



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Matemáticas

**Introducción a la integración numérica de ecuaciones diferenciales
algebraicas de índice 1**

Autor: José Ignacio Segovia Martín

Tutora: M.P. Calvo Cabrero

Índice general

Introducción	3
1. Ecuaciones Diferenciales Algebraicas (EDAs)	5
1.1. Índice de una EDA	5
1.1.1. Índice de perturbación	5
1.1.2. Índice de diferenciación	6
1.2. El Péndulo	9
1.2.1. El péndulo como un sistema de EDOs	10
1.2.2. El péndulo como un sistema de EDAs	11
2. Métodos de Rosenbrock para EDAs de índice 1	15
2.1. Definición del Método	15
2.1.1. Métodos de Rosenbrock para EDOs	15
2.1.2. Métodos de Rosenbrock para un sistema de EDAs de índice 1	17
2.1.3. Derivadas de la solución exacta	20
2.2. Condiciones de Orden	21
2.2.1. Árboles y Diferenciales Elementales	21
2.2.2. Desarrollo de Taylor de la solución exacta	23
2.2.3. Desarrollo de Taylor de la solución numérica	25
2.2.4. Convergencia	34
2.3. Implementación	37
3. Resultados numéricos	41
3.1. Problemas	41
3.1.1. Péndulo matemático	41
3.1.2. Amplificador	42
3.2. Métodos	46
3.3. Resultados	48
3.3.1. Péndulo matemático	48
3.3.2. Amplificador	51

Bibliografía	55
A. Código MATLAB	57
A.1. Función principal: Rosenbrock.m	57
A.2. Funciones para el péndulo	60
A.3. Funciones para el amplificador	61
A.4. ErrorCondicionAlgebraica.m	62
A.5. U5Ue.m	64

Introducción

El objetivo de este trabajo es, tras introducir los conceptos y resultados básicos relacionados con las ecuaciones diferenciales algebraicas, el estudio de la integración numérica de las ecuaciones de índice 1 mediante métodos de tipo Rosenbrock. Ni el tipo de ecuaciones diferenciales consideradas, ni la familia de métodos que se utilizan en este trabajo para su integración numérica se han estudiado en las asignaturas del Grado en Matemáticas de la Universidad de Valladolid.

En el Capítulo 1 han introducido los conceptos básicos sobre las ecuaciones diferenciales algebraicas, junto con un ejemplo sencillo para ilustrar el concepto de índice.

El Capítulo 2, el más extenso de todo el trabajo, se ha dedicado a estudiar detalladamente los métodos de Rosenbrock para integrar EDAs de índice 1. Se han estudiado con detalle las condiciones de orden de dichos métodos, para lo cual nos hemos ayudado de una *teoría de árboles* adecuada.

En el Capítulo 3 se presentan dos problemas que conducen a sistemas de EDAs de índice 1: el péndulo y un amplificador con un transistor. Se han implementado un método de orden 3 y otro de orden 4 de la familia estudiada en el Capítulo 2 y se ilustra su comportamiento sobre ambos problemas, aunque para el circuito únicamente se ha utilizado el método de orden 4.

En el Apéndice se incluyen los programas en MATLAB que implementan los métodos elegidos y que se han utilizado para obtener los resultados numéricos recogidos en el Capítulo 3.

Agradecer a mi tutora, María Paz Calvo Cabrero por su trabajo de dirección realizado que me ha permitido finalizar el trabajo a tiempo a pesar de haber estado fuera todo el curso 2018/2019.

Capítulo 1

Ecuaciones Diferenciales Algebraicas (EDAs)

Consideramos en este trabajo ecuaciones diferenciales algebraicas, es decir, problemas cuya forma general es un sistema de ecuaciones diferenciales en forma implícita,

$$F(y', y) = 0, \quad (1.1)$$

donde $F : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}^D$, $y : \mathbb{R} \rightarrow \mathbb{R}^D$ e $y' = dy/dx$ denota la derivada de y respecto de la variable independiente x .

Si tenemos un sistema no autónomo $F(y', y, x) = 0$, podemos convertirlo en uno de la forma (1.1) añadiendo x como una componente adicional a y , y añadiendo la ecuación $x' - 1 = 0$. Suponemos que el valor inicial $y(0)$ es conocido, y que buscamos la solución $y(x)$ en el intervalo acotado $[0, \bar{x}]$. En el caso de que $\partial F / \partial y'$ sea invertible, podemos resolver formalmente para y' en (1.1) y obtener así una ecuación diferencial ordinaria. Pero lo que nos interesa en este trabajo es estudiar el caso en el que $\partial F / \partial y'$ es singular. A continuación clasificaremos este tipo de problemas introduciendo el concepto de índice.

1.1. Índice de una EDA

1.1.1. Índice de perturbación

Introducimos el concepto de índice de perturbación como medida de la sensibilidad de las soluciones al perturbar la ecuación.

Definición 1.1 Decimos que la ecuación (1.1) tiene un índice de perturbación m en la solución y en $[0, \bar{x}]$ si m es el entero más pequeño tal que, para todas las funciones \hat{y} con

$$F(\hat{y}', \hat{y}) = \delta(x), \quad (1.2)$$

existe en $[0, \bar{x}]$ una estimación,

$$\|\hat{y}(x) - y(x)\| \leq C(\|\hat{y}(0) - y(0)\| + \max_{0 \leq \xi \leq x} \|\delta(\xi)\| + \dots + \max_{0 \leq \xi \leq x} \|\delta^{(m-1)}(\xi)\|), \quad (1.3)$$

cuando la expresión del lado derecho es suficientemente pequeña. C es una constante que únicamente depende de F y del tamaño del intervalo $[0, \bar{x}]$.

Diremos que una ecuación tiene índice de perturbación m cuando lo tiene a lo largo de todas las soluciones. Por lo tanto, el índice de perturbación no puede ser menor que 1.

El caso de índice 0 se puede incluir si interpretamos $\delta^{(-1)}(\xi)$ como una integral sobre δ . Entonces, decimos que (1.1) tiene índice de perturbación 0 si existe una estimación de la forma

$$\|\hat{y}(x) - y(x)\| \leq C(\|\hat{y}(0) - y(0)\| + \max_{0 \leq \xi \leq x} \|\int_0^\xi \delta(t) dt\|).$$

1.1.2. Índice de diferenciación

El índice de diferenciación del sistema (1.1) es el mínimo entero m tal que el sistema de ecuaciones (1.1) y

$$\frac{dF(y', y)}{dx} = 0, \quad \dots, \quad \frac{d^m F(y', y)}{dx^m} = 0,$$

puede ser resuelto para obtener y' como función de y , $y'(y)$.

El índice de diferenciación nos indica por tanto el número de veces que debemos diferenciar una EDA para transformarla en una ecuación diferencial ordinaria (EDO).

A continuación conoceremos más específicamente el formato de los sistemas de índices de diferenciación 1, 2 y 3.

Sistemas de índice 1 El más sencillo, tiene la forma

$$y' = f(y, z), \quad (1.4a)$$

$$0 = g(y, z), \quad (1.4b)$$

donde f, g son suficientemente diferenciables (siempre lo daremos por supuesto) y g_z tiene inversa acotada ($g_z = \partial g / \partial z$).

Derivando (1.4b) respecto de nuestra variable independiente x y sustituyendo y' por su expresión en (1.4a) obtenemos

$$0 = g_y(y, z)f(y, z) + g_z(y, z)z'.$$

Como hemos supuesto que g_z es invertible en un entorno de la solución, tenemos

$$z' = -g_z^{-1}(y, z)g_y(y, z)f(y, z).$$

Por lo tanto, el sistema (1.4a)-(1.4b) tiene índice de diferenciación 1 pues al derivar una vez, obtenemos el sistema de EDOs

$$\frac{d}{dt}(y, z) = (f(y, z), -g_y(y, z)f(y, z)g_z^{-1}(y, z)).$$

Diremos que el valor inicial (y_0, z_0) es consistente si $g(y_0, z_0) = 0$. Los problemas de la forma

$$BY' = a(Y), \quad (1.5)$$

donde B es una matriz singular constante, pueden ser también transformados en uno de la forma (1.4a)-(1.4b) al descomponer B como

$$B = S \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} T,$$

con S, T invertibles.

Al premultiplicar por S^{-1} en (1.5) y usar

$$TY = \begin{pmatrix} y \\ z \end{pmatrix},$$

obtenemos

$$\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} TY' = S^{-1}a(Y),$$

que es un sistema de la forma (1.4a)-(1.4b).

La condición de que g_z tiene inversa en un entorno de la solución se traduce en que $[(\frac{\partial}{\partial Y}(S^{-1}a))T^{-1}]_{22}$ tiene inversa en un entorno de la solución, donde $[\cdots]_{22}$ indica el bloque derecho más bajo de la matriz, acorde con la descomposición de B . En este caso, la condición inicial Y_0 es consistente si $a(Y_0)$ está en el rango de B .

Sistemas de índice 2 Consideramos ahora el problema

$$y' = f(y, z), \quad (1.6a)$$

$$0 = g(y), \quad (1.6b)$$

asumiendo que $g_y f_z$ tiene inversa en un entorno de la solución. Tenemos que, derivando (1.6b) respecto de y y sustituyendo y' según (1.6a),

$$0 = g_y(y)f(y, z).$$

Ahora derivando otra vez respecto de y tenemos

$$0 = g_{yy}(y)(f(y, z), f(y, z)) + g_y(y)[f_y(y, z)f(y, z) + f_z(y, z)z'],$$

y de aquí obtenemos que

$$z' = -(g_y(y)f_z(y, z))^{-1}(g_{yy}(y)(f(y, z), f(y, z)) + g_y(y)f_y(y, z)f(y, z)).$$

Luego (1.6a)-(1.6b) tiene índice de diferenciación 2, pues al derivar dos veces obtenemos la EDO

$$\frac{d}{dx}(y, z) = (f(y, z), -(g_y(y)f_z(y, z))^{-1}(g_{yy}(y)(f(y, z), f(y, z)) + g_y(y)f_y(y, z)f(y, z))).$$

Diremos que los valores iniciales (y_0, z_0) son consistentes si satisfacen la restricción algebraica $g(y_0) = 0$.

Sistemas de índice 3 Los problemas de la forma

$$y' = f(y, z), \tag{1.7a}$$

$$z' = k(y, z, u), \tag{1.7b}$$

$$0 = g(y), \tag{1.7c}$$

tienen índice 3 si $g_y f_z k_u$ tiene inversa en un entorno de la solución.

Diferenciando (1.7c) respecto de nuestra variable independiente x , obtenemos

$$0 = g_y(y)f(y, z), \tag{1.8a}$$

$$0 = g_{yy}(y)(f(y, z), f(y, z)) + g_y(y)f_y(y, z)f(y, z) + g_y(y)f_z(y, z)k(y, z, u). \tag{1.8b}$$

Ahora, si consideramos las ecuaciones (1.7a), (1.7b) y (1.8b) obtenemos un sistema de índice 1, por lo que podemos asegurar que (1.7a)-(1.7c) es un sistema de índice 3.

Para que los valores iniciales (y_0, z_0, u_0) sean consistentes han de satisfacer (1.7c), (1.8a) y (1.8b). El índice de diferenciación y el de perturbación son iguales si la derivada de la solución aparece multiplicada por una matriz constante, como en el caso de (1.4a)-(1.4b), (1.6a)-(1.6b) y (1.7a)-(1.7c) donde está multiplicada por una matriz con entradas llenas de unos y ceros.

Podemos encontrar una investigación detallada acerca de la relación entre ambos índices en [1], aunque en este trabajo no hemos profundizado en este aspecto. El principal resultado que relaciona ambos índices es que

$$\text{índice diferenciación} \leq \text{índice perturbación} \leq \text{índice diferenciación} + 1.$$

1.2. El Péndulo

En esta sección ilustraremos con el ejemplo del péndulo los tres tipos de ecuaciones diferenciales algebraicas que hemos visto en la sección anterior, atendiendo al índice de diferenciación.

Un péndulo matemático o simple consta de una varilla de longitud l , con uno de los extremos fijo y en el otro extremo un cuerpo puntual de masa m . Observamos que al desplazar la varilla de su posición de equilibrio, el péndulo comienza a oscilar. En ausencia de rozamiento las fuerzas que intervienen en el movimiento son las siguientes:

- la tensión del hilo T ,
- el peso del cuerpo P .

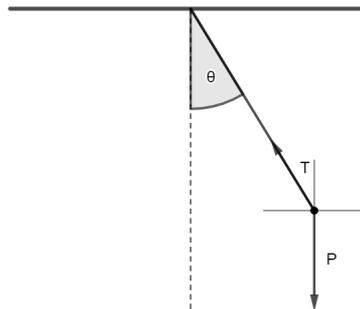


Figura 1.1: Esquema del péndulo simple

Separando las componentes de la tensión obtenemos:

$$\mathbf{T}_x = -T \sin \theta, \quad (1.9a)$$

$$\mathbf{T}_y = -T \cos \theta, \quad (1.9b)$$

donde T es el módulo de la tensión. Además, $\sin \theta = \frac{x}{l}$, $\cos \theta = \frac{y}{l}$ y, por lo tanto,

$$\mathbf{T}_x = -T \frac{x}{l}, \quad (1.10a)$$

$$\mathbf{T}_y = -T \frac{y}{l}. \quad (1.10b)$$

Aplicando las leyes de Newton, tenemos

$$mx'' = -T \frac{x}{l}, \quad (1.11a)$$

$$my'' = -T \frac{y}{l} + m\mathbf{g}. \quad (1.11b)$$

donde \mathbf{g} es la aceleración de la gravedad. Transformándolo ahora en un sistema de primer orden, tenemos

$$x' = u, \quad y' = v, \quad (1.12a)$$

$$u' = -\frac{Tx}{lm}, \quad v' = -\frac{T y}{lm} + \mathbf{g}. \quad (1.12b)$$

Además, sabemos que la varilla siempre tiene la misma longitud, por lo tanto, tenemos que $x^2 + y^2 = l^2$. Consecuentemente, el sistema final que obtenemos es

$$x' = u, \quad y' = v, \quad (1.13a)$$

$$u' = -\frac{Tx}{lm}, \quad v' = -\frac{T y}{lm} + \mathbf{g}, \quad (1.13b)$$

$$0 = x^2 + y^2 - l^2. \quad (1.13c)$$

1.2.1. El péndulo como un sistema de EDOs

Ahora nos disponemos a considerar el péndulo como una EDO.

Como ya dijimos antes, $\sin \theta = \frac{x}{l}$ y $\cos \theta = \frac{y}{l}$ por lo que, si consideramos el cambio de variable $x = l \sin \theta$, $y = l \cos \theta$ vemos que la última ecuación del sistema anterior se garantiza siempre

$$x^2 + y^2 - l^2 = l^2 \sin^2 \theta + l^2 \cos^2 \theta - l^2 = l^2 - l^2 = 0.$$

Las derivadas de primer y segundo orden de las variables x e y respecto de θ son las siguientes

$$x = l \sin \theta, \quad y = l \cos \theta, \quad (1.14a)$$

$$x' = l \theta' \cos \theta, \quad y' = -l \theta' \sin \theta, \quad (1.14b)$$

$$x'' = l \theta'' \cos \theta - l (\theta')^2 \sin \theta, \quad y'' = -l \theta'' \sin \theta - l (\theta')^2 \cos \theta. \quad (1.14c)$$

Sustituyendo ahora en las ecuaciones anteriores, tenemos

$$l \theta'' \cos \theta - l (\theta')^2 \sin \theta = -\frac{T}{lm} l \sin \theta, \quad (1.15a)$$

$$-l \theta'' \sin \theta - l (\theta')^2 \cos \theta = -\frac{T}{lm} l \cos \theta + g, \quad (1.15b)$$

que, tras simplificar, se reduce a

$$\left(\frac{T}{l} - m(\theta')^2\right)l \sin \theta + ml\theta'' \cos \theta = 0, \quad (1.16a)$$

$$\left(\frac{T}{l} - m(\theta')^2\right)l \cos \theta - ml\theta'' \sin \theta - mg = 0. \quad (1.16b)$$

Ahora, multiplicando (1.16a) por $\cos \theta$ y (1.16b) por $-\sin \theta$ y sumando ambas igualdades, obtenemos

$$ml\theta'' + mg \sin \theta = 0,$$

o, equivalentemente,

$$\theta'' = -\frac{g}{l} \sin \theta, \quad (1.17)$$

conocida EDO de segundo orden para el péndulo.

1.2.2. El péndulo como un sistema de EDAs

Por último consideramos el péndulo reescrito como una EDA de índice 3, 2 y 1.

Índice 3 Con la formulación correspondiente al sistema (1.13a)-(1.13c), el péndulo es un sistema de EDAs de índice 3 donde tenemos que agrupando las variables según

$$\mathbf{y} = [x, y]^T, \quad (1.18a)$$

$$\mathbf{z} = [u, v]^T, \quad (1.18b)$$

$$\mathbf{u} = T, \quad (1.18c)$$

las ecuaciones del sistema son las siguientes

$$\mathbf{y}' = \mathbf{z}, \quad (1.19a)$$

$$\mathbf{z}' = \left[-\frac{T}{lm}x, -\frac{T}{lm}y + \mathbf{g}\right]^T, \quad (1.19b)$$

$$0 = x^2 + y^2 - l^2. \quad (1.19c)$$

Vemos que tienen la forma de (1.7a)-(1.7c) con

$$f(\mathbf{y}, \mathbf{z}) = \mathbf{z}, \quad (1.20a)$$

$$k(\mathbf{y}, \mathbf{z}, \mathbf{u}) = \left[-\frac{T}{lm}x, -\frac{T}{lm}y + \mathbf{g}\right]^T, \quad (1.20b)$$

$$g(\mathbf{y}) = x^2 + y^2 - l^2. \quad (1.20c)$$

Tenemos

$$g_y = [2x, 2y], \quad f_z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad k_u = \left[-\frac{x}{ml}, -\frac{y}{ml}\right]^T,$$

luego $g_y f_z k_u = -\frac{2}{ml}(x^2 + y^2)$, que vemos que es invertible en un entorno de la solución.

Índice 2 Ahora pasamos a formularlo como una EDA de índice 2.

Para ello comenzamos derivando $x^2 + y^2 - l^2 = 0$ respecto de nuestra variable independiente (t , el tiempo)

$$2xx' + 2yy' = 0,$$

que, tras sustituir $x' = u$ e $y' = v$ nos queda:

$$xu + yv = 0. \tag{1.21}$$

Las ecuaciones

$$x' = u, \quad y' = v, \tag{1.22a}$$

$$u' = -\frac{Tx}{ml}, \quad v' = -\frac{Ty}{ml} + \mathbf{g}, \tag{1.22b}$$

$$0 = xu + yv, \tag{1.22c}$$

forman un sistema de EDAs de índice 2 con $\mathbf{y} = [x, y, u, v]$ y $\mathbf{z} = T$, que se pueden reagrupar como

$$\mathbf{y}' = [u, v, -\frac{T}{ml}x, -\frac{T}{ml}y + \mathbf{g}]^T, \tag{1.23a}$$

$$0 = xu + yv, \tag{1.23b}$$

con

$$f(\mathbf{y}, \mathbf{z}) = [u, v, -\frac{T}{ml}x, -\frac{T}{ml}y + \mathbf{g}]^T, \tag{1.24a}$$

$$g(\mathbf{y}) = xu + yv. \tag{1.24b}$$

En esta ocasión, tenemos que

$$g_y = [u, v, x, y], \quad f_z = [0, 0, -\frac{x}{ml}, -\frac{y}{ml}]^T.$$

Luego $g_y f_z = -\frac{1}{ml}(x^2 + y^2)$, que es invertible en un entorno de la solución.

Índice 1 Por último, lo formulamos como una EDA de índice 1. Derivando $0 = xu + yv$ respecto de la variable independiente, obtenemos

$$u^2 + xu' + v^2 + yv' = 0.$$

Sustituyendo en la igualdad anterior u' y v' por las expresiones (1.13b) obtenemos

$$u^2 + v^2 + x\left(-\frac{Tx}{ml}\right) + y\left(-\frac{T}{ml} + \mathbf{g}\right) = 0,$$

que implica

$$u^2 + v^2 - \frac{T}{ml}(x^2 + y^2) + \mathbf{g}y = 0.$$

Sabemos que $x^2 + y^2 = l^2$, luego $u^2 + v^2 - \frac{Tl}{m} + \mathbf{g}y = 0$ por lo que

$$m(u^2 + v^2) - Tl + \mathbf{g}my = 0. \quad (1.25)$$

El sistema

$$x' = u, \quad y' = v, \quad (1.26a)$$

$$u' = -\frac{Tx}{ml}, \quad v' = -\frac{T}{ml} + \mathbf{g}, \quad (1.26b)$$

$$0 = m(u^2 + v^2) - Tl + \mathbf{g}my, \quad (1.26c)$$

es un sistema de EDAs de índice 1 donde $\mathbf{y} = [x, y, u, v]^T$ y $\mathbf{z} = T$, y las ecuaciones son

$$\mathbf{y}' = [u, v, -\frac{T}{ml}x, -\frac{T}{ml}y + \mathbf{g}]^T, \quad (1.27a)$$

$$0 = m(u^2 + v^2) - Tl + \mathbf{g}my, \quad (1.27b)$$

con

$$f(\mathbf{y}, \mathbf{z}) = [u, v, -\frac{T}{ml}x, -\frac{T}{ml}y + \mathbf{g}]^T, \quad (1.28a)$$

$$g(\mathbf{y}, \mathbf{z}) = m(u^2 + v^2) - Tl + \mathbf{g}my, \quad (1.28b)$$

donde $g_z = -l$, que es invertible.

Capítulo 2

Métodos de Rosenbrock para EDAs de índice 1

2.1. Definición del Método

2.1.1. Métodos de Rosenbrock para EDOs

Comenzamos, en primer lugar, con el problema de valores iniciales,

$$y'(x) = f(y(x)), \quad y(x_0) = y_0, \quad (2.1)$$

definido sobre un espacio real con $f: \mathbb{R}^D \rightarrow \mathbb{R}^D$, $y: [x_0, x_f] \rightarrow \mathbb{R}^D$ e $y_0 \in \mathbb{R}^D$. Dada una longitud de paso h , queremos encontrar una aproximación y_1 de $y(x_0 + h)$, la solución de (2.1) tras h unidades. Para ello, aplicamos inicialmente un método Runge-Kutta diagonalmente implícito

$$k_i = hf \left(y_0 + \sum_{j=1}^{i-1} a_{ij} k_j + a_{ii} k_i \right) \quad i = 1, \dots, s; \quad (2.2a)$$

$$y_1 = y_0 + \sum_{i=1}^s b_i k_i. \quad (2.2b)$$

La idea es pasar del sistema no lineal (2.2a) a un sistema lineal. Linealizando el lado derecho de (2.2a) para cada i en torno a $y_0 + \sum_{j=1}^{i-1} a_{ij} k_j$ nos lleva a reemplazar (2.2a) por

$$k_i = hf(g_i) + hf'(g_i) a_{ii} k_i, \quad (2.3a)$$

$$g_i = y_0 + \sum_{j=1}^{i-1} a_{ij} k_j, \quad (2.3b)$$

que podemos interpretar como una iteración de Newton para la etapa i -ésima definida en (2.2a), tomando como iterante inicial $k_i^{(0)} = 0$. Las relaciones (2.3a)-(2.3b) resultantes, junto con (2.2b), proporcionan una nueva clase de métodos que se pueden generalizar.

Definición 2.1 *Un método de Rosenbrock de s etapas viene dado por*

$$k_i = hf(y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j) + hJ \sum_{j=1}^i \gamma_{ij} k_j, \quad i = 1, \dots, s; \quad (2.4a)$$

$$y_1 = y_0 + \sum_{i=1}^s b_i k_i, \quad (2.4b)$$

donde α_{ij}, γ_{ij} y b_j , con $1 \leq i, j \leq s$, son los coeficientes del método, y $J = f'(y_0)$.

Si en (2.4a) se pasa restando al lado izquierdo el término $hJ\gamma_{ii}k_i$, se observa que en cada etapa del método hay que resolver un sistema lineal con matriz $I - h\gamma_{ii}J$ y término independiente $hf(y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j) + hJ \sum_{j=1}^{i-1} \gamma_{ij} k_j$ cuya solución es precisamente k_i . Nos interesa estudiar los métodos en los que $\gamma_{11} = \dots = \gamma_{ss} = \gamma$, pues así solo será necesario realizar una factorización LU por cada paso que será utilizada en las s etapas.

Al igual que antes, consideramos problemas de valores iniciales

$$y'(x) = f(y(x)), \quad y(x_0) = y_0. \quad (2.5)$$

En este caso, el correspondiente método de Rosenbrock está definido de la siguiente forma, [6], [5]:

$$Ek_1 = hf(y_0), \quad (2.6a)$$

$$Ek_i = hf\left(y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j\right) + hJ \sum_{j=1}^{i-1} \gamma_{ij} k_j, \quad (i = 2, \dots, s), \quad (2.6b)$$

$$y_1 = y_0 + \sum_{i=1}^s b_i k_i, \quad (2.6c)$$

donde $E = I - \gamma hJ$ (I denota la matriz identidad de tamaño $D \times D$ y J la matriz jacobiana de f evaluada en la condición inicial y_0).

Para los problemas en los cuales f depende explícitamente de la variable independiente x , conocidos como problemas no autónomos,

$$y'(x) = f(x, y(x)), \quad y(x_0) = y_0, \quad (2.7)$$

podemos convertirlos en problemas autónomos añadiendo una nueva variable dependiente y tomando $x' = 1$. El método (2.4a)-(2.4b) aplicado a nuestro nuevo sistema $[y', x']^T = [f(x, y), 1]^T$ nos da las fórmulas

$$k_i = hf \left(x_0 + \alpha_i h, y_0 + h \sum_{j=1}^{i-1} \alpha_{ij} k_j \right) + \gamma_i h^2 \frac{\partial f}{\partial x}(x_0, y_0) + h \frac{\partial f}{\partial y}(x_0, y_0) \sum_{j=1}^i \gamma_{ij} k_j, \quad (i = 1, \dots, s), \quad (2.8a)$$

$$y_1 = y_0 + \sum_{i=1}^s b_i k_i, \quad (2.8b)$$

donde $\alpha_i = \sum_{j=1}^{i-1} \alpha_{ij}$, $\gamma_i = \sum_{j=1}^i \gamma_{ij}$.

Para un sistema diferencial de la forma

$$My'(x) = f(x, y(x)), \quad y(x_0) = y_0, \quad (2.9)$$

con M una matriz constante y no singular, si premultiplicamos (2.9) por M^{-1} , aplicamos el método de Rosenbrock (2.8a)-(2.8b), y después premultiplicamos el método resultante por M , obtenemos

$$Mk_i = hf \left(x_0 + \alpha_i h, y_0 + h \sum_{j=1}^{i-1} \alpha_{ij} k_j \right) + \gamma_i h^2 \frac{\partial f}{\partial x}(x_0, y_0) + h \frac{\partial f}{\partial y}(x_0, y_0) \sum_{j=1}^i \gamma_{ij} k_j, \quad (i = 1, \dots, s), \quad (2.10a)$$

$$y_1 = y_0 + \sum_{i=1}^s b_i k_i. \quad (2.10b)$$

La ventaja de esta formulación es que evitamos invertir la matriz M .

2.1.2. Métodos de Rosenbrock para un sistema de EDAs de índice 1

Ahora, extenderemos los métodos de Rosenbrock a los sistemas de ecuaciones diferenciales algebraicas de forma semiexplícita

$$y' = f(y, z), \quad y(x_0) = y_0, \quad (2.11a)$$

$$0 = g(y, z), \quad z(x_0) = z_0, \quad (2.11b)$$

con $y : [x_0, x_f] \rightarrow \mathbb{R}^D$, $z : [x_0, x_f] \rightarrow \mathbb{R}^{D'}$, $f : \mathbb{R}^D \times \mathbb{R}^{D'} \rightarrow \mathbb{R}^D$ y $g : \mathbb{R}^D \times \mathbb{R}^{D'} \rightarrow \mathbb{R}^{D'}$, con f y g suficientemente regulares. Para ello seguiremos [5], [8].

Suponemos que g_z es invertible, así que el problema es de índice 1, y que las condiciones iniciales son consistentes, es decir, que $g(y_0, z_0) = 0$.

La principal ventaja de los métodos de Rosenbrock sobre los métodos Runge-Kutta implícitos es que podemos evitar totalmente la resolución de sistemas no lineales.

Aunque aplicando el teorema de la función implícita, el sistema (2.11a)-(2.11b) se puede transformar en $y' = f(y, G(y))$ (pues sabemos que (2.11b) posee una solución única local $z = G(y)$, la cual podemos introducir en (2.11a) para obtener así la fórmula anterior), la aplicación del método de Rosenbrock en el formato de EDO podría destruir la ventaja anteriormente mencionada.

Para describir la aplicación de un método de Rosenbrock a (2.11a)-(2.11b), consideramos el sistema de perturbación singular asociado,

$$y' = f(y, z), \quad (2.12a)$$

$$\varepsilon z' = g(y, z), \quad (2.12b)$$

donde ε es un parámetro real pequeño. Al hacer tender ε hacia 0 se obtiene el correspondiente sistema reducido, que no es otro que el sistema de EDAs (2.11a)-(2.11b).

Aplicando el método de Rosenbrock en su formulación (2.4a)-(2.4b) al problema que se obtiene de (2.12a)-(2.12b) dividiendo la segunda ecuación por ε , se llega a

$$\begin{pmatrix} k_i \\ l_i \end{pmatrix} = h \begin{pmatrix} f(v_i, w_i) \\ \varepsilon^{-1} g(v_i, w_i) \end{pmatrix} + h \begin{pmatrix} f_y & f_z \\ \varepsilon^{-1} g_y & \varepsilon^{-1} g_z \end{pmatrix} (y_0, z_0) \sum_{j=1}^i \gamma_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad (2.13a)$$

$$\begin{pmatrix} v_i \\ w_i \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{j=1}^{i-1} \alpha_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad 1 \leq i \leq s, \quad (2.13b)$$

$$\begin{pmatrix} y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{i=1}^s b_i \begin{pmatrix} k_i \\ l_i \end{pmatrix}. \quad (2.13c)$$

Ahora, si multiplicamos la segunda línea de (2.13a) por ε y hacemos $\varepsilon = 0$, ob-

tenemos, para $i = 1, \dots, s$

$$\begin{pmatrix} k_i \\ 0 \end{pmatrix} = h \begin{pmatrix} f(v_i, w_i) \\ g(v_i, w_i) \end{pmatrix} + h \begin{pmatrix} f_y & f_z \\ g_y & g_z \end{pmatrix} (y_0, z_0) \sum_{j=1}^i \gamma_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad (2.14a)$$

$$\begin{pmatrix} v_i \\ w_i \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{j=1}^{i-1} \alpha_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad (2.14b)$$

y, finalmente,

$$\begin{pmatrix} y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{i=1}^s b_i \begin{pmatrix} k_i \\ l_i \end{pmatrix}. \quad (2.14c)$$

Las relaciones (2.14a)-(2.14c) definen la aplicación de los métodos de Rosenbrock al problema (2.11a)-(2.11b).

Estos métodos fueron considerados por primera vez por Michelsen en 1976 (ver [5] y las referencias allí citadas) y estudiados posteriormente por Roche en 1988 [8].

Cabe destacar que el cálculo de (k_i, l_i) , con $1 \leq i \leq s$, en (2.14a) requiere calcular la solución del sistema lineal con matriz

$$\begin{pmatrix} I - \gamma h f_y & -\gamma h f_z \\ -\gamma h g_y & -\gamma h g_z \end{pmatrix}, \quad (2.15)$$

donde $\gamma = \gamma_{ii}$, $1 \leq i \leq s$, y todas las derivadas parciales las evaluamos en el punto (y_0, z_0) . Para g_z no singular, $\gamma \neq 0$, y $h > 0$ suficientemente pequeño, esta matriz es invertible. Esto se puede observar dividiendo los dos bloques de abajo por γh y haciendo $h = 0$.

Si las funciones f y g en (2.11a)-(2.11b) dependen también de x , podemos reemplazar (2.14a) por

$$\begin{pmatrix} k_i \\ 0 \end{pmatrix} = h \begin{pmatrix} f(x_0 + \alpha_i h, v_i, w_i) \\ g(x_0 + \alpha_i h, v_i, w_i) \end{pmatrix} + h \begin{pmatrix} f_y & f_z \\ g_y & g_z \end{pmatrix} \sum_{j=1}^i \gamma_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix} + h^2 \gamma_i \begin{pmatrix} f_x \\ g_x \end{pmatrix},$$

donde todas las derivadas son evaluadas en el valor inicial (x_0, y_0, z_0) .

Una vez introducida la nueva clase de métodos, debemos estudiar sus condiciones de orden. Como de costumbre, esto se hace calculando los desarrollos de Taylor de la solución exacta y de la solución numérica. Para ello, obtendremos una relación entre las derivadas de dichas soluciones y una serie de árboles con raíz con dos tipos distintos de vértices [8].

2.1.3. Derivadas de la solución exacta

Para esta parte, simplificaremos la notación y escribiremos las derivadas como aplicaciones multilineales. Por ejemplo, la expresión

$$\sum_{j=1}^D \sum_{k=1}^{D'} \frac{\partial^2 g}{\partial y_j \partial z_k} u_j v_k$$

la escribiremos como

$$g_{yz}(u, v),$$

lo que nos facilitará bastante las siguientes fórmulas.

Derivamos (2.11b) respecto de x para obtener $0 = g_y y' + g_z z'$ que, puesto que $y' = f$ y g_z es invertible, permite despejar z' como

$$z' = (-g_z^{-1}) g_y f. \quad (2.16)$$

Derivando respecto de x (2.11a) obtenemos

$$y'' = f_y y' + f_z z' = f_y f + f_z (-g_z^{-1}) g_y f. \quad (2.17)$$

Antes de derivar (2.16) respecto de la variable independiente x , tenemos en cuenta que, utilizando $(A^{-1}(x))' = -A^{-1}(x)A'(x)A^{-1}(x)$ (pues $0 = (AA^{-1})' = A'A^{-1} + A(A^{-1})'$) y la regla de la cadena, obtenemos la siguiente fórmula que usaremos más adelante

$$(-g_z^{-1})' u = (-g_z^{-1})(g_{zy}((-g_z^{-1})u, f) + g_{zz}((-g_z^{-1})u, (-g_z^{-1})g_y f)). \quad (2.18)$$

De esta manera, obtenemos

$$\begin{aligned} z'' &= (-g_z^{-1})g_{zy}((-g_z^{-1})g_y f, f) + (-g_z^{-1})g_{zz}((-g_z^{-1})g_y f, (-g_z^{-1})g_y f) \\ &\quad + (-g_z^{-1})g_{yy}(f, f) + (-g_z^{-1})g_{yz}(f, (-g_z^{-1})g_y f) \\ &\quad + (-g_z^{-1})g_y f_y f + (-g_z^{-1})f_z (-g_z^{-1})g_y f. \end{aligned} \quad (2.19)$$

Claramente, estas expresiones pasan a ser muy complicadas en cuanto se deriva un par de veces, por lo que será deseable una representación gráfica de los términos (2.17) y (2.19) que permita obtener expresiones gráficas de las derivadas de orden más alto. Esta representación gráfica estará basada en la construcción de un conjunto de árboles adecuado para este tipo de problema y método (ver [2], [5], [3] y [4] para otros problemas y diferentes tipos de métodos).

2.2. Condiciones de Orden

2.2.1. Árboles y Diferenciales Elementales

Ahora debemos identificar cada f con vértice negro, y una derivada k -ésima de f con k ramas ascendentes que parten del vértice negro (al usar árboles en este trabajo los representaremos de tal forma que en el nivel abajo del todo hay un único vértice). La expresión $(-g_z^{-1})g$ se identificará con un vértice blanco, y una derivada k -ésima g será identificada, de nuevo, por k ramas ascendentes que parten del vértice blanco.

Las fórmulas anteriores de y' , z' , y'' y z'' pasan a ser representadas por

$$y' = \bullet \quad (2.20a)$$

$$z' = \circ \quad (2.20b)$$

$$y'' = \begin{array}{c} \bullet \\ \diagup \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagdown \\ \bullet \end{array} \quad (2.20c)$$

$$z'' = \begin{array}{c} \bullet \\ \diagup \\ \circ \\ \diagdown \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagdown \\ \circ \\ \diagup \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagup \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagdown \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagup \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagdown \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagup \\ \bullet \end{array} \quad \begin{array}{c} \bullet \\ \diagdown \\ \bullet \end{array} \quad (2.20d)$$

La primera y cuarta expresión en (2.19) son idénticas, puesto que $g_{zy}(u, v) = g_{yz}(v, u)$, lo cual concuerda perfectamente con que los árboles correspondientes a estas expresiones son topológicamente equivalentes. El vértice que se encuentre en el nivel más bajo de un árbol lo llamaremos la raíz.

Vemos que las derivadas de y están caracterizadas por árboles con una raíz que es un vértice negro. Estos árboles los denotaremos por t o t_i , y el árbol que solo está formado por la raíz (el correspondiente a y') será denotado por τ_y . Las derivadas de z están caracterizadas por árboles cuya raíz es un vértice blanco. Estos árboles los denotaremos por u o u_i , y el árbol correspondiente a z' lo denotamos por τ_z .

A continuación damos una definición precisa del conjunto de árboles con los que vamos a trabajar.

Definición 2.2 Sea $DAT = DAT_y \cup DAT_z$ el conjunto de árboles (con raíz diferencial algebraica) definidos de forma recursiva por

a) $\tau_y \in DAT_y, \tau_z \in DAT_z$;

b) $[t_1, \dots, t_m, u_1, \dots, u_n]_y \in DAT_y$ si $t_1, \dots, t_m \in DAT_y$ y $u_1, \dots, u_n \in DAT_z$;

c) $[t_1, \dots, t_m, u_1, \dots, u_n]_z \in DAT_z$ si $t_1, \dots, t_m \in DAT_y, u_1, \dots, u_n \in DAT_z$, y el par (m, n) es distinto de $(0, 1)$.

Aquí $[t_1, \dots, t_m, u_1, \dots, u_n]_y$ y $[t_1, \dots, t_m, u_1, \dots, u_n]_z$ representan $(m + n)$ -uplas desordenadas de árboles de DAT.

La representación gráfica de estos árboles es la siguiente: si conectamos las raíces de $t_1, \dots, t_m, u_1, \dots, u_n$ por $m + n$ ramas a un nuevo vértice negro (la nueva raíz), obtenemos $[t_1, \dots, t_m, u_1, \dots, u_n]_y$; mientras que si lo conectamos a un vértice blanco, obtenemos $[t_1, \dots, t_m, u_1, \dots, u_n]_z$. Por ejemplo, los dos árboles de la Figura 2.1 se pueden escribir como $[\tau_y]_y$ y $[\tau_z, \tau_z]_z$.

La condición c) implica que si un nodo blanco no tiene ramificaciones, el nodo que está encima de él tiene que ser necesariamente negro. Por otro lado, la estructura de τ_y y τ_z y las definiciones recursivas b) y c) garantizan que los árboles DAT no tienen nodos blancos finales.

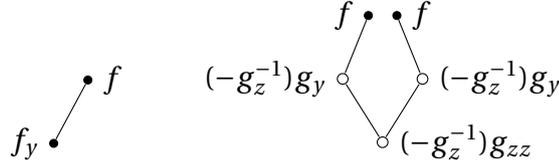


Figura 2.1: Ejemplos de representación gráfica de diferenciales elementales.

Definición 2.3 El orden de un árbol $t \in \text{DAT}_y$ o $u \in \text{DAT}_z$, denotado por $\rho(t)$ o $\rho(u)$, es el número de vértices negros que tiene el árbol.

Vemos en (2.20a)-(2.20d) que con esta definición de orden, el orden de un árbol coincide con el orden de la derivada de y o z en cuya representación aparece.

A continuación daremos una definición recursiva de la relación uno a uno de los árboles en (2.20a)-(2.20d) y las expresiones en (2.17) y (2.19).

Definición 2.4 Las diferenciales elementales $F(t)$ (o $F(u)$) correspondientes a los árboles de DAT están definidas de la siguiente forma:

a) $F(\tau_y) = f;$

b) $F(\tau_z) = (-g_z^{-1})g_y f;$

c) $F(t) = \frac{\partial^{m+n} f}{\partial y^m \partial z^n}(F(t_1), \dots, F(t_m), F(u_1), \dots, F(u_n)),$

si $t = [t_1, \dots, t_m, u_1, \dots, u_n]_y \in \text{DAT}_y$;

d) $F(u) = (-g_z)^{-1} \frac{\partial^{m+n} g}{\partial y^m \partial z^n}(F(t_1), \dots, F(t_m), F(u_1), \dots, F(u_n)),$

si $u = [t_1, \dots, t_m, u_1, \dots, u_n]_z \in \text{DAT}_z$.

A causa de la simetría de las derivadas parciales, esta definición no se ve afectada por la permutación de $t_1, \dots, t_m, u_1, \dots, u_n$ y, por lo tanto, las funciones $F(t)$ y $F(u)$ están bien definidas.

Por ejemplo, el primer árbol de la Figura 2.1 se corresponde con $F([\tau_y]_y)$, que se corresponde a su vez con el primer sumando de (2.17). El segundo árbol de la Figura 2.1 se corresponde con $F([\tau_z, \tau_z]_z)$, que se corresponde a su vez con el segundo sumando de (2.19).

2.2.2. Desarrollo de Taylor de la solución exacta

Para entender mejor el proceso de obtención de (2.17) y (2.19) estudiaremos la derivación de una diferencial elemental respecto de nuestra variable independiente x . Por la regla de Leibniz, la diferenciación de $F(t)$ (o $F(u)$) nos da una suma de nuevas diferenciales elementales asociadas a árboles de DAT que son obtenidos a partir de t (o de u) por las siguientes cuatro reglas:

- I) añadir a un vértice del árbol una rama con τ_y (derivada de $F(t)$ o $F(u)$ respecto de y , y la incorporación del factor $y' = f$);
- II) añadir a un vértice del árbol una rama con τ_z (derivada de $F(t)$ o $F(u)$ respecto de z , y la incorporación del factor $z' = (-g_z^{-1})g_y f$);
- III) dividir un vértice blanco en dos nuevos vértices blancos (unidos por una nueva rama) y pegar al vértice que se sitúe más abajo una nueva rama con τ_y ;
- IV) como en el anterior, dividir un vértice blanco en dos nuevos vértices blancos (unidos por una nueva rama), pero esta vez pegamos al vértice que se sitúe más abajo una rama con τ_z .

La tercera y cuarta regla corresponden a derivar $(-g_z^{-1})$ como en (2.18), el primer sumando de (2.18) se corresponde con la regla III mientras que el segundo sumando de (2.18) se corresponde con la regla IV.

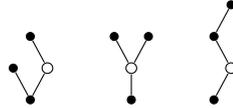
Observamos que la diferenciación de un árbol de orden q (o, de manera más precisa, de su correspondiente diferencial elemental) genera varios árboles de orden $q + 1$.

Por ejemplo, partiendo del árbol

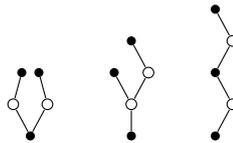


tenemos lo siguiente:

aplicar la regla I equivale a obtener los árboles



aplicar la regla II equivale a obtener los árboles



aplicar la regla III genera el árbol



y, finalmente, la regla IV equivale a generar el árbol



Algunos de estos árboles aparecen varias veces en la derivada (como el primer y cuarto árbol para z'' en (2.20d)). Para distinguir todos estos árboles, indicamos el orden de creación de los vértices negros mediante etiquetas. Ya que en el proceso de diferenciación, el nuevo vértice negro es siempre un vértice final del árbol, el etiquetado deberá, obligatoriamente, ser creciente desde la raíz hacia arriba en cada rama. Dicho etiquetado es lo que llamamos *etiquetado monótono*.

En la Figura 2.2 se muestran dos etiquetados monótonos distintos de un mismo árbol.

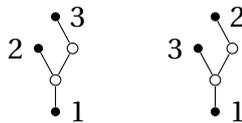


Figura 2.2: Ejemplo de dos etiquetados monótonos distintos de un mismo árbol.

Definición 2.5 Un árbol $t \in \text{DAT}_y$ (o $u \in \text{DAT}_z$) junto con un etiquetado monótono de sus vértices negros se conoce como árbol monótonamente etiquetado. El conjunto de todos los árboles monótonamente etiquetados se denota por LDAT_y (si la raíz es negra), LDAT_z (si la raíz es blanca) y $\text{LDAT} = \text{LDAT}_y \cup \text{LDAT}_z$.

La Definición 2.3 (orden de un árbol) y la Definición 2.4 (diferenciales elementales) pueden ser extendidas de manera natural para árboles monótonamente etiquetados.

Tras estos preliminares, podemos escribir las derivadas de la solución exacta como sigue:

Teorema 2.1 (Roche, 1988). Para la solución exacta de (2.11a)-(2.11b), tenemos que:

$$y^{(q)}(x_0) = \sum_{\substack{t \in \text{LDAT}_y \\ \varrho(t) = q}} F(t)(y_0, z_0) = \sum_{\substack{t \in \text{DAT}_y \\ \varrho(t) = q}} \alpha(t) F(t)(y_0, z_0), \quad (2.21a)$$

$$z^{(q)}(x_0) = \sum_{\substack{u \in \text{LDAT}_z \\ \varrho(u) = q}} F(u)(y_0, z_0) = \sum_{\substack{u \in \text{DAT}_z \\ \varrho(u) = q}} \alpha(u) F(u)(y_0, z_0). \quad (2.21b)$$

Los coeficientes enteros $\alpha(t)$ y $\alpha(u)$ indican el número de formas distintas de etiquetar monótonamente un árbol.

Demostración. Para $q = 1$ y $q = 2$ tan solo es (2.11a), (2.16), (2.17) y (2.19).

Para un q general, el proceso anterior de derivación de diferenciales elementales asociadas a árboles genera todos los elementos de LDAT , cada uno una sola vez. Si tomamos la suma sobre DAT_y y DAT_z , hay que añadir los factores $\alpha(t)$ y $\alpha(u)$.

2.2.3. Desarrollo de Taylor de la solución numérica

Lo que buscamos ahora es probar un teorema análogo al Teorema 2.1 pero para la solución numérica generada con el método de Rosenbrock (2.14a)-(2.14c). De la expresión (2.14c), tenemos que

$$\frac{d^q}{dh^q} y_1(0) = \sum_{i=1}^s b_i \frac{d^q}{dh^q} k_i(0), \quad \frac{d^q}{dh^q} z_1(0) = \sum_{i=1}^s b_i \frac{d^q}{dh^q} l_i(0). \quad (2.22)$$

Usando la regla de Leibniz, que nos dice que

$$\frac{d^q}{dh^q} (h\varphi(h))|_{h=0} = q \frac{d^{q-1}}{dh^{q-1}} (\varphi(h))|_{h=0},$$

y derivando la primera línea de (2.14a) respecto de h , tenemos que

$$\begin{aligned} \frac{d^q}{dh^q} k_i &= q \frac{d^{q-1}}{dh^{q-1}} (f(v_i, w_i)) \\ &+ q(f_y)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^{q-1}}{dh^{q-1}} k_j + q(f_z)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^{q-1}}{dh^{q-1}} l_j, \end{aligned} \quad (2.23)$$

donde, cuando escribimos $(f_y)_0$ y $(f_z)_0$ nos referimos a que las derivadas están evaluadas en el punto (y_0, z_0) y todas las derivadas respecto de h están evaluadas en $h = 0$.

Para la segunda línea de (2.14a), dividiremos por h antes de derivar, pues el lado izquierdo de la expresión es 0, por lo que simplificamos el factor h del lado derecho. Esto nos da, para $h = 0$,

$$0 = \frac{d^q}{dh^q} (g(v_i, w_i)) + q(g_y)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^q}{dh^q} k_j + q(g_z)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^q}{dh^q} l_j. \quad (2.24)$$

En el caso en el que no dividiéramos por h y lo hiciéramos como para la primera línea de (2.14a), obtendríamos

$$0 = p \frac{d^{p-1}}{dh^{p-1}} (g(v_i, w_i)) + p(g_y)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^{p-1}}{dh^{p-1}} k_j + p(g_z)_0 \sum_{j=1}^i \gamma_{ij} \frac{d^{p-1}}{dh^{p-1}} l_j,$$

que es lo mismo tomando $p = q + 1$ y dividiendo por p .

Para calcular las derivadas de f y g necesitaremos usar la Fórmula de Faà di Bruno.

Lema 2.1 (Fórmula de Faà di Bruno). *Para $q \geq 1$, tenemos que*

$$(f^J(g))^{(q-1)} = \sum_{u \in \text{LS}_q} \sum_{k_1, \dots, k_m} f_{k_1, \dots, k_m}^J(g) (g^{k_1})^{(\delta_1)} \dots (g^{k_m})^{(\delta_m)},$$

donde LS_q denota el conjunto de árboles especiales etiquetados de orden q que no tienen ramificaciones salvo en la raíz.

Para $u \in \text{LS}_q$, m es el número de ramas que salen de la raíz y $\delta_1, \dots, \delta_m$ es el número de nodos en cada uno de las m ramas, tal que $q = 1 + \delta_1 + \dots + \delta_m$.

Aplicando el Lema 2.1 para calcular las derivadas de la función f , tenemos

$$\frac{d^{q-1}}{dh^{q-1}} (f(v_i, w_i)) = \sum \frac{\partial^{m+n} f(v_i, w_i)}{\partial y^m \partial z^n} (v_i^{(\mu_1)}, \dots, v_i^{(\mu_m)}, w_i^{(\nu_1)}, \dots, w_i^{(\nu_n)}), \quad (2.25)$$

donde, como en el Lema 2.1, la suma es sobre todos los árboles especiales LDAT_y de orden q . Estos árboles son árboles monótonamente etiquetados,

$[t_1, \dots, t_m, u_1, \dots, u_n]$ donde t_j y u_j no tienen ramificaciones y todos los vértices son negros con la excepción de las raíces de u_1, \dots, u_n . Los enteros μ_j y ν_j son el orden de t_j y u_j respectivamente. Satisfacen que $\mu_1 + \dots + \mu_m + \nu_1 + \dots + \nu_n = q-1$, al igual que en el Lema 2.1.

De manera similar, aplicando el Lema 2.1 a g obtenemos

$$\frac{d^{q-1}}{dh^{q-1}}(g(v_i, w_i)) = \sum \frac{\partial^{m+n} g(v_i, w_i)}{\partial y^m \partial z^n} (v_i^{(\mu_1)}, \dots, v_i^{(\mu_m)}, w_i^{(\nu_1)}, \dots, w_i^{(\nu_n)}) + g_z(v_i, w_i) w_i^{(q)}, \quad (2.26)$$

donde el sumatorio es sobre todos los árboles especiales LDAT_z de orden q . Están definidos al igual que antes pero tenemos vértices blancos. Los enteros μ_j y ν_j satisfacen $\mu_1 + \dots + \mu_m + \nu_1 + \dots + \nu_n = q$. El término que contiene a g_z lo escribimos aparte, pues, por como definimos los árboles $\text{LDAT}_z, [u_1]_z$ no es un árbol admisible (apartado c) de la Definición 2.2).

Ahora ya podemos comenzar a calcular las derivadas de k_i y l_i . Para ello, nos conviene introducir la notación

$$\beta_{ii} = \gamma, \quad 1 \leq i \leq s; \quad \beta_{ij} = \alpha_{ij} + \gamma_{ij}, \quad 1 \leq j < i \leq s. \quad (2.27)$$

Introducimos esta notación pues los coeficientes que acompañan a $[k_j, l_j]^T$ en (2.14a) y (2.14b) son los α_{ij} y γ_{ij} .

Además, necesitaremos la inversa de la matriz (β_{ij}) , cuyos coeficientes denotaremos por ω_{ij} :

$$(\omega_{ij})_{1 \leq i, j \leq s} = (\beta_{ij})_{1 \leq i, j \leq s}^{-1}. \quad (2.28)$$

Definición 2.6 Los pesos elementales $\Phi_i(t)$ y $\Phi_i(u)$ asociados al método vienen dados por $\Phi_i(\tau_y) = 1$, $\Phi_i(\tau_z) = 1$ y

$$\Phi_i(t) = \begin{cases} \sum_{\mu_1, \dots, \mu_m, \nu_1, \dots, \nu_n} \alpha_{i\mu_1} \cdots \alpha_{i\mu_m} \alpha_{i\nu_1} \cdots \alpha_{i\nu_n} \\ \quad \cdot \Phi_{\mu_1}(t_1) \cdots \Phi_{\mu_m}(t_m) \Phi_{\nu_1}(u_1) \cdots \Phi_{\nu_n}(u_n) \\ \text{si } t = [t_1, \dots, t_m, u_1, \dots, u_n]_y \quad \text{y } m+n \geq 2, \\ \sum_j \beta_{ij} \Phi_j(t_1) \quad \text{si } t = [t_1]_y, \\ \sum_j \beta_{ij} \Phi_j(u_1) \quad \text{si } t = [u_1]_y, \end{cases}$$

$$\Phi_i(u) = \begin{cases} \sum_{j, \mu_1, \dots, \mu_m, \nu_1, \dots, \nu_n} \omega_{ij} \alpha_{j\mu_1} \cdots \alpha_{j\mu_m} \alpha_{j\nu_1} \cdots \alpha_{j\nu_n} \\ \quad \cdot \Phi_{\mu_1}(t_1) \cdots \Phi_{\mu_m}(t_m) \Phi_{\nu_1}(u_1) \cdots \Phi_{\nu_n}(u_n) \\ \text{si } u = [t_1, \dots, t_m, u_1, \dots, u_n]_z \quad \text{y } m+n \geq 2, \\ \Phi_i(t_1) \quad \text{si } u = [t_1]_z. \end{cases}$$

Por ejemplo el peso elemental del primer árbol de la Figura 2.1 es $\sum_j \beta_{ij}$, mientras que el peso elemental del segundo árbol representado de la misma forma es $\sum_{i,j,k,l,m,n,p,q} \omega_{ij} \alpha_{jk} \omega_{kl} \beta_{lq} \alpha_{jm} \omega_{mn} \beta_{np}$.

Definición 2.7 La función de densidad $\gamma(t)$ y $\gamma(u)$ correspondiente a los árboles de DAT está definida por $\gamma(\tau_y) = 1$, $\gamma(\tau_z) = 1$ y

$$\gamma(t) = \rho(t) \gamma(t_1) \cdots \gamma(t_m) \gamma(u_1) \cdots \gamma(u_n) \quad \text{si } t = [t_1, \dots, t_m, u_1, \dots, u_n]_y, \quad (2.29a)$$

$$\gamma(u) = \gamma(t_1) \cdots \gamma(t_m) \gamma(u_1) \cdots \gamma(u_n) \quad \text{si } u = [t_1, \dots, t_m, u_1, \dots, u_n]_z. \quad (2.29b)$$

Por ejemplo, para la Figura 2.1, la función de densidad del primer árbol es 2, mientras que la función de densidad del segundo árbol es 1.

Teorema 2.2 Las derivadas de k_i y l_i satisfacen:

$$\frac{d^q}{dh^q} k_i(0) = \sum_{\substack{t \in \text{LDAT}_y \\ \rho(t) = q}} \gamma(t) \Phi_i(t) F(t)(y_0, z_0), \quad (2.30a)$$

$$\frac{d^q}{dh^q} l_i(0) = \sum_{\substack{t \in \text{LDAT}_z \\ \rho(u) = q}} \gamma(u) \Phi_i(u) F(u)(y_0, z_0), \quad (2.30b)$$

$$(2.30c)$$

Demostración. Por (2.14a)-(2.14c), tenemos que

$$\frac{d^\mu}{dh^\mu} v_i = \sum_{j=1}^{i-1} \alpha_{ij} \frac{d^\mu}{dh^\mu} k_j, \quad \frac{d^\nu}{dh^\nu} w_i = \sum_{j=1}^{i-1} \alpha_{ij} \frac{d^\nu}{dh^\nu} l_j. \quad (2.31)$$

Ahora, sustituimos (2.31) en (2.25),

$$\begin{aligned} \frac{d^{q-1}}{dh^{q-1}} (f(v_i, w_i)) &= \sum_{m+n \geq 2} \frac{\partial^{m+n} f(v_i, w_i)}{\partial y^m \partial z^n} \\ &\cdot \left(\sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\mu_1}}{dh^{\mu_1}} k_j, \dots, \sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\nu_1}}{dh^{\nu_1}} l_j, \dots \right), \end{aligned} \quad (2.32)$$

y, de nuevo, sustituimos esta expresión en (2.23), lo cual nos lleva a obtener (recordamos que todas las expresiones están evaluadas en $h = 0$)

$$\begin{aligned} \frac{d^q}{dh^q} k_i &= q \sum_{m+n \geq 2} \frac{\partial^{m+n} f(y_0, z_0)}{\partial y^m \partial z^n} \left(\sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\mu_1}}{dh^{\mu_1}} k_j, \dots, \sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\nu_1}}{dh^{\nu_1}} l_j, \dots \right) \\ &+ q(f_y)_0 \sum_{j=1}^i \beta_{ij} \frac{d^{q-1}}{dh^{q-1}} k_j + q(f_z)_0 \sum_{j=1}^i \beta_{ij} \frac{d^{q-1}}{dh^{q-1}} l_j. \end{aligned} \quad (2.33)$$

Realizando el mismo análisis para la segunda componente, obtenemos

$$0 = \sum_{m+n \geq 2} \frac{\partial^{m+n} g(y_0, z_0)}{\partial y^m \partial z^n} \left(\sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\mu_1}}{dh^{\mu_1}} k_j, \dots, \sum_{j=1}^{i-1} \alpha_{ij} \frac{d^{\nu_1}}{dh^{\nu_1}} l_j, \dots \right) \\ + (g_y)_0 \sum_{j=1}^i \beta_{ij} \frac{d^q}{dh^q} k_j + (g_z)_0 \sum_{j=1}^i \beta_{ij} \frac{d^q}{dh^q} l_j. \quad (2.34)$$

Al igual que en (2.25) y (2.26), los sumatorios de (2.33) y (2.34) son sobre elementos de LDAT. Además, si usamos la inversa de (β_{ij}) , podemos obtener los $\frac{d^q}{dh^q} l_i|_{h=0}$ a partir de (2.34) como

$$\frac{d^q}{dh^q} l_i = (-g_z)_0^{-1} \sum_{j=1}^i \omega_{ij} \\ \cdot \sum_{m+n \geq 2} \frac{\partial^{m+n} g(y_0, z_0)}{\partial y^m \partial z^n} \left(\sum_{k=1}^{j-1} \alpha_{jk} \frac{d^{\mu_1}}{dh^{\mu_1}} k_k, \dots, \sum_{k=1}^{j-1} \alpha_{jk} \frac{d^{\nu_1}}{dh^{\nu_1}} l_k, \dots \right) \\ + ((-g_z^{-1})g_y)_0 \frac{d^q}{dh^q} k_i. \quad (2.35)$$

La prueba de (2.30a)-(2.30b) se hace por inducción sobre q . En el caso $q = 1$ se ve claramente gracias a (2.23) y (2.24), y para q general, insertamos la hipótesis de inducción en (2.33) y (2.35). Finalmente, sustituyendo las expresiones de (2.30a)-(2.30b) en (2.22) obtenemos las derivadas de la solución numérica.

Teorema 2.3 (Roche, 1988). *La solución numérica de (2.11a)-(2.11b) calculada mediante un paso de longitud h con (2.14a)-(2.14c) satisface:*

$$\frac{d^q}{dh^q} y_1(0) = \sum_{\substack{t \in \text{LDAT}_y \\ \varrho(t) = q}} \gamma(t) \sum_{i=1}^s b_i \Phi_i(t) F(t)(y_0, z_0), \quad (2.36a)$$

$$\frac{d^q}{dh^q} z_1(0) = \sum_{\substack{u \in \text{LDAT}_z \\ \varrho(u) = q}} \gamma(u) \sum_{i=1}^s b_i \Phi_i(u) F(u)(y_0, z_0), \quad (2.36b)$$

donde los coeficientes Φ_i son como en la Definición 2.6 y los coeficientes γ son como en la Definición 2.7.

Comparando el Teorema 2.1 para las derivadas sucesivas de la solución exacta con el Teorema 2.3 para las derivadas de la solución numérica, obtenemos el siguiente teorema que proporciona las condiciones de orden para los métodos de Rosenbrock.

Teorema 2.4 Para el método de Rosenbrock (2.14a)-(2.14c), tenemos:

1. $y(x_0 + h) - y_1 = O(h^{p+1})$ si y solo si $\sum_{i=1}^s b_i \Phi_i(t) = \frac{1}{\gamma(t)}$ para cada $t \in \text{DAT}_y$,
con $\varrho(t) \leq p$
2. $z(x_0 + h) - z_1 = O(h^{q+1})$ si y solo si $\sum_{i=1}^s b_i \Phi_i(u) = \frac{1}{\gamma(u)}$ para cada $u \in \text{DAT}_z$,
con $\varrho(t) \leq q$;

donde los coeficientes Φ_i y γ vienen dados por la Definición 2.6 y la Definición 2.7, respectivamente.

Si repetimos de forma recursiva la definición de Φ_i en el Teorema 2.2, obtenemos el algoritmo que se describe a continuación.

Para construir el lado izquierdo de las condiciones de orden hay que asignar a cada vértice negro un índice, y a cada vértice blanco dos índices (uno encima del otro). Entonces, el lado izquierdo de las condiciones de orden es una suma sobre todos los índices de un producto con factores

- b_i si i es el índice de la raíz (el menor índice si la raíz es blanca);
- α_{ij} si j es hijo de i pero no es el único;
- β_{ij} si j es el único hijo de i ;
- ω_{ij} si i, j son los dos índices de un vértice blanco (i anterior a j).

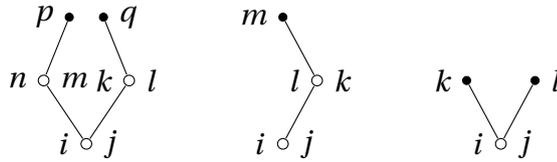


Figura 2.3: Árboles con su etiquetado

Como ejemplo, presentamos las condiciones de orden para los dos primeros árboles de la Figura 2.3.

$$\sum_{i,j,k,l,m,n,p,q} b_i \omega_{ij} \alpha_{jk} \omega_{kl} \beta_{lq} \alpha_{jm} \omega_{mn} \beta_{np} = 1, \tag{2.37}$$

$$\sum_{i,j,k,l,m} b_i \omega_{ij} \beta_{jk} \omega_{kl} \beta_{lm} = 1. \tag{2.38}$$

Las condiciones de (2.37) se pueden simplificar aun más si utilizamos el hecho de que (ω_{ij}) es la inversa de la matriz (β_{ij}) .

$$\begin{aligned}
1 &= \sum_{i,j,k,l,m,n,p,q} b_i \omega_{ij} \alpha_{jk} \omega_{kl} \beta_{lq} \alpha_{jm} \omega_{mn} \beta_{np} \\
&= \sum_{i,j,k,m,p,q} b_i \omega_{ij} \alpha_{jk} \left(\sum_l \omega_{kl} \beta_{lq} \right) \alpha_{jm} \left(\sum_n \omega_{mn} \beta_{np} \right) \\
&= \sum_{i,j,k,m,p,q} b_i \omega_{ij} \alpha_{jk} \delta_{kq} \alpha_{jm} \delta_{mp} \\
&= \sum_{i,j,k,m} b_i \omega_{ij} \alpha_{jk} \alpha_{jm}.
\end{aligned} \tag{2.39}$$

En efecto, (2.37) equivale a

$$\sum_{i,j,k,l} b_i \omega_{ij} \alpha_{jk} \alpha_{jl} = 1,$$

que es la condición de orden para el tercer árbol de la Figura 2.3.

Aprovechando de forma sistemática esta reducción, obtenemos el siguiente resultado.

Lema 2.2 *Para el método de Rosenbrock (2.14a)-(2.14c), las condiciones de orden asociadas a árboles con algunas de las siguientes características son redundantes, en el sentido de que se reducen a condiciones de orden asociadas a árboles de orden menor, o del mismo orden pero con menos vértices:*

- a) *un vértice blanco tiene una única ramificación;*
- b) *un vértice con una única ramificación viene seguido de un vértice blanco.*

Esto se debe a que el vértice blanco va a tener dos índices i, j , por lo que nos generará un elemento ω_{ij} y al tener una única ramificación nos generará también un elemento β_{jk} , y al ser una matriz inversa de la otra se cancelan. Para el caso *b*) primero tenemos un vértice (da igual de qué tipo), con una ramificación a un vértice blanco, lo que nos genera el coeficiente β_{ij} y como el vértice es blanco nos genera el coeficiente ω_{jk} , que junto con el coeficiente β_{ij} produce un elemento δ_{ik} .

En la Figura 2.3 se puede ver un ejemplo. Las condiciones de orden del primer grafo se corresponden con el caso *a*) del Lema 2.2, por lo que se pueden simplificar, resultando las condiciones de orden del tercer árbol de la Figura 2.3. Por otro lado, la condición de orden del segundo árbol es redundante por *b*). De hecho se reduce a $\sum b_i = 1$ que es la condición de orden asociada a τ_γ .

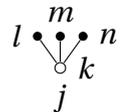
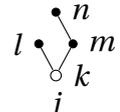
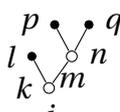
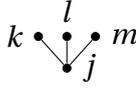
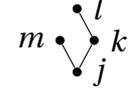
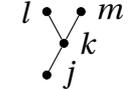
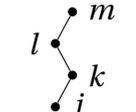
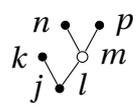
El subconjunto de DAT_γ formado por los árboles que únicamente poseen vértices negros es denotado por T .

Por lo tanto, un método de Rosenbrock de orden p tiene que satisfacer todas las condiciones de orden clásicas para métodos Runge-Kutta y, además, unas cuantas condiciones *algebraicas*. En la Tabla 2.1 se pueden ver las condiciones de orden no redundantes hasta orden 3 tanto para la variable diferencial como para la variable algebraica.

Tabla 2.1: Árboles y condiciones de orden hasta orden 3.

$\varrho(t)$	t	árbol	$\gamma(t)$	$\Phi_j(t)$
1	τ_y	$\bullet j$	1	1
2	t_{21}	$\begin{array}{c} \bullet k \\ \diagdown \\ \bullet j \end{array}$	2	$\sum_k \beta_{jk}$
	u_{21}	$\begin{array}{c} \bullet l \quad \bullet m \\ \diagdown \quad \diagup \\ \circ k \\ \diagup \\ \bullet j \end{array}$	1	$\sum_{k,l,m} \omega_{jk} \alpha_{kl} \alpha_{km}$
3	t_{31}	$\begin{array}{c} \bullet k \quad \bullet l \\ \diagdown \quad \diagup \\ \bullet j \end{array}$	3	$\sum_{k,l} \alpha_{jk} \alpha_{jl}$
	t_{32}	$\begin{array}{c} \bullet l \\ \diagdown \\ \bullet k \\ \diagdown \\ \bullet j \end{array}$	6	$\sum_{k,l} \beta_{jk} \beta_{kl}$

Tabla 2.2: Árboles y condiciones de orden hasta orden 4.

$\varrho(t)$	t	árbol	$\gamma(t)$	$\Phi_j(t)$
3	u_{31}		1	$\sum_{k,l,m,n} \omega_{jk} \alpha_{kl} \alpha_{km} \alpha_{kn}$
	u_{32}		2	$\sum_{k,l,m,n} \omega_{jk} \alpha_{kl} \alpha_{km} \beta_{mn}$
	u_{33}		1	$\sum_{k,l,m,n,p,q} \omega_{jk} \alpha_{kl} \alpha_{km} \omega_{mn} \alpha_{np} \alpha_{nq}$
4	t_{41}		4	$\sum_{k,l,m} \alpha_{jk} \alpha_{jl} \alpha_{jm}$
	t_{42}		8	$\sum_{k,l,m} \alpha_{jk} \beta_{kl} \alpha_{jm}$
	t_{43}		12	$\sum_{k,l,m} \beta_{jk} \alpha_{kl} \alpha_{km}$
	t_{44}		24	$\sum_{k,l,m} \beta_{jk} \beta_{kl} \beta_{lm}$
	t_{45}		4	$\sum_{k,l,m,n,p} \alpha_{jk} \alpha_{jl} \omega_{lm} \alpha_{mn} \alpha_{mp}$

2.2.4. Convergencia

Ahora procederemos a estudiar las propiedades de convergencia del método de Rosenbrock. Para ello necesitaremos conocer la función de estabilidad $R(z)$, que se define por la relación $y_1 = R(z)y_0$, siendo y_1 la solución numérica de la ecuación test escalar $y' = \lambda y$, $Re(\lambda) < 0$, tras un paso de longitud h partiendo de y_0 , y denotando $z = \lambda h$.

Aplicando el método de Rosenbrock con $f(y) = \lambda y$ y condición inicial $y(0) = y_0$ tenemos que

$$k_1 = hf(y_0) + h\gamma Jk_1,$$

que, tras pasar al lado izquierdo el término que contiene k_1 , y teniendo en cuenta que el jacobiano es constante igual a λ , que $f(y_0) = \lambda y_0$ y que $z = \lambda h$, obtenemos la expresión

$$(1 - z\gamma)k_1 = zy_0.$$

Ahora para k_i , $2 \leq i \leq s$, tenemos

$$k_i = h\lambda(y_0 + \sum_{j=1}^{i-1} \alpha_{ij}k_j) + h\lambda \sum_{j=1}^i \gamma_{ij}k_j,$$

que, operando como antes, y teniendo además en cuenta que $\alpha_{ij} + \gamma_{ij} = \beta_{ij}$, nos lleva a la siguiente expresión

$$(1 - z\gamma)k_i = zy_0 + \sum_{j=1}^{i-1} \beta_{ij}zk_j.$$

Si consideramos como incógnitas los k_i , obtenemos el siguiente sistema de ecuaciones lineales

$$\begin{pmatrix} 1 - z\gamma & 0 & \cdots & 0 \\ -z\beta_{21} & 1 - z\gamma & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -z\beta_{s1} & \cdots & \cdots & 1 - z\gamma \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_s \end{pmatrix} = \begin{pmatrix} zy_0 \\ zy_0 \\ \vdots \\ zy_0 \end{pmatrix}$$

cuya solución es

$$(k_1, \dots, k_s)^T = (I - zB)^{-1}zy_0\mathbb{1},$$

donde $B = (\beta_{ij})_{i,j=1}^s$ y $\mathbb{1} = [1, \dots, 1]^T \in \mathbb{R}^s$. Como $y_1 = y_0 + \sum_{i=1}^s b_i k_i$, tenemos que

$$\begin{aligned} y_1 &= y_0 + \sum_{i=1}^s b_i k_i \\ &= y_0 + b^T (I - zB)^{-1} zy_0 \mathbb{1} \\ &= [1 + zb^T (I - zB)^{-1} \mathbb{1}] y_0. \end{aligned}$$

Luego la función de estabilidad de los métodos de Rosenbrock está definida por

$$R(z) = 1 + zb^T(I - zB)^{-1}\mathbb{1},$$

donde usamos la notación $b^T = [b_1, \dots, b_s]$.

Haciendo tender z a $-\infty$ obtenemos la expresión

$$R(\infty) = 1 - b^T B^{-1}\mathbb{1} = 1 - \sum_{i,j} b_i \omega_{ij}, \quad (2.40)$$

que utilizaremos en el siguiente teorema.

Denotamos el error local del método de Rosenbrock (2.14a)-(2.14c) por

$$\delta y_h(x) = y_1 - y(x+h), \quad \delta z_h(x) = z_1 - z(x+h), \quad (2.41)$$

donde y_1, z_1 es la solución numérica obtenida tras un paso de longitud h con los valores iniciales exactos $y_0 = y(x_0), z_0 = z(x_0)$.

Teorema 2.5 *Suponemos que g_z es regular en un entorno de la solución $(y(x), z(x))$ de (2.11a)-(2.11b) y que los valores iniciales (y_0, z_0) son consistentes. Si la función de estabilidad es tal que $|R(\infty)| < 1$, y el error local satisface que*

$$\delta y_h(x) = O(h^{p+1}), \quad \delta z_h(x) = O(h^p), \quad (2.42)$$

entonces el método de Rosenbrock (2.14a)-(2.14c) es convergente de orden p ; i.e.,

$$y_n - y(x_n) = O(h^p), \quad z_n - z(x_n) = O(h^p), \quad \text{para} \quad x_n - x_0 = nh \leq Cte.$$

Demostración. Como g_z es regular, tenemos

$$\|g_z^{-1}(y, z)g(y, z)\| \leq \delta \quad (2.43)$$

para (y, z) en un entorno compacto U de la solución. El h -valor independiente de δ podemos hacerlo tan pequeño como queramos restringiendo U . También supondremos, de momento, que la solución y todos los pasos intermedios para obtenerla permanecen en este entorno. Ahora lo dividiremos en dos, en la parte a) estudiaremos la propagación del error local y en la parte b) su acumulación sobre todo el intervalo.

a) Consideramos dos pares de valores iniciales (y_0, z_0) y (\hat{y}_0, \hat{z}_0) , y aplicaremos el método a cada par (estos valores pueden ser inconsistentes, pero asumiremos que están en U). Debemos probar que

$$\|y_1 - \hat{y}_1\| \leq (1 + hL) \|y_0 - \hat{y}_0\| + hM \|z_0 - \hat{z}_0\|, \quad (2.44a)$$

$$\|z_1 - \hat{z}_1\| \leq N \|y_0 - \hat{y}_0\| + K \|z_0 - \hat{z}_0\|, \quad (2.44b)$$

donde $K < 1$. Para ello consideramos un paso de tamaño h lo suficientemente pequeño, y consideramos y_1 , z_1 , k_i y l_i como funciones de (y_0, z_0) .

Debemos probar que

$$\frac{\partial y_1}{\partial y_0} = I + O(h), \quad \frac{\partial y_1}{\partial z_0} = O(h), \quad (2.45a)$$

$$\frac{\partial z_1}{\partial y_0} = O(1), \quad \frac{\partial z_1}{\partial z_0} = R(\infty)I + O(h + \delta). \quad (2.45b)$$

El teorema del valor intermedio implica entonces (2.44a)-(2.44b).

Primero estimaremos k_i y l_i , definidos en (2.14a). Usando (2.43) calculamos l_i de la segunda línea y lo insertamos en la primera. Esto nos da, de manera sucesiva, que $k_i = O(h)$ y $l_i = O(h + \delta)$ para todas las etapas internas. Después, diferenciamos (2.14b) una vez respecto de y_0 y otra respecto de z_0 . Un análisis similar a este para k_i y l_i nos da

$$\frac{\partial k_i}{\partial y_0} = O(h), \quad \frac{\partial k_i}{\partial z_0} = O(h), \quad (2.46a)$$

$$\frac{\partial l_i}{\partial y_0} = O(1), \quad \frac{\partial l_i}{\partial z_0} = -\sum_j \omega_{ij} I + O(h + \delta), \quad (2.46b)$$

y las estimaciones (2.45a)-(2.45b) se siguen de (2.14b)-(2.14c) y (2.40).

b) Para este apartado necesitaremos usar el siguiente lema:

Lema 2.3 *Sea u_n , v_n dos sucesiones de números no negativos satisfaciendo (componente a componente)*

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} \leq \begin{pmatrix} 1 + O(h) & O(\varepsilon) \\ O(1) & \alpha + O(\varepsilon) \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix} + M \begin{pmatrix} h \\ 1 \end{pmatrix}, \quad (2.47)$$

con $0 \leq \alpha < 1$ y $M \geq 0$. Entonces, las siguientes estimaciones son sólidas para $\varepsilon \leq ch$, $h \leq h_0$ y $nh \leq \text{Const}$

$$u_n \leq C(u_0 + \varepsilon v_0 + M), \quad (2.48a)$$

$$v_n \leq C(u_0 + (\varepsilon + \alpha^n) v_0 + M). \quad (2.48b)$$

Demostración. Transformamos la matriz de (2.47) en una matriz diagonal y obtenemos así

$$\begin{pmatrix} u_n \\ v_n \end{pmatrix} \leq T^{-1} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} T \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + M \sum_{j=1}^n T^{-1} \begin{pmatrix} \lambda_1^{n-j} & 0 \\ 0 & \lambda_2^{n-j} \end{pmatrix} T \begin{pmatrix} h \\ 1 \end{pmatrix},$$

donde $\lambda_1 = 1 + O(h)$, $\lambda_2 = \alpha + O(\varepsilon)$ son los autovalores y la matriz de transformación T (formada por los autovectores correspondientes a los autovalores anteriores) satisface

$$T = \begin{pmatrix} 1 & O(\varepsilon) \\ O(1) & 1 \end{pmatrix}.$$

Las estimaciones buscadas se siguen del hecho de que $(\alpha + O(\varepsilon))^n = O(\alpha^n) + O(\varepsilon)$ para $\varepsilon \leq ch$ y $nh \leq Const$.

Como consecuencia de este lema, la propagación de los errores locales de la solución en x_n , $\delta y_h(x_{j-1})$, $\delta z_h(x_{j-1})$ se pueden acotar por

$$C(\|\delta y_h(x_{j-1})\| + (h + K^{n-j}) \|\delta z_h(x_{j-1})\|). \quad (2.49)$$

Sumando estos términos desde $j = 1$ hasta $j = n$ y usando (2.42) nos da las acotaciones para el error global, porque $\sum_{j=1}^n (h + K^{n-j}) \leq Const$.

Nuestra suposición de que la solución numérica y las fases intermedias caen en U se puede justificar fácilmente ahora por inducción sobre el número de etapas. La solución numérica permanece a una distancia de orden $O(h^p)$ de la solución exacta y, por tanto, para un h suficientemente pequeño la solución numérica seguirá en U . Esto implica que $g(y_j, z_j) = O(h^p)$ para todo j y por lo tanto también $l_i = O(h)$. Como consecuencia, (v_i, w_i) están tan cerca de la solución exacta como queremos.

2.3. Implementación

En esta sección explicaremos cómo programar el método de Rosenbrock definido en (2.14a)-(2.14c) con paso fijo para resolver EDAs de la forma (2.11a)-(2.11b) en MATLAB.

Como ya vimos antes, las ecuaciones del método de Rosenbrock corresponden al sistema

$$\begin{pmatrix} k_i \\ 0 \end{pmatrix} = h \begin{pmatrix} f(v_i, w_i) \\ g(v_i, w_i) \end{pmatrix} + h \begin{pmatrix} f_y & f_z \\ g_y & g_z \end{pmatrix} (y_0, z_0) \sum_{j=1}^i \gamma_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad (2.50a)$$

$$\begin{pmatrix} v_i \\ w_i \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{j=1}^{i-1} \alpha_{ij} \begin{pmatrix} k_j \\ l_j \end{pmatrix}, \quad (2.50b)$$

$$\begin{pmatrix} y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \sum_{i=1}^s b_i \begin{pmatrix} k_i \\ l_i \end{pmatrix}. \quad (2.50c)$$

Sabiendo esto, los parámetros de entrada requeridos en nuestro programa serán los siguientes:

- el valor inicial (y_0, z_0) sobre el que comenzaremos a aplicar nuestro método;
- el número de pasos N que debemos avanzar con el método;
- el tiempo inicial y final, que llamaremos t y T respectivamente (en general tenemos que $t = 0$, pero no tiene porqué, por eso lo pedimos). Además, como pedimos N y T , podemos calcular la longitud de paso h , pues sabemos que $Nh = T$;
- un programa, creado por el usuario, en el que el parámetro de entrada sea el par (v_i, w_i) con $i = 1, \dots, s$ y los parámetros de salida sean las correspondientes funciones f y g , definidas en (2.11a)-(2.11b), evaluadas en el punto;
- otro programa, creado por el usuario, cuyo parámetro de entrada sea el par (y_j, z_j) con $j = 1, \dots, N$ y los parámetros de salida sean los jacobianos f_y, f_z, g_y y g_z evaluados en ese punto.

Pedimos al usuario dos programas distintos, pues en cada paso llamaremos al primer programa mencionado un total de s veces, pues i va desde 1 hasta s mientras que solo es necesario llamar una única vez al segundo programa porque solo necesitamos aplicarlo al (y_p, z_p) correspondiente al paso en el que nos encontremos. Como hay que avanzar N pasos, al pedir dos programas distintos, reducimos el costo operativo en un total de $N(s - 1)$ evaluaciones del jacobiano, así que nuestro programa será más eficiente.

- Por último, pediremos los coeficientes $\alpha_{i,j}$, b_i y γ_{ij} .

Como parámetro de salida, daremos (y_n, z_n) , $n = 0, 1, 2, \dots, N$ obtenidos al avanzar los N pasos con nuestro método a partir del valor inicial (y_0, z_0) .

En cada paso de nuestro método necesitaremos calcular los vectores (k_i, l_i) con $i = 1, \dots, s$. Para ello, los despejamos de (2.14a), teniendo en cuenta que queremos expresar los (k_i, l_i) en función de todos los anteriores. Tenemos

$$k_i = hf(v_i, w_i) + h(f_y)_0 \sum_{j=1}^i \gamma_{ij} k_j + h(f_z)_0 \sum_{j=1}^i \gamma_{ij} l_j, \quad (2.51a)$$

$$0 = hg(v_i, w_i) + h(g_y)_0 \sum_{j=1}^i \gamma_{ij} k_j + h(g_z)_0 \sum_{j=1}^i \gamma_{ij} l_j. \quad (2.51b)$$

Ahora ponemos todos los términos que contengan a k_i o a l_i al lado izquierdo, y los que no, al derecho

$$k_i - h(f_y)_0 \gamma_{ii} k_i - h(f_z)_0 \gamma_{ii} l_i = hf(v_i, w_i) + h(f_y)_0 \sum_{j=1}^{i-1} \gamma_{ij} k_j + h(f_z)_0 \sum_{j=1}^{i-1} \gamma_{ij} l_j, \quad (2.52a)$$

$$-h(g_y)_0 \gamma_{ii} k_i - h(g_z)_0 \gamma_{ii} l_i = hg(v_i, w_i) + h(g_y)_0 \sum_{j=1}^{i-1} \gamma_{ij} k_j + h(g_z)_0 \sum_{j=1}^{i-1} \gamma_{ij} l_j. \quad (2.52b)$$

Si observamos bien, el sistema de ecuaciones (2.52a)-(2.52b) se puede escribir como un sistema de ecuaciones lineales de dimensión $(D + D') \times (D + D')$ de la forma

$$\begin{pmatrix} I - h(f_y)_0 \gamma & -h(f_z)_0 \gamma \\ -h(g_y)_0 \gamma & -h(g_z)_0 \gamma \end{pmatrix} \begin{pmatrix} k_i \\ l_i \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \quad (2.53)$$

con $\gamma = \gamma_{ii}$,

$$\eta_1 = hf(v_i, w_i) + (f_y)_0 \sum_{j=1}^{i-1} \gamma_{ij} k_j + (f_z)_0 \sum_{j=1}^{i-1} \gamma_{ij} l_j, \quad (2.54a)$$

$$\eta_2 = hg(v_i, w_i) + (g_y)_0 \sum_{j=1}^{i-1} \gamma_{ij} k_j + (g_z)_0 \sum_{j=1}^{i-1} \gamma_{ij} l_j. \quad (2.54b)$$

Ahora necesitaremos realizar la factorización LU, estudiada en [9], de la matriz de (2.53) para resolver el sistema. Notemos que dicha matriz es la misma para todas las etapas por lo que es precisa una única factorización LU por paso, seguida de s sustituciones progresivas y regresivas (una para cada etapa del método).

Observamos que para ello necesitamos conocer los (v_i, w_i) ; los cuales podemos obtener gracias a (2.50b). Además, tenemos que $(v_1, w_1) = (y_0, z_0)$. Por lo tanto, nuestro programa procederá como sigue

1. Para $n = 0$, comienza con nuestro valor inicial (y_0, z_0) .

Para $n = 1, \dots, N$ repetir:

2. Calcula la factorización LU.

Para $i = 1, \dots, s$ repetir:

3. Calcula (v_i, w_i) .

4. Calcula (k_i, l_i) (sustitución progresiva y regresiva).

5. Calcular (y_n, z_n) .

Capítulo 3

Resultados numéricos

En este capítulo procederemos a resolver dos ejemplos concretos de sistemas de EDAs de índice 1 mediante métodos de Rosenbrock estudiados en el Capítulo 2. Los problemas que trataremos serán el péndulo, del que ya hablamos en el Capítulo 1, y un circuito eléctrico correspondiente a un amplificador.

3.1. Problemas

A continuación describimos los dos problemas con los que se va a trabajar.

3.1.1. Péndulo matemático

El movimiento del péndulo matemático, descrito en la Figura 1.1, cuando se utilizan coordenadas cartesianas (x, y) junto a la restricción considerada en el Capítulo 1, $x^2 + y^2 - l^2 = 0$, está descrito por las ecuaciones siguientes

$$x' = u, \quad y' = v, \quad (3.1a)$$

$$u' = -\frac{Tx}{lm}, \quad v' = -\frac{T y}{lm} + g, \quad (3.1b)$$

$$0 = x^2 + y^2 - l^2, \quad (3.1c)$$

donde, recordamos, m es la masa puntual suspendida al final de una varilla sin masa, l es la longitud de dicha varilla, g es la constante de gravedad, y T es el módulo de la tensión de la varilla que mantiene la masa puntual a una distancia constante del origen. Agrupando las variables dependientes como

$$\mathbf{y} = [x, y]^T, \quad \mathbf{z} = [u, v]^T, \quad \lambda = T,$$

el sistema (3.1a)-(3.1c) corresponde a un sistema de EDAs de índice 3.

En el Capítulo 1 vimos que derivando dos veces la restricción (3.1c) respecto de nuestra variable independiente, obtenemos el sistema de ecuaciones

$$x' = u, \quad y' = v, \quad (3.2a)$$

$$u' = -\frac{Tx}{ml}, \quad v' = -\frac{Ty}{ml} + g, \quad (3.2b)$$

$$0 = m(u^2 + v^2) - Tl + gmy, \quad (3.2c)$$

que tras reagrupar las variables dependientes como

$$\mathbf{y} = [x, y, u, v]^T, \quad \mathbf{z} = T,$$

constituye un sistema de EDAs de índice 1 al que podemos aplicar los métodos de Rosenbrock estudiados en el Capítulo 2.

Las ecuaciones de nuestro sistema son, por tanto,

$$\mathbf{y}' = \left[u, v, -\frac{T}{lm}x, -\frac{T}{lm}y + g \right]^T, \quad (3.3a)$$

$$0 = m(u^2 + v^2) - Tl + gmy, \quad (3.3b)$$

con

$$\mathbf{f}(\mathbf{y}, \mathbf{z}) = \left[u, v, -\frac{T}{lm}x, -\frac{T}{lm}y + g \right]^T, \quad (3.4a)$$

$$\mathbf{g}(\mathbf{y}, \mathbf{z}) = m(u^2 + v^2) - Tl + gmy. \quad (3.4b)$$

En las simulaciones que realizaremos utilizaremos los parámetros

$$m = 1\text{kg}, \quad l = 1\text{m}, \quad g = 9,81\text{m/s}^2,$$

y los valores iniciales

$$\mathbf{y}(0) = [0, 1, 6, 0]^T, \quad \mathbf{z}(0) = 6,19.$$

La elección de $\mathbf{z}(0)$ ha sido realizada para partir de condiciones iniciales consistentes con la restricción (3.4b), es decir, $\mathbf{z}(0)$ se ha calculado como

$$\mathbf{z}(0) = \frac{m(u(0)^2 + v(0)^2) + gmy(0)}{l}.$$

3.1.2. Amplificador

A continuación consideramos el circuito representado en la Figura 3.1, obtenida en [5], correspondiente a un amplificador, donde $U_e(t)$ es el voltaje de

entrada, $U_b = 6$ es el voltaje operativo, $U_i(t)$, con $i = 1, \dots, 5$, son los voltajes en los nodos $1, \dots, 5$, y $U_5(t)$ es el voltaje de salida.

La corriente a través de una resistencia satisface, según la Ley de Ohm, $I = U/R$, mientras que la corriente a través de un condensador satisface que $I = C \cdot dU/dt$, donde R y C son constantes y U es el voltaje, que depende del tiempo t . El transistor actúa como amplificador, de forma que la corriente del nodo 4 al nodo 3 es 99 veces mayor que la del nodo 2 al nodo 3, y depende de la diferencia de voltajes $U_3 - U_2$ de forma no lineal.

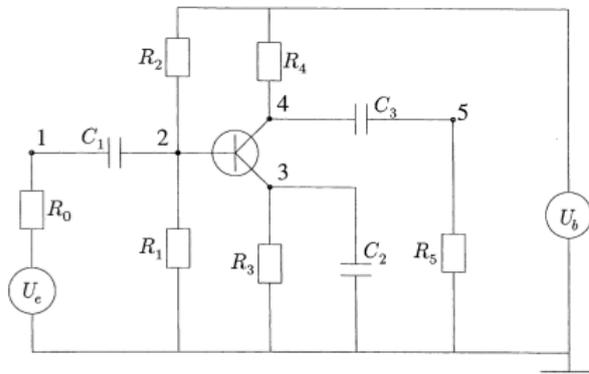


Figura 3.1: Amplificador con un transistor.

La Ley de Kirchhoff nos dice que la suma de las corrientes que hay en un nodo es nula, por lo que aplicando dicha ley a cada uno de los nodos considerados obtenemos las siguientes ecuaciones diferenciales:

$$\text{Nodo 1: } \frac{U_e}{R_0} - \frac{U_1}{R_0} + C_1(U_2' - U_1') = 0, \quad (3.5a)$$

$$\text{Nodo 2: } \frac{U_b}{R_2} - U_2 \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + C_1(U_1' - U_2') - 0,01f(U_2 - U_3) = 0, \quad (3.5b)$$

$$\text{Nodo 3: } f(U_2 - U_3) - \frac{U_3}{R_3} - C_2U_3' = 0, \quad (3.5c)$$

$$\text{Nodo 4: } \frac{U_b}{R_4} - \frac{U_4}{R_4} + C_3(U_5' - U_4') - 0,99f(U_2 - U_3) = 0, \quad (3.5d)$$

$$\text{Nodo 5: } -\frac{U_5}{R_5} + C_3(U_4' - U_5') = 0. \quad (3.5e)$$

Los valores de las constantes que hemos utilizado son los que aparecen en [7]

Sustituyendo ahora estas expresiones en (3.5a)-(3.5e), obtenemos el sistema

$$\frac{U_e}{R_0} - \frac{z_1}{R_0} - C_1 y_1' = 0, \quad (3.10a)$$

$$\frac{U_b}{R_2} + (y_1 - z_1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + C_1 y_1' - 0,01 f(z_1 - y_1 - y_2) = 0, \quad (3.10b)$$

$$f(z_1 - y_1 - y_2) - \frac{y_2}{R_3} - C_2 y_2' = 0, \quad (3.10c)$$

$$\frac{U_b}{R_4} - \frac{z_2}{R_4} - C_3 y_3' - 0,99 f(z_1 - y_1 - y_2) = 0, \quad (3.10d)$$

$$\frac{y_3 - z_2}{R_5} + C_3 y_3' = 0. \quad (3.10e)$$

Ahora, si despejamos y_1' , y_2' , y_3' de la primera, tercera y quinta ecuación respectivamente, y tomamos $\mathbf{y} = [y_1, y_2, y_3]^T$, $\mathbf{z} = [z_1, z_2]^T$, obtenemos un sistema diferencial de la forma (3.9a), con

$$f(t, \mathbf{y}, \mathbf{z}) = \left(\frac{U_e - z_1}{R_0 C_1}, \frac{f(z_1 - y_1 - y_2)}{C_2} - \frac{y_2}{C_2 R_3}, \frac{z_2 - y_3}{C_3 R_5} \right)^T. \quad (3.11)$$

Por último, sumando en (3.10a)-(3.10e) la primera con la segunda ecuación, y la cuarta con la quinta, obtenemos un par de restricciones algebraicas de la forma (3.9b), con

$$g_1(t, \mathbf{y}, \mathbf{z}) = \left(\frac{U_e - z_1}{R_0} + \frac{U_b}{R_2} + (y_1 - z_1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right) - \frac{f(z_1 - y_1 - y_2)}{100} \right), \quad (3.12a)$$

$$g_2(t, \mathbf{y}, \mathbf{z}) = \left(\frac{U_b - z_2}{R_4} - 0,99 f(z_1 - y_1 - y_2) + \frac{y_3 - z_2}{R_5} \right). \quad (3.12b)$$

El sistema (3.11), (3.12a)-(3.12b) es un sistema de la forma (3.9a)-(3.9b), pero tal y como está escrito es un sistema de EDAs de índice 1 no autónomo, pues U_e depende de la variable independiente t . Para poder utilizar de manera directa el método de Rosnebrock programado, necesitamos convertirlo en un sistema de ecuaciones diferenciales algebraicas autónomo. Para ello introducimos la nueva variable $y_4 = t$, y llamando $\tilde{\mathbf{y}} = [y_1, y_2, y_3, y_4]^T$,

$$f(\tilde{\mathbf{y}}, \mathbf{z}) = \left[\frac{U_e(y_4) - z_1}{R_0 C_1}, \frac{f(z_1 - y_1 - y_2)}{C_2} - \frac{y_2}{C_2 R_3}, \frac{z_2 - y_3}{C_3 R_5}, 1 \right]^T, \quad (3.13a)$$

$$g_1(\tilde{\mathbf{y}}, \mathbf{z}) = \frac{U_e(y_4) - z_1}{R_0} + \frac{U_b}{R_2} + (y_1 - z_1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right) - \frac{f(z_1 - y_1 - y_2)}{100}, \quad (3.13b)$$

$$g_2(\tilde{\mathbf{y}}, \mathbf{z}) = \frac{U_b - z_2}{R_4} - 0,99 f(z_1 - y_1 - y_2) + \frac{y_3 - z_2}{R_5}, \quad (3.13c)$$

tenemos ya un sistema de EDAs de índice 1 autónomo, con $\tilde{\mathbf{y}}' = f(\tilde{\mathbf{y}}, \mathbf{z})$, $0 = g_1(\tilde{\mathbf{y}}, \mathbf{z})$ y $0 = g_2(\tilde{\mathbf{y}}, \mathbf{z})$.

Los valores iniciales consistentes deben satisfacer las ecuaciones $\varphi_1(u) + \varphi_2(u) = 0$ y $\varphi_4(u) + \varphi_5(u) = 0$, donde $\varphi_1, \dots, \varphi_5$ son las componentes de $\varphi(u)$ en (3.8).

Si tomamos $U_2(0) = U_3(0)$, tenemos que $f(U_2(0) - U_3(0)) = 0$, y deberían satisfacer

$$0 = \frac{U_1(0)}{R_0} - \frac{U_e(0)}{R_0} - \frac{U_b}{R_2} + U_2(0) \left(\frac{1}{R_1} + \frac{1}{R_2} \right),$$

$$0 = -\frac{U_b}{R_4} + \frac{U_4(0)}{R_4} + \frac{U_5(0)}{R_5}.$$

Como $U_e(0) = 0$, tomando además $U_1(0) = U_5(0) = 0$, basta despejar $U_2(0)$ y $U_4(0)$ de las relaciones anteriores, lo que nos lleva a tomar los valores iniciales consistentes

$$U_1(0) = 0, \quad U_2(0) = U_3(0) = \frac{U_b R_1}{R_1 + R_2}, \quad U_4(0) = U_b, \quad U_5(0) = 0.$$

3.2. Métodos

Para obtener soluciones numéricas de los problemas planteados en la sección anterior se han implementado en MATLAB dos métodos de Rosenbrock de la forma (2.4a)-(2.4b), que permiten la integración numérica de los sistemas de EDAs de índice 1. En el Apéndice podemos ver el código de los programas realizados con los que se han obtenido los resultados numéricos que se presentarán en la Sección 3.3.

Se ha implementado el método de orden 3, conocido como ROWDA3 [8], cuyos coeficientes, de acuerdo con la Definición 2.1, se encuentran recogidos en la Tabla 3.1.

También se ha implementado un método de Rosenbrock de orden 4 utilizado en [8], y cuyos coeficientes aparecen en la Tabla 3.2.

Tabla 3.1: Coeficientes del método conocido como ROWDA3.

$$\gamma = 0,435866521508459$$

$$b_1 = 0,3197278911564624$$

$$b_2 = 0,7714777906171382$$

$$b_3 = -0,09120568177360061$$

$$\alpha_{21} = 0,7$$

$$\alpha_{31} = 0,7$$

$$\alpha_{32} = 0$$

$$\gamma_{21} = 0,1685887625570998$$

$$\gamma_{31} = 4,943922277836421$$

$$\gamma_{32} = 1$$

Tabla 3.2: Coeficientes de un método de Rosenbrock de orden 4.

$$\gamma = 0,70751226521$$

$$b_1 = 0,2523628037277470$$

$$b_2 = -0,2209698738798533$$

$$b_3 = -0,2256411840923124$$

$$b_4 = 0,3179133966013711$$

$$b_5 = 0,8763348576430476$$

$$\alpha_{21} = 1,233311380872013$$

$$\alpha_{31} = 0,6535453813273382$$

$$\alpha_{32} = 0,2295950748229277$$

$$\alpha_{41} = 2,681059792907162$$

$$\alpha_{42} = -1,554590259558157$$

$$\alpha_{43} = -0,9682496302574051$$

$$\alpha_{51} = -0,6021422614217772$$

$$\alpha_{52} = 0,2994399056322287$$

$$\alpha_{53} = 0,4792338650945191$$

$$\alpha_{54} = 0,8010415023569842$$

$$\gamma_{21} = -1,818714325256271$$

$$\gamma_{31} = -0,4589460040608732$$

$$\gamma_{32} = 0,3613323897595465$$

$$\gamma_{41} = -3,424045164556574$$

$$\gamma_{42} = 1,553491448551290$$

$$\gamma_{43} = 1,249712740807497$$

$$\gamma_{51} = -0,2261466054228607$$

$$\gamma_{52} = -0,3882326103473952$$

$$\gamma_{53} = -0,3589041115714489$$

$$\gamma_{54} = -0,01860845389367294$$

3.3. Resultados

En esta sección procederemos a mostrar los resultados numéricos obtenidos cuando los problemas planteados en la Sección 3.1 se integran numéricamente mediante los métodos de Rosenbrock programados en MATLAB.

3.3.1. Péndulo matemático

Tanto las ecuaciones del movimiento del péndulo como los valores de los parámetros y de las condiciones iniciales utilizadas se han fijado en la Subsección 3.1.1. El tiempo de integración que se ha considerado es $[0,5]$ y con longitud de paso $h = 0,05$ los resultados producen errores inferiores a una décima.

En la Figura 3.2 puede observarse la evolución de la variable z (que, como ya se ha indicado, se corresponde con la tensión) con respecto al tiempo.

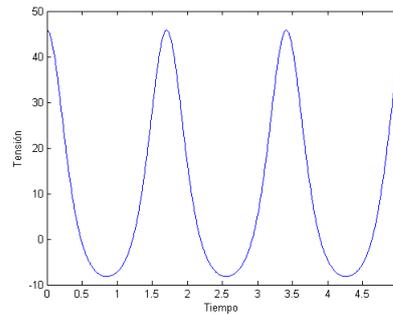


Figura 3.2: Evolución de la tensión con el tiempo.

En la Figura 3.3 vemos dos gráficas. La gráfica superior se corresponde con las dos primeras componentes de y , relativas a la posición del péndulo. La línea de color azul representa la abscisa del péndulo, mientras que la de color rojo representa la ordenada. La gráfica inferior se corresponde con la tercera y la cuarta componente de y , relativas a la velocidad. La línea de color azul representa la componente de la velocidad en el eje x , mientras que la de color rojo representa la componente de la velocidad en el eje y .

Ambas figuras han sido obtenidas mediante el método de Rosenbrock de orden 4 con longitud de paso $h = 0,05$.

Otro aspecto que nos interesa comprobar es si se mantiene con el tiempo la condición algebraica $g(\mathbf{y}, \mathbf{z}) = 0$. En la gráfica superior de la Figura 3.4, que se muestra en escala logarítmica para el eje y , podemos observar el valor absoluto de la diferencia entre el valor que toma la condición algebraica a lo largo de la solución numérica y el que debería tomar (constante igual a $g(\mathbf{y}_0, \mathbf{z}_0) = 0$).

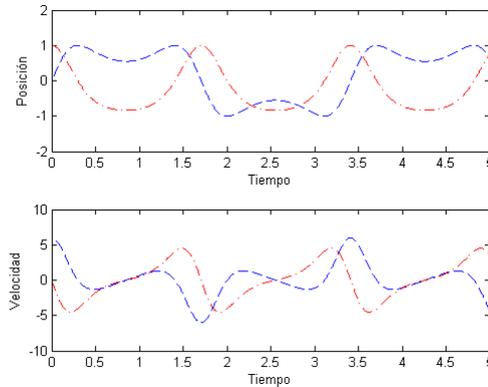


Figura 3.3: Evolución de la posición y velocidad con el tiempo.

Este valor está por debajo de 10^{-4} para el método de Rosenbrock de orden 3 (en negro en la gráfica) y por debajo de 10^{-4} también, aunque algo menor, para el método de Rosenbrock de orden 4 (en rojo en la gráfica).

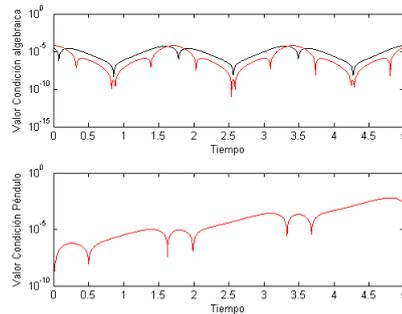


Figura 3.4: Evolución de la condición algebraica con el tiempo.

Hay que tener en cuenta que la condición del péndulo original es $x^2 + y^2 - l = 0$, es decir, que la longitud de la varilla sea siempre la misma. En la gráfica inferior de la Figura 3.4 podemos observar la evolución con el tiempo del valor de $x^2 + y^2 - l$ a lo largo de la solución numérica calculada con el método de Rosenbrock de orden 4. Vemos que el error crece al crecer el tiempo, comenzando en valores inferiores a 10^{-6} y llegando hasta 10^{-2} al final de la integración.

Como ya vimos en la Sección 1.2, el problema del péndulo se puede resolver considerando el péndulo como una EDO de segundo orden (1.18a)-(1.18c). Utilizando la función de MATLAB ode45, hemos creado la Figura 3.5 en la que puede observarse la evolución con respecto al tiempo de las dos componentes de la posición, en la parte superior la de x y en la parte inferior la de y . Pa-

ra ambas gráficas, la línea roja se corresponde con el método ode45, mientras que la línea azul se corresponde con el método de Rosenbrock de orden 4, con $h = 0,05$. No se observa apenas diferencia entre las dos soluciones numéricas obtenidas por ambos procedimientos.

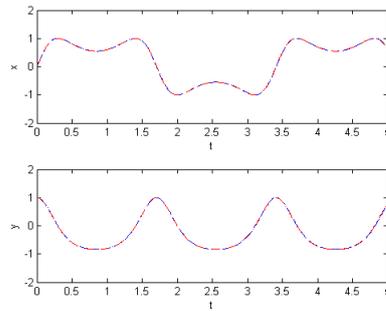


Figura 3.5: Problema del péndulo como EDO y como sistema de EDAs.

Pasamos por último a analizar, valiéndonos de la Figura 3.6, el orden de los métodos de Rosenbrock al integrar numéricamente el péndulo para las variables diferenciales y y la variable algebraica z .

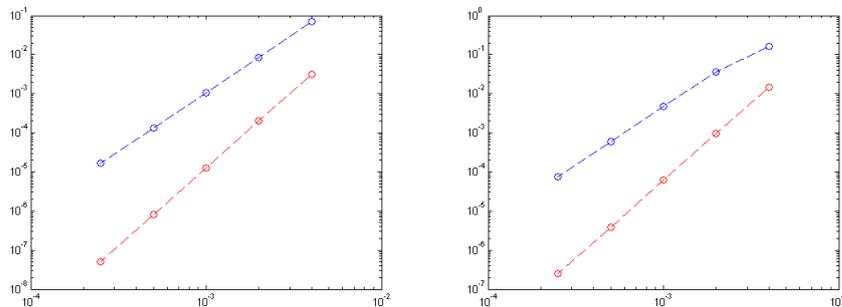


Figura 3.6: Error en las variables diferenciales frente a la longitud de paso h .

Como solución de referencia se ha tomado la obtenida por la función ode45 de MATLAB, con tolerancia para el error relativo de 10^{-12} y de 10^{-12} para el error absoluto, considerando las ecuaciones de segundo del péndulo (1.17) reescritas como un sistema de EDOs de primer orden. Las cinco longitudes de paso utilizadas van desde $h = 5/1000$ (la más grande), hasta $h = 5/16000$, reduciéndose a la mitad en cada ejecución.

La gráfica de la izquierda muestra en escala doblemente logarítmica el error de las dos primeras componentes de la solución numérica al final de la integración, medido en la norma infinito de \mathbb{R}^2 . En la gráfica de la derecha se muestra,

en escala doblemente logarítmica, el valor absoluto del error en la variable algebraica (la tensión T) al final de la integración.

En las dos gráficas de la Figura 3.6, la línea azul, correspondiente al método de Rosenbrock de orden 3, tiene pendiente aproximadamente 3, mientras que la línea roja, correspondiente al método de Rosenbrock de orden 4, tiene pendiente aproximadamente 4. En ambos casos observamos que la pendiente de las líneas coincide con el orden del método con el que se han generado.

Si el orden del método es p , el error global se comportará aproximadamente como $E(h) \approx Kh^p$. Por tanto la pendiente aproximada de la recta entre dos puntos correspondientes a pasos $h_1 < h_2$ es

$$\frac{\log(E(h_2)) - \log(E(h_1))}{\log(h_2) - \log(h_1)} \approx \frac{p \log(h_2) - p \log(h_1)}{\log(h_2) - \log(h_1)} = p,$$

el orden del método. Es decir, al dividir h por 2,

$$E(h/2) \approx K(h/2)^p = E(h) \frac{1}{2^p},$$

y, en consecuencia, el error global se divide por 2^p .

3.3.2. Amplificador

Tanto las ecuaciones del amplificador como los valores de los parámetros y las condiciones iniciales utilizadas en la simulación se han fijado en la Subsección 3.1.2. Siguiendo [5], el tiempo de integración que se ha considerado es $[0, 0.2]$ y se han utilizado varias longitudes de paso, aunque las gráficas que se muestran corresponden a $h = 0,0002$.

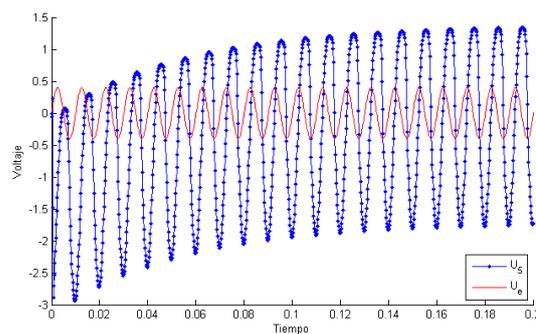


Figura 3.7: Evolución de los voltajes U_5 y U_e con el tiempo.

Utilizando el método de Rosenbrock de orden 4 descrito en la Sección 3.2, con el tiempo de integración y la longitud de paso anteriormente descritas, hemos obtenido las gráficas que se muestran en la Figura 3.7 y la Figura 3.8, donde

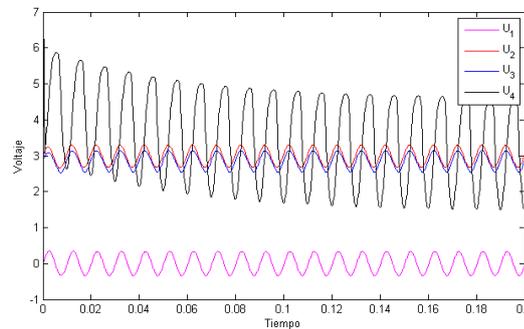


Figura 3.8: Evolución de los voltajes U_1 , U_2 , U_3 y U_4 con el tiempo.

se observa la evolución de los distintos voltajes con el tiempo. La gráfica de la Figura 3.7 coincide con la que aparece en [5, p. 379].

Observamos que se trata de un problema cuya solución es altamente oscilatoria, debido a la naturaleza del voltaje U_e fijado en (3.7), lo que hace que la longitud del paso de integración tenga que ser muy pequeña (si se compara con las utilizadas para el péndulo) para poder reproducir con precisión dichas oscilaciones.

Igual que hicimos en el problema del péndulo, nos interesa ver el grado de cumplimiento de la condición algebraica $(g_1, g_2) = \mathbf{0}$. En la gráfica de la Figura 3.9, que se muestra en escala logarítmica para el eje vertical, podemos observar el valor que toman las dos componentes de la condición algebraica a lo largo de la solución numérica. Este valor está por debajo de 10^{-6} para la primera componente de g y por debajo de 10^{-8} para la segunda.

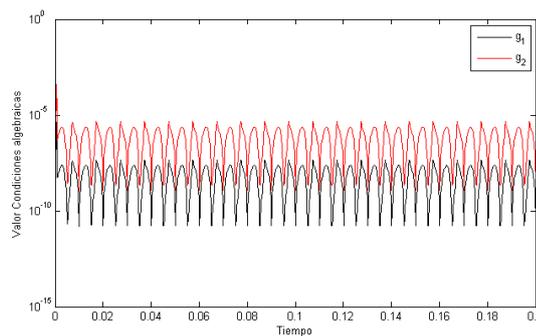


Figura 3.9: Valor de las condiciones algebraicas.

Si cambiamos el voltaje de entrada a $U_e(t) = 0,4 \sin(20\pi t)$ la solución sigue presentando oscilaciones, pero de menor frecuencia, como se puede ver en la Figura 3.10 y en la Figura 3.11.

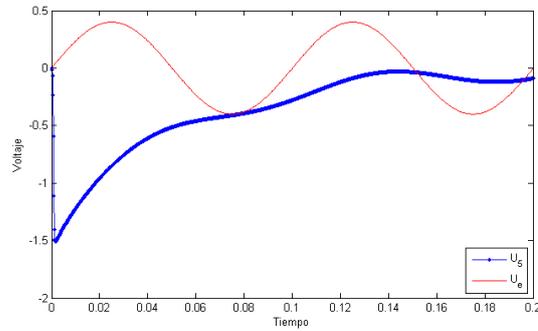


Figura 3.10: Evolución de los voltajes U_5 y U_e con el tiempo.

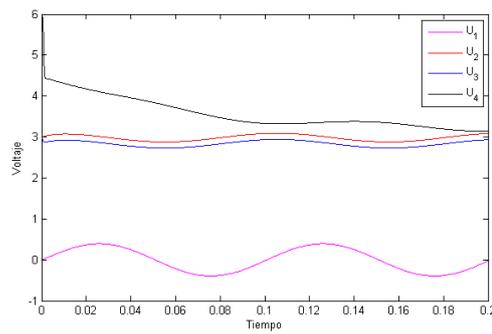


Figura 3.11: Evolución de los voltajes U_1 , U_2 , U_3 y U_4 con el tiempo.

Bibliografía

- [1] C.W. Gear *Differential-algebraic equation index transformations*. SIAM J. Sci. Statist. Comput. 9 (1988), pp. 39-47.
- [2] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations. I. Nonstiff Problems*. Second edition. Springer Series in Computational Mathematics 8. Springer-Verlag, Berlin (1987).
- [3] E. Hairer and G. Wanner *Multistep-Multistage-Multiderivative Methods for Ordinary Differential Equations*. Computing 11 (1973), pp. 287-303.
- [4] E. Hairer and G. Wanner *On the Butcher Group and General Multi-Value Methods*. Computing 13 (1974), pp. 1-15.
- [5] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations. II. Stiff and Differential-Algebraic Problems*. Second edition. Springer Series in Computational Mathematics 14. Springer-Verlag, Berlin (1996).
- [6] P. Kaps and G. Wanner, *A Study of Rosenbrock-Type Methods of High Order*. Numerische Mathematik 38 (1981), pp. 279-298.
- [7] P. Rentrop, M. Roche and G. Steinebach, *The application of Rosenbrock-Wanner type methods with stepsize control in differential-algebraic equations*. Numerische Mathematik 55 (1989), pp. 545-563.
- [8] M. Roche, *Rosenbrock Methods for Differential Algebraic Equations*. Numerische Mathematik 52 (1988), pp. 45-63.
- [9] J.M. Sanz-Serna, *Diez lecciones de cálculo numérico*. Segunda edición. Manuales y textos universitarios, Ciencias No. 41. Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial, Valladolid (2010).

Apéndice A

Código MATLAB

En este apéndice mostramos el código de los programas de MATLAB que hemos realizado para la implementación de los métodos de Rosenbrock de orden 3 y 4 cuyos coeficientes se muestran en la Subsección 3.2, y su aplicación a los dos problemas considerados en la Sección 3.1 y cuyos resultados numéricos mostramos en la Sección 3.3: el péndulo matemático y el amplificador.

A.1. Función principal: Rosenbrock.m

El programa principal implementa, con paso fijo, los métodos de Rosenbrock de orden 3 y 4 para la resolución de sistemas de EDAs de índice 1. Necesita dos programas auxiliares, que contienen, dependiendo del problema, las funciones necesarias.

```
1 function [y,z]=Rosenbrock(y0,z0,N,t0,T,fg,jac,...
2                               alpha,betta,gamma);
3 % Programa principal para la integracion numerica de
4 % sistemas de EDAs de indice 1, utilizando metodos
5 % de Rosenbrock.
6 %
7 % Uso:
8 %[y,z]=Rosenbrock(y0,z0,N,t0,T,fg,jac,
9 %                 alpha,betta,gamma)
10 %
11 % y0, z0 son las condiciones iniciales
12 %
13 % t0 es el tiempo inicial, T es el tiempo final
14 %
15 % N es el numero de pasos que se dan con el metodo
```

```

16 %
17 % fg es la funcion auxiliar en la
18 % que se evaluan las funciones
19 % f y g asociadas a nuestro problema
20 %
21 % jac es la funcion auxiliar en la que se evaluan
22 % las derivadas parciales de f y g
23 %
24 % alpha, betta, gamma son los coeficientes del metodo
25 % de Rosenbrock que se elija
26 %
27 % y, z contienen por filas las
28 % componentes correspondientes a la solucion
29 % numerica del problema
30 %
31 h = (T-t0)/N; % longitud de paso
32 % inicializacion
33 y=zeros(N,length(y0));
34 z=zeros(N,length(z0));
35 k=zeros(length(betta),length(y0));
36 l=zeros(length(betta),length(z0));
37
38 for p=1:N
39     y(p,:)=y0;
40     z(p,:)=z0;
41     [fy,fz,gy,gz]=jac(y0,z0);
42     [f,g]=fg(y0,z0);
43
44     M=[eye(length(y0))-gamma(1,1)*h*fy ...
45         -gamma(1,1)*h*fz;-gamma(1,1)*h*gy -gamma(1,1)*
46         h*gz];
47     [L,U]=lu(M);
48     ab=U\(L\[h*f;h*g]);
49     k(1,:)=ab(1:length(y0))';
50     l(1,:)=ab(length(y0)+1:length(y0)+length(z0))';
51     for i=2:length(betta)
52         [f,g]=fg(y0+alpha(i,1:(i-1))*k(1:(i-1),:),...
53             z0+alpha(i,1:(i-1))*l(1:(i-1),:));
54
55         n1=h*f+h*fy*(gamma(i,1:(i-1))*k(1:(i-1),:))' ...

```

```
56         +h*fz*(gamma(i,1:(i-1))*l(1:(i-1),:))';
57     n2=h*g+h*gy*(gamma(i,1:(i-1))*k(1:(i-1),:))'...
58         +h*gz*(gamma(i,1:(i-1))*l(1:(i-1),:))';
59     ab=U\(L\[n1;n2]);
60
61     k(i,:)=ab(1:length(y0))';
62     l(i,:)=ab(length(y0)+1:length(y0)+length(z0))';
63     end
64     y0=y0+betta*k;
65     z0=z0+betta*l;
66     end
67     y(N,:)=y0;
68     z(N,:)=z0;
69     end
```

A.2. Funciones para el péndulo

Las funciones `fgPendulo.m` y `jacPendulo.m` contienen los parámetros y funciones del péndulo matemático.

```

1 function [f,g] = fgPendulo(y,z);
2 global m l gra %constantes del problema
3
4 f=[y(3) y(4) (-z/(m*l))*y(1) (-z/(m*l))*y(2)+gra]';
5 g=m*((y(3))^2+(y(4))^2)-z*l+gra*m*(y(2));
6
7 end

1 function [fy,fz,gy,gz] = jacPendulo(y,z);
2 global m l gra %constantes del problema
3
4 fy=[0 0 1 0;0 0 0 1;(-z/(m*l)) 0 0 0;...
5     0 (-z/(m*l)) 0 0];
6 fz=[0,0,-y(1)/(m*l),-y(2)/(m*l)]';
7 gy=[0, gra*m, 2*m*y(3), 2*m*y(4)];
8 gz=-l;
9
10 end

```

A.3. Funciones para el amplificador

Las funciones fgCircuito.m y jacCircuito.m contienen los parámetros y funciones del amplificador.

```

1 function [f,g] = fgCircuito(y,z);
2 global R0 R1 R2 R3 R4 R5 C1 C2 C3 Ub %constantes
3 ue=(0.4*sin(200*pi*y(4))-z(1));
4 fzy=10^(-6)*(exp((z(1)-y(1)-y(2))/0.026)-1);
5
6 f=[ue/(R0*C1),fzy/C2-(y(2))/(C2*R3),...
7   (z(2)-y(3))/(C3*R5),1]';
8 g=[ue/R0+Ub/R2+(y(1)-z(1))*(1/R1+1/R2)-0.01*fzy,...
9   (Ub-z(2))/R4-0.99*fzy+(y(3)-z(2))/R5]';
10
11 end

1 function [fy,fz,gy,gz] = jacCircuito(y,z);
2 global R0 R1 R2 R3 R4 R5 C1 C2 C3 Ub %constantes
3 fu=(10^(-6)*(exp((z(1)-y(1)-y(2))/0.026))...
4   *(1/0.026));
5 uep=80*pi*cos(200*pi*y(4));
6
7 fy=[0 0 0 uep/(R0*C1);
8   -fu/C2 -fu/C2-1/(C2*R3) 0 0;
9   0 0 -1/(C3*R5) 0;
10  0 0 0 0];
11 fz=[-1/(R0*C1) 0;fu/C2 0;0 1/(C3*R5); 0 0];
12 gy=[1/R1+1/R2+0.01*fu 0.01*fu 0 uep/R0;...
13   0.99*fu 0.99*fu 1/R5 0];
14 gz=[-1/R0-1/R1-1/R2-0.01*fu 0;-0.99*fu -1/R4-1/R5];
15
16 end

```

A.4. ErrorCondicionAlgebraica.m

En este programa se muestra cómo, utilizando las funciones Rosenbrock, fgPendulo y jacPendulo que hemos creado, obtenemos la solución numérica con los dos métodos de orden 3 y 4 y representamos gráficamente los valores de la condición algebraica a lo largo de las dos soluciones numéricas obtenidas. Se ha utilizado para generar la gráfica superior de la Figura 3.4.

```

1 clear variables
2 global m l gra
3
4 % definimos los coeficientes de los
5 % metodos de Rosenbrock de orden 3 y 4
6 betta=[0.3197278911564624,0.7714777906171382,...
7     -0.09120568177360061];
8 alpha=[0,0,0;0.7,0,0;0.7,0,0];
9 gamma=[0.435866521508459,0,0;...
10     0.1685887625570998,0.435866521508459,0;...
11     4.943922277836421,1,0.435866521508459];
12 betta2=[0.2523628037277470,-0.2209698738798533,...
13     -0.2256411840923124,0.3179133966013711,...
14     0.8763348576430476];
15 alpha2=[0,0,0,0,0;1.233311380872013,0,0,0,0;...
16     0.6535453813273382,0.2295950748229277,0,0,0;...
17     2.681059792907162,-1.554590259558157,...
18     -0.9682496302574051,0,0;...
19     -0.6021422614217772,0.2994399056322287,...
20     0.4792338650945191,0.8010415023569842,0];
21 gamma2=[0.70751226521,0,0,0,0;...
22     -1.818714325256271,0.70751226521,0,0,0;...
23     -0.4589460040608732,0.3613323897595465,...
24     0.70751226521,0,0;...
25     -3.424045164556574,1.553491448551290,...
26     1.249712740807497,0.70751226521,0;...
27     -0.2261466054228607,-0.3882326103473952,...
28     -0.3589041115714489,-0.01860845389367294,...
29     0.70751226521];
30
31 % Ahora se definen las variables globales
32 % del problema
33 m=1;

```

```
34 l=1;
35 gra=9.81;
36
37 t0=0;
38 T=5;
39 N=1000; %h=(T-t0)/N
40 y0=[0 1 6 0];
41 z0=(m*(y0(3).^2+y0(4).^2)+gra*m*y0(2))/l;
42
43 [y,z] = Rosenbrock(y0,z0,N,t0,T,...
44     @fgPendulo,@jacPendulo,alpha,betta,gamma);
45
46 [y2,z2] = Rosenbrock(y0,z0,N,t0,T,...
47     @fgPendulo,@jacPendulo,alpha2,betta2,gamma2);
48     t=linspace(t0,T,N);
49
50 figure(3)
51 semilogy(t,abs(m*(y(:,3).^2+y(:,4).^2)-z*l+gra*m*y
52     (:,2)), 'k-',...
53     t,abs(m*(y2(:,3).^2+y2(:,4).^2)-z2*l+gra*m*y2
54     (:,2)), 'r-')
```

A.5. U5Ue.m

En este programa se muestra cómo, utilizando las funciones Rosenbrock, fgCircuito y jacCircuito que hemos creado, obtenemos la solución numérica con el método de Rosenbrock de orden 4 y representamos gráficamente las componentes que aproximan a $U_e(t)$ y $U_5(t)$ a lo largo del intervalo $[0, 0.2]$. Se ha utilizado para generar la Figura 3.7.

```

1 clear variables
2 global R0 R1 R2 R3 R4 R5 C1 C2 C3 Ub
3 format long
4 %Coeficientes del metodo de Rosenbrock de orden 4
5 betta2=[0.2523628037277470 , -0.2209698738798533, ...
6         -0.2256411840923124 , 0.3179133966013711, ...
7         0.8763348576430476];
8 alpha2=[0,0,0,0,0; 1.233311380872013,0,0,0,0; ...
9         0.6535453813273382,0.2295950748229277,0,0,0; ...
10        2.681059792907162, -1.554590259558157, ...
11        -0.9682496302574051,0,0; ...
12        -0.6021422614217772,0.2994399056322287, ...
13        0.4792338650945191,0.8010415023569842,0];
14 gamma2=[0.70751226521,0,0,0,0; ...
15         -1.818714325256271,0.70751226521,0,0,0; ...
16         -0.4589460040608732,0.3613323897595465, ...
17         0.70751226521,0,0; ...
18         -3.424045164556574,1.553491448551290, ...
19         1.249712740807497,0.70751226521,0; ...
20         -0.2261466054228607, -0.3882326103473952, ...
21         -0.3589041115714489, -0.01860845389367294, ...
22         0.70751226521];
23
24 % A continuacion se definen las variables globales
25 % del problema
26 R0=1000;
27 R1=9000;
28 R2=9000;
29 R3=9000;
30 R4=9000;
31 R5=9000;
32 C1=10^(-6);
33 C2=2*10^(-6);

```

```
34 C3=3*10^(-6);
35 Ub=6;
36
37 y0=[ -(Ub*R1)/(R1+R2) (Ub*R1)/(R1+R2) Ub 0];
38 z0=[0 Ub];
39 t0=0;
40 T=0.2;
41 N=1000;
42
43 [y,z] = Rosenbrock(y0,z0,N,t0,T,...
44             @fgCircuito,@jacCircuito,alpha2,betta2,gamma2)
45             ;
46 U(:,1)=z(:,1);
47 U(:,2)=z(:,1)-y(:,1);
48 U(:,3)=y(:,2);
49 U(:,4)=z(:,2);
50 U(:,5)=z(:,2)-y(:,3);
51 t=linspace(t0,T,N);
52 % Generamos la figura que muestra Ue y U5
53
54 figure(1)
55 U(:,6)=0.4*sin(200*pi*y(:,4));
56 plot(t,U(:,5),'-b.',t,U(:,6),'r-');
57 legend('U_5','U_e');
58 xlabel('Tiempo')
59 ylabel('Voltaje')
```