UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO

Facultad de Ciencias

Una Variante en los Métodos Numéricos de Solución de Sistemas de Ecuaciones Diferenciales

TESIS

Que para obtener el título de:

MATEMATICO

Presenta:

Alejandro López Ortiz

Julio de 1989

Indice

INTRODUCCION

En 1981 Charles W. Gear quien resolvio el problema de hallar un algoritmo eficiente para resolver Sistemas de Ecuaciones Diferenciales Rigidas, se plantea en el SIAM Review[1] la pregunta :

SOLUCIONES NUMERICAS DE ECUACIONES DIFERENCIALES :
QUEDA ALGO MAS POR HACER ?

y responde a esto realizando un analisis de los distintos pasos en el desarrollo de la solucion de un problema de soluciones numericas de ecuaciones diferenciales ordinarias desde el momento en que es identificado por el «usuario» hasta el momento en el cual existen programas que resuelven este problema, para posteriormente describir el avance en la solucion de varios problemas de interes en esta area. Citamos aqui sus comentarios con respecto al avance de a los metodos conocidos como metodos de poca precision :

> Metodos de poca precision.- Los metodos actuales de solucion de Ecuaciones Diferenciales Ordinarias estan basados en el concepto del orden polinomial. Esto aproxima la solucion numerica a la verdadera solucion de una manera asintotica. Desafortunadamente, para incrementos grandes aparecen varios problemas. El primero es que las estimaciones de error son poco confiables debido a que tambien se encuentran basados en

---

[1] GEAR, C. W.
NUMERICAL SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS: IS THERE ANYTHING LEFT TO DO?
SIAM Review, Vol. 23, No. 1, January 1981.

teoria asintotica. El segundo es que los metodos actuales son ineficientes y muestran pocas propiedades de estabilidad. Metodos de poca precision son requeridos en un rango amplio de aplicaciones, en particular, simulaciones en tiempo real o integracion en el tiempo de Ecuaciones Diferenciales. En la integracion de tiempo real una de las propiedades mas importantes es la correspondencia entre la estabilidad real del sistema y la del modelo numerico dado que las simulaciones frecuentemente se hacen para verificar las propiedades de manejo de un dispositivo dado. El metodo debe ser tan estable como el dispositivo lo es.

Es uno de estos metodos de baja precision el tema de esta tesis. Dicho metodo surgio como una solucion al problema de graficacion de soluciones de Sistemas de Ecuaciones Diferenciales por computadora en las que se deben destacar propiedades de periodicidad, estabilidad y de tipo geometrico.

En el capitulo 1 se hace un estudio teorico de las propiedades del metodo propuesto obteniendose algunos teoremas que describen el comportamiento de este metodo.

En el capitulo 2 se realiza una comparacion entre distintos metodos numericos de resolucion de sistemas de ecuaciones diferenciales.

El capitulo 3 trata acerca de un programa graficador de sistemas de ecuaciones diferenciales desarrollado por el autor en el Instituto de Investigaciones en Matematicas Aplicadas y Sistemas y en el Instituto de Matematicas de la UNAM. Cabe mencionar que la mayor parte de las ilustraciones de este trabajo fueron genereadas por dicho programa.

CAPITULO 1

1.1. PRELIMINARES

Sistemas de Ecuaciones Diferenciales
Sistemas Autonomos
Soluciones Periodicas
Primera Integral

1.2. METODOS DE UN SOLO PASO

Definiciones
Criterios de Convergencia
Calculo del Error
Series de Taylor

1.3. METODO DE EULER

1.4. METODO DE EULER 2

Sistemas Lineales
Sistemas no Lineales
Algunos Resultados

SECCION 1.1
PRELIMINARES

Sistemas de Ecuaciones Diferenciales.- A lo largo de este texto se debe entender por Sistema de Ecuaciones Diferenciales el siguiente Sistema de Ecuaciones Diferenciales Ordinarias de Primer Orden :

$$
(1.1) \qquad
\begin{aligned}
\dot{u}_1 &= f_1(t, u_1, u_2, \dots, u_s) & \qquad u_1(t_0) &= u_{10} \\
\dot{u}_2 &= f_2(t, u_1, u_2, \dots, u_s) & \qquad u_2(t_0) &= u_{20} \\
&\vdots & &\vdots \\
\dot{u}_s &= f_s(t, u_1, u_2, \dots, u_s) & \qquad u_s(t_0) &= u_{s0},
\end{aligned}
$$

donde $u_{10}, u_{20}, \dots, u_{s0}$ son valores dados y

$$f_i : A \subset \mathbb{R}^{s+1} \longrightarrow \mathbb{R} \qquad i = 1 \dots s$$

Al conjunto de funciones $u_1(t)$, $u_2(t)$ ... $u_s(t)$ que satisfacen para toda t las ecuaciones

$$
\begin{aligned}
\dot{u}_1(t) &= f_1(t, u_1, u_2, \dots, u_s) \\
\dot{u}_2(t) &= f_2(t, u_1, u_2, \dots, u_s) \\
&\vdots \\
\dot{u}_s(t) &= f_s(t, u_1, u_2, \dots, u_s)
\end{aligned}
$$

se le llama la solucion general del Sistema.

Al problema de encontrar una solucion al sistema (1.1) que satisfaga las condiciones iniciales dadas

$$u_i(t_0) = \eta_i \qquad i = 1, 2, \ldots, s \ ,$$

donde los los numeros η_i son variables dadas se le llama el problema con condiciones iniciales.

Sistemas Autonomos .- Un sistema es *autonomo* si $\dot{u}_i$ no aparece en terminos del tiempo en (1.1) para toda i, es decir, si

$$\frac{\partial f_i}{\partial t} = 0 \quad \forall\ i = 1 \ldots s \ ,$$

por lo que si (1.1) es un Sistema Autonomo puede ser escrito de la siguiente manera :

$$\begin{array}{ll}
\dot{u}_1 = f_1(u_1, u_2, \ldots, u_s) & u_1(t_0) = u_{10} \\
\dot{u}_2 = f_2(u_1, u_2, \ldots, u_s) & u_2(t_0) = u_{20} \\
\vdots & \vdots \\
\dot{u}_s = f_s(u_1, u_2, \ldots, u_s) & u_s(t_0) = u_{s0} \ ,
\end{array}$$

donde $u_{10}, u_{20}, \ldots, u_{s0}$ son valores dados y

$$f_i : A \subset \mathbb{R}^s \longrightarrow \mathbb{R} \qquad i = 1 \ldots s$$

Tambien los Sistemas de Ecuaciones Diferenciales suelen ser denotados vectorialmente:

$$\dot{\bar{u}} = f\ (t, \bar{u}), \qquad \bar{u}(t_0) = \bar{u}_0$$

donde

$$\bar{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad \bar{u}_0 = \begin{bmatrix} u_{10} \\ u_{20} \\ \vdots \\ u_{n0} \end{bmatrix} \qquad f\,(t\,,\,\bar{u}) = \begin{bmatrix} f_1(t\,,\,\bar{u}) \\ f_2(t\,,\,\bar{u}) \\ \vdots \\ f_n(t\,,\,\bar{u}) \end{bmatrix}$$

Ahora bien, si tenememos un sistema no autonomo de n ecuaciones podemos obtener un sistema autonomo de $n+1$ ecuaciones con soluciones equivalentes, por lo que en general trabajaremos con sistemas de la forma :

$$\dot{\bar{u}}(t) = f(\bar{u}), \qquad \bar{u}(t_0) = \bar{u}_0$$

Para el caso de $n = 2$ usaremos una notacion un poco distinta aunque convencional en el caso de Sistemas Autonomos en el plano

$$\begin{aligned} \bar{u} &= (x,\ y) \\ \dot{x} &= f_1(x,y) \\ \dot{y} &= f_2(x,y) \end{aligned}$$

y con $\bar{u}_0 = (x_0\,,\ y_0)$ como condicion inicial al tiempo t_0. Frecuentemente trabajaremos con sistemas de dos variables puesto que uno de los objetivos de este trabajo es la graficacion por computadora de soluciones de ecuaciones diferenciales y en general solo se grafican dos o a lo mas tres variables de las que componen el sistema.

Soluciones Periodicas .- Una solucion al sistema

$$\dot{\bar{u}} = f\ (t,\ \bar{u}) \qquad \bar{u}\ (t_0) = \bar{u}_0$$

se dice que es una *solucion periodica* si existe un numero $T > 0$ tal que $\bar{u}(t + T) = \bar{u}(t)$ para toda $t \geq 0$. T es llamado un periodo de la solucion. A la menor de todos las T que satisfacen la condicion $\bar{u}(t + T) \equiv \bar{u}(t)$ se le llama *el periodo* de la solucion, y si $T \neq 0$ se dice que la solucion periodica es *no trivial* .

Ejemplo 1.1.- Considere el sistema de ecuaciones diferenciales :

$$\dot{x} = -y \qquad \dot{y} = x$$

con solucion general :

$$x(t) = C_1\cos(t) + C_2\operatorname{sen}(t)$$
$$y(t) = C_1\operatorname{sen}(t) + C_2\cos(t)$$

$$f(t) = (C_1\cos(t) + C_2\operatorname{sen}(t),\ C_1\operatorname{sen}(t) + C_2\cos(t))$$

Se tiene $(C_1\cos(t) + C_2\operatorname{sen}(t),\ C_1\operatorname{sen}(t) + C_2\cos(t)) = (C_1\cos(t + 2\pi) + C_2\operatorname{sen}(t + 2\pi),\ C_1\operatorname{sen}(t + 2\pi) + C_2\cos(t + 2\pi))$ por lo que es un sistema con soluciones periodicas, hecho que se puede apreciar en la figura 1 donde estan graficadas varias soluciones de este sistema.

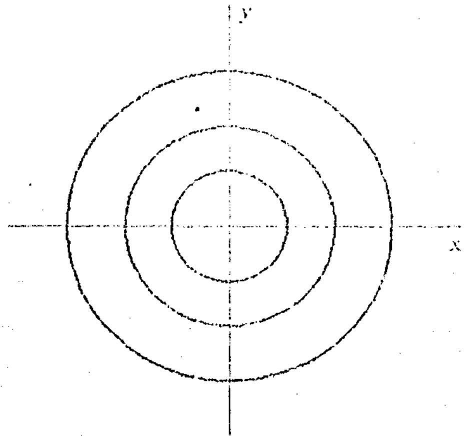


Figura 1

Soluciones de $(\dot{x},\dot{y}) = (-y,x)$

Un punto $\bar{u}_o$ es un *punto critico* del sistema $\dot{\bar{u}} = f(\bar{u})$ si

$$f(\bar{u}_o) = \bar{0}$$

Primera Integral .- Una funcion U definida en un compacto K que toma valores en los reales con derivadas parciales U_x y U_y continuas en K y que no se anulan salvo posiblemente en los puntos criticos del sistema $(\dot{x}, \dot{y}) = f(x, y)$, es llamada una *primera integral del sistema en* K si $U(x(t), y(t)) \equiv cte.$ para toda solucion $(x(t), y(t))$ del sistema de ecuaciones diferenciales que este contenida en K. Las curvas de nivel $U(x, y) \equiv cte.$ son llamadas las *curvas integrales* del sistema. Citamos aqui un par de lemas sin demostracion [Martin, 1983], que seran de utilidad para determinar soluciones periodicas.

Lema 1.1. Si $U(x, y)$ es una primera integral de la ecuacion diferencial total $f_1(x, y)\,dx - f_2(x, y)\,dy = 0$ en K, entonces U es una primera integral de $(\dot{x}, \dot{y}) = f(x, y)$ en K. ◻

Lema 1.2. Sea $U = U(x, y)$ una primera integral de un sistema de ecuaciones diferenciales en dos variables, y si la curva de nivel $C = \{(x, y) : U(x, y) = k\}$ es una curva cerrada y no contiene ningun punto critico del sistema, entonces cualquier solucion del sistema con valor inicial en C es periodica y la curva de nivel C es su trayectoria. ◻

SECCION 12
METODOS DE UN SOLO PASO

Definiciones.-Sea $t \in [a,b]$; sea $\overline{u}_o$ un vector arbitrario y $\overline{u}(t)$ la solucion del Sistema de Ecuaciones Diferenciales $\dot{\overline{u}} = f(t,\overline{u})$ que satisface $\overline{u}(t_o) = \overline{u}_o$, definimos el *incremento exacto* relativo a la solucion $\overline{u}$ como

$$\Delta(t, \overline{u}; h) = \begin{cases} \dfrac{\overline{u}(t+h) - \overline{u}(t)}{h} & h \neq 0 \\ f(t,\overline{u}) & h = 0 \end{cases}$$

Un *metodo de un paso* para la solucion de un Sistema de Ecuaciones Diferenciales se define por las ecuaciones :

$$\overline{u}_o = \overline{\eta}$$

$$\overline{u}_{n+1} = \overline{u}_n + h\Phi(t_n, \overline{u}_n; h) \qquad n = 0, 1, \ldots ,$$

donde $\overline{u}_n$ aproxima a $\overline{u}(t_o + nh)$. La funcion $\Phi(t, \overline{u}; h)$ es llamada la *funcion incremento*.

Es intuitivamente razonable suponer que $\Phi(t, \overline{u}; h)$ se aproxime a $\Delta(t, \overline{u}; h)$ tanto como sea posible siempre que h sea suficientemente pequena, es decir :

$$\lim_{h \to 0} | \Phi(t, \overline{u}; h) - \Delta(t, \overline{u}; h) | = 0$$

La mayoria de las funciones incremento comunmente usadas satisfacen la condicion de arriba. Esto nos lleva a la siguiente definicion :

Un metodo definido por la funcion incremento $\Phi(t, \bar{u}; h)$ es *convergente* si para toda condicion inicial $\bar{u}_0$ y para todo t :

$$\lim_{\substack{h \to 0 \\ t_n = t}} \bar{u}_n = \bar{u}(t),$$

donde $t_n = t_0 + nh$.

Un metodo es *consistente* si para toda $\bar{u}$ y todo t :

$$\Phi(t, \bar{u}; 0) = f(t, \bar{u})$$

Un metodo es de *orden* p si p es el mayor entero para el cual se cumple :

$$\Phi(t, \bar{u}; h) - \Delta(t, \bar{u}; h) = O(h^p)$$

Por otra parte, si nos restringimos a sistemas autonomos entonces Φ y Δ dependen unicamente de $\bar{u}$ y h, por lo que se suprime el argumento t al escribirlas.

Veamos finalmente un teorema de gran utilidad para el estudio de sucesiones, que usaremos ampliamente sobre puntos de la solucion del sistema.

Lema 1.3. Considere la sucesion ξ_n ($n = 0, 1, 2 \ldots$) que satisface la siguiente desigualdad para $n = 0, 1, 2, \ldots, N - 1$:

$$|\xi_{n+1}| \le A|\xi_n| + B,$$

donde A y B son constantes no negativas que no dependen de n entonces

$$|\xi_n| \le A^n|\xi_0| + \begin{cases} \dfrac{A^n - 1}{A - 1} B & A \ne 1 \\ nB & A = 1 \end{cases} \qquad n = 1, 2, \ldots, N \tag{1.2}$$

Demostracion. Demostraremos este hecho por induccion matematica; primer caso si $A \neq 1$:

i) Base de la Induccion.- Para $n = 1$, tenemos que (1.2) se reduce a $|\xi_1| \leq A|\xi_0| + B$, lo cual es cierto por hipotesis.

ii) Paso de Induccion.- Supongamos que (1.2) es cierto para $n < N$, entonces por hipotesis del lema

$$|\xi_{n+1}| \leq A|\xi_n| + B$$

y como por hipotesis de induccion

$$|\xi_n| \leq A^n|\xi_0| + \frac{A^n - 1}{A - 1} B$$

sustituimos $|\xi_n|$ en la penultima desigualdad y se tiene

$$|\xi_{n+1}| \leq A \left[A^n|\xi_0| + \frac{A^n - 1}{A - 1} B \right] + B$$

$$= A^{n+1}|\xi_0| + \left[A \frac{A^n - 1}{A - 1} + 1 \right] B$$

$$= A^{n+1}|\xi_0| + \frac{A^{n+1} - 1}{A - 1} B$$

lo cual es lo que se queria demostrar. Ahora bien para el caso $A = 1$ demostraremos el resultado tambien por induccion, a saber

i)Base de la induccion.- Para $n = 1$ tenemos que (1.2) se reduce
a $|\xi_1| \leq |\xi_0| + B$ lo cual es cierto por hipotesis.

ii) Paso de Induccion.- Supongamos que (1.2) es cierto para $n < N$, entonces por hipotesis del lema

$$|\xi_{n+1}| \leq |\xi_n| + B$$

y como por hipotesis de induccion

$$|\xi_n| \leq |\xi_o| + nB$$

sustituimos $|\xi_n|$ en la penultima desigualdad y se tiene :

$$|\xi_{n+1}| \leq |\xi_o| + nB + B = |\xi_o| + (n+1)B$$

lo cual es lo que se queria demostrar. ∎

Criterios de Convergencia.- Es de utilidad el tener una prueba sencilla sobre la convergencia de un metodo pues verificar las condiciones de la definicion es en la mayoria de los casos un problema muy complicado. En este sentido el siguiente Teorema nos da un criterio de facil aplicacion.

Teorema 1.4. Sea la funcion $\Phi(t, \bar{u}; h)$ continua y tal que existe una constante L que cumple :

$$|\Phi(t, \bar{z}; h) - \Phi(t, \bar{u}; h)| \leq L|\bar{z} - \bar{u}|$$

para todo $(t, \bar{z}; h)$ y $(t, \bar{u}; h)$ y tal que $h \leq h_o$, donde $h_o \geq 0$ es una constante dada. Entonces una condicion necesaria y suficiente para la convergencia del metodo definido por Φ es que el metodo sea consistente.

Demostracion. Por definicion, un metodo es convergente si

$$\lim_{\substack{n\to\infty \\ t_n = s}} \bar{u}_n = \bar{u}(s) \quad ,$$

o equivalentemente si

$$\lim_{\substack{n\to\infty \\ nh = s - t_o}} |\bar{u}_n - \bar{u}(s)| = 0$$

Si $t_i = ih + t_0$ se tiene :

$$\begin{aligned}
\overline{u}_n - \overline{u}(s) &= \overline{u}_n - \overline{u}(t_n) \\
&= \overline{u}_{n-1} + h\Phi(t_{n-1}, \overline{u}_{n-1}; h) \\
&\quad - \overline{u}(t_{n-1}) - h\Delta(t_{n-1}, \overline{u}(t_{n-1}); h) \\
&= \overline{u}_{n-1} - \overline{u}(t_{n-1}) \\
&\quad + h[\Phi(t_{n-1}, \overline{u}_{n-1}; h) - \Delta(t_{n-1}, \overline{u}(t_{n-1}); h)] \\
&= \overline{u}_{n-1} - \overline{u}(t_{n-1}) \\
&\quad + h\big[\Phi(t_{n-1}, \overline{u}_{n-1}; h) - \Phi(t_{n-1}, \overline{u}(t_{n-1}); h) \\
&\quad + \Phi(t_{n-1}, \overline{u}(t_{n-1}); h) - \Phi(t_{n-1}, \overline{u}(t_{n-1}); 0) \\
&\quad + \Phi(t_{n-1}, \overline{u}(t_{n-1}); 0) - \Delta(t_{n-1}, \overline{u}(t_{n-1}); h)\big] ;
\end{aligned}$$

aplicando la desigualdad del triangulo

$$\begin{aligned}
|\overline{u}_n - \overline{u}(t_n)| \leq |\overline{u}_{n-1} - \overline{u}(t_{n-1})| & \\
+ h\big[|\Phi(t_{n-1}, \overline{u}_{n-1}; h) &- \Phi(t_{n-1}, \overline{u}(t_{n-1}); h)| \\
+ |\Phi(t_{n-1}, \overline{u}(t_{n-1}); h) &- \Phi(t_{n-1}, \overline{u}(t_{n-1}); 0)| \\
+ |\Phi(t_{n-1}, \overline{u}(t_{n-1}); 0) &- \Delta(t_{n-1}, \overline{u}(t_{n-1}); h)|\big];
\end{aligned}$$

y por las hipotesis tenemos que

i) $|\Phi(t_{n-1}, \overline{u}_{n-1}; h) - \Phi(t_{n-1}, \overline{u}(t_{n-1}); h)| \leq L|\overline{u}_{n-1} - \overline{u}(t_{n-1})|$

ii) por ser Φ continua es uniformente continua en el compacto definido por : $t_0 \leq t \leq t_n$ $\quad \overline{u} = \overline{u}(t)$ $\quad 0 \leq h \leq h_0$ y por tanto

$$|\Phi(t_{n-1}, \overline{u}(t_{n-1}); h) - \Phi(t_{n-1}, \overline{u}(t_{n-1}); 0)| \to 0 \qquad n \to \infty$$

uniformemente para toda t, es decir, como Φ es uniformemente continua en K se tiene que para toda $\varepsilon > 0$ existe $\delta > 0$ tal que para todos $(s, \overline{u}, h) \in K$, $(s_0, \overline{u}_0, h_0) \in K$ tales que :

$$|(s, \overline{u}, h) - (s_0, \overline{u}_0, h_0)| < \delta \qquad \text{entonces}$$

$$|\Phi(s, \overline{u}; h) - \Phi(s_0, \overline{u}_0; h_0)| < \varepsilon$$

por lo que para el caso de $s = s_0$, $\bar{u} = \bar{u}_0$, $h_0 = 0$ obtenemos que la expresion anterior implica que

$$|h| < \delta \longrightarrow |\Phi(t_{n-1}, \bar{u}(t_{n-1}); 0) - \Phi(t_{n-1}, \bar{u}(t_{n-1}); h)| < \varepsilon$$

por lo que esta ultima diferencia es tan cercana a cero como se quiera siempre que h sea lo suficientemente pequena independientemente del valor de t_{n-1} .

iii) como el metodo es consistente

$$\Phi(t_{n-1}, \bar{u}(t_{n-1}); 0) = f(t_{n-1}, \bar{u}(t_{n-1})),$$

y por la definicion y el teorema del valor medio :

$$\begin{aligned} \Delta(t_{n-1}, \bar{u}(t_{n-1}); h) &= \frac{\bar{u}(t_n) - \bar{u}(t_{n-1})}{h} \\ &= \dot{\bar{u}}\ (t_{n-1} + \xi) \\ &= f(t_{n-1} + \xi, \bar{u}(t_{n-1} + \xi)) \end{aligned}$$

para $0 \le \xi \le t_n - t_{n-1}$, obteniendo

$$|\Phi(t_{n-1}, \bar{u}(t_{n-1}); 0) - \Delta(t_{n-1}, \bar{u}(t_{n-1}); h)| =$$

$$|f(t_{n-1}, \bar{u}(t_{n-1})) - f(t_{n-1} + \xi, \bar{u}(t_{n-1} + \xi))|,$$

lo cual tiende a cero uniformemente cuando $n \to \infty$.

Si definimos

$$\begin{aligned} A &= (1 + hL) \\ B &= h\big[|\Phi(t_{n-1}, \bar{u}(t_{n-1}); h) - \Phi(t_{n-1}, \bar{u}(t_{n-1}); 0)| \\ &\quad + |f(t_{n-1}, \bar{u}(t_{n-1})) - f(t_{n-1} + \xi, \bar{u}(t_{n-1} + \xi))|\big], \end{aligned}$$

tenemos

$$|\bar{u}_n - \bar{u}(t_n)| \le |\bar{u}_{n-1} - \bar{u}(t_{n-1})| + hL|\bar{u}_{n-1} - \bar{u}(t_{n-1})| + B$$
$$= A|\bar{u}_{n-1} - \bar{u}(t_{n-1})| + B$$

y por el lema 1.1 :

$$|\bar{u}_n - \bar{u}(t_n)| \le A^n|\bar{u}_0 - \bar{u}(t_0)| + \frac{A^n - 1}{A - 1} B$$

y dado que $\bar{u}_0 = \bar{u}(t_0)$

$$|\bar{u}_n - \bar{u}(t_n)| \le \frac{A^n - 1}{A - 1} B$$

Tomando el limite obtenemos

$$\lim_{n\to\infty} |\bar{u}_n - \bar{u}(s)| \le \lim_{n\to\infty} \frac{(1 + hL)^n - 1}{1 + hL - 1} B$$

$$= \lim_{n\to\infty} \frac{\left(1 + \frac{(s - t_0)L}{n}\right)^n - 1}{hL} \cdot h \cdot$$

$$\cdot \left[|\Phi(t_{n-1}, \bar{u}(t_{n-1}); h) - \Phi(t_{n-1}, \bar{u}(t_{n-1}); 0)| + |f(t_{n-1}, \bar{u}(t_{n-1})) - f(t_{n-1} + \xi, \bar{u}(t_{n-1} + \xi))|\right]$$

$$= \frac{e^{(s-t_0)L} - 1}{L} \cdot$$

$$\cdot \lim_{n\to\infty} \left[|\Phi(t_{n-1}, \bar{u}(t_{n-1}); h) - \Phi(t_{n-1}, \bar{u}(t_{n-1}); 0)| + |f(t_{n-1}, \bar{u}(t_{n-1})) - f(t_{n-1} + \xi, \bar{u}(t_{n-1} + \xi))|\right]$$

$$= \frac{e^{(s-t_0)L} - 1}{L} \cdot 0$$

$$= 0$$

Con esto hemos demostrado la suficiencia de las condiciones del teorema para que un metodo sea convergente. A continuacion se demostrara que la consistencia es una condicion necesaria para que un metodo sea convergente.

Supongamos que el metodo determinado por la funcion incremento Φ es convergente con respecto al sistema

$$\dot{\bar{u}} = f(t, \bar{u}) \qquad \bar{u}(t_0) = \bar{u}_0,$$

y sea z la solucion al sistema

$$\dot{\bar{z}} = g(t, \bar{z}) \qquad \bar{z}(t_0) = \bar{u}_0,$$

donde $g(t,\bar{z}) := \Phi(t, \bar{z}; 0)$. Por la definicion misma de g se tiene que el metodo determinado por Φ visto como solucion del sistema $\dot{\bar{z}} = g(t, z)$ es consistente, y por tanto por la primera parte de este teorema se tiene que los valores $\bar{u}_n$ dados por el metodo convergen a la solucion $\bar{z}$ del sistema. Por otra parte tenemos que por hipotesis el metodo es convergente con respecto a $f(t, \bar{u})$, por lo que los valores $\bar{u}_n$ convergen a la solucion $\bar{u}$. Entonces por la unicidad del limite se tiene que necesariamente $\bar{z}(t) \equiv \bar{u}(t)$, lo que implica que :

$$f(t, \bar{u}) = \dot{\bar{u}}(t) = \dot{\bar{z}}(t) = g(t, \bar{z}) = \Phi(t, \bar{z}; 0) = \Phi(t, \bar{u}; 0)$$

lo cual prueba la consistencia del metodo. ∎

Calculo del Error.- El conocer de antemano el error en la aproximacion de la solucion de un metodo es de gran importancia, por lo que se citamos sin demostracion un teorema [Henrici, 1983] en este sentido.

Teorema 1.5. Sea $\Phi(t, \overline{u}; h)$ como en el teorema 1.4. y sean $N \geq 0$, $p \geq 0$ y $h \geq 0$ constantes tales que :

$$|\phi(t, \overline{u}(t); h) - \Delta(t, \overline{u}(t); h)| \leq Nh^{p},$$
$$t \in [a, b] \quad h \leq h_o$$

y si $\overline{v}_n$ es una sucesion de vectores que satisface

$$\begin{aligned} \overline{u}_o &= \overline{\eta} & n &= 0,1,2,\ldots, \\ \overline{u}_{n+1} &= \overline{u}_n + h[\Phi(t_n, \overline{u}_n; h) + h^q\, k\, \theta_n] & t_n &\in [a, b] \end{aligned}$$

donde $k \geq 0$ y $q \geq 0$ son constantes, y los vectores θ_n satisfacen $|\theta_n| \leq 1$. Entonces, para $t_n \in [a, b]$ y $h \leq h_o$:

$$|\overline{v}_n - \overline{u}(t_n)| \leq h^r\, N_1\, E_L(t - a),$$

donde

$$r = \min(p, q),$$

$$N_1 = Nh_o^{p-r} + kh_o^{q-r},$$

$$E_L(t) = \begin{cases} \dfrac{e^{Lt} - 1}{L} & L \geq 0 \\ t & L = 0 \end{cases} \qquad \blacksquare .$$

Series de Taylor.- Si la funcion $\overline{u}(t)$ es una solucion del Sistema Ecuaciones Diferenciales, y si las componentes de f son lo suficientemente diferenciables, entonces las derivadas superiores de $\overline{u}(t)$ pueden ser expresadas en terminos de f y de sus derivadas, esto es

$$\overline{u}^{(k+1)}(t) = \frac{d^k}{dt^k} f(\overline{u}(t)) = f^{(k)}(\overline{u}(t))$$

Si $\overline{u}(t)$ es infinitamente diferenciable -y en general supondremos que lo es- podemos aplicar el Teorema de Taylor obteniendo :

$$\overline{u}(t+h) = \overline{u}(t) + h\dot{\overline{u}}(t) + \frac{h^2}{2!}\ddot{\overline{u}}(t) + \frac{h^3}{3!}\dddot{\overline{u}}(t) + \ldots$$

$$= \overline{u}(t) + hf(\overline{u}) + \frac{h^2}{2!}f'(\overline{u}) + \frac{h^3}{3!}f''(\overline{u}) + \ldots$$

por lo que :

$$\Delta(\overline{u};\, h) = \frac{\overline{u}(t+h) - \overline{u}(t)}{h} = f(\overline{u}) + \frac{h}{2}f'(\overline{u}) + \frac{h^2}{3!}f''(\overline{u}) + \ldots$$

y por lo tanto un metodo de integracion puede estar basado en un truncamiento de la serie que describe a $\Delta(\overline{u};\, h)$, es decir, una funcion incremento de orden p esta dada por :

$$\Phi(\overline{u};\, h) = f(\overline{u}) + \frac{h}{2}f'(\overline{u}) + \ldots + \frac{h^{p-1}}{p!}f^{(p-1)}(\overline{u})$$

y en el caso de $p = 1$ tenemos el Metodo de Euler :

$$\Phi(\overline{u};\, h) = f(\overline{u})$$

$$\overline{u}_0 = \overline{\eta}$$

$$\overline{u}_{n+1} = \overline{u}_n + hf(\overline{u}_n)$$

El usar truncamientos de la Serie de Taylor con $p > 1$ lleva a metodos poco practicos pues se necesita calcular y evaluar derivadas superiores de $f(\overline{u})$. Los metodos de un solo paso intentan aproximar $\Delta(\overline{u};\, h)$ a traves de expresiones en que no aparezca explicitamente ninguna derivada de f. Este es el caso de los metodos de Runge-Kutta. Veamos el caso de los metodos de Runge-Kutta simplificados o de segundo orden.

Estos metodos estan dados por funciones incremento de la forma

$$\Phi(\bar{u};\ h) = a_1 f(\bar{u}) + a_2 f(\bar{u} + \rho h f(\bar{u})), \qquad \text{donde}$$

$$a_1 + a_2 = 1, \qquad a_2\rho = 1/2$$

con lo que obtenemos la solucion general :

$$\Phi(\bar{u};\ h) = (1 - \alpha) f(\bar{u}) + \alpha f(\bar{u} + \frac{h}{2\alpha} f(\bar{u})) \qquad \alpha \neq 0$$

que cosiste en un metodo de segundo orden en el que se requieren dos evaluaciones de f para obtener esta solucion.

SECCION 1.3
EL METODO DE EULER

Este metodo es poco preciso sin embargo es conceptualmente simple y de facil estudio por lo que es uno de los metodos mas conocidos. En el caso de un sistema autonomo de dos variables tenemos que el metodo de Euler para dos variables nos da la siguiente aproximacion :

$$x_n = x_{n-1} + h\, f_1(x_{n-1}, y_{n-1})$$
$$y_n = y_{n-1} + h\, f_2(x_{n-1}, y_{n-1}),$$

donde

$$\dot{x} = f_1(x,\ y) \qquad\qquad \text{y } (x_0, y_0) \text{ es la condicion inicial}$$
$$\dot{y} = f_2(x,\ y)$$

De aqui se deduce que la funcion incremento del metodo de Euler $\Phi(t,\ \bar{u})$ coincide con la funcion $f(\bar{u})$ y por tanto no depende de h .

Teorema 1.6. El metodo de Euler es consistente y convergente siempre que f tenga primera derivada continua.

Demostracion. Claramente el metodo de Euler es consistente pues $\Phi(t,\ \bar{u};\ h) = f(t,\ \bar{u})$ para toda h, en particular para $h = 0$. Ahora bien, el metodo de Euler es convergente, pues si se restringe el dominio de f' al compacto $t_0 \leq t \leq s$; $\bar{u} \in K$ donde K es compacto, entonces f' es acotada por ser imagen continua de un compacto y por lo tanto f $(=\Phi)$ es de Lipschitz, es decir :

$$|\Phi(t,\ \bar{z}(t)) - \Phi(t,\ \bar{u}(t))| \leq L|\bar{z} - \bar{u}|$$

Esta ultima desigualdad junto con la consistencia del Metodo son las hipotesis requeridas para aplicar el teorema 1.4., por lo que el metodo es convergente ∎.

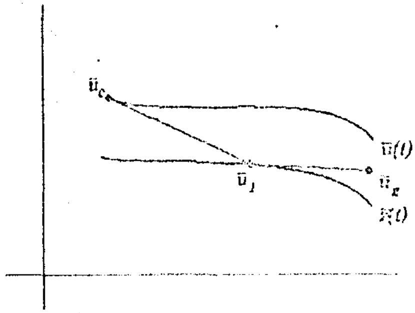


Figura 2
Metodo de Euler

Recordemos que el Metodo de Euler es un metodo de primer orden, pues es un truncamiento en el primer termino de la expansion de Taylor de la funcion incremento exacto Δ. Graficamente, el metodo de Euler tiene la interpretacion dada por la figura 2, donde $\bar{u}_0$ es la condicion inicial, $\bar{u}(t)$ la solucion al sistema con $\bar{u}_0$ como condicion inicial y $\bar{z}(t)$ es otra solucion al mismo sistema con diferente condicion inicial. Entonces el metodo de Euler en t_0 nos da $\bar{u}_0$ como primer punto de la solucion, y para $t_0 + h$ obtenemos $\bar{u}_1 = \bar{u}_0 + hf(\bar{u}_0)$ $\bar{u}_1$ se encuentra ubicado en la tangente a $\bar{u}(t)$ en $\bar{u}_0$, y por tanto no es probable que $\bar{u}_1$ pertenezca a la grafica de $\bar{u}(t)$, por lo que podemos considerar que estamos frente a un nuevo problema con condicion inicial u_1 con solucion $\bar{z}(t)$ y ahora tendremos que $\bar{u}_2$ se encuentra sobre la tangente a $\bar{z}(t)$ en $\bar{u}_1$, y asi sucesivamente.

Una rutina en pseudocodigo que implemente el Metodo de Euler en para dos variables podria ser

$$
\begin{array}{l}
x_n := x_0; \\
y_n := y_0; \\
\text{Repeat} \\
\quad x_{n-1} := x_n; \\
\quad y_{n-1} := y_n; \\
\quad x_n := x_{n-1} + hf_1(x_{n-1}, y_{n-1}); \\
\quad y_n := y_{n-1} + hf_2(x_{n-1}, y_{n-1}); \\
\text{Endrepeat;}
\end{array}
$$

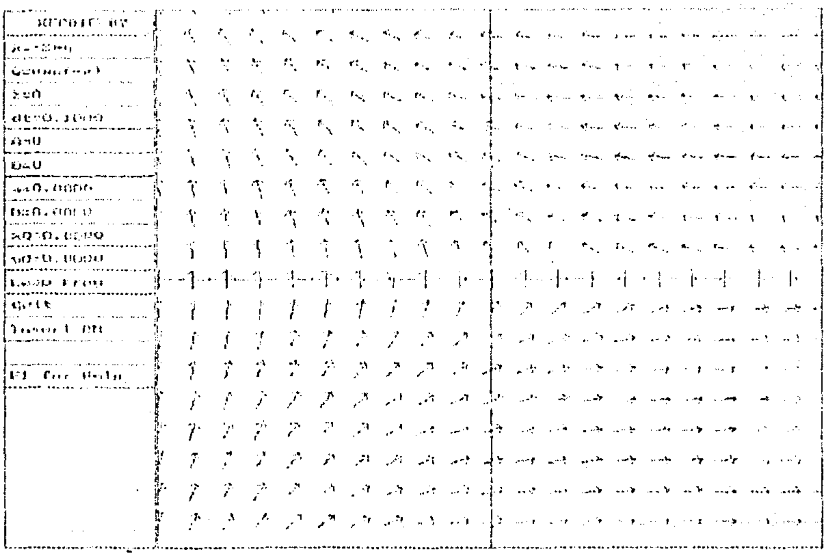


Figura 9
Grafica de flujo de un sistema

Ejemplo 1.2. Considere el sistema descrito por la grafica del flujo de la figura 3, con ecuacion

$$\begin{aligned} \dot{x} &= -2y \\ \dot{y} &= e^{-x} \end{aligned} \qquad \text{y valor inicial } t_o = 1,\ x_o = 0,\ y_o = -1,$$

con solucion $(x, y) = (\ln t^2 \ , -1/t)$.

En la tabla 1 se encuentran los puntos de la solucion exacta, la solucion del metodo de Euler y la diferencia entre ambas con un incremento h = 0.1

n	t	$x(t)$	$y(t)$	x	y	$\lvert x-x(t)\rvert$	$\lvert y-y(t)\rvert$
0	1.0	0.0000	-1.0000	0.000	-1.000	0.0000	0.0000
1	1.1	0.1906	-0.9090	0.200	-0.900	0.0093	0.0090
2	1.2	0.3646	-0.8333	0.380	-0.818	0.0153	0.0152
3	1.3	0.5247	-0.7692	0.543	-0.749	0.0188	0.0194
4	1.4	0.6729	-0.7142	0.693	-0.691	0.0206	0.0226
5	1.5	0.8109	-0.6666	0.831	-0.641	0.0209	0.0249
6	1.6	0.9400	-0.6250	0.960	-0.598	0.0202	0.0268
7	1.7	1.0612	-0.5882	1.079	-0.559	0.0186	0.0283
8	1.8	1.1755	-0.5555	1.191	-0.525	0.0162	0.0296
9	1.9	1.2837	-0.5263	1.297	-0.495	0.0133	0.0307
10	2.0	1.3862	-0.5000	1.386	-0.468	0.0098	0.0317
15	2.5	1.8325	-0.4000	1.819	-0.363	0.0135	0.0361
20	3.0	2.1972	-0.3333	2.152	-0.292	0.0446	0.0403
30	4.0	2.7725	-0.2500	2.647	-0.200	0.1250	0.0495
40	5.0	3.2188	-0.2000	2.990	-0.140	0.2282	0.0599
50	6.0	3.5835	-0.1666	3.228	-0.095	0.3549	0.0715
75	8.5	4.2801	-0.1176	3.491	-0.100	0.7884	0.1070
100	11.0	4.7957	-0.0909	3.360	-0.067	1.4356	0.1588

Tabla 1
Calculo de la solucion

Si graficamos ambas soluciones obtenemos

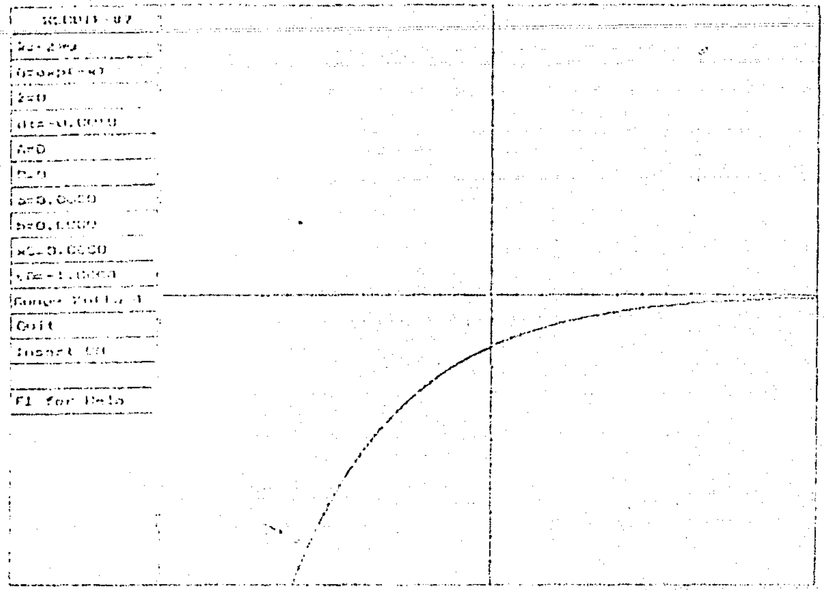


Figura 4
Solucion del Sistema

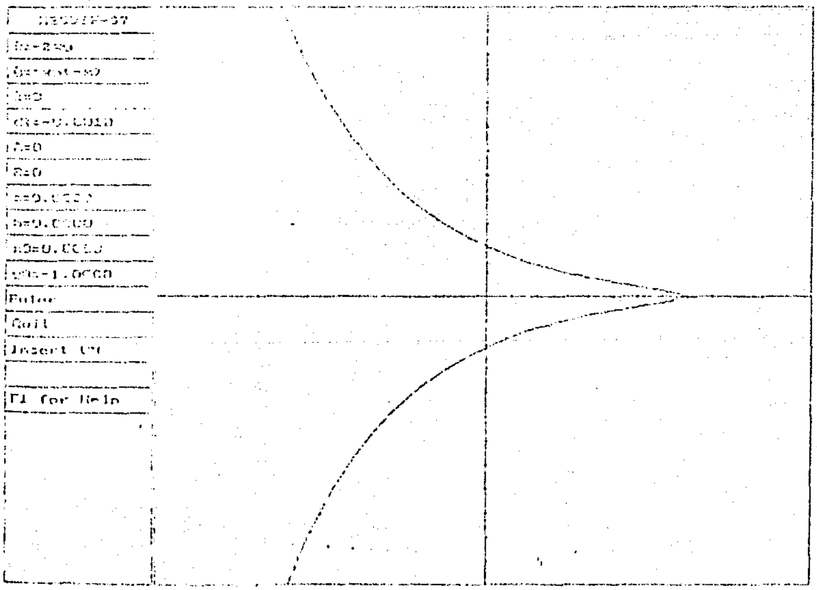


Figura 5
Metodo de Euler

SECCION 1.4
EL METODO DE EULER 2

Durante la implementacion del metodo de Euler en dos variables en Pascal para graficar soluciones de Ecuaciones Diferenciales surge de manera natural la pregunta de como hacer mas eficiente dicho metodo. Una de las primeras ideas que surgio fue eliminar las lineas 4 y 5 de la rutina listada en la pagina 24, obteniendo

```
x   := x0;
y   := y0;
Repeat
      x   := x + hf1(x, y);
      y   := y + hf2(x, y);
Endrepeat;
```

Claramente esta modificacion en la codificacion conlleva una modificacion en el metodo de Euler pues formalmente obtenemos el siguiente metodo :

$$x_n = x_{n-1} + h\ f_1(x_{n-1}, y_{n-1})$$
$$y_n = y_{n-1} + h\ f_2(x_n, y_{n-1})$$

observe n en lugar de $n-1$

Lo verdaderamente interesante de esta modificacion no es en si la supresion de dos lineas de codigo, que no son realmente muy costosas en terminos de tiempo maquina, sino el incremento en precision aparente al graficar soluciones periodicas de sistemas de ecuaciones diferenciales en el plano. Las graficas correspondientes a las figuras 6 y 7 son soluciones aproximadas de un Sistemas de Ecuaciones Diferenciales calculadas por el metodo de Euler y el metodo de Euler 2 con el mismo paso o incremento h que muestran la dramatica diferencia entre estos metodos pues la solucion «real» esta dada por la grafica obtenida con el metodo de Euler 2. De aqui surge la idea de estudiar las propiedades del metodo de Euler 2.

Ejemplo 1.3. Graficas de la solucion del sistema

$$\dot{x} = -y + 2 \qquad\qquad \dot{y} = 2x - 4$$

con valores iniciales $x_0 = 3$, $y_0 = 2$.

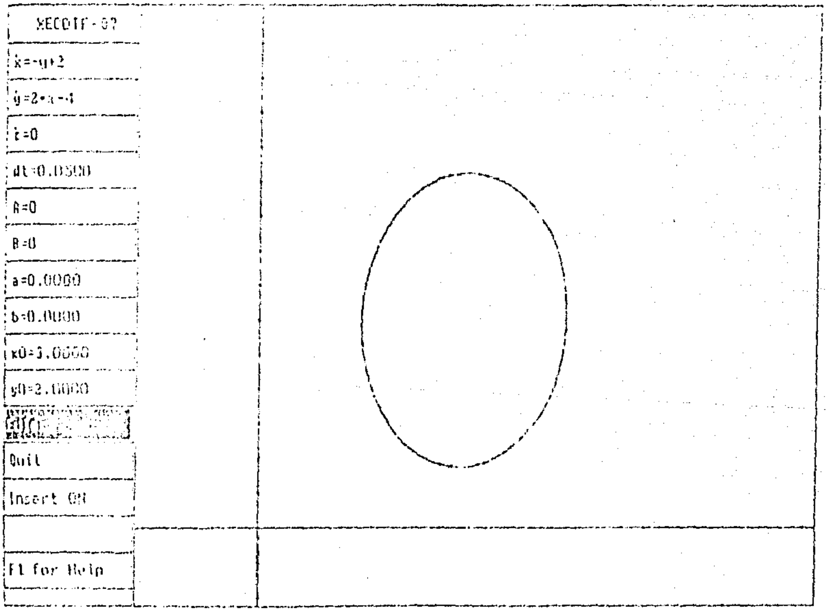


Figura 6
Metodo de Euler 2

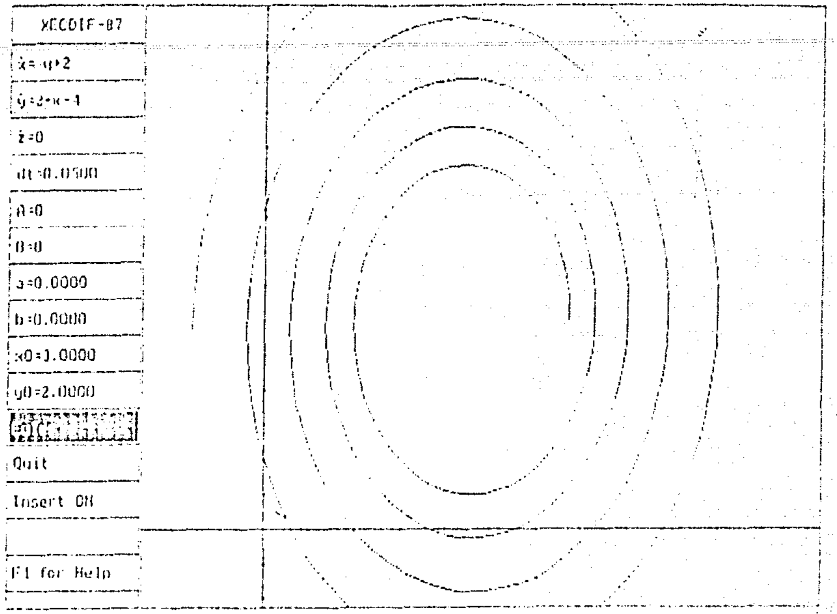


Figura 7
Metodo de Euler

Primeramente tenemos que en estricto el metodo de Euler 2 no es de un solo paso, pues y_n se encuentra expresada en terminos de (x_n, y_{n-1}) y no de (x_{n-1}, y_{n-1}), como corresponde a un metodo de un solo paso. De hecho, a primera vista, el metodo de Euler 2 semeja en ciertos aspectos a un metodo predictor-corrector. Sin embargo, reexpresando este metodo, veremos sus similitudes con el metodo de Euler y de ahi que se le haya llamado a esta modificacion metodo de Euler 2.

Verifiquemos esta informacion : si sustituimos a x_n por $x_{n-1} + hf_1(x_{n-1}, y_{n-1})$ en la expresion de y_n obtenemos :

$$x_n = x_{n-1} + hf_1(x_{n-1}, y_{n-1})$$

$$y_n = y_{n-1} + hf_1(x_{n-1} + hf_1(x_{n-1}, y_{n-1}), y_{n-1}) \;,$$

con lo que Φ esta determinada por

$$\Phi((x_n, y_n); h) = \begin{bmatrix} x_{n-1} + hf_1(x_{n-1}, y_{n-1}) \\ y_{n-1} + hf_1(x_{n-1} + hf_1(x_{n-1}, y_{n-1}), y_{n-1}) \end{bmatrix}$$

mostrando esto que el metodo de Euler 2 es un metodo de un solo paso.

Sistemas Lineales. Ahora bien estudiemos el comportamiento del Metodo de Euler 2 en el caso de Sistemas de Ecuaciones Diferenciales Lineales en dos variables, es decir, estudiaremos el sistema :

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{o equivalentemente}$$

$$\begin{aligned} \dot{x} &= ax + by \\ \dot{y} &= cx + dy \end{aligned}$$

Por Euler 2 obtenemos :

$$x_n = x_{n-1} + h(ax_{n-1} + by_{n-1}) \qquad \ldots(1.3)$$
$$y_n = y_{n-1} + h(cx_n + dy_{n-1}) \qquad \ldots(1.4)$$

sustituyendo x_n (1.3) en la ecuacion (1.4) se tiene :

$$y_n = y_{n-1} + h(c[x_{n-1} + h(ax_{n-1} + by_{n-1})] + dy_{n-1})$$

y reorganizando terminos :

$$y_n = y_{n-1} + h(cx_{n-1} + dy_{n-1}) + h^2c(ax_{n-1} + by_{n-1})$$

donde los primeros dos terminos de la parte derecha de la ecuacion son el valor propuesto por el metodo de Euler, y el tercero un factor corrector de y_n.

Ejemplo 1.4. Considere el sistema :

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Este sistema tiene como soluciones circulos centrados en el origen, y el Metodo de Euler 2 da en este caso

$$\begin{aligned} x \ &= x_{n-1} - hy_{n-1} \\ y_n &= y_{n-1} + hx_{n-1} - h^2 y_{n-1} \end{aligned}$$

Por otro lado, del metodo de Euler resulta la aproximacion :

$$\begin{aligned} x \ &= x_{n-1} - hy_{n-1} \\ y_n &= y_{n-1} + hx_{n-1} \end{aligned}$$

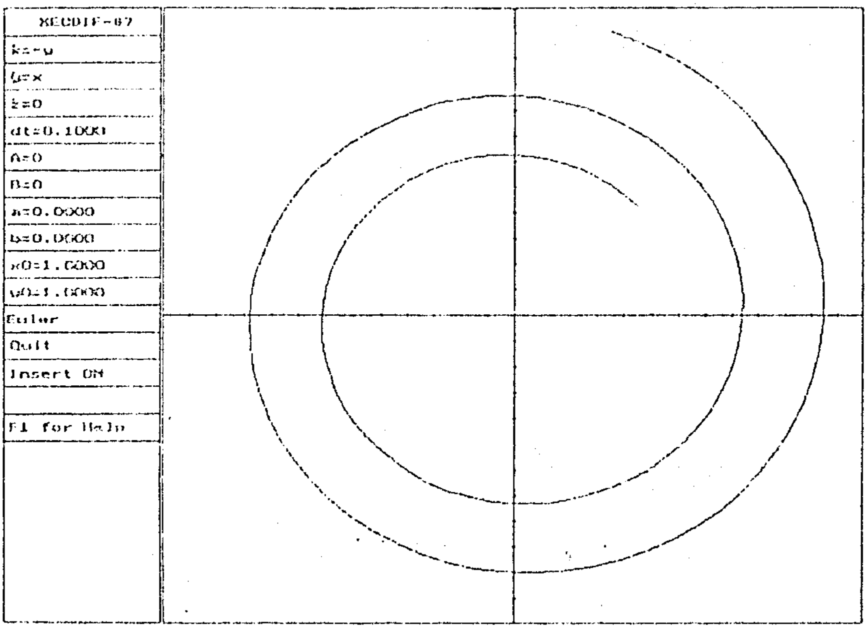


Fig. 8

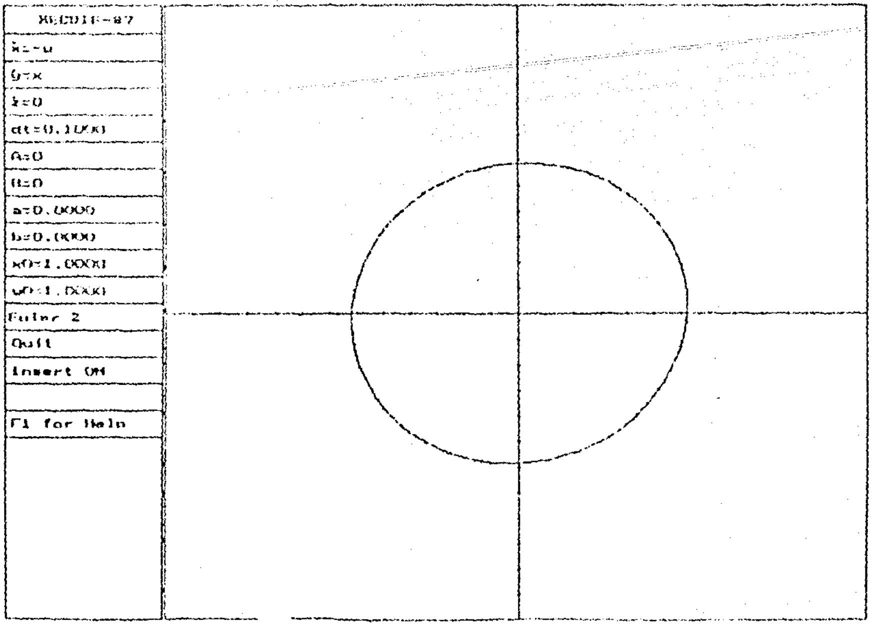


Figura 9
Metodo de Euler 2

Las figuras 8 y 9 muestran la diferencia grafica entre ambos metodos originada por el factor corrector de y_n.

Sistemas No Lineales. Veamos el caso para Sistemas No Lineales, Tenemos :

$$\begin{aligned} x_n &= x_{n-1} + h\, f_1(x_{n-1}, y_{n-1}) \\ y_n &= y_{n-1} + h\, f_2(x_n, y_{n-1}) \end{aligned} \qquad \ldots(1.5)$$

Si f es lisa, como funcion de x tenemos que para h suficientemente pequena :

$$f_2(x_n, y_{n-1}) = f_2(x_{n-1}, y_{n-1}) + f_{2x}(x_{n-1}, y_{n-1})(x_n - x_{n-1}) + \frac{1}{2!} f_{2xx}(\xi)(x_n - x_{n-1})^2$$

pero como :

$$x_n = x_{n-1} + hf_1(x_{n-1}, \ y_{n-1}) \qquad \text{por tanto}$$

$$x_n - x_{n-1} = hf_1(x_{n-1}, y_{n-1}) \ ,$$

y se tiene :

$$\begin{aligned} f_2(x_{n-1}, y_{n-1}) = f_2(x_{n-1}, y_{n-1}) \\ + hf_{2x}(x_{n-1}, y_{n-1})f_1(x_{n-1}, y_{n-1}) \\ + \frac{h^2}{2!} f_{2xx}(\xi)f_1(x_{n-1}, y_{n-1})^2 \end{aligned}$$

Sustituyendo en (1.5) :

$$\begin{aligned} y_n = y_{n-1} + h\Big[f_2(x_{n-1}, y_{n-1}) \\ + hf_{2x}(x_{n-1}, y_{n-1})f_1(x_{n-1}, y_{n-1}) \\ + \frac{h^2}{2!} f_{2xx}(\xi)f_1(x_{n-1}, y_{n-1})^2\Big] \end{aligned}$$

y reordenando se tiene :

$$\begin{aligned} y_n = y_{n-1} + hf_2(x_{n-1}, y_{n-1}) \\ + h^2 f_1(x_{n-1}, y_{n-1})f_{2x}(x_{n-1}, y_{n-1}) \\ + \frac{h^3}{2!} f_{2xx}(\xi)f_1(x_{n-1}, y_{n-1})^2 \end{aligned}$$

El coeficiente de h^2 sera llamado en ocasiones la *correccion de* y_n o simplemente la *correccion*.

Este factor corrector, que tambien fue encontrado en el caso de los sistemas lineales, tiene una justificacion intuitiva interesante. Pensemos en el metodo de Euler como de dos pasos relativamente independientes el uno del otro : el primero consiste en calcular x_n y el segundo en calcular y_n. Entonces, de acuerdo con la visto en 1.3 el metodo de Euler avanza por tangentes a las soluciones del sistema donde en cada paso es muy probable que el valor propuesto como siguiente aproximacion de la solucion este en una solucion distinta que el inmediato anterior. Por lo que se puede pensar que el metodo de Euler da aproximaciones equivocadas en cada uno de los dos pasos que se mencionan arriba. Entonces en este sentido el metodo de Euler 2 tambien da una aproximacion equivocada en el primer paso, a saber, el calculo de x_n, por lo que se debe entonces hacer una correccion en el segundo paso del metodo de Euler, es decir, en el calculo de y_n y esta correccion debe depender del error que se cometio en el primer paso, o sea, debe estar en terminos de x. Como se ve, este es el caso del factor corrector del metodo de Euler 2. Mas adelante en esta misma seccion se demostraran algunas propiedades de este factor corrector.

Por ahora se demostraran algunas propiedades del metodo de Euler 2 en su conjunto.

Teorema 1.7. El metodo de Euler 2 es consistente y convergente.

Demostracion. El metodo de Euler 2 es consistente pues dado que :

$$\Phi((x_n, y_n); h) =$$

$$= \begin{bmatrix} f_1(x_n, y_n) \\ f_2(x_n, y_n) + hf_1(x_n, y_n)f_{2x}(x_n, y_n) \\ \qquad + \frac{h^2}{2!} f_{2xx}(\xi)f_1(x_n, y_n)^2 \end{bmatrix}$$

para $h = 0$ se tiene $\Phi((x_n, y_n); h) = f(x_n, y_n)$.

Ahora bien, el metodo de Euler 2 es convergente pues si se restringe el dominio de f' al compacto $t_0 \leq t \leq s$; $\bar{u} \in K$ donde K es compacto, entonces f y su derivadas son acotadas por ser imagen continua de un compacto, y por lo tanto las suma y multiplicacion de f_1, f_2 y f_{2x} $(=\Phi)$ es de Lipschitz, es decir :

$$|\Phi(t\,,\ \bar{z}(t)) - \Phi(t\,,\ \bar{u}(t))| \le L\,|\bar{z} - \bar{u}|$$

Esta ultima desigualdad, junto con la consistencia del Metodo son las hipotesis requeridas para aplicar el teorema 1.2., por lo que el metodo es convergente. ∎

Teorema 1.8. Sea $\bar{u}$ la solucion al sistem $\dot{\bar{u}} = f(\bar{u})$ con condicion inicial $\bar{u}_0$, y $\{\bar{u}_n\}_{n=1,2,...}$ la solucion de valores obtenidos por el metodo de Euler 2. Entonces para $t \in [a,\ b]$, $h \le h_0$, se tiene que :

$$|\bar{u}_n - \bar{u}(t_n)| \le hN_1E_L(t - a)$$

donde N_1 y E_L estan definidos como en el teorema 1.5.

Demostracion.- Definamos

$$\Phi^*(t\,,\ (x,\ y);\ h) = \begin{cases} f_1(x,\ y) \\ f_2(x,\ y) + hf_1(x,\ y)\ f_{2x}(x,\ y) \end{cases}$$

y como se tiene que

$$\Delta(t\,,\ \bar{u};\ h) = \frac{\bar{u}(t + h) - \bar{u}(t)}{h}$$

y por tanto

$$|\Phi^*(t\,,\ \bar{u};\ h) - \Delta(t\,,\ \bar{u};\ h)| =$$

$$\left| \left(f_1(\bar{u})\,,\ f_2(\bar{u}) + hf_1(\bar{u})f_{2x}(\bar{u})\right) - \frac{\bar{u}(t + h) - \bar{u}(t)}{h} \right|$$

y por lo tanto

$$= \left| \left[f_1(\overline{u}),\ f_2(\overline{u}) + hf_1(\overline{u})f_{2x}(\overline{u}) \right] \right.$$

$$\left. - \frac{\overline{u}(t) + hf(\overline{u}) + \frac{h^2}{2} f''(\overline{u}(\xi)) - \overline{u}(t)}{h} \right|$$

$$= \left| \left[f_1(\overline{u}),\ f_2(\overline{u}) + hf_1(\overline{u})f_{2x}(\overline{u}) \right] - \frac{hf(\overline{u}) + \frac{h^2}{2} f''(\overline{u}(\xi))}{h} \right|$$

$$= h\ |(0,\ f_1(\overline{u})f_{2x}(\overline{u})) - 1/2\ f''(\overline{u}(\xi))|$$

Dado que esta ultima expresion esta definida en un compacto tanto $f_1(\overline{u})$, $f_{2x}(\overline{u})$ y $f''(\overline{u}(\xi))$ estan acotadas. y de ahi que :

$$h\ |(0,\ f_1(\overline{u})f_{2x}(\overline{u})) - 1/2\ f''(\overline{u}(\xi))|$$

$$\leq h|(0,\ f_1(\overline{u})f_{2x}(\overline{u}))| + |\ 1/2\ f''(\overline{u}(\xi))|$$

$$\leq h\ \left[M_1M_2 + 1/2\ M_3 \right] = hN,$$

donde

$$|f_1(\overline{u})| \leq M_1$$

$$|f_{2x}(\overline{u})| \leq M_2$$

$$|f''(\overline{u})| \leq M_3$$

$$N = M_1M_2 + 1/2M_3$$

Si por otra parte definimos :

$$\overline{u}_0 = \overline{\eta} \qquad\qquad n = 0,\ 1,\ 2, \ldots$$

$$\overline{u}_{n+1} = \overline{u}_n + h[\Phi^*(t_n,\ \overline{u}_n;\ h) + h^2\ k\ \theta_n] \qquad t_n \in [a,\ b],$$

donde

$$|f_{2xx}(\xi)f_1(x_n,\ y_n)^2| \leq k$$

$$\theta_n = \frac{1}{k}\ f_{2xx}(\xi)f_1(x_n,\ y_n)^2$$

aplicando el teorema 1.5. con $p = 1$, $q = 2$, $r = 1$, $N_1 = N + Mh_0$ obtenemos lo que se queria demostrar. ∎

Ahora bien estudiemos el valor del factor de segundo orden en la expresion de y_n (la correccion) segun el metodo de Euler 2. Principalmente analizaremos cualitativamente el comportamiento de la correccion $h^2 f_1(x_n, y_n) f_{2x}(x_n, y_n)$.

Inicialmente veamos el caso de soluciones periodicas pues es donde el metodo de Euler 2 es propuesto. En cualquier solucion cerrada destacaremos al menos cuatro puntos, a saber, aquellos donde se anule $\dot{x}$ o $\dot{y}$ (ver figura 10).

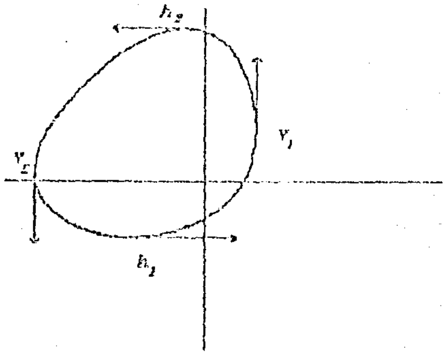


Figura 10
Solucion Periodica

Lema 1.9. Sea $\bar{u}(t)$ una solucion periodica al sistema $\dot{y} = f(y)$ y ea $s \in [t_0, t_f]$ tal que $\bar{u}(s) = (0, v)$. Entonces el factor de correccion del metodo de Euler 2 coincide con la correccion del metodo de Euler.

Demostracion .- La expresion de x_n del metodo de Euler 2 coincide con la expresion de x_n del metodo de Euler. Solo resta ver el caso de y_n, para el caso de Euler 2, Factorizando f_1 tenemos :

$$y_n = y_{n-1} + hf_2(x_{n-1}, y_{n-1}) +$$

$$+ f_1(x_{n-1}, y_{n-1}) \left[h^2 f_{2x}(x_n, y_n) + \frac{h^3}{2!} f_{2xx}(\xi) f_1(x_{n-1}, y_{n-1}) \right]$$

y como por hipotesis f_1 es cero tenemos que

$$y_n = y_{n-1} + hf_2(x_{n-1}, y_{n-1})$$

que es la expresion de y_n segun el metodo de Euler. □

El siguiente teorema afirma que el factor de correccion de y en el metodo de Euler 2 es positivo cuando la correccion real dada por $u_2(s + h) - \left[u_2(s) + hf_2(\bar{u}(s))\right]$ lo es, y viceversa cuando este factor es negativo tambien lo es la correccion dada por el metodo de Euler 2.

Lema 1.10. Sea $\bar{u}(t)$ una solucion periodica al sistema $\dot{\bar{u}} = f(\bar{u})$, y sea s $\in [t_0, t_f]$ tal que $u(\bar{s}) = (v, 0)$. Entonces, el factor de correccion del metodo de Euler 2 en este punto tiene el signo correcto, es decir :

$$\left[h\, f_{2x}(\bar{u}(s))f_1(\bar{u}(s))\right] \left[u_2(s + h) - \left[u_2(s) + hf_2(\bar{u}(s))\right]\right] \geq 0$$

Demostracion .- Como por hipotesis $f_2(\bar{u}(s)) = 0$ es equivalente demostrar que

$$\left[h\, f_{2x}(\bar{u}(s))f_1(\bar{u}(s))\right] \left[u_2(s + h) - u_2(s)\right] \geq 0,$$

o tambien es equivalente

$$\left[f_{2x}(\overline{u}(s))f_1(\overline{u}(s))\right] \Delta_2(\overline{u}(s);\ h) \geq 0$$

por definicion :

$$\Delta(\overline{u}(s);\ h) = \frac{\overline{u}(s+h)-\overline{u}(s)}{h}, \qquad \text{y por tanto}$$

$$\Delta_2(\overline{u}(s);\ h) = \frac{u_2(s+h)-u_2(s)}{h}$$

obteniendo

$$\left[h\ f_{2x}(\overline{u}(s))f_1(\overline{u}(s))\right] \Delta_2(\overline{u}(s);\ h) =$$

$$= f_{2x}(\overline{u}(s))\ f_1(\overline{u}(s))\ \left[u_2(s+h) - u_2(s)\right]$$

y expandiendo en un termino de la serie de Taylor a $u_2(s+h)$ tenemos :

$$u_2(s+h) = u_2(s) + h\dot{u}_2(s + \xi h) \quad 0 \leq \xi \leq 1,$$

de lo que se obtiene

$$f_{2x}(\overline{u}(s))\ f_1(\overline{u}(s))\ \left[u_2(s) + h\dot{u}_2(s + \xi h) - u_2(s)\right] =$$

$$f_{2x}(\overline{u}(s))\ f_1(\overline{u}(s))\ h\ \dot{u}_2(s + \xi\ h)$$

Como $\overline{u}$ es solucion del sistema de ecuaciones diferenciales, y dado que $h > 0$, es equivalente demostrar que $f_{2x}(\overline{u}(s))\quad f_1(\overline{u}(s))\ f_2(\overline{u}(s + \xi\ h)) \geq 0$

Ahora bien aplicando la regla de la cadena

$$\frac{d}{dt} f_2 = f_{2x}\ \dot{x} + f_{2y}\ \dot{y} = f_{2x}\ f_1 + f_{2y}\ f_2$$

y como por hipotesis $f_2(\bar{u}(s)) = 0$ se tiene

$$\frac{d}{dt} f_2(\bar{u}(s)) = f_{2x}(\bar{u}(s))\, f_1(\bar{u}(s)) + f_{2y}(\bar{u}(s))\, f_2(\bar{u}(s))$$
$$= f_{2x}(\bar{u}(s))\, f_1(\bar{u}(s))$$

y como $f_2(\bar{u}(s)) = 0$ y $h > 0$ se tiene que

$$\frac{d}{dt} f_2 = f_{2x}(\bar{u}(s))\, f_1(\bar{u}(s)) > 0 \quad \Longleftrightarrow \quad f_2(\bar{u}(s + \xi h)) > 0$$

pues dado que f_2 se anula en $\bar{u}(s)$, entonces si f_2 tiene derivada positiva en este puento por fuerza ha de ser creciente en una vecindad y de ahi que $f_2(\bar{u}(s + \xi h)) > 0$. Inversamente si f_2 tiene derivada negativa entonces es decreciente en una vecindad y se tiene $f_2(\bar{u}(s + \xi h)) < 0$.

Con lo que hemos obtenido que el producto

$$f_{2x}(\bar{u}(s))\, f_1(\bar{u}(s))\, f_2(\bar{u}(s + \xi h))$$

es positivo. ◻

Lema 1.11. Una condicion necesaria para que el metodo de Euler 2 de mejores aproximaciones a la verdadera solucion es que

$$\left[h\, f_{2x}(\bar{u}(s))/f_1(\bar{u}(s))\right] \left[u_2(s + h) - \left[u_2(s) + h f_2(\bar{u}(s))\right]\right] \geq 0$$

Demostracion.- Este hecho se sigue de manera evidente de las argumentaciones hechas en la demostracion del Lema 1.10. ◻

C A P I T U L O 2

2.1. COMPARACIONES DE METODOS

Problemas

1ª Clase de Problemas

2ª Clase de Problemas

2.2. OTROS RESULTADOS

Aplicaciones a otros metodos

SECCION 2.1[2]
COMPARACIONES DE METODOS

Restringiremos nuestra atencion a metodos numericos que resuelven sistemas de primer orden de ecuaciones diferenciales. Los problemas de prueba y los criterios de comparacion fueron escogidos de tal manera que los resultados obtenidos sobre un metodo en particular dependan principalmente en que tan bien puede realizar los pasos de integracion relativamente rutinarios bajo diversos requerimientos de precision en los resultados. Por lo tanto, deliberadamente se ha intentado suprimir los efectos de caracteristicas especiales de los metodos, como podrian ser su habilidad para escoger el paso inicial, el manejo de discontinuidades, o el trato de ecuaciones rigidas. Estas caracteristicas son importantes, pero pueden ser consideradas separadamente. En particular, aunque algunas de las ecuaciones consideradas son ligeramente rigidas, los metodos disenados especialmente para este tipo de problemas no fueron considerados aqui.

Desde luego, los problemas seleccionados fueron escogidos en gran manera arbitrariamente. Se intento escoger problemas realistas y representativos, sin embargo, como ya se indico, se han evitado problemas con singularidades en su solucion o que son mas que moderadamente rigidos. Los problemas estan agrupados en 4 grandes clases, 5 problemas en cada clase. Cada problema esta definido por una ecuacion diferencial y una tolerancia en el

---

Este capitulo esta fuertemente basado en los resultados propuestos en el articulo COMPARING NUMERICAL METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS por T. E. HULL, W. H. ENRIGHT, B. M. FELLEN Y E. SEDGEWICK publicado en el SIAM JOURNAL ON NUMERICAL ANALYSIS VOL. 9, NO. 4 DICIEMBRE 1972, en diversas ocasiones se presentan paragrafos y resultados de dicho articulo en el texto.

error. El problema es encontrar una solucion aproximada manteniendo el error local menor a la tolerancia. Se han mantenido las tolerancias lo suficientemente grandes para que el error de redondeo de la computadora no cause ningun problema.

Problemas.- Un problema p queda determinado por seis elementos $p = <f, t_0, \bar{u}_0, t_f, \tau, h_{max}>$, donde los tres primeros terminos definen la ecuacion diferencial $\dot{\bar{u}} = f(t, \bar{u})$ con condicion inicial $\bar{u}(t_0) = \bar{u}_0$. El cuarto elemento t_f es el valor final de t e indica que la integracion debe ser realizada sobre el intervalo $[t_0, t_f]$. El siguiente termino τ es la tolerancia en el error. Finalmente, el ultimo termino en la especificacion de un problema es el tamano maximo del paso h_{max}.

Una vez definido un problema podemos considerar clases de problemas P, con una medida de probabilidad asociada con la clase de problemas. Entonces la clase se convierte en un espacio muestral de problemas. Asi mismo, podemos considerar metodos m y clases de metodos M. Una funcion costo c es una medida sobre parejas (p, m) donde $c(p, m)$ es el costo de resolver el problema p con el metodo m. Finalmente definiremos una funcion costo de un metodo relativa a una clase de problemas, es decir, $C(P, m)$. Por lo que se puede afirmar que m es mejor que m' en la clase de problemas P de acuerdo con el costo C si $C(P,m) < C(P,m')$.

En este caso en particular se usara la funcion de costo determinada por el tiempo total de calculo de la solucion.

Comparemos ahora bajo estos criterios los metodos de Adams-Bashforth de orden 2, Runge-Kutta de orden 4, Euler 2 y Heun, las definiciones particulares de estos metodos fueron obtenidas en [Burden, 1985], [King, 1984], [Henrici, 1983] y [Pizer, 1983].La comparacion se hara sobre cuatro familias de ecuaciones, cada familia con 5 ecuaciones. En este caso particular la tolerancia no esta determinada por un valor especifico de τ, sino de una manera visual en terminos de que tan «cerca» esta la solucion aproximada de la real en la grafica. El tamano del paso esta ajustado para obtener soluciones aproximadamente equivalentes.

Clase A. Ecuaciones de una sola variable.[a]

A1: El inverso de la exponencial

$$\dot{y} = -y \qquad y(0) = 1 \qquad (y = C\, e^{-x},\ C = 1).$$

A2: Un caso especial de la ecuacion de Riccati

$$\dot{y} = -y^3/2 \qquad y(0) = 1 \qquad (y = 1/\sqrt{x + C}\ ,\ C = 1).$$

A3: Un problema oscilatorio

$$\dot{y} = y\cos(x) \qquad y(0) = 1 \qquad (y = C\, e^{\operatorname{sen} x},\ C = 1).$$

A4: Una curva logistica

$$\dot{y} = \frac{y}{4}\left(1 - \frac{y}{20}\right) \qquad y(0) = 1 \qquad \left(y = \frac{20}{1+19Ce^{-x/4}},\ C = 1\right).$$

A5: Una curva espiral

$$\dot{y} = \frac{y - x}{y + x} \qquad y(0) = 4 \qquad (r = Ce^{-\theta},\ C = 4e^{\pi/2}).$$

---

[a] En esta clase consideraremos $x' = 1$, o lo que es equivalente $x = t$

Clase B. Sistemas de ecuaciones de pocas variables.

B1: El crecimiento de dos poblaciones en conflicto

$$\dot{x} = 2(x - xy) \qquad x(0) = 1$$
$$\dot{y} = -(y - xy) \qquad y(0) = 3.$$

B2: Una reaccion quimica lineal

$$\dot{x} = -x + y \qquad x(0) = 2$$
$$\dot{y} = x - 2y + z \qquad y(0) = 0$$
$$\dot{z} = y - z \qquad z(0) = 1.$$

B3: Una reaccion quimica no lineal

$$\dot{x} = -x \qquad x(0) = 1$$
$$\dot{y} = x - y^2 \qquad y(0) = 0$$
$$\dot{z} = y^2 \qquad z(0) = 0.$$

B4: La integral de superficie de toro

$$\dot{x} = -y - xz / (x^2 + y^2)^{1/2} \qquad x(0) = 3$$
$$\dot{y} = x - yz / (x^2 + y^2)^{1/2} \qquad y(0) = 0$$
$$\dot{z} = x / (x^2 + y^2)^{1/2} \qquad z(0) = 0.$$

B5: Ecuaciones de Euler del movimento de un cuerpo rigido sin fuerzas externas

$$\dot{x} = yz \qquad x(0) = 0$$
$$\dot{y} = -xz \qquad y(0) = 1$$
$$\dot{z} = -.51\, xy \qquad z(0) = 1.$$

Clase C. Ecuaciones de Orbita.

C1:

$$\dot{x} = z \qquad x(0) = 1 - \varepsilon$$

$$\dot{y} = u \qquad y(0) = 0$$

$$\dot{z} = -x/(x^2 + y^2)^{3/2} \qquad z(0) = 0$$

$$\dot{u} = -y/(x^2 + y^2)^{3/2} \qquad u(0) = \left(\frac{1 + \varepsilon}{1 - \varepsilon}\right)^{1/2}$$

$\varepsilon = .1$ (ε es la excentricidad de la orbita)

C2: Como C1 con $\varepsilon = .3$.

C3: Como C1 con $\varepsilon = .5$.

C4: Como C1 con $\varepsilon = .7$.

C5: Como C1 con $\varepsilon = .9$.

Clase D. Equaciones de orden superior.

D1: Derivada de las ecuaciones de Bessel de orden 1/2 con el origen recorrido una unidad a la izquierda

$$\dot{x} = y \qquad\qquad x(0) = 0.6713967...$$
$$y = -\left[\frac{y}{t+1} + \left(1 - \frac{0.25}{(t+1)^2}\right) x \right] \qquad y(0) = 0.0954995...$$

D2: Derivada de la ecuacion de Van der Pol

$$\dot{x} = y \qquad\qquad x(0) = 2$$
$$y = (1 - x^2)y - x \qquad\qquad y(0) = 0 .$$

D3: Derivada de la ecuacion de Duffing

$$\dot{x} = y \qquad\qquad x(0) = 0$$
$$y = x^3/6 - x + 2\,\mathrm{sen}(2.78535x) \qquad\qquad y(0) = 0 .$$

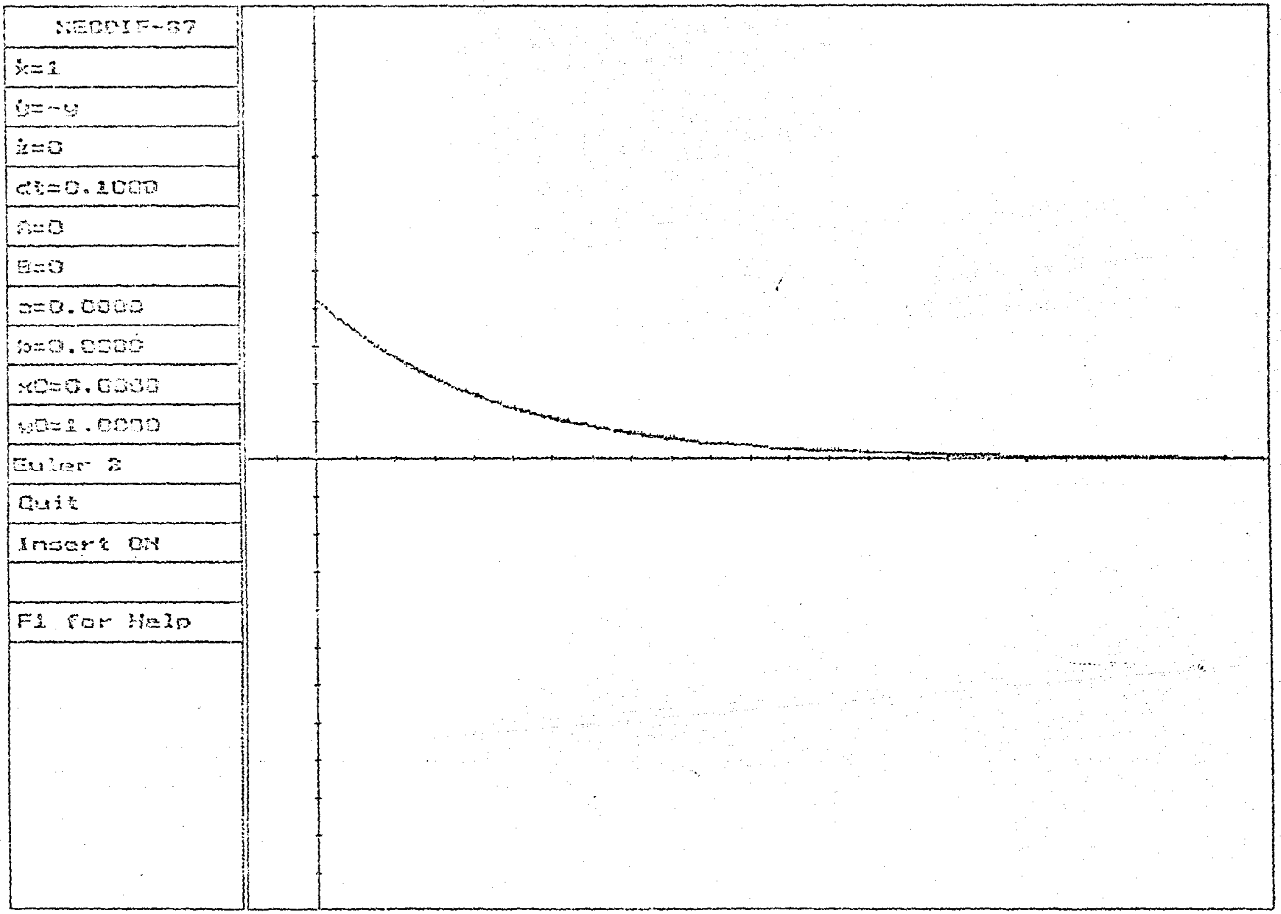
D4: Derivada de la ecuacion de un cuerpo en caida libre.

$$\dot{x} = y \qquad\qquad x(0) = 30$$
$$y = 0.032 - 0.4y^2 \qquad\qquad y(0) = 0 .$$

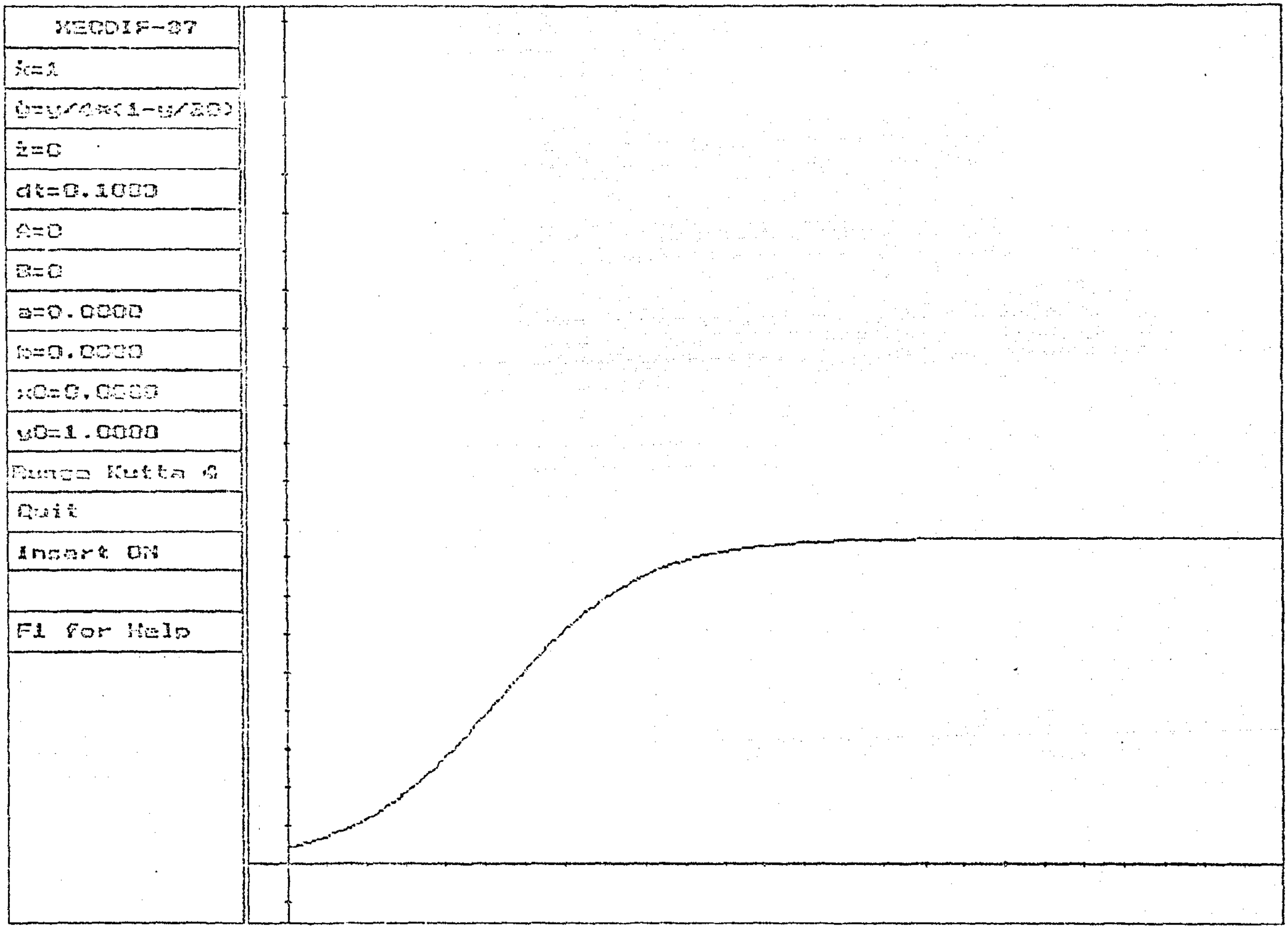
D5: Derivada de la ecuacion lineal de seguimiento

$$\dot{x} = y \qquad\qquad x(0) = 0$$
$$y = (1 + y^2)^{1/2}/(25 - t) \qquad\qquad y(0) = 0 .$$

A1

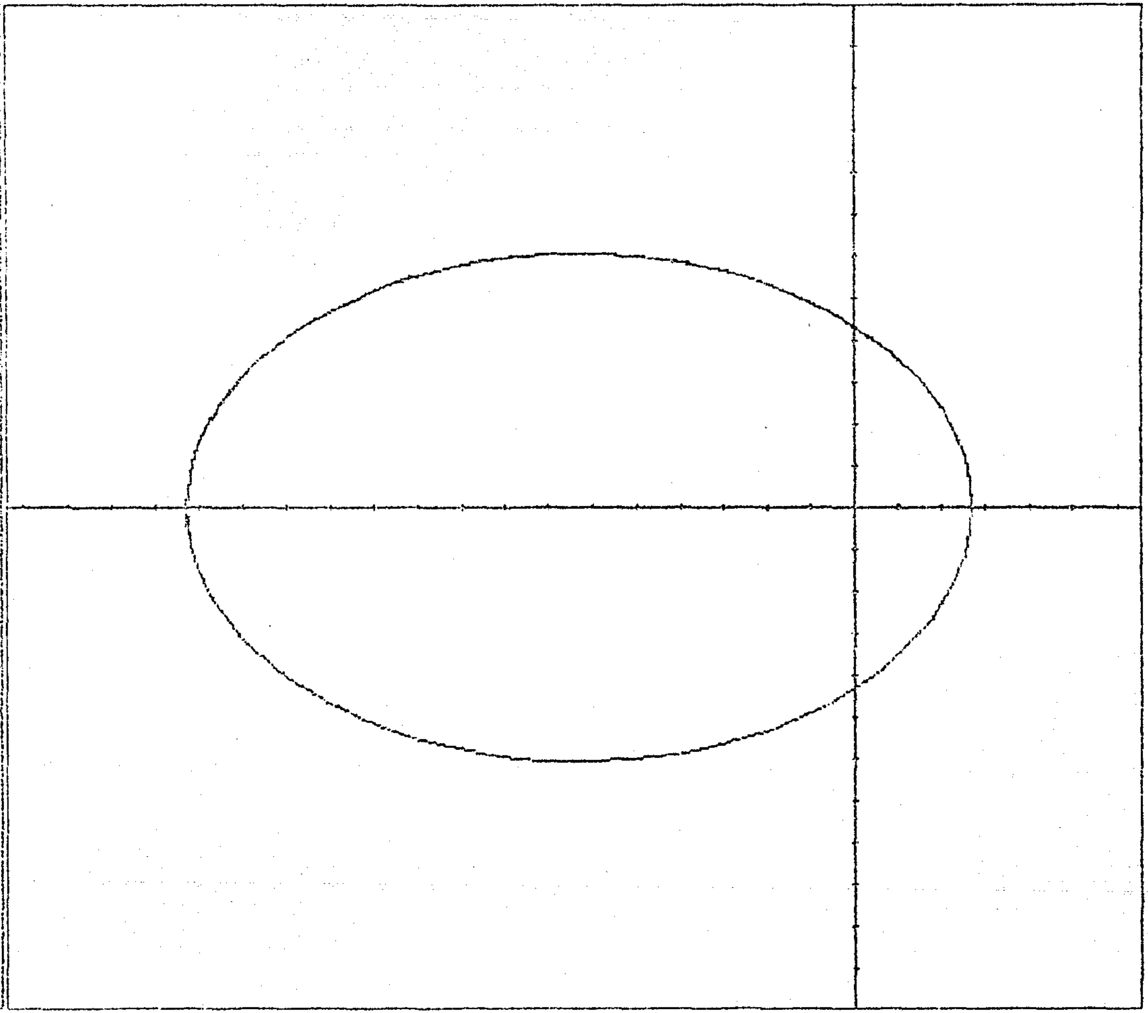

NEODIF-87
ẋ=1
ẏ=-y
ż=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=0.0000
y0=1.0000
Euler 2
Quit
Insert ON
F1 for Help

A4


MECDIF-87
ẋ=1
ẏ=y/4*(1-y/20)
ż=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=0.0000
y0=1.0000
Runge Kutta 4
Quit
Insert ON
F1 for Help

C5

MECDIF-87
ẋ=z
ẏ=u
ż=−x/A
dt=0.0500
A=(x*x+y*y)^1.
B=sqrt((1+a)/(
a=0.9000
b=0.0000
x0=0.3000
y0=0.0000
Runge Kutta 4
Quit
Insert ON
F1 for Help



El costo de resolver los problemas anteriores esta dado en la tabla 2 donde AB = Adams-Bashforth de orden 2, RK = Runge-Kutta de orden 4, E2 = Euler 2 y H = Heun.

ρ	AB	RK	E2	H
A1	3	7	2	3
A2	2	3	1	2
A3	2	5	1	3
A4	3	4	2	3
A5	1	3	1	2
B1	3	7	2	4
B2	1	2	1	1
B3	4	12	3	5
B4	16	3	9	17
B5	2	9	22	2

ρ	AB	RK	E2	H
C1	7	7	3	7
C2	16	7	8	17
C3	17	7	9	17
C4	17	7	8	17
C5	16	4	9	18
D1	2	3	1	2
D2	2	5	1	2
D3	1	3	0	1
D4	4	11	2	5
D5	2	4	1	2

Tabla 2
Tiempos Comparativos

Por otro lado y de manera independiente a las comparaciones anteriores se probo el metodo de Euler 2 en una serie de ejemplos obtenidos en diversos textos de ecuaciones diferenciales ordinarias que se citan en la bibliografia. Presentamos aqui la lista de ellos, y a continuacion las graficas de algunas soluciones particulares.

P1: $\dot{x} = -2x - y^2$
$\dot{y} = -y - x^2$

P2: $\dot{x} = -2x(x - 1)^2$
$\dot{y} = -2y$

P3: $\dot{x} = -y\cos(x)$
$\dot{y} = -\operatorname{sen}(x)$

P4: $\dot{x} = -y(1 - z)$
$\dot{y} = x(1 - z)$
$\dot{z} = 1 - x^2 - y^2$

P5: $\dot{x} = y - x^3 + x$
$\dot{y} = -x$

P6: $\dot{x} = y - x^2$
$\dot{y} = -x$

P7: $\dot{x} = y - x^{1/2}$
$\dot{y} = -x$

P8: $\dot{x} = y - \cos(x)$
$\dot{y} = \operatorname{sen}(x)$

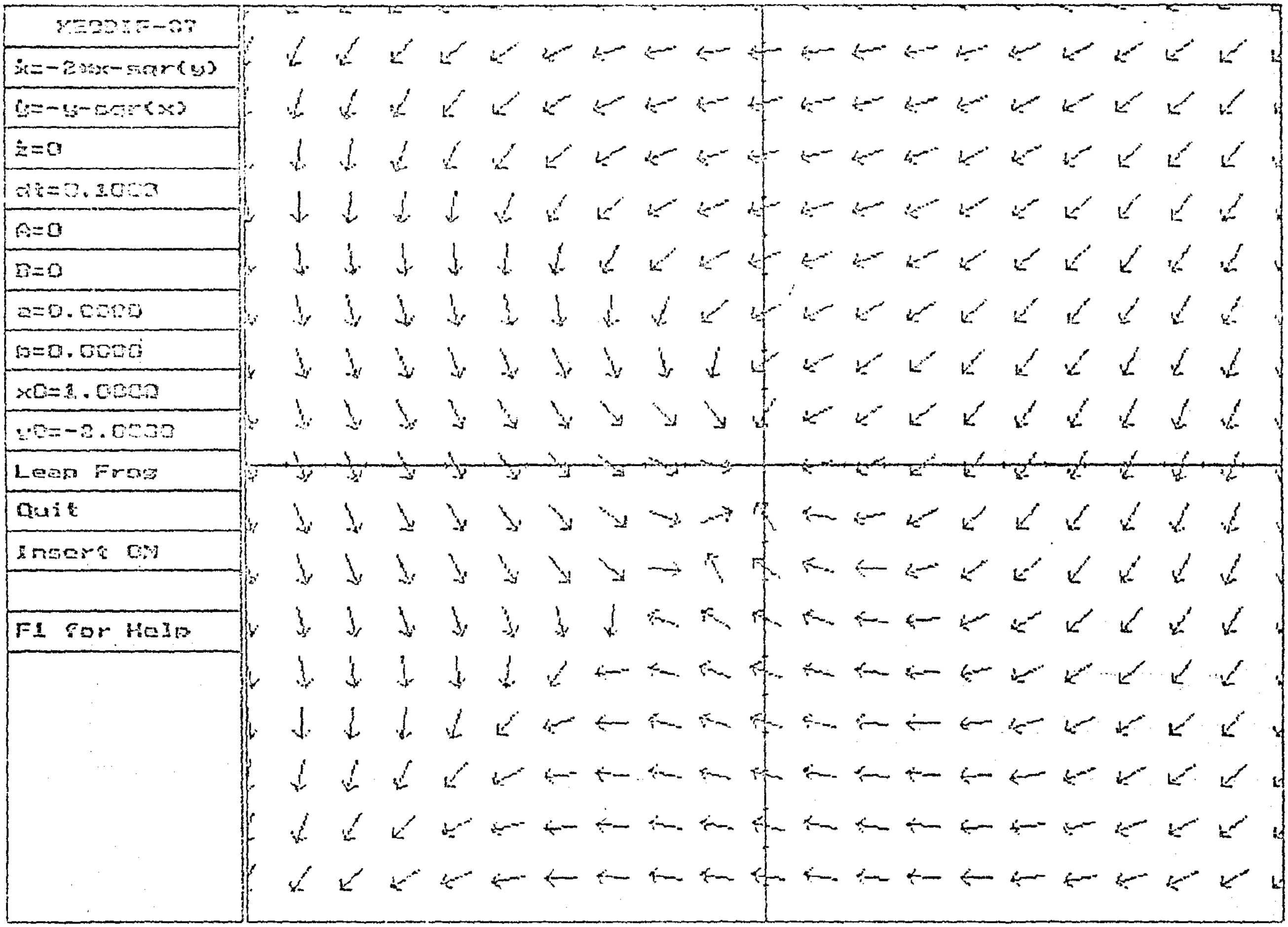
P9: $\dot{x} = y^3$
$\dot{y} = -x^3$

P10: $\dot{x} = x - xy$
$\dot{y} = -y + xy$

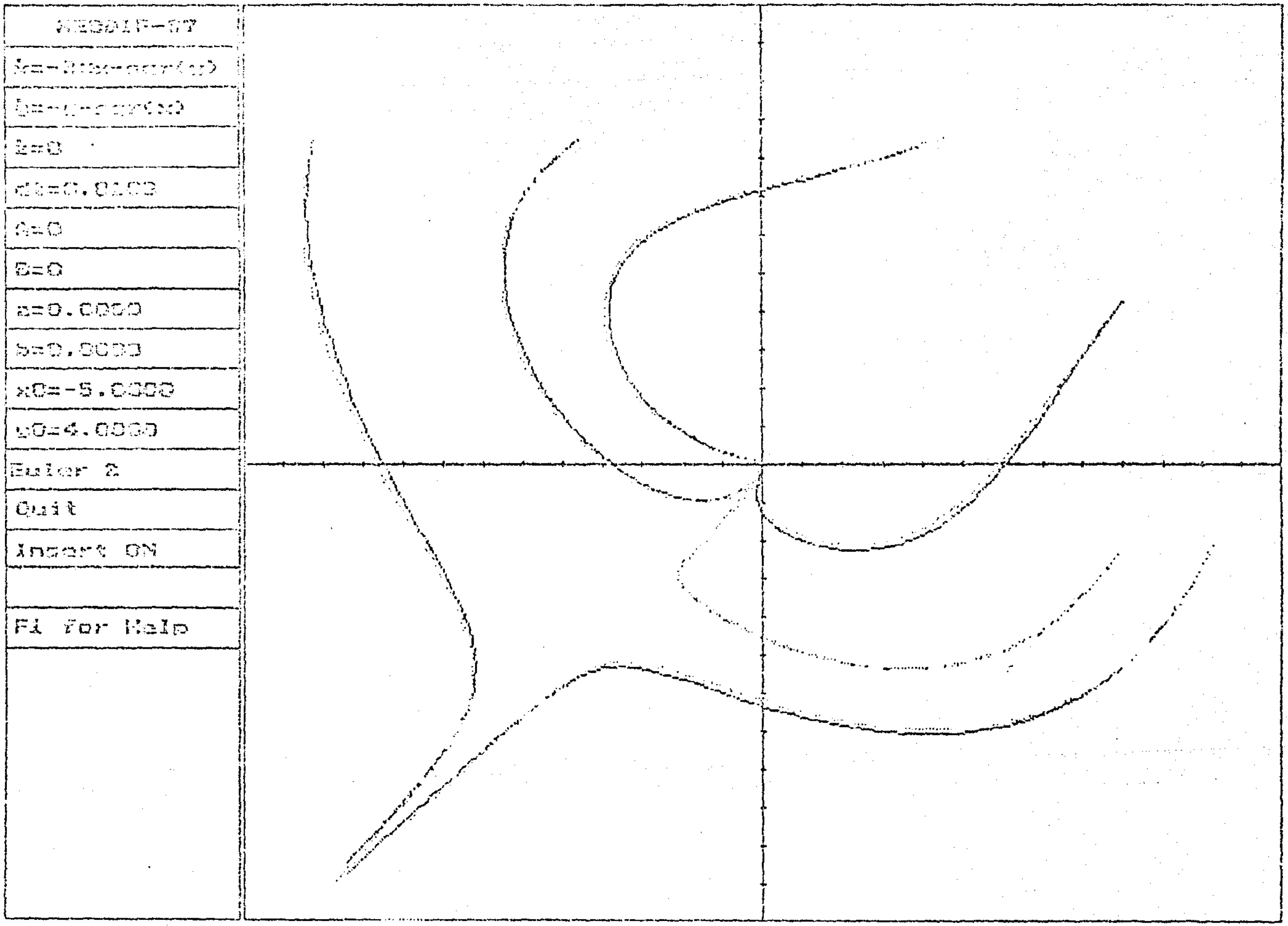
P11: $\dot{x} = y$
$\dot{y} = -(g/b)\operatorname{sen}(x)$

P12: $\dot{x} = y(1 + y)$
$\dot{y} = -\frac{1}{2}x + \frac{1}{5}y - xy + \frac{6}{5}y^2$

De las graficas de las soluciones a los diversos sistemas de ecuaciones diferenciales se concluye que el metodo de Euler 2 es casi tan bueno como el metodo de Runge-Kutta y por lo general mejor que los metodos de orden 2 usados.

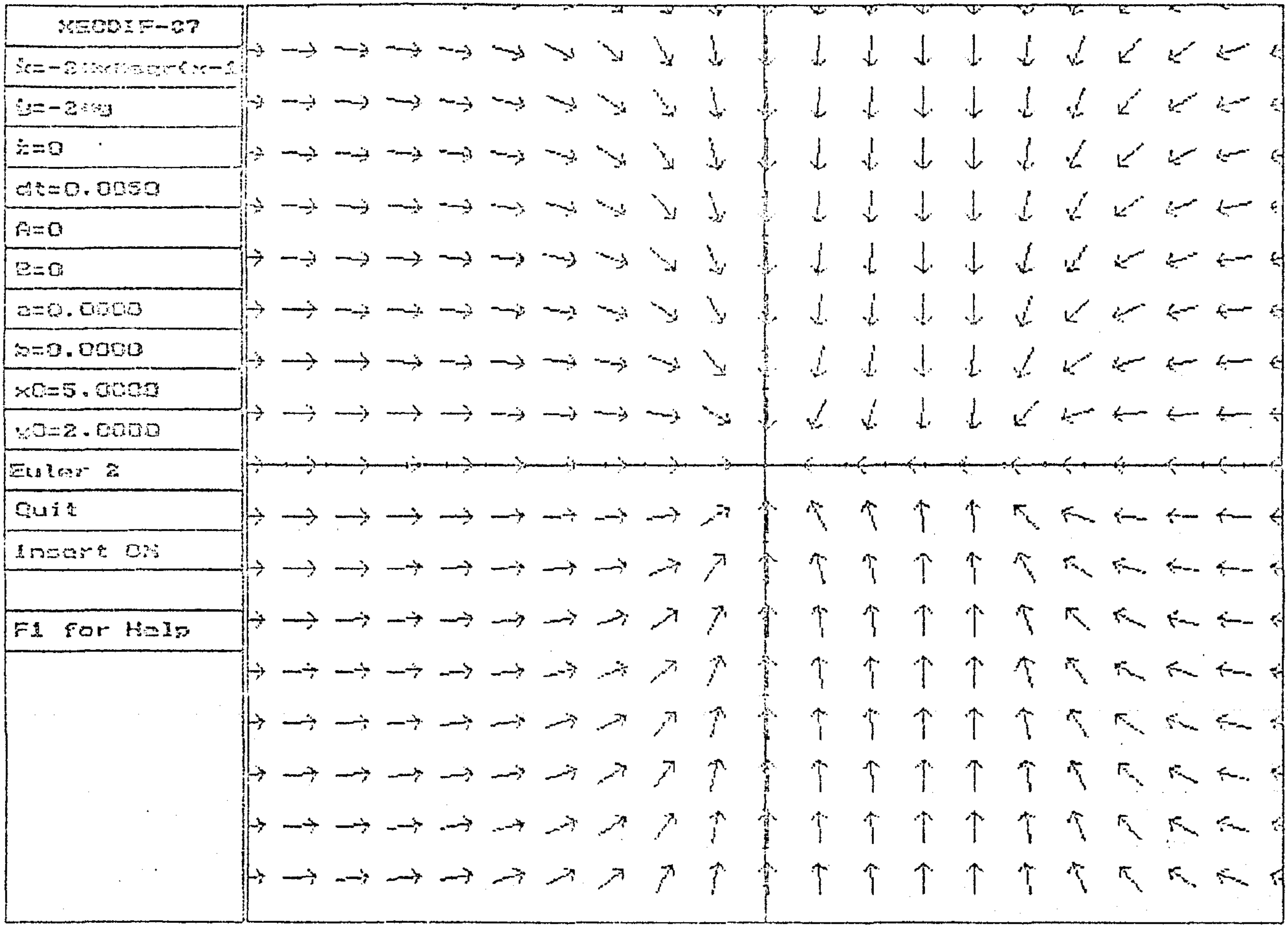

KEDDIF-87
ẋ=-2*x-sqr(y)
ẏ=-y-sqr(x)
ż=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=1.0000
y0=-2.0000
Leap Frog
Quit
Insert ON
F1 for Help

01



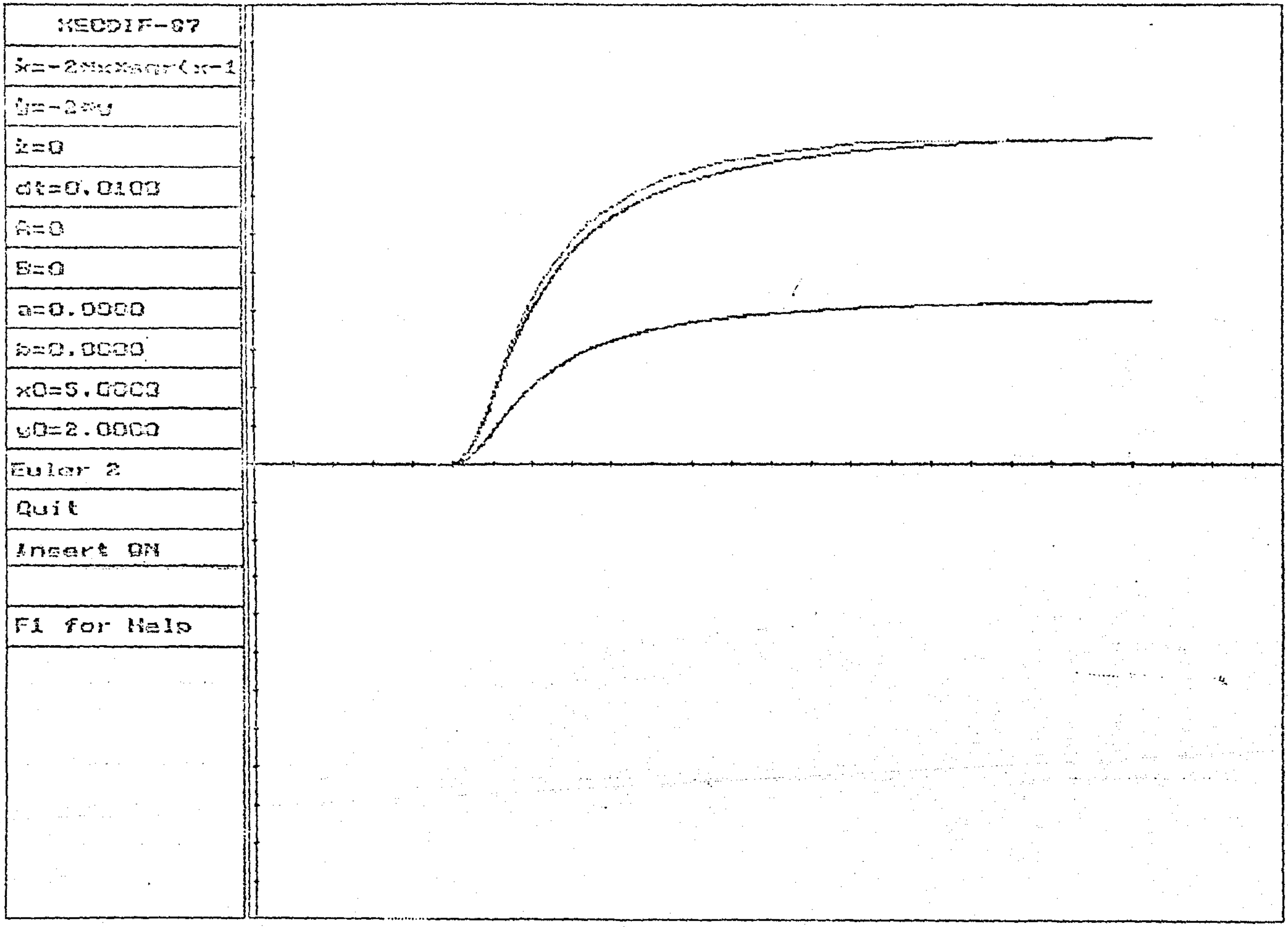
dt=0.0100
A=0
B=0
a=0.0000
b=0.0000
x0=-5.0000
y0=4.0000
Euler 2
Quit
Insert ON
F1 for Help

P2



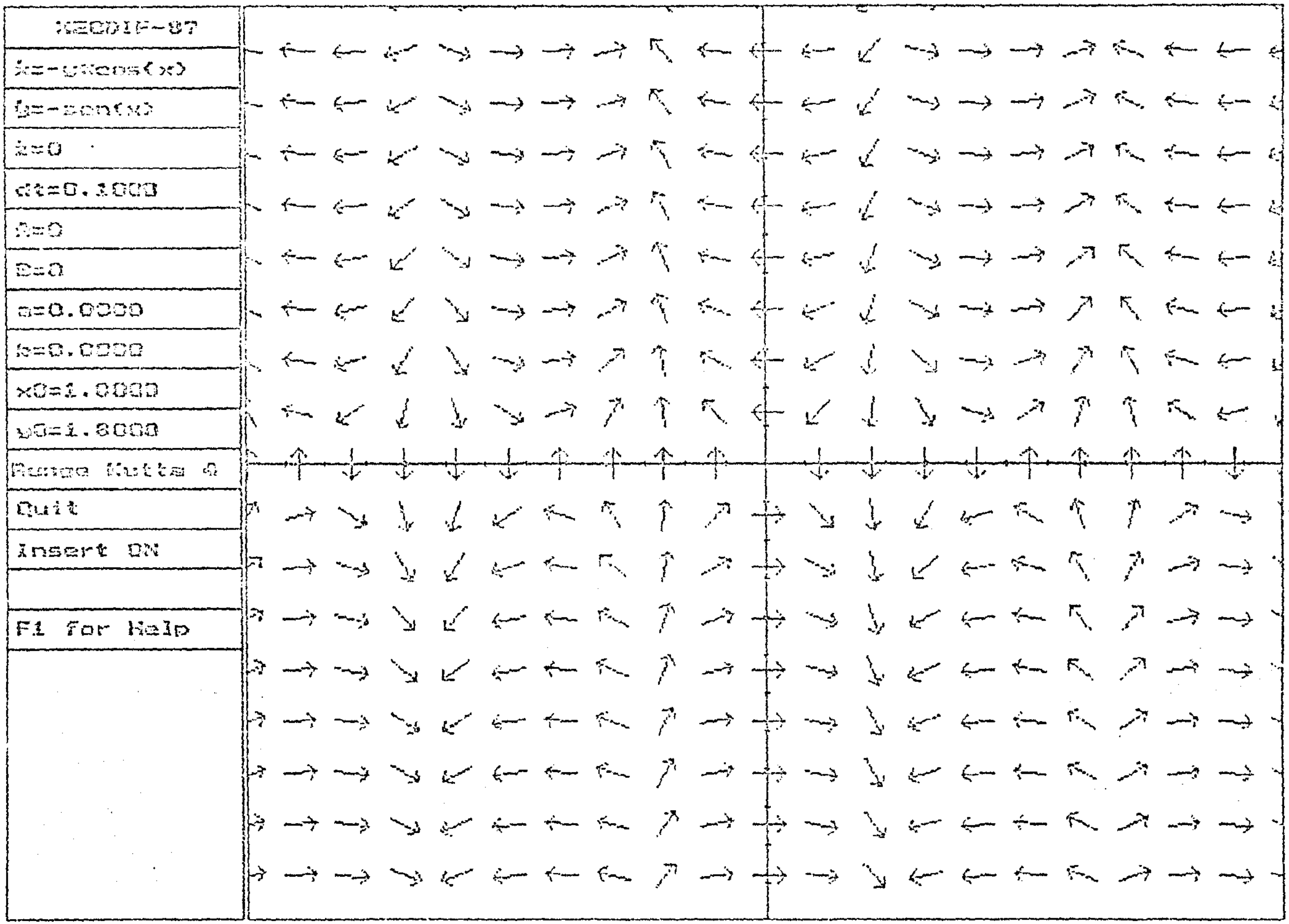
dt=0.0050
A=0
B=0
a=0.0000
b=0.0000
x0=5.0000
y0=2.0000
Euler 2
Quit
Insert ON
F1 for Help

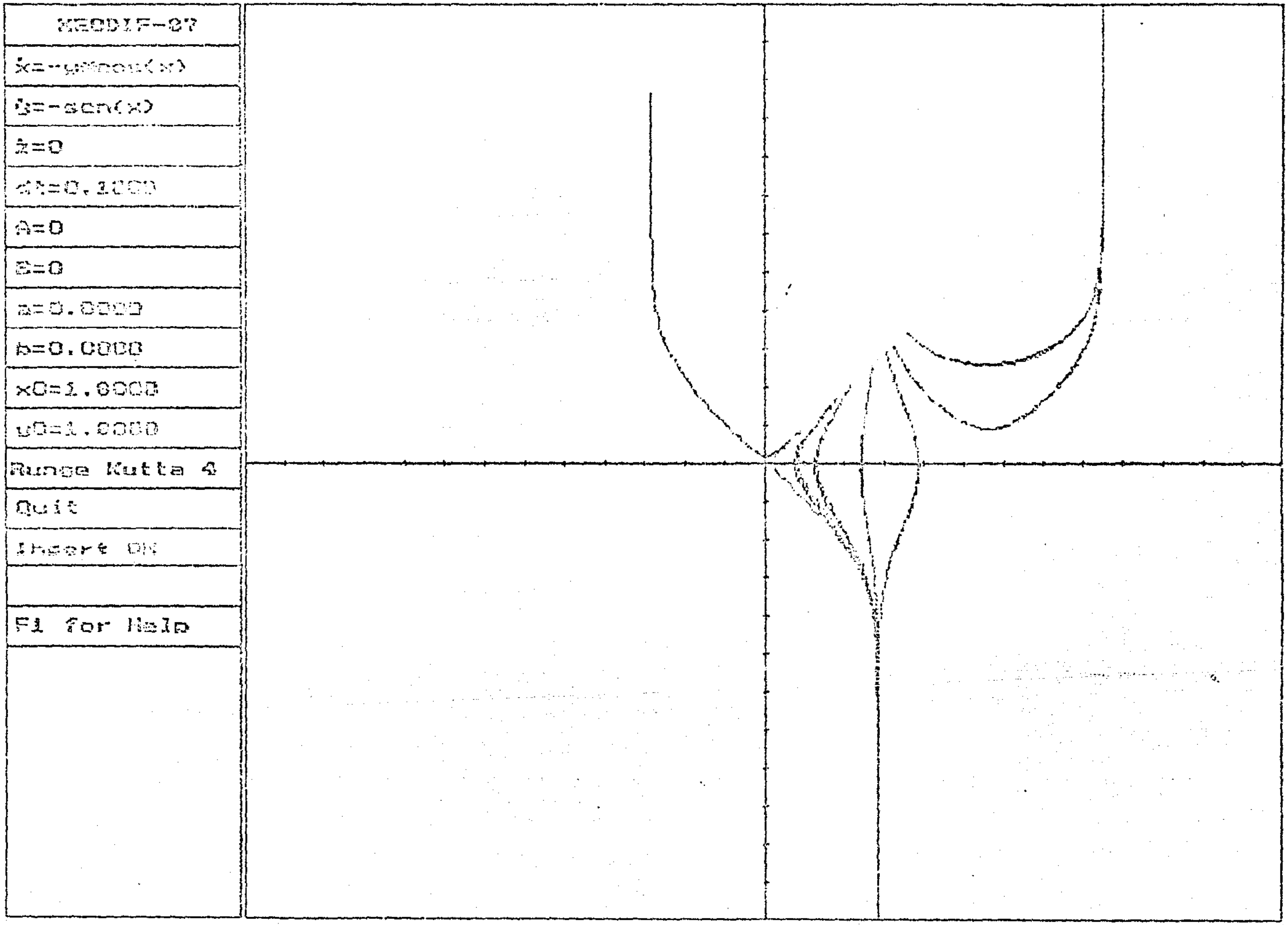
D2



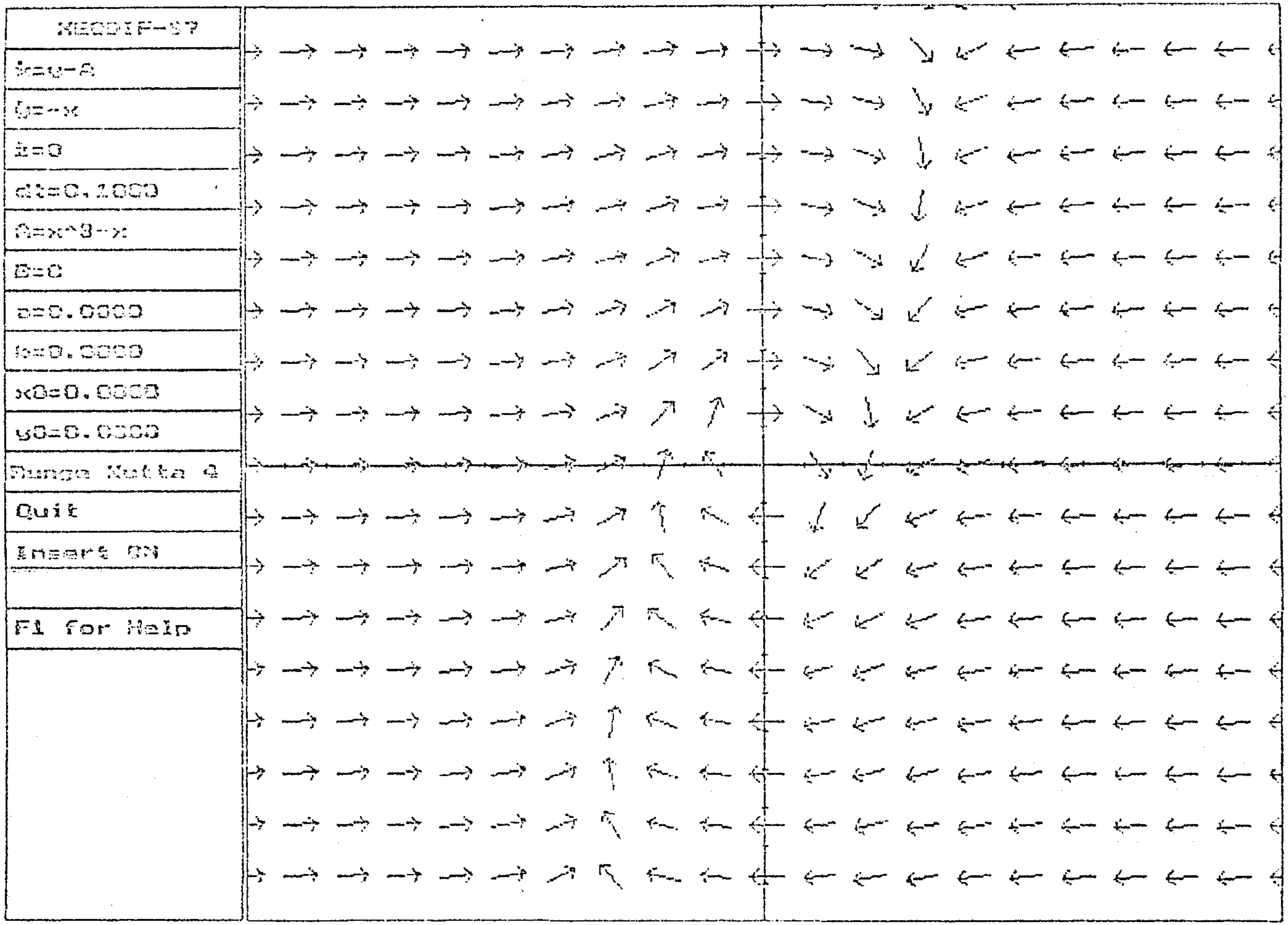
KECDIF-87
ẋ=-2*x*sqr(x-1
ẏ=-2*y
ż=0
dt=0.0100
A=0
B=0
a=0.0000
b=0.0000
x0=5.0000
y0=2.0000
Euler 2
Quit
Insert ON
F1 for Help

P3


MECDIF-87
ẋ=-y*cos(x)
ẏ=-sen(x)
h=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=1.0000
y0=1.8000
Runge Kutta 4
Quit
Insert ON
F1 for Help


ẏ=-sen(x)
t=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=1.0000
y0=1.0000
Runge Kutta 4
Quit
Insert ON
F1 for Help

P5



MEDIF-87
ẋ=y-f
ẏ=-x
t=0
dt=0.1000
f=x^3-x
g=0
a=0.0000
b=0.0000
x0=0.0000
y0=0.0000
Runge Kutta 4
Quit
Insert ON
F1 for Help

MECDIF-87

$\dot{x}=y$

$\dot{y}=-(g/b)*sen(x$

$a=0$

dt=-0.1000

A=0

B=0

a=0.0000

b=5.0000

x0=0.0000

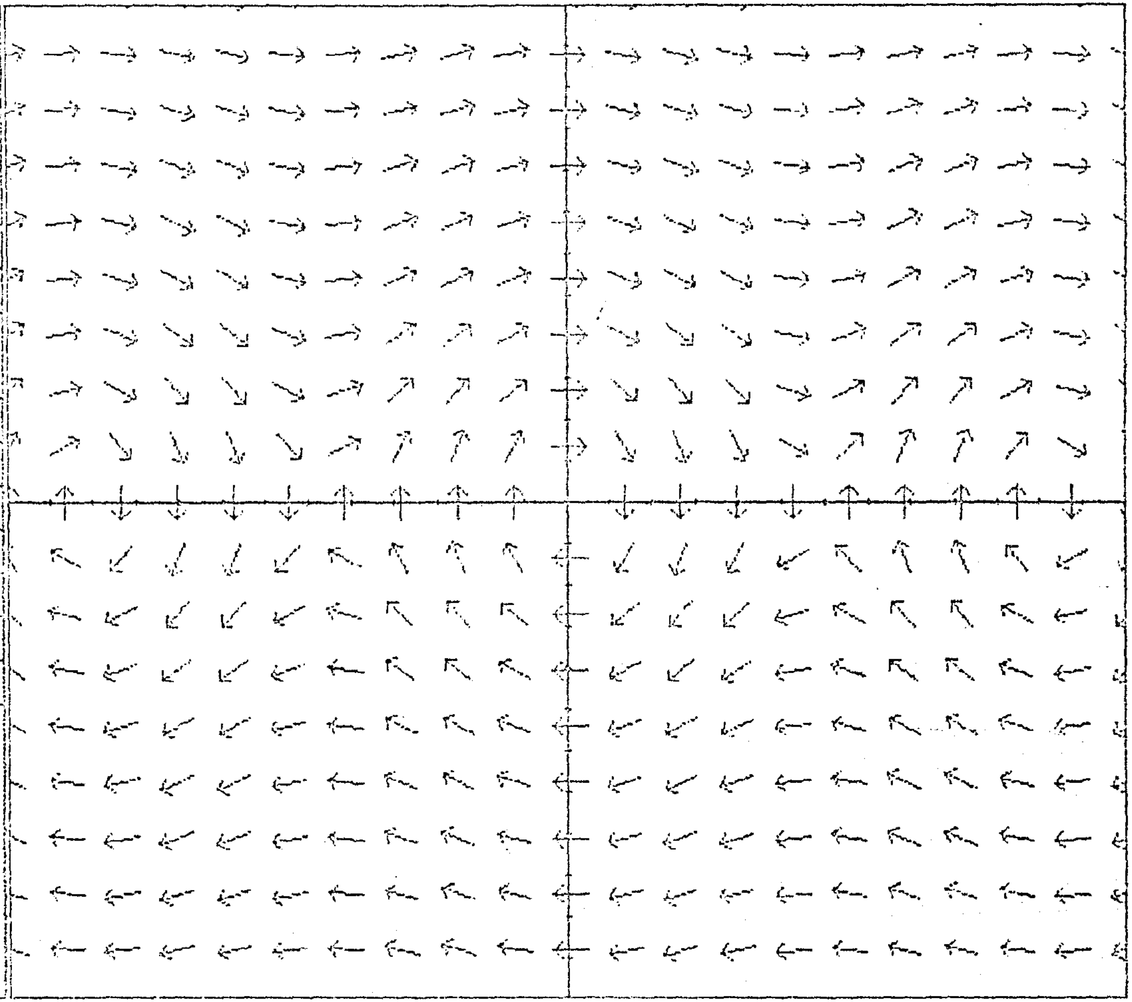
y0=4.0000

Euler 2

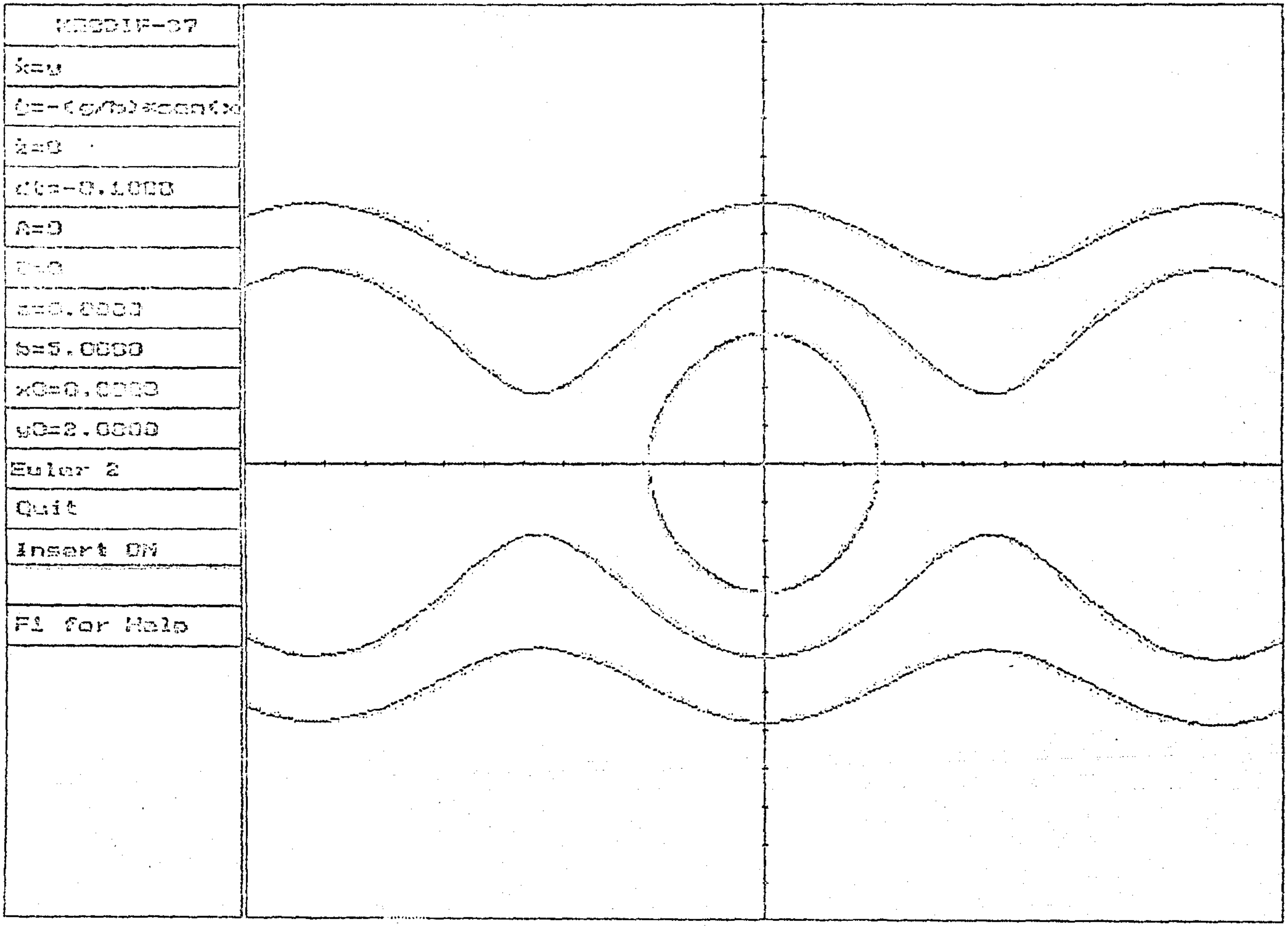
Quit

Insert ON

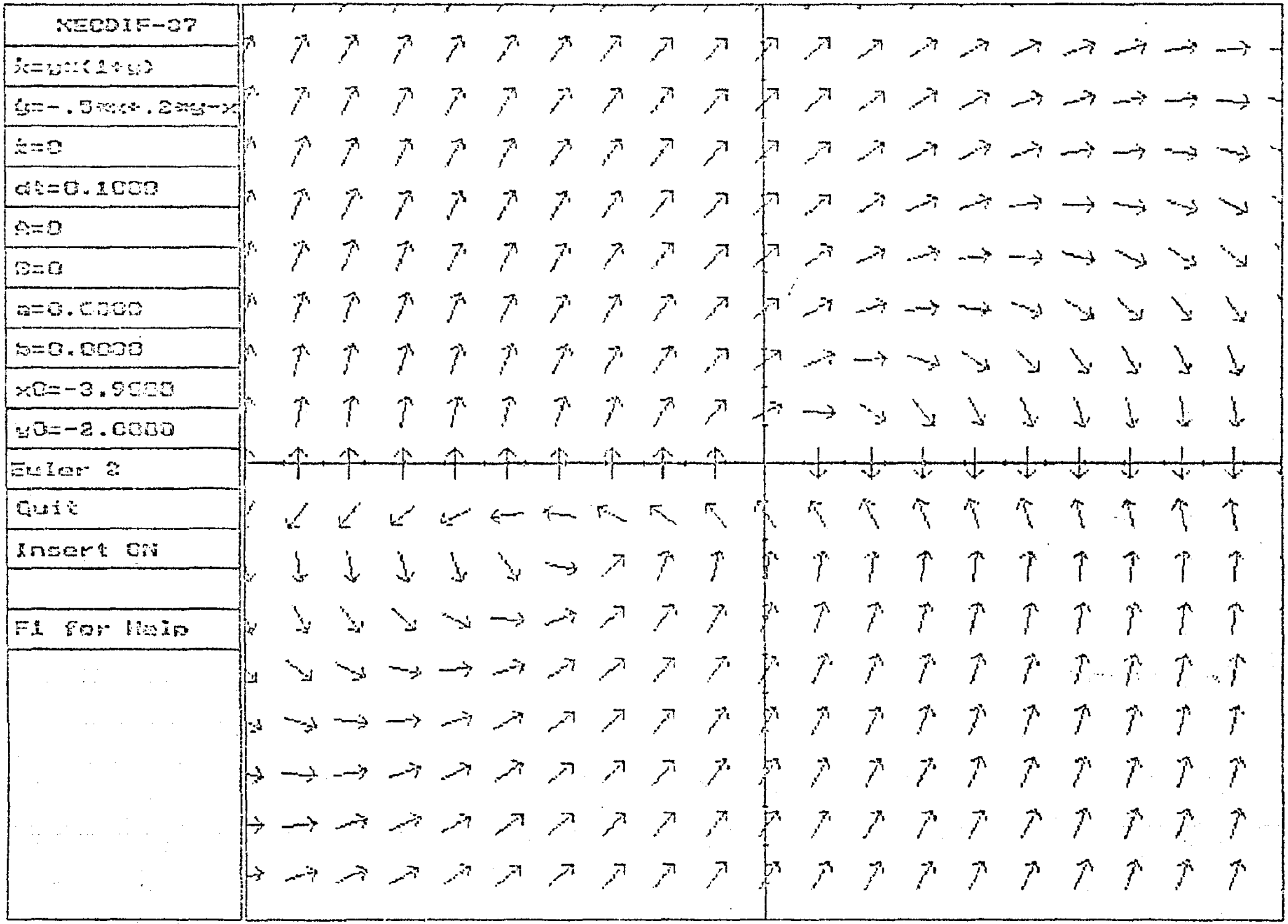
F1 for Help



p11



ẋ=y
ż=0
dt=-0.1000
A=0
a=0.0000
b=5.0000
x0=0.0000
y0=2.0000
Euler 2
Quit
Insert ON
F1 for Help


NECDIF-37
ẋ=y*(1+y)
ẏ=-.5*x+.2*y-x
t=0
dt=0.1000
A=0
B=0
a=0.0000
b=0.0000
x0=-3.9000
y0=-2.0000
Euler 2
Quit
Insert ON
F1 for Help

Seccion 22
Otros Resultados

La modificacion que se hizo al metodo de Euler se puede aplicar a la resolucion de sistemas de mas de dos variables. Por ejemplo, en el caso de sistemas de tres variables usando en el calculo de y_{n+1} los valores (x_{n+1}, y_n, z_n) y para el calculo de z_{n+1} los valores (x_{n+1}, y_{n+1}, z_n). En general, para cualquier numero de variables, utilizando en cada paso los valores mas actuales obtenidos $\bar{u}^a_n$.

Mas aun de acuerdo a como se definio un metodo de un solo paso en 1.2, este queda determinado por la funcion incremento $\Phi(t_n, \bar{u}_n; h)$, y por tanto se puede, a la manera del metodo de Euler 2, reemplazar el vector $\bar{u}_n = (x_n, y_n)$ por el vector $\bar{u}^2_n = (x_{n+1}, y_n)$ en el calculo de y_n. De manera analoga, aun para sistemas de mas de dos variables, se puede extender esta modificacion. Incluso para metodos predictores-correctores o de mas de un paso, siempre que se haga referencia a $\bar{u}_n$, se puede sustituir por el vector de valores mas actuales computados $\bar{u}^a_n$.

En este sentido se modificaron los metodos de Heun y Leap Frog observandose en la practica una mejora en la aproximacion numerica y mejores propiedades de estabilidad con respecto a soluciones periodicas.

C A P I T U L O 3

3.1. Implementacion de las Modificaciones

Parser
Clipping
Editor de Caracter
Menu de Comandos y Codigos de Control
Implementacion de Algoritmos

3.2. Manual de Manejo del Programa

SECCION 3.1
IMPLEMENTACION DE LAS MODIFICACIONES

XECDIF-87 es un programa para microcomputadoras IBM PC compatibles que ofrece al usuario un laboratorio para el estudio de sistemas de ecuaciones diferenciales. El programa fue desarrollado en el Instituto de Matematicas y en el Instituto de Investigaciones en Matematicas Aplicadas y Sistemas de la UNAM con el proposito de estudiar propiedades geometricas de las soluciones de sistemas de ecuaciones diferenciales.

Este programa recibe como entrada las funciones que describen el sistema de ecuaciones diferenciales y los valores iniciales del problema y tiene como salida la grafica de la solucion del sistema de ecuaciones, o bien la grafica del flujo del sistema de ecuaciones diferenciales.

En el programa XECDIF-87 se destacan algunos puntos desde el punto de vista computacional, a los cuales nos referiremos en este capitulo.

Parser de Funciones.- Para que el usuario pueda introducir las funciones que determinan el sistema de ecuaciones diferenciales en el momento de la corrida del programa sin tener que modificar el programa fuente se requirio escribir un parser de funciones que recibiera como entrada una cadena de caracteres conteniendo la definicion matematica de una funcion. Este la revise para verificar que no haya ninguna inconsistencia, y genera un formato evaluable por la computadora.

El formato evaluable utilizado es el conocido como notacion polaca inversa o notacion sufija. La notacion sufija, a diferencia de la notacion infija usual que pone los operadores entre los operandos, coloca primero los operandos y despues los operadores. Veamos algunos ejemplos :

Notacion Infija	Notacion Sufija
$a + b$	$a\ b\ +$
$a * (b + c)$	$a\ b\ c\ +\ *$
$a * b + c$	$a\ b\ *\ c\ +$
$a * b^c$	$a\ b\ c$ ^ $*$
$a/(b - c)$	$a\ b\ c\ -\ /$
$(a - b)/c$	$a\ b\ -\ c\ /$

La evaluacion se realiza leyendo la expresion de izquierda a derecha efectuando las operaciones en el orden de escritura en contraste a la notacion infija que evalua las expresiones de acuerdo a complicadas reglas de prioridad de operadores y uso de parentesis.

La interpretacion de la definicion de una funcion capturada desde el teclado esta formada por dos procesos : primero, la conversion de la cadena en notacion infija a la forma evaluable en notacion sufija que se hace una sola vez al momento de la captura de la funcion, y segundo, la evaluacion en si de esta forma sufija para valores especificos. Con respecto a este ultimo paso es de gran importancia que la evaluacion se haga rapidamente, pues en la resolucion de un sistema de ecuaciones diferenciales por metodos numericos se evalua al menos una vez la funcion que lo define en cada paso de la aproximacion a la solucion. Una vez escrita la primera version del evaluador de funciones de notacion sufija, se procedio a depurar el codigo con el efecto de hacerlo mas eficiente, obteniendose que se evalua una funcion en notacion sufija en el 103% - 113% del tiempo que tarda en hacerlo si la misma funcion esta escrita en el codigo fuente y compilada. Es decir, se obtuvo un sobrecosto de entre el 3% y el 13%, sobrecosto que no afecta significativamente el tiempo de resolucion de los sistemas de ecuaciones.

Ademas, en tanto que la evaluacion y los calculos son relizados desde programa se tomaron algunas medidas para evitar operaciones que puedan resultar en desborde aritmetico. Para este fin se usaron formulas de estimacion del resultado de las distintas operaciones aritmeticas posibles. Si denotamos por MAXREAL es el maximo numero real que no causa desborde y por a, b numeros menores que MAXREAL tenemos que las formulas utilizadas son :

Suma, Resta .- Si $-$ MAXREAL/2 < $a/2 \pm b/2$ < MAXREAL/2 entonces $a \pm b$ no resulta en desborde aritmetico.

Demostracion.- Claramente si :

$$- \text{MAXREAL}/2 < a/2 \pm b/2 < \text{MAXREAL}/2$$

se tiene que $a \pm b$ < MAXREAL y ademas el calculo de $a/2 \pm b/2$ no genera desborde puesto que a < MAXREAL y b < MAXREAL y por tanto $a/2 \pm b/2$ < MAXREAL.

Multiplicacion .- Si $a < 1$ o $b < 1$ o bien

$$\ln(|a|) + \ln(|b|) < \ln(\text{MAXREAL})$$

entonces $a * b$ no resulta en desborde aritmetico.

Demostracion.- En el caso de $|a| < 1$ o $|b| < 1$ tenemos

$$|a * b| < |a| < \text{MAXREAL} \qquad |a * b| < |b| < \text{MAXREAL}$$

y para el caso de $\ln(|a|) + \ln(|b|) < \ln(\text{MAXREAL})$ tenemos que $\ln(|a|) + \ln(|b|) = \ln(|a * b|)$, ahora bien, como la funcion exponencial es estrictamente creciente resulta $|a * b|$ < MAXREAL y nuevamente el calculo de $\ln(|a|) + \ln(|b|)$ no genera desborde puesto que $1 < |a|$ < MAXREAL, $1 < |b|$ < MAXREAL, $\ln(x) < x/2$ por tanto $\ln(a) + \ln(b) < a/2 + b/2$ < MAXREAL/2 + MAXREAL/2 < MAXREAL.

Division .- Si $|a|$/MAXREAL < $|b|$ entonces a/b no resulta en desborde aritmetico.

Demostracion .- Claramente de $|a|/MAXREAL < |b|$ se deduce que $|a/b| < MAXREAL$ y obviamente el calculo no genera desborde puesto que $|a|/MAXREAL < |a| < MAXREAL$.

Logaritmo\Exponencial .- Si $exp(-MAXREAL) < a \setminus a > 0$ ^ $a < ln(MAXREAL)$ entonces $ln(a) \setminus exp(a)$ no genera desborde aritmetico.

Demostracion .- Dado que logaritmo\exponencial es una funcion estrictamente creciente y ademas $ln(x) < x \setminus exp(-|x|) < 1$ se sigue de manera inmediata que no genera desborde.

Clipping.- El efecto de clipping consiste en tomar las coordenadas de un punto en la pantalla de graficacion como relativas a algun otro punto. Este efecto fue usado para tres distintos propositos en el programa :

a) el permitir que simultaneamente se despliegue el menu de opciones y la ventana de graficacion,

b) centrar las coordenadas de acuerdo a los ejes cartesianos,

c)y para permitir «perseguir» una solucion que abandone la ventana de graficacion moviendo la ventana en la direccion de la particula.

El clipping que permite desplegar tanto el menu como la ventana de graficacion es realizado por rutinas de graficacion incluidas en TURBO Pascal, en tanto que el centrado de coordenadas y la «persecucion» de soluciones son calculadas en el programa.

En este mismo sentido se aplico una escalacion a la ventana de graficacion la cual se puede variar desde el programa, es decir, la porcion del plano de graficacion que se esta graficando se puede alterar como si se viera de cerca o de lejos el plano de graficacion.

Editor de Caracter.- Con el objeto de hacer facil la captura y modificacion de las ecuaciones del sistema, de las condiciones iniciales o de parametros del problema, se incluyeron una serie de constantes predefinidas y funciones a definir por el usuario, asi como un editor de caracter que facilita cualquier consulta o modificacion de la definicion de alguna de las funciones existentes a traves de insercion o reescritura de caracteres, valor por omision de la cadena si no es introducido valor alguno, etc.

Ejemplo 3.1. Supongamos que el usuario tecleo la funcion X equivocadamente como $x - x + 3 * y$ en lugar de $x - (x + 3) * y$ entonces, el usuario puede corregir esta expresion tecleando :

i) X	para invocar la definicion de la funcion
ii) →→	para posicionarse sobre el 3er caracter
iii) Ins	para prender el modo de Insercion
iv) (	inserta el parentesis
v) →→→→	para posicionarse sobre el 9^{o} caracter
vi))	inserta el parentesis
vii) End	para posicionarse al final de la cadena
viii) Return	para confirmar la captura

Menu de Comandos y Codigos de Control.- El programa puede ser controlado en gran medida por el menu que aparece a la izquierda en la pantalla de trabajo y a traves de codigos de control, de hecho el modo menu genera codigos de control que el programa interpreta. Asi se tiene, por ejemplo, que oprimir <return> con la barra brillante en $\dot{x}$ es equivalente a dar X desde el teclado. Esto esta de acuerdo con la tendencia mas aceptada actualmente consistente en permitir el manejo de un programa por menu o por comandos de control.

Implementacion de Algoritmos.- En el momento de ejecucion se puede seleccionar el metodo numerico con el que se resolvera el sistema de ecuaciones diferenciales, y cuando se da la senal de inicio de calculo de la solucion, el programa determina cual es el metodo actual y ejecuta la rutina de solucion del sistema con el metodo respectivo. En el calculo de la solucion se determina si el incremento dt es variable (inversamente proporcional a la longitud del vector $f(\bar{u})$) o fijo, y si el sistema es de menos de tres o de mas de tres variables, puesto que en el primer caso solo evalua las funciones $\dot{x}$, $\dot{y}$, $\dot{z}$. El punto que se esta graficando se hace centellear un breve instante para mostrar en donde esta la solucion actual, sin embargo, si entre un paso y el siguiente no hay cambio en la posicion en pantalla del punto este no se grafica ni se hace centellear, por lo que una solucion con condicion inicial en un punto critico se vera como un punto estatico. Cuando el tiempo final t_f es alcanzado se hace sonar una alarma y t es reinicializada a 0.

Seccion 3.2[1]

Manual de Manejo del Programa

XECDIF-87 es una herramienta util tanto para el investigador como para el estudiante que desee estudiar cualitativamente el comportamiento de las soluciones de ecuaciones diferenciales no lineales, que suele ser complicado. La visualizacion de soluciones aproximadas ayuda a adquirir experiencia y a desarrollar una intuicion que puede guiar hacia un estudio y/o compresion profundos del tema.

El sistema de Ecuaciones Diferenciales Ordinarias que resuelve XECDIF-87 para el caso de tres o menos variables es :

$$(\dot{x}, \dot{y}, \dot{z}) = f(t, x, y, z)$$

y para el caso de mas de tres variables es :

$$(\dot{x}, \dot{y}, \dot{z}, \dot{u}, \dot{v}, \dot{w}) = f(t, x, y, z, u, v, w)$$

donde t es la variable independiente, y x, y, z, u, v, w son las variables dependientes. El valor inicial esta dado por t_i, x_0, y_0, z_0, u_0, v_0, w_0.

Para usar el program XECDIF-87, el usuario debe determinar primero el sistema de ecuaciones especifico que desea estudiar. Para esto el programa le permite definir los terminos del sistema de ecuaciones mediante el editor de caracter mencionado en 3.1. Posteriormente el usuario debe determinar los valores iniciales ti, $x0$, $y0$, $z0$, $u0$, $v0$, $w0$ tambien mediante el editor de caracter.

---

[1] Consultese: MP, MP-87, Movimiento de Particulas. Manual de operacion y aplicaciones. J.L. Abreu; M. Garza; A. Lopez-Ortiz; M. Olivero. Publicado por en Comunicaciones Tecnicas Serie Amarilla No. 72 en el Instituto de Investigaciones en Matematicas Aplicadas y Sistemas de la UNAM.

Es recomendable observar inicialmente la grafica del flujo (oprimiendo ctrl-k), y en base a esto, modificar las escalas de graficacion p y q donde reducir p a la mitad implica que el segmento del eje x observado se duplica y analogamente para q y el eje y.

El programa cuenta con tres tipos diferentes de numeros reales : constantes, parametros especiales y variables.

Las constantes son a, b, c, ..., f inicializadas con el valor 0 y e, g, π inicializadas con el numero e, la constante universal de gravitacion en sistema cgs y el numero π; los parametros especiales estan dados por p, q, dt, estos parametros afectan la ejecucion del programa; las variables de XECDIF son t, x, y, z, u, v, w.

Se cuentan con dos tipos de funciones : Funciones Auxiliares y terminos del sistema de ecuaciones diferenciales.

Las funciones auxiliares son A, B, ..., F inicializadas con el valor 0; y los terminos del sistema X, Y, Z, U, V, W.

La pantalla de observacion de XECDIF esta desplegada a la derecha del menu, y desde ella se pueden consultar y cambiar todos los numeros y funciones. El centro de la pantalla es el origen del sistema de coordenadas. Este puede moverse libremente si se oprime la tecla de Insert y oprimen las flechas del cursor. Cada movimiento desplaza los ejes 64 pixels. El desplazamiento puede reducirse a la mitad oprimiendo '-' y duplicarse oprimiendo '+'. La tecla 'Home' regresa el origen al centro de la pantalla. Hay que notar que poder usar el menu se tiene que volver a oprimir Insert.

Los comandos con que cuenta el programa son :

CTRL-A	Despliega la imagen anterior.
CTRL-B	Inicializa las variables y funciones del programa.
CTRL-C	Pone el incremento dt en terminos de la velocidad.
CTRL-H	Despliega una ayuda en la pantalla.
CTRL-I	Repone los valores iniciales del problema.
CTRL-J	Pone o quita los ejes.
CTRL-K	Dibuja el flujo del sistema de ecuaciones.
CTRL-R	Sobrepone una red en la pantalla.
CTRL-W	Limpia la pantalla.
CTRL-X	Pone el incremento dt fijo.
CTRL-Y	Cambia el color de la trayectoria.
\b	Grafica una solucion.
+	Aumenta el tamano del desplazamiento de los ejes.
-	Disminuye el tamano del desplazamiento de los ejes.
F1	Despliega una explicacion del programa.
Alt-M	Cambia el metodo de resolucion.
X, Y, Z, U, V, W	Captura de las funciones $\dot{x}$, $\dot{y}$, $\dot{z}$, $\dot{u}$, $\dot{v}$, $\dot{w}$.
x, y, z, u, v, w	Valores iniciales x_o, y_o, z_o, u_o, v_o, w_o.
Alt-d	Valor dt
p, q	Parametros de escalacion.

Ejemplo 3.2. Si se desea estudiar el sistema de ecuaciones diferenciales :

$$\dot{x} = x^2 + y^2 - 20$$
$$\dot{y} = 5(xy - 1)$$

y las soluciones con valor inicial

$x_o = 3$ $\qquad$ $y_o = -0.2, -0.1, 0, 0.1, \ldots, 0.8, 0.9.$

Entonces se oprime la tecla X, aparece el recuadro que esta ilustrado en la figura 11 en el se escribe sqr(x)+sqr(y)-20 y se

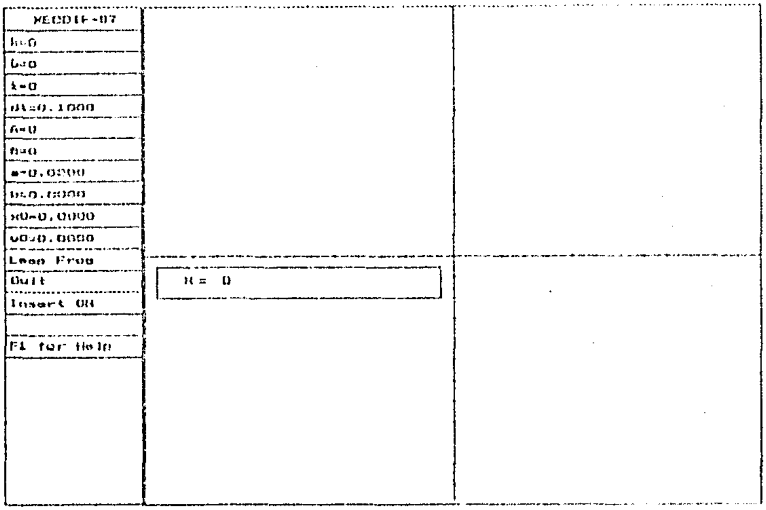


Figura 11
Captura de una funcion

termina con <return>, despues se oprime Y y aparece nuevamente un recuadro en el que se debe escribir 5*(x*y - 1) terminado con <return>, despues para ver el flujo del sistema oprimimos CTRL-K y aparece la grafica de la figura 12 de la que apreciamos que no se han visto todas las caracteristicas interesantes, por lo que escalamos para ver una porcion del plano 10 veces mayor, es decir, oprimimos p, aparece un recuadro y damos el valor 9.5, despues se oprime q, y se da el valor 8.5, damos CTRL-W para limpiar la pantalla y CTRL-K para que se grafique el flujo, aqui se aprecia que la vista es muy lejana por lo que con p = 36 y q = 34 obtenemos la grafica de la figura 13 la cual es una buena aproximacion pues aparecen graficados los cuatro puntos criticos del sistema.

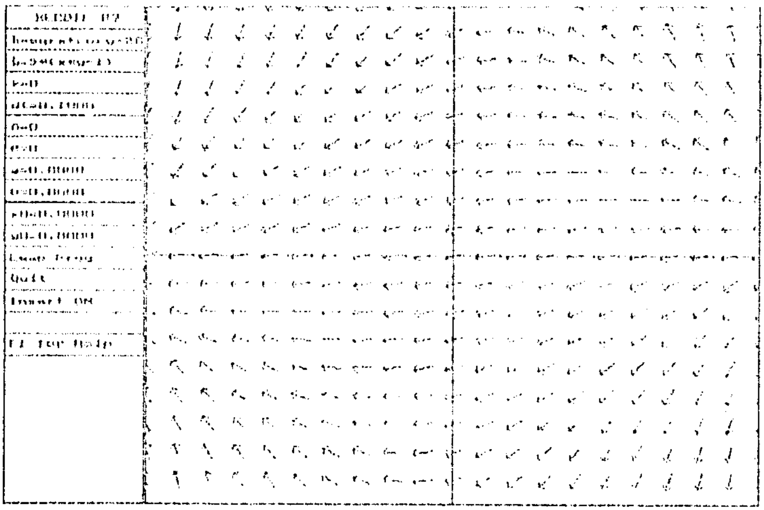


Figura 12

p = 95, q = 85

Para graficar las soluciones particulares oprimimos x y el valor inicial correspondiente 3, despues tecleamos y y el valor inicial -0.2, se oprime la barra espaciadora y se grafica la solucion, como esta no queda muy clara se cambia el incremento de tiempo con Alt-d y damos 0.01 y para tener una buena aproximacion usamos el metodo de Runge-Kutta 4 que se obtiene oprimiendo repetidas veces Alt-M hasta que aparezca en el menu de la izquierda como el metodo actual. Para estudiar toda la solucion damos Alt-d con el valor -0.01 con lo que la solucion va a ir hacia atras en el tiempo y asi sucesivamente para cada valor inicial de y, obteniendose la grafica de la figura 14.

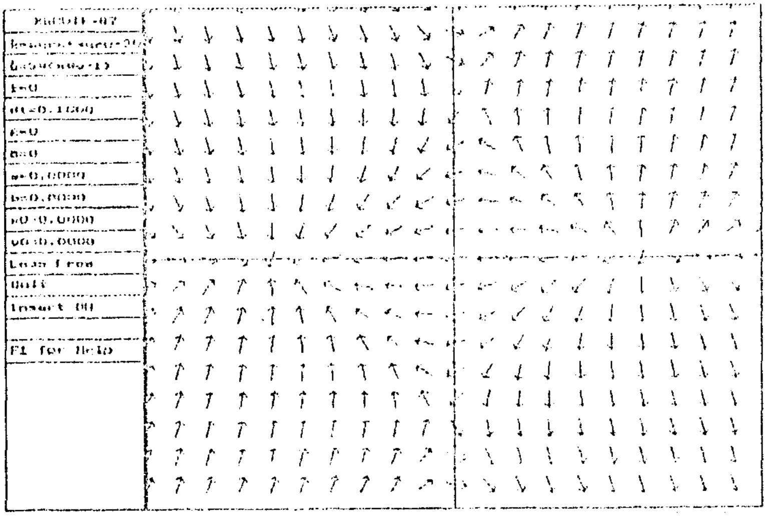


Figura 13
p = 36, q = 34

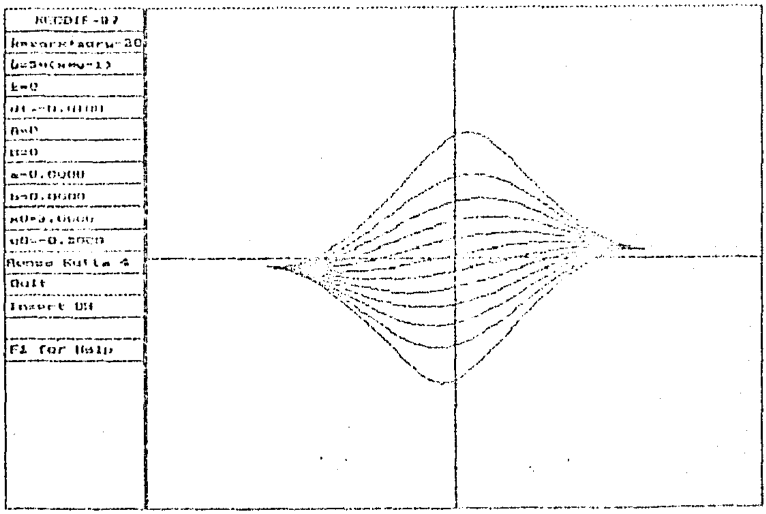


Figura 14
Soluciones del sistema de ODE's.

Finalmente para abandonar el programa oprimimos CTRL-Q. Hay que notar que el programa no efectua ningun tipo de comprobacion de esta instruccion, por lo que se debe ser cuidadoso al utilizarla.

CONCLUSIONES

El objetivo inicial de esta tesis era estudiar y analizar matematicamente las propiedades del mètodo de Euler 2. Se buscaba tener una fundamentacion teorica del uso de este metodo en el programa XECDIF-87.

En la etapa de documentacion, que suele anteceder a todo trabajo escrito, se observo que el metodo propuesto era radicalmente distinto en las ideas que lo justifican (un factor corrector) con respecto a aquellas que fundamentan los metodos usuales (aproximacion de la serie de Taylor). De ahi que aun cuando en ocasiones se utilizo un enfoque canonico en su estudio, en una gran medida este enfoque no es de gran utilidad. Mas aun, puesto que el metodo se destaca por su tratamiento de las propiedades geometricas de las soluciones y no tanto por su precision numerica, los analisis usuales del error son algo vanos.

En lo general el objetivo inicial fue alcanzado, utilizando una nueva perspectiva en el estudio de algunos metodos numericos : la preservacion de propiedades geometricas de la solucion que esta siendo aproximada. Incluso se logro fundamentar desde el punto de vista heuristico, a traves del uso de enfoques heuristicos generales, la superioridad aparente del metodo de Euler 2 con respecto a otros metodos numericos de orden superior.

Por otra parte, el programa que motivo este estudio valida y aplica los resultados de esta tesis. Este programa, mas alla del uso que aqui se le dio, queda como una herramienta para otras aplicaciones. Algunas de las ideas computacionales originadas en este programa han sido utilizadas en otras aplicaciones desarrolladas en el Instituto de Matematicas.

No resta mas que agradecer a las personas que aportaron ideas al contenido de esta tesis : Dr. Salvador Perez Esteva, M. en C. Hans Fetter Nathansky, M. en C. Maria Garza Vigil y Dr. Jose Luis Abreu Leon.