada por z_1 .

Observación 3.3. Si se considera un MPU consistente, se tiene que el polinomio característico para este método es de grado 1 y la única raíz de este polinomio es $z = 1$. Por lo tanto, por la Definición 3.7 de estabilidad, si un MPU es consistente entonces es estable.

Teorema 3.2. *La consistencia y estabilidad del MPML (3.1.1) son condiciones necesarias y suficientes para que el método sea convergente.*

La complejidad de la demostración del Teorema 3.2 va más allá de la teoría tratada en este trabajo, por lo tanto no se incluye pero puede ser consultada en [7].

3.6. Estabilidad absoluta

La estabilidad absoluta de los MPML al igual que en el caso de MPU tiene que ver con el problema de elegir un tamaño de paso adecuado, ya que a pesar de que la teoría asegura buenos resultados para tamaño de paso h que sean próximos a cero, en la práctica tomar tamaños de paso pequeños supone una gran cantidad de cálculos, lo cual genera altos costos computacionales y grandes errores de redondeo. Por lo tanto, en este capítulo se presentan algunas técnicas que permiten calcular tamaños de paso para MPML que aseguren buenas aproximaciones numéricas.

Análogo al estudio de estabilidad para los MPU en (2.5.3), la estabilidad absoluta para los MPML se estudia en un problema de Cauchy escalar de la forma

$$y'(x) = \lambda y(x). \quad (3.6.1)$$

Aplicando un MPML como (3.1.1) al problema (3.6.1), se tiene

$$\sum_{j=0}^N \alpha_j y_{k+j} = \lambda h \sum_{j=0}^N \beta_j y_{k+j},$$

es decir

$$\sum_{j=0}^N \alpha_j y_{k+j} - \lambda h \sum_{j=0}^N \beta_j y_{k+j} = 0. \quad (3.6.2)$$

Por otro lado calculando el error local de discretización, se tiene

$$\tau_k = \frac{\sum_{j=0}^N \alpha_j y(x_k + jh)}{h} - \lambda \sum_{j=0}^N \beta_j \lambda y_{k+j}. \quad (3.6.3)$$

Si se multiplica (3.6.3) por h y a este resultado se le suma la expresión (3.6.2), se tiene,

$$\begin{aligned} h\tau_k &= \sum_{j=0}^N \alpha_j y(x_k + jh) - \lambda h \sum_{j=0}^N \beta_j y_{k+j} - \sum_{j=0}^N \alpha_j y_{k+j} + \lambda h \sum_{j=0}^N \beta_j y_{k+j} \\ &= \left[\sum_{j=0}^N \alpha_j y(x_k + jh) - \sum_{j=0}^N \alpha_j y_{k+j} \right] - \lambda h \left[\sum_{j=0}^N \beta_j y_{k+j} - \sum_{j=0}^N \beta_j y_{k+j} \right] \\ &= \sum_{j=0}^N \alpha_j (y(x_k + jh) - y_{k+j}) - \lambda h \sum_{j=0}^N \beta_j (y(x_k + jh) - y_{k+j}) \\ &= \sum_{j=0}^N (\alpha_j - \lambda h \beta_j) (y(x_k + jh) - y_{k+j}). \end{aligned}$$

Si se toma $\bar{h} = \lambda h$ y se supone que $h\tau_k = \psi$ es constante, entonces se tiene la ecuación de diferencias

$$\sum_{j=0}^N (\alpha_j - \bar{h}\beta_j) e_{k+j} = \psi. \quad (3.6.4)$$

Por el Apéndice A.3, la solución general de (3.6.4) es una sucesión $\{e_n\}$, donde cada término es de la forma

$$e_k = \sum_{i=0}^N A\bar{r}_i^k + \psi,$$

donde \bar{z}_i es una raíz del polinomio $\sum_{j=0}^N (\alpha_j - \bar{h}\beta_j)z^j$ para $i = 0, 1, \dots, N$. Note que las raíces $\bar{z}_i \neq 1$, ya que si $\bar{z}_i = 1$ se tiene que

$$\sum_{j=0}^N (\alpha_j - \bar{h}\beta_j),$$

lo que implica que

$$\sum_{j=0}^N \alpha_j = \bar{h} \sum_{j=0}^N \beta_j. \quad (3.6.5)$$

Para que el MPML que se aplicó sea convergente se debe de cumplir que $\sum_{j=0}^N \alpha_j = 0$, por tanto de la ecuación (3.6.5) se tiene que

$$\sum_{j=0}^N \beta_j = 0 = \sum_{j=0}^N j\alpha_j,$$

por lo tanto si $\bar{z}_i = 1$, el método no será convergente.

Nuevamente del Apéndice A.3 se tiene que una solución particular de la ecuación de diferencias (3.6.4) es

$$\psi = \frac{\phi}{\sum_{j=0}^N (\alpha_j - \bar{h}\beta_j)}.$$

Pero

$$\sum_{j=0}^N \alpha_j = 0,$$

por tanto

$$\Psi = -\frac{\phi}{\bar{h} \sum_{j=0}^N \beta_j}.$$

Así, la solución general de (3.6.4) es

$$e_k = \sum_{i=0}^N A\bar{r}_i^k - \frac{\phi}{\bar{h} \sum_{j=0}^N \beta_j}. \quad (3.6.6)$$

Esta expresión permite conocer el comportamiento del error. Se observa que el error depende significativamente del primer sumando, de manera que si una raíz tiene modulo mayor que 1, el error crece si N crece.

Definición 3.8. Se denomina **polinomio de estabilidad absoluta** del MPML al polinomio

$$\pi(z, \bar{h}) = \rho(z) - \bar{h}\sigma(z) = \sum_{j=0}^N (\alpha_j - \bar{h}\beta_j)z^j = 0, \quad (3.6.7)$$

donde $\bar{h} = \lambda h$.

Definición 3.9. Se dice que el MPML es **absolutamente estable** para un valor de $\bar{h} = \lambda h$ dado si para este \bar{h} , todas las raíces \bar{z}_i del polinomio de estabilidad absoluta (3.6.7) satisfacen que

$$|\bar{z}_i| < 1,$$

para $i = 1, 2, \dots, N$. El conjunto

$$\Omega = \{\bar{h} \in \mathbb{C} / |\bar{z}_i| < 1 \text{ para toda raíz } z_i \text{ de } \pi(z, \bar{h})\},$$

se conoce como **región de estabilidad absoluta** y la intersección de esta región con el eje real se conoce como **intervalo de estabilidad absoluta**.

Ejemplo 3.3. Para el método de Adams Bashforth 2, el polinomio de estabilidad absoluta es

$$\pi(z, \bar{h}) = z^2 - \left(1 + \frac{3\bar{h}}{2}\right)z + \frac{\bar{h}}{2}. \quad (3.6.8)$$

En la Figura 3.1 se muestra la región de estabilidad para el método de Adams Bashforth 2, la cual fue diseñada usando Matlab. Inicialmente se calcularon las raíces del polinomio (3.6.8) para un conjunto de valores de \bar{h} en las vecindades del origen del plano complejo, luego para estas raíces se presentó gráficamente las que cumplieron con la condición $|\bar{z}_i| < 1$.

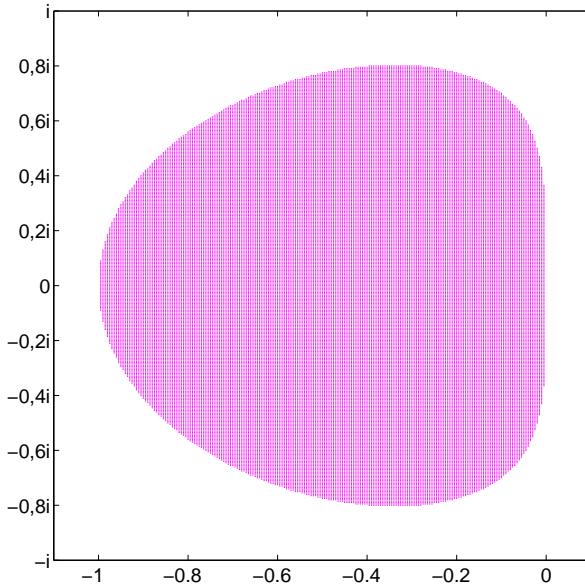


Figura 3.1: Región de estabilidad absoluta Adams Bashforth 2. □

Capítulo 4

Resultados Numéricos

En este capítulo se realizan diferentes simulaciones numéricas con el objetivo de verificar los resultados teóricos presentados en los capítulos anteriores respecto a consistencia, estabilidad y convergencia de los métodos estudiados al compararse con EDO a las que se les conoce la solución exacta. Además, se aplicarán algunos de estos métodos en un modelo para el VIH/SIDA descrito en [17], el cual no posee solución analítica.

4.1. Validación de implementaciones

Algunos de los problemas con los que se han realizado las pruebas numéricas son para EDO reales e imaginarias y otros para SEDOP, esto con el fin de verificar las propiedades teóricas de los métodos en cada caso. A seguir se presentan los problemas que se usarán para verificar las implementaciones realizadas en Matlab.

Problema 4.1.

$$\begin{cases} y'(x) = -y(x) + 2\cos(x), \\ y(0) = 1, \quad x \in [0, 1]. \end{cases}$$

Con solución exacta $y(x) = \sin(x) + \cos(x)$ y $y(1) = 1.3818$.

Problema 4.2.

$$\begin{cases} y'(x) = \cos(x)y(x), \\ y(0) = 1, \quad x \in [0, 20]. \end{cases}$$

Con solución exacta $y(x) = e^{\sin(x)}$ y $y(20) = 2.4916$.

Problema 4.3.

$$\begin{cases} y'(x) = -15y(x), \\ y(0) = 1, \quad x \in [0, 20]. \end{cases}$$

Con solución exacta $y(x) = e^{-15x}$ y $y(20) = 5.1482 \times 10^{-131}$.

Problema 4.4.

$$\begin{cases} y'(x) = y(x)^2 + 1, \\ y(0) = 0, \quad x \in [0, 1.2]. \end{cases}$$

Con solución exacta $y(x) = \tan(x)$ y $y(1.2) = 2.5721$.

Problema 4.5.

$$\begin{cases} y'(x) = (-1 + 2i)y(x), \\ y(0) = 1, \quad x \in [0, 20]. \end{cases}$$

Con solución exacta $y(x) = e^{(-1+2i)x}$ y $y(20) = -1.3746 \times 10^{-9} + 1.5357 \times 10^{-9}i$.

Problema 4.6.

$$\begin{cases} u'(x) = u(x) - 2v(x) + 4\cos(x) - 2\sin(x) \\ v'(x) = 3u(x) - 4v(x) + 5\cos(x) - 5\sin(x), \\ u(0) = 1, \quad v(0) = 2, \quad x \in [0, 1]. \end{cases}$$

Con solución exacta $u(x) = \sin(x) + \cos(x)$, $v(x) = 2\cos(x)$ y $u(1) = 1.3818$, $v(1) = 1.0806$.

Problema 4.7.

$$\begin{cases} u'(x) = -2u(x) + v(x) \\ v'(x) = u(x) - 2v(x), \\ u(0) = 0, \quad v(0) = 2, \quad x \in [0, 20]. \end{cases}$$

Con solución exacta $u(x) = e^{-x} - e^{-3x}$, $v(x) = e^{-x} + e^{-3x}$ y $u(20) = 2.0611 \times 10^{-9}$, $v(20) = 2.0611 \times 10^{-9}$.

En las pruebas numéricas se calculará el error global de discretización, el cual en el caso escalar es $e_k = |y(x_k) - y_k|$, es decir la diferencia entre la solución exacta y la solución aproximada. En el caso vectorial el error global es $e_k = \|\mathbf{y}(x_k) - \mathbf{y}_k\|_\infty$, donde se considera la norma infinito (1.3.8). Para calcular el orden de convergencia p se usa la expresión

$$p = \log_2 \left(\frac{e_{(k,h)}}{e_{(k,h/2)}} \right),$$

donde $e_{(k,h)}$ denota el error global en x_k obtenido con tamaño de paso h .

4.1.1. Métodos de paso único

Cada uno de los problemas 4.1 - 4.5, fueron seleccionadas con la idea de explorar ciertas propiedades como son el comportamiento de la solución numérica en EDO lineales, no-lineales, reales y complejas. Por otro lado, a fin de tener soluciones exactas en el caso de SEDO se tienen los problemas 4.6 y 4.7, ambos son lineales con coeficientes constantes pero se diferencian en que el

primero es no homogéneo y el segundo es homogéneo según como se definió en la Sección 1.2. Las aproximaciones en un SEDO no lineal, en el que no se tiene una solución analítica, se presentarán en la Sección 4.2 con la aplicación de un modelo para el VIH-SIDA. Se han implementado los métodos de Euler, Taylor 2, RK22, RK33, RK44 y RK56. El cálculo de derivadas parciales, necesarias en los métodos de Taylor, se realizaron analíticamente con el fin de centrarnos en los errores numéricos y propiedades específicas del método y no de la aproximación de estas derivadas. En el transcurso de esta sección, presentaremos diferentes simulaciones numéricas con las implementaciones realizadas, a fin de analizar ventajas y desventajas de usar un determinado método.

La primera de las propiedades que se verificará es la convergencia y el orden que ella tiene en cada uno de los métodos según lo presentado en la Sección 2.2.

Convergencia

Los métodos numéricos implementados son convergentes, por lo tanto las implementaciones deben cumplir que a medida que se elijan tamaños de paso cada vez más pequeños, el error global tome valores cada vez más pequeños. Además de verificar que el error está disminuyendo es importante verificar si está disminuyendo con el orden esperado para cada método.

Las tablas 4.1, 4.2 y 4.3 corresponden a los errores y ordenes estimados en la solución de los problemas 4.1, 4.2 y 4.6, respectivamente, mediante la aplicación de algunos MPU. En los tres problemas se determinó una solución numérica para cada uno de los métodos considerados usando un tamaño de paso inicial de $h = 2^{-1}$ y posteriormente para encontrar una nueva aproximación se redujo este tamaño de paso a la mitad. En estas tablas se puede observar que el error está bajando con el orden esperado para cada método.

Si se consideran tamaños de paso demasiado pequeños pueden generarse grandes errores de redondeo que afectan a la solución numérica. La Figura 4.1 incluye el comportamiento del error absoluto al comparar la solución numérica y la exacta del Problema 4.1, para los diferentes métodos. A diferencia de la Tabla 4.1 en la que el tamaño de paso se reduce de 2^{-1} a 2^{-6} , en esta figura el tamaño de paso se reduce desde 2^{-1} hasta 2^{-13} . Se resalta que el método de Euler tuvo una reducción del error desde un orden de 10^{-1} hasta 10^{-4} , lo cual es una variación pobre al compararla con la de los otros métodos de orden superior para los que el error llega a reducirse entre una precisión de 10^{-8} y 10^{-5} . Además, podemos concluir que efectivamente los métodos de Taylor 2 y RK22 tienen el mismo orden, dado que las rectas azul y verde respectivamente relacionadas con los resultados de estos métodos, son paralelas. Incluso al comparar estos dos métodos, Taylor 2 produce un error menor que RK22, esto se debe a que el cálculo en las derivadas parciales para el método de Taylor 2 se realizó de forma analítica. En esta figura, debido a las escalas logarítmicas en ambos

Método	h	Error e_k	Orden	Método	h	Error e_k	Orden
Euler	2^{-1}	2.45809×10^{-1}		RK33	2^{-1}	5.41752×10^{-3}	
	2^{-2}	1.13014×10^{-1}	1.12102		2^{-2}	6.13386×10^{-4}	3.14276
	2^{-3}	5.44431×10^{-2}	1.05369		2^{-3}	7.25708×10^{-5}	3.07933
	2^{-4}	2.67465×10^{-2}	1.02539		2^{-4}	8.81497×10^{-6}	3.04136
	2^{-5}	1.32591×10^{-2}	1.01236		2^{-5}	1.08590×10^{-6}	3.02106
Taylor 2	2^{-1}	1.59322×10^{-2}		RK44	2^{-1}	6.24409×10^{-4}	
	2^{-2}	2.71879×10^{-3}	2.55090		2^{-2}	3.47545×10^{-5}	4.16722
	2^{-3}	5.54569×10^{-4}	2.29353		2^{-3}	2.04165×10^{-6}	4.08938
	2^{-4}	1.24872×10^{-4}	2.15091		2^{-4}	1.23608×10^{-7}	4.04589
	2^{-5}	2.96060×10^{-5}	2.07649		2^{-5}	7.60216×10^{-9}	4.02322
RK22	2^{-1}	7.11069×10^{-2}		RK56	2^{-1}	1.26283×10^{-5}	
	2^{-2}	1.58148×10^{-2}	2.16870		2^{-2}	3.82736×10^{-7}	5.04417
	2^{-3}	3.72500×10^{-3}	2.08596		2^{-3}	1.16370×10^{-8}	5.03955
	2^{-4}	9.04064×10^{-4}	2.04274		2^{-4}	3.57799×10^{-10}	5.02342
	2^{-5}	2.22710×10^{-4}	2.02125		2^{-5}	1.10846×10^{-11}	5.01254

Tabla 4.1: Convergencia MPU, Problema 4.1.

Método	h	Error e_k	Orden	Método	h	Error e_k	Orden
Euler	2^{-1}	2.31889		RK33	2^{-1}	4.15463×10^{-2}	
	2^{-2}	1.76439	0.39426		2^{-2}	3.84758×10^{-3}	3.43269
	2^{-3}	1.12976	0.64316		2^{-3}	4.43460×10^{-4}	3.11707
	2^{-4}	6.46712×10^{-1}	0.80481		2^{-4}	5.45722×10^{-5}	3.02256
	2^{-5}	3.47126×10^{-1}	0.89766		2^{-5}	6.81450×10^{-6}	3.00148
Taylor 2	2^{-1}	3.49363×10^{-2}		RK44	2^{-1}	2.19803×10^{-3}	
	2^{-2}	1.92863×10^{-2}	0.85714		2^{-2}	8.34044×10^{-5}	4.71994
	2^{-3}	7.37975×10^{-3}	1.38593		2^{-3}	3.82879×10^{-6}	4.44516
	2^{-4}	2.14251×10^{-3}	1.78427		2^{-4}	1.97936×10^{-7}	4.27378
	2^{-5}	5.72599×10^{-4}	1.90370		2^{-5}	1.10839×10^{-8}	4.15848
RK22	2^{-1}	2.22447×10^{-1}		RK56	2^{-1}	1.05354×10^{-3}	
	2^{-2}	4.00038×10^{-2}	2.47525		2^{-2}	3.37597×10^{-5}	4.96380
	2^{-3}	8.55285×10^{-3}	2.22566		2^{-3}	1.06307×10^{-6}	4.98898
	2^{-4}	1.97299×10^{-3}	2.11601		2^{-4}	3.33351×10^{-8}	4.99505
	2^{-5}	4.72907×10^{-4}	2.06075		2^{-5}	1.04345×10^{-9}	4.99760

Tabla 4.2: Convergencia MPU, Problema 4.2.

Método	h	Error e_k	Orden	Método	h	Error e_k	Orden
Euler	2^{-1}	2.18818×10^{-1}		RK33	2^{-1}	1.94718×10^{-2}	
	2^{-2}	1.20406×10^{-1}	0.86181		2^{-2}	1.52701×10^{-3}	3.67260
	2^{-3}	6.05322×10^{-2}	0.99213		2^{-3}	1.47435×10^{-4}	3.37256
	2^{-4}	3.02705×10^{-2}	0.99979		2^{-4}	1.61565×10^{-5}	3.18988
	2^{-5}	1.51303×10^{-2}	1.00046		2^{-5}	1.89099×10^{-6}	3.09490
Taylor 2	2^{-1}	2.17215×10^{-2}		RK44	2^{-1}	2.99866×10^{-3}	
	2^{-2}	6.56278×10^{-3}	1.72674		2^{-2}	7.68028×10^{-5}	5.28701
	2^{-3}	1.73316×10^{-3}	1.92090		2^{-3}	2.59629×10^{-6}	4.88663
	2^{-4}	4.40489×10^{-4}	1.97622		2^{-4}	1.23637×10^{-7}	4.39226
	2^{-5}	1.10771×10^{-4}	1.99152		2^{-5}	6.76312×10^{-9}	4.19228
RK22	2^{-1}	1.18706×10^{-1}		RK56	2^{-1}	3.30900×10^{-4}	
	2^{-2}	1.97107×10^{-2}	2.59034		2^{-2}	5.78134×10^{-6}	5.83884
	2^{-3}	4.32157×10^{-3}	2.18935		2^{-3}	1.27089×10^{-7}	5.50749
	2^{-4}	1.03594×10^{-3}	2.06061		2^{-4}	3.26119×10^{-9}	5.28429
	2^{-5}	2.54185×10^{-4}	2.02699		2^{-5}	9.17443×10^{-11}	5.15163

Tabla 4.3: Convergencia MPU, Problema 4.6.

ejes, también es posible relacionar la pendiente de las rectas con el orden de convergencia, entre mayor inclinación mayor es el orden del método.

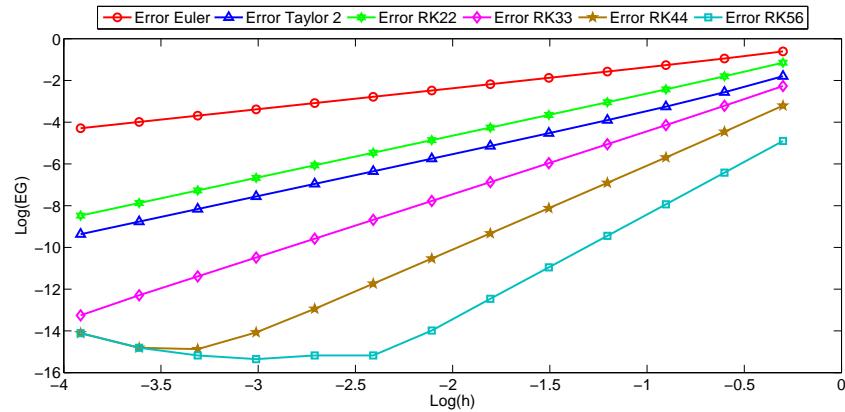


Figura 4.1: Convergencia MPU, Problema 4.1.

De la Figura 4.1, también se observa que para el método RK56 no es recomendable usar un tamaño de paso tan pequeño, ya que para tamaños de paso menores que $h = 2^{-8}$ la precisión finita del computador hace que se produzcan errores de redondeo que afectan significativamente al error global de discretización. De manera similar para el método RK44 no se producen buenos resultados

para tamaños de paso menores que 2^{-11} . Los valores de aproximaciones obtenidas con estos dos métodos, luego de pasar el cero computacional de 10^{-16} ya no son confiables y deben ser descartados.

En las implementaciones de los métodos la elección de tamaños de paso demasiado pequeños no siempre produce buenos resultados, por tal motivo se evidencia en la práctica la importancia que tiene el estudio de estabilidad absoluta para los MPU.

Estabilidad absoluta

Con el fin de evaluar la región de estabilidad absoluta para el método de Euler en problemas que involucren parámetros complejos, siguiendo los conceptos presentados en la Sección 2.5, se considera el Problema 4.5. Dado que para este estudio se toma como modelo la EDO (2.5.3), se concluye que para el Problema 4.5 se tiene $\lambda = -1 + 2i$ y por lo tanto $\lambda h = -h + 2hi$. Para que λh esté dentro de la región de estabilidad absoluta del método de Euler, que es el interior de un círculo de centro en $(-1, 0)$ en el plano complejo (2.5.5), se debe cumplir que

$$[(-h) - (-1)]^2 + (2h)^2 < 1,$$

es decir que $h(5h - 2) < 0$. Por lo tanto para asegurar la estabilidad absoluta al aplicar el método de Euler para aproximar la solución de este problema, se debe elegir $h \in (0, 0.4)$.

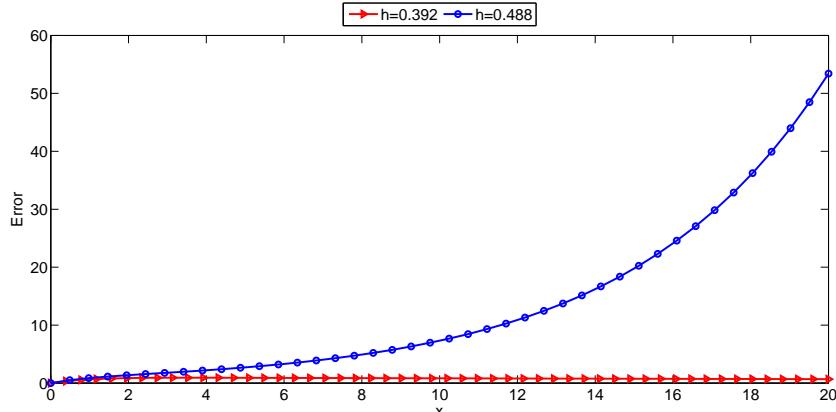


Figura 4.2: Estabilidad absoluta Euler, Problema 4.5.

En la Figura 4.2 se presenta el error obtenido para dos valores de h , uno dentro del intervalo $(0, 0.4)$ y otro fuera de este intervalo. Se observa que para h fuera del intervalo, el cual tiene el error de color azul, se obtienen grandes errores, en este caso para $x = 20$ el error fue de 53.41, mientras que para h dentro del intervalo se observa que el error disminuye considerablemente, en $x = 20$ el error fue 0.67. Con esta simulación, se verifica numéricamente que para obtener buenas

aproximaciones numéricas se deben tomar tamaños de paso dentro de la región de estabilidad absoluta.

Para los problemas 4.3 y 4.7 se realizarán diferentes simulaciones numéricas con los métodos de Euler y RK33, para tamaños de paso dentro y fuera de las regiones de estabilidad absoluta, presentadas en la Tabla 2.1 y en la Figura 2.1 del Capítulo 2.

En el Problema 4.3 se tiene que $\lambda = -15$, por lo tanto los intervalos de estabilidad absoluta para los métodos de Euler y Rk33 en este problema son respectivamente

- $-2 < -15h < 0 \Rightarrow 0 < h < 0.1334$.
- $-2.51 < -15h < 0 \Rightarrow 0 < h < 0.1673$.

En la Figura 4.3 se presenta en color rojo la solución exacta del Problema 4.3, junto con dos soluciones numéricas, de color azul, obtenidas de aplicar el método de Euler con tamaños de paso $h_{1E} = 0.1428$ y $h_{2E} = 0.1307$, que están dentro y fuera del intervalo de estabilidad absoluta respectivamente. Se verifica que para h_{1E} los resultados numéricos son inaceptables ya que la diferencia máxima entre la solución numérica y la solución exacta es del orden de 10^8 , mientras que para h_{2E} , a pesar de que en principio se presentan algunas oscilaciones la diferencia máxima entre la solución numérica es de 1 unidad y esta diferencia disminuye conforme avanza el proceso de integración. Sin embargo, aunque h_{2E} se encuentra dentro del intervalo de estabilidad, es muy cercano al valor extremo de este intervalo lo que justifica la inestabilidad inicial que se observa en la Figura 4.3 (b).

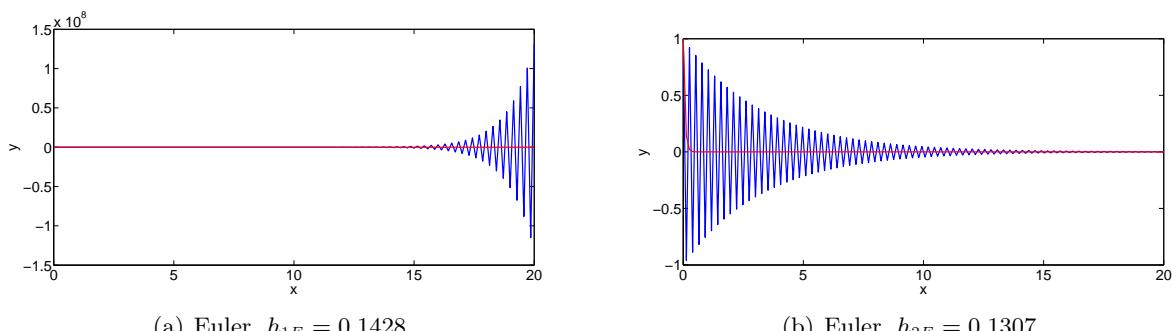


Figura 4.3: Estabilidad absoluta método de Euler, valores de h críticos, Problema 4.3.

Para mejorar los resultados obtenidos usando h_{2E} , la Figura 4.4 muestra los resultados de aplicar el método de Euler con tamaños de paso h_{3E} y h_{4E} , los cuales están dentro del intervalo de

estabilidad absoluta y son de menor magnitud que h_{2E} . En esta figura, se hace un acercamiento en la región que se presenta la diferencia máxima entre la solución exacta y la solución numérica. Se concluye que para $h_{4E} = 0.0667$, el cual se encuentra en la mitad del intervalo $(0, 0.1334)$, la aproximación se acerca en forma más suave a la solución exacta, como se observa en la Figura 4.4 (b). Aunque $h_{3E} = 0.1000$ también aproxima con buena precisión la solución, por lo obtenido en la Figura 4.4 (a), se observa que al ser un valor de paso más cercano al extremo derecho del intervalo de estabilidad para x pequeño la convergencia se da de forma oscilante.

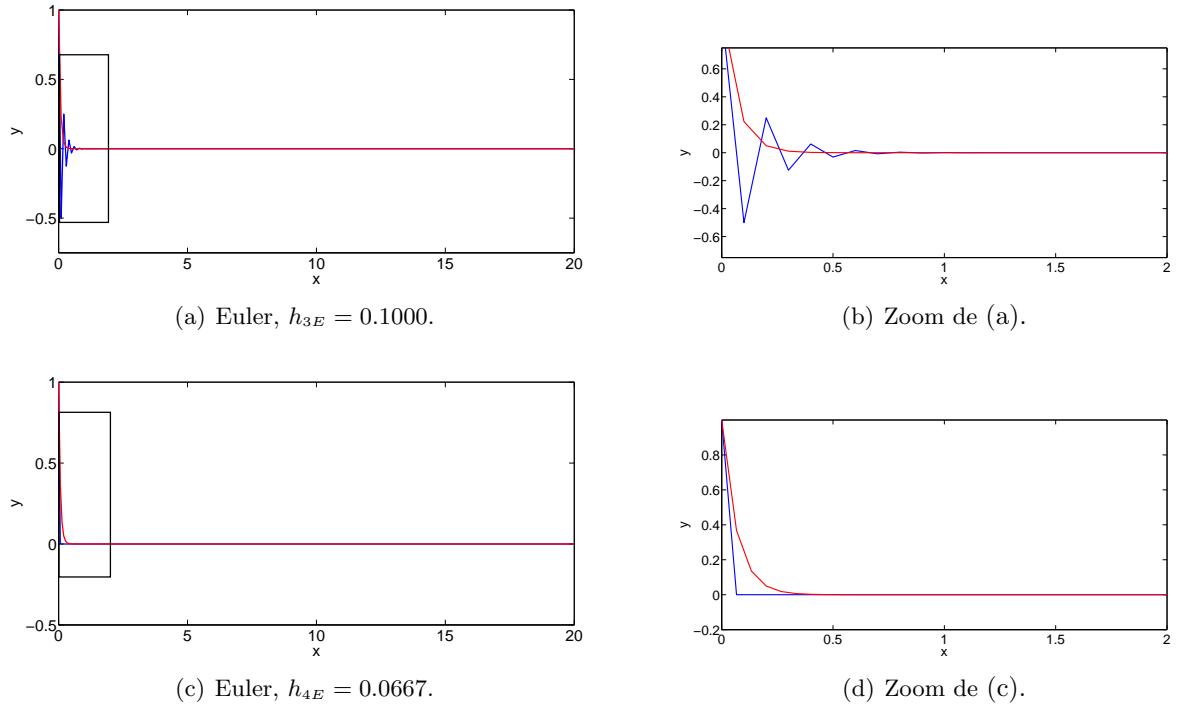
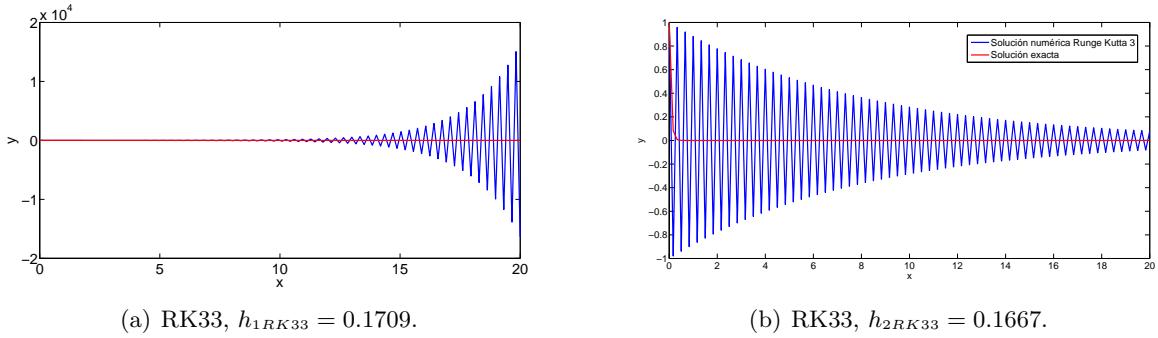
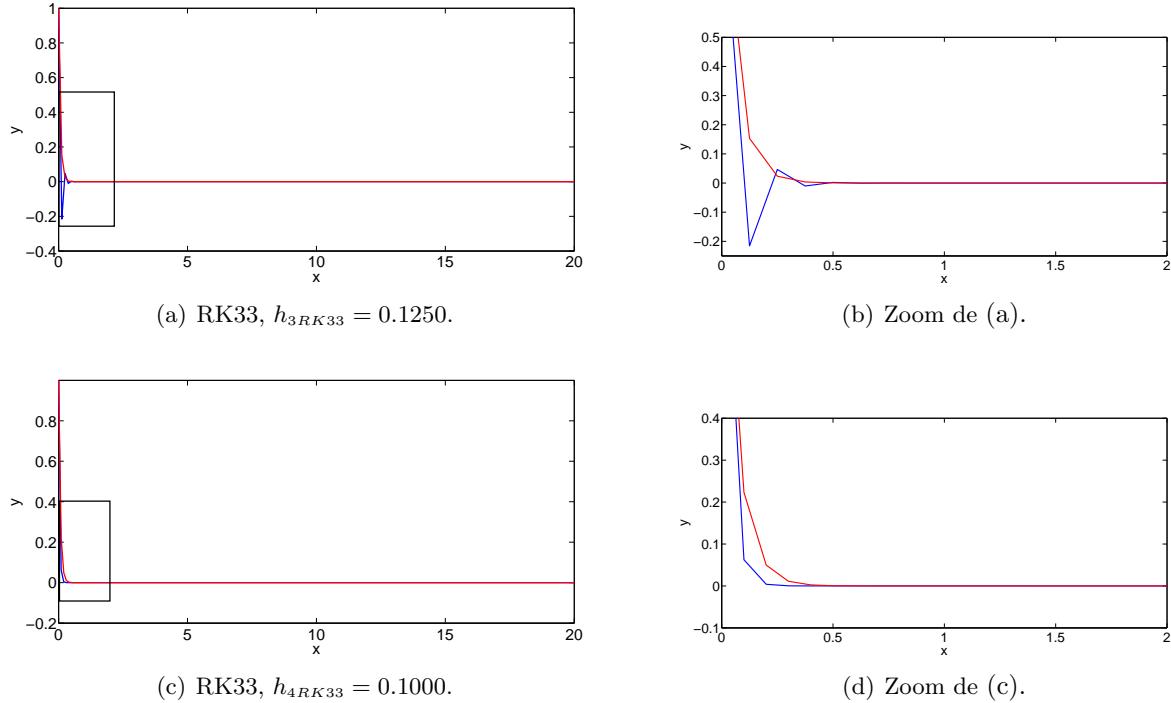


Figura 4.4: Estabilidad absoluta método de Euler, h en la región de estabilidad, Problema 4.3.

En el caso de la estabilidad absoluta para el método RK33, en la Figura 4.5 se presenta en color rojo la solución exacta del Problema 4.3 junto con dos soluciones numéricas en azul con tamaños de paso $h_{1RK33} = 0.1709$, $h_{2RK33} = 0.1667$, que están dentro y fuera del intervalo de estabilidad absoluta respectivamente. Al igual que lo sucedido con el método de Euler, se observa que para el tamaño de paso fuera del intervalo de estabilidad absoluta la diferencia entre la solución numérica y la solución exacta es muy grande, mientras que para un tamaño de paso ligeramente dentro del intervalo de estabilidad, en inicio se presentan grandes oscilaciones que disminuyen su magnitud conforme avanza el proceso de integración.

Figura 4.5: Estabilidad absoluta método RK33, valores de h críticos, Problema 4.3.

En la Figura 4.6 (a) y (b) se presentan mejores aproximaciones obtenidas de aplicar el método RK33 respectivamente, con tamaños de paso $h_{3RK33} = 0.1250$ y $h_{4RK33} = 0.1000$ que están dentro del intervalo de estabilidad absoluta y no son tan cercanos a los extremos del intervalo. En esta figura se hace un acercamiento de la región en la que se presenta la diferencia máxima entre la solución exacta y la solución numérica.

Figura 4.6: Estabilidad absoluta método RK33, h en la región de estabilidad, Problema 4.3.

Las simulaciones numéricas realizadas al Problema 4.3 mediante los métodos de Euler y RK33 evidencian numéricamente la importancia del estudio de estabilidad absoluta para los MPU, ya que a pesar de que los métodos implementados cumplen las propiedades teóricas de consistencia, convergencia y estabilidad, es indispensable elegir un tamaño de paso adecuado para obtener buenos resultados.

En el caso de sistemas EDO, teóricamente el intervalo de estabilidad absoluta está influenciado por uno de los valores propios del sistema, para verificar numéricamente este hecho se considera el Problema 4.7 que es un sistema de primer orden con dos EDO. Según los valores propios del Problema 4.7, $\lambda_1 = -1$ y $\lambda_2 = -3$, los intervalos de estabilidad absoluta para los métodos de Euler y Rk33 en este problema son respectivamente

- $-2 < \lambda_1 h < 0 \Rightarrow -2 < -h < 0 \Rightarrow 0 < h < 2,$
 $-2 < \lambda_2 h < 0 \Rightarrow -2 < -3h < 0 \Rightarrow 0 < h < 0.6667.$
- $-2.51 < \lambda_1 h < 0 \Rightarrow -2.51 < -h < 0 \Rightarrow 0 < h < 2.51,$
 $-2.51 < \lambda_2 h < 0 \Rightarrow -2.51 < -3h < 0 \Rightarrow 0 < h < 0.8366.$

En las figuras 4.7 y 4.8 se presentan las soluciones exactas del sistema (rojo), junto con las soluciones numéricas (u-azul, v-verde) obtenidas de aplicar los métodos de Euler y RK33 respectivamente. Se han utilizado diferentes tamaños de paso, dentro y fuera de los intervalos de estabilidad absoluta. Estas simulaciones numéricas permiten observar que tanto para el método de Euler como para RK33 el valor propio $\lambda = -3$ es el que determina el intervalo de estabilidad absoluta para el Problema 4.7. Esto era de esperarse ya que al observar, por el ejemplo para el método de Euler, el tamaño de paso debía cumplir dos condiciones

$$0 < h < 2 \quad \text{y} \quad 0 < h < 0.6667,$$

donde la condición con mayor restricción era la asociada con el valor propio $\lambda = -3$. Es decir que para garantizar la estabilidad absoluta para los métodos de Euler y RK33, respectivamente el tamaño de paso debe pertenecer a los intervalos $(0, 0.6667)$ y $(0, 0.8366)$.

Dado que los intervalos de estabilidad absoluta son mayores para los métodos implícitos que para los explícitos, es interesante estudiar este comportamiento y así se incluye la simulación numérica de un método implícito. Un estudio sobre región de estabilidad para métodos implícitos puede ser estudiado en [7].

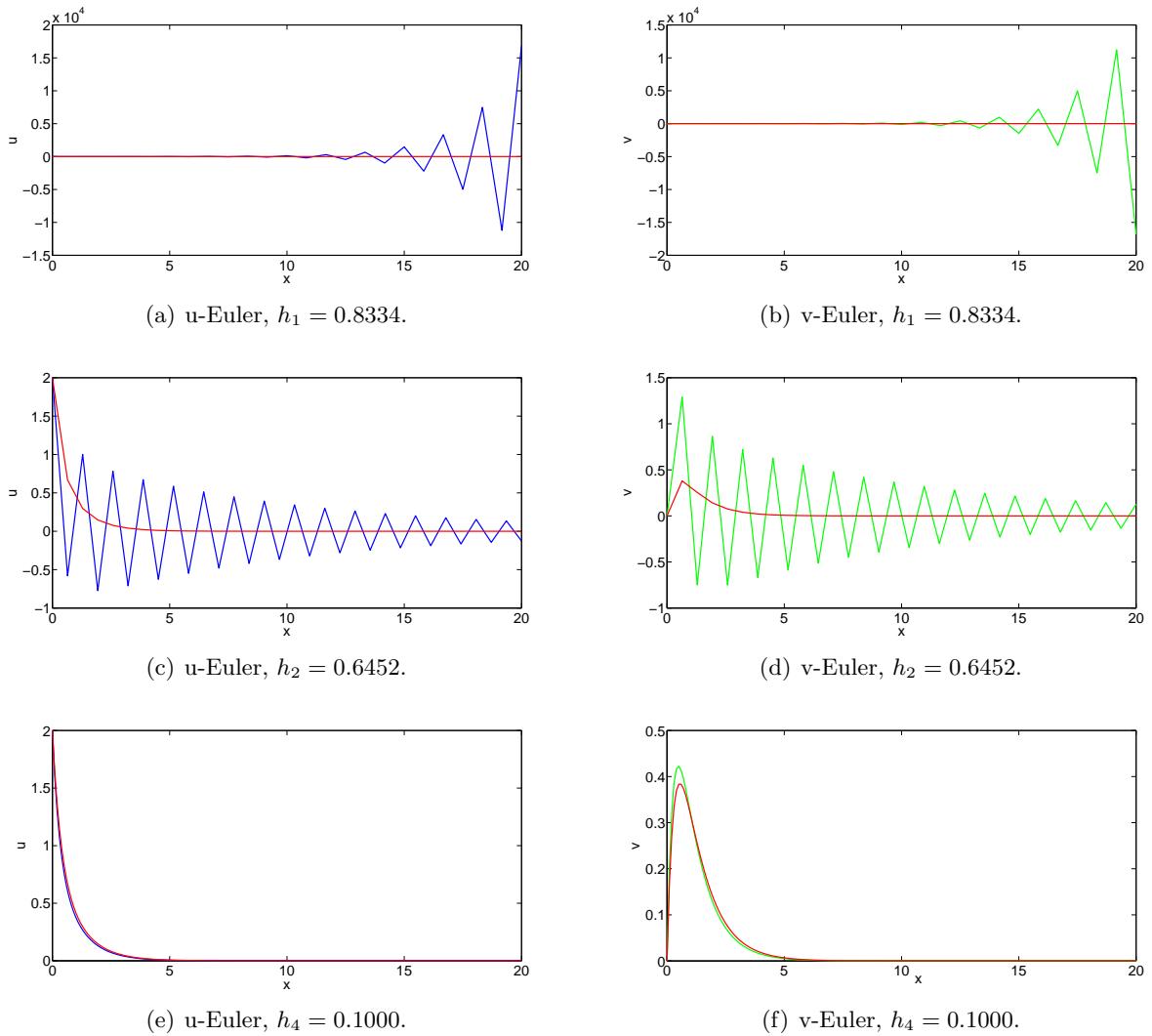


Figura 4.7: Estabilidad absoluta método de Euler, Problema 4.7.

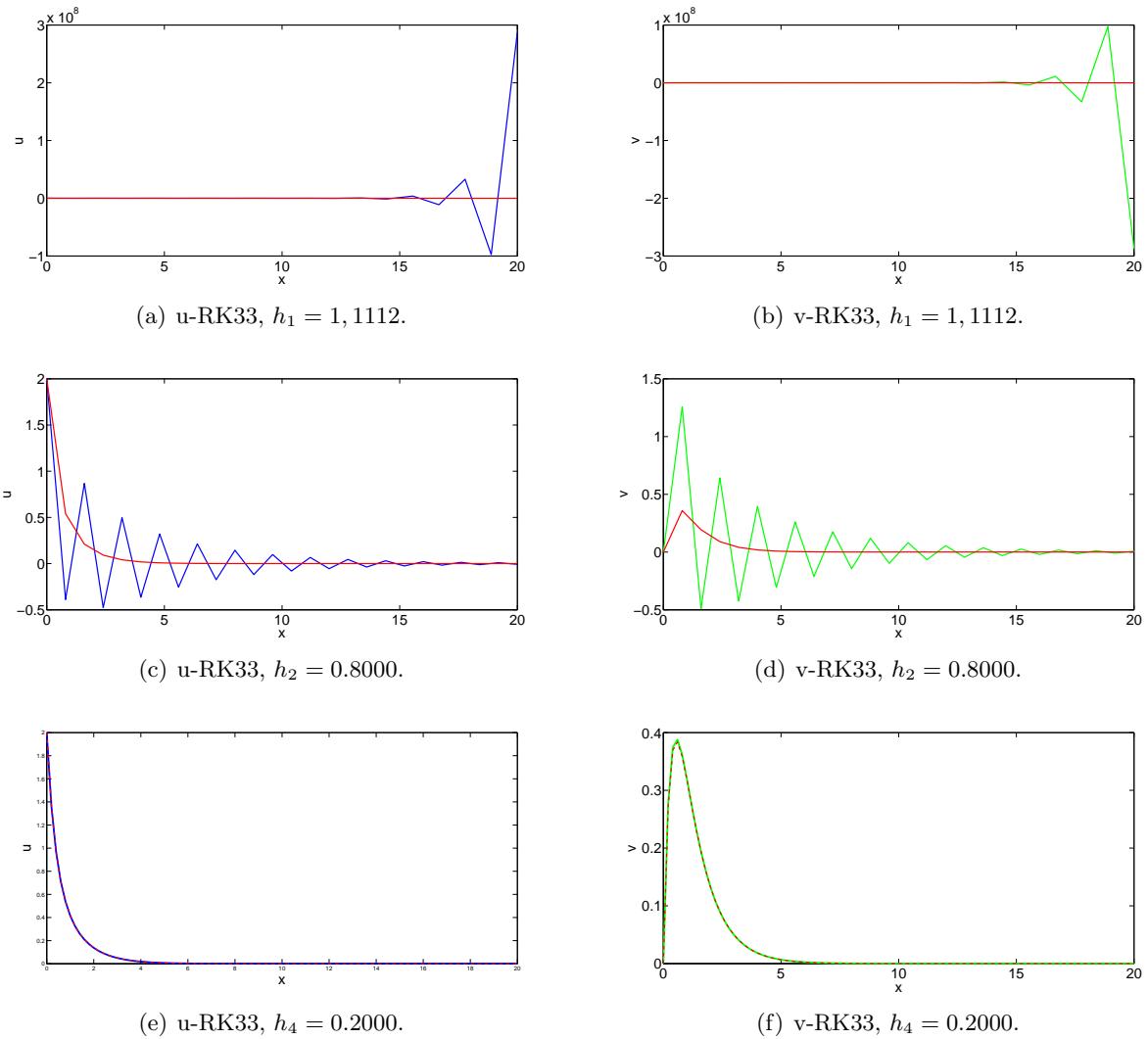


Figura 4.8: Estabilidad absoluta método RK33, Problema 4.7.

Aplicando el método de Euler implícito (1.3.11) al Problema 4.3 se tiene que

$$y_{k+1} = y_k - 15hy_{k+1}.$$

Luego despejando y_{k+1} se obtiene la expresión

$$y_{k+1} = \frac{y_k}{1 + 15h},$$

que es una fórmula explícita para integral el Problema 4.3.

En la Figura 4.9 se observa que para los valores de $h_{1E} = 0.1428$ y $h_{2E} = 0.1307$, que están fuera y dentro de las regiones de estabilidad del método de Euler en el mismo problema, el método de Euler implícito ofrece buenos resultados. Esto se debe a que las regiones de estabilidad absoluta para los métodos implícitos es mayor que en los métodos explícitos, sin embargo los métodos implícitos no son de uso común en la práctica porque no siempre se puede despejar el valor de y_{k+1} , por lo cual para obtener un valor de y_{k+1} de una ecuación implícita se debe usar métodos numéricos para resolver ecuaciones como el método de Newton, lo cual puede aportar errores numéricos considerables.

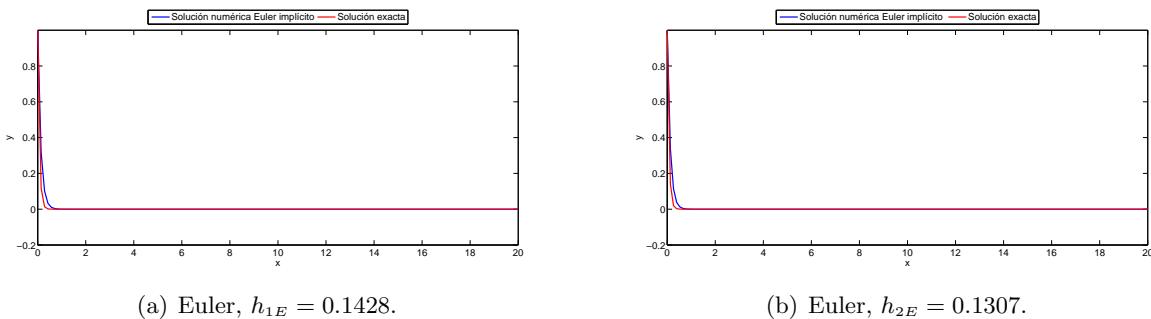


Figura 4.9: Estabilidad absoluta método de Euler implícito, Problema 4.3.

Comparando los métodos de Euler y RK33 en cuanto a las simulaciones presentadas se concluye que es más conveniente usar métodos de alto orden ya que estos presentan regiones de estabilidad más grandes es decir que se puede obtener buenos resultados con tamaños de paso más grandes, lo cual disminuye costos computacionales.

Prueba de tiempo

El costo computacional tiene que ver con la cantidad de recursos computacionales que usa un algoritmo para resolver un problema, los recursos de mayor interés en este tipo de problemas son la memoria que emplea el algoritmo y el tiempo que tarda el algoritmo en ejecutarse. En este trabajo

para comparar el desempeño de los métodos implementados solo consideraremos el costo temporal.

Dado que la aproximación numérica de los problemas tratados en este trabajo no demandan tiempos considerables, se considera el Problema 4.1 sobre el intervalo de integración $[0, 200]$ para aumentar el tiempo de cómputo. En la Figura 4.10 se muestra la relación obtenida entre el tiempo de cómputo (en segundos) y los diferentes errores producidos en la aproximación a $y(200)$ para el Problema 4.1, partiendo de un paso de integración inicial $h = 2$ y disminuyendo este valor a la mitad en cada aproximación. En esta figura los errores se tienen en escala logarítmica para centrarnos en el exponente del error y se observa que para disminuir el error es necesario mayor tiempo de cómputo. Se observa de la figura que el método de Euler es el que toma menor tiempo computacional pero los errores son de baja precisión, por otro lado los métodos RK44 y RK56 son los que producen resultados con mayor precisión y demandan mayor tiempo de cómputo.

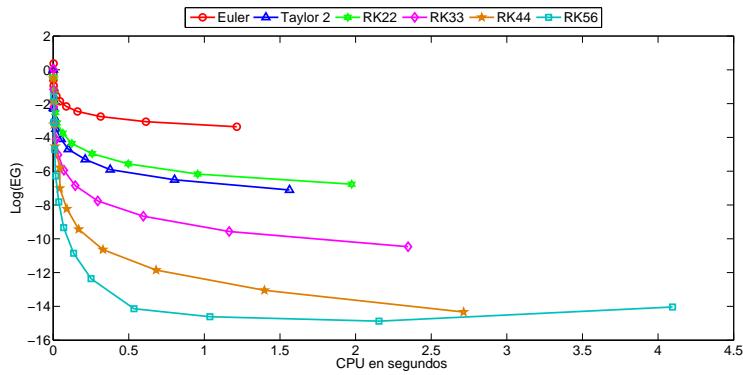


Figura 4.10: Prueba de tiempos MPU, Problema 4.1.

Comparando los métodos de RK44 y RK56 en la Figura 4.10 se observarse que a pesar de que el método RK56 emplea más tiempo en el cálculo que el método RK44, este ofrece resultados más precisos, sin embargo la última aproximación el error global es menor en el método RK44 y esto se debe a que el RK56 alcanzó el orden de máquina y los valores finales dejan de ser confiables. En la última aproximación que corresponde al tamaño de paso $h = 0.00024$ el método RK44 presenta un error global de 4.66293×10^{-15} y la aproximación fue obtenida en 2.714 segundos, mientras que para el método RK56 el error fue de 9.21485×10^{-15} y la aproximación fue obtenida en 4.095, por tanto se concluye que para tamaños de paso pequeños el método RK44 ofrece resultados más precisos que el método RK56 y en menor tiempo. La diferencia de tiempos de cálculo entre los métodos RK44 y RK56 se justifica en el hecho de que mientras que en el método de RK44 se realizan 4 evaluaciones en la función f por cada paso, en el método RK56 es necesario realizar 6 evaluaciones de la función f , lo que obviamente demanda más tiempo de cálculo.

Runge Kutta Fehlberg

En la práctica con frecuencia se presentan problemas para los cuales la solución que se quiere aproximar es suave en algunas regiones, por lo tanto para evitar costos computacionales en estas regiones no es necesario usar tamaños de paso tan pequeños para obtener buenos resultados. Este tipo de problemas da origen a los métodos adaptativos, los cuales eligen el tamaño de paso de acuerdo al problema tratado. En este trabajo se ha implementado el método de Runge Kutta Fehlber, el cual es una combinación de los métodos RK44 y RK45 capaz de variar el tamaño de paso a fin de que el error se mantenga en una tolerancia predefinida, el cual se estudia en el Apéndice A.2.

Para validar la implementación del método de Runge Kutta Fehlbert se considera el Problema 4.4. La gráfica de la solución exacta a este problema se muestra en la Figura 4.11. En la gráfica se

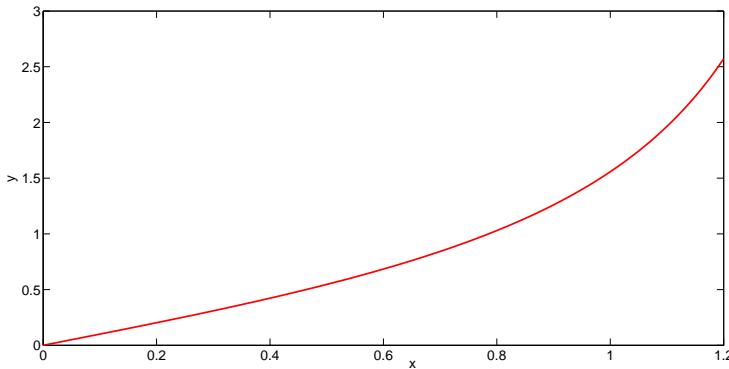


Figura 4.11: Solución exacta Problema 4.4.

observa que la solución exacta en general es suave, pero el crecimiento de la función aumenta en forma un poco más acelerada en la parte final del intervalo. Aplicando el método RKF al Problema 4.4 con un tamaño de paso $h \in (0.01, 0.25)$ y una tolerancia para el error de 10^{-6} , se obtienen los resultados presentados en la Tabla 4.4, donde a partir de $x = 0.8$ es necesario disminuir el tamaño de paso para alcanzar la tolerancia fijada.

4.1.2. Métodos de paso múltiple

De forma análoga a la sección anterior, en esta sección se muestran algunos resultados numéricos en base a las propiedades teóricas correspondientes a los MPM del Capítulo 3 de este trabajo. Se han implementado los métodos de Adams Bashford de ordenes 2, 3 y 4 (AB2, AB3 y AB4, respectivamente), que corresponden a los métodos (3.4.4), (3.4.6) y (3.4.7). Las simulaciones numéricas obtenidas de estas implementaciones permiten determinar ventajas y desventajas de usar un determinado método.

x_k	h_k	y_k	Error e_k
0.25	0.25	0.25534	7.11008×10^{-8}
0.5	0.25	0.54630	1.21433×10^{-6}
0.75	0.25	0.93160	9.42345×10^{-6}
0.872	0.122	1.19008	1.28408×10^{-5}
0.967	0.095	1.45125	1.70797×10^{-5}
1.041	0.074	1.70959	2.19929×10^{-5}
1.100	0.059	1.96845	2.76557×10^{-5}
1.149	0.048	2.23063	3.41606×10^{-5}
1.190	0.040	2.49801	4.16013×10^{-5}
1.2	0.009	2.57219	4.37629×10^{-5}

Tabla 4.4: Solución numérica Problema 4.4, método RKF.

Antes de empezar es importante mencionar que en algunos de los resultados que se presentan a seguir se supone que todos los valores iniciales para los métodos de Adams son calculados de forma exacta a no ser que se diga lo contrario.

Convergencia

Para verificar numéricamente la convergencia se ha considerado los problemas 4.1, 4.2 y 4.6 partiendo de un tamaño de paso inicial $h = 2^{-3}$ y reduciendo este valor a la mitad en cada aproximación. En las tablas 4.5, 4.6 y 4.7, se puede observar que el error está bajando con el orden esperado para cada método, sin embargo si se consideran tamaños de paso demasiado pequeños pueden generarse grandes errores de redondeo que afectan a la solución numérica. La Figura 4.12 corresponde a la solución numérica del Problema 4.1 mediante los métodos de Adams, aquí se observa que el método AB4, a partir de cierto tamaño de paso el error no decrece linealmente y tiende a aumentar.

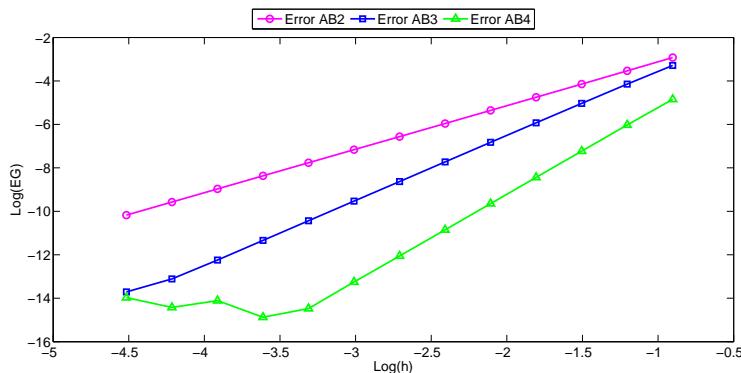


Figura 4.12: Convergencia MPM, Problema 4.1.

Método	h	Error e_k	Orden
AB2	2^{-3}	1.19334×10^{-3}	
	2^{-4}	2.91899×10^{-4}	2.03146
	2^{-5}	7.17306×10^{-5}	2.02481
	2^{-6}	1.77467×10^{-5}	2.01503
	2^{-7}	4.41148×10^{-6}	2.00822
AB3	2^{-3}	5.17271×10^{-4}	
	2^{-4}	7.13303×10^{-5}	2.85833
	2^{-5}	9.28717×10^{-6}	2.94120
	2^{-6}	1.18274×10^{-6}	2.97310
	2^{-7}	1.49168×10^{-7}	2.98712
AB4	2^{-3}	1.42537×10^{-5}	
	2^{-4}	9.44054×10^{-7}	3.91632
	2^{-5}	5.88113×10^{-8}	4.00470
	2^{-6}	3.63915×10^{-9}	4.01442
	2^{-7}	2.25836×10^{-10}	4.01024

Tabla 4.5: Convergencia MPM, Problema 4.1.

Método	n	Error e_k	Orden
AB2	2^{-3}	1.82998×10^{-2}	
	2^{-4}	5.34438×10^{-3}	1.77573
	2^{-5}	1.43079×10^{-3}	1.90120
	2^{-6}	3.69455×10^{-4}	1.953334
	2^{-7}	9.38294×10^{-5}	1.97729
AB3	2^{-3}	8.47887×10^{-3}	
	2^{-4}	1.08230×10^{-3}	2.96976
	2^{-5}	1.37559×10^{-4}	2.97597
	2^{-6}	1.73601×10^{-5}	2.98620
	2^{-7}	2.18102×10^{-6}	2.99270
AB4	2^{-3}	1.16350×10^{-3}	
	2^{-4}	5.37009×10^{-5}	4.43739
	2^{-5}	2.71931×10^{-6}	4.30363
	2^{-6}	1.49348×10^{-7}	4.18648
	2^{-7}	8.67899×10^{-9}	4.10501

Tabla 4.6: Convergencia MPM, Problema 4.2.

Método	n	Error e_k	Orden
AB2	2^{-3}	4.09536×10^{-3}	
	2^{-4}	1.07441×10^{-3}	1.93043
	2^{-5}	2.74846×10^{-4}	1.96686
	2^{-6}	6.94833×10^{-5}	1.98388
	2^{-7}	1.74666×10^{-5}	1.99206
AB3	2^{-3}	644371×10^{-4}	
	2^{-4}	8.52866×10^{-5}	2.91750
	2^{-5}	1.08846×10^{-5}	2.97002
	2^{-6}	1.37301×10^{-6}	2.98688
	2^{-7}	1.72358×10^{-7}	2.99386
AB4	2^{-3}	3.66991×10^{-5}	
	2^{-4}	3.07576×10^{-6}	3.57673
	2^{-5}	2.11800×10^{-7}	3.86016
	2^{-6}	1.38082×10^{-8}	3.93910
	2^{-7}	8.80160×10^{-10}	3.97162

Tabla 4.7: Convergencia MPM, Problema 4.6.

Prueba de tiempos

Para realizar una prueba de tiempos se considera el Problema 4.1 sobre el intervalo de integración $[0, 200]$. En la Figura 4.13 se muestran los diferentes errores producidos en la aproximación a $y(200)$ para el Problema 4.1, partiendo de un paso de integración inicial $h = 2^{-1}$ y disminuyendo este valor a la mitad en cada aproximación. De las implementaciones realizadas, el método AB4 ofrece mejores

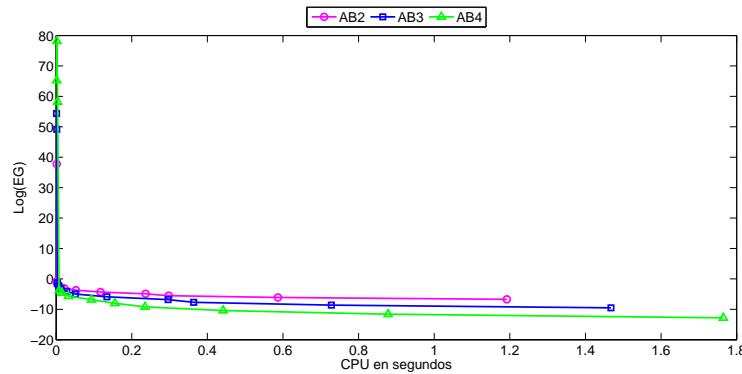


Figura 4.13: Prueba de tiempos MPM, Problema 4.1.

resultados y la diferencia de tiempos entre cada uno de los métodos de Adams no es significativa, ya que en las implementaciones de los métodos de Adams, a partir de cierta iteración se utiliza las evaluaciones de f de pasos anteriores, por lo cual sólo se hace una evaluación de la función f en

cada nueva iteración.

Valores iniciales Adams Bashford

En las simulaciones anteriores se ha calculado los valores iniciales para los métodos de Adams usando la solución exacta del problema, en la práctica esto no es posible. Generalmente se suele usar un MPU para calcular los valores iniciales de un MPM. En la Tabla 4.1.2 se muestra algunos resultados obtenidos aplicando el método de AB4 al Problema 4.2, donde el cálculo de los valores iniciales se ha realizado mediante los métodos de Euler y RK44. Inicialmente se usó el método de Euler y RK4 con el mismo tamaño de paso que el método de AB4, $h = 2^{-3}$. Puede observarse que se obtiene mejores resultados usando el método RK44 en el cálculo de los valores iniciales. A fin de mejorar los resultados de usar el método de Euler en el cálculo de los valores iniciales se usa un tamaño de paso de $h = 2^{-5}$. De acuerdo al interés de la investigación en un problema práctico, si se desea mayor precisión se debe usar el método de RK44, pero si lo que se quiere es disminuir costos computacionales, se observa que el uso del método de Euler con tamaños de paso pequeños ofrece buenos resultados en el cálculo de valores iniciales para MPM.

x_k	Error e_k , AB4	Error e_k , E-AB4	Error e_k , E-AB4 $h/4$	Error e_k , RK44-AB4
0.1	0	4.98683×10^{-3}	1.29167×10^{-3}	8.46348×10^{-8}
0.2	0	1.03280×10^{-2}	2.67299×10^{-3}	1.82482×10^{-7}
0.3	0	1.58405×10^{-2}	4.09359×10^{-3}	2.90106×10^{-7}
0.4	2.48135×10^{-5}	1.75603×10^{-2}	4.51912×10^{-3}	2.44909×10^{-5}
0.5	4.89248×10^{-5}	1.89021×10^{-2}	4.84841×10^{-3}	4.85778×10^{-5}
0.6	6.24625×10^{-5}	2.06717×10^{-2}	5.29558×10^{-3}	6.20827×10^{-5}
0.7	6.34433×10^{-5}	2.24334×10^{-2}	5.75005×10^{-3}	6.30311×10^{-5}
0.8	4.77791×10^{-5}	2.41420×10^{-2}	6.20319×10^{-3}	4.73359×10^{-5}
0.9	1.21518×10^{-5}	2.58244×10^{-2}	6.66437×10^{-3}	1.16785×10^{-5}
1.0	4.41085×10^{-5}	2.74290×10^{-2}	7.12075×10^{-3}	4.46102×10^{-5}

Tabla 4.8: Valores iniciales AB4 con Euler y RK44, Problema 4.2.

4.1.3. Comparación entre los métodos de paso único y paso múltiple

La primera desventaja de usar los métodos de paso múltiple es el cálculo de los valores iniciales, ya que en la práctica tales valores se deben calcular mediante un método auxiliar que puede ser un MPU, pero como se observó en los experimentos numéricos dicho método auxiliar debe tener alto orden para que el MPM ofrezca buenas aproximaciones. Los MPU son autónomos en el sentido de que no depende de métodos auxiliares para iniciar.

Una característica particular de los MPU de la familia Runge Kutta es que mediante la combinación de dos métodos de Runge Kutta de diferente orden es posible estimar numéricamente el error local de discretización y con ello controlar el paso de integración en cada iteración, lo cual reduce costo computacionales. En los problemas de prueba se utilizó el método de Runge Kutta Felber que es una combinación de los métodos de RK44 y RK45.

En la prueba de tiempo para el Problema 4.1, en la Figura 4.13, se observa que los métodos de Adams son más eficientes que los métodos de Runge Kutta. Esto se debe a que las evaluaciones de la función f en los métodos de Adams son menos tediosas que en los métodos de Runge Kutta. Por lo tanto si en un problema práctico la función f es costosa computacionalmente, sera conveniente usar un MPM.

4.2. Aplicación VIH/SIDA.

4.2.1. Generalidades VIH/SIDA

El Síndrome de Inmunodeficiencia Adquirida **SIDA** es la fase final de la enfermedad producida por el Virus de Inmunodeficiencia Humana **VIH**. Su principal forma de transmisión es por contacto sexual, sin embargo también se puede transmitir por transfusiones de sangre, por usar un instrumento quirúrgico contaminado con VIH e incluso los bebés que nacen de una madre infectada pueden infectarse durante el embarazo, parto o lactancia.

El virus del VIH ataca el sistema inmunitario y debilita los sistemas de vigilancia y defensa contra las enfermedades infecciosas. El virus afecta principalmente a las células $TCD4^+$, que son células que controlan la respuesta del sistema inmunológico. En la etapa final de la infección por VIH, el número de células $TCD4^+$ funcionales disminuye y el sistema inmunológico falla dejando al individuo portador del virus expuesto a diversas infecciones y enfermedades que normalmente personas con un sistema inmunitario saludable pueden combatir fácilmente.

Según el Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA, ONUSIDA [15], entre el 2000 y 2015 más de 40 millones de personas en el mundo han muerto por esta enfermedad y en la actualidad el número de infectados a nivel mundial es de 36 millones. El VIH/SIDA es un grave problema de salud pública por tanto es de suma importancia diseñar modelos matemáticos que permitan determinar características generales en el desarrollo de esta enfermedad a fin de establecer estrategias de control.

Infección de las células $TCD4^+$ por el virus del VIH

Como los virus no son capaces de reproducirse por sí mismos necesitan utilizar a otros seres vivos

para poder multiplicarse y sobrevivir, para lo cual usan a las células de nuestro sistema inmune, específicamente a las células $TCD4^+$. Inicialmente el virus del VIH se adhiere a la superficie de la célula, una vez fijo el virus libera su material genético en el interior de la célula, el material genético del VIH es **ARN** pero para poder actuar sobre la célula tiene que convertirse en **ADN**, de esto se encarga la proteína viral denominada Transcriptasa, pues esta convierte la cadena simple de ARN vírico en una cadena doble de ADN. El nuevo ADN vírico entra en el núcleo de la célula y se adhiere al ADN celular formando un pro-virus, este madura y genera varias copias de virus VIH. Terminado el proceso de formación del nuevo virus en el interior de la célula $TCD4^+$ infectada, estos brotan por la superficie de la célula. Una ilustración de esta dinámica se presenta en la Figura 4.14, e información complementaria se tiene en [3, 10] y [17].

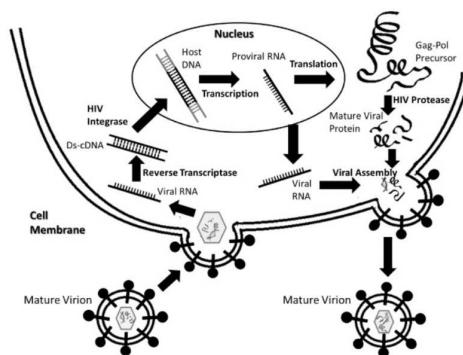


Figura 4.14: Ciclo de vida del VIH [22].

Se cree que la salida de las partículas virales hace que la membrana de la célula $TCD4^+$ se torne más permeable lo cual permite la entrada excesiva de diferentes sustancias que provocan la muerte de la célula. Como las células $TCD4^+$ son esenciales para el correcto funcionamiento del sistema inmunológico, la muerte de estas por acción del **VIH** deja al sistema inmune incapaz de reaccionar de forma eficiente frente a diferentes enfermedades infecciosas, esta etapa de la enfermedad se conoce como SIDA.

4.2.2. Formulación del modelo

Cuando se quiere determinar si un individuo está infectado con el **VIH** se hace una medición o conteo del número de células $TCD4^+$ presentes en un milímetro cúbico de sangre (mm^3), que para un individuo infectado suele estar por debajo de las 200 células por mm^3 , el recuento de células $TCD4^+$ es indispensable pues permite tomar decisiones importantes acerca del tratamiento y la atención médica del VIH como se muestra en la Tabla 4.9. Para modelar esta dinámica numéricamente, se usará el modelo matemático propuesto en [17], el cual ha sido estudiado, modificado y simulado a profundidad en [3, 10, 14] y [19]. El modelo presentado en [17] ha sido el punto de partida de

modelos actuales y en el se considera la variación en la concentración de tres poblaciones T , I y V , que corresponden a la población de células $TCD4^+$ sanas, células $TCD4^+$ infectadas y virus VIH libre, respectivamente.

Concentración células T	Diagnóstico
Entre 500 y 1200 células/ mm^3	Lo normal entre personas sin VIH.
Por encima de 350 células/ mm^3	No se recomienda, en general, el tratamiento anti-VIH.
Por debajo de 350 células/ mm^3	Si se recomienda iniciar el tratamiento anti-VIH.
Por debajo de 200 células/ mm^3	Se diagnostica SIDA. Existe mayor riesgo de infecciones y enfermedades, se recomienda tratamiento anti-VIH.

Tabla 4.9: Conteo de células T sanas y diagnóstico [19].

Con el fin de describir el modelo, se presentan las siguientes hipótesis

- El número de células sanas T producidas por el organismo en cada instante de tiempo es s .
- Las células sanas T se multiplican a una tasa pT y mueren de forma natural a una tasa dT .
- Las células sanas T mueren por acción del virus del VIH a una tasa cI .
- Cada célula infectada I produce N nuevos virus al morir.
- La tasa de infección al contacto entre un virus libre V y una célula sana T es dada por kVT .
- Los virus libres V son removidos del organismo a una tasa uV .

Al estudiar la variación de T , I y V en base a las hipótesis anteriores se tiene un modelo de EDO no lineal que describe la dinámica del VIH, dado por

$$\begin{cases} T'(t) = s + pT - dT - kVT \\ I'(t) = kVT - cI \\ V'(t) = NcI - uV. \end{cases} \quad (4.2.1)$$

Sin embargo, a fin de tener un modelo más acorde con la realidad, se cambia el término que controla el aumento de células T sanas pT por el término logístico $pT(1 - \frac{T+\alpha I}{T_{max}})$, donde T_{max} es el número máximo de células T sanas en el organismo y α es un valor constante. De aquí, al modificar el modelo (4.2.1), se obtiene el modelo de interés

$$\begin{cases} T'(t) = s + pT(1 - \frac{T+\alpha I}{T_{max}}) - dT - kVT \\ I'(t) = kVT - cI \\ V'(t) = NcI - uV. \end{cases} \quad (4.2.2)$$

Este nuevo modelo es más realista pues existen límites biológicos para la producción de células T , sin esta sustitución puede pasar que el número de células T crezca indefinidamente, lo cual no acompaña a la realidad. La constante α permite que este modelo sea versátil con la literatura dado que algunas investigaciones ignoran el efecto de la población de células infectadas en el término logístico al ser muy pequeño, del orden de 10^{-4} a 10^{-5} , en este caso tomamos $\alpha = 0$ o en caso contrario $\alpha = 1$.

4.2.3. Modelo con estrategia de control.

El modelo (4.2.2) no tiene en cuenta el comportamiento de la enfermedad al suministrarse medicamentos, pero es claro que al enterarse un paciente de la enfermedad intentará combatirla. De aquí es importante el estudio de modelos con estrategias de control. Se sigue la estrategia presentada en [10] y [17].

Supongamos que se aplica un tratamiento anti viral que afecta la tasa de infección al contacto entre células T y virus V . El tratamiento consiste en la administración de un medicamento que impide que un virus infecte nuevas células. Si denotamos por D la eficiencia del medicamento, se puede modelar la acción de este médicalemente en el organismo sustituyendo en (4.2.2) el término kVT por el término $(1 - D)kVT$, obteniendo así el siguiente modelo con estrategia de control de la infección.

$$\begin{cases} T'(t) = s + pT \left(1 - \frac{T+\alpha I}{T_{max}}\right) - dT - (1 - D)kVT \\ I'(t) = (1 - D)kVT - cI \\ V'(t) = NcI - uV. \end{cases} \quad (4.2.3)$$

4.2.4. Simulaciones numéricas para el modelo VIH.

El objetivo en esta sección se centra en presentar los resultados numéricos obtenidos al aplicar a los modelos (4.2.2) y (4.2.3) algunos de los métodos numéricos implementados. Debido a que el interés es sobre los métodos numéricos, no se presentarán detalles analíticos ni sugerencias sobre los modelos tratados.

Inicialmente se comparan aproximaciones al modelo (4.2.2) presentadas en [10, 14] y [19] con aproximaciones obtenidas de aplicar la implementación RK44, e incluso como varios investigadores usan el Toolbox ode45 de Matlab también se realizan comparaciones con este método, los parámetros usados en cada caso se muestran en la Tabla 4.10. Luego se encuentran aproximaciones usando algunos de los métodos numéricos estudiados a fin de concluir eficiencia y coherencia en los resultados para cada método. Finalmente, se aplica el método más eficiente al modelo (4.2.3) para simular

la acción de un tratamiento en la dinámica de la enfermedad. En esta sección también se presenta el comportamiento del tamaño de paso en un método adaptativo en relación a la variación de células T , I y virus libre V en una simulación.

Parámetros	Valor ref[10]	Valor ref[14]	Valor ref[16]	Valor ref[19]
$s (dia^{-1}mm^{-3})$	10	0.1	10	10
$p (dia^{-1})$	0.03	3	0.03	0.022
$T_{max} (mm^{-3})$	1500	1500	1500	1500
$d (dia^{-1})$	0.02	0.02	0.02	0.02
$k (mm^3dia^{-1})$	2.4×10^{-5}	2.7×10^{-3}	2.4×10^{-5}	2.9274×10^{-5}
$c (dia^{-1})$	0.24	0.3	0.24	0.3
$u (dia^{-1})$	2.4	2.4	2.4	2.4
N	230	10	774	128
α	0	1	0	0
$T(0) (mm^{-3})$	1000	0.1	1000	500
$I(0) (mm^{-3})$	0	0	0	50
$V(0) (mm^{-3})$	10^{-3}	0.1	10^{-3}	800

Tabla 4.10: Parámetros del modelo [16].

Para empezar se aplica la implementación RK44 al modelo (4.2.2) sujeto a los parámetros presentados en la columna 3 de la Tabla 4.10, esto con el fin de comparar los resultados de la implementación realizada contra los resultados numéricos presentados en el artículo [14]. Aunque en este artículo también se presentan resultados de métodos que no se han estudiado, únicamente se considera los resultados presentados para un método de Runge Kutta de orden 4.

En el artículo [14] no se da detalles sobre el tamaño de paso que se usó con el método de Runge Kutta de orden 4, por lo tanto para comparar estos resultados contra los resultados de la implementación RK44 inicialmente se tomó un tamaño de paso de $h = 0.2$. A pesar de que se observó similitud entre los resultados con una diferencia máxima del orden 10^{-1} , al reducir el tamaño de paso a $h = 0.01$ se obtuvo resultados más similares, con una diferencia máxima del orden 10^{-7} , tal y como se puede observar en las tablas 4.11-4.13. Además en estas tablas se incluyen resultados de aplicar la función ode45 de Matlab con una tolerancia de 10^{-10} , donde se observa que la diferencia máxima con los resultados de la implementación RK44 es del orden 10^{-5} . Por lo tanto se considera que los resultados obtenidos de la implementación RK44 son consistentes con los resultados disponibles en el artículo [14] y los obtenidos de aplicar la función ode45 de Matlab.

t	RK44 h=0.2	RK44 h = 0.01	ref [14]	ode45
0.2	0.2087153941	0.2088080834	0.2088080833	0.2088080843
0.4	0.4059043039	0.4062405394	0.4062405393	0.4062405456
0.6	0.7635094172	0.7644238893	0.7644238890	0.7644239510
0.8	1.411837346	1.414046829	1.414046831	1.414047303
1.0	2.586595129	2.591594801	2.591594802	2.591596560

Tabla 4.11: Comparación numérica para $T(t)$.

t	RK44 h = 0.2	RK44 h = 0.01	ref [14]	ode45
0.2	$0.6019536413 \times 10^{-5}$	$0.6032702150 \times 10^{-5}$	$0.6032702150 \times 10^{-5}$	$0.6032702243 \times 10^{-5}$
0.4	$0.1312991789 \times 10^{-4}$	$0.1315834072 \times 10^{-4}$	$0.1315834073 \times 10^{-4}$	$0.1315834121 \times 10^{-4}$
0.6	$0.2117812628 \times 10^{-4}$	$0.2122378507 \times 10^{-4}$	$0.2122378506 \times 10^{-4}$	$0.2122378974 \times 10^{-4}$
0.8	$0.3011243552 \times 10^{-4}$	$0.3017741955 \times 10^{-4}$	$0.3017741955 \times 10^{-4}$	$0.3017744898 \times 10^{-4}$
1.0	$0.3995116228 \times 10^{-4}$	$0.4003781469 \times 10^{-4}$	$0.4003781468 \times 10^{-4}$	$0.4003789327 \times 10^{-4}$

Tabla 4.12: Comparación numérica para $I(t)$.

t	RK44 h = 0.2	RK44 h = 0.01	ref [14]	ode45
0.2	0.06189950205	0.06187984331	0.06187984331	0.06187984323
0.4	0.03831922214	0.03829488787	0.03829488788	0.03829488787
0.6	0.02372713945	0.02370455013	0.02370455014	0.02370455079
0.8	0.01469899971	0.01468036376	0.01468036377	0.01468036633
1.0	0.00911525297	0.00910084505	0.00910084504	0.00910084895

Tabla 4.13: Comparación numérica para $V(t)$.

Seguidamente, usando la implementación RK44 con los parámetros de las columnas 2 y 5 de la Tabla 4.10 que corresponden a los artículos [10] y [19] respectivamente, se obtienen las figuras 4.15 y 4.16. En ambos casos al comparar con las gráficas incluidas en cada artículo se observa comportamientos similares.

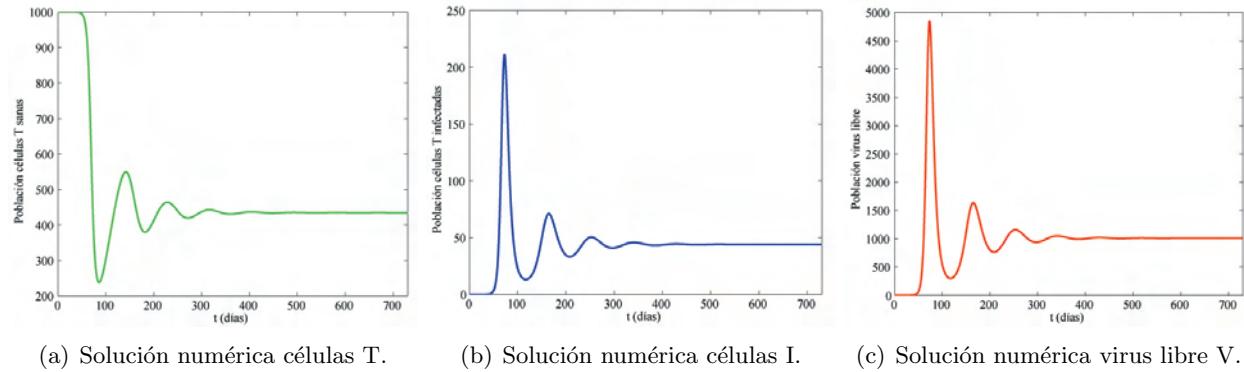


Figura 4.15: Comparación en la solución numérica del método Rk44 y [10].

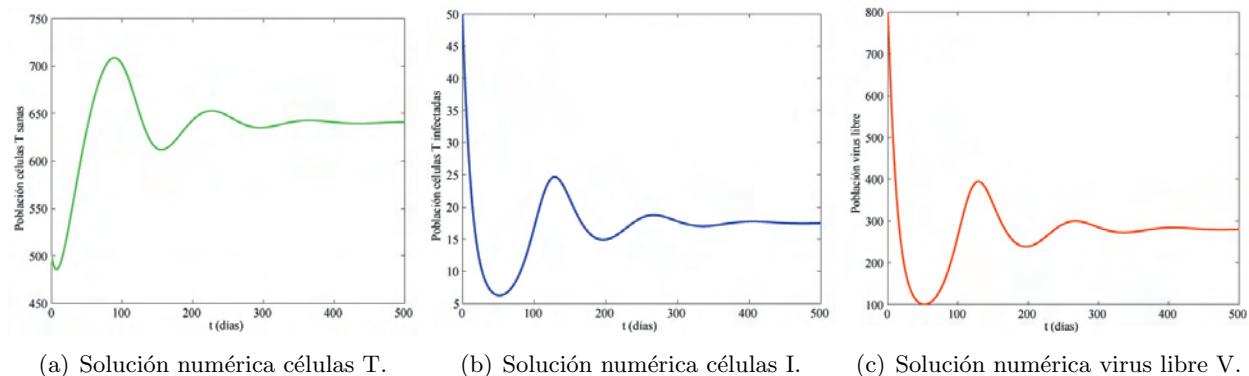


Figura 4.16: Comparación en la solución numérica del método Rk44 y [19].

Los parámetros de la columna 4 en la Tabla 4.10 corresponden a una persona contagiada con VIH, aplicando la implementación RK44 al modelo (4.2.2) sujeto a estos parámetros con tamaño de paso $h = 0.1$ se obtiene la Figura 4.17, donde se observa que el modelo describe importantes características de la enfermedad para una persona infectada en el periodo de tiempo considerado. Inicialmente el número de células $TCD4^+$ sanas disminuye rápidamente mientras que la población de virus libre en la sangre aumenta de forma exponencial. Después de alcanzar un mínimo la población de células $TCD4^+$ sanas aumenta lentamente manteniéndose en niveles muy bajos y el número de virus libre después de alcanzar un punto máximo disminuye rápidamente. Finalmente las poblaciones se estabilizan y permanecen así hasta el fin del periodo de tiempo considerado.

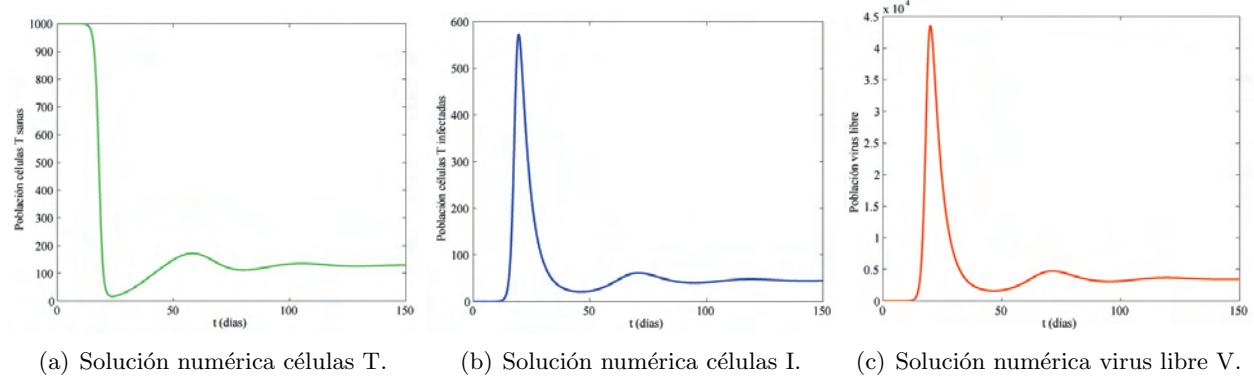


Figura 4.17: Solución numérica para modelo VIH/SIDA, método Rk44.

Para este mismo caso, la Tabla 4.14 presenta las aproximaciones numéricas a T para determinados períodos de tiempo usando diferentes métodos con el mismo tamaño de paso $h = 0.1$. La diferencia entre las soluciones numéricas de cada método parece no ser significativa, sin embargo en la Tabla 4.15, se observa que en el intervalo de tiempo (15, 22) se presentan grandes diferencias, así por ejemplo considerando la interpretación biológica del modelo para el método de Euler en el día 19 de acuerdo a la Tabla 4.9 se concluye que no es conveniente que el paciente reciba tratamiento, mientras que si se observa los resultados obtenidos con métodos de mayor orden como RK44 por ejemplo, se concluye que el paciente debe recibir tratamiento anti-VIH. Por lo tanto el método de Euler no ofrece buenos resultados en el modelo con el tamaño de paso considerado. Si se aplica el método de Euler con tamaño de paso $h = 0.01$ se obtiene que para el día 19 el número de células T sanas es de 242.114808, por lo tanto con este tamaño de paso el método de Euler ofrece buenos resultados. Este fenómeno se debe a que la solución exacta del modelo presenta un cambio brusco en el periodo de tiempo (15, 22) tal y como se muestra en la Figura 4.17, por lo tanto se debería usar un tamaño de paso más pequeño en este intervalo.

t	Euler	RK22	RK33	RK44	RK56	RK44-AB4
10	999.867020	999.780968	999.777113	999.777011	999.777008	999.778740
30	31.119532	34.396355	34.509947	34.513049	34.513111	34.634371
50	146.393706	149.122702	149.268031	149.271438	149.271512	149.484933
70	135.912513	133.152404	133.036372	133.033476	133.033414	132.868112
90	120.987027	122.140415	122.180238	122.181294	122.181316	122.238010
110	135.659506	134.967235	134.954927	134.954559	134.954551	134.937244
130	125.985445	126.291317	126.292036	126.292084	126.292084	126.292926
150	130.439863	130.330041	130.332391	130.332441	130.332442	130.336054

Tabla 4.14: Solución numérica modelo VIH/SIDA.

t	Euler	RK22	RK33	RK44	RK56	RK44-AB4
15	980.570607	959.740511	958.717880	958.690738	958.690163	960.263332
16	948.548818	892.019087	889.304917	889.233252	889.231730	894.896136
17	869.393706	738.070435	732.311699	732.161499	732.158290	748.753427
18	700.295172	482.377732	474.595957	474.397487	474.393194	504.292239
19	436.744082	230.448930	224.813031	224.675766	224.672715	246.861543
20	194.509817	91.544408	89.056851	89.002172	89.000859	93.399884
21	72.397152	39.654764	38.756125	38.739342	38.738879	37.786878
22	31.248216	22.849677	22.535457	22.530898	22.530741	21.610893

Tabla 4.15: Solución numérica modelo VIH/SIDA, falla Euler.

Aplicando el método RF, en la Figura 4.18 se observa que para el periodo de tiempo (15, 22) se considera tamaños de paso más pequeños.

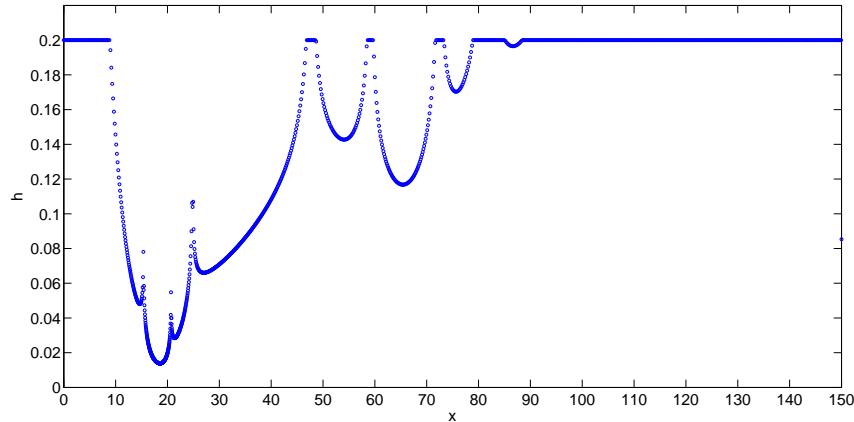


Figura 4.18: Variación del tamaño de paso modelo VIH/SIDA, método RKF.

Solución numérica del modelo con estrategia de control

De las pruebas anteriores se puede concluir que la implementación RK44 ofrece buenos resultados, aplicando esta implementación al modelo (4.2.3) sujeto a los parámetros en la columna 4 de la Tabla 4.10 con tamaño de paso $h = 0.1$, se observa características importantes del efecto de un medicamente en la dinámica del VIH, en la Figura 4.19 y la Tabla 4.16 se puede observar como el tratamiento anti viral retrasa la infección de células $TCDA^+$ sanas y hace que esta población no alcance niveles tan bajos. Por otro lado el tratamiento retrasa el crecimiento en las poblaciones de virus libres y células infectadas, haciendo que no alcancen niveles tan altos.

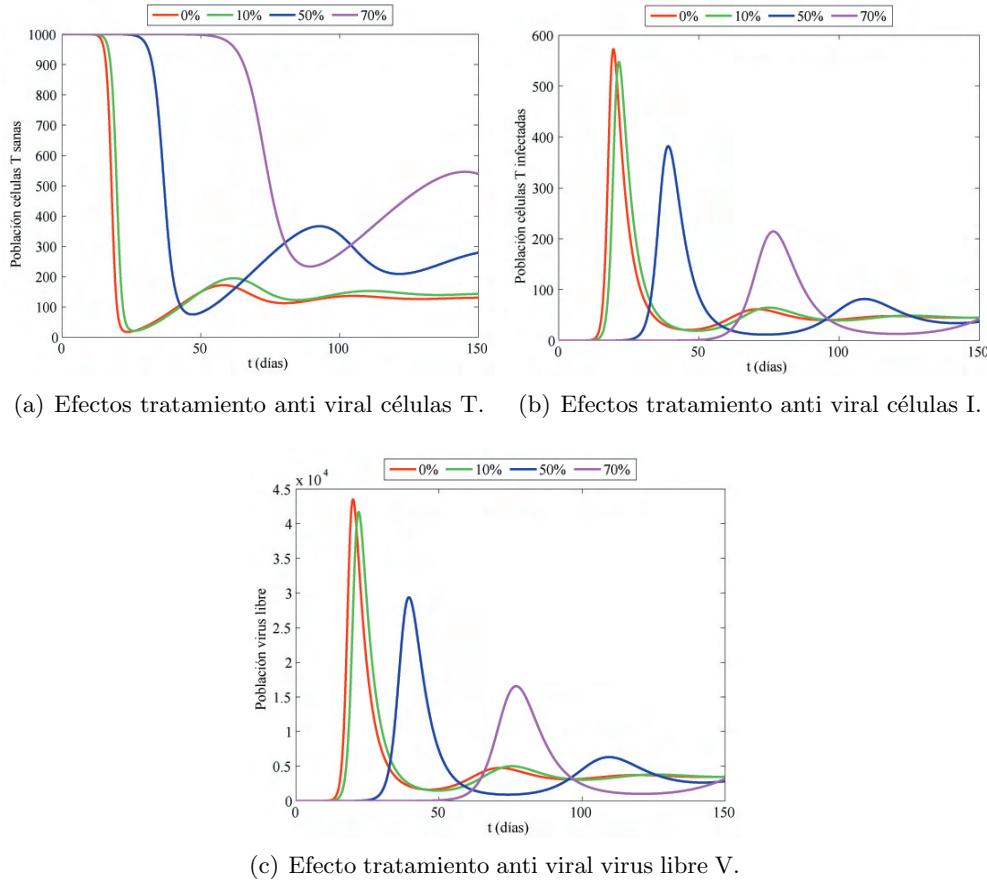


Figura 4.19: Solución numérica para modelo VIH/SIDA con estrategia de control, tratamiento con 0 %, 10 %, 50 % y 70 % de eficiencia, método Rk44.

Eficiencia	Mínimo T	Máximo I	Máximo V
0 %	17.01	575.6	4.35241×10^4
10 %	20.99	549.2	3.49712×10^4
50 %	75.07	381.9	2.93308×10^4
70 %	233.3	214.3	1.65542×10^4

Tabla 4.16: Solución numérica para modelo VIH/SIDA, eficiencia tratamiento, método Rk44.

Capítulo 5

Conclusiones y trabajos futuros

5.1. Conclusiones

- Los conceptos teóricos e implementaciones presentadas en este trabajo se han enfocado a sistemas de ecuaciones diferenciales ordinarias de primer orden, esto permite aplicar estos métodos a una gran variedad de problemas prácticos que tengan que ver con SEDO y EDO de orden superior.
- A pesar de que estamos en una era en que los computadores realizan cálculos rápidamente, en la práctica existen problemas de gran complejidad que demandan altos costos computacionales, por lo tanto de acuerdo al problema que se esté trabajando es importante considerar las ventajas y desventajas al usar MNEDO tal y como se presentó en el Capítulo 4. De acuerdo a los intereses del investigador se puede sacrificar precisión para ganar eficiencia.
- Al momento de aplicar un método numérico, además de verificar las propiedades de consistencia, convergencia y estabilidad, se debe tener en cuenta que los computadores trabajan con una aritmética finita, por lo tanto además de los propios errores generados por el método se están generando errores de redondeo. En los métodos presentados el error de redondeo aumentaba cuando se elegía tamaños de paso muy pequeños y al usar métodos de alto orden que alcanzan el cero computacional rápidamente tal y como se presentó en las figuras 4.1 y 4.12 con los métodos de RK44, RK56 y AB4. Para una elección adecuada del tamaño de paso es importante realizar un estudio de estabilidad absoluta.
- De los resultados numéricos obtenidos en el Capítulo 4 se concluye que de los métodos implementados, el método RK44 en la práctica ofrece una óptima relación entre costos computacionales y precisión.
- Muchos de los modelos prácticos incluyen EDO no lineales para los que no se ha determinado una solución analítica. En este tipo de modelos las simulaciones numéricas permiten concluir

características generales para la solución, tal y como se muestra en la sección 4.2, donde se estudia numéricamente el impacto de un medicamento contra para el VIH/SIDA.

5.2. Trabajos futuros

Este trabajo abre las puertas a una gran variedad de proyectos futuros, ya que por un lado se puede aplicar los métodos presentados en diferentes problemas prácticos, así como también se puede profundizar en los temas tratados y presentar métodos más avanzados para resolver el problema de Cauchy. A continuación se presentan algunos temas sobre los cuales se puede profundizar

- Métodos para estimar el error global de discretización.
- Métodos Implícitos.
- Adaptatividad en MNEDO.
- Métodos predictor corrector.
- Solución numérica en problemas rígidos.
- Métodos de paso múltiple no lineales.

Apéndice

A.1. Interpolación

El problema de la interpolación consiste en encontrar una función real g cuyo gráfico coincida con un conjunto de puntos dado $(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$; es decir que

$$g(x_i) = f(x_i) \text{ para } 0 \leq i \leq n. \quad (\text{A.1.1})$$

En otras palabras se busca una función g que aproxime a la función f en el intervalo $[x_0, x_n]$.

El proceso de interpolación es de suma importancia en diferentes tópicos del análisis numérico tales como la integración numérica, el cálculo de raíces en ecuaciones no lineales. Incluso es frecuentemente usado en la práctica para encontrar valores intermedios que no están en una tabla de datos experimentales. En este trabajo usaremos el concepto de interpolación para la deducción de MPM para aproximar la solución al problema de Cauchy (1.3.1). El método de interpolación a usarse es el método de Lagrange.

Considerando que los polinomios son funciones sencillas y ampliamente estudiadas es conveniente considerar que la función g en (A.1.1) sea un polinomio. Por lo tanto el problema de interpolación se ha transformado en la búsqueda de un polinomio, llamado **polinomio interpolador**, que cumpla con las condiciones (A.1.1), este tipo de interpolación se conoce como **interpolación polinómica**.

Definición A.1.1. Sean $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ un conjunto de números reales o complejos, siendo todos los valores x_i diferentes. El problema de **interpolación polinómica** consiste en encontrar un polinomio $P(x)$ de grado $m \leq n$, tal que

$$P(x_i) = y_i, \text{ para } i = 0, 1, 2, \dots, n.$$

El siguiente teorema muestra que este polinomio existe y que es único.

Teorema A.1.1. *Dados $n + 1$ puntos distintos x_0, x_1, \dots, x_n , reales, y sus correspondientes valores y_0, y_1, \dots, y_n . Existe uno y solo un polinomio, $P(x)$, de grado menor o igual a n , tal que*

$$P(x_i) = y_i, \text{ para } i = 0, 1, 2, \dots, n. \quad (\text{A.1.2})$$

Demostración. Sea $P(x) = a_0 + a_1x + \dots + a_nx^n$, un polinomio de grado menor o igual a n , con los $n+1$ coeficientes a_0, a_1, \dots, a_n a ser determinados. Usando la ecuación (A.1.2) se tiene

$$\begin{aligned} a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n &= y_0, \\ a_0 + a_1x_1 + a_2x_1^2 + \dots + a_nx_1^n &= y_1, \\ &\vdots \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_nx_n^n &= y_n. \end{aligned}$$

Este es un sistema lineal de $n+1$ ecuaciones con $n+1$ incógnitas. En notación matricial este sistema se puede escribir como

$$V\mathbf{a} = \mathbf{y}, \quad (\text{A.1.3})$$

donde

$$V = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Claramente resolver el sistema (A.1.3) es equivalente a encontrar el polinomio interpolador $P(x)$. Para garantizar la existencia y unicidad del polinomio $P(x)$ basta probar que el sistema (A.1.3) tiene solución única, es decir, se debe probar que el determinante de la matriz V es diferente de 0. La matriz V es llamada **matriz de Vandermonde** [2], y su determinante es

$$\det(V) = \prod_{0 \leq j < i \leq n} (x_i - x_j),$$

y por hipótesis los puntos x_0, x_1, \dots, x_n son distintos, por lo tanto se concluye que $\det(V) \neq 0$. \square

El Teorema A.1.1 asegura la existencia y unicidad de un polinomio interpolador, sin embargo no podemos usar este teorema para obtener tal polinomio. A seguir presentamos una de las técnicas más básicas que permite calcular un polinomio interpolador, conocida como **fórmula de Lagrange**.

El problema es construir un polinomio de grado no máximo n cuyo gráfico pase por los $n+1$ puntos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, a este fin para $k = 0, 1, \dots, n$ se construye una función $L_{n,k}(x)$ tal que

$$L_{n,k}(x_i) = \begin{cases} 0, & \text{si } k \neq i \\ 1 & \text{si } k = i. \end{cases}$$

La función que cumple con esta propiedad es

$$L_{n,k}(x_i) = \frac{(x - x_0) \dots (x - x_k)(x - x_{k+1}) \dots (x - x_n)}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)} = \prod_{i=0}^n \frac{(x - x_i)}{(x_k - x_i)}. \quad (\text{A.1.4})$$

El siguiente teorema describe el proceso para obtener el polinomio de Lagrange usando (A.1.4).

Teorema A.1.2. Si x_0, x_1, \dots, x_n son $n + 1$ números distintos y si f es una función cuyos valores están dados en estos números, entonces existe un único polinomio $P(x)$ de grado menor o igual que n , con la propiedad

$$P(x_k) = f(x_k) \text{ para cada } k = 0, 1, \dots, n.$$

Este polinomio está dado por

$$P(x) = f(x_0)L_{n,0}(x) + \dots + f(x_n)L_{n,n}(x) = \sum_{k=0}^n f(x_k)L_{n,k}(x) \quad (\text{A.1.5})$$

y se conoce como **polinomio interpolador de Lagrange**.

Teorema A.1.3. Sean x_0, x_1, \dots, x_n , $n + 1$ números distintos del intervalo $[a, b]$ y $f \in C^{n+1}$ en $[a, b]$. Entonces, para cada $x \in [a, b]$ existe un número $\varsigma(x)$ en (a, b) tal que

$$f(x) = P(x) + \frac{f^{(n+1)}(\varsigma(x))}{(n+1)!} \prod_{i=0}^n (x - x_i),$$

donde $P(x)$ es el polinomio interpolador de Lagrange (A.1.5).

Del Teorema A.1.3 se obtiene una fórmula para el error del método de Lagrange, esta expresión es de suma importancia en la deducción de métodos de integración numérica tal y como se presenta en [5].

A.2. Adaptatividad y estimación para el error

En este apéndice se estudian métodos en los cuales se controla en cada paso los errores mediante la variación del tamaño de paso, los métodos que tienen esta propiedad se conocen como **métodos adaptativos**.

Estimativa del error global de discretización

Muchas de las técnicas para estimar el error global de discretización se basan en el cálculo de dos soluciones numéricas del problema de Cauchy, de manera que la diferencia entre estas dos soluciones pueda darnos una idea de la magnitud y el comportamiento del error durante el proceso de integración. Una de las técnicas para aproximar el error global de discretización se conoce como **método de extrapolación de Richardson** presentado en [7], el cual se deduce del siguiente teorema tomado de [20].

Teorema A.2.1. Sea la función $f(x, y)$ con $N + 2$ derivadas parciales con relación a y continuas y acotadas en (a, b) y sea $y_{(h)}$ la solución numérica del problema de Cauchy obtenida por la aplicación de un MPU de orden p , con $p \leq N$ determinada con paso de integración h . Bajo estas condiciones la solución numérica $y_{(h)}$ admite una expansión asintótica de la forma

$$y_{(h)} = y(x) + h_p g_p(x) + h^{p+1} g_{p+1}(x) + \dots + h^N g_N(x) + h^{(N+1)} G_{(N+1)}(x, h), \quad (\text{A.2.1})$$

con $g_j(x_0) = 0$ para $j = p, p + 1, \dots$, válidas para todo $x \in (a, b)$ y para todo $h > 0$.

Nótese que las funciones g_j son independientes del valor de h y que $G_{N+1}(x, h)$ es un residuo.

Aplicando la expansión asintótica (A.2.1) podemos escribir una expresión para el error global de discretización así

$$e_k = y(x_k) - y_{(k,h)} = - \left[\sum_{j=p}^N h^j g_j(x) + h^{N+1} G_{N+1}(x, h) \right]. \quad (\text{A.2.2})$$

La expresión (A.2.2) será usada para obtener una aproximación numérica al error global de discretización. El proceso es el siguiente.

Para un paso de integración h obtenemos la solución numérica $y_{(h,k)}$, para el mismo x_k se obtiene la solución numérica $y_{(h/2,k)}$. Luego para h lo suficientemente pequeño se tiene que

$$e_{(k,h)} = y(x_k) - y_{(k,h)} \approx -g_p(x)h^p, \quad (\text{A.2.3})$$

$$e_{(k,h/2)} = y(x_k) - y_{(k,h/2)} \approx -g_p(x) \left(\frac{h}{2}\right)^p. \quad (\text{A.2.4})$$

Restando (A.2.3) y (A.2.4), se tiene

$$y_{(k,h)} - y_{(k,h/2)} \approx g_p(x) \left[h^p - \left(\frac{h}{2}\right)^p \right] = g_p(x) \left(\frac{h}{2}\right)^p (2^p - 1). \quad (\text{A.2.5})$$

Por lo tanto

$$g_p(x)(h/2)^p \approx \frac{y_{(k,h)} - y_{(k,h/2)}}{2^p - 1}. \quad (\text{A.2.6})$$

Sustituyendo (A.2.6) en (A.2.4), se tiene una expresión para aproximar el error global de discretización

$$e_{(k,h/2)} \approx -\frac{y_{(k,h)} - y_{(k,h/2)}}{2^p - 1}.$$

Estimativa del error local de discretización

La idea consiste en obtener una aproximación numérica del error local de discretización a partir de la aplicación de dos MPU, uno de orden n y otro de orden $n+1$. Este método fue desarrollado por Merson en 1957 y cobra una particular importancia al aplicarse en los métodos de Runge kutta.

Consideremos dos MPU, uno de orden n y otro de orden $n+1$. Supongamos que al aplicar estos métodos al problema de Cauchy (1.3.1), con tamaño de paso fijo h , se obtienen respectivamente las soluciones numéricas

$$y_{k+1} = y_k + h\phi(x_k, y_k, h), \quad (\text{A.2.7})$$

$$\bar{y}_{k+1} = \bar{y}_k + h\bar{\phi}(x_k, \bar{y}_k, h). \quad (\text{A.2.8})$$

Observe que para la aproximación (A.2.7) de orden n el error local $\tau_{k+1} = O(h^n)$, mientras que para la aproximación (A.2.8) de orden $n+1$ el error local $\bar{\tau}_{k+1}(h) = O(h^{n+1})$. Los dos errores locales de discretización satisfacen la ecuación (2.2.3), por lo tanto

$$\tau_{k+1}(h) = \frac{1}{h} (y(x_{k+1}) - y_{k+1}), \quad (\text{A.2.9})$$

$$\bar{\tau}_{k+1}(h) = \frac{1}{h} (y(x_{k+1}) - \bar{y}_{k+1}). \quad (\text{A.2.10})$$

Sumando y restando \bar{y}_{k+1} en la expresión (A.2.9) y agrupando términos se tiene

$$\begin{aligned}\tau_{k+1} &= \frac{1}{h} [(y(x_{k+1}) - \bar{y}_{k+1}) + (\bar{y}_{k+1} - y_{k+1})] \\ &= \bar{\tau}_{k+1} + \frac{1}{h} (\bar{y}_{k+1} - y_{k+1}).\end{aligned}$$

Pero τ_{k+1} es $O(h^n)$ y $\bar{\tau}_{k+1}$ es $O(h^{n+1})$, por lo tanto el término $\bar{\tau}_{k+1}$ no es significativo en la expresión de τ_{k+1} . Si se omite este término, se obtiene una expresión para aproximar el error local de discretización para el método $O(h^n)$ tal que

$$\tau_{k+1} \approx \frac{1}{h} (\bar{y}_{k+1} - y_{k+1}). \quad (\text{A.2.11})$$

Al igual que el método de extrapolación de Richardson este método para aproximar el error local de discretización tiene un elevado costo computacional. Sin embargo, si aplicamos este método a los métodos de Runge Kutta es posible calcular una expresión más eficiente para aproximar el error local, evitando el cálculo de las dos soluciones numéricas en la expresión (A.2.11) tal y como se presenta en [20].

Ejemplo .1. Considérese el método de Euler modificado (2.4.9) y el método Runge-Kutta de orden 3 (2.4.10). Aplicando la ecuación (A.2.11) se tiene

$$\begin{aligned}\tau_{k+1} &\approx \frac{1}{h} \left[y_k + \frac{h}{6} (k_1 + 4k_2 + k_3) - (y_k + hk_2) \right] \\ &= \frac{1}{6} (k_1 - 2k_2 + k_3).\end{aligned}$$

Por lo tanto, se determinó una expresión para aproximar el error local de discretización para el método de Euler modificado

$$\tau_{k+1} \approx \frac{1}{6} (k_1 - 2k_2 + k_3).$$

La aproximación del error local obtenida de combinar los métodos RK44 y RK45 es muy importante en la práctica ya que de esta se obtiene el método de Runge Kutta Fehlberg, RKF, un método que controla automáticamente el tamaño de paso.

Control del tamaño de paso

Tanto el método de extrapolación de Richardson como el método de Melene pueden ser usados para controlar el paso de integración, el objetivo es modificar el tamaño de paso de tal forma que no se permita un error por encima de cierta tolerancia fijada para el problema de estudio, los métodos numéricos que tienen esta propiedad se conocen como **métodos adaptativos**.

Las modificaciones en el tamaño de paso consisten en multiplicarlo por un factor adecuado δ , que aumenta o disminuye el tamaño de paso. Evidentemente el error disminuye en la medida en que el tamaño de paso h disminuye, por tanto si el error estimado sobrepasa la tolerancia fijada ϵ , se debe disminuir el tamaño de paso. Por otro lado, trabajar con tamaños de paso muy pequeños

en lugares donde la función que define el problema de Cauchy presenta un buen comportamiento conlleva a realizar esfuerzos innecesarios, lo cual se traduce en altos costos computacionales por lo cual en el caso en el que los errores estimados no sean significativos se puede aumentar el tamaño de paso.

Una idea general de estas conclusiones, se presentan en los siguientes pasos algorítmicos:

1. Estimar el error (puede ser el error global o local).
2. si $error > \epsilon$ (ϵ es la tolerancia fijada) entonces
reducir el tamaño de paso $h \leftarrow \delta h$, $0 < \delta < 1$ y regresar a 1.
Caso contrario ir a 3.
3. Almacenar aproximación numérica en el vector y_k , hacer $x_k \leftarrow \min\{x_k + h, b\}$ y
aumentar el tamaño de paso $h \leftarrow \delta h$, $\delta > 1$.
4. si $0 < x_k < b$ entonces regresar a 1. Caso contrario parar

Es de especial interés mantener controlado el error global de discretización ya que en gran medida la validez de los resultados de la integración numérica dependen de este valor, sin embargo estimar numéricamente el error global de discretización demanda un alto costo computacional. Por esta razón, en muchos métodos adaptativos se trabaja con técnicas de estimaciones del error local de discretización como por ejemplo el método de Melene que aplicado a los métodos de Runge Kutta resulta ser más eficiente que el método de extrapolación de Richardson tal y como se muestra en [20].

Usando el método de Melene para los métodos RK45 y RK56 se puede construir un método adaptativo muy usado en la práctica conocido como el método de **Runge Kutta Fehlberg**, RKF. La expresión para aproximar el error local de discretización en este método es:

$$\tau_{k+1} = \frac{1}{360}k_1 - \frac{128}{4275}k_3 - \frac{2197}{75240}k_4 + \frac{1}{50}k_5 + \frac{2}{55}k_6. \quad (\text{A.2.12})$$

Una ventaja de este método es que únicamente requiere 6 evaluaciones de la función f por paso, ya que para los métodos RK45 y RK56 las evaluaciones de esta función k_1, k_2, k_3, k_4 y k_5 coinciden, por lo tanto es necesario calcularlas una sola vez junto con k_6 para obtener una aproximación numérica del error local usando (A.2.12).

La forma más sencilla de variar el tamaño de paso es duplicar o partir a la mitad el tamaño de paso, es decir que $\delta = 2$ o $\delta = 0,5$ según sea el caso, sin embargo esta técnica no es tan eficiente y se debe analizar que por cada intento fallido se debe repetir el proceso y esto demanda altos costos computacionales. Una forma mas sofisticada de elegir el factor δ según el cual se aumenta o disminuye el tamaño de paso puede verse en . Esta técnica se deduce del conocimiento del orden del método utilizado. Suponga que el método tiene orden n , de modo que τ_{k+1} es $O(h^n)$, por tanto existe un K independiente de h tal que

$$\tau_{k+1} = Kh^n.$$

Usando el método de Melene con tamaño de paso δh , se tiene

$$\tau_{k+1} \approx K(\delta h)^n = \delta^n(Kh^n) \approx \delta^n \tau_{k+1} \approx \frac{\delta^n}{h}(\bar{y}_{k+1} - y_{k+1}).$$

La aproximación del error local τ_{k+1} debe de estar acotada por la tolerancia fijada ϵ , por tanto

$$\tau_{k+1} \approx \frac{\delta^n}{h} (\bar{y}_{k+1} - y_{k+1}) \leq \epsilon,$$

es decir

$$\delta \leq \left(\frac{\epsilon h}{|\bar{y}_{k+1} - y_{k+1}|} \right)^{\frac{1}{n}}.$$

Se calculó una cota para el factor δ , la cual es para el método de Runge Kutta Fehlberg, la elección común del factor δ es

$$\delta = 0.84 \left(\frac{\epsilon h}{|\bar{y}_{k+1} - y_{k+1}|} \right)^{1/4}.$$

A.3. Ecuaciones de diferencias

Mientras que las ecuaciones diferenciales se ocupan de las funciones de una variable continua, las ecuaciones de diferencias se ocupan de las funciones de una variable discreta. En lugar de una fórmula para la derivada de una función escrita en términos de la función en sí, tenemos que considerar las sucesiones para las cuales cada miembro está relacionado de una manera específica con su predecesor inmediato o varios de sus predecesores más recientes, tal y como se muestra en [7]. Así podemos escribir

$$y_j = \phi_j(y_{j-1}, y_{j-2}, \dots, y_{j-k}) \text{ para } 0 \leq j \leq k,$$

donde k es el orden de la ecuación de diferencias.

En este trabajo se usa las ecuaciones de diferencias para analizar la estabilidad de los MPML, ya que las expresiones que definen los MPML son ecuaciones de diferencias lineales, las cuales se definen a seguir.

Definición A.3.1. Se denomina **ecuación de diferencias lineal con coeficientes constantes** a una expresión de la forma

$$\gamma_N y_{k+N} + \gamma_{N-1} y_{k+N-1} + \dots + \gamma_0 y_k = \phi_k \text{ para } 0 \leq j \leq k, \quad (\text{A.3.1})$$

donde cada γ_j , $0 \leq j \leq N$, son constantes, con $\gamma_0 \neq 0$ y $\gamma_N \neq 0$.

la solución de (A.3.1) es una sucesión

$$y_k, y_{k+1}, y_{k+2}, \dots,$$

lo cual se denota $\{y_j\}_{j \in \mathbb{N}}$.

La ecuación (A.3.1) se denomina ecuación de diferencias homogénea si $\phi_k = 0$ para todo $k \in \mathbb{N}$, caso contrario se denomina ecuación de diferencias no homogénea.

Considérese la ecuación de diferencias homogénea

$$\gamma_N y_{k+N} + \gamma_{N-1} y_{k+N-1} + \dots + \gamma_0 y_k = 0. \quad (\text{A.3.2})$$

Si la solución de (A.3.2) se denotamos por $\{\hat{y}_j\}_{j \in \mathbb{N}}$ y una solución particular de la ecuación de diferencias no homogénea (A.3.1), se denota por $\{\psi_j\}_{j \in \mathbb{N}}$. La solución general de (A.3.1) es entonces una sucesión $\{y_j\}_{j \in \mathbb{N}}$ donde cada elemento está dado por

$$y_j = \hat{y}_j + \psi_j, \quad j \in \mathbb{N}.$$

Un caso básico para calcular la solución particular de (A.3.1) se presenta cuando ϕ_k es constante, es decir cuando $\phi_k = \phi \in \mathbb{R}$ para todo $k \in \mathbb{N}$. En este caso una solución particular de la ecuación en diferencias no homogénea viene dada por

$$\psi_k = \frac{\phi}{\sum_{j=0}^N \gamma_j}, \quad k = 0, 1, 2, \dots,$$

siempre que el denominador sea diferente de cero.

Definición A.3.2. El conjunto de K soluciones de la ecuación de diferencias homogénea (A.3.2), denotado por $\{y_{(N,x)}\}$, $x = 1, 2, \dots, K$; se dice **linealmente independiente** si la combinación lineal

$$a_1 y_{j,1} + a_2 y_{j,2} + \dots + a_K y_{j,K} = 0 \quad \text{para } j \in \mathbb{N},$$

implica que $a_x = 0$ para $x = 1, 2, \dots, K$.

El conjunto de K soluciones $\{y_{n,x}\}$, $x = 1, 2, \dots, K$; linealmente independientes, forma un sistema fundamental y toda solución es dada por

$$\left\{ \sum_{x=1}^K d_x y_{x,t} \right\}_{t \in \mathbb{N}}. \quad (\text{A.3.3})$$

Una técnica para calcular la solución de la ecuación de diferencias homogénea (A.3.2) es suponer que la solución es de la forma

$$y_k = z^k. \quad (\text{A.3.4})$$

Remplazando (A.3.4) en (A.3.3) se tiene

$$z^k \sum_{j=0}^N \alpha_j z^j. \quad (\text{A.3.5})$$

La expresión (A.3.5) puede escribirse como

$$z^k \rho_N(z),$$

donde $\rho_n(z) = \sum_{j=0}^N \alpha_j z^j$.

Por lo tanto una solución trivial de (A.3.5) será $z^k = 0$ y la solución general se presenta cuando z es una raíz del polinomio $\rho_N(z)$, luego la solución general de (A.3.3) es

$$y_k = \sum_{j=1}^N \alpha_j z_j^k.$$

En el caso de raíces múltiples, la solución se modifica. Si $\rho(z)$ tiene raíces r_j , $j = 1, 2, \dots, p$; y la raíz r_j tiene multiplicidad m_j , la solución general de (A.3.3) es $\{y_n\}$, donde

$$\begin{aligned} y_k &= [d_{1,1} + d_{1,2}k + \dots + d_{1,m_1}k(k-1)(k-2)\dots(k-m_1+2)]r_1^k + \\ &\quad + [d_{2,1} + d_{2,2}k + \dots + d_{2,m_2}k(k-1)(k-2)\dots(k-m_2+2)]r_2^k + \\ &\quad \vdots \\ &\quad + [d_{p,1} + d_{p,2}k + \dots + d_{p,m_p}k(k-1)(k-2)\dots(k-m_p+2)]r_p^k. \end{aligned}$$

A.4. Implementaciones

En este apéndice se presentan las implementaciones para los siguientes MNEDO usando el software MATLAB: el método de Euler y los métodos Runge Kutta 44 tanto para el caso real como para sistemas. Las implementaciones de los otros métodos se incluyen en el DVD anexo de este trabajo de grado.

Las implementaciones de los métodos de Euler y Runge Kutta 44 resuelven el problema de Cauchy

$$\begin{cases} y'(x) = f(x, y(x)) \\ y(x_0) = y_0, \end{cases}$$

donde $f : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$ y $x \in [a, b]$.

1. Implementación método de Euler (1.3.3)

```

1 % Método de Euler
2 % Parámetros de entrada
3 % a: Extremo inicial del intervalo de integración.
4 % b: Extremo final del intervalo de integración.
5 % n: Número de divisiones del intervalo de integración.
6 % y0: Condición inicial.
7 % f: Función problema de Cauchy  $y' = f(x, y)$ . Esta función se define
8 % en un archivo exterior.
9 % Salida
10 % x: Vector de puntos del intervalo  $[a, b]$  para la discretización.
11 % y: Vector solución numérica para los puntos x considerados.
12
13 function [x,y]=Euler(a,b,n,y0)
14
15     h = (b-a)/n; % Tamaño de paso.
16     x = zeros(1,n+1);
17     y = zeros(1,n+1);
18
19     y(1) = y0;
20     x(1) = a;
21
22     for k = 1:n
23         x(k+1) = x(k) + h;
24         y(k+1) = y(k) + h*f(x(k),y(k));
25     end
26 end

```

2. Implementación método de Runge Kutta 44 (2.4.11)

```

1 % Método de Runge Kutta 44
2 % Parámetros de entrada
3 % a: Extremo inicial del intervalo de integración .
4 % b: Extremo final del intervalo de integración .
5 % n: Número de divisiones del intervalo de integración .
6 % y0: Condición inicial .
7 % f: Función problema de Cauchy  $y' = f(x, y)$ . Esta función se define
8 % en un archivo exterior .
9 % Salida
10 % x: Vector de puntos del intervalo  $[a, b]$  para la discretización .
11 % y: Vector solución numérica para los puntos x considerados .
12
13 function [x,y]=Runge_Kutta_44(a,b,n,y0)
14
15         h = (b-a)/n; % Tamaño de paso .
16         x = zeros(1,n+1);
17         y = zeros(1,n+1);
18
19         y(1) = y0;
20         x(1) = a;
21
22         for k = 1:n
23             x(k+1) = x(k) + h;
24             k1 = f(x(k),y(k));
25             k2 = f(x(k)+h*0.5,y(k)+h*0.5*k1);
26             k3 = f(x(k)+h*0.5,y(k)+h*0.5*k2);
27             k4 = f(x(k)+h,y(k)+h*k3);
28
29             y(k+1) = y(k) + h*(k1+2.0*k2+2.0*k3+k4)/6.0;
30         end
31
32 end

```

La siguiente implementación corresponde al método de Runge Kutta 44 para sistemas resuelve el problema de Cauchy

$$\begin{cases} \mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x)) \\ \mathbf{y}(\mathbf{x}_0) = \mathbf{y}_0, \end{cases}$$

donde $\mathbf{f} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ y $x \in [a, b]$.

3. Implementación método de Runge Kutta 44 sistema (2.4.11)

```

1 % Método de Runge Kutta 44 para sistemas de 3 EDO.
2 % Parámetros de entrada
3 % a: Extremo inicial del intervalo de integración .
4 % b: Extremo final del intervalo de integración .
5 % n: Número de divisiones del intervalo de integración .
6 % y0: Condición inicial para la primera ecuación del sistema .
7 % y1: Condición inicial para la segunda ecuación del sistema .
8 % y2: Condición inicial para la tercera ecuación del sistema .
9 % f1: Primera función del sistema en el problema de Cauchy  $y' = f(x, y)$ .
10 % f2: Segunda función del sistema en el problema de Cauchy  $y' = f(x, y)$ .
11 % f3: Tercera función del sistema en el problema de Cauchy  $y' = f(x, y)$ .

```

```
12 % Las funciones f1 , f2 y f3 se definen en archivos exteriores .
13 % Salida
14 % x: Vector de puntos del intervalo [a,b] para la discretización .
15 % y: Matriz solución numérica del sistema para los puntos x considerados .
16
17 function [y ,x]=Runge_Kutta_44_Sistema(a ,b ,n ,y0 ,y1 ,y2)
18
19     h = (b-a)/n; % Tamaño de paso .
20     x = zeros(1 ,n+1);
21     y = zeros(3 ,n+1);
22
23     y(1 ,1) = y0 ;
24     y(2 ,1) = y1 ;
25     y(3 ,1)= y2;
26     x(1) = a;
27
28 for k = 1:n
29     x(k+1) = x(k) + h;
30     % Cálculo vector K1
31     k1 = f1(x(k) ,y(1,k) ,y(2,k) , y(3 ,k));
32     w1 = f2(x(k) ,y(1,k) ,y(2,k) , y(3 ,k));
33     z1 = f3(x(k) ,y(1,k) ,y(2,k) , y(3 ,k));
34     K1 = [k1; w1; z1]; % Vector K1
35
36     % Cálculo vector K2
37     k2 = f1(x(k)+0.5*h,y(1,k)+0.5*h*k1 , y(2 ,k)+0.5*h*w1 , y(3 ,k)+0.5*h*z1 );
38     w2 = f2(x(k)+0.5*h,y(1,k)+0.5*h*k1 , y(2 ,k)+0.5*h*w1 , y(3 ,k)+0.5*h*z1 );
39     z2 = f3(x(k)+0.5*h,y(1,k)+0.5*h*k1 , y(2 ,k)+0.5*h*w1 , y(3 ,k)+0.5*h*z1 );
40     K2 = [k2; w2; z2]; % Vector K2
41
42     % Cálculo vector K3
43     k3 = f1(x(k)+0.5*h,y(1,k)+0.5*h*k2 , y(2 ,k)+0.5*h*w2 , y(3 ,k)+0.5*h*z2 );
44     w3 = f2(x(k)+0.5*h,y(1,k)+0.5*h*k2 , y(2 ,k)+0.5*h*w2 , y(3 ,k)+0.5*h*z2 );
45     z3 = f3(x(k)+0.5*h,y(1,k)+0.5*h*k2 , y(2 ,k)+0.5*h*w2 , y(3 ,k)+0.5*h*z2 );
46     K3 = [k3; w3; z3]; % Vector K3
47
48     % Cálculo vector K4
49     k4 = f1(x(k+1) ,y(1,k)+h*k3 , y(2 ,k)+h*w3 , y(3 ,k)+h*z3 );
50     w4 = f2(x(k+1) ,y(1,k)+h*k3 , y(2 ,k)+h*w3 , y(3 ,k)+h*z3 );
51     z4 = f3(x(k+1) ,y(1,k)+h*k3 , y(2 ,k)+h*w3 , y(3 ,k)+h*z3 );
52     K4 = [k3; w3; z3]; % Vector K4
53
54     y(:,k+1) = y(:,k) + h*(K1 + 2*K2 + 2*K3 + K4) / 6.0;
55 end
56 end
```

Referencias

- [1] Apostol, T. M. (1996). *Análisis matemático*. Barcelona: Editorial Reverté.
- [2] Atkinson, K., Han, W., & Stewart, D. E. (2011). *Numerical solution of ordinary differential equations*. New York: John Wiley & Sons.
- [3] Bevilacqua, J. S., Rafikov, M., & Guedes, C. D. (2002). *Modelagem em biomatemática*. São José do Rio Preto: USP.
- [4] Braun, M. (1992). *Ecuaciones diferenciales y sus aplicaciones*. México: Grupo Editorial Iberoamérica.
- [5] Burden, R. L., Faires, J. D., & Mochon, S. (1985). *Análisis numérico*. México: Grupo Editorial Iberoamericana.
- [6] Butcher, J. C. (2000). *Numerical methods for ordinary differential equations in the 20th century*. Journal of Computational and Applied Mathematics, 125(1), 1-29.
- [7] Butcher, J.C. (1933). *Numerical methods for ordinary differential equations*. New York: John Wiley & Sons.
- [8] Fatunla, S. O. (1982). *Nonlinear multistep methods for initial value problems*. Computers & Mathematics with Applications, 8(3), 231-239.
- [9] Fatunla, S. O. (2014). *Numerical methods for initial value problems in ordinary differential equations*. New York: Academic Press.
- [10] Grapiglia, G. N., Angelossi, K. H., & Rizzi, R. L. (2008). *Modelos Matemáticos da Dinâmica do HIV*. Anais da XXII Semana Académica da Matemática. Cascavel: Universidade Estadual do Oeste do Paraná.
- [11] Grossman, S. I., & Soto, F. P. (1983). *Álgebra lineal*. Grupo Editorial Iberoamericana.
- [12] Lambert, J. D. (1973). *Computational methods in ordinary differential equations*. New York: John Wiley & Sons.
- [13] Nagle, R. K., Saff, E. B., & Snider, A. D. (2001). *Ecuaciones diferenciales y problemas con valores en la frontera*. México: Pearson Educación.
- [14] Ongun, M. Y. (2011). *The Laplace adomian decomposition method for solving a model for HIV infection of CD4+ T cells*. Mathematical and Computer Modelling, 53(5), 597-603.

- [15] ONUSIDA. *Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA*. Disponible en <http://www.unaids.org/es/dataanalysis>, 2016.
- [16] Perelson, A. S., Kirschner, D. E., & De Boer, R. (1993). *Dynamics of HIV infection of CD4+ T cells*. Mathematical biosciences, 114(1), 81-125.
- [17] Perelson, A. S., & Nelson, P. W. (1999). *Mathematical analysis of HIV-1 dynamics in vivo*. SIAM review, 41(1), 3-44.
- [18] Perko, L. (2013). *Differential equations and dynamical*. New York: Springer Science & Business Media.
- [19] Ramírez, J. A., Estrada, J. L., De León, C. V., & Cano, M. V. (2016). *Un modelo matemático para el VIH/SIDA*. Journal of Basic Sciences, 1(3).
- [20] Roma, A. M., & Bevilacqua, J. S. (2004). *Métodos Lineares de n-passos para Solução Numérica de Equações Diferenciais Ordinárias*. São Carlos: SBMAC.
- [21] Spiegel, M. R. (1983). *Ecuaciones diferenciales aplicadas*. New York: Prentice Hall.
- [22] Yang, H., Nkeze, J., & Zhao, R. Y. (2012). *Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy*. Cell & bioscience, 2(1),1.
- [23] Zill, D. G. (2002). *Ecuaciones diferenciales con aplicaciones de modelado*. México: Thomson Learning.

Índice alfabético

- Condición de Lipschitz, 4, 5
- Condiciones iniciales, 3, 26, 46, 47
- Consistencia, 25, 26, 46, 49
- Convergencia, 25, 26, 46, 49, 55, 68
- Costo computacional, 14, 29, 65, 71
- Discretización, 14
- Ecuación de diferencias, 48, 90
- Ecuación diferencial de orden superior, 2
- Ecuación diferencial ordinaria, 1
- Error global de discretización, 24, 86
- Error local de discretización, 24, 45, 87
- Estabilidad, 26, 46, 49
- Estabilidad absoluta, 32, 49, 58
- Integración numérica, 40
- Interpolación, 40, 43, 84
- Intervalo de estabilidad absoluta, 35, 59
- Método de Adams Bashforth, 43, 44, 71
- Método de Adams Moulton, 43, 44
- Método de Euler, 14, 71
- Método de Euler implícito, 22, 62
- Método de Euler mejorado, 31
- Método de Euler modificado, 31
- Método de Simpson, 22, 42
- Método del trapecio, 22
- Métodos de paso único, 23, 54, 71
- Métodos de paso múltiple, 37, 67, 71
- Métodos de Runge Kutta, 29, 71
- Métodos de Taylor, 28
- Métodos implícitos, 21
- Orden de consistencia, 25, 32, 46
- Polinomio característico, 47, 49
- Polinomio de Lagrange, 85
- Problema de Cauchy, 4, 5
- Raíz principal, 49
- Región de estabilidad absoluta, 35, 58, 65
- Runge Kutta Fehlberg, 66, 89
- Series de Taylor, 14, 38, 45
- Simulación numérica, 78, 80
- Sistema de ecuaciones diferenciales ordinarias de primer orden, 2, 4
- Tamaño de paso, 14, 50, 67, 88
- Valores propios, 8, 33
- Vectores propios, 8
- VIH/SIDA, 72, 75
- Wronskiano, 7