



UNIVERSIDAD DE LOS ANDES  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE MATEMÁTICAS  
GRUPO CIENCIAS DE LA COMPUTACIÓN

---

## Solución Numérica de Problemas de Valor de Frontera para Ecuaciones Diferenciales Ordinarias<sup>1</sup>

---

BR. LUIS JOSÉ BERBESÍ MÁRQUEZ

TRABAJO ESPECIAL DE GRADO  
PARA OPTAR AL TÍTULO DE  
LICENCIADO EN MATEMÁTICAS  
TUTOR: DR. GIOVANNI CALDERÓN<sup>2</sup>

MÉRIDA-VENEZUELA  
JULIO 2010

---

<sup>1</sup>Este trabajo fue parcialmente financiado por Consejo de Desarrollo Científico Humanístico y Tecnológico, CDCHT - ULA, bajo el proyecto C - 1687 - 10 - 02 - ED

<sup>2</sup>Grupo Ciencias de la Computación, Departamento de Matemáticas, Edificio Teórico de Ciencias, Facultad de Ciencias, Universidad de Los Andes, La Hechicera, Mérida 5101, Estado Mérida, Venezuela. e-mail: giovanni@ula.ve, Teléfono: 0274-2403404



## RESUMEN

Dado el Problema de Valor de Frontera (PVF) de segundo orden:

$$\begin{cases} y'' = f(x, y, y'), \\ y(a) = \alpha, \\ y(b) = \beta. \end{cases}$$

se quiere encontrar soluciones aproximadas mediante métodos numéricos. En este trabajo se estudian cinco métodos para estimar la solución del problema planteado. Estos métodos numéricos son: los métodos del Disparo y de Diferencias Finitas, considerados clásicos en la literatura puesto que están presentes en la mayoría de los textos de Análisis Numérico que abordan el problema propuesto; el método de Elementos Finitos y de Galerkin Discontinuo, los cuales parten de la teoría de Espacios Normados y del Análisis Funcional; y el método **Bvp4c**, un método que usa la idea de superposición y el cual se encuentra implementado en MATLAB. Se discute y compara la eficiencia y costo computacional de estos métodos en un conjunto de ejemplos lineales y no lineales.



<b>Introducción</b>	<b>III</b>
<b>1. Preliminares</b>	<b>1</b>
1.1. Orden de aproximación $\mathcal{O}(h^n)$ . . . . .	1
1.2. Teoría de Álgebra Lineal . . . . .	2
1.3. Teoría de Ecuaciones Diferenciales . . . . .	2
1.4. Teoría de Análisis Funcional . . . . .	5
1.5. Método de Newton Raphson y Runge Kutta . . . . .	8
<b>2. Métodos Numéricos para aproximar la solución de un Problema de Valor de Frontera de segundo orden</b>	<b>11</b>
2.1. Método del Disparo . . . . .	11
2.1.1. Planteamiento para el caso $y'' = p(x)y' + q(x)y + r(x)$ . . . . .	11
2.1.2. Planteamiento para el caso $y'' = f(x, y, y')$ . . . . .	14
2.2. Método de Diferencias Finitas . . . . .	17
2.2.1. Planteamiento para el caso $y'' = p(x)y' + q(x)y + r(x)$ . . . . .	17
2.2.2. Planteamiento para el caso $y'' = f(x, y, y')$ . . . . .	20
2.3. Método de Elementos Finitos . . . . .	22
2.3.1. Formulación débil de un problema modelo unidimensional . . . . .	23
2.3.2. El MEF para el problema modelo con funciones lineales a trozos . . . . .	24
2.3.3. Integración numérica . . . . .	26
2.4. Método Galerkin Discontinuo . . . . .	27
2.4.1. Formulación débil de un problema modelo unidimensional . . . . .	28
2.4.2. Derivación del sistema lineal $A\alpha = b$ mediante el uso de polinomios discontinuos a trozos . . . . .	30
2.4.3. Construcción de la matriz $A$ . . . . .	31
2.4.4. Construcción del vector $b$ . . . . .	33
2.5. Método Bvp4c . . . . .	35
<b>3. Experimentación Numérica</b>	<b>37</b>
Ejemplo 1 . . . . .	38

## Índice general

---

Ejemplo 2 . . . . .	40
Ejemplo 3 . . . . .	43
Ejemplo 4 . . . . .	45
Ejemplo 5 . . . . .	48
Ejemplo 6 . . . . .	50
Ejemplo 7 . . . . .	52
Consideraciones Finales . . . . .	54
<b>A. Pruebas de algunas afirmaciones</b>	<b>57</b>
A.1. Correspondientes al Método del Disparo . . . . .	57
A.2. Correspondientes al Método de Diferencias Finitas . . . . .	58
<b>Referencias</b>	<b>61</b>

En muchas áreas de las Ciencias Básicas e Ingeniería existen problemas donde resulta necesario encontrar la solución de una Ecuación Diferencial Ordinaria (EDO). Estas ecuaciones describen fenómenos que cambian frecuentemente. Comúnmente, una solución de interés está determinada especificando los valores de todas sus componentes en un punto  $x = a$ . Esto es un Problema de Valor Inicial. Sin embargo, en muchas ocasiones, una solución está determinada en más de un punto. Un problema de este tipo es denominado Problema de Valor de Frontera (PVF). Un PVF de interés en la actualidad son los de segundo orden, es decir, los PVF que se especifican en dos puntos:

$$\begin{cases} y'' &= f(x, y, y'), \\ y(a) &= \alpha, \\ y(b) &= \beta. \end{cases}$$

Los PVF de segundo orden suelen ser comunes en todas las ramas de las Ciencias Experimentales. Por ejemplo, en Física, las leyes de Newton y muchas otras se expresan como un problema de este tipo, en Biología aparecen en el modelado de dinámica de poblaciones, en la Química surgen en la evaluación de las concentraciones de diversos reactivos durante una reacción, etc.

A lo largo de los años se han desarrollado técnicas para encontrar la solución analítica a Ecuaciones Diferenciales Ordinarias, y por ende, la solución de los PVF. Sin embargo, resulta habitual que en la mayoría de los casos no se conozca la solución analítica de la misma, y sólo estudios cualitativos de dichas ecuaciones son presentados en la literatura. Debido a esto, en la práctica, resulta impetuoso usar métodos numéricos para dar aproximaciones numéricas de la solución (aproximaciones tan buenas como se pueda).

Hoy en día, en la literatura, existen muchos métodos numéricos que definen aproximaciones de la solución de un PVF de segundo orden. Este trabajo se concentrará en el estudio (eficiencia y costo computacional) de cinco de ellos. Estos métodos son: los métodos del Disparo y de Diferencias Finitas (ver [9, 18]), considerados clásicos dentro de la literatura, los métodos de Elementos Finitos y de Galerkin Discontinuo (ver [3, 17] y [1], respectivamente), los cuales se fundamentan en resultados del Análisis Funcional, y el método Bvp4c (ver [14]), un método que usa la idea de superposición y el mismo se encuentra implementado en MATLAB.

Las implementaciones numéricas de los cinco métodos mencionados son llevadas a cabo en MATLAB. No obstante, por considerarlo irrelevante, los códigos no son dados explícitamente en este documento. Sin embargo, en el formato electrónico existe un directorio con los códigos de las implementaciones.

En general, el trabajo queda dividido de la siguiente manera: en el capítulo 1 se desarrollan los preliminares teóricos necesarios para abordar adecuadamente el estudio de los cinco métodos. Se introducen resultados del Álgebra Lineal (ver [2]), Ecuaciones Diferenciales (ver [7]) y Análisis Funcional (ver [3, 17]). Así mismo, se introducen el método de Newton Raphson, tanto para una ecuación como para un sistema de ecuaciones no lineales, y el método de Runge Kutta, para resolver Problemas de Valor Inicial de Ecuaciones Diferenciales Ordinarias, necesarios al momento de definir los métodos numéricos. En el Capítulo 2 se expondrán cada uno de los cinco métodos que se emplearán para aproximar la solución de un PVF de segundo orden. Se presenta la

construcción de cada método, y en mucho de ellos, el análisis del error. En el último capítulo se hará una breve discusión, más que comparación, sobre la eficiencia y propiedades de cada uno de estos métodos, esto en vista de que no se presentan las condiciones adecuadas para realizar una comparación detallada y óptima entre ellos. Por ejemplo, el método **Bvp4c** realiza proceso adaptativo con el fin de controlar el error de la aproximación, mientras que los restantes cuatro métodos no realizan dicho proceso adaptativo. En el apéndice A se realizan las demostraciones de varios teoremas presentes en el capítulo 2. Por último, la bibliografía presenta las referencias literarias usadas.

En este capítulo se expondrán los preliminares matemáticos necesarios para abordar adecuadamente el estudio de los métodos numéricos usados para aproximar la solución de problema de valor de frontera de segundo orden. Así pues, el objetivo principal de este capítulo es el desarrollo de las herramientas matemáticas comúnmente usadas en el planteamiento y análisis de estos métodos numéricos. Al mismo tiempo, se introducen el método de Newton Raphson, tanto para una ecuación como para un sistema de ecuaciones no lineales, y el método de Runge Kutta para resolver Problemas de Valor Inicial de Ecuaciones Diferenciales Ordinarias, estudiados en un curso básico de Análisis Numérico, necesarios al momento de definir los métodos usados en este trabajo.

Es importante resaltar que los resultados que se desarrollan en el presente capítulo no serán tratados de forma detallada por completo, pues el objetivo es solamente dejar asentado un material de repaso, cuyo conocimiento es importante para entender el resto de este trabajo. Muchos de estos temas se tratan de modo más amplio en algunos textos de Análisis Numérico (ver [6, 12, 18]), Álgebra Lineal (ver [2]), Ecuaciones Diferenciales (ver [7]) y Análisis Funcional (ver [3]).

## 1.1. Orden de aproximación $\mathcal{O}(h^n)$

**Definición 1.1** Se dice que  $f(h)$  es de orden  $g(h)$  cuando  $h \rightarrow 0$  y  $h \neq 0$ , se denotará por  $f(h) = \mathcal{O}(g(h))$ , si existen números reales  $M > 0$  y  $k > 0$  tales que,

$$|f(h)| \leq M|g(h)|, \text{ siempre que } |h| < k.$$

**Ejemplo 1.1** Consideremos las funciones  $f(x) = x^3 + 2x^2$  y  $g(x) = x^2$ . Puesto que,  $x^3 < x^2$  para  $|x| \leq 1$ , se obtiene,  $|x^3 + 2x^2| < 3|x^2|$ . Por lo tanto,  $f(x) = \mathcal{O}(g(x))$ .

**Definición 1.2** Sean  $p$  y  $f$  funciones, se dice que  $p(h)$  aproxima a  $f(h)$  con un orden de aproximación  $\mathcal{O}(h^n)$ , lo que se denota por  $f(h) = p(h) + \mathcal{O}(h^n)$ , si existe un número real  $M > 0$  y un número natural  $n$  tales que,

$$\frac{|f(h) - p(h)|}{|h^n|} \leq M, \text{ para } h \text{ suficientemente pequeño.}$$

Consideremos el caso en que  $p(x)$  es la  $n$ -ésima aproximación por polinomios de Taylor a  $f(x)$  alrededor de  $x_0$ , es decir

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{f^{n+1}(c)}{(n+1)!} (x - x_0)^{n+1},$$

para algún  $c$  entre  $x$  y  $x_0$ . Cuando  $x - x_0 \rightarrow 0$ , entonces, por la Definición 1.1, se tiene que

$$\mathcal{O}((x - x_0)^{n+1}) = \frac{f^{n+1}(c)}{(n+1)!} (x - x_0)^{n+1},$$

así,  $f(x) = p(x) + \mathcal{O}((x - x_0)^{n+1})$ , es decir,  $p(x)$ , con  $p(x)$  el primer término del lado derecho de la igualdad, se aproxima a  $f(x)$  con un orden de aproximación  $\mathcal{O}((x - x_0)^{n+1})$ . Luego, el Teorema de Taylor se puede enunciar de la siguiente forma:

**Teorema 1.1 (Teorema de Taylor)** Si  $f \in C^{n+1}[a, b]$  y  $x_0 \in [a, b]$ , entonces para cada  $x \in [a, b]$ ,

$$f(x_0 + h) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} h^k + \mathcal{O}(h^{n+1}), \text{ donde } h = x - x_0.$$

## 1.2. Teoría de Álgebra Lineal

**Definición 1.3** Se dice que una matriz  $A$  de  $n \times n$  es **invertible**, si existe una matriz  $A^{-1}$  de  $n \times n$ , con  $AA^{-1} = A^{-1}A = I$ , donde  $I$  es la matriz identidad. La matriz  $A^{-1}$  se llama inversa de  $A$ . Una matriz que no tiene inversa se le da el nombre de **no invertible o singular**.

**Definición 1.4 (Matriz tridiagonal)** Una matriz  $A = (a_{ij})$  de  $n \times n$  se dice **tridiagonal** si  $a_{ij} = 0$ , siempre que  $|i - j| > 1$ . Esto es,

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & a_{n-1,1} \\ 0 & \cdots & \cdots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}.$$

Ahora bien, en muchas ocasiones estamos interesados en saber qué condiciones debe cumplir una matriz tridiagonal para que sea invertible. Dichas condiciones nos las da el siguiente teorema.

**Teorema 1.2** Sea  $A$  una matriz tridiagonal, con entradas  $[A]_{ij} = a_{ij}$ . Para cada  $i = 2, 3, \dots, n-1$ , supongamos que se cumple  $a_{i,i-1}a_{i,i+1} \neq 0$ . Si  $|a_{11}| > |a_{12}|$ ,  $|a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|$ , para  $i = 2, 3, \dots, n-1$  y  $|a_{nn}| > |a_{n,n-1}|$ , entonces  $A$  es invertible.

**Demostración:** Ver referencia [9], Capítulo 2, Sección 3, pág. 56. ■

## 1.3. Teoría de Ecuaciones Diferenciales

Es bien conocido en el quehacer científico, bien sea en las ciencias puras o la ingeniería, la utilidad de las ecuaciones diferenciales. El estudio de las ecuaciones diferenciales han influenciado el desarrollo de varias áreas de las matemáticas, como es el caso de la Topología, el Análisis Real, y nuestra área de estudio: el Análisis Numérico.

A lo largo de los años se han desarrollado técnicas para encontrar soluciones analíticas a ecuaciones diferenciales ordinarias que modelan un sin fin de problemas que surgen de cualquiera de las áreas antes mencionadas. Sin embargo, resulta habitual que en la mayoría de los casos no se conozca la solución analítica de la misma, y sólo estudios cualitativos de dichas ecuaciones son presentados en la literatura. Debido a esto, en la práctica, resulta imperioso definir métodos numéricos para dar aproximaciones numéricas de la solución (aproximaciones tan buenas como se quiera). Vale señalar que muchos de estos estudios cualitativos se basan en aproximaciones numéricas para iniciar la investigación del problema. Antes de continuar, es importante recordar algunas definiciones y conceptos básicos de ecuaciones diferenciales.

Una **ecuación diferencial**, el cual denotaremos mediante ED, es una ecuación que involucra una o más variables dependientes, con respecto a una o más variables independientes. Por ejemplo, en la ecuación:

$$\frac{d^2x}{dt^2} + 16x = 0,$$

se tiene que el símbolo  $x$  representa una variable dependiente mientras que la variable independiente es  $t$ .

Si una ecuación contiene sólo las derivadas ordinarias de una o más variables dependientes con respecto a una sola variable independiente, entonces se dice que es una **ecuación diferencial ordinaria** (EDO).

Las ecuaciones:

- $\frac{dy}{dx} + 5y = e^x,$
- $\frac{d^2y}{dx^2} - \frac{dy}{dx} + 6y = 0,$
- $\frac{dx}{dt} + \frac{dy}{dt} = 2x + y,$

son ejemplos de ecuaciones diferenciales ordinarias. Es importante resaltar que en este trabajo las derivadas ordinarias se escribirán con la **notación de Leibniz**:  $\frac{dy}{dx}, \frac{d^2y}{dx^2}, \frac{d^3y}{dx^3}, \dots$ , o bien la notación del símbolo de prima:  $y', y'', y'''$ . En realidad, la notación de prima se emplea para denotar sólo las tres primeras derivadas, la cuarta derivada se escribirá  $y^{(4)}$  en lugar de  $y''''$ ; más general, la  $n$ -ésima derivada de  $y$  se escribirá  $\frac{d^n y}{dx^n}$  o  $y^{(n)}$ . Por otro lado, el orden de la más alta derivada de una ED se llama **orden** de la ecuación. Por ejemplo:

$$\frac{d^2y}{dx^2} + 5\left(\frac{dy}{dx}\right)^3 - 4y = e^x,$$

es una EDO de segundo orden.

En símbolos, una EDO de  $n$ -ésimo orden de una variable dependiente, se puede expresar mediante la forma general

$$F(x, y, y', \dots, y^{(n)}) = 0, \tag{1.1}$$

donde  $F$  es una función de valores reales de  $n + 2$  variables:  $x, y, y', \dots, y^{(n)}$ . Por razones prácticas y teóricas supondremos de aquí en adelante que es posible despejar en una EDO en forma única la derivada superior  $y^{(n)}$  de una ED que esté en la forma (1.1), en términos de  $n + 1$  variables restantes. La ecuación diferencial

$$\frac{d^n y}{dx^n} = f(x, y, y', \dots, y^{(n-1)}),$$

donde  $f$  es una función continua de valores reales, se denomina **forma normal** de (1.1). Así, cuando sea conveniente, se usan las formas normales

$$\frac{dy}{dx} = f(x, y) \quad \text{y} \quad \frac{d^2y}{dx^2} = f(x, y, y')$$

para representar las EDO de primer y segundo orden.

Un concepto importante dentro del marco de las EDO es el concepto de linealidad. Se dice que una EDO de orden  $n$  es **lineal** si  $F$  es lineal en  $y, y', \dots, y^{(n)}$ . Esto significa que una EDO de orden  $n$  es lineal cuando (1.1) se puede reescribir como

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_1(x)y' + a_0(x)y = g(x).$$

Las ecuaciones diferenciales lineales se caracterizan por dos propiedades:

1. la variable independiente  $y$  junto con todas sus derivadas son de primer grado, es decir, la potencia de cada término en  $y$  es 1,
2. los coeficientes  $a_0, a_1, \dots, a_n$ , de  $y, y', \dots, y^{(n)}$ , dependen sólo de la variable independiente  $x$ .

Una ecuación que no es lineal se dice **no lineal**.

Con frecuencia nos interesa problemas en los que se busca una solución  $y(x)$  de una ED de modo que  $y(x)$  satisfaga condiciones adicionales prescritas, es decir, condiciones impuestas en la  $y(x)$  desconocida o sus derivadas. Dos conceptos importantes dentro del marco de las EDO son los conceptos de *problema de valor inicial* y *problema de valor de frontera*.

Para una ED, un Problema de Valor Inicial (el cual se denotará como PVI) de  $n$ -ésimo orden es resolver

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

sujeta a:

$$y(x_0) = y_0, y'(x_0) = y_1, \dots, y^{(n-1)}(x_0) = y_{n-1}.$$

Otro tipo de problemas consiste en resolver una ED de orden dos o mayor en el que la variable dependiente  $y$  o sus derivadas se especifican en diferentes puntos. Un problema que consiste en resolver

$$y'' = f(x, y, y')$$

sujeta a:

$$y(a) = y_0, y(b) = y_1,$$

se llama Problema de Valor de Frontera (el cual se denotará mediante PVF). Los valores preescritos  $y(a) = y_0$  y  $y(b) = y_1$  se llaman condiciones de frontera.

Ahora bien, estamos interesados en saber bajo qué condiciones un PVF de segundo orden tiene solución. El siguiente teorema establece condiciones generales que garantizan la existencia y unicidad de la solución a dicho problema de segundo orden.

**Teorema 1.3** *Supongamos que la función  $f$  en el PVF*

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

*es continua en el conjunto*

$$D = \{(x, y, y') / a \leq x \leq b, -\infty < y < \infty, -\infty < y' < \infty\},$$

*y que  $f_y$  y  $f_{y'}$  también son continuas en  $D$ . Si*

1.  $f_y(x, y, y') > 0$ , para toda  $(x, y, y') \in D$ , y
2. existe una constante  $M$ , con  $|f_{y'}(x, y, y')| \leq M$ , para todo  $(x, y, y') \in D$ ,

*entonces el PVF tiene solución única.*

**Demostración:** Ver referencia [13]. ■

Como consecuencia del teorema anterior, tenemos el siguiente resultado.

**Corolario 1.1** *Si el problema lineal con valor en la frontera*

$$y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

*satisface:*

1.  $p(x)$ ,  $q(x)$  y  $r(x)$  son continuas en  $[a, b]$ , y
2.  $q(x) > 0$  en  $[a, b]$ ,

*entonces el problema tiene solución única.*

## 1.4. Teoría de Análisis Funcional

Para considerar los aspectos referentes al Método de Elementos Finitos y al Método Galerkin Discontinuo, es importante conocer algunos resultados del Análisis Funcional.

### Los espacios normados

En un espacio lineal podemos asignar a cada elemento  $x$  la noción de longitud por medio del número real  $\|x\|$  que es llamado norma de  $x$ . En particular, la *norma* es una función de valor real definida en un espacio lineal  $\mathbb{E}$  que satisface las siguientes condiciones o axiomas:

1.  $\|x\| \geq 0$ , con la particularidad de que  $\|x\| = 0$  sólo si  $x = 0$ .
2.  $\|\alpha x\| = |\alpha| \|x\|$ , donde  $\alpha$  es un número real.
3.  $\|x + y\| \leq \|x\| + \|y\|$ ,  $x, y \in \mathbb{E}$ .

**Ejemplo 1.2** Si  $\mathbb{E} = \mathbb{R}^2$ , la selección usual para la longitud o norma  $\|\cdot\|_2$  de un vector  $x = (x_1, x_2)$  en el plano real es

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2},$$

que es llamada norma 2 o norma euclidiana. Otro ejemplo de norma es la norma 1 que se define como

$$\|x\|_1 = |x_1| + |x_2|,$$

o la norma infinita que se define como

$$\|x\|_\infty = \text{máx}\{|x_1|, |x_2|\}.$$

**Ejemplo 1.3** Sea  $\mathbb{E} = C[a, b]$ . La función de valor real definida por

$$\|f\| = \text{máx}_{a \leq t \leq b} |f(t)|$$

es una norma en  $\mathbb{E}$ , conocida como norma uniforme o norma Chebyshev de  $f$  en el espacio  $C[a, b]$ . Una segunda norma en  $C[a, b]$ , que es también muy importante, es la llamada norma  $L_2$ , que se denota

$$\|\cdot\|_{L_2} = \left[ \int_a^b [f(t)]^2 dt \right]^{1/2}.$$

Sea  $\mathbb{E}$  un espacio normado. La sucesión  $\{u_n\}_{n=1}^\infty \subset \mathbb{E}$  se dice *acotada* en  $\mathbb{E}$  si existe un  $C > 0$  tal que  $\|u_n\| < C$ , para todo  $n$ . La sucesión se dice *convergente* en  $\mathbb{E}$  si existe un elemento  $u \in \mathbb{E}$  tal que para todo  $0 < \varepsilon \in \mathbb{R}$ , existe un  $N_0 \in \mathbb{N}$  tal que  $\|u - u_n\| < \varepsilon$ , para todo  $n > N_0$ . El elemento  $u$  es el límite de la sucesión  $\{u_n\}_{n=1}^\infty$ . Usualmente se escribirá de forma indistinta

$$\lim_{n \rightarrow \infty} u_n = u \quad \text{o} \quad \|u - u_n\| \rightarrow 0, \text{ para } n \rightarrow \infty.$$

La sucesión  $\{u_n\}_{n=1}^\infty \subset \mathbb{E}$  será una *sucesión de Cauchy* si para todo  $\varepsilon > 0$ , existe  $N_0 \in \mathbb{N}$  tal que

$$\|u_n - u_m\| \leq \varepsilon, \text{ para todo } n, m \geq N_0.$$

Un espacio lineal normado  $\mathbb{E}$  se llama *completo*, cuando toda sucesión de Cauchy de este espacio converge hacia un cierto elemento  $u \in \mathbb{E}$ . Los espacios lineales normados completos se llaman *espacios de Banach*. Todo espacio lineal normado de dimensión finita es completo.

## El espacio euclidiano y el espacio unitario

Una forma conocida de introducir una norma en un espacio lineal consiste en definir en éste el producto escalar o producto interior. Se llama producto escalar en un espacio lineal real (o complejo)  $\mathbb{E}$  a una función  $(\cdot, \cdot) : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$  (o  $\mathbb{C}$ ) con las siguientes propiedades, para todo par de elementos  $x, y \in \mathbb{E}$ :

1.  $(x, y) = \overline{(y, x)}$ ,
2.  $(x + y, z) = (x, z) + (y, z)$ ,
3.  $(\lambda x, y) = \lambda(x, y)$ ,  $\lambda \in \mathbb{R}$  (o  $\mathbb{C}$ ),
4.  $(x, x) \geq 0$ , y  $(x, x) = 0$  sólo si  $x = 0$ .

En el caso de un espacio lineal real el primer axioma se reduce a la simetría  $(x, y) = (y, x)$ .

**Ejemplo 1.4** Sean  $x(t)$  e  $y(t)$  dos funciones de  $C^2[a, b]$ . Se pueden definir varios productos escalares:

- $(x, y) = \int_a^b x(t)y(t)dt$ ,
- $(x, y) = \int_a^b [x(t)y(t) + x'(t)y'(t)] dt$ ,
- $(x, y) = \int_a^b [x(t)y(t) + x'(t)y'(t) + x''(t)y''(t)] dt$ ,

donde las integrales son integrales de Riemann.

**Lema 1.1** Sea  $\mathbb{E}$  un espacio con producto interno. Entonces la función

$$\|u\| = \sqrt{(u, u)}, \quad u \in \mathbb{E}$$

es una norma en  $\mathbb{E}$ .

Los espacios dotados de un producto interno son llamados *espacio producto interno*. Un espacio lineal real normado  $\mathbb{E}$ , en el cual la norma está generada por un producto escalar, es decir,

$$\|u\| = \sqrt{(u, u)},$$

se llama *espacio euclidiano*; en el caso complejo recibe el nombre de *espacio unitario*. Un espacio unitario (euclidiano) completo es llamado *espacio de Hilbert*  $\mathbf{H}$ .

Ahora se introducen algunos espacios de Hilbert que resultan naturales para la formulación variacional de los problemas de valor de frontera que serán considerados. Si  $\Omega$  es un dominio acotado en  $\mathbb{R}^2$ , se define el espacio de funciones de cuadrado integrables sobre  $\Omega$ :

$$L^2(\Omega) = \{v : \Omega \rightarrow \mathbb{R} \text{ tal que } \int_{\Omega} v^2 d\Omega < \infty\}.$$

El espacio  $L^2(\Omega)$  es un espacio de Hilbert con el producto escalar

$$(u, v)_{L^2(\Omega)} = \int_{\Omega} vud\Omega,$$

y la correspondiente norma

$$\|v\|_{L^2(\Omega)} = \left[ \int_{\Omega} [v(x)]^2 dx \right]^{1/2} = (v, v)_{L^2(\Omega)}^{1/2}.$$

Se introduce además el espacio

$$\mathbf{H}^1(\Omega) = \left\{ v \in L^2(\Omega) : \frac{\partial v}{\partial x_i} \in L^2(\Omega), i = 1, 2 \right\},$$

dotado del producto escalar y la correspondiente norma

$$(v, w)_{\mathbf{H}^1(\Omega)} = \int_{\Omega} (vw + \nabla v \nabla w) d\Omega, \quad \|\cdot\|_{\mathbf{H}^1(\Omega)} = (\cdot, \cdot)_{\mathbf{H}^1(\Omega)}^{1/2}.$$

El espacio  $\mathbf{H}^1(\Omega)$  consiste de todas las funciones  $v$  definidas en  $\Omega$  junto con sus primeras derivadas que son de cuadrado integrable, es decir, pertenecen a  $L^2(\Omega)$ . Además, se define el espacio

$$\mathbf{H}_0^1 = \{v \in \mathbf{H}^1(\Omega) : v = 0 \text{ en } \partial v\},$$

que se dota del mismo producto interno y norma que el espacio  $\mathbf{H}^1(\Omega)$ . Para evitar sobrecargar la notación, siempre y cuando no se lleve a confusión, se utilizará la notación  $(\cdot, \cdot)$  y  $\|\cdot\|$  en lugar de  $(\cdot, \cdot)_{L^2(\Omega)}$  y  $\|\cdot\|_{L^2(\Omega)}$ . De forma análoga,  $(\cdot, \cdot)_1$  y  $\|\cdot\|_1$  denotarán  $(\cdot, \cdot)_{\mathbf{H}^1(\Omega)}$  y  $\|\cdot\|_{\mathbf{H}^1(\Omega)}$ , respectivamente.

Los espacios  $L^2(\Omega)$  y  $\mathbf{H}^1(\Omega)$  pertenecen a familias de espacios más generales, conocidas como espacios  $L^p(\Omega)$ , con  $p \geq 1$ , y espacios de *Sobolev*, respectivamente.

### Funcionales lineales

Un *funcional lineal* es un caso particular de un operador lineal, es decir, es un operador lineal que transforma el espacio dado  $\mathbb{E}$  en valores de  $\mathbb{R}$  o  $\mathbb{C}$ . En general, si  $\mathbb{E}$  es un espacio lineal normado sobre un cuerpo  $\mathbf{K}$ , entonces un funcional lineal  $f$  es una aplicación lineal de  $\mathbb{E} \rightarrow \mathbf{K}$  tal que

1.  $f(u + w) = f(u) + f(w), \forall u, w \in \mathbb{E}$ ,
2.  $f(\alpha u) = \alpha f(u), \forall u \in \mathbb{E}, \alpha \in \mathbf{K}$ .

El espacio de todos los funcionales lineales de un espacio normado  $\mathbb{E}$  sobre su cuerpo,  $\mathcal{L}(\mathbb{E}, \mathbf{K})$ , se le llama espacio dual de  $\mathbb{E}$ , y suele denotarse por  $\mathbb{E}'$ .

Consideremos algunos ejemplos de funcionales.

**Ejemplo 1.5** Sea  $\mathbb{E} = \mathbb{R}^n$ . El operador  $f : \mathbb{E} \rightarrow \mathbb{R}$  definido como el promedio de las componentes del vector

$$f(v) = \frac{1}{n} \sum_{i=1}^n v_i, \quad \text{para todo } v \in \mathbb{E},$$

es un funcional lineal sobre  $\mathbb{E}$ , es decir,  $f \in \mathbb{E}'$ .

**Ejemplo 1.6** El operador integral  $A : C[a, b] \rightarrow \mathbb{R}$ , definido por

$$A(f) = \int_a^b f(x) dx, \quad \text{para todo } f \in C[a, b],$$

es un funcional lineal sobre  $C[a, b]$ .

**Teorema 1.4 (Teorema de Representación de Riesz)** Sea  $\mathbf{H}$  un espacio de Hilbert y sea  $F$  un funcional lineal continuo de  $\mathbf{H}$ . Entonces existe un único elemento  $u \in \mathbf{H}$  tal que

$$F(v) = (u, v), \quad \forall v \in \mathbf{H}.$$

Además  $\|F\| = \|u\|$

## Formas bilineales

Otro tipo especial de operador que pueden ocurrir muy frecuentemente en el estudio de problemas de valor de frontera es uno que mapea un par de elementos a los números reales, y que es lineal en cada uno de estos. Un operador  $B : \mathbf{U} \times \mathbf{V} \rightarrow \mathbb{R}$  ( $\mathbf{U}, \mathbf{V}$  espacios lineales) se dice que es una *forma bilineal* si:

1.  $B(\alpha u + \beta w, v) = \alpha B(u, v) + \beta B(w, v), \quad u, w \in \mathbf{U}, v \in \mathbf{V}$
2.  $B(u, \alpha v + \beta w) = \alpha B(u, v) + \beta B(u, w), \quad u \in \mathbf{U}, v, w \in \mathbf{V}$

**Formas bilineales continuas.** Consideremos una forma bilineal  $B : \mathbf{U} \times \mathbf{V} \rightarrow \mathbb{R}$ , donde  $\mathbf{U}$  y  $\mathbf{V}$  son espacios lineales normados. Si existe una constante  $k > 0$  tal que

$$|B(u, v)| \leq k \|u\| \|v\| \quad \text{para todo } u \in \mathbf{U}, v \in \mathbf{V},$$

entonces  $B$  es llamada una *forma bilineal continua*.

**Formas bilineales  $\mathbf{H}$  - elípticas.** Dada una forma bilineal  $B : \mathbf{H} \times \mathbf{H} \rightarrow \mathbb{R}$ , donde  $\mathbf{H}$  es un espacio con producto interno, se dice que  $B$  es  $\mathbf{H}$  - elíptica si existe una constante  $\alpha > 0$  tal que

$$B(v, v) \geq \alpha \|v\|^2, \quad \forall v \in \mathbf{H}.$$

Así una forma  $\mathbf{H}$  - elíptica es acotada inferiormente.

## El Teorema de Lax - Milgram

**Teorema 1.5 (El Teorema de Lax - Milgram)** Sea  $\mathbf{H}$  un espacio de Hilbert y sea  $B : \mathbf{H} \times \mathbf{H} \rightarrow \mathbb{R}$  una forma bilineal,  $\mathbf{H}$  - elíptica, continua, definida en  $\mathbf{H}$ . Entonces, dada cualquier funcional lineal  $F$  definida en  $\mathbf{H}$ , existe un único elemento  $u \in \mathbf{H}$  tal que

$$B(u, v) = F(v), \quad \forall v \in \mathbf{H},$$

además,

$$\|u\|_1 \leq \alpha^{-1} \|F\|_{\mathbf{H}'}$$

## 1.5. Método de Newton Raphson y Runge Kutta

En esta sección se introducen dos métodos que serán necesarios para el estudio de los métodos que se emplearán para aproximar la solución de un PVF de segundo orden. Estos métodos son el de Newton Raphson, tanto para una ecuación como para un sistema de ecuaciones no lineales, y el de Runge Kutta, para resolver PVI de ecuaciones diferenciales ordinarias.

**Método de Newton Raphson.** Sea  $f$  una función dos veces diferenciable. Supongamos que  $\alpha$  es una raíz simple de  $f(x) = 0$  y  $x_0$  es una aproximación inicial a  $\alpha$ . El método de Newton Raphson consiste en conseguir la aproximación usando la recta tangente a la curva  $f$ , que pasa por el punto  $(x_0, f(x_0))$ . Esta recta intersecta al eje  $OX$  en el punto  $(x_1, 0)$ , donde  $x_1$  será la aproximación a la raíz de  $f$ , luego

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Para continuar, se renombra  $x_0 = x_1$  y se repite el proceso tanto como desee.

**Teorema 1.6 (Convergencia del Método de Newton Rapson)** Sean  $f \in C^2[a, b]$  y  $\alpha \in [a, b]$  tal que  $f(\alpha) = 0$ . Si  $f'(\alpha) \neq 0$ , entonces existe  $\delta > 0$  tal que la sucesión  $\{x_k\}_{k=0}^{\infty}$  definida por el proceso iterativo,

$$x_k = g(x_{k-1}) = x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}, \text{ para } k = 1, 2, 3, \dots,$$

converge a  $\alpha$  cualesquiera que sea la aproximación inicial  $x_0 \in [\alpha - \delta, \alpha + \delta]$ .

**Demostración:** Ver la referencia [18], Capítulo 2, Sección 2.1, pág. 69. ■

**Método de Newton para sistemas de ecuaciones no lineales.** En muchas ocasiones aparece la necesidad de resolver sistemas de ecuaciones algebraicas no lineales de la forma

$$\left. \begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0. \end{aligned} \right\}$$

El método de Newton para sistema de ecuaciones no lineales es una extensión directa del método del mismo nombre para buscar ceros de funciones de una variable. La idea es realizar el desarrollo de Taylor en  $n$  variables en el entorno de una raíz,

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^{(0)}) + \left( \frac{\partial f_i}{\partial x_j} \right)_{\mathbf{x}^{(0)}} (\mathbf{x} - \mathbf{x}^{(0)}) + \mathcal{O}(\|\mathbf{x} - \mathbf{x}^{(0)}\|^2).$$

Si  $\mathbf{x}$  es una raíz aproximada, definiendo la matriz jacobiana como  $J(\mathbf{x}^{(0)}) = \left( \frac{\partial f_i}{\partial x_j} \right)_{\mathbf{x}^{(0)}}$  encontramos que

$$\mathbf{x} \approx \mathbf{x}^{(0)} - J^{-1}(\mathbf{x}^{(0)})\mathbf{f}(\mathbf{x}^{(0)}).$$

En la práctica, esta fórmula nos permite definir un proceso iterativo,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - J^{-1}(\mathbf{x}^{(k)})\mathbf{f}(\mathbf{x}^{(k)}),$$

que es el método de Newton en varias variables. A la hora de implementar el método, es costoso tener que calcular la inversa de  $J$  y por eso, normalmente, lo que se hace es resolver el sistema de ecuaciones lineales

$$J(\mathbf{x}^{(k)})\left(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\right) = \mathbf{f}(\mathbf{x}^{(k)}),$$

del cual se puede despejar  $\mathbf{x}^{(k+1)}$  una vez conocida la solución.

En [18], Capítulo 10, Sección 10.2, se hace el estudio de la convergencia y la implementación del método.

**Método de Runge Kutta.** Es un método de resolución numérica de ecuaciones diferenciales ordinarias, el cual fue inicialmente desarrollado alrededor del año 1900 por los matemáticos *C. Runge*<sup>1</sup> y *M. W. Kutta*<sup>2</sup>. El método de *Runge Kutta* no es sólo un método sino una importante familia de métodos iterativos tanto implícitos como explícitos. Un método clásico es el método de RK explícito de cuarto orden. Para introducirlo definamos el PVI:

$$y' = f(t, y), \quad y(t_0) = y_0.$$

---

<sup>1</sup>**C. Runge** nació el 30 de agosto de 1856 y murió el 03 de enero de 1927. Fue un matemático, físico y espectroscopista alemán. Su primer nombre suele escribirse Carl.

<sup>2</sup>**Martin Wilhelm Kutta** fue un físico y matemático alemán, nació el 03 de noviembre de 1867 en Pitschen, Alta Silecia (en la actualidad pertenece a Polonia) y murió el 25 de diciembre de 1944 en Fürstenfeldbruck, Alemania.

Entonces el método de RK de cuarto orden para este problema está dado por la siguiente ecuación:

$$y_{n+1} = y_n + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4),$$

donde:

$$k_1 = f(t_n, y_n),$$

$$k_2 = f\left(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$

$$k_3 = f\left(t_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right),$$

$$k_4 = f(t_n + h, y_n + k_3).$$

Por ser de cuarto orden, se tiene que el error de paso es de orden  $\mathcal{O}(h^5)$ , mientras que el error total acumulado tiene orden  $\mathcal{O}(h^4)$ . Para mayores detalles del método, ver [6], Capítulo 8, Sección 8.3, pág. 514.

## Métodos Numéricos para aproximar la solución de un Problema de Valor de Frontera de segundo orden

Las Ecuaciones Diferenciales Ordinarias describen fenómenos que cambian frecuentemente. Estas surgen en modelos totalmente matemáticos, físicos o químicos. Un sistema de ecuaciones diferenciales ordinarias puede tener muchas soluciones. Comúnmente, una solución de interés está determinada especificando los valores de todas sus componentes en un punto  $x = a$ . Esto es un Problema de Valor Inicial. Sin embargo, en muchas aplicaciones, una solución está determinada en más de un punto. Un problema de este tipo es denominado Problema de Valor de Frontera, el cual es nuestro problema a resolver, específicamente en dos puntos:

$$\begin{cases} y'' = f(x, y, y'), \\ y(a) = \alpha, \\ y(b) = \beta. \end{cases}$$

Es importante resaltar que si el PVF anterior se expresa mediante

$$y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

entonces se dice que el PVF es del tipo lineal, en caso contrario, se dice que el PVF es del tipo no lineal. Problemas de este tipo suelen ser comunes en muchas áreas de la vida cotidiana, como, por ejemplo, en las Ciencias básicas o en la Ingeniería. Sin embargo, su principal dificultad radica en que la solución a dichos problemas no son fáciles de obtener analíticamente. Debido a esto, resulta importante definir métodos numéricos para dar estimaciones numéricas de la solución. En este capítulo se expondrán cinco métodos numéricos que nos ayudarán a encontrar estas estimaciones: el Método del Disparo y el Método de Diferencias Finitas, los cuales son presentados en las referencias [9, 18], el Método de Elementos Finitos, planteado en [3, 17], el Método Galerkin Discontinuo, planteado en la referencia [1], y el Método Bvp4c, el cual se presenta en la referencia [14]. Es importante resaltar que, cuando el PVF es del tipo lineal, usaremos los cinco métodos mencionados, mientras que si el PVF es del tipo no lineal, entonces se emplearán sólo los métodos de Disparo, Diferencias Finitas y Bvp4c, en vista de que los métodos de Elementos Finitos y Galerkin Discontinuo ameritan un estudio mucho más amplio que para el caso lineal, y esto conlleva a que la implementación de los códigos requieran más tiempo.

### 2.1. Método del Disparo

Uno de los métodos numéricos, considerado clásico en la literatura, para aproximar la solución de un Problema de Valor de Frontera de segundo orden, es el método del Disparo. Este método, el cual se plantea en las referencias [9, 18], se reduce a resolver varios PVI para encontrar la solución al problema planteado.

#### 2.1.1. Planteamiento para el caso $y'' = p(x)y' + q(x)y + r(x)$

Para aproximar la solución única del problema lineal

$$y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta, \quad (2.1)$$

garantizada por el cumplimiento de las hipótesis del Corolario 1.1, primero se consideran los PVI:

$$u'' = p(x)u' + q(x)u + r(x), \quad a \leq x \leq b, \quad u(a) = \alpha, \quad u'(a) = 0 \quad (2.2)$$

y

$$v'' = p(x)v' + q(x)v, \quad a \leq x \leq b, \quad v(a) = 0, \quad v'(a) = 1. \quad (2.3)$$

Según las hipótesis del Corolario 1.1, tanto (2.2) como (2.3) tienen solución única. La solución de (2.1) viene dada por

$$y(x) = u(x) + \left( \frac{\beta - u(b)}{v(b)} \right) v(x). \quad (2.4)$$

A partir de ésta,

$$y'(x) = u'(x) + \left( \frac{\beta - u(b)}{v(b)} \right) v'(x), \quad y''(x) = u''(x) + \left( \frac{\beta - u(b)}{v(b)} \right) v''(x).$$

Así,

$$\begin{aligned} y''(x) &= p(x)u'(x) + q(x)u(x) + r(x) + \left( \frac{\beta - u(b)}{v(b)} \right) (p(x)v'(x) + q(x)v(x)) \\ &= p(x) \left( u'(x) + \left( \frac{\beta - u(b)}{v(b)} \right) v'(x) \right) + q(x) \left( u(x) + \left( \frac{\beta - u(b)}{v(b)} \right) v(x) \right) + r(x) \\ &= p(x)y'(x) + q(x)y(x) + r(x). \end{aligned}$$

Como se puede apreciar, la solución (2.4) satisface las condiciones de frontera. En efecto,

$$\begin{aligned} y(a) &= u(a) + \left( \frac{\beta - u(b)}{v(b)} \right) v(a) \\ &= \alpha + \left( \frac{\beta - u(b)}{v(b)} \right) 0 \\ &= \alpha, \end{aligned}$$

y

$$\begin{aligned} y(b) &= u(b) + \left( \frac{\beta - u(b)}{v(b)} \right) v(b) \\ &= u(b) + \beta - u(b) \\ &= \beta. \end{aligned}$$

Por tanto,  $y(x)$  es la solución única al problema de valor de frontera (sujeta a que  $v(b) \neq 0$ ).

El método del Disparo para las ecuaciones lineales se basa en la sustitución del PVF por dos PVI. En el estudio de PVI para ecuaciones diferenciales ordinarias, se describen muchos métodos con los cuales se pueden aproximar las soluciones  $u(x)$  y  $v(x)$  (como ejemplo: el método de Euler, Trapecio, Punto Medio, Runge Kutta), y una vez que se cuenta con estas aproximaciones, la solución del PVF se aproxima por medio de la ecuación (2.4). Para efecto de nuestro estudio, el método que se usará para aproximar (2.2) y (2.3) será el método de Runge Kutta.

Desde el punto de vista gráfico, el método tiene el aspecto que se observa en la Figura 2.1.

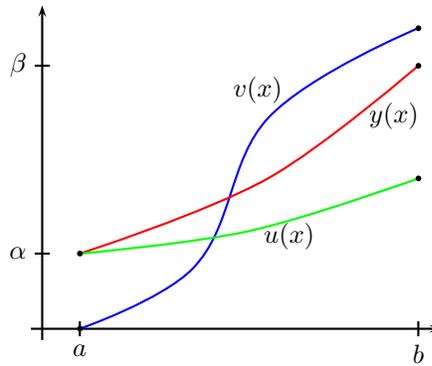


Figura 2.1: Método del Disparo para el caso lineal.

### Algoritmo del Método del Disparo para el caso lineal.

**Entrada:** extremos  $a, b$ ; condiciones de frontera  $\alpha, \beta$ ; número de subintervalos  $N$ .

**Salida:** aproximaciones  $w_{1,i}$  a  $y(x_i)$ ;  $w_{2,i}$  a  $y'(x_i)$ , para cada  $i = 0, \dots, N$ .

1. Tomar  $h = (b - a)/N$ ;  $u_{1,0} = \alpha$ ;  $u_{2,0} = 0$ ;  $v_{1,0} = 0$ ;  $v_{2,0} = 1$ .
2. Para  $i = 1, \dots, N - 1$ , hacer los pasos 3 y 4.
  3. Tomar  $x = a + ih$ .
  4. Aplicar Runge Kutta para obtener  $u_{1,i+1}, u_{2,i+1}, v_{1,i+1}, v_{2,i+1}$ .
5. Tomar  $w_{1,0} = \alpha$ ;  $w_{2,0} = (\beta - u_{1,N})/v_{1,N}$ . Mostrar  $(a, w_{1,0}, w_{2,0})$ .
6. Para  $i = 1, \dots, N$ , tomar

$$\begin{aligned} w_{1,i} &= u_{1,i} + w_{2,0}v_{1,i}, \\ w_{2,i} &= u_{2,i} + w_{2,0}v_{2,i}, \\ x &= a + ih. \end{aligned}$$

Mostrar  $(x, w_{1,i}, w_{2,i})$ .

7. Parar.

**Problemas por errores de redondeo.** Desafortunadamente, esta técnica, por errores de redondeo, puede contener problemas ocultos. Si  $u(x)$  crece rápidamente cuando  $x$  recorre  $[a, b]$ , entonces  $u_{1,N}$  es grande; si además,  $\beta$  es pequeño en comparación con  $u_{1,N}$ , entonces

$$\begin{aligned} w_{2,0} &= \frac{\beta - u_{1,N}}{v_{1,N}} \\ &\approx -\frac{u_{1,N}}{v_{1,N}}. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} w_{1,i} &= u_{1,i} + w_{2,0}v_{1,i} \\ &\approx u_{1,i} - \frac{u_{1,N}}{v_{1,N}}v_{1,i}, \end{aligned}$$

lo cual permite una posible pérdida de dígitos significativos debido a la cancelación. Pero como  $u_{1,i}$  es una aproximación a  $u(x_i)$ , se puede entonces vigilar el comportamiento de  $u(x)$ , y si  $u_{1,i}$  aumenta rápidamente de

a a b, podemos aplicar hacia atrás el método del Disparo, esto es, resolver en su lugar los problemas de valor inicial

$$u'' = p(x)u' + q(x)u + r(x), \quad a \leq x \leq b, \quad u(b) = \beta, \quad u'(b) = 0,$$

y

$$v'' = p(x)v' + q(x)v, \quad a \leq x \leq b, \quad v(b) = 0, \quad v'(b) = 1.$$

Si este método de Disparo inverso todavía presenta la eliminación de los dígitos significativos y si el aumento de precisión no produce mayor exactitud, será necesario utilizar otros métodos numéricos, como las que explicaremos más adelante.

**Análisis del error.** En muchas ocasiones estamos interesados en encontrar cotas para el error que se está produciendo al momento de aproximar. El siguiente teorema nos proporciona dichas cotas. Para no recargar demasiado la exposición, presentaremos la prueba en el Apéndice A, al que referimos al lector interesado.

**Teorema 2.1** Sea  $h = (b - a)/N$ ,  $x_i = a + nh$ , con  $i = 0, 1, \dots, N$ . Sean  $y(x)$ ,  $u(x)$  y  $v(x)$  y las soluciones respectivas de (2.1), (2.2) y (2.3). Si  $\tilde{u}_i$  y  $\tilde{v}_i$  son aproximaciones de  $\mathcal{O}(h^n)$  para  $u_i = u(x_i)$  y  $v_i = v(x_i)$ , respectivamente, entonces  $w_i$  será una aproximación de  $\mathcal{O}(h^n)$  para  $y_i = y(x_i)$ . En particular,

$$|w_i - y_i| \leq Kh^n \left| 1 + \frac{v_i}{v_N} \right|,$$

para alguna  $K$  apropiada.

**Demostración:** Ver Apéndice A, Sección A.1. ■

### 2.1.2. Planteamiento para el caso $y'' = f(x, y, y')$

El método del Disparo para el PVF no lineal

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta, \tag{2.5}$$

utiliza las soluciones de una sucesión de PVI de la forma que contenga un parámetro  $t$ , para aproximar la solución al PVF

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = t. \tag{2.6}$$

Esto se hace escogiendo los parámetros  $t = t_k$  de tal forma que

$$\lim_{k \rightarrow \infty} y(b, t_k) = y(b) = \beta,$$

donde  $y(b, t_k)$  denotará la solución de (2.6) con  $t = t_k$ , mientras  $y(x)$  denota la solución de (2.5). Esta técnica se conoce con el nombre de método del Disparo, por la analogía con el procedimiento de dispararle a objetos situados en un blanco fijo (ver Figura 2.2). Se comienza con un parámetro  $t_0$  que determina la elevación inicial a la cual se le dispara al objeto desde el punto  $(a, \alpha)$  y a lo largo de la curva descrita por la solución al PVI:

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = t_0.$$

Si  $y(b, t_0)$  no está suficientemente cerca de  $\beta$ , se cambia la aproximación seleccionando las elevaciones  $t_1, t_2$ , y así sucesivamente, hasta que  $y(b, t_k)$  esté bastante cerca de acertar en el blanco  $\beta$  (Ver Figura 2.3). Para determinar los parámetros  $t_k$ , supongamos que (2.5) satisface las hipótesis del Teorema 1.3. Si  $y(x, t)$  denota la solución de (2.6), el problema consistirá en determinar  $t$  tal que

$$y(b, t) - \beta = 0. \tag{2.7}$$

Esta es una ecuación no lineal, y para resolverlo se disponen de varios métodos.

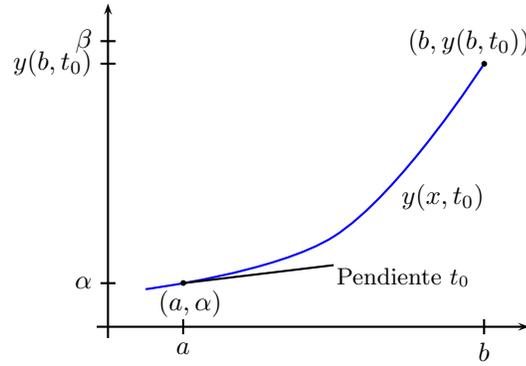


Figura 2.2: Método del Disparo para el caso no lineal.

**Método de la Secante.** Para emplear este método, se necesitan elegir las aproximaciones iniciales  $t_0$  y  $t_1$ , y luego generar los términos restantes de la sucesión mediante

$$t_k = t_{k-1} - \frac{(y(b, t_{k-1}) - \beta)(t_{k-1} - t_{k-2})}{y(b, t_{k-1}) - y(b, t_{k-2})}, \quad k = 2, 3, \dots$$

**Método de Newton.** Para generar la sucesión  $\{t_k\}$  con este método, que es más poderoso, sólo se necesita una aproximación inicial  $t_0$ . Sin embargo, la iteración tiene la forma

$$t_k = t_{k-1} - \frac{y(b, t_{k-1}) - \beta}{\frac{dy}{dt}(b, t_{k-1})}, \quad (2.8)$$

y requiere conocer  $\frac{dy}{dt}(b, t_{k-1})$ . Esto presenta un problema porque no se conoce una representación explícita de  $y(b, t)$ , sólo se conocen los valores  $y(b, t_0), y(b, t_1), \dots, y(b, t_{k-1})$ . Supongamos que se reescribe la ecuación (2.6), haciendo énfasis en que la solución se basa tanto en  $x$  como en el parámetro  $t$ :

$$y''(x, t) = f(x, y(x, t), y'(x, t)), \quad a \leq x \leq b, \quad y(a, t) = \alpha, \quad y'(a, t) = t. \quad (2.9)$$

Se conserva la notación prima para indicar la derivada respecto a  $x$ . Puesto que se necesita determinar  $\frac{dy}{dt}(b, t)$  cuando  $t = t_{k-1}$ , primero se toma la derivada parcial de (2.9) respecto a  $t$ . Esto significa que

$$\begin{aligned} \frac{\partial y''}{\partial t}(x, t) &= \frac{\partial f}{\partial t}(x, y(x, t), y'(x, t)) \\ &= \frac{\partial f}{\partial x}(x, y(x, t), y'(x, t)) \frac{\partial x}{\partial t}(x, t) + \frac{\partial f}{\partial y}(x, y(x, t), y'(x, t)) \frac{\partial y}{\partial t}(x, t) \\ &\quad + \frac{\partial f}{\partial y'}(x, y(x, t), y'(x, t)) \frac{\partial y'}{\partial t}(x, t). \end{aligned}$$

Dado que  $x$  y  $t$  son independientes,  $\frac{\partial x}{\partial t} = 0$ , y

$$\frac{\partial y''}{\partial t}(x, t) = \frac{\partial f}{\partial y}(x, y(x, t), y'(x, t)) \frac{\partial y}{\partial t}(x, t) + \frac{\partial f}{\partial y'}(x, y(x, t), y'(x, t)) \frac{\partial y'}{\partial t}(x, t), \quad (2.10)$$

con  $a \leq x \leq b$ . Las condiciones iniciales dan

$$\begin{aligned} \frac{\partial y}{\partial t}(a, t) &= 0, \\ \frac{\partial y'}{\partial t}(a, t) &= 1. \end{aligned}$$

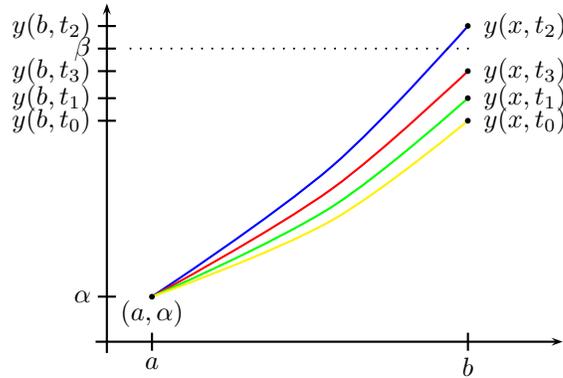


Figura 2.3: Escogencia de los parámetros  $t_k$ .

Si se simplifica la notación usando  $z(x, t)$  para denotar  $\frac{\partial y}{\partial t}(x, t)$  y si se supone que el orden de derivación de  $x$  y  $t$  puede invertirse, con las condiciones iniciales, la ecuación (2.10) se convierte en el PVI:

$$z''(x, t) = \frac{\partial f}{\partial y'}(x, y, y')z'(x, t) + \frac{\partial f}{\partial y}(x, y, y')z(x, t), \quad a \leq x \leq b, \quad z(a, t) = 0, \quad z'(a, t) = 1. \quad (2.11)$$

Así pues, el método de Newton requiere que las ecuaciones (2.9) y (2.11) sean resueltas en cada iteración. Entonces, conforme a la ecuación (2.8),

$$t_k = t_{k-1} - \frac{y(b, t_{k-1}) - \beta}{z(b, t_{k-1})}. \quad (2.12)$$

Para efecto de nuestro estudio, se empleará el método de Runge Kutta de cuarto orden para aproximar las soluciones de (2.9) y (2.11).

### Algoritmo del Método del Disparo para el caso no lineal.

**Entrada:** extremos  $a, b$ ; condiciones de frontera  $\alpha, \beta$ ; número de subintervalos  $N \geq 2$ ; tolerancia  $\text{Tol}$ , número máximo de iteraciones  $M$ .

**Salida:** aproximaciones  $w_{1,i}$  a  $y(x_i)$ ;  $w_{2,i}$  a  $y'(x_i)$ , para cada  $i = 0, \dots, N$ , o bien un mensaje de que se excedió el número máximo de iteraciones.

1. Tomar  $h = (b - a)/N$ ;  $u_{1,0} = \alpha$ ;  $k = 1$ ,  $TK = (\beta - \alpha)/(b - a)$ .
2. Mientras  $k \leq M$ , hace los pasos 3 - 10.
  3. Tomar  $w_{1,0} = \alpha$ ,  $w_{2,0} = TK$ ,  $u_1 = 0$ ,  $u_2 = 1$ .
  4. Para  $i = 1, \dots, N$ , hacer los pasos 5 - 6.
  5. Tomar  $x = a + (i - 1)h$ .
  6. Aplicar Runge Kutta para obtener  $w_{1,i}$ ,  $w_{2,i}$ ,  $u_1$ ,  $u_2$ .
  7. Si  $|w_{1,N} - \beta| < \text{Tol}$ , entonces hacer los pasos 8 y 9.
    8. Para  $i = 0, 1, \dots, N$ , tomar  $x = a + ih$ , mostrar  $(x, w_{1,i}, w_{2,i})$ .
    9. Procedimiento terminado, PARAR.
  10. Tomar  $k = k + 1$  y  $TK = TK - (w_{1,N} - \beta)/u_1$ . El método de Newton se emplea para calcular  $TK$ .
11. Número máximo de iteraciones excedido, procedimiento terminado sin éxito, PARAR.

El valor  $t_0 = TK$  escogido en el paso 1 es la pendiente de la recta que pasa por  $(a, \alpha)$  y por  $(b, \beta)$ . Como el método del Disparo emplea el método de Runge Kutta de cuarto orden para aproximar la solución de dos problemas de valor inicial, entonces la solución encontrada aproxima la solución exacta con un orden de aproximación  $\mathcal{O}(h^4)$  (ver [9], Capítulo 8, Sección 7.2, pág. 426).

## 2.2. Método de Diferencias Finitas

El método de Diferencias Finitas, al igual que Disparo, es considerado clásico dentro de la literatura. Este método, presentado en las referencias [9, 18], se basa en la idea de aproximar las derivadas que aparecen en un problema de contorno o valor de frontera mediante una fórmula de diferencias sobre una malla de puntos. Luego se sustituye dichas aproximaciones en la ecuación diferencial para posteriormente resolver un sistema de ecuaciones lineales o no lineales, dependiendo si el PVF es lineal o no lineal.

### 2.2.1. Planteamiento para el caso $y'' = p(x)y' + q(x)y + r(x)$

Para el problema de contorno

$$y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

se procede a discretizar el intervalo  $[a, b]$  en una malla de puntos  $a = x_0 < x_1 < \dots < x_{n+1} = b$ , el cual se tomará uniforme, esto es,  $x_{i+1} - x_i = h$ , aunque, en general, los puntos no necesitan estar distribuidos de esta manera.

En los puntos interiores de la malla ( $x_i = a + ih, i = 1, 2, \dots, n$ ), la ecuación diferencial a aproximar es

$$y''(x_i) = p(x_i)y'(x_i) + q(x_i)y(x_i) + r(x_i). \quad (2.13)$$

Si se usa el polinomio de Taylor de orden 3 para  $y(x)$ , alrededor de  $x_i$ , y se evalúa en  $x_{i+1}$  y  $x_{i-1}$ , entonces, suponiendo que  $y \in C^4[x_{i-1}, x_{i+1}]$ , se tiene

$$\begin{aligned} y(x_{i+1}) &= y(x_i + h) \\ &= y(x_i) + hy'(x_i) + \frac{h^2}{2}y''(x_i) + \frac{h^3}{6}y^{(3)}(x_i) + \frac{h^4}{24}y^{(4)}(\xi_i^+), \end{aligned}$$

para algún  $\xi_i^+$  en  $(x_i, x_{i+1})$ , y

$$\begin{aligned} y(x_{i-1}) &= y(x_i - h) \\ &= y(x_i) - hy'(x_i) + \frac{h^2}{2}y''(x_i) - \frac{h^3}{6}y^{(3)}(x_i) + \frac{h^4}{24}y^{(4)}(\xi_i^-), \end{aligned}$$

para algún  $\xi_i^-$  en  $(x_{i-1}, x_i)$ .

Si sumamos adecuadamente estas dos últimas igualdades,

$$y(x_{i+1}) + y(x_{i-1}) = 2y(x_i) + h^2y''(x_i) + \frac{h^2}{24}\left(y^{(4)}(\xi_i^+) + y^{(4)}(\xi_i^-)\right),$$

y al despejar  $y''(x_i)$ ,

$$y''(x_i) = \frac{1}{h^2}\left(y(x_{i+1}) - 2y(x_i) + y(x_{i-1})) - \frac{h^2}{24}\left(y^{(4)}(\xi_i^+) + y^{(4)}(\xi_i^-)\right).$$

Ahora, aplicando el teorema de valor intermedio, tenemos que

$$y''(x_i) = \frac{1}{h^2}\left(y(x_{i+1}) - 2y(x_i) + y(x_{i-1})) - \frac{h^2}{12}y^{(4)}(\xi_i), \quad (2.14)$$

para algún  $\xi_i$  en  $(x_{i-1}, x_{i+1})$ . A este esquema se le llama *fórmula de diferencias centradas* para  $y''(x_i)$ . De manera semejante se obtiene una fórmula de este tipo para  $y'(x_i)$ , el cual viene dada por

$$y'(x_i) = \frac{1}{2h} \left( y(x_{i+1}) - y(x_{i-1}) \right) - \frac{h^2}{6} y^{(3)}(\eta_i), \quad (2.15)$$

para algún  $\eta_i$  en  $(x_{i-1}, x_{i+1})$ .

**Formación del sistema de ecuaciones.** Sustituyendo (2.14) y (2.15) en (2.13), entonces se obtiene la igualdad

$$\begin{aligned} \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} &= p(x_i) \left( \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} \right) + q(x_i)y(x_i) + r(x_i) \\ &\quad - \frac{h^2}{12} \left( 2p(x_i)y^{(3)}(\eta_i) - y^{(4)}(\xi_i) \right). \end{aligned}$$

El método de Diferencias Finitas con error de truncamiento de orden  $\mathcal{O}(h^2)$  se obtiene empleando esta ecuación junto con las condiciones de frontera  $y(a) = \alpha$  y  $y(b) = \beta$  para definir

$$w_0 = \alpha, \quad w_{n+1} = \beta$$

y

$$\frac{w_{i+1} - 2w_i + w_{i-1}}{h^2} - p(x_i) \left( \frac{w_{i+1} - w_{i-1}}{2h} \right) - q(x_i)w_i = r(x_i), \quad (2.16)$$

para cada  $i = 1, 2, \dots, n$ . Reescribiendo la ecuación anterior, se tiene que

$$-\left(1 + \frac{h}{2}p(x_i)\right)w_{i-1} + \left(2 + h^2q(x_i)\right)w_i - \left(1 - \frac{h}{2}p(x_i)\right)w_{i+1} = -h^2r(x_i).$$

El sistema de ecuaciones resultante se expresa en la forma

$$Aw = z, \quad (2.17)$$

donde,

$$A = \begin{pmatrix} \gamma_1 & \delta_1 & 0 & \cdots & \cdots & 0 \\ \epsilon_2 & \gamma_2 & \delta_2 & \ddots & & \vdots \\ 0 & \epsilon_3 & \gamma_3 & \delta_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \delta_{n-1} \\ 0 & \cdots & \cdots & 0 & \epsilon_n & \gamma_n \end{pmatrix}, \quad \begin{aligned} \gamma_i &= 2 + h^2q(x_i), & 1 \leq i \leq n, \\ \delta_i &= -1 + \frac{h}{2}p(x_i), & 1 \leq i \leq n-1, \\ \epsilon_i &= -1 - \frac{h}{2}p(x_i), & 2 \leq i \leq n, \end{aligned}$$

$$w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix} \quad \text{y} \quad z = \begin{pmatrix} -h^2r(x_1) + \left(1 + \frac{h}{2}p(x_1)\right)w_0 \\ -h^2r(x_2) \\ -h^2r(x_3) \\ \vdots \\ -h^2r(x_{n-1}) \\ -h^2r(x_n) + \left(1 + \frac{h}{2}p(x_n)\right)w_{n+1} \end{pmatrix}.$$

El siguiente teorema establece bajo cuáles condiciones el sistema lineal (2.17) tiene solución única. Su demostración es consecuencia del Teorema 1.2.

**Teorema 2.2** *Supongamos que  $p(x)$ ,  $q(x)$  y  $r(x)$  son continuas en  $[a, b]$ . Si  $q(x) \geq 0$  en  $[a, b]$ , entonces el sistema lineal tridiagonal (2.17) tiene una solución única siempre y cuando  $h < \frac{2}{L}$ , donde  $L = \max_{a \leq x \leq b} |p(x)|$ .*

**Demostración:** Ver Sección A.2 del Apéndice A. ■

**Algoritmo del Método de Diferencias Finitas para el caso lineal.**

**Entrada:** extremos  $a, b$ ; condiciones de frontera  $\alpha, \beta$ ;  $N \geq 2$ .

**Salida:** aproximaciones  $w_i$  a  $y(x_i)$ , para cada  $i = 0, \dots, N + 1$ .

1. Tomar

$$\begin{aligned} h &= (b - a)/(N + 1), \\ x &= a + h, \\ a_1 &= 2 + h^2q(x), \\ b_1 &= -1 + (h/2)p(x), \\ d_1 &= -h^2r(x) + (1 + (h/2)p(x))\alpha. \end{aligned}$$

2. Para  $i = 2, \dots, N - 1$ , tomar

$$\begin{aligned} x &= a + ih, \\ a_i &= 2 + h^2q(x), \\ b_i &= -1 + (h/2)p(x), \\ c_i &= -1 - (h/2)p(x), \\ d_i &= -h^2r(x). \end{aligned}$$

3. Tomar

$$\begin{aligned} x &= b - h, \\ a_N &= 2 + h^2q(x), \\ c_N &= -1 - (h/2)p(x), \\ d_N &= -h^2r(x) + (1 - (h/2)p(x))\beta. \end{aligned}$$

4. Tomar

$$\begin{aligned} l_1 &= a_1, \\ u_1 &= b_1/a_1, \\ z_1 &= d_1/l_1. \end{aligned}$$

(Los pasos 4 - 8 resuelven un sistema lineal tridiagonal utilizando el algoritmo de Crout).

5. Para  $i = 2, \dots, N - 1$ , tomar

$$\begin{aligned} l_i &= a_i - c_i u_{i-1}, \\ u_i &= b_i/l_i \\ z_i &= (d_i - c_i z_{i-1})/l_i. \end{aligned}$$

6. Tomar

$$\begin{aligned} l_N &= a_N - c_N u_{N-1}, \\ z_N &= (d_N - c_N z_{N-1})/l_N. \end{aligned}$$

7. Tomar

$$\begin{aligned} w_0 &= \alpha, \\ w_{N+1} &= \beta, \\ w_N &= z_N. \end{aligned}$$

8. Para  $N - 1, \dots, 1$ , tomar  $w_i = z_i - u_i w_{i+1}$ .

9. Para  $i = 0, \dots, N + 1$ , tomar  $x = a + ih$ . Mostrar  $(x, w_i)$ .

10. Parar.

**Análisis del error** El siguiente teorema nos ayuda a deducir cotas para el error de la aproximación de las ecuaciones diferenciales mediante el esquema de Diferencias Finitas de segundo orden. La demostración está detallada en la Sección A.2 del Apéndice A.

**Teorema 2.3** Sean  $Q^*$  tal que  $q(x) \geq Q^* > 0$ , para  $a \leq x \leq b$ ,  $L = \max_{a \leq x \leq b} |p(x)|$ ,  $M_3 = \max_{a \leq x \leq b} |y^{(3)}(x)|$  y  $M_4 = \max_{a \leq x \leq b} |y^{(4)}(x)|$ . Entonces, para  $i = 0, 1, \dots, n+1$ , se cumple:

$$|w_i - y(x_i)| \leq h^2 \left( \frac{M_4 + 2LM_3}{12Q^*} \right).$$

**Demostración:** Ver Apéndice A, Sección A.2. ■

El teorema anterior nos dice que la solución de Diferencias Finitas converge a la solución exacta cuando  $h \rightarrow 0$ , y el error es  $\mathcal{O}(h^2)$ .

**Observación.** Cuando, en la ecuación diferencial lineal,  $p(x) = 0$ , entonces se puede demostrar que el error cometido es del orden  $\mathcal{O}(h^4)$ .

### 2.2.2. Planteamiento para el caso $y'' = f(x, y, y')$

Cuando el PVF es no lineal,

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

el método de Diferencias Finitas se parece al que se aplicó en la subsección anterior para el caso lineal. Sin embargo, el sistema de ecuaciones será no lineal y, por tanto, se requiere un proceso iterativo para resolverlo. Nuevamente, se procede a discretizar el intervalo  $[a, b]$  en una malla de puntos  $a = x_0 < x_1 < \dots < x_{n+1} = b$ , que tomaremos equiespaciada o uniforme, es decir,  $x_{i+1} - x_i = h$ .

En los puntos interiores de la malla ( $x_i = a + ih$ ,  $i = 1, 2, \dots, n$ ), la ecuación diferencial a aproximar es:

$$y''(x_i) = f(x_i, y(x_i), y'(x_i)). \tag{2.18}$$

**Formación del sistema de ecuaciones.** Sustituyendo (2.14) y (2.15) en (2.18), obtendremos:

$$\frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} = f\left(x_i, y(x_i), \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} - \frac{h^2}{6}y^{(3)}(\eta_i)\right) + \frac{h^2}{12}y^{(4)}(\xi_i),$$

para algún  $\xi_i$  y  $\eta_i$  en el intervalo  $(x_{i-1}, x_{i+1})$ , con  $i = 1, 2, \dots, n$ . El método de Diferencias Finitas se obtiene empleando esta ecuación junto con las condiciones de frontera  $y(a) = \alpha$  y  $y(b) = \beta$  para definir

$$w_0 = \alpha, \quad w_{n+1} = \beta,$$

y

$$\frac{w_{i+1} - 2w_i + w_{i-1}}{h^2} - f\left(x_i, w_i, \frac{w_{i+1} - w_{i-1}}{2h}\right) = 0,$$

para cada  $i = 1, 2, \dots, n$ . El sistema no lineal de  $n \times n$  obtenido con este método

$$F_i(w_1, \dots, w_n) = 0, \tag{2.19}$$

con

$$F_i(w_1, \dots, w_n) = w_{i+1} - 2w_i + w_{i-1} - h^2 f\left(x_i, w_i, \frac{w_{i+1} - w_{i-1}}{2h}\right), \quad i = 1, \dots, n,$$

tiene solución única si  $h < \frac{2}{L}$ , donde  $L$  es tal que  $\left| \frac{\partial f}{\partial y'}(x, y, y') \right| \leq L$  (ver [13], pág. 86).

**Resolución del sistema no lineal.** Para resolver el sistema no lineal (2.19), se usa el método de Newton para sistemas de ecuaciones no lineales. Si tomamos inicialmente  $w^{(0)} = (w_1^{(0)}, \dots, w_n^{(0)})^T$ , la iteración mediante este método es:

$$w^{(j+1)} = w^{(j)} - J^{-1}(w^{(j)})\mathbf{F}(w^{(j)}),$$

donde  $J(w^{(j)})$  es la matriz jacobiana. En la práctica es más conveniente a efectos computacionales iterar resolviendo las ecuaciones:

$$J(w^{(j)})(w^{(j+1)} - w^{(j)}) = -\mathbf{F}(w^{(j)}).$$

Calculemos el jacobiano  $[J(w^{(j)})]_{p,q} = \frac{\partial F_p}{\partial w_q}$ :

$$\begin{aligned} [J]_{k,k} &= -2 - h^2 \frac{\partial f}{\partial y} \left( x_k, w_k, \frac{w_{k+1} - w_{k-1}}{2h} \right), \\ [J]_{k,k+1} &= 1 - \frac{h}{2} \frac{\partial f}{\partial y'} \left( x_k, w_k, \frac{w_{k+1} - w_{k-1}}{2h} \right), \\ [J]_{k+1,k} &= 1 + \frac{h}{2} \frac{\partial f}{\partial y'} \left( x_k, w_k, \frac{w_{k+1} - w_{k-1}}{2h} \right), \end{aligned}$$

con  $k = 1, \dots, n$ . El resto de las entradas de la matriz  $J$  son nulos. En particular, como todos los elementos con  $|p - q| > 1$  son nulos, entonces  $J$  es tridiagonal. En cada iteración del sistema lineal

$$J(w^{(j)})v = -\mathbf{F}(w^{(j)}),$$

con  $v = (v_1, \dots, v_n)^T$ , se requiere obtener  $v$ , porque

$$w^{(j+1)} = w^{(j)} + v.$$

Puesto que  $J$  es tridiagonal, se puede utilizar un método para resolver sistemas de ecuaciones lineales. Para efecto de la implementación del código, se usará el algoritmo de reducción de Crout para sistemas tridiagonales (ver referencia [18], Capítulo 6, Sección 6.6, pág. 408).

### Algoritmo del Método de Diferencias Finitas para el caso no lineal.

**Entrada:** extremos  $a, b$ ; condiciones de frontera  $\alpha, \beta$ ; entero  $N \geq 2$ ; tolerancia Tol, número máximo de iteraciones  $M$ .

**Salida:** aproximaciones  $w_i$  a  $y(x_i)$ , para cada  $i = 0, \dots, N + 1$ , o un mensaje de que se excedió el número máximo de iteraciones.

1. Tomar  $h = (b - a)/(N + 1)$ ;  $w_0 = \alpha$ ;  $w_{N+1} = \beta$ .
2. Para  $i = 1, \dots, N$ , tomar  $w_i = \alpha + i \left( \frac{\beta - \alpha}{b - a} \right) h$ .
3. Tomar  $k = 1$ .
4. Mientras  $k \leq M$ , hacer los pasos 5 - 16.

5. Tomar

$$\begin{aligned}x &= a + h, \\t &= (w_2 - \alpha)/2h, \\a_1 &= 2 + h^2 f_y(x, w_1, t), \\b_1 &= -1 + (h/2) f_{y'}(x, w_1, t), \\d_1 &= -(2w_1 - w_2 - \alpha + h^2 f(x, w_1, t)).\end{aligned}$$

6. Para  $i = 2, \dots, N - 1$ , tomar

$$\begin{aligned}x &= a + ih, \\t &= (w_{i+1} - w_{i-1})/2h, \\a_i &= 2 + h^2 f_y(x, w_i, t), \\b_i &= -1 + (h/2) f_{y'}(x, w_i, t), \\c_i &= -1 - (h/2) f_{y'}(x, w_i, t), \\d_i &= -(2w_i - w_{i+1} - w_{i-1} + h^2 f(x, w_i, t)).\end{aligned}$$

7. Tomar

$$\begin{aligned}x &= b - h, \\t &= (\beta - w_{N-1})/2h, \\a_N &= 2 + h^2 f_y(x, w_N, t), \\c_N &= -1 - (h/2) f_{y'}(x, w_N, t), \\d_N &= -(2w_N - w_{N-1} - \beta + h^2 f(x, w_N, t)).\end{aligned}$$

8. Tomar  $l_1 = a_1$ ;  $u_1 = b_1/a_1$ ;  $z_1 = d_1/l_1$ . (Los pasos 8-12 resuelven un sistema lineal tridiagonal utilizando el algoritmo de Crout).

9. Para  $i = 2, \dots, N - 1$ , tomar  $l_i = a_i - c_i u_{i-1}$ ;  $u_i = b_i/l_i$ ;  $z_i = (d_i - c_i z_{i-1})/l_i$ .

10. Tomar  $l_N = a_N - c_N u_{N-1}$ ;  $z_N = (d_N - c_N z_{N-1})/l_N$ .

11. Tomar  $v_N = z_N$ ;  $w_N = w_N + v_N$ .

12. Para  $i = N - 1, \dots, 1$ , tomar  $v_i = z_i - u_i v_{i+1}$ ;  $w_i = w_i + v_i$ .

13. Si  $\|v\| \leq \text{Tol}$ , hacer los pasos 14-15.

14. Para  $i = 0, \dots, N + 1$ , tomar  $x = a + ih$ . Mostrar  $(x, w_i)$ .

15. Parar. Procedimiento terminado con éxito.

16. Tomar  $k = k + 1$ .

17. Máximo de iteraciones, procedimiento terminado sin éxito. Parar.

Al igual que el caso lineal, la solución encontrada por el método de Diferencias Finitas aproxima a la solución exacta con un orden de aproximación  $\mathcal{O}(h^2)$  (ver [9], Capítulo 8, Sección 7.2, pág. 432 - 433).

### 2.3. Método de Elementos Finitos

El método de Elementos Finitos (MEF), el cual se plantea en las referencias [3, 17], constituye hoy en día el procedimiento habitual para la aproximación numérica de la solución de problemas prácticos que surgen en Ingeniería o las Ciencias. Se trata de un método general para la solución de problemas de contorno de ecuaciones diferenciales ordinarias o parciales. En esencia se trata de una técnica que sustituye el problema diferencial por otro algebraico, aproximadamente equivalente, para el cual se conocen técnicas generales de resolución. Para ello hace uso de la discretización o subdivisión de una región sobre la cual están definidas las ecuaciones en formas geométricas simples denominadas elementos finitos. Una de las ventajas de este método es su facilidad de implementación en un programa computacional, que a su vez es una condición básica para su utilización ya que el tratamiento de un problema, en particular, debe efectuarse en un número muy elevado de operaciones para resolver sistemas algebraicos del orden de cientos o miles de ecuaciones. No obstante, esta cantidad no es una limitación con las computadoras estándar de hoy.

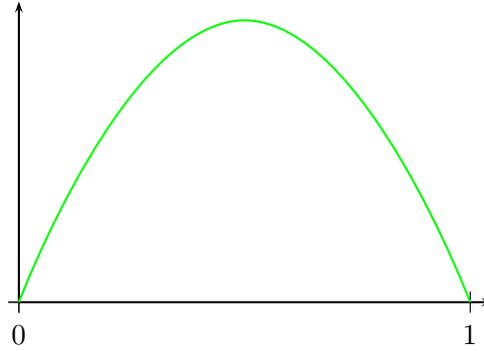


Figura 2.4: La función  $y(x) = -3x^2 + 3x$  es un ejemplo de una función en  $V$ .

El MEF fue al principio desarrollado en 1943 por R. Courant, quién utilizó el método de Ritz de análisis numérico y minimización de las variables de cálculo para obtener soluciones aproximadas a un sistema de vibración. En la década de 1960 el método fue generalizado para la solución aproximada de problemas de análisis de tensión, flujo de fluidos y transferencia de calor. El primer libro sobre Elementos Finitos fue publicado en 1967 por Zienkiewicz y Cheung ([19]). En la década de 1970 el método fue extendido al análisis de problemas no lineales de la mecánica del continuo. Hoy el método permite resolver prácticamente cualquier situación física que pueda formularse mediante un sistema de ecuaciones diferenciales.

### 2.3.1. Formulación débil de un problema modelo unidimensional

Consideremos el problema de valor de frontera siguiente:

$$\begin{aligned} -u''(x) &= f(x), \\ u(0) &= 0, \\ u(1) &= 0, \end{aligned} \tag{2.20}$$

donde  $v' = \frac{dv}{dx}$  y  $f$  es una función continua. Se introduce el siguiente espacio vectorial de funciones:

$$V = \{v : v \in C[0, 1], v' \text{ es continua a trozos y acotada en } [0, 1], v(0) = v(1) = 0\}.$$

Si se multiplica la ecuación diferencial dada en (2.20) por una función cualquiera de  $V$  (ver Figura 2.4) e integramos en el intervalo  $[0, 1]$  usando la fórmula de integración por partes, entonces se tiene:

$$-\int_0^1 u''(x)v(x)dx = -u'(1)v(1) + u'(0)v(0) + \int_0^1 u'(x)v'(x)dx = \int_0^1 u'(x)v'(x)dx,$$

y concluimos que para toda función  $v \in V$ , la solución  $u$  del problema modelo (2.20) verifica:

$$\mathcal{B}(u, v) = \mathcal{F}(v), \tag{2.21}$$

donde  $\mathcal{B}(u, v) = \int_0^1 u'(x)v'(x)dx$  es una forma bilineal y  $\mathcal{F}(v) = \int_0^1 f(x)v(x)dx$  es un funcional lineal. La formulación anterior (2.21) se llama *formulación débil o variacional* del problema de partida (2.20). Las dos formulaciones no son estrictamente equivalentes. En efecto, toda solución de (2.20) es solución de (2.21). Sin embargo, para que toda solución de (2.21) sea solución de (2.20), se debe exigir que dicha solución sea dos veces continuamente diferenciable. La existencia y unicidad de la solución de (2.21) queda garantizada en virtud del Teorema de Lax Milgram y el Teorema de Representación de Riesz (Ver Preliminares, Teoría de Análisis Funcional).

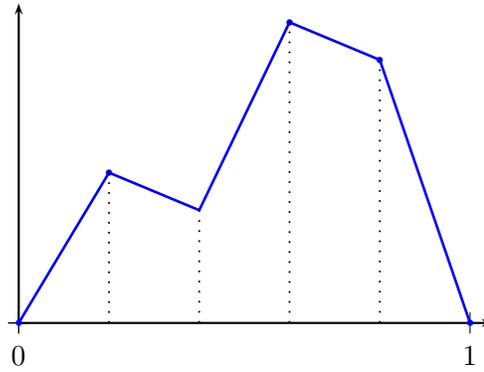


Figura 2.5: Ejemplo de una función  $v_h \in V_h$ , 5 subintervalos.

### 2.3.2. El MEF para el problema modelo con funciones lineales a trozos

El espacio  $V$  es un espacio de funciones de dimensión infinita. La idea del método de Elementos Finitos es buscar soluciones de (2.21) en un subespacio de  $V$  más sencillo, en particular, en un espacio vectorial de funciones que sea de dimensión finita. Esto permitirá representar cualquier función de este subespacio como una combinación lineal de elementos de una base. En primer lugar, construiremos un subespacio  $V_h$  de  $V$ . Para ello sea  $0 = x_0 < x_1 < \dots < x_M < x_{M+1} = 1$ , una partición del intervalo  $(0, 1)$  en subintervalos  $I_j = (x_{j-1}, x_j)$  de longitud  $h_j = x_j - x_{j-1}$ ,  $j = 1, 2, \dots, M + 1$ , y sea  $h = \max h_j$ . La cantidad  $h$  es una medida de lo fina que es la partición. Consideremos ahora

$$V_h = \{v : v \text{ es lineal en cada subintervalo } I_j, v \in C[0, 1] \text{ y } v(0) = v(1) = 0\}.$$

Un ejemplo de una función en  $V_h$  se aprecia en la Figura 2.5.

Una base del espacio  $V_h$  está constituido por el siguiente conjunto de funciones  $\varphi_j \in V_h$ ,  $j = 1, \dots, M$ , definidas por:

$$\varphi_j(x_i) = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

es decir,  $\varphi_j$  es continua y lineal a trozos, y toma el valor 1 en el nodo  $x_j$  y el valor 0 en los otros nodos. De forma más precisa:

$$\varphi_j(x) = \begin{cases} 0 & \text{si } x \leq x_{j-1}, \\ \frac{x - x_{j-1}}{x_j - x_{j-1}} & \text{si } x_{j-1} < x \leq x_j, \\ \frac{x_{j+1} - x}{x_{j+1} - x_j} & \text{si } x_j < x < x_{j+1}, \\ 0 & \text{si } x_{j+1} \leq x. \end{cases}$$

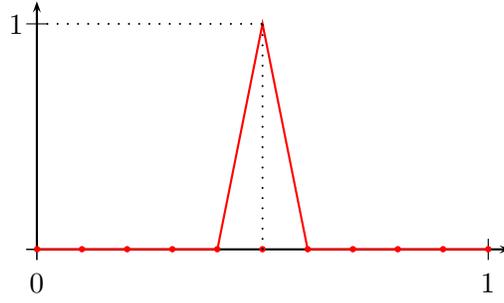
Una función de esta base se aprecia en la Figura 2.6.

Una función  $v_h \in V_h$  se escribe de forma única como combinación lineal de funciones de la base  $\{\varphi_j\}_{j=1}^M$ ,

$$v_h(x) = \sum_{j=1}^M v_h(x_j) \varphi_j(x), \quad x \in [0, 1].$$

El método de Elementos Finitos para el problema (2.20) se formula de la siguiente manera: encontrar  $u_h \in V_h$  tal que, para todo  $v_h \in V_h$ , verifique

$$\mathcal{B}(u_h, v_h) = \mathcal{F}(v_h). \quad (2.22)$$


 Figura 2.6: Ejemplo de una función de la base de  $V_h$ , 10 subintervalos.

Este problema, como vamos a ver, es equivalente a resolver un sistema algebraico lineal. En efecto, por una parte, la solución  $u_h(x)$  se expresa en función de los elementos de la base  $\{\varphi_i\}_{i=1}^M$ , esto es,

$$u_h(x) = \sum_{i=1}^M u_i \varphi_i(x),$$

donde  $u_i = u_h(x_i)$ . Por otro lado, para que se verifique (2.22) con cualquier  $v_h(x)$ , es necesario y suficiente que se verifique para cualquier función de la base. Teniendo en cuenta la linealidad de la integral, resulta que el problema a resolver es:

$$\sum_{i=1}^M \mathcal{B}(\varphi_i(x), \varphi_j(x)) u_i = \mathcal{F}(\varphi_j(x)), \quad \text{para } j = 1, \dots, M.$$

En forma matricial, el sistema se puede escribir como:

$$Au = b, \tag{2.23}$$

donde  $[A]_{ij} = \mathcal{B}(\varphi_i(x), \varphi_j(x))$  y  $b_j = \mathcal{F}(\varphi_j(x))$ .

Los elementos  $[A]_{ij}$  se calculan fácilmente. Primero notemos que si  $|i - j| > 1$ , entonces  $[A]_{ij} = 0$ , pues, en este caso, para  $x \in [0, 1]$ ,  $\varphi_i(x)$  o  $\varphi_j(x)$  es igual a cero. Por lo tanto, la matriz  $A$  es tridiagonal, esto es, únicamente los elementos de la diagonal principal y los elementos de las dos diagonales adyacentes pueden ser diferentes de cero. Así, para  $j = 1, \dots, M$ ,

$$\begin{aligned} \mathcal{B}(\varphi_j(x), \varphi_j(x)) &= \int_0^1 \varphi_j'(x) \varphi_j'(x) dx \\ &= \int_{x_{j-1}}^{x_j} \varphi_j'(x) \varphi_j'(x) dx + \int_{x_j}^{x_{j+1}} \varphi_j'(x) \varphi_j'(x) dx \\ &= \frac{1}{h_j} + \frac{1}{h_{j+1}}, \end{aligned}$$

y para  $j = 2, \dots, M$ ,

$$\begin{aligned} \mathcal{B}(\varphi_j(x), \varphi_{j-1}(x)) &= \int_0^1 \varphi_j'(x) \varphi_{j-1}'(x) dx \\ &= \int_{x_{j-1}}^{x_j} \frac{1}{h_j^2} dx \\ &= -\frac{1}{h_j}. \end{aligned}$$

La matriz  $A$  es simétrica y definida positiva, esto es,  $xAx^T > 0$ , para todo vector fila  $x$  de dimensión  $M$  no nulo. En efecto, notemos que  $[A]_{ij} = \mathcal{B}(\varphi_i(x), \varphi_j(x)) = \mathcal{B}(\varphi_j(x), \varphi_i(x)) = [A]_{ji}$ . Por otro lado, para un vector fila  $x$  arbitrario no nulo, tenemos

$$\begin{aligned} xAx^T &= \sum_{k=1}^M \left( \sum_{p=1}^M v_p \mathcal{B}(\varphi_p(x), \varphi_k(x)) \right) v_k \\ &= \sum_{k=1}^M \mathcal{B} \left( \sum_{p=1}^M v_p \varphi_p(x), \varphi_k(x) \right) v_k \\ &= \mathcal{B} \left( \sum_{p=1}^M v_p \varphi_p(x), \sum_{k=1}^M v_k \varphi_k(x) \right), \\ &= \mathcal{B}(\tilde{v}(x), \tilde{v}(x)) \\ &> 0. \end{aligned}$$

Así, como  $A$  es definida positiva, entonces  $A$  es invertible, por tanto, el sistema (2.23) tiene solución. En el caso especial en que los subintervalos  $I_j$  tengan todos la misma longitud  $h = \frac{1}{M+1}$ , el sistema tiene la forma:

$$\frac{1}{h} \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & \vdots \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 2 \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_M \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}.$$

**Observación:** Si consideramos la ecuación diferencial  $u''(x) = -f(x)$ , con  $u(0) = u(1) = 0$ , entonces la matriz  $A$ , del método de Diferencias Finitas para el caso lineal, coincide con la matriz  $A$  del método de Elementos Finitos.

### 2.3.3. Integración numérica

A la hora de calcular el segundo miembro del sistema (2.23), aparecen integrales de la forma

$$\int_{x_{j-1}}^{x_j} f(x) \varphi_j(x) dx + \int_{x_j}^{x_{j+1}} f(x) \varphi_j(x) dx, \quad (2.24)$$

con  $j = 1, \dots, M$ . Salvo en el caso en que la función  $f$  sea una función constante a trozos, la forma práctica de calcular estas integrales será mediante la integración numérica. Para efecto de este trabajo, se usará la fórmula de cuadratura Gaussiana, específicamente la de dos puntos (Ver [12], Capítulo 5, Sección 5.3, pág. 270). Dicha fórmula dice que

$$\int_{-1}^1 f(\xi) d\xi \approx f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right).$$

Por otro lado, una integral  $\int_a^b g(x) dx$  en un intervalo arbitrario  $[a, b]$ , se puede transformar en otra en  $[-1, 1]$ , usando el cambio de variables

$$t = \frac{2x - a - b}{b - a},$$

el cual equivale a decir que

$$x = \frac{1}{2} \left[ (b - a)t + a + b \right].$$

Esto permite aplicar la fórmula de cuadratura Gaussiana a cualquier intervalo  $[a, b]$ , ya que

$$\int_a^b g(x)dx = \frac{b-a}{2} \int_{-1}^1 g\left(\frac{(b-a)t + b + a}{2}\right) dt.$$

Por comodidad pongamos  $\psi_j(x) = f(x)\varphi_j(x)$ ,  $j = 1, \dots, M$ . Para (2.24), se cumple

$$\begin{aligned} b_j &= \int_{x_{j-1}}^{x_j} \psi_j(x)dx + \int_{x_j}^{x_{j+1}} \psi_j(x)dx \\ &= \frac{h_j}{2} \int_{-1}^1 \psi_j\left(\frac{h_j t + x_j + x_{j-1}}{2}\right) dt + \frac{h_{j+1}}{2} \int_{-1}^1 \psi_j\left(\frac{h_{j+1} t + x_{j+1} + x_j}{2}\right) dt \\ &\approx \frac{h_j}{2} \left( \psi_j\left(\frac{-h_j \tilde{p} + x_j + x_{j-1}}{2}\right) + \psi_j\left(\frac{h_j \tilde{p} + x_j + x_{j-1}}{2}\right) \right) \\ &\quad + \frac{h_{j+1}}{2} \left( \psi_j\left(\frac{-h_{j+1} \tilde{p} + x_{j+1} + x_j}{2}\right) + \psi_j\left(\frac{h_{j+1} \tilde{p} + x_{j+1} + x_j}{2}\right) \right), \end{aligned}$$

donde  $\tilde{p} = \frac{1}{\sqrt{3}}$ .

**Descripción de un algoritmo del Método de Elementos Finitos.** El desarrollo de un algoritmo del método de Elementos Finitos requiere, en general, de cuatro etapas:

1. El problema debe reformularse en forma variacional.
2. El dominio de la variable independiente debe dividirse mediante una partición en subdominios, llamados elementos finitos. Asociada a la partición anterior se construye un espacio vectorial de dimensión finita, llamado espacio de elementos finitos, siendo la solución numérica una combinación lineal en dicho espacio vectorial.
3. Se obtiene la proyección del problema variacional original sobre el espacio de elementos finitos obtenido de la partición. Esto da lugar a un finito sistema de ecuaciones con un número elevado de incógnitas. El número de ecuaciones y de incógnitas será igual a la dimensión del espacio vectorial. En general, cuanto mayor sea dicha dimensión mejor será la aproximación numérica obtenida.
4. El último paso es el cálculo numérico de la solución del sistema de ecuaciones.

Es importante resaltar que la solución estimada por el método de Elementos Finitos aproxima la solución exacta con un orden de aproximación  $\mathcal{O}(h^2)$  (ver [17]).

## 2.4. Método Galerkin Discontinuo

El método Galerkin Discontinuo (GD), presentado en la referencia [1], forma parte de una clase de métodos numéricos para solucionar ecuaciones diferenciales parciales y ordinarias. Combina muchas características del método de Elementos Finitos y se ha aplicado con éxito a problemas hiperbólicos, elípticos y parabólicos. Este método fue propuesto y analizado en los años 70 como una técnica para solucionar numéricamente ecuaciones diferenciales parciales que surgen en muchas aplicaciones.

Este método numérico sigue la misma estructura del método de Elementos Finitos. La idea consiste en reformular el problema dado en forma débil o forma variacional. Se procede a dividir el dominio de la variable independiente en subdominios. Asociada a la partición anterior se construye un espacio vectorial de dimensión finita, siendo la solución numérica una combinación lineal en dicho espacio vectorial, lo que conllevará a un sistema de ecuaciones lineales, es decir, encontrar la solución de una ecuación diferencial se limita a encontrar la solución de un sistema de ecuaciones lineales.

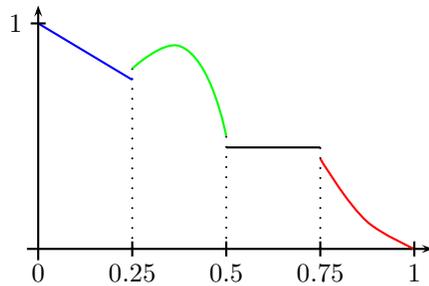


Figura 2.7: Ejemplo de una función  $v$  en  $\mathcal{D}_2(\varepsilon_h)$ , 4 subintervalos de igual longitud. En  $I_0$  e  $I_2$   $v$  es un segmento de recta, mientras que en  $I_1$  e  $I_3$   $v$  es una parábola cuadrática.

### 2.4.1. Formulación débil de un problema modelo unidimensional

Consideremos el problema siguiente en el intervalo  $(0, 1)$ :

$$\begin{aligned} -(K(x)p'(x))' &= f(x), \\ p(0) &= 1, \\ p(1) &= 0, \end{aligned} \tag{2.25}$$

donde  $K \in C^1(0, 1)$ , es positiva y acotada, y  $f \in C^0(0, 1)$ . Vamos a reformular el problema modelo, pero antes se introducirá algunos conceptos que serán útiles para todo lo que sigue.

Sea  $0 = x_0 < x_1 < \dots < x_N = 1$  una partición  $\varepsilon_h$  de  $(0, 1)$ , pongamos  $I_n = (x_n, x_{n+1})$ , y definamos

$$h_n = x_{n+1} - x_n, \quad h_{n-1,n} = \max(h_{n-1}, h_n), \quad h = \max_{0 \leq n \leq N-1} h_n.$$

Denotemos por  $\mathcal{D}_k(\varepsilon_h)$  el espacio de los polinomios discontinuos a trozos de grado  $k$ :

$$\mathcal{D}_k(\varepsilon_h) = \{v : v|_{I_n} \in \mathbb{P}_k(I_n), \forall n = 0, \dots, N-1\},$$

donde  $\mathbb{P}_k(I_n)$  es el espacio de los polinomios de grado  $k$  en el intervalo  $I_n$ . Sean  $v(x_n^+) = \lim_{\epsilon \rightarrow 0} v(x_n + \epsilon)$  y  $v(x_n^-) = \lim_{\epsilon \rightarrow 0} v(x_n - \epsilon)$ , donde  $\epsilon > 0$ . Definiremos el *salto y promedio de  $v$  en los extremos de  $I_n$* , respectivamente, como

$$[v(x_n)] = v(x_n^-) - v(x_n^+), \quad \{v(x_n)\} = \frac{1}{2}(v(x_n^-) + v(x_n^+)), \quad n = 1, \dots, N-1,$$

mientras que el *salto y promedio en los extremos del intervalo  $(0, 1)$*  se define como:

$$[v(x_0)] = -v(x_0^+), \quad \{v(x_0)\} = v(x_0^+), \quad [v(x_N)] = v(x_N^-), \quad \{v(x_N)\} = v(x_N^-).$$

Ahora, se introducen los términos de *salto de la solución y su derivada*. Estos términos vienen dados por

$$J_0(v, w) = \sum_{n=0}^N \frac{\sigma^0}{h_{n-1,n}} [v(x_n)] [w(x_n)], \quad J_1(v, w) = \sum_{n=1}^{N-1} \frac{\sigma^1}{h_{n-1,n}} [v'(x_n)] [w'(x_n)],$$

donde  $\sigma^0$  y  $\sigma^1$  son dos números reales no negativos.

Ahora bien, ya estamos en condición de reformular el problema modelo. Sea  $v$  una función en  $\mathcal{D}_k(\varepsilon_h)$  (Ver Figura 2.7). Si se multiplica la ecuación diferencial dada en (2.25) por  $v$  y se integra en  $I_n$ , con  $n = 0, \dots, N-1$ , entonces, usando la fórmula de integración por partes, se obtiene que

$$\int_{x_n}^{x_{n+1}} K(x)p'(x)v'(x)dx - K(x_{n+1})p'(x_{n+1})v(x_{n+1}^-) + K(x_n)p'(x_n)v(x_n^+) = \int_{x_n}^{x_{n+1}} f(x)v(x)dx.$$

Sumando las  $N$  ecuaciones anteriores y usando la definición de salto para  $v$  en cada nodo de la partición, se tiene

$$\sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)p'(x)v'(x)dx - \sum_{n=0}^N [K(x_n)p'(x_n)v(x_n)] = \int_0^1 f(x)v(x)dx.$$

Notemos que, para  $n = 1, \dots, N-1$ , se cumple

$$[K(x_n)p'(x_n)v(x_n)] = \{K(x_n)p'(x_n)\}[v(x_n)] + \{v(x_n)\}[K(x_n)p'(x_n)]. \quad (2.26)$$

Si se usa (2.26) y sabiendo que la solución exacta  $p$  satisface  $[K(x_n)p'(x_n)] = 0$  en los nodos interiores de la partición  $\varepsilon_h$  (esto gracias a la continuidad de  $p'$ ), se obtiene

$$\sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)p'(x)v'(x)dx - \sum_{n=0}^N \{K(x_n)p'(x_n)\}[v(x_n)] = \int_0^1 f(x)v(x)dx.$$

Como la solución exacta  $p$  es también continua, entonces  $[p(x_n)] = 0$ . Notemos que  $p(x_0) = p(0) = 1$  y  $p(x_N) = p(1) = 0$ , con lo que

$$\begin{aligned} & \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)p'(x)v'(x)dx - \sum_{n=0}^N \{K(x_n)p'(x_n)\}[v(x_n)] + \epsilon \sum_{n=0}^N \{K(x_n)v'(x_n)\}[p(x_n)] \\ &= \int_0^1 f(x)v(x)dx - \epsilon K(x_0)v'(x_0)p(x_0) + \epsilon K(x_N)v'(x_N)p(x_N) \\ &= \int_0^1 f(x)v(x)dx - \epsilon K(x_0)v'(x_0). \end{aligned}$$

El tercer término del primer miembro de la igualdad anterior es casi siempre cero. Además,  $\epsilon$  puede ser cualquier número real, sin embargo, nos restringiremos al caso en que  $\epsilon = -1$ ,  $\epsilon = 0$  o  $\epsilon = 1$ . Así, definimos la forma bilineal del método Galerkin Discontinuo  $\mathcal{B}_\epsilon : \mathcal{D}_k(\varepsilon_h) \times \mathcal{D}_k(\varepsilon_h) \rightarrow \mathbb{R}$ , como

$$\begin{aligned} \mathcal{B}_\epsilon(w, v) &= \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)w'(x)v'(x)dx - \sum_{n=0}^N \{K(x_n)w'(x_n)\}[v(x_n)] \\ &+ \epsilon \sum_{n=0}^N \{K(x_n)v'(x_n)\}[w(x_n)] + J_0(w, v) + J_1(w, v). \end{aligned}$$

La forma bilineal  $\mathcal{B}_\epsilon$  tiene las siguientes propiedades:

1. Para  $\epsilon = -1$ , la forma es simétrica, y además

$$\mathcal{B}_{-1}(v, v) = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)(v'(x))^2 dx - 2 \sum_{n=0}^N \{K(x_n)v'(x_n)\}[v(x_n)] + J_0(v, v) + J_1(v, v).$$

2. Para  $\epsilon = 0$  o  $\epsilon = 1$ , la forma es no simétrica, y cumple

$$\mathcal{B}_1(v, v) = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)(v'(x))^2 dx + J_0(v, v) + J_1(v, v) \geq 0,$$

$$\mathcal{B}_0(v, v) = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)(v'(x))^2 dx - \sum_{n=0}^N \{K(x_n)v'(x_n)\}[v(x_n)] + J_0(v, v) + J_1(v, v).$$

La idea de este método consiste en encontrar  $\tilde{P} \in \mathcal{D}_k(\varepsilon_h)$  tal que

$$\mathcal{B}_\epsilon(\tilde{P}, v) = \mathcal{L}(v), \quad (2.27)$$

para cada  $v \in \mathcal{D}_k(\varepsilon_h)$ , donde  $\mathcal{L} : \mathcal{D}_k(\varepsilon_h) \rightarrow \mathbb{R}$  es el funcional lineal definido por

$$\mathcal{L}(v) = \int_0^1 f(x)v(x)dx - \epsilon K(x_0)v'(x_0) + \frac{\sigma^0}{h_{0,1}}v(x_0).$$

La existencia y unicidad de la solución de (2.27) queda garantizada en virtud del Teorema de Lax Milgram y el Teorema de Representación de Riesz (Ver Preliminares, Teoría de Análisis Funcional).

Dependiendo de la elección de los parámetros  $\epsilon$ ,  $\sigma^0$  y  $\sigma^1$ , el método Galerkin Discontinuo recibe varios nombres.

- Si  $\epsilon = -1$ ,  $\sigma^1 = 0$  y  $\sigma^0$  está acotado inferiormente por un número de gran magnitud, el método resultante es llamado **symmetric interior penalty Galerkin (SIPG)**, introducido después de 1970 por Wheeler (ver [16]) y Arnold (ver [8]).
- Si  $\epsilon = -1$  y  $\sigma^0 = \sigma^1 = 0$ , el método resultante es llamado **global element**, introducido en 1979 por Delves y Hall (ver [15]). Sin embargo, la matriz asociada a la forma bilineal es indefinida, la parte real de los autovalores no son todos positivos y, por tanto, el método es inestable.
- Si  $\epsilon = 1$ ,  $\sigma^1 = 0$  y  $\sigma^0 = 1$ , el método resultante es llamado **nonsymmetric interior penalty Galerkin (NIPG)**, introducido en 1999 por Rivière, Wheeler and Girault (ver [4]).
- Si  $\epsilon = +1$  y  $\sigma^0 = \sigma^1 = 0$ , el método resultante fue introducido por Oden, Babuška y Baumann en 1998 (ver [10]). Haremos referencia a este método como **NIPG 0**, ya que corresponde al caso particular de **NIPG**, con  $\sigma^0 = 0$ .
- Si  $\epsilon = 0$ , obtenemos el método **incomplete interior penalty Galerkin (IIPG)**, el cual fue introducido por Dawson, Sun and Wheeler en 2004 (ver [5]).

Cabe preguntarnos, ¿qué pasa si  $\epsilon = 0$  y  $\sigma^0 = \sigma^1 = 0$ ?. Resulta que el método no es convergente y es inestable, más aún, no se puede probar la existencia y unicidad de la solución.

#### 2.4.2. Derivación del sistema lineal $A\alpha = b$ mediante el uso de polinomios discontinuos a trozos

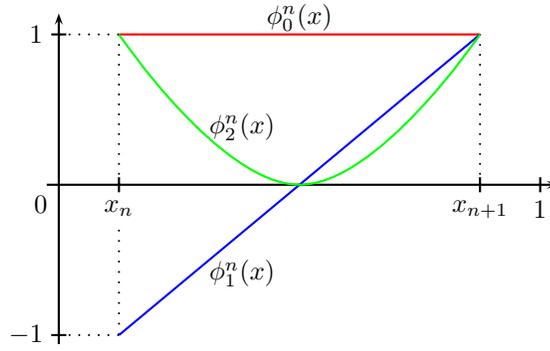
En este apartado se deriva el sistema lineal que se obtiene al emplear el esquema Galerkin Discontinuo para el problema modelo, tomando el caso simple cuando  $K(x)$  es idénticamente igual a 1 y  $\sigma^1 = 0$ . Al mismo tiempo, y por efectos de cálculo, se consideran los polinomios cuadráticos discontinuos a trozos, es decir, cuando se trabaja en  $\mathcal{D}_2(\varepsilon_h)$ .

Una base del espacio  $\mathbb{P}_2(I_n)$  está constituido por el siguiente conjunto de funciones  $\phi_j^n \in \mathbb{P}_2(I_n)$ ,  $j = 0, 1, 2$ , definidas por

$$\phi_0^n(x) = 1, \quad \phi_1^n(x) = 2 \frac{x - \tilde{x}_n}{x_{n+1} - x_n}, \quad \phi_2^n(x) = 4 \frac{(x - \tilde{x}_n)^2}{(x_{n+1} - x_n)^2},$$

donde  $\tilde{x}_n = \frac{1}{2}(x_n + x_{n+1})$  es el punto medio del intervalo  $I_n$ . Notemos que  $\phi_0^n$ ,  $\phi_1^n$  y  $\phi_2^n$  son las traslaciones respectivas de las funciones  $g_0(x) = 1$ ,  $g_1(x) = x$  y  $g_2(x) = x^2$ , desde el intervalo  $(-1, 1)$  al intervalo  $I_n$  (Ver Figura 2.8). Con efecto de simplificar un poco los cálculos, supongamos que  $\varepsilon_h$  es una partición uniforme de  $(0, 1)$ . Por tanto, las funciones de base local y sus derivadas se pueden simplificar:

$$\phi_0^n(x) = 1, \quad \phi_1^n(x) = \frac{2}{h} \left( x - \left( n + \frac{1}{2} \right) h \right), \quad \phi_2^n(x) = \frac{4}{h^2} \left( x - \left( n + \frac{1}{2} \right) h \right)^2, \quad (2.28)$$


 Figura 2.8: Funciones de base local para el espacio  $\mathbb{P}_2(I_n)$ .

$$(\phi_0^n)'(x) = 0, \quad (\phi_1^n)'(x) = \frac{2}{h}, \quad (\phi_2^n)'(x) = \frac{8}{h^2} \left( x - \left( n + \frac{1}{2} \right) h \right). \quad (2.29)$$

Las funciones de base global  $\{\Phi_i^n\}$  para el espacio  $\mathcal{D}_2(\varepsilon_h)$  son obtenidas de las funciones de base local extendiéndolas a cero:

$$\Phi_i^n(x) = \begin{cases} \phi_i^n(x), & x \in I_n, \\ 0, & x \notin I_n. \end{cases}$$

Ahora, para cada  $x \in (0, 1)$ , la solución aproximada  $\tilde{P}$  viene dada por

$$\tilde{P}(x) = \sum_{m=0}^{N-1} \sum_{i=0}^2 \alpha_i^m \Phi_i^m(x), \quad (2.30)$$

donde los coeficientes  $\alpha_i^m$  son desconocidos y requieren ser calculados. Si se sustituye (2.30) en (2.27), entonces, gracias a la linealidad de  $\mathcal{B}_\epsilon$  y  $\mathcal{L}$ , se tiene

$$\sum_{m=0}^{N-1} \sum_{i=0}^2 \alpha_i^m \mathcal{B}_\epsilon(\Phi_i^m(x), \Phi_j^n(x)) = \mathcal{L}(\Phi_j^n(x)),$$

para  $n = 0, \dots, N-1$  y  $j = 0, 1, 2$ . La igualdad anterior no es más que un sistema lineal de la forma  $A\alpha = b$ , donde  $\alpha$  es el vector con componentes  $\alpha_i^m$ ,  $A$  es la matriz con entradas  $\mathcal{B}_\epsilon(\Phi_i^m(x), \Phi_j^n(x))$ , y  $b$  es el vector con componentes  $\mathcal{L}(\Phi_j^n(x))$ .

### 2.4.3. Construcción de la matriz $A$

Las entradas de la matriz global  $A$  pueden ser obtenidas calculando y ensamblando matrices locales, por tanto, vamos a describir cómo calcular dichas matrices. Se reagruparán los términos de la definición  $\mathcal{B}_\epsilon$  en tres grupos: los términos que involucran las integrales sobre  $I_n$ , los términos que involucran los nodos interiores  $x_n$ , y los relacionados con los nodos de la frontera  $x_0$  y  $x_N$ . Primero se considera los términos correspondientes a las integrales sobre los intervalos  $I_n$ . En cada elemento  $I_n$ , la solución  $\tilde{P}$  es un polinomio de grado 2, y se puede expresar como

$$\tilde{P}(x) = \alpha_0^n \phi_0^n(x) + \alpha_1^n \phi_1^n(x) + \alpha_2^n \phi_2^n(x), \quad (2.31)$$

para cada  $x \in I_n$ . Usando (2.31) y eligiendo  $v = \phi_j^n(x)$ , para cada  $j = 0, 1, 2$ , se obtiene

$$\int_{I_n} \tilde{P}'(x) (\phi_j^n)'(x) dx = \sum_{i=0}^2 \alpha_i^n \int_{I_n} (\phi_i^n)'(x) (\phi_j^n)'(x) dx,$$

Este sistema lineal puede ser reescrito como  $A_n \alpha^n$ , donde

$$\alpha^n = \begin{pmatrix} \alpha_0^n \\ \alpha_1^n \\ \alpha_2^n \end{pmatrix}, \quad [A_n]_{ij} = \int_{I_n} (\phi_i^n)'(x) (\phi_j^n)'(x) dx.$$

La matriz  $A_n$  es fácil de determinar,

$$A_n = \frac{1}{h} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & \frac{16}{3} \end{pmatrix}.$$

En segundo lugar, se considera los términos que involucran los nodos interiores  $x_n$ . Usando la definición promedio y salto de la función  $v$  para dichos nodos, se tiene

$$-\{\tilde{P}'(x_n)\}[v(x_n)] + \epsilon\{v'(x_n)\}[\tilde{P}(x_n)] + \frac{\sigma^0}{h}[\tilde{P}(x_n)][v(x_n)] = b_n + c_n + d_n + e_n,$$

donde

$$\begin{aligned} b_n &= \frac{1}{2}\tilde{P}'(x_n^+)v(x_n^+) - \frac{\epsilon}{2}\tilde{P}(x_n^+)v'(x_n^+) + \frac{\sigma^0}{h}\tilde{P}(x_n^+)v(x_n^+), \\ c_n &= -\frac{1}{2}\tilde{P}'(x_n^-)v(x_n^-) + \frac{\epsilon}{2}\tilde{P}(x_n^-)v'(x_n^-) + \frac{\sigma^0}{h}\tilde{P}(x_n^-)v(x_n^-), \\ d_n &= -\frac{1}{2}\tilde{P}'(x_n^+)v(x_n^-) - \frac{\epsilon}{2}\tilde{P}(x_n^+)v'(x_n^-) - \frac{\sigma^0}{h}\tilde{P}(x_n^+)v(x_n^-), \\ e_n &= \frac{1}{2}\tilde{P}'(x_n^-)v(x_n^+) + \frac{\epsilon}{2}\tilde{P}(x_n^-)v'(x_n^+) - \frac{\sigma^0}{h}\tilde{P}(x_n^-)v(x_n^+). \end{aligned}$$

Usando (2.31) y con el cambio  $v = \phi_j^n(x)$ , los cuatros términos anteriores determinarán las matrices locales  $B_n$ ,  $C_n$ ,  $D_n$  y  $E_n$ , respectivamente. Las entradas de dichas matrices son

$$\begin{aligned} [B_n]_{ij} &= \frac{1}{2}(\phi_i^n)'(x_n^+)\phi_j^n(x_n^+) - \frac{\epsilon}{2}\phi_i^n(x_n^+)(\phi_j^n)'(x_n^+) + \frac{\sigma^0}{h}\phi_i^n(x_n^+)\phi_j^n(x_n^+), \\ [C_n]_{ij} &= -\frac{1}{2}(\phi_i^n)'(x_n^-)\phi_j^n(x_n^-) + \frac{\epsilon}{2}\phi_i^n(x_n^-)(\phi_j^n)'(x_n^-) + \frac{\sigma^0}{h}\phi_i^n(x_n^-)\phi_j^n(x_n^-), \\ [D_n]_{ij} &= -\frac{1}{2}(\phi_i^n)'(x_n^+)\phi_j^n(x_n^-) - \frac{\epsilon}{2}\phi_i^n(x_n^+)(\phi_j^n)'(x_n^-) - \frac{\sigma^0}{h}\phi_i^n(x_n^+)\phi_j^n(x_n^-), \\ [E_n]_{ij} &= \frac{1}{2}(\phi_i^n)'(x_n^-)\phi_j^n(x_n^+) + \frac{\epsilon}{2}\phi_i^n(x_n^-)(\phi_j^n)'(x_n^+) - \frac{\sigma^0}{h}\phi_i^n(x_n^-)\phi_j^n(x_n^+). \end{aligned}$$

Las entradas  $ij$  de cada matriz local  $3 \times 3$  se puede calcular fácilmente usando (2.28) y (2.29):

$$\begin{aligned} B_n &= \frac{1}{h} \begin{pmatrix} \sigma^0 & 1 - \sigma^0 & -2 + \sigma^0 \\ -\epsilon - \sigma^0 & -1 + \epsilon + \sigma^0 & 2 - \epsilon - \sigma^0 \\ 2\epsilon + \sigma^0 & 1 - 2\epsilon - \sigma^0 & -2 + 2\epsilon - \sigma^0 \end{pmatrix}, \\ C_n &= \frac{1}{h} \begin{pmatrix} \sigma^0 & -1 + \sigma^0 & -2 + \sigma^0 \\ \epsilon + \sigma^0 & -1 + \epsilon + \sigma^0 & -2 + \epsilon + \sigma^0 \\ 2\epsilon + \sigma^0 & -1 + 2\epsilon + \sigma^0 & -2 + 2\epsilon + \sigma^0 \end{pmatrix}, \\ D_n &= \frac{1}{h} \begin{pmatrix} -\sigma^0 & -1 + \sigma^0 & 2 - \sigma^0 \\ -\epsilon - \sigma^0 & -1 + \epsilon + \sigma^0 & 2 - \epsilon - \sigma^0 \\ -2\epsilon - \sigma^0 & -1 + 2\epsilon + \sigma^0 & 2 - 2\epsilon - \sigma^0 \end{pmatrix}, \\ E_n &= \frac{1}{h} \begin{pmatrix} -\sigma^0 & 1 - \sigma^0 & 2 - \sigma^0 \\ \epsilon + \sigma^0 & -1 + \epsilon + \sigma^0 & -2 + \epsilon + \sigma^0 \\ -2\epsilon - \sigma^0 & 1 - 2\epsilon - \sigma^0 & 2 - 2\epsilon - \sigma^0 \end{pmatrix}. \end{aligned}$$



Salvo en el caso en que la función  $f$  sea una función constante a trozos, la forma práctica de calcular la integral anterior será mediante la integración numérica, y, al igual que en el MEF, usaremos la fórmula de cuadratura Gaussiana de dos puntos (Ver [12], Capítulo 5, Sección 5.3, pág. 270). Por lo tanto

$$\int_0^1 f(x)\Phi_j^n(x)dx \approx \frac{h}{2} \sum_{p=1}^2 f\left(\frac{h}{2}s_p + \left(n + \frac{1}{2}\right)h\right)(s_p)^j,$$

donde  $s_1 = -\frac{1}{\sqrt{3}}$  y  $s_2 = \frac{1}{\sqrt{3}}$ . Escribiremos las componentes del vector  $b$  en un orden consecuente con el orden de las incógnitas  $\alpha_i^n$ :

$$\left(b_0^0, b_1^0, b_2^0, b_0^1, b_1^1, b_2^1, b_0^2, b_1^2, b_2^2, \dots, b_0^{N-1}, b_1^{N-1}, b_2^{N-1}\right),$$

donde las tres primeras componentes vienen dadas por

$$b_0^0 = \frac{h}{2} \sum_{p=1}^2 f\left(\frac{h}{2}s_p + \frac{h}{2}\right) + \frac{\sigma^0}{h},$$

$$b_1^0 = \frac{h}{2} \sum_{p=1}^2 f\left(\frac{h}{2}s_p + \frac{h}{2}\right)s_p - \epsilon \frac{2}{h} - \frac{\sigma^0}{h},$$

$$b_2^0 = \frac{h}{2} \sum_{p=1}^2 f\left(\frac{h}{2}s_p + \frac{h}{2}\right)(s_p)^2 + \epsilon \frac{4}{h} + \frac{\sigma^0}{h},$$

mientras que el resto de las  $3(N-1)$  componentes son

$$b_i^n = \frac{h}{2} \sum_{p=1}^2 f\left(\frac{h}{2}s_p + \left(n + \frac{1}{2}\right)h\right)(s_p)^i.$$

A continuación se presenta un breve algoritmo de este método, el cual emplea las ideas básicas anteriormente descritas.

### Algoritmo del Método Galerkin Discontinuo.

**Entrada:** Número de subintervalos  $N \geq 2$ ; parámetros  $\epsilon$  y  $\sigma^0$ .

**Salida:** Vector de incógnitas  $\alpha_i^n$ ,  $i = 0, 1, 2$ ;  $n = 0, \dots, N$ .

1. Tomar  $h = 1/N$ .
2. Construir las matrices locales  $A_n, B_n, C_n, D_n, E_n, F_0, F_N$ .
3. Ensamblar las matrices locales dentro de la matriz global  $A$ .
4. Aplicar cuadratura Gaussiana de dos puntos para calcular y construir el vector  $b$ .
5. Tomar  $A^{-1}b$ .

Para efecto de la implementación de los códigos, se trabajará con SIPG, esto es, cuando  $\epsilon = -1$  y  $\sigma^0 = 10^6$ . A raíz de esto, la solución encontrada por este método aproxima la solución exacta con un orden de aproximación  $\mathcal{O}(h^3)$  (ver [1], pág. 13).

## 2.5. Método Bvp4c

Hasta los momentos se han planteado cuatro métodos para aproximar la solución de un Problema de Valor de Frontera de segundo orden: los métodos de Disparo y Diferencias Finitas son los métodos comúnmente más usados, con frecuencia aparecen en todas las referencias básicas de Análisis Numérico, mientras que los métodos de Elementos Finitos y Galerkin Discontinuo, un poco más complejos que los anteriores, parten de la teoría de espacios normados y se fundamentan en resultados del Análisis Funcional. En esta sección se planteará el último de los métodos numéricos que se empleará para la resolución del problema planteado: el Método Bvp4c, presentado en [14], el cual se encuentra implementado en MATLAB.

El Método Bvp4c pone en práctica el principio de superposición para aproximar la solución del Problema de Valor de Frontera de segundo orden:

$$\begin{cases} y'' &= f(x, y, y'), \\ y(a) &= \alpha, \\ y(b) &= \beta. \end{cases} \quad (2.32)$$

Para conseguir esto, se reescribe (2.32) como un sistema de ecuaciones diferenciales ordinarias de primer orden:

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad a \leq x \leq b, \quad (2.33)$$

sujeta a condiciones generales de frontera no lineales:

$$\mathbf{g}(\mathbf{y}(a), \mathbf{y}(b)) = \mathbf{0}.$$

La solución aproximada del sistema (2.33),  $\mathbf{S}(x)$ , es una función continua. Más en general, la solución  $\mathbf{S}(x)$  es un polinomio cúbico en cada subintervalo  $I_n = [x_n, x_{n+1}]$ , de una malla de puntos  $a = x_0 < x_1 < \dots < x_N = b$ . Esta solución aproximada satisface las condiciones de frontera que cumple  $\mathbf{y}$ , es decir:

$$\mathbf{g}(\mathbf{S}(a), \mathbf{S}(b)) = \mathbf{0}.$$

Así mismo, la solución  $\mathbf{S}(x)$  satisface la ecuación diferencial en los extremos y punto medio de cada subintervalo  $I_n$ , esto es,

$$\begin{aligned} \mathbf{S}'(x_n) &= \mathbf{f}(x_n, \mathbf{S}(x_n)), \\ \mathbf{S}'\left(\frac{x_n + x_{n+1}}{2}\right) &= \mathbf{f}\left(\frac{x_n + x_{n+1}}{2}, \mathbf{S}\left(\frac{x_n + x_{n+1}}{2}\right)\right), \\ \mathbf{S}'(x_{n+1}) &= \mathbf{f}(x_{n+1}, \mathbf{S}(x_{n+1})). \end{aligned}$$

Estas ecuaciones conllevan a un sistema de ecuaciones algebraicas no lineales para los coeficientes de  $\mathbf{S}(x)$ . A diferencia del método del Disparo para el caso no lineal, la solución  $\mathbf{y}(x)$  es aproximada sobre todo el intervalo  $[a, b]$  y las condiciones de frontera son tomadas en cuenta todo el tiempo. Las ecuaciones algebraicas no lineales son resueltas iterativamente por linealización. La solución  $\mathbf{S}(x)$  es una aproximación de cuarto orden para  $\mathbf{y}(x)$ , es decir,  $\|\mathbf{y}(x) - \mathbf{S}(x)\| \leq Ch^4$  (ver [11]). Aquí,  $h$  es el máximo tamaño de paso  $h_n = x_{n+1} - x_n$  y  $C$  una constante.

Este método adapta la malla para obtener así una solución numérica acertada con un modesto número de puntos de dicha malla. Poder hacer una aproximación suficientemente buena es la parte más difícil al momento de resolver un PVF. La continuidad de  $\mathbf{S}(x)$  en  $[a, b]$  y la superposición en los extremos y punto medio de cada intervalo implica que la derivada de  $\mathbf{S}(x)$  también es continua en  $[a, b]$ .

El método Bvp4c controla el error cometido al momento de aproximar  $\mathbf{y}(x)$  mediante  $\mathbf{S}(x)$ . Para el control de dicho error, se introduce el residuo o resto de la ecuación diferencial. Este viene dado por

$$\mathbf{r}(x) = \mathbf{S}'(x) - \mathbf{f}(x, \mathbf{S}(x)).$$

Como se puede apreciar, en los extremos y punto medio de cada intervalo  $I_n$ , el resto es cero.

**Sintaxis.** Como se dijo, el método `Bvp4c` es un método que usa la idea de superposición y el mismo se encuentra implementado en MATLAB, el cual permite resolver numéricamente un PVF del tipo

$$y''(x) = f(x, y(x), y'(x)), \quad x \in [a, b],$$

sujeta a

$$g(y(a), y(b)) = 0.$$

La sintaxis básica, empleando MATLAB, es la siguiente:

```
sol = bvp4c(@ odefun, @ bcfun, solinit),
```

donde,

1. `@odefun` es el nombre del fichero de la ecuación diferencial ordinaria que se quiere resolver, expresado como un sistema de ecuaciones diferenciales ordinarias de primer orden,

$$dv = h(x, v).$$

Aquí  $x$  es un escalar, y además,

$$\begin{aligned} v &= [v_1 \ v_2]^T = [y(x) \ y'(x)]^T, \\ dv &= [v'_1 \ v'_2]^T = [y'(x), f(x, y(x), y'(x))]^T. \end{aligned}$$

2. `@bcfun` es el nombre del fichero que contiene las condiciones de contorno.
3. `solinit` es el dato inicial de la ecuación diferencial en un conjunto de nodos ordenados  $a = x_0 < x_1 < \dots < x_N = b$ . Para ello se usa la función `bvpinit`:

```
solinit = bvpinit(xmesh, yinit),
```

para una malla inicial.

La salida `sol` es una estructura con

- `sol.x`, que es el mallado seleccionado por `Bvp4c`,
- `sol.y`, aproximación a  $y(x)$  en los puntos del mallado `sol.x`,
- `sol.yp`, aproximación a  $y'(x)$  en los puntos del mallado `sol.x`.

Una vez presentados y analizados los métodos para Problemas de Valor de Frontera, en este capítulo nos centraremos en el análisis y comparación desde un punto de vista experimental. Para esto, primero que todo, se implementa y valida cada método en MATLAB. No obstante, por considerarlo irrelevante, los códigos no son dados explícitamente en este trabajo. Un conjunto de ejemplos, lineales y no lineales, es considerado para la validación y comparación de los métodos. Sin embargo, debido a las características que presenta cada método, la implementación para problemas no lineales no se llevará a cabo para el Método de Elementos Finitos (MEF) y el Método de Galerkin Discontinuo.

Debido a que no todas las versiones de los métodos serán codificadas, la comparación será realizada sólo parcialmente y, en algunos casos, se proyectarán conclusiones de los casos estudiados a los casos más generales. Por ejemplo, el método `Bvp4c`, se encuentra implementado usando un proceso adaptativo (control del error) en MATLAB (ver [14]). Por lo cual, resultará más apropiado en cuanto al tamaño del paso a tomar. Sin embargo, proyectaremos resultados de otros que no tienen implementado un proceso adaptativo sin que para esto tomemos en cuenta el tamaño de paso óptimo. Otro caso, es en cuanto a la linealidad o no de los ejemplos a tratar, pues como ya dijimos, el método de Elementos Finitos y el método de Galerkin Discontinuo no se implementarán para el caso no lineal, y su comparación con los otros métodos sólo se hará en ejemplos lineales.

El error cometido en cada uno de los métodos se medirá usando las normas del máximo,  $\|\cdot\|_\infty$ , y la norma  $L_2$ ,  $\|\cdot\|_{L_2}$ . Esto es, si  $x$  es un vector de  $n$  componentes, entonces

$$\|x\|_\infty = \text{máx}\{|x_1|, |x_2|, \dots, |x_n|\},$$

Por otro lado, la norma  $L_2$  queda dada por

$$\|e_h(x)\|_{L_2} = \left( \int_a^b (e_h(x))^2 dx \right)^{1/2},$$

con  $e_h(x) = u(x) - \tilde{u}(x)$ ,  $u(x)$  la solución exacta del PVF planteado y  $\tilde{u}(x)$  su aproximación. En el caso en que se emplee el método Galerkin Discontinuo,  $\tilde{u}(x)$  es un polinomio cuadrático en cada subintervalo de la partición, para el caso del MEF es un polinomio lineal, y en todos los otros casos la solución es puntual en cada subintervalo de dicha partición (constante en cada subintervalo). Además, en muchas ocasiones estaremos interesados en tener conocimiento sobre costo de error y costo computacional (tiempo de CPU usado) que ocurre al momento de usar uno de los métodos anteriormente mencionados, es decir, nos interesa saber qué tan grande es la magnitud del error que se comete cuando se aproxima la solución y el tiempo de máquina que se requiere para realizar tal aproximación.

En lo que sigue, se ponen a prueba los códigos de los métodos numéricos sobre un conjunto de siete ejemplos. Los cuatro primeros son del tipo lineal, lo que permite emplear los cinco métodos, mientras que los tres ejemplos restantes son del tipo no lineal, lo que conlleva a usar sólo los métodos del Disparo, de Diferencias Finitas y `Bvp4c`. En vista de que los métodos del Disparo y de Diferencias Finitas, para el caso no lineal, usan un método iterativo para encontrar el cero de una función no lineal y resolver un sistema de ecuaciones no lineales,

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 10$	Disparo	0.041	$1.915 \times 10^{-6}$	$1.115 \times 10^{-3}$
	Dif. Finitas	0.023	$1.203 \times 10^{-3}$	$8.653 \times 10^{-4}$
	Elementos Finitos	0.061	$1.277 \times 10^{-6}$	$1.114 \times 10^{-3}$
	G. Discontinuo	0.115	$1.277 \times 10^{-6}$	$3.796 \times 10^{-5}$
	Bvp4c	0.987	$1.915 \times 10^{-6}$	$1.115 \times 10^{-3}$
$N = 50$	Disparo	0.045	$3.042 \times 10^{-9}$	$4.472 \times 10^{-5}$
	Dif. Finitas	0.024	$4.845 \times 10^{-5}$	$3.455 \times 10^{-5}$
	Elementos Finitos	0.084	$2.028 \times 10^{-9}$	$4.472 \times 10^{-5}$
	G. Discontinuo	0.141	$1.755 \times 10^{-9}$	$3.041 \times 10^{-7}$
	Bvp4c	1.018	$3.042 \times 10^{-9}$	$4.472 \times 10^{-5}$
$N = 100$	Disparo	0.049	$1.902 \times 10^{-10}$	$1.118 \times 10^{-5}$
	Dif. Finitas	0.027	$1.211 \times 10^{-5}$	$8.637 \times 10^{-6}$
	Elementos Finitos	0.103	$1.268 \times 10^{-10}$	$1.118 \times 10^{-5}$
	G. Discontinuo	0.151	$9.361 \times 10^{-9}$	$3.846 \times 10^{-8}$
	Bvp4c	1.075	$1.902 \times 10^{-10}$	$1.118 \times 10^{-5}$

Cuadro 3.1: Ejemplo 1. Tiempo empleado y errores obtenidos para mallas de 10, 50 y 100 subintervalos.

respectivamente, estos se ejecutarán usando una tolerancia de  $10^{-3}$ . De la misma manera, como el método Bvp4c realiza proceso adaptativo, este se ejecutará usando una tolerancia también de  $10^{-3}$ . Para el método de Galerkin Discontinuo, se trabajará en el caso particular en que  $\varepsilon = -1$  y  $\sigma^0 = 10^6$ , el cual se definió como SIPG (ver sección 2.4 del capítulo anterior). Se estimará el error mediante la norma del máximo  $\|\cdot\|_\infty$  y la norma  $\|\cdot\|_{L_2}$ , así como también el tiempo de máquina, en segundos, requerido para realizar las distintas operaciones.

Aunque la idea es comparar los errores alcanzados por los métodos propuestos, por efecto de visualización los presentaremos por separados. Pues se quiere ver el comportamiento del error de cada método, y por efecto de escala estos se solapan y sólo uno puede visualizarse, quedando los otros superpuestos en el eje coordenado. Así pues, para el análisis, se deben observar las gráficas de forma global y no por separado.

### Ejemplo 1

Consideremos el siguiente problema de valor de frontera lineal:

$$\begin{aligned} y'' &= -(4x^3 - 4x^2 - 6x + 2)e^{-x^2}, \\ y(0) &= 1, \\ y(1) &= 0, \end{aligned}$$

con solución

$$y(x) = (1 - x)e^{-x^2}.$$

La gráfica de la solución exacta de este problema lineal se muestra en la Figura 3.1. Los resultados para una partición del intervalo  $[0, 1]$  en 10, 50 y 100 subintervalos de igual longitud son mostrados en el Cuadro 3.1. Para el caso del método de Diferencias Finitas, los resultados son menos exactos que los restantes métodos, esto se debe a que presenta un error del orden  $\mathcal{O}(h^2)$ , cosa que no sucede, por citar, con el método del Disparo o Bvp4c, que presentan un error del orden  $\mathcal{O}(h^4)$ .

El comportamiento del error relativo al dividir el intervalo  $[0, 1]$  en 50 subintervalos de igual longitud se ilustra en la Figura 3.1. Se puede apreciar en las gráficas como, para los métodos del Disparo, Diferencias Finitas, Elementos Finitos y Bvp4c, el error crece pero mantiene su tolerancia dentro de parámetros aceptables (el error en el punto final fue omitido debido a que en este la solución exacta es cero). Para el método Galerkin

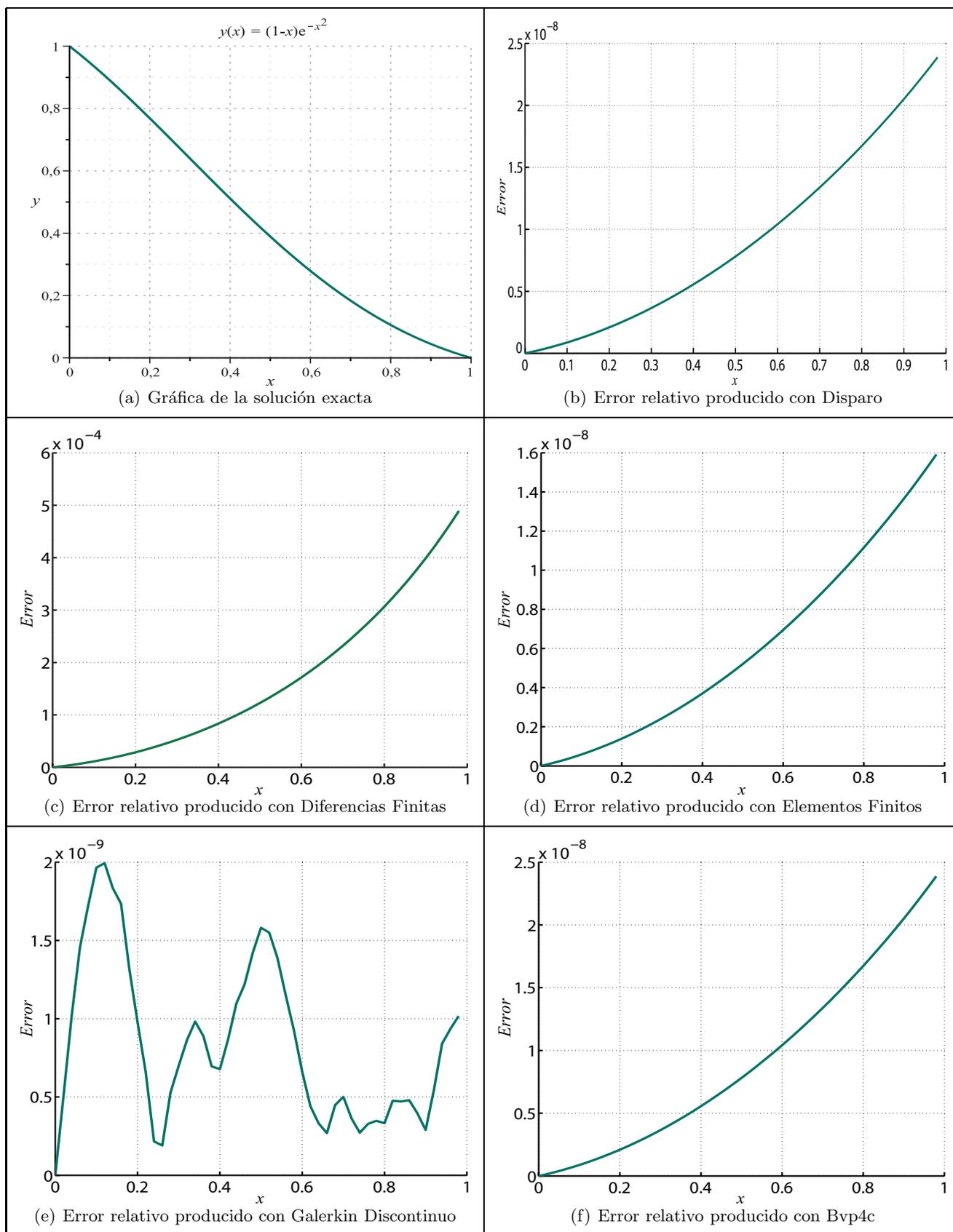


Figura 3.1: Ejemplo 1. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas, Elementos Finitos, Galerkin Discontinuo y Bvp4c, para un mallado de 50 subintervalos, empleando la norma del máximo.

Discontinuo, el error se comporta de manera oscilatoria pero estable y con mejor precisión que los otros métodos. El método Galerkin Discontinuo logra dar una mejor aproximación a la solución del PVF empleando la norma  $L_2$ , por encima de los otros cuatro métodos. Este hecho es debido a que el método Galerkin Discontinuo emplea polinomios de interpolación de grado 2, a diferencia de los otros métodos cuya interpolación se hace mediante polinomios lineales.

Para cada una de las tres mallas que se usan en el Cuadro 3.1, el método Bvp4c no realiza proceso de adaptatividad, ya que en cada subintervalo de esas tres mallas, la norma del residuo, que resulta de aproximar la solución mediante polinomios cúbicos, está por debajo de la tolerancia dada. Tomando como referencia la malla de 10 subintervalos de igual longitud, se tiene a partir de los resultados mostrados en el Cuadro 3.1, que el método Bvp4c emplea 0.987 segundos para obtener la aproximación con un error máximo de  $1.915 \times 10^{-6}$ . Para que los restantes cuatro métodos consuman un tiempo relativamente igual, necesitan emplear el número de intervalos que se expresan en el Cuadro 3.2.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3554	$4.796 \times 10^{-14}$
Dif. Finitas	5559	$3.918 \times 10^{-9}$
Elementos Finitos	1108	$3.221 \times 10^{-12}$
G. Discontinuo	514	$3.160 \times 10^{-7}$

Cuadro 3.2: Ejemplo 1. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 0.987 segundos (tiempo empleado por Bvp4c).

Como resultado se tiene, empleando el Cuadro 3.1 y el Cuadro 3.2, que tanto el método del Disparo como Diferencias Finitas requieren una cantidad considerable de intervalos para usar el tiempo empleado por Bvp4c para 10 subintervalos, esto es debido a que el costo computacional de cada subintervalo es más económico para estos métodos que para los restantes métodos. Debido a esto, los errores empleando la norma del máximo (o la norma  $L_2$ ) disminuyen considerablemente. Por otro lado, y a diferencia del método del Disparo y Diferencias Finitas, tanto el método de Elementos Finitos como el método Galerkin Discontinuo requieren menos intervalos para obtener el tiempo de prueba (la evaluación en cada subintervalo es más costosa) obteniendo buenas aproximaciones en los dos métodos. Como es de apreciar, el método de Elementos Finitos aproxima mejor la solución en menos intervalos en comparación con Diferencias Finitas.

Los resultados del Cuadro 3.2 permiten concluir, que cualquiera de los métodos sin proceso adaptativo implementado logra superar la precisión del método Bvp4c usando el mismo tiempo de cálculo. En otras palabras, el proceso adaptativo sólo permite tener control del error y no mejora el tiempo de calculo al utilizar menos nodos de discretización a lo largo del cálculo.

## Ejemplo 2

Trabajemos ahora con el siguiente problema de valor de frontera:

$$\begin{aligned} y'' &= -27x + \frac{28}{3}, \\ y(0) &= 1, \\ y(1) &= 0, \end{aligned}$$

el cual tiene por solución exacta

$$y(x) = -\frac{9}{2}x^3 + \frac{14}{3}x^2 - \frac{7}{6}x + 1.$$

La gráfica de la solución exacta  $y(x)$  se refleja en la Figura 3.2. Los resultados para una partición del intervalo  $[0, 1]$  en 10, 50 y 100 subintervalos de igual longitud son mostrados en el Cuadro 3.3. Se logra apreciar que los errores, empleando la norma del máximo, son muy buenos en todos los métodos, excepto en el método GD. En este caso, aunque el error resulta apropiado, existe una diferencia significativa con los otros métodos. En

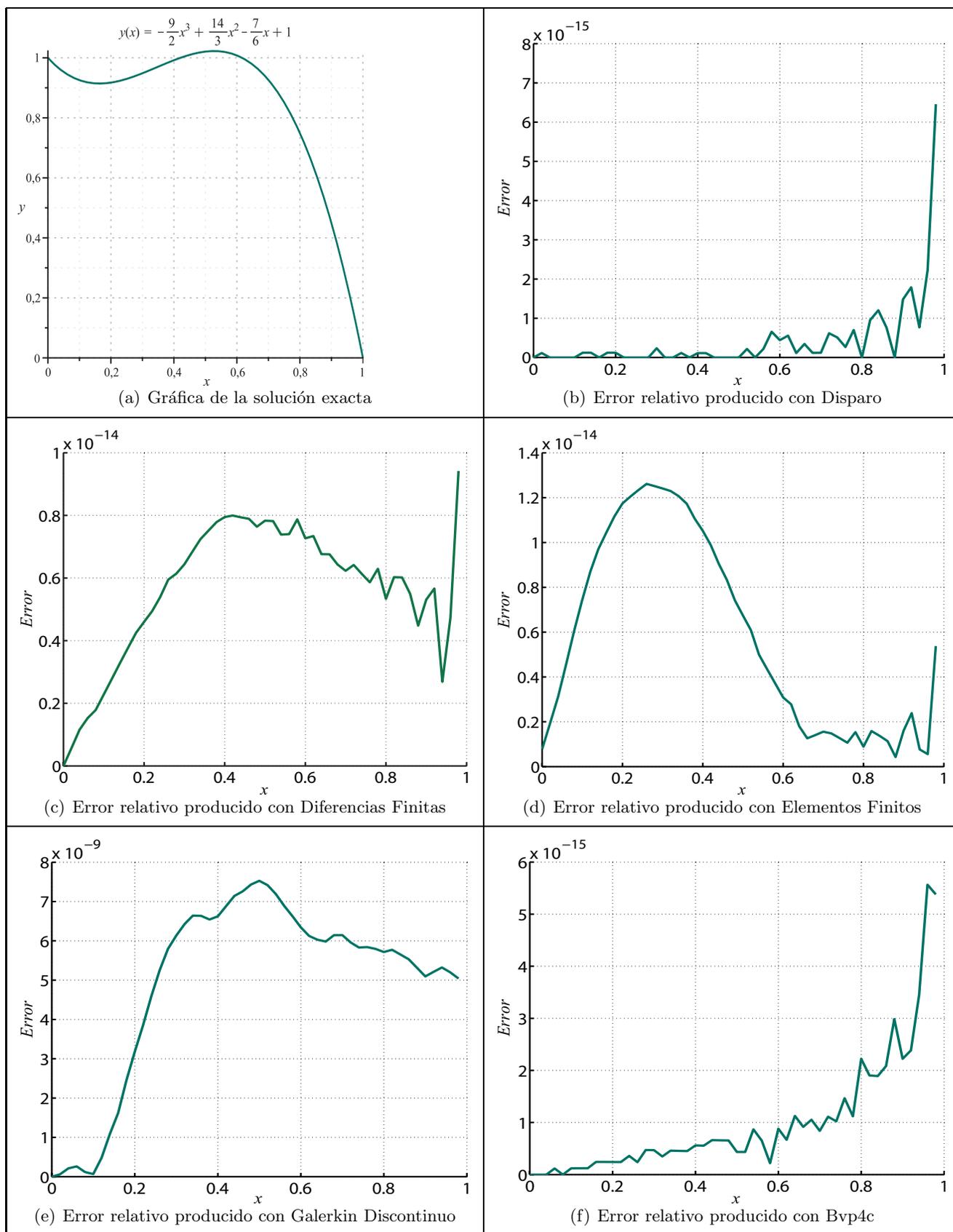


Figura 3.2: Ejemplo 2. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas, Elementos Finitos, Galerkin Discontinuo y Bvp4c, para una malla de 50 subintervalos, empleando la norma del máximo.

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 10$	Disparo	0.043	$4.441 \times 10^{-16}$	$7.340 \times 10^{-3}$
	Dif. Finitas	0.021	$5.551 \times 10^{-16}$	$7.340 \times 10^{-3}$
	Elementos Finitos	0.062	$8.882 \times 10^{-16}$	$7.340 \times 10^{-3}$
	G. Discontinuo	0.115	$1.788 \times 10^{-9}$	$2.165 \times 10^{-4}$
	Bvp4c	0.979	$7.772 \times 10^{-16}$	$7.340 \times 10^{-3}$
$N = 50$	Disparo	0.046	$7.772 \times 10^{-16}$	$2.946 \times 10^{-4}$
	Dif. Finitas	0.024	$7.994 \times 10^{-15}$	$2.946 \times 10^{-4}$
	Elementos Finitos	0.080	$1.177 \times 10^{-14}$	$2.946 \times 10^{-4}$
	G. Discontinuo	0.127	$7.703 \times 10^{-9}$	$1.732 \times 10^{-6}$
	Bvp4c	1.028	$1.665 \times 10^{-15}$	$2.946 \times 10^{-4}$
$N = 100$	Disparo	0.049	$6.661 \times 10^{-16}$	$7.365 \times 10^{-5}$
	Dif. Finitas	0.026	$1.088 \times 10^{-14}$	$7.365 \times 10^{-5}$
	Elementos Finitos	0.104	$1.410 \times 10^{-14}$	$7.365 \times 10^{-5}$
	G. Discontinuo	0.148	$2.195 \times 10^{-8}$	$2.170 \times 10^{-7}$
	Bvp4c	1.064	$2.442 \times 10^{-15}$	$7.365 \times 10^{-5}$

Cuadro 3.3: Ejemplo 2. Tiempo empleado y errores obtenidos para un mallado de 10, 50 y 100 subintervalos.

norma infinita, el método del Disparo logra obtener una mejor aproximación que los restantes métodos, a pesar que emplea casi el doble del tiempo que el método de Diferencias Finitas, que resulta ser el más rápido. Esto es previsible debido a que el método del Disparo emplea el método de Runge-Kutta de cuarto orden, el cual ofrece una exactitud  $\mathcal{O}(h^4)$  a las soluciones de los problemas de valor inicial. Mientras, para el error empleando norma  $L_2$ , el método Galerkin Discontinuo logra obtener una mejor aproximación sobre los otros métodos, en vista de que este emplea polinomios de grado 2 (como ya se discutió anteriormente).

Para un tamaño de paso de  $h = 0.02$ , el comportamiento del error relativo cometido por cada uno de los métodos se ilustra en la Figura 3.2. Es importante resaltar nuevamente, que el método Bvp4c no realiza proceso de adaptatividad en las tres mallas iniciales dadas, ya que en cada subintervalo de cada una de las tres particiones, la norma del residuo, que resulta de aproximar la solución mediante polinomios cúbicos, está por debajo de la tolerancia dada. Si se toma como referencia la malla de 10 subintervalos de igual longitud, entonces, en vista del Cuadro 3.3, Bvp4c emplea 0.979 segundos para obtener la aproximación con un error máximo de  $7.772 \times 10^{-16}$ . Para que los restantes cuatro métodos requieran un tiempo relativamente igual, necesitan emplear el número de intervalos que vienen dados en el Cuadro 3.4.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3610	$6.151 \times 10^{-14}$
Dif. Finitas	5635	$5.358 \times 10^{-12}$
Elementos Finitos	1100	$7.480 \times 10^{-13}$
G. Discontinuo	524	$1.570 \times 10^{-7}$

Cuadro 3.4: Ejemplo 2. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 0.979 segundos.

Nuevamente, se puede apreciar del Cuadro 3.4, que los métodos pueden trabajar con particiones muy finas utilizando la misma cantidad de tiempo que el empleado por el método Bvp4c. Sin embargo, en este ejemplo, debido a la precisión alcanzada para 10 intervalos, los métodos no logran superar este error al aumentar el número de intervalos. Llegando incluso a perder precisión debido a la inestabilidad numérica del proceso de diferenciación.

## Ejemplo 3

Consideremos el problema de valor de frontera lineal:

$$\begin{aligned} y'' &= -4 \frac{\sin(2x) + 4 \cos^2(x) - 4x \cos^2(x) - 5 + 5x}{\cos^6(x)}, \\ y(0) &= 1, \\ y(1) &= 0, \end{aligned}$$

con solución exacta

$$y(x) = \frac{1-x}{\cos^4(x)}.$$

La gráfica de la solución exacta  $y(x)$  viene dada en la Figura 3.3. En el Cuadro 3.5, se reflejan los datos obtenidos para una partición uniforme de  $[0, 1]$  en 10, 50 y 100 subintervalos. Para la malla inicial de 10

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 10$	Disparo	0.043	$1.706 \times 10^{-3}$	$2.340 \times 10^{-2}$
	Dif. Finitas	0.021	$7.850 \times 10^{-2}$	$6.295 \times 10^{-2}$
	Elementos Finitos	0.059	$1.128 \times 10^{-3}$	$2.274 \times 10^{-2}$
	G. Discontinuo	0.116	$1.128 \times 10^{-3}$	$2.877 \times 10^{-3}$
$N = 15$	Bvp4c	1.118	$1.120 \times 10^{-4}$	$6.232 \times 10^{-3}$
$N = 50$	Disparo	0.045	$3.073 \times 10^{-6}$	$9.846 \times 10^{-4}$
	Dif. Finitas	0.024	$3.407 \times 10^{-3}$	$2.765 \times 10^{-3}$
	Elementos Finitos	0.077	$2.048 \times 10^{-6}$	$9.833 \times 10^{-4}$
	G. Discontinuo	0.129	$2.052 \times 10^{-6}$	$2.321 \times 10^{-5}$
	Bvp4c	1.045	$3.073 \times 10^{-6}$	$9.846 \times 10^{-4}$
$N = 100$	Disparo	0.049	$1.928 \times 10^{-7}$	$2.466 \times 10^{-4}$
	Dif. Finitas	0.027	$8.543 \times 10^{-4}$	$6.933 \times 10^{-5}$
	Elementos Finitos	0.102	$1.285 \times 10^{-7}$	$2.465 \times 10^{-4}$
	G. Discontinuo	0.148	$1.159 \times 10^{-7}$	$2.901 \times 10^{-6}$
	Bvp4c	1.070	$1.928 \times 10^{-7}$	$2.466 \times 10^{-4}$

Cuadro 3.5: Ejemplo 3. Tiempo empleado y errores obtenidos para mallas de 10, 50 y 100 subintervalos. Para la malla inicial de 10 subintervalos, el método Bvp4c adapta la malla con la finalidad de mejorar la estimación de la solución, obteniendo así una malla final de 15 subintervalos.

subintervalos, el método Bvp4c adapta la malla con la finalidad de que el residuo satisfaga la tolerancia dada, obteniendo así una nueva malla de 15 subintervalos (no todos de la misma longitud). Al igual que los dos ejemplos anteriores, las aproximaciones arrojadas por cada método son relativamente óptimas, haciendo énfasis, nuevamente, en que el método de Diferencias Finitas proporciona errores un poco mayor que los dados por los otros métodos. Esto, en parte, al bajo orden de aproximación ( $\mathcal{O}(h^2)$ ) que presenta dicho método. Es importante resaltar que el proceso de adaptatividad por parte de Bvp4c conlleva a que se realicen más operaciones con el fin de mejorar los resultados. Es por eso, por ejemplo, que el tiempo empleado al usar la malla inicial de 10 subintervalos es mucho mayor que el tiempo usado en la malla de 50 o 100 subintervalos. De nuevo, queda de manifiesto que el proceso adaptativo controla el error y no el tiempo de cálculo.

Para el caso particular en que la malla consta de 51 puntos distribuidos uniformemente, el comportamiento del error relativo cometido se refleja en la Figura 3.3. Aquí se puede notar como los errores, para cada uno de los cinco métodos, se comportan de manera similar. En esta malla, el método Bvp4c usa un tiempo de 1.045 segundos para realizar las operaciones involucradas en la aproximación, cometiendo un error máximo de  $3.073 \times 10^{-6}$ . Para tener un tiempo aproximadamente igual, los cuatros métodos restantes necesitarán aproximar

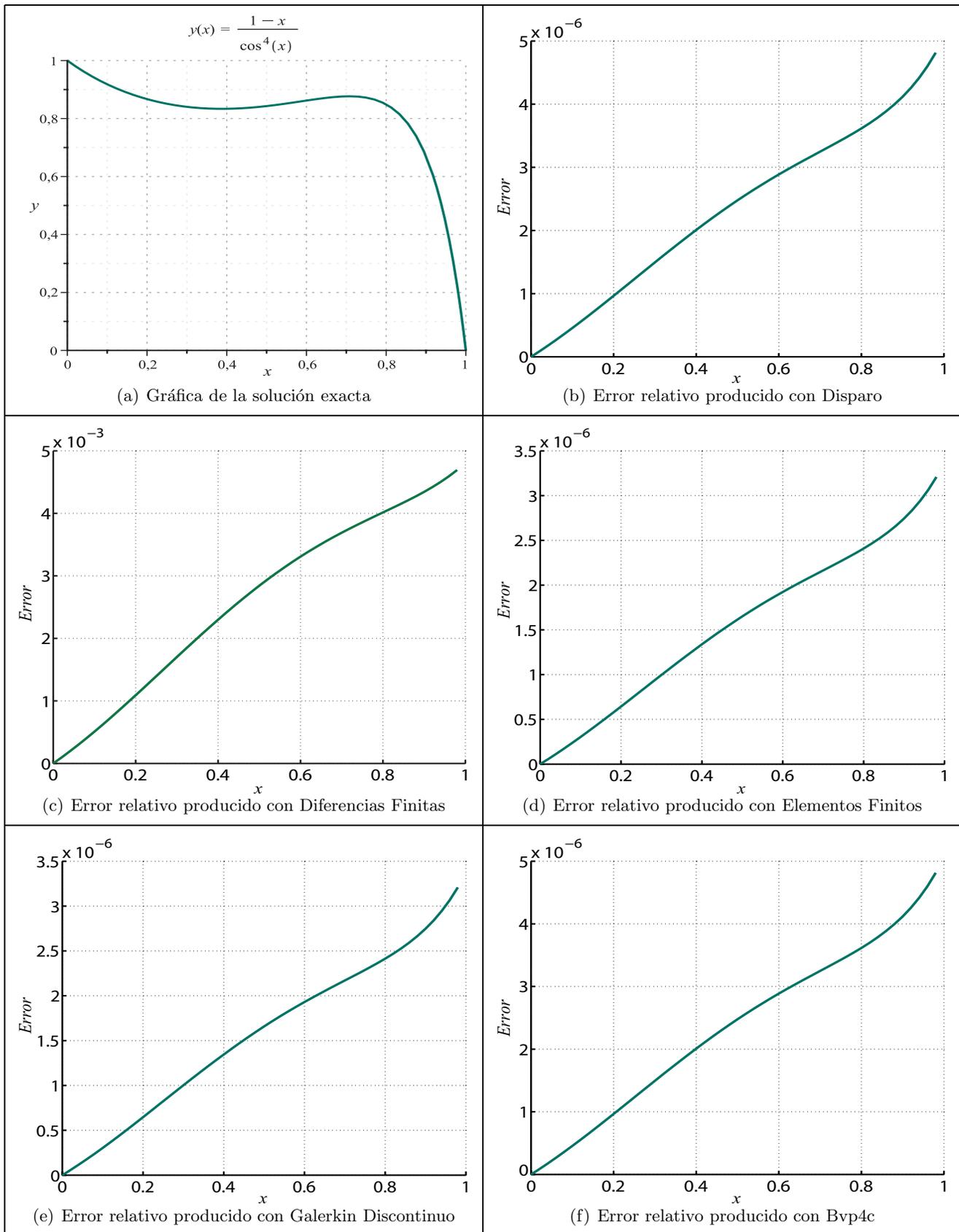


Figura 3.3: Ejemplo 3. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas, Elementos Finitos, Galerkin Discontinuo y Bvp4c, para un mallado de 50 subintervalos, empleando la norma del máximo.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3654	$1.020 \times 10^{-13}$
Dif. Finitas	5801	$2.542 \times 10^{-7}$
Elementos Finitos	1109	$8.607 \times 10^{-12}$
G. Discontinuo	533	$2.698 \times 10^{-7}$

Cuadro 3.6: Ejemplo 3. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 1.045 segundos.

la solución sobre el número de intervalos que se reflejan en el Cuadro 3.6. Al igual que en el Ejemplo 1, se obtiene, en virtud del Cuadro 3.5 y del Cuadro 3.6, que los errores disminuyen a medida que aumentan los intervalos de la partición. Tanto el método del Disparo como Elementos Finitos aportan errores relativamente iguales, con la diferencia de que Elementos Finitos requiere menos intervalos para aproximar la solución. De la misma manera, tanto Diferencias Finitas como Galerkin Discontinuo obtienen errores similares, pero se diferencian en el hecho de que Galerkin Discontinuo necesita menos de la décima parte del total de intervalos que necesita Diferencias Finitas.

## Ejemplo 4

Trabajemos ahora con la siguiente ecuación diferencial:

$$\begin{aligned} y'' &= \frac{-12(2x-1)p}{(p+(2x-1)^2)^{5/2}}, \\ y(0) &= 1, \\ y(1) &= 0, \end{aligned}$$

el cual tiene por solución exacta

$$y(x) = \frac{2x-1}{\sqrt{p+(2x-1)^2}} - \frac{1}{2} \frac{(2\sqrt{p+1}+p+1)(2x-1)}{p+1} + \frac{1}{2},$$

donde  $p = 10^{-3}$ . En la Figura 3.4, se ilustra la gráfica de la solución  $y(x)$ .

Para una partición uniforme del intervalo  $[0, 1]$  en 100, 200 y 500 subintervalos, los errores y tiempo requerido por cada método se expresan en el Cuadro 3.7. Es importante mencionar que para la malla inicial que consta de 101 puntos distribuidos uniformemente, el método `Bvp4c` realiza proceso adaptativo, para obtener así una malla final que consta de 65 puntos, mientras que para la malla de 201 puntos, este método realiza nuevamente proceso adaptativo para obtener una malla constituido por 139 puntos, no necesariamente distribuidos uniformemente.

En la Figura 3.4 y Figura 3.5 se muestran las gráficas de los distintos errores absolutos y relativos al emplear cada uno de los cinco métodos numéricos, trabajando con una malla de 200 subintervalos. Como se dijo anteriormente, para esta malla el método `Bvp4c` realiza adaptatividad con la finalidad de controlar el error (ver Figura 3.4 (f) donde el error alcanza la tolerancia impuesta al error máximo). Para ello crea una malla final que va a constar de 139 nodos. Para realizar dicha proceso adaptativo, este método emplea 1.669 segundos, cometiendo un error máximo de  $9.849 \times 10^{-4}$ . Para que los restantes cuatro métodos consuman un tiempo similar en el cálculo, necesitarán usar aproximadamente el número de subintervalos que se ilustran en el Cuadro 3.8.

Los resultados mostrados en la Figura 3.5 permiten ver el buen funcionamiento de los métodos: Disparo, Elementos Finitos y Galerkin Discontinuo. El método `Bvp4c` realiza el proceso adaptativo llevando el control del error absoluto y no del error relativo. Por tal motivo, el error relativo falla para este método, llegando incluso a alcanzar errores de hasta un 20% en los puntos de mayor pendiente de la solución.

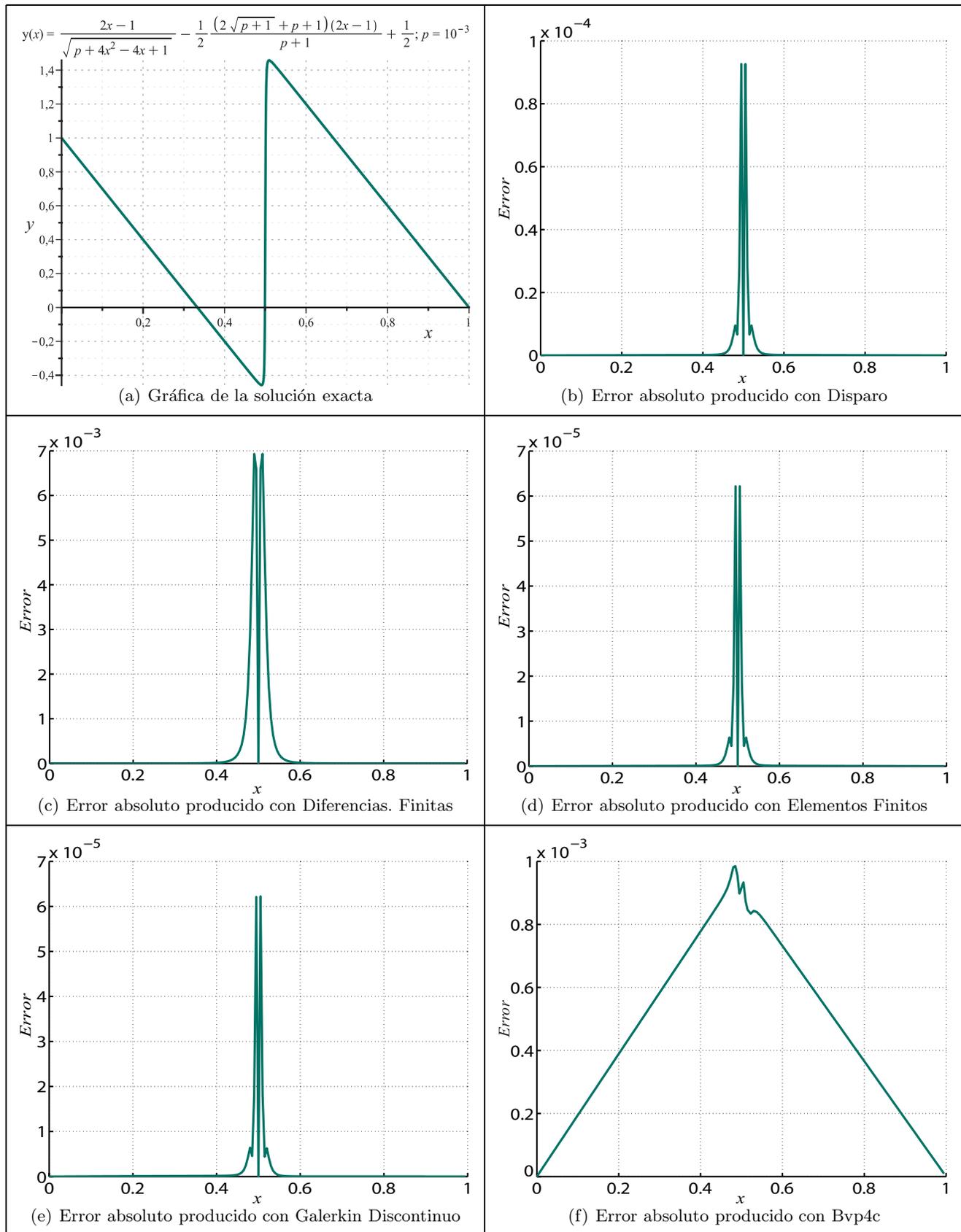


Figura 3.4: Ejemplo 4. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error absoluto) obtenidos con Disparo, Diferencias Finitas, Elementos Finitos, Galerkin Discontinuo y Bvp4c, para una malla de 200 subintervalos, empleando la norma del máximo.

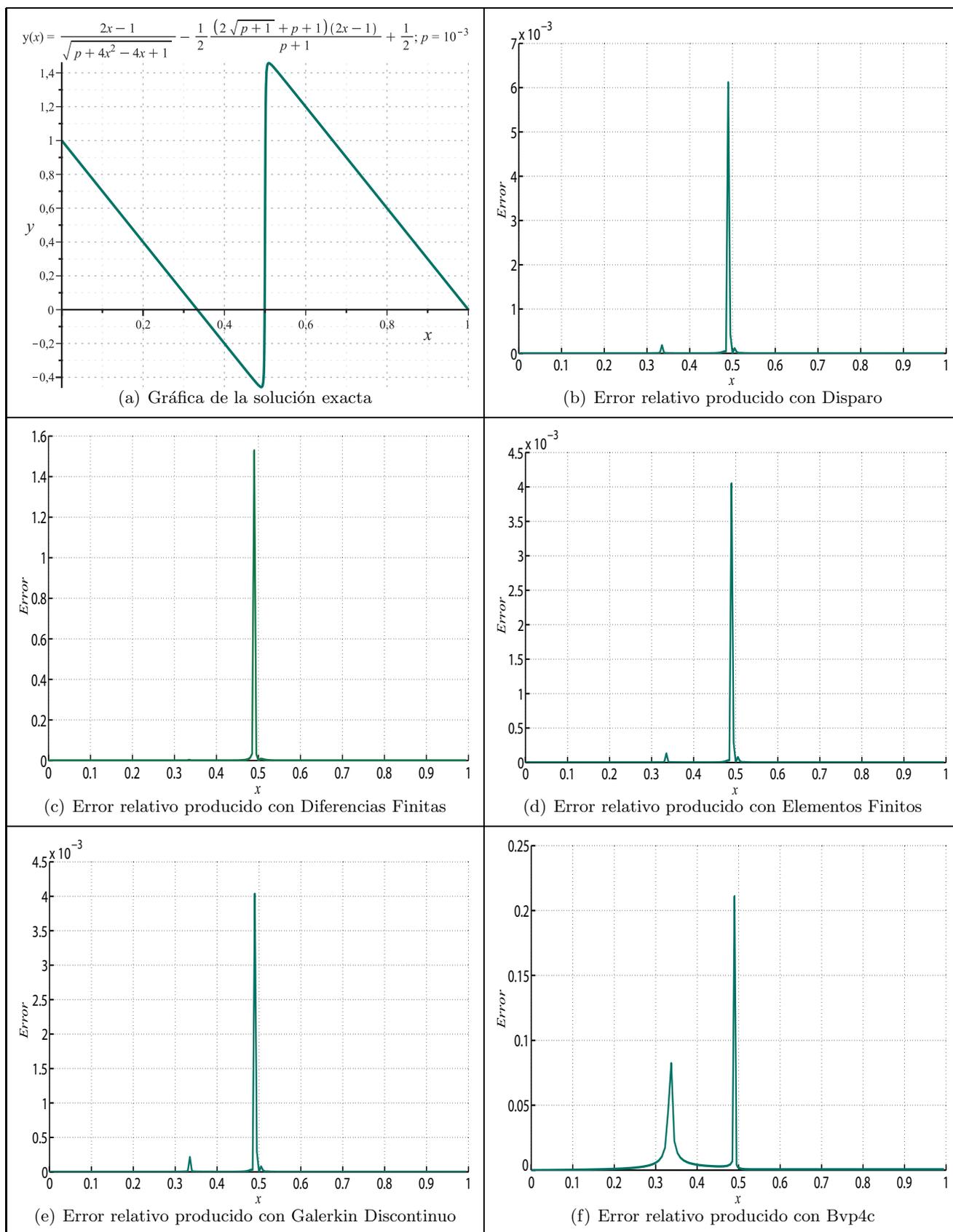


Figura 3.5: Ejemplo 4. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas, Elementos Finitos, Galerkin Discontinuo y Bvp4c, para una malla de 200 subintervalos, empleando la norma del máximo.

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 100$	Disparo	0.066	$1.230 \times 10^{-3}$	$4.154 \times 10^{-3}$
	Dif. Finitas	0.032	$2.522 \times 10^{-2}$	$2.438 \times 10^{-3}$
	Elementos Finitos	0.120	$8.797 \times 10^{-4}$	$4.479 \times 10^{-3}$
	G. Discontinuo	0.199	$8.796 \times 10^{-4}$	$4.479 \times 10^{-3}$
$N = 64$	Bvp4c	1.582	$1.115 \times 10^{-3}$	$1.408 \times 10^{-3}$
$N = 200$	Disparo	0.091	$9.264 \times 10^{-5}$	$1.096 \times 10^{-3}$
	Dif. Finitas	0.047	$6.931 \times 10^{-3}$	$1.479 \times 10^{-4}$
	Elementos Finitos	0.164	$6.216 \times 10^{-5}$	$1.087 \times 10^{-3}$
	G. Discontinuo	0.376	$6.223 \times 10^{-5}$	$1.087 \times 10^{-3}$
$N = 138$	Bvp4c	1.669	$9.849 \times 10^{-4}$	$1.404 \times 10^{-3}$
$N = 500$	Disparo	0.171	$2.144 \times 10^{-6}$	$1.761 \times 10^{-4}$
	Dif. Finitas	0.064	$1.150 \times 10^{-3}$	$9.045 \times 10^{-6}$
	Elementos Finitos	0.417	$1.431 \times 10^{-6}$	$1.758 \times 10^{-4}$
	G. Discontinuo	1.231	$1.435 \times 10^{-6}$	$1.758 \times 10^{-4}$
	Bvp4c	1.744	$2.144 \times 10^{-6}$	$1.761 \times 10^{-4}$

Cuadro 3.7: Ejemplo 4. Tiempo empleado y errores obtenidos para un mallado de 100, 200 y 500 subintervalos. Para las mallas iniciales de 100 y 200 subintervalos, el método Bvp4c adapta dichas mallas con la finalidad de mejorar la estimación de la solución, obteniendo así dos mallas finales de 64 y 138 subintervalos, respectivamente.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3841	$6.173 \times 10^{-10}$
Dif. Finitas	6930	$5.960 \times 10^{-6}$
Elementos Finitos	1301	$3.123 \times 10^{-8}$
G. Discontinuo	631	$9.876 \times 10^{-7}$

Cuadro 3.8: Ejemplo 4. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 1.669 segundos (tiempo usado por Bvp4c).

### Ejemplo 5

Consideremos la siguiente ecuación diferencial:

$$y'' = -y^3 + \frac{3x^3 - 6x}{(1 + x^2)^3},$$

$$y(0.2) = \frac{0.2}{1.04},$$

$$y(0.8) = \frac{0.8}{1.64},$$

el cual tiene por solución exacta

$$y(x) = \frac{x}{1 + x^2}.$$

La ecuación planteada es un problema de valor de frontera del tipo no lineal. En la Figura 3.6, se ilustra la gráfica de la solución  $y(x)$ . Dividiendo uniformemente el intervalo  $[0.2, 0.8]$  en 20, 40 y 80 subintervalos, se puede apreciar del Cuadro 3.9, que los errores, empleando las dos normas, son suficientemente pequeños. Además, se debe notar que las soluciones aportadas tanto por Disparo como por Bvp4c son mejores que las aportadas por

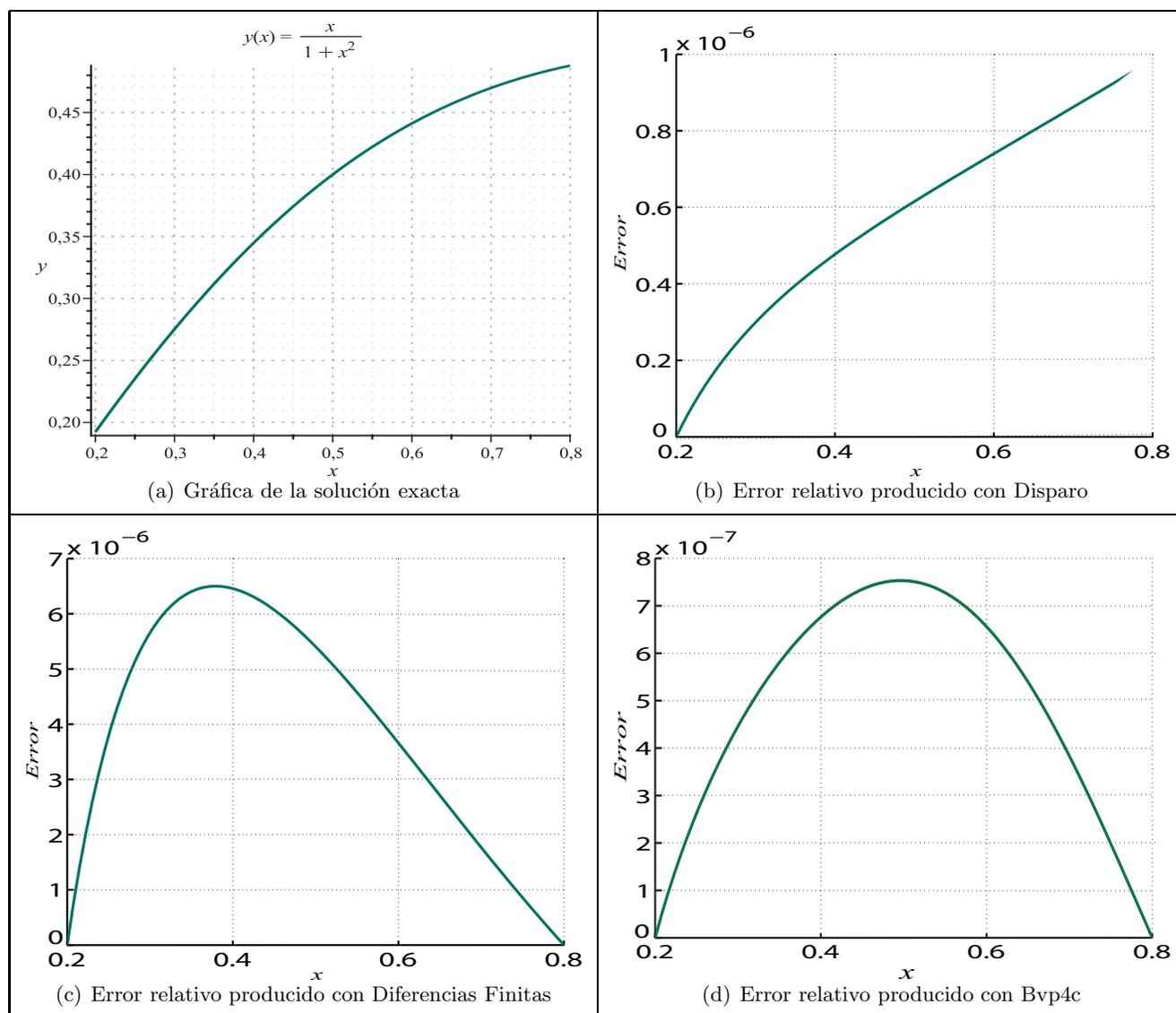


Figura 3.6: Ejemplo 5. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenido con Disparo, Diferencias Finitas y Bvp4c, para una malla de 80 subintervalos, usando la norma del máximo.

Diferencias Finitas (recuerde, nuevamente, que el orden de aproximación de los dos primeros es el doble que el orden de aproximación de Diferencias Finitas,  $\mathcal{O}(h^4)$  contra  $\mathcal{O}(h^2)$ , respectivamente).

Para un tamaño de paso  $h = 0.0075$ , el comportamiento del error relativo que se comete con los tres métodos se refleja en la Figura 3.6. Como es de apreciar, los errores, tanto para Diferencias Finitas (que amerita dos iteraciones por parte del método de Newton Raphson) como Bvp4c, empiezan a aumentar considerablemente, alcanzando el máximo alrededor de  $x = 0.4$  y  $x = 0.5$ , respectivamente, para luego empezar a disminuir. En el caso del método del Disparo (que requiere tres iteraciones del método Newton Raphson) el error crece de forma lineal. Es importante resaltar que tanto en el método de Diferencias Finitas como en Bvp4c, la solución es aproximada sobre todo el intervalo  $[0.2, 0.8]$  y las condiciones de frontera son tomadas todo el tiempo, cosa que no ocurre con Disparo. Es por esta razón, para este último método, que el error producido para  $x = 0.8$  no necesariamente es cero, como se podría pensar.

Para las tres mallas iniciales dadas, el método Bvp4c no realiza proceso de adaptatividad, lo que hace pensar que las aproximaciones usando los polinomios cúbicos son buenas. Particularmente, cuando se trabaja en una malla uniforme que consta de 21 nodos, el método Bvp4c usa un tiempo de 0.886 segundos para realizar las operaciones involucradas en la aproximación, cometiendo un error máximo de  $3.087 \times 10^{-7}$ . Para que Disparo

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 20$	Disparo	0.045	$1.672 \times 10^{-6}$	$7.431 \times 10^{-5}$
	Dif. Finitas	0.042	$3.643 \times 10^{-5}$	$5.530 \times 10^{-5}$
	Bvp4c	0.886	$3.087 \times 10^{-7}$	$7.384 \times 10^{-5}$
$N = 40$	Disparo	0.052	$8.711 \times 10^{-7}$	$1.875 \times 10^{-5}$
	Dif. Finitas	0.047	$9.106 \times 10^{-6}$	$1.382 \times 10^{-5}$
	Bvp4c	1.046	$3.077 \times 10^{-7}$	$1.857 \times 10^{-5}$
$N = 80$	Disparo	0.059	$4.807 \times 10^{-7}$	$4.785 \times 10^{-6}$
	Dif. Finitas	0.052	$2.276 \times 10^{-6}$	$3.454 \times 10^{-6}$
	Bvp4c	1.087	$3.079 \times 10^{-7}$	$4.759 \times 10^{-6}$

Cuadro 3.9: Ejemplo 5. Tiempo empleado y errores obtenidos para una malla de 20, 40 y 80 subintervalos.

y Diferencias Finitas requieran un tiempo similar, necesitan emplear el número de intervalos que vienen dados en el Cuadro 3.10.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3601	$1.050 \times 10^{-7}$
Dif. Finitas	3391	$1.267 \times 10^{-9}$

Cuadro 3.10: Ejemplo 5. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 0.886 segundos.

Empleando el Cuadro 3.9 y el Cuadro 3.10, se logra apreciar que el error obtenido por el método del Disparo mejora poco, pese a aumentar la partición del dominio de la ecuación diferencial. Por el contrario, las estimaciones aportadas por Diferencias Finitas son mejores a medida que se incrementa la partición del dominio. El error que comete Diferencias Finitas es mucho menor en comparación con el método del Disparo.

### Ejemplo 6

Consideremos el siguiente PVF no lineal:

$$\begin{aligned}
 y'' &= -yy' - 2 \sin(x) - x \cos(x) + x \cos^2(x) - \frac{x^2 \sin(2x)}{2}, \\
 y(-1) &= -\cos(-1), \\
 y(2) &= 2 \cos(2),
 \end{aligned}$$

con solución exacta

$$y(x) = x \cos(x).$$

La gráfica de la solución del PVF planteado se muestra en la Figura 3.7. Se realiza una partición uniforme del intervalo  $[-1, 2]$  en 20, 40 y 80 subintervalos. En virtud del Cuadro 3.11, se logra notar que los errores usando cada una de las dos normas introducidas son aceptables. Las estimaciones aportadas por Disparo y Bvp4c siguen siendo mejores que las aportadas por Diferencias Finitas, debido a que este último método posee un orden de aproximación bajo, en comparación con los otros dos.

Para un tamaño de paso  $h = 0.0375$ , es decir 80 subintervalos, el comportamiento del error relativo cometido viene dado en la Figura 3.7. Para el método del Disparo, que amerita cuatro iteraciones por parte del método de Newton Raphson, el error se mantiene estable, exceptuando los picos que presenta; éstos son debido a los ceros de la solución. En el extremo derecho el error es distinto de cero debido al mismo razonamiento hecho en el ejemplo anterior. Para Diferencias Finitas, que requiere cuatro iteraciones por parte del método de Newton, el error presenta un comportamiento análogo al método del Disparo pero con un error mayor en magnitud. Para

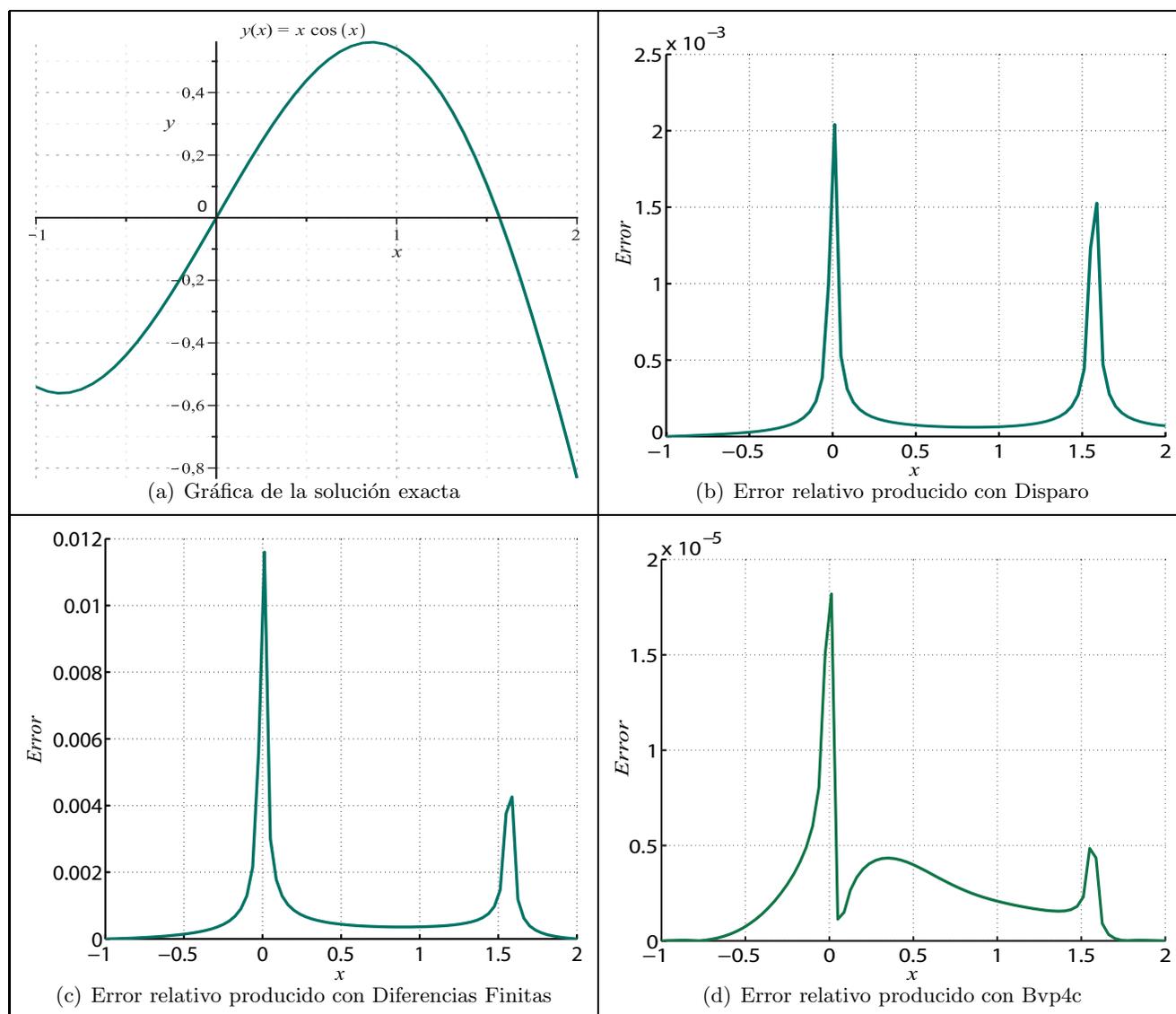


Figura 3.7: Ejemplo 6. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas y **Bvp4c**, para una malla de 80 subintervalos, usando la norma del máximo.

el método **Bvp4c**, el error es análogo pero con magnitudes mucho menores que las presentadas por el método del Disparo.

Es importante acotar, que para las tres mallas iniciales dadas, el método **Bvp4c** no realiza proceso de adaptatividad, con lo que se concluye que las estimaciones usando polinomios cúbicos son buenas. En particular, trabajando en una malla uniforme de 21 puntos, el método **Bvp4c** usa un tiempo de 1.070 segundos para realizar las operaciones relacionadas con el cálculo de la aproximación de la solución, cometiendo un error máximo de  $3.095 \times 10^{-6}$ . Para requerir un tiempo aproximadamente igual, tanto Disparo como Diferencias Finitas necesitarán aproximar la solución sobre el número de intervalos que se muestran en el Cuadro 3.12.

En virtud del Cuadro 3.11 y del Cuadro 3.12, se obtiene, al igual que el Ejemplo 5, que el error que proporciona el método del Disparo se sigue manteniendo igual, pese a aumentar considerablemente el total de subintervalos de la partición del dominio donde está definido el PVF. Todo lo contrario ocurre con Diferencias Finitas, ya que su error disminuyó relativamente a medida que se consideró más intervalos en la partición del dominio.

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 20$	Disparo	0.049	$6.304 \times 10^{-4}$	$5.778 \times 10^{-3}$
	Dif. Finitas	0.046	$3.210 \times 10^{-3}$	$4.085 \times 10^{-3}$
	Bvp4c	1.070	$3.095 \times 10^{-6}$	$5.356 \times 10^{-3}$
$N = 40$	Disparo	0.053	$8.948 \times 10^{-5}$	$1.282 \times 10^{-3}$
	Dif. Finitas	0.049	$8.038 \times 10^{-4}$	$1.020 \times 10^{-3}$
	Bvp4c	1.136	$1.901 \times 10^{-6}$	$1.340 \times 10^{-3}$
$N = 80$	Disparo	0.062	$5.792 \times 10^{-5}$	$2.995 \times 10^{-4}$
	Dif. Finitas	0.059	$2.010 \times 10^{-4}$	$2.551 \times 10^{-4}$
	Bvp4c	1.230	$1.769 \times 10^{-6}$	$3.359 \times 10^{-4}$

Cuadro 3.11: Ejemplo 6. Tiempo empleado y errores obtenidos para una malla de 20, 40 y 80 subintervalos.

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	3841	$1.021 \times 10^{-5}$
Dif. Finitas	3092	$1.347 \times 10^{-7}$

Cuadro 3.12: Ejemplo 6. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 1.070 segundos.

### Ejemplo 7

Para finalizar, consideremos ahora el PVF no lineal:

$$y'' = -y^2 + \frac{x(px\sqrt{p+x^2} - 3p + x^3\sqrt{p+x^2})}{\sqrt{(p+x^2)^5}},$$

$$y(-0.1) = -\frac{0.1}{\sqrt{p+0.01}},$$

$$y(0.1) = \frac{0.1}{\sqrt{p+0.01}},$$

con solución

$$y(x) = \frac{x}{\sqrt{p+x^2}},$$

donde  $p = 10^{-5}$ .

La solución gráfica del PVF planteado se ilustra en la Figura 3.8. Se realiza una partición uniforme del intervalo  $[-1, 2]$  en 200, 300 y 500 subintervalos. En virtud del Cuadro 3.13, se aprecia que los errores, empleando cada una de las dos normas introducidas, no son los esperados para la gran cantidad de subintervalos de las tres particiones. De nuevo, las estimaciones aportadas por Disparo y Bvp4c siguen siendo mejores que las aportadas por Diferencias Finitas.

Usando un tamaño de paso  $h = 0.0004$ , es decir, 500 subintervalos, el comportamiento del error que se produce se refleja en la Figura 3.8. Para el método del Disparo, que amerita tres iteraciones por parte del método de Newton Raphson, el error tiene un comportamiento lineal, pero alrededor de  $x = 0$  presenta oscilaciones, esto debido a la alta pendiente que presenta la solución en un entorno de ese punto. Para el método de Diferencias Finitas (que requiere tres iteraciones por parte del método de Newton), y el método Bvp4c, el error tiene un comportamiento estable en casi todo en intervalo  $[-0.1, 0.1]$ , presentando una mayor magnitud en un entorno de  $x = 0$ , donde presenta oscilaciones considerablemente altas.

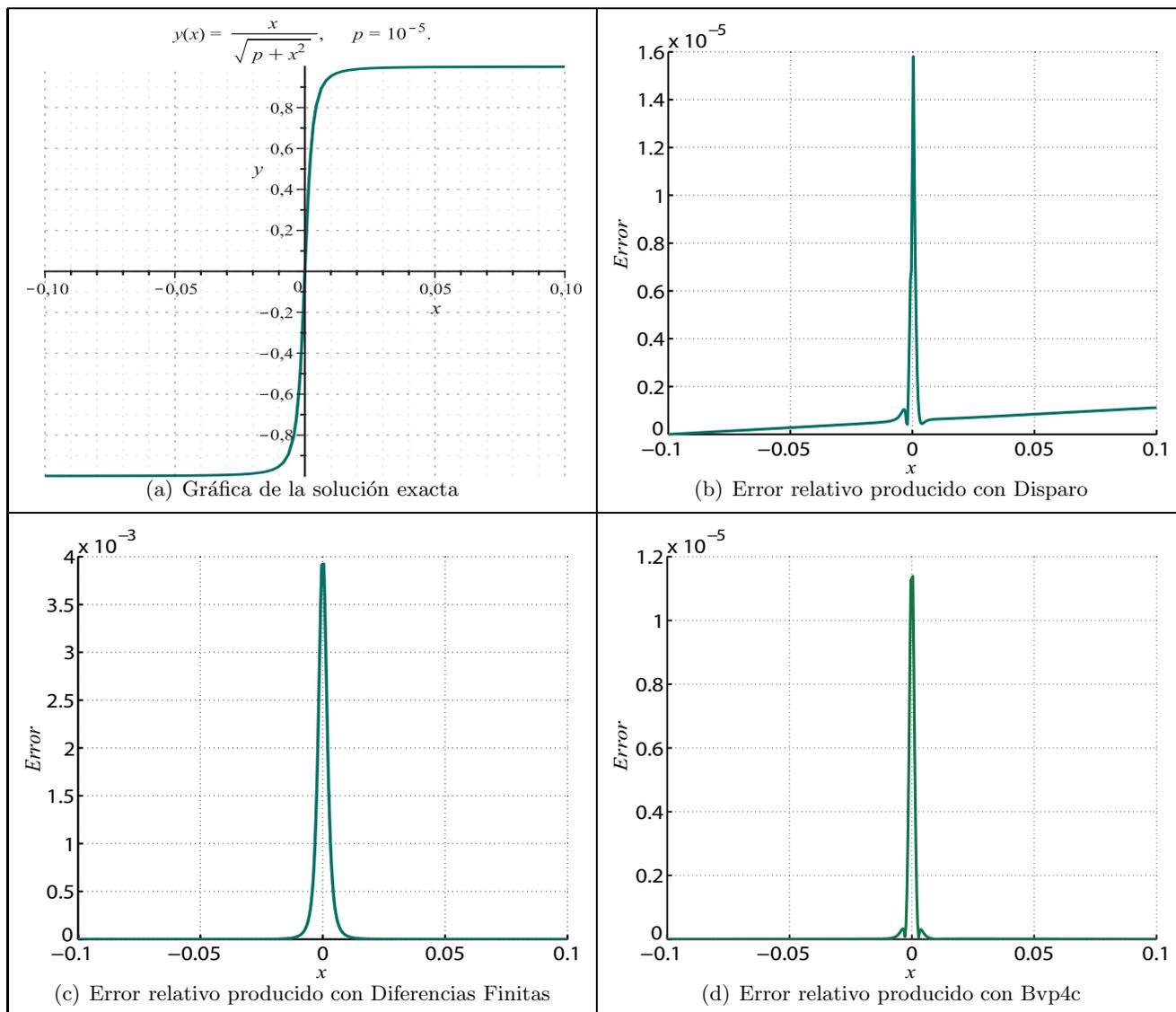


Figura 3.8: Ejemplo 7. De arriba hacia abajo y de izquierda a derecha se muestran la solución exacta y el error (error relativo) obtenidos con Disparo, Diferencias Finitas y Bvp4c, para una malla de 500 subintervalos, usando la norma del máximo.

	Método	Tiempo	$\ \varepsilon\ _\infty$	$\ \varepsilon\ _{L_2}$
$N = 200$	Disparo	0.157	$9.359 \times 10^{-5}$	$4.901 \times 10^{-4}$
	Dif. Finitas	0.079	$6.934 \times 10^{-3}$	$6.614 \times 10^{-5}$
	Bvp4c	1.249	$9.266 \times 10^{-5}$	$4.901 \times 10^{-4}$
$N = 300$	Disparo	0.214	$1.685 \times 10^{-5}$	$2.184 \times 10^{-4}$
	Dif. Finitas	0.097	$3.171 \times 10^{-3}$	$1.901 \times 10^{-5}$
	Bvp4c	1.383	$1.613 \times 10^{-5}$	$2.184 \times 10^{-4}$
$N = 500$	Disparo	0.333	$2.709 \times 10^{-6}$	$7.874 \times 10^{-5}$
	Dif. Finitas	0.134	$1.151 \times 10^{-3}$	$4.047 \times 10^{-6}$
	Bvp4c	1.618	$2.153 \times 10^{-6}$	$7.874 \times 10^{-5}$

Cuadro 3.13: Ejemplo 7. Tiempo empleado y errores obtenidos para una malla de 200, 300 y 500 subintervalos.

Nuevamente, en las tres mallas iniciales dadas, el método **Bvp4c** no realiza proceso de adaptatividad. En particular, cuando se trabaja en una malla uniforme que consta de 201 nodos, el método **Bvp4c** usa un tiempo de 1.249 segundos para realizar las operaciones involucradas, cometiendo un error máximo de  $9.266 \times 10^{-5}$ . Para que Disparo y Diferencias Finitas requieran un tiempo similar, necesitan emplear el número de intervalos que vienen dados en el Cuadro 3.14. Como se logra apreciar, pese a aumentar ampliamente la cantidad de intervalos

Método	Intervalos	$\ \varepsilon\ _\infty$
Disparo	1889	$7.399 \times 10^{-7}$
Dif. Finitas	4250	$1.585 \times 10^{-5}$

Cuadro 3.14: Ejemplo 7. Números de intervalos necesarios y error obtenido por los métodos para requerir un tiempo próximo a 1.249 segundos.

de la partición del dominio donde está definida la solución del PVF, los errores sólo disminuyen levemente. El método del Disparo logra aportar mejores estimaciones en menos subintervalos que el método de Diferencias Finitas.

### Consideraciones Finales

En vista de los resultados obtenidos en los ejemplos anteriores podemos dar algunas consideraciones sobre los métodos numéricos estudiados. En primer lugar, los cuatro métodos que no hacen adaptatividad (todos menos **Bvp4c**) son métodos con un costo computacional muy bajo, lo que permite aumentar el total de subintervalos de la partición con la finalidad de mejorar mucho más las aproximaciones obtenidas. Es decir, refinando la malla de forma uniforme (sin ningún criterio de optimalidad) hasta un punto que estos métodos usen el mismo tiempo de cálculo que el usado por el método **Bvp4c**, se obtienen mejores resultados. En otras palabras, si no se requiere la garantía de la precisión de la solución (cota del error) cualquiera de estos métodos resulta apropiado, incluso en problemas con singularidades muy grandes.

Dadas las propiedades que presenta el método Galerkin Discontinuo, se tiene que su comportamiento en norma  $L_2$  resulta ser el método más robusto. Pues, en todos los ejemplos mostrados, su error es el menor y decrece a medida que aumenta el número de nodos en la partición de la malla. Debemos tener claro que el error en norma infinita de éste método no es representativo, dado que este método representa la solución por un polinomio de segundo grado en cada intervalo del dominio.

Se debe revisar el proceso adaptativo en el método **Bvp4c**. Del Ejemplo 4 se concluye que el error relativo no es representativo aún cuando se cumple el error absoluto requerido en el proceso adaptativo (ver Figuras 3.4 y 3.5 del Ejemplo 4).

Para el caso en el que el PVF es no lineal, no hay mayor diferencia con respecto al caso lineal. La mayor

relevancia está en que la precisión de las soluciones dadas por los métodos del Disparo, Diferencias Finitas y Bvp4c son un poco menores en comparación con las soluciones aportadas por estos métodos para el caso lineal.

Como continuación de este trabajo, se plantea la posibilidad de implementar un proceso adaptativo para los métodos del Disparo, Diferencias Finitas, Elementos Finitos y Galerkin Discontinuo. De la misma manera, se recomienda hacer el estudio del método de Elementos Finitos y Galerkin Discontinuo para el caso en que el PVF es no lineal, ya que, por efecto de tiempo de elaboración de este trabajo, no se pudo realizar.



### A.1. Correspondientes al Método del Disparo

**Teorema A.1** Sea  $h = (b - a)/N$ ,  $x_i = a + nh$ , con  $i = 0, 1, \dots, N$ . Sean  $y(x)$ ,  $u(x)$  y  $v(x)$  las soluciones respectivas de (2.1), (2.2) y (2.3). Si  $\tilde{u}_i$  y  $\tilde{v}_i$  son aproximaciones de  $\mathcal{O}(h^n)$  para  $u_i = u(x_i)$  y  $v_i = v(x_i)$ , respectivamente, entonces  $w_i$  será una aproximación de  $\mathcal{O}(h^n)$  para  $y_i = y(x_i)$ . En particular,

$$|w_i - y_i| \leq Kh^n \left| 1 + \frac{v_i}{v_N} \right|,$$

para alguna  $K$  apropiada.

**Demostración:** Definamos

$$e_i = w_i - y_i,$$

$$e_i^{(u)} = \tilde{u}_i - u_i,$$

$$e_i^{(v)} = \tilde{v}_i - v_i,$$

para  $i = 0, 1, \dots, N$ . En virtud de (2.4), en cada nodo  $x_i$ , se tiene,

$$y_i = u_i + Cv_i, \tag{A.1}$$

donde

$$C = \frac{\beta - u(b)}{v(b)}.$$

Por otra parte,

$$w_i = \tilde{u}_i + \tilde{C}\tilde{v}_i, \tag{A.2}$$

donde

$$\tilde{C} = \frac{\beta - \tilde{u}_N}{\tilde{v}_N}.$$

Restando (A.1) a (A.2),

$$\begin{aligned} w_i - y_i &= \tilde{u}_i - u_i + \tilde{C}\tilde{v}_i - Cv_i \\ &= \tilde{u}_i - u_i + \tilde{C}\tilde{v}_i - \tilde{C}v_i + \tilde{C}v_i - Cv_i \\ &= \tilde{u}_i - u_i + \tilde{C}(\tilde{v}_i - v_i) + (\tilde{C} - C)v_i. \end{aligned}$$

Por tanto,

$$e_i = e_i^{(u)} + \tilde{C}e_i^{(v)} + (\tilde{C} - C)v_i. \tag{A.3}$$

Para  $j = N$ , notemos que  $e_N = 0$ , así,

$$\tilde{C} - C = -\frac{e_N^{(u)} + \tilde{C}e_N^{(v)}}{v_N}. \quad (\text{A.4})$$

Sustituyendo (A.4) en (A.3),

$$e_i = e_i^{(u)} + \tilde{C}e_i^{(v)} - \left( e_N^{(u)} + \tilde{C}e_N^{(v)} \right) \frac{v_i}{v_N}.$$

Tomando valor absoluto, y sabiendo que  $|e_i^{(u)}| \leq M_1 h^n$  y  $|e_i^{(v)}| \leq M_2 h^n$ , para ciertas constantes positivas  $M_1$  y  $M_2$ , tenemos

$$\begin{aligned} |e_i| &\leq M_1 h^n + |\tilde{C}| M_2 h^n + \left( M_1 h^n + |\tilde{C}| M_2 h^n \right) \left| \frac{v_i}{v_N} \right| \\ &= \left( M_1 + |\tilde{C}| M_2 \right) h^n \left( 1 + \left| \frac{v_i}{v_N} \right| \right). \end{aligned}$$

Así, eligiendo  $K = M_1 + |\tilde{C}| M_2$ , concluimos que  $w_i$  aproxima a  $y_i$  con un orden  $\mathcal{O}(h^n)$ . ■

## A.2. Correspondientes al Método de Diferencias Finitas

**Teorema A.2** *Supongamos que  $p(x)$ ,  $q(x)$  y  $r(x)$  son continuas en  $[a, b]$ . Si  $q(x) \geq 0$  en  $[a, b]$ , entonces el sistema lineal tridiagonal (2.17) tiene una solución única siempre y cuando  $h < 2/L$ , donde  $L = \max_{a \leq x \leq b} |p(x)|$ .*

**Demostración:** Basta verificar que se cumplen las hipótesis del teorema (1.2).

Como  $L < 2/h$ , entonces, para cada  $i = 1, 2, \dots, n$ , se tiene que  $|p(x_i)| < 2/h$ , con lo que,

$$-1 < \frac{h}{2} p(x_i) < 1 \quad (\text{A.5})$$

De la última desigualdad se obtiene que  $0 < 1 - \frac{h}{2} p(x_i) = |\delta_i|$  y  $0 < 1 + \frac{h}{2} p(x_i) = |\epsilon_i|$ . Por tanto,

1. Para  $i = 2, 3, \dots, n-1$ , se verifica que  $\epsilon_i \delta_i \neq 0$ .
2. Usando (A.5) y sabiendo que  $q(x) \geq 0$ , en  $[a, b]$ , se tiene,

$$\begin{aligned} |\delta_1| &= 1 - \frac{h}{2} p(x_1) \\ &< 2 \\ &\leq 2 + h^2 q(x_1) \\ &= |\gamma_1|. \end{aligned}$$

3. Para  $i = 2, \dots, n-1$ ,

$$\begin{aligned} |\epsilon_i| + |\delta_i| &= 1 + \frac{h}{2} p(x_i) + 1 - \frac{h}{2} p(x_i) \\ &= 2 \\ &\leq 2 + h^2 q(x_i) \\ &= |\gamma_i|. \end{aligned}$$

4. Nuevamente, usando (A.5), se obtiene,

$$\begin{aligned} |\epsilon_n| &= 1 + \frac{h}{2} p(x_n) \\ &< 2 \\ &\leq 2 + h^2 q(x_n) \\ &= |\gamma_n|. \end{aligned}$$

Así, el sistema (2.17) tiene solución única. ■

**Teorema A.3** Sean  $Q^*$  tal que  $q(x) \geq Q^* > 0$ , para  $a \leq x \leq b$ ,  $L = \max_{a \leq x \leq b} |p(x)|$ ,  $M_3 = \max_{a \leq x \leq b} |y^{(3)}(x)|$  y  $M_4 = \max_{a \leq x \leq b} |y^{(4)}(x)|$ . Entonces, para  $i = 0, 1, \dots, n+1$ , se cumple:

$$|w_i - y(x_i)| \leq h^2 \left( \frac{M_4 + 2LM_3}{12Q^*} \right).$$

**Demostración:** Para  $i = 0, 1, \dots, n+1$ , pongamos  $e_i = w_i - y(x_i)$ , y sea  $e = \max_{0 \leq i \leq n+1} |e_i|$ .

Existen  $\xi_i, \eta_i \in (x_{i-1}, x_{i+1})$  tales que,

$$\begin{aligned} \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} &= p(x_i) \left( \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} \right) + q(x_i)y(x_i) + r(x_i) \\ &\quad - \frac{h^2}{12} \left( 2p(x_i)y^{(3)}(\eta_i) - y^{(4)}(\xi_i) \right), \end{aligned} \tag{A.6}$$

Se sabe que,

$$\frac{w_{i+1} - 2w_i + w_{i-1}}{h^2} = p(x_i) \left( \frac{w_{i+1} - w_{i-1}}{2h} \right) + q(x_i)w_i + r(x_i). \tag{A.7}$$

Restando (A.6) de (A.7) y usando la definición de  $e_i$ ,

$$\frac{e_{i+1} - 2e_i + e_{i-1}}{h^2} = p(x_i) \left( \frac{e_{i+1} - e_{i-1}}{2h} \right) + q(x_i)e_i + \frac{h^2}{12} \left( 2p(x_i)y^{(3)}(\eta_i) - y^{(4)}(\xi_i) \right).$$

Multiplicando la igualdad anterior por  $h^2$  y manipulando algebraicamente se obtiene,

$$(2 + h^2q(x_i))e_i = \left( 1 - \frac{h}{2}p(x_i) \right) e_{i+1} + \left( 1 + \frac{h}{2}p(x_i) \right) e_{i-1} - \frac{h^4}{12} \left( y^{(4)}(\xi_i) - 2p(x_i)y^{(3)}(\eta_i) \right),$$

para  $i = 1, 2, \dots, n$ . Por tanto,

$$\begin{aligned} (2 + h^2q(x_i))|e_i| &\leq \left( 1 - \frac{h}{2}p(x_i) \right) e + \left( 1 + \frac{h}{2}p(x_i) \right) e + \frac{h^4}{12} \left( |y^{(4)}(\xi_i)| + 2|p(x_i)||y^{(3)}(\eta_i)| \right) \\ &\leq 2e + \frac{h^4}{12} \left( M_4 + 2LM_3 \right), \end{aligned}$$

Notemos que  $e_0 = e_{n+1} = 0$ . Así, la desigualdad anterior es válida para todo  $i = 0, 1, \dots, n+1$ . Además  $h^2q(x_i) \geq h^2Q^*$ . Por lo tanto,

$$h^2Q^*e \leq \frac{h^4}{12} \left( M_4 + 2LM_3 \right),$$

de donde,

$$e \leq \frac{h^4}{12Q^*} \left( M_4 + 2LM_3 \right).$$

■



## Referencias

- [1] Béatrice Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equation*, SIAM (2008) 3-17.
- [2] Charles W. Curtis, *Linear Algebra: An Introductory Approach*, Editorial Board, 1974.
- [3] B. Dayanad Reddy, *Functional Analysis and Boundary-Value Problems: an Introductory treatment*. John Wiley & Sons, Inc. 1986.
- [4] B. Rivière, M. Wheeler, and V. Girault, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I*, Computational Geosciences, 3 (1999), pp. 337 - 360.
- [5] C. Dawson, S. Sun, and M. Wheeler, *Compatible algorithms for coupled flow and transport*, Computer Methods in Applied Mechanics and Engineering, 193 (2004), pp. 2565 - 2580.
- [6] David Kincaid y Ward Cheney, *Análisis Numérico: Las matemáticas del cálculo científico*. Addison - Wesley Iberoamericana, 1994.
- [7] Dennis G. Zill, *Ecuaciones diferenciales con aplicaciones de modelado*, 7a. ed., Thomson Learning, 2002.
- [8] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM Journal on Numerical Analysis, 19 (1982), pp. 742 - 760.
- [9] Eugene Isaacson y Helbert Bishop Keller, *Analysis of Numerical Methods*, John Wiley & Sons, Inc., 1966.
- [10] J. Oden, I. Babuška, and C. Baumann, *A discontinuous hp finite element method for diffusion problems*, Journal of Computational Physics, 146 (1998), pp. 491 - 519.
- [11] J. Kierzenka, *Studies in the Numerical Solution of Ordinary Differential Equations*, PhD thesis, Southern Methodist University, Dallas, TX, 1998.
- [12] Kendall E. Atkinson, *An introduction to NUMERICAL ANALYSIS*, Second Edition, John Wiley & Sons, Inc., 1989.
- [13] Keller, H. B., *Numerical methods for two-point boundary-value problems*, Blaisdell, Waltham, MA, 1968.
- [14] Lawrence F. Shampine, Jacek Kierzenka and Mark W. Reichelt, *Solving Boundary Value Problems for Ordinary Differential Equations in MATLAB with bvp4c*. El tutorial y programas están disponibles en [http://www.mathworks.com/bvp\\_tutorial](http://www.mathworks.com/bvp_tutorial).

- [15] L. Delves and C. Hall, *An implicit matching principle for global element calculations*, Journal of the Institute of Mathematics and its Applications, 23 (1979), pp. 223 - 234.
- [16] M. F. Wheeler, *An elliptic collocation - finite element method with interior penalties*, SIAM Journal on Numerical Analysis, 15 (1978), pp. 152 - 161.
- [17] Pavel Šolín, *Partial Differential Equations and the Finite Element Method*. John Wiley & Sons, Ltd. 2006
- [18] Richard L. Burden y J. Douglas Faires, *Análisis numérico*, 7a. ed., Thomson Learning, 2002.
- [19] Zienkiewicz and Cheung, *The Finite Element Method in Structural and Continuum Mechanics*, 1967.