

ECUACIONES DIFERENCIALES I

M.A. RODRÍGUEZ
Departamento de Física Teórica
Universidad Complutense de Madrid

13 de octubre de 1999

ÍNDICE

| | |
|---|-----------|
| Prólogo | v |
| 1 Introducción | 1 |
| 1.1 Primeras definiciones | 1 |
| 1.2 Problemas a resolver | 3 |
| 1.2.1 Existencia y unicidad | 4 |
| 1.2.2 Dependencia continua | 4 |
| 1.2.3 Prolongación | 5 |
| 1.2.4 Solución general | 7 |
| 1.2.5 Ecuaciones en forma implícita | 7 |
| 1.3 Interpretación geométrica. Isoclinas | 7 |
| 1.4 Métodos elementales de integración | 10 |
| 1.4.1 Ecuaciones exactas | 10 |
| 1.4.2 Factores integrantes | 11 |
| 1.4.3 Ecuaciones de variables separadas | 12 |
| 1.4.4 Cambios de variable | 13 |
| 1.4.5 La ecuación lineal de primer orden | 15 |
| 1.5 Métodos aproximados en ecuaciones de primer orden | 17 |
| 1.5.1 Poligonal de Euler | 17 |
| 1.5.2 Método de Runge-Kutta | 18 |
| 1.6 Existencia y unicidad | 19 |
| 1.7 Estabilidad | 22 |
| 1.8 Ecuaciones autónomas | 23 |
| 2 Sistemas y ecuaciones lineales | 27 |
| 2.1 Sistemas lineales | 27 |
| 2.2 El espacio de soluciones de la ecuación lineal | 28 |
| 2.3 Sistemas lineales con coeficientes constantes | 30 |
| 2.3.1 Cálculo de exponenciales de matrices | 31 |
| 2.4 Solución general | 36 |
| 2.5 Ecuaciones lineales de orden n | 39 |
| 2.6 Ecuaciones con coeficientes constantes | 42 |
| 2.7 Coeficientes indeterminados | 45 |
| 2.8 Estabilidad de ecuaciones lineales | 47 |
| 2.9 El criterio de Routh-Hurwitz | 49 |
| 3 Sistemas dinámicos | 53 |
| 3.1 Introducción | 53 |
| 3.2 Definiciones y resultados elementales | 53 |
| 3.3 Sistemas lineales | 56 |
| 3.4 Estabilidad de sistemas autónomos | 60 |
| 3.5 Clasificación de puntos críticos de sistemas lineales | 61 |
| 3.6 Puntos críticos no elementales | 69 |

| | | |
|----------|--|-----------|
| 3.7 | Sistemas autónomos no lineales planos | 70 |
| 3.8 | Simetrías y campos de vectores | 73 |
| 3.9 | Integrales primeras | 78 |
| 3.10 | Sistemas hamiltonianos | 85 |
| 4 | Soluciones en forma de serie | 89 |
| 4.1 | Introducción | 89 |
| 4.2 | Ecuaciones diferenciales lineales de segundo orden | 90 |
| 4.3 | Puntos singulares | 97 |
| 4.4 | Singularidades en el infinito | 106 |
| 4.5 | La ecuación hipergeométrica | 107 |
| 4.6 | La ecuación hipergeométrica confluyente | 109 |
| 4.7 | El átomo de hidrógeno | 110 |

Lista de Figuras

| | |
|------|----|
| 1.1 | 2 |
| 1.2 | 4 |
| 1.3 | 5 |
| 1.4 | 6 |
| 1.5 | 6 |
| 1.6 | 7 |
| 1.7 | 8 |
| 1.8 | 9 |
| 1.9 | 9 |
| 1.10 | 24 |
| 1.11 | 25 |
| 1.12 | 26 |
| 1.13 | 26 |
| 3.1 | 57 |
| 3.2 | 57 |
| 3.3 | 58 |
| 3.4 | 59 |
| 3.5 | 60 |
| 3.6 | 61 |
| 3.7 | 62 |
| 3.8 | 63 |
| 3.9 | 64 |
| 3.10 | 65 |
| 3.11 | 66 |
| 3.12 | 67 |
| 3.13 | 68 |
| 3.14 | 69 |
| 3.15 | 70 |
| 3.16 | 71 |
| 3.17 | 72 |
| 3.18 | 73 |
| 3.19 | 74 |
| 3.20 | 75 |
| 3.21 | 76 |
| 3.22 | 77 |
| 3.23 | 78 |
| 3.24 | 79 |
| 3.25 | 80 |
| 3.26 | 81 |
| 3.27 | 83 |
| 3.28 | 84 |
| 3.29 | 85 |

| | | |
|------|-------|----|
| 3.30 | | 86 |
| 3.31 | | 88 |
| 3.32 | | 88 |

Prólogo

Las Ecuaciones Diferenciales (ED) es uno de los campos de la Matemática donde se siente con más fuerza la influencia mutua con la Física. Muchas de las leyes de la Física adoptan en su enunciado matemático la forma de una ecuación diferencial. Los grandes desarrollos en el mundo de las ED son motivados en su mayor parte por la necesidad de estudiar el significado de esas leyes para poder predecir comportamientos de los sistemas físicos.

Desde muy pronto en la historia de la matemática moderna, es decir, los tiempos de Euler, las ecuaciones diferenciales aparecen como una de las ramas con un crecimiento espectacular aunque existan discontinuidades claras. Desde los objetivos iniciales de Euler y los Bernouilli por resolver de manera explícita el mayor número de ecuaciones posibles, hasta los supuestamente más modestos de obtener el mayor número de resultados posibles aún sin saber la solución, ha pasado un largo periodo en el que el desarrollo de las otras ramas de la matemática, singularmente el análisis, han permitido conseguir un nivel de conocimientos realmente impresionante en este tema.

En estas notas sobre ecuaciones diferenciales ordinarias se pretende dar una visión adaptada al curso de Ecuaciones Diferenciales I de la licenciatura de Ciencias Físicas. Representan una adaptación de las notas sobre Ecuaciones Diferenciales Ordinarias, elaboradas para el antiguo curso de Ecuaciones Diferenciales (Métodos II). Se han suprimido varios capítulos y se ha resumido el contenido de otros para poderlas utilizar de manera apropiada en este segundo año de licenciatura.

Tema 1

Introducción

1.1 Primeras definiciones

Sin entrar en mayores precisiones, una ecuación diferencial es una relación entre las derivadas de una función y sus variables, es decir:

$$f(x, u, u_x, u_{xx}, \dots) = 0$$

donde tanto x como u pueden ser vectores de espacios de dimensión finita, al menos en lo que concierne a estas notas.

En la primera parte de estas notas desarrollaremos el estudio de las llamadas ecuaciones diferenciales ordinarias (EDO) en las que x es una variable escalar, aunque u puede ser un vector. El orden de una ecuación es el de la derivada que lo tiene máximo. Así por ejemplo, la ecuación:

$$y' = y$$

es una ecuación de primer orden en la que la función incógnita es y , la variable independiente es x , que no aparece en la ecuación, y donde:

$$y' = \frac{dy}{dx}$$

El teorema fundamental del cálculo permite obtener la solución de ecuaciones diferenciales que se encuentran reducidas a cuadraturas, es decir aquellas como:

$$y' = f(x)$$

en las que todas las soluciones son simplemente:

$$y(x) = \int f + C$$

donde el signo integral se refiere a la integral indefinida de f y C es una constante arbitraria. Sin embargo, en el ejemplo anterior, $y' = y$, la solución no es tan trivial. De hecho, se puede interpretar esa ecuación, junto con una condición adecuada, como la definición de la función exponencial. En efecto, si se supone que se busca una solución que verifique $y(0) = 1$, se encuentra como única solución:

$$y(x) = e^x.$$

Todas las soluciones de esta ecuación tienen la forma:

$$y(x) = Ce^x \tag{1.1}$$

donde C es una constante cualquiera (figura 1.1). En efecto, sea $y(x)$ una solución de la ecuación, y multipliquémosla por $\exp(-x)$. Derivando respecto a x , es fácil obtener:

$$\frac{d}{dx}(y(x)e^{-x}) = 0$$

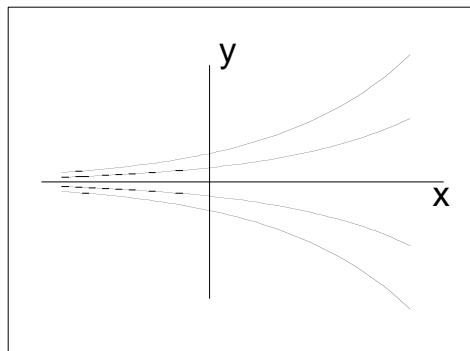


Figura 1.1

de donde se deduce la forma dada en (1.1).

Para finalizar este ejemplo, demostremos como las propiedades de la función exponencial pueden deducirse del estudio de la ecuación diferencial de la que es solución; consideremos por ejemplo:

$$e^{a+b} = e^a e^b.$$

Si $y(x)$ es solución de esta ecuación, es obvio, debido a la independencia en t que $y(x+t)$ también lo es. Así pues, como todas las soluciones son de la forma dada en (1.1), tendremos:

$$y(x+t) = C e^x$$

donde C es una constante, determinada evidentemente por t . Igualando:

$$y(x+t) = e^{x+t} = C(t) e^x$$

de donde se obtiene el resultado anunciado anteriormente sin más que tomar $x = 0$.

Otra forma de ver las cosas es la siguiente. Supongamos que el objetivo es determinar todas las funciones que verifican la propiedad:

$$f(a+b) = f(a)f(b)$$

es decir, que representan el grupo aditivo de los reales en el grupo multiplicativo. Esta idea permite tratar las funciones especiales de la Física Matemática desde el punto de vista de la teoría de la representación de los grupos. Para calcular f procedamos de la siguiente forma: supongamos que $a = t$ y derivemos con respecto a esta variable:

$$f'(t+b) = f'(t)f(b)$$

si ahora sustituimos t por 0 y tratamos b como una variable x , tendremos:

$$f'(x) = f'(0)f(x)$$

llamando $y = f(x)$: $y' = Ky$ donde $K = f'(0)$. La ecuación obtenida es la tratada anteriormente salvo una constante K .

Se pueden generar ED de la siguiente forma: sea

$$f(x, y, c_1, c_2, \dots, c_n) = 0$$

una ecuación, en la que se supone que las operaciones a realizar son válidas. Derivemos n veces respecto a la variable x , considerando a y como función implícita de x . Esto permite obtener $n+1$ ecuaciones de

donde es posible, en principio, eliminar las constantes c_i , obteniéndose una ecuación diferencial de grado n . El orden de esta ecuación puede ser menor si las constantes satisfacen algunas relaciones. La solución de esta ED será la ecuación de partida, siempre que y pueda ser escrita como función de x , y dependerá de n constantes de integración. Esto es una solución general en el sentido habitual. Sin embargo es posible que alguna solución pueda no estar representada en ella. Consideremos el siguiente ejemplo para una ecuación con una constante: $f(x, y, c) = 0$. Si derivamos con respecto a x obtendremos:

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} y' = 0$$

donde hemos sustituido c por la expresión obtenida de $f(x, y, c) = 0$. Si derivamos esta expresión considerando a x, y, c como variables independientes tendremos:

$$\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial c} dc = 0$$

(también con c eliminada), y de ambas ecuaciones llegamos a:

$$\frac{\partial f}{\partial c} dc = 0$$

es decir, o bien $dc = 0$ con lo cual c es constante y obtenemos la solución general $f(x, y, c) = 0$, o bien:

$$\frac{\partial f}{\partial c} = 0$$

que puede en ocasiones ser una solución de la ED y no estar incluida en la solución general. Se trata de una solución singular, y desde un punto de vista geométrico es la envolvente de la familia de curvas dada por la solución general.

La mayor parte de los resultados sobre ecuaciones diferenciales se refieren a aquellas que están escritas en la llamada forma normal, es decir aquellas en las que la derivada de mayor orden está despejada:

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

Si y es una función vectorial se dice que tenemos un sistema de ecuaciones. Existe una relación entre los sistemas de primer orden y las ecuaciones de orden n . En efecto, una ecuación diferencial de orden n , dada en forma normal siempre se puede escribir como un sistema de ecuaciones de primer orden sin más que definir las derivadas sucesivas de y como nuevas funciones incógnitas: $y = y_1, y' = y_2$ (así que $y'_1 = y_2$), etc. con lo que la propia ecuación se transforma en:

$$y'_n = f(x, y_1, y_2, \dots, y_n).$$

El proceso inverso, es decir, pasar de un sistema a una ecuación, no siempre es tan evidente y resulta necesario examinar cuidadosamente el sistema. Por ejemplo:

$$\begin{cases} y'_1 = y_1 \\ y'_2 = y_2 \end{cases}$$

es un sistema de primer orden que no es equivalente a una ecuación de orden dos. Incluso aunque se pueda encontrar esa ecuación, cuestiones como la derivabilidad de las soluciones deben ser tratadas separadamente.

1.2 Problemas a resolver

Los problemas que surgen en el estudio de las ecuaciones diferenciales son muy variados. Algunos serán tratados en los próximos ejemplos.

1.2.1 Existencia y unicidad

La existencia y unicidad de soluciones de una ecuación serán las primeras dificultades a discutir. Como ya hemos visto, una ED tiene en general un número infinito de soluciones. Sin embargo si sometemos dichas soluciones a alguna condición de tipo a especificar, podemos elegir entre ellas una sola. Las restricciones que aparecen sobre la ecuación y la condición para que esto sea cierto se estudiarán más adelante. Los ejemplos que siguen muestran varias ED con diferentes comportamientos:

Ejemplo 1.2.1 $y' = y$, $y(0) = 1$ tiene una única solución, como ya hemos visto, $y(x) = e^x$

Ejemplo 1.2.2 $y' = 3y^{2/3}$ tiene más de una solución cuando $y(0) = 0$. Por ejemplo: $y(x) = 0$, $y(x) = x^3$ verifican tanto la ecuación como la condición establecida. La solución general de esta ecuación es:

$$y(x) = (x + c)^3$$

aunque $y(x) = 0$ no aparece en ella. Esta solución es la envolvente de la familia de curvas dada en la solución general (figura 1.2).

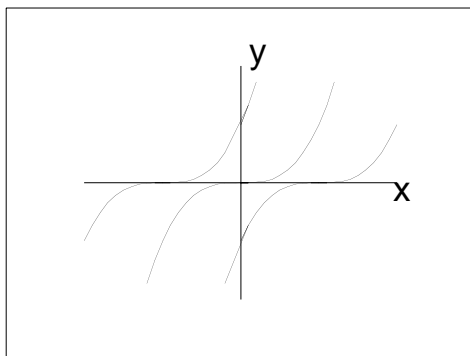


Figura 1.2

Ejemplo 1.2.3 $y' = -x/y$, $y(0) = 0$, es un problema que no tiene solución. La solución general viene dada por:

$$x^2 + y^2 = c$$

donde c es una constante positiva. Ninguna solución pasa por $(0, 0)$. Nótese que esta expresión no es una solución en sentido estricto pues y no es función de x en todos los puntos. Sin embargo contiene a las soluciones de la ecuación. Con más propiedad la podemos llamar una integral de la ecuación (figura 1.3).

1.2.2 Dependencia continua

Puesto que muchos fenómenos físicos pueden representarse por modelos que encuentran su formulación matemática en una ED, resulta de suma importancia saber como dependen sus soluciones de los parámetros y de los datos. En general éstos no se conocerán con precisión absoluta y pequeñas variaciones podrían provocar grandes cambios en las soluciones. La dependencia respecto a parámetros y datos, y la estabilidad de las soluciones son temas de importancia fundamental. Los siguientes ejemplos muestran como ecuaciones sencillas presentan distintos comportamientos.

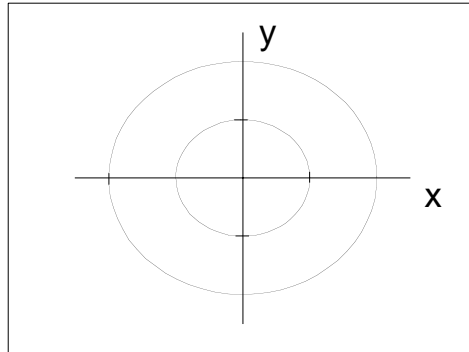


Figura 1.3

Ejemplo 1.2.4 $y' = ay$ donde a es una constante real. Las soluciones de esta ecuación dependen continuamente de los datos, para toda constante a . Es decir cuando se consideran soluciones con dato inicial $y(x_0) = y_0$ y se varía y_0 la solución depende de y_0 de forma continua en un intervalo $[x_0, x_1]$ con x_1 finito. La demostración es muy sencilla.

Ejemplo 1.2.5 En la misma ecuación del ejemplo anterior uno puede estudiar cómo las soluciones dependen continuamente respecto al parámetro a . El comportamiento cualitativo de las soluciones es muy similar según varía a . Se puede considerar el punto $a = 0$ como un punto de bifurcación. Para $a > 0$ todas las soluciones tienden a ∞ cuando $x \rightarrow \infty$. Sin embargo cuando $a < 0$ todas tienden a 0 cuando $x \rightarrow a$. Cuando $a = 0$ las soluciones son rectas paralelas al eje x .

Ejemplo 1.2.6 Siguiendo con el ejemplo anterior, la solución $y(x) = 0$ es una solución estable asintóticamente cuando $x \rightarrow \infty$, si $a < 0$. Si $a = 0$ esta solución sigue siendo estable pero no asintóticamente. Pero cuando $a > 0$ la solución es inestable. Aún sin entrar en mayores precisiones sobre estos conceptos, resulta claro del dibujo de las soluciones el significado de estas palabras (figura 1.4).

1.2.3 Prolongación

Al ser una solución una función $y = y(x)$, puede ocurrir que no esté bien definida para cualquier valor de x . En general uno considera soluciones definidas en el mayor intervalo posible. Se llaman soluciones maximales en el sentido que el intervalo no se puede ampliar más. Este intervalo puede ser, por supuesto, no acotado por la derecha, izquierda o por ambas simultáneamente. Los siguientes ejemplos muestran soluciones que no se pueden prolongar hasta infinito, es decir, el intervalo máximo de definición no llega hasta ∞ .

Ejemplo 1.2.7 $y' = -x/y$. Como hemos visto antes, las soluciones de esta ecuación están contenidas en las curvas de nivel de la función:

$$x^2 + y^2 = c.$$

Por tanto, una solución será:

$$y(x) = \sqrt{c - x^2}.$$

Está claro que esta solución no se puede prolongar hacia la derecha (x crecientes) más allá de $c^{1/2}$. Un razonamiento similar muestra que no se puede prolongar hacia la izquierda más allá de $-c^{1/2}$. Por

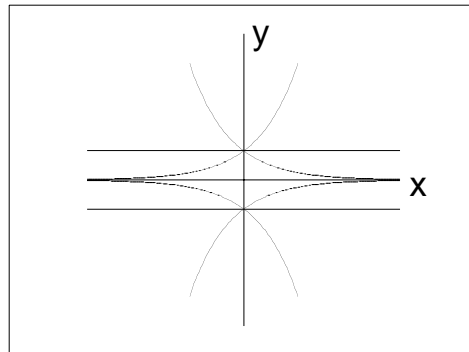


Figura 1.4

tanto esta solución está definida en el intervalo $(-c^{1/2}, c^{1/2})$ y es no prolongable en el sentido dicho anteriormente (figura 1.3).

Ejemplo 1.2.8 $y' = y^2$ tiene como solución general $y(x) = 1/(c - x)$. Para cualquier c la solución no se puede prolongar más allá de c .

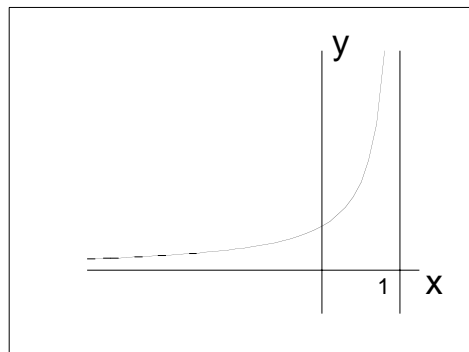


Figura 1.5

Por ejemplo, para fijar ideas, consideremos una solución particular que verifica $y(0) = 1$, es decir, $y(x) = 1/(1 - x)$. Está claro que esta solución no puede prolongarse más allá de $x = 1$, donde tiene una asíntota vertical. Por la izquierda no presenta ninguna dificultad así que el intervalo máximo de definición es $(-\infty, 1)$ (figura 1.5). Se dice muchas veces que la solución no es prolongable hacia la derecha y sí lo es hacia la izquierda, para indicar que la solución maximal tiene como intervalo de definición el anteriormente citado. Esta misma ecuación tiene soluciones con intervalo de definición del tipo (c, ∞) , e incluso una solución definida en toda la recta, $y(x) = 0$.

Las dificultades de extensión del intervalo, son distintas en los dos ejemplos expuestos. En el primero de ellos la derivada se hace infinito en un punto mientras que en el segundo existe una asíntota vertical pero la derivada no se hace infinito en ningún punto en el que la función esté definida.

1.2.4 Solución general

Se ha hablado de soluciones generales como aquellas que contienen todas las soluciones de la ecuación, que se obtienen al dar valores a una constante que aparece en ellas. Sin embargo es posible que alguna solución escape de la solución general y no se obtenga para ningún valor real de la constante (o constantes en otros casos). Por ejemplo, en el caso anterior, $y(x) = 0$ no aparece en la solución general para ningún valor de la constante c .

1.2.5 Ecuaciones en forma implícita

Finalmente, existe un tipo de ecuaciones diferenciales no resueltas en la derivada de orden superior. En general equivalen a una ecuación cuando se puede despejar esa derivada, por ejemplo cuando se puede aplicar el teorema de la función implícita

Ejemplo 1.2.9 $(y')^2 = 4y$ es un caso sencillo de este tipo de ecuaciones que equivale a dos:

$$y' = 2y^{1/2}; \quad y' = -2y^{1/2}.$$

Las soluciones de ambas ecuaciones son el haz de parábolas $y = (x + c)^2$. La solución $y(x) = 0$ es la envolvente del haz y las otras soluciones presentan un comportamiento peculiar en el eje x , donde no existe solución única (figura 1.6).

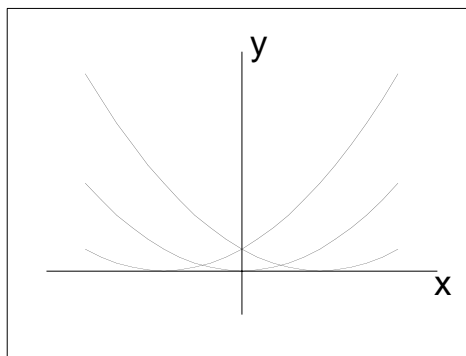


Figura 1.6

1.3 Interpretación geométrica. Isoclinas

En esta sección consideraremos tan solo ecuaciones de primer orden resueltas en la derivada. Aunque similares afirmaciones podrían hacerse para sistemas de ecuaciones de primer orden, solo para las ecuaciones pueden hacerse representaciones gráficas sencillas. Así, sea $y' = f(x, y)$ una ED de primer orden. A cada punto del plano x, y podemos asociarle un segmento de pendiente $f(x, y)$. De esta manera obtendremos un campo de direcciones en la región en la cuál esté definida f . Se denominan isoclinas a las curvas que

unen los puntos en los que la pendiente es constante. De acuerdo con la ED $y' = f(x, y)$, una solución de esta ecuación es tangente en cada punto al campo de direcciones. Es posible obtener de esta forma una gráfica aproximada de las soluciones, mediante el dibujo de las isoclinas de la ecuación, curvas de nivel de la función f . Es posible que en algún punto la función f tenga un valor infinito. En este caso no puede hablarse con propiedad de solución. Sin embargo, con respecto al dibujo aproximado de las curvas integrales, este hecho puede obviarse mediante la consideración de la ecuación asociada $x' = 1/f(x, y)$, donde $x' = dx/dy$, es decir considerando a x como función de y . De esta forma, cuando aparezca un punto de pendiente infinita para la ecuación inicial, se puede considerar esta otra ecuación y tratar a la curva integral como una solución suya. No siempre es sencillo dibujar las isoclinas y por tanto las soluciones.

Los siguientes ejemplos muestran cuales son los pasos más adecuados para llevarlo a cabo, aunque cada situación puede necesitar un estudio particular.

Ejemplo 1.3.1 $y' = y - x$ es una ecuación lineal. Las isoclinas son muy sencillas en este caso: $y - x = m$ son rectas de pendiente 1 y ordenada en el origen m , es decir la pendiente de las soluciones que pasan por ellas. El dibujo de isoclinas y soluciones es el que aparece en la figura 1.7.

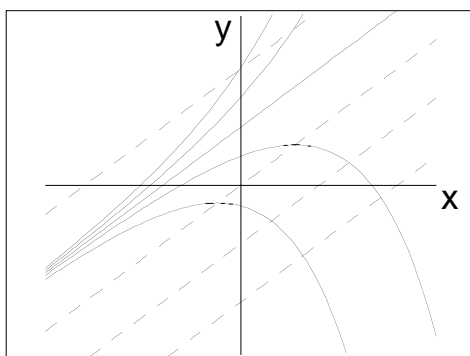


Figura 1.7

Ejemplo 1.3.2 $xy' + (1 - x)y = 0$. Las isoclinas son ahora $(x - 1)y/x = m$, o si despejamos y :

$$y = \frac{mx}{x - 1}$$

curvas que pasan por $(0,0)$ y tienen una asíntota vertical en el punto $x = 1$. La pendiente en el origen no está definida. De la expresión obtenida en primer lugar, uno puede deducir que la recta $x = 1$ es una isoclina de pendiente = 0. Asimismo $x = 0$ es una isoclina de la ecuación asociada (dx/dy) con pendiente igual a 0 (o si se quiere de la ecuación inicial con pendiente ∞). El dibujo de las isoclinas y soluciones puede verse en la figura 1.8.

Otras características de las soluciones pueden obtenerse de la propia ecuación. Por ejemplo, el conjunto de puntos de inflexión está contenido en el conjunto de puntos que anulan a la segunda derivada y puede obtenerse de la ecuación sin más que derivar (suponiendo que se pueda):

$$y'' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} f(x, y) = 0.$$

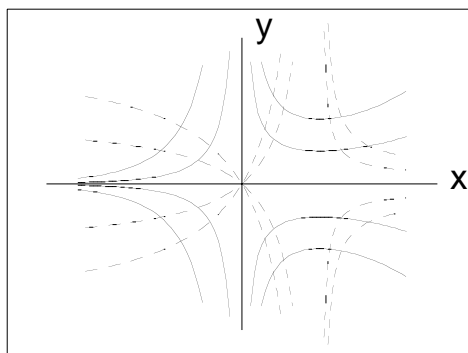


Figura 1.8

Ejemplo 1.3.3 $y' = y^2 - x$. Las isoclinas de esta ecuación son las curvas $y^2 - x = m$, es decir parábolas con eje horizontal. La derivada segunda de una solución es:

$$y'' = 2y'y - 1 = 2y(y^2 - x) - 1$$

y por tanto, la ecuación de los posibles puntos de inflexión es la expresión anterior igualada a cero es decir:

$$x = y^2 - \frac{1}{2y}.$$

En la figura 1.9 pueden verse las isoclinas, curva de puntos de inflexión y soluciones.

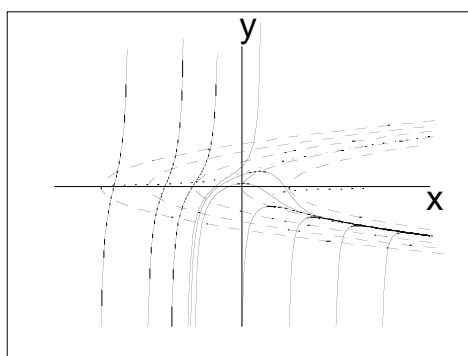


Figura 1.9

Con esto terminamos esta introducción al estudio de las ecuaciones diferenciales. En la siguiente sección estudiaremos diversos métodos de resolución de ED de primer orden.

1.4 Métodos elementales de integración

Aún cuando no cabe esperar resolver todas las ED de forma completa es posible en ciertos casos obtener la solución general de una forma explícita. A continuación estudiaremos algunos casos de ecuaciones diferenciales de primer orden para los que se puede encontrar la solución general.

1.4.1 Ecuaciones exactas

Consideremos una ED escrita en la forma siguiente:

$$P(x, y)dx + Q(x, y)dy = 0. \quad (1.2)$$

Desde el punto de vista de la teoría de formas diferenciales, si ω es una 1-forma en \mathbb{R}^2 (o en un abierto de \mathbb{R}^2), su expresión en coordenadas es:

$$\omega = P(x, y)dx + Q(x, y)dy$$

por lo tanto la ecuación (1.2) es $\omega = 0$. Cuando ω sea una 1-forma exacta, existirá una función $f(x, y)$ tal que

$$\omega = df$$

y por lo tanto, al ser $\omega = 0$, $f(x, y) = c$ es la solución general de la ecuación. Una condición necesaria para que ω sea exacta es que sea cerrada, es decir: $d\omega = 0$. En coordenadas, esta restricción se traduce en:

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0 \quad (1.3)$$

suponiendo que las funciones $P(x, y)$, $Q(x, y)$ son diferenciables. Esta condición es también suficiente en el caso en que el abierto de \mathbb{R}^2 en el que trabajamos sea, por ejemplo, simplemente conexo, o haciendo consideraciones de tipo local, como haremos de hecho.

De acuerdo con esta interpretación, llamaremos ecuación diferencial exacta a aquella que escrita en la forma (1.2) verifique la condición (1.3). Las ecuaciones exactas son inmediatamente resolubles, pues basta encontrar una función $f(x, y)$ cuyas derivadas con respecto a x e y coincidan respectivamente con $P(x, y)$ y $Q(x, y)$. Veamos como se puede encontrar $f(x, y)$.

De la primera condición:

$$\begin{aligned} \frac{\partial f}{\partial x} &= P(x, y) \\ f(x, y) &= \int_{x_0}^x P(s, y)ds + h(y) \end{aligned}$$

si ahora derivamos con respecto a y :

$$\frac{\partial f}{\partial y} = \int_{x_0}^x \frac{\partial P}{\partial y}(s, y)ds + h'(y) = Q(x, y) - Q(x_0, y) + h'(y) = Q(x, y)$$

de donde:

$$h'(y) = Q(x_0, y)$$

ecuación que se integra fácilmente dando:

$$f(x, y) = \int_{x_0}^x P(s, y)ds - \int_{y_0}^y Q(x_0, s)ds.$$

Con la elección de extremos de la integral que se ha hecho, $f(x, y) = 0$ es la solución que pasa por (x_0, y_0) . En la práctica se consideran las integrales indefinidas y la solución general viene dada por $f(x, y) = c$ en la cuál solo aparece una constante como era de esperar.

Ejemplo 1.4.1

$$y - x^3 + (x + y^3)y' = 0.$$

La ecuación es exacta como se comprueba fácilmente:

$$P(x, y) = y - x^3, \quad Q(x, y) = x + y^3$$

luego:

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} = 1.$$

Sea $f(x, y)$ la función cuyas curvas de nivel contienen a las soluciones de la ecuación:

$$\frac{\partial f}{\partial x} = P(x, y)$$

$$f(x, y) = yx - \frac{1}{4}x^4 + h(y)$$

$$\frac{\partial f}{\partial y} = x + h'(y) = x + y^3$$

por tanto:

$$h(y) = \frac{y^4}{4}, \quad f(x, y) = yx - \frac{x^4 - y^4}{4}$$

y las curvas integrales son $f(x, y) = c$. Nótese que $Q(x, y) = 0$ cuando $x + y^3 = 0$ por lo que en estos puntos no es obvio que y pueda definirse como función de x (al menos no se verifican las hipótesis del teorema de la función implícita). En los demás puntos esa ecuación define a y como función de x , y es una solución de la ecuación.

1.4.2 Factores integrantes

No toda ecuación es exacta, pero podría pensarse que la multiplicación por un factor global $\mu(x, y)$ haría que la nueva ecuación, equivalente a la anterior al menos si existe el inverso de μ , lo fuera. Es decir aunque ω no sea exacta, una elección adecuada de μ hace que $\mu\omega$ lo sea. Para ello debería verificarse la condición de exactitud dada anteriormente:

$$\frac{\partial}{\partial y} \mu P = \frac{\partial}{\partial x} \mu Q$$

lo que equivale a una ecuación en derivadas parciales de primer orden para la función μ . La obtención de la solución de esta ecuación diferencial es equivalente en dificultad a la resolución de la ecuación de partida (el uso del método de características lleva a la ecuación original) pero no es necesario conocerla. Basta hallar una función que verifique esa ecuación, es decir, basta una solución particular. Se dice que $\mu(x, y)$ es un factor integrante para esa ecuación.

Ejemplo 1.4.2 $y' = (y - 2x)/(2y + x)$. Escribiendo esta ecuación como hemos hecho antes:

$$(2x - y)dx + (2y + x)dy = 0$$

que no es una ecuación exacta. Tratemos de buscar un factor integrante para ella. La ecuación que debe verificar $\mu(x, y)$ es:

$$(2y + x)\mu_x - (2x - y)\mu_y = -2\mu.$$

Esta ecuación en derivadas parciales de primer orden lleva a la ecuación de partida. Sin embargo, puede probarse fácilmente que la función:

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

es una solución particular. Aunque el método aquí descrito no sea muy general, no deja de tener interés. Hagamos el siguiente cambio de variables:

$$\begin{aligned}x &= \rho \cos \theta \\y &= \rho \operatorname{sen} \theta\end{aligned}$$

la nueva ecuación es:

$$\rho\mu_\rho + 2\mu_\theta = -2\mu$$

y suponiendo que μ no depende de θ , $\rho\mu_\rho = -2\mu$, ecuación de la que se obtiene como solución particular $\mu(\rho) = 1/\rho^2$, es decir el factor integrante antes dado. Si se supone que μ solo depende de θ , se obtiene la ecuación: $\mu_\theta = -\mu$, con solución:

$$\mu = e^{-\theta}$$

es decir: $\mu = e^{-\arctan(y/x)}$ que es otro factor integrante para esta ecuación. Escojamos el primer factor integrante. La ecuación es ahora:

$$\frac{2x-y}{x^2+y^2}dx + \frac{2y+x}{x^2+y^2}dy = 0$$

que es exacta y tiene como solución general:

$$\log(x^2 + y^2) - \arctan \frac{x}{y} = C.$$

Esta función no define en todos los puntos a y como función de x . De acuerdo con el teorema de la función implícita (que da condiciones suficientes) cuando $2y+x=0$, no se puede asegurar que exista esa función implícita. En esos puntos es donde $f(x, y) = (y-2x)/(2y+x)$ es discontinua.

En el ejemplo siguiente veremos como una hipótesis adicional sobre μ , a saber, que dependa tan solo de la variable x , permite obtener fácilmente una ecuación exacta.

Ejemplo 1.4.3 $(x^2 + y^2 + 2x + 1)dx + y(x + 1)dy = 0$

Esta ecuación no es exacta, $P(x, y) = x^2 + y^2 + 2x + 1$, $Q(x, y) = y(x + 1)$ así que:

$$\frac{\partial P}{\partial y} = 2y; \quad \frac{\partial Q}{\partial x} = y.$$

Si suponemos que existe un factor integrante que solo es función de x , $\mu(x)$:

$$(x+1)\mu' = \mu$$

ecuación que tiene por solución $\mu(x) = x + 1$. De esta forma uno obtiene fácilmente la solución general

$$(x+1)^2(2y^2 + (x+1)^2) = C.$$

El uso de factores integrantes no es, aunque pueda parecerlo de los ejemplos precedentes, un método muy efectivo. De hecho, toda ecuación que tiene una única solución (para una condición inicial dada) admite una familia infinita de factores integrantes. Es obvio que no sabemos calcular la solución de todas las ED.

1.4.3 Ecuaciones de variables separadas

Se dice que una ecuación diferencial es de variables separadas si se puede escribir en la forma:

$$y' = \frac{f(x)}{g(y)}.$$

Se trata de una ecuación exacta, escribiéndola en la forma (1.2):

$$f(x)dx - g(y)dy = 0$$

y por tanto se pueden usar los métodos expuestos anteriormente. Sin embargo, en la práctica resulta más directo prescindir de estos planteamientos y escribir:

$$\int g(y)y'(x)dx = \int f(x)dx + C$$

y cambiando la variable de integración en la primera integral:

$$\int g(y)dy = \int f(x)dx + C$$

relación que una vez integrada, permitirá encontrar la solución.

Ejemplo 1.4.4 $y' = -(1+y^2)/(1+x^2)$ es una ecuación de variables separadas, con $f(x) = -1/(1+x^2)$, $g(y) = 1/(1+y^2)$, por lo que su solución se encuentra inmediatamente como:

$$-\int \frac{dx}{1+x^2} = \int \frac{dy}{1+y^2}$$

es decir: $-\arctan(x) = \arctan(y) + C$, ecuación de la que se puede despejar $y(x)$ obteniéndose la solución general:

$$y(x) = \frac{k-x}{1+kx}.$$

Además, $y(x) = 1/x$ también es solución, aunque no se obtenga para ningún valor de $k \in \mathbb{R}$ ($k = \infty$ permite obtener esa solución).

1.4.4 Cambios de variable

Uno de los métodos más interesantes para la resolución de ED es el de cambios de variable, tanto independiente como dependiente. Sin embargo, no siempre es sencillo adivinar cual es el mejor cambio que simplifica la ecuación y la hace resoluble por algún método elemental. Veremos a continuación algunos tipos de ecuaciones y los cambios de variable que ayudan a su solución.

Ecuaciones homogéneas

Supongamos que la función $f(x, y)$ es homogénea de grado 0, es decir:

$$f(tx, ty) = f(x, y).$$

En este caso, si se escribe $y = zx$, donde z es la nueva incógnita, la nueva ecuación es:

$$z'x + z = f(x, zx) = f(1, z)$$

que es una ecuación de variables separadas:

$$z' = \frac{f(1, z) - z}{x}$$

y que puede resolverse con el método dado anteriormente.

Ejemplo 1.4.5 $y - 2x - (2y + x)y' = 0$.

Esta ecuación fue resuelta antes buscando un factor integrante que la hiciera exacta. Sin embargo es más sencillo calcular su solución considerándola como una ecuación homogénea. Si $y = zx$:

$$xz' = -2\frac{z^2 + 1}{2z + 1}$$

que es una ecuación de variables separadas:

$$\int \frac{2z + 1}{z^2 + 1} dz = -2 \int \frac{dx}{x}$$

y por tanto:

$$\log(x^2 + y^2) + \arctan \frac{y}{x} = C$$

que es la misma solución anterior (aunque el significado de la constante C sea distinto).

Ecuación de Bernoulli

El siguiente tipo de ecuaciones se puede resolver mediante un cambio en la variable dependiente:

$$y' = f(x)y + g(x)y^r$$

donde r es un número real. Si se pone $z = y^{1-r}$, la nueva ecuación es:

$$z' = (1-r)(f(x)z + g(x))$$

que es una ecuación lineal de primer orden sobre la que hablaremos a continuación.

Ejemplo 1.4.6 $y' = y - 4y^2 \cos^2(x/2)$. Poniendo $z = 1/y$, obtenemos una ecuación lineal de primer orden:

$$z' = -z + 4 \cos^2 \frac{x}{2}$$

que será resuelta en la próxima sección.

La ecuación de Riccati

Se trata de una ecuación con términos no lineales cuadráticos:

$$y' = a(x) + b(x)y + c(x)y^2.$$

En general no se puede reducir a cuadraturas, y solo cuando uno conoce una solución particular puede convertirla mediante un cambio de variable en una ecuación de Bernoulli para desde allí pasar a una ecuación lineal de primer orden. La ecuación de Riccati es un caso interesantísimo que puede ser estudiado desde numerosos puntos de vista. Veamos primero como el conocimiento de una solución particular simplifica la ecuación. Sea $y_1(x)$ esa solución y llamemos:

$$u(x) = y(x) - y_1(x).$$

Sustituyendo en la ecuación hacemos desaparecer el término que no depende de y :

$$u' = (b(x) + 2y_1(x)c(x))u + c(x)u^2$$

que es una ecuación de Bernoulli, convirtiéndose en una ecuación lineal con el cambio $u = 1/z$.

La ecuación de Riccati aparece relacionada con una ecuación lineal de orden 2; un cambio de variable linealiza a la ecuación, aunque no es una linealización estricta ya que el orden de la ecuación aumenta. Para conseguir esto se pone: $v = yc(x)$ con lo que se obtiene:

$$v' = c(x)a(x) + \left(b(x) + \frac{c'(x)}{c(x)}\right)v + v^2$$

ecuación que se transforma en la deseada por el cambio $v = -z'/z$

$$z'' - (b + c'/c)z' + acz = 0.$$

La solución general de esta ecuación se puede poner como una combinación lineal de dos soluciones linealmente independientes:

$$z(x) = k_1 z_1(x) + k_2 z_2(x)$$

con lo que la solución de la ecuación de Riccati es:

$$y(x) = -\frac{z'}{cz}$$

es decir:

$$y(x) = -\frac{k_1 z_1'(x) + k_2 z_2'(x)}{c(x)(k_1 z_1(x) + k_2 z_2(x))}$$

que no depende de dos constantes, sino solo de una, por ejemplo del cociente $k_1/k_2 = k$. En cualquier caso, el cálculo de las soluciones de esta ecuación lineal no es más sencillo que el problema inicial.

Una propiedad de la ecuación de Riccati es que la solución general se puede expresar en función de tres soluciones particulares distintas y de una constante, o lo que es lo mismo, la siguiente expresión es constante:

$$\frac{y(x) - y_1(x)}{y(x) - y_2(x)} \frac{y_3(x) - y_2(x)}{y_3(x) - y_1(x)} = C$$

de donde uno puede despejar $y(x)$ suponiendo que conoce las otras tres soluciones. Esta propiedad es debida a la relación que existe entre la ecuación de Riccati y el grupo de matrices 2×2 con coeficientes reales y determinante igual a uno, hecho también relacionado con la ecuación lineal de segundo orden que antes hemos estudiado. No podemos seguir aquí esta línea de razonamiento pues necesitaríamos un aparato matemático del que carecemos.

Ejemplo 1.4.7 $y' = y^2 - 2/x^2$.

Es fácil comprobar que $y(x) = 1/x$ es una solución particular de esta ecuación. Más difícil es establecer esa conjetura, aunque en este caso se puede pensar que una función del tipo k/x podría ser solución. Introduciéndola en la ecuación se obtienen dos valores diferentes de k para los que esta función es solución, $k = 1$ y $k = 2$. En cualquier caso, puesto que la ecuación es de Riccati, una vez conocida una solución, las demás se pueden obtener fácilmente desarrollando las ideas del apartado anterior.

$$u(x) = y(x) - \frac{1}{x}$$

la nueva ecuación es:

$$u' = u^2 + \frac{2u}{x}$$

ecuación de Bernouilli, que se resuelve poniendo $z = 1/u$:

$$z' = -1 - \frac{2z}{x}$$

una ecuación lineal no homogénea, resuelta en el próximo apartado.

1.4.5 La ecuación lineal de primer orden

La forma más general de una ecuación lineal de primer orden es:

$$y' = a(x)y + b(x) \tag{1.4}$$

donde $a(x)$ y $b(x)$ son funciones exclusivamente de x . Esta ecuación no es exacta, aunque un factor integrante que solo dependa de x se puede encontrar fácilmente. Sin embargo, procederemos de otra forma para calcular su solución. En primer lugar consideremos la llamada ecuación homogénea:

$$y' = a(x)y$$

que es una ecuación de variables separadas y por lo tanto integrable por cuadraturas:

$$\int \frac{dy}{y} = \int a(x) dx$$

es decir:

$$\log |y| = \int^x a(s) ds + C$$

o despejando $y(x)$:

$$y(x) = K e^{\int^x a(s) ds} \tag{1.5}$$

donde K es una constante real (lo que nos permite eliminar el valor absoluto que aparecía en el logaritmo neperiano). Esta es la solución general de la ecuación homogénea. Como se ve, todas las soluciones de

esta ecuación son proporcionales, o dicho de otra manera, forman un espacio vectorial real de dimensión igual a 1. El cálculo de la solución general de la ecuación completa puede hacerse utilizando el llamado método de variación de constantes, que consiste en probar soluciones de la forma:

$$y(x) = K(x)e^{\int^x a(s)ds} \quad (1.6)$$

sustituyendo en la ecuación (1.4), tenemos:

$$K'(x) = b(x)e^{-\int^x a(s)ds}$$

e integrando, se llega a una solución particular de la ecuación completa que se puede poner como:

$$y_p(x) = \left[\int^x b(t) \exp\left(-\int^t a(s)ds\right) dt \right] \exp\left(\int^x a(s)ds\right).$$

La solución general de la ecuación completa (1.4), es la suma de la solución general de la ecuación homogénea más una solución particular de la completa, es decir:

$$y(x) = Ke^{\int^x a(s)ds} + y_p(x). \quad (1.7)$$

Este resultado se puede demostrar fácilmente. Dada una solución de (1.4), $y_1(x)$, es claro que $y_1(x) - y_p(x)$ es una solución de la ecuación homogénea. Por tanto es de la forma dada en (1.5) para algún valor de K . Es evidente también que la función dada en (1.7) es solución de (1.4) sin más que revisar su cálculo.

Ejemplo 1.4.8 $z' = -z + 4\cos^2(x/2)$.

Esta ecuación se obtuvo a partir de una ecuación de Bernoulli. La solución general de la ecuación homogénea es:

$$z' = -z, \quad z(x) = Ke^{-x}$$

y una particular de la completa, obtenida por el método de variación de constantes:

$$K' = 4e^x \cos^2 \frac{x}{2}, \quad K(x) = e^x(2 + \operatorname{sen} x + \cos x)$$

por lo que la solución general es:

$$z(x) = Ke^{-x} + 2 + \operatorname{sen} x + \cos x.$$

Ejemplo 1.4.9 $z' = -1 - 2z/x$

Esta es la ecuación que obtuvimos al tratar de calcular las soluciones de una ecuación de Riccati. La ecuación homogénea correspondiente es:

$$z' = -\frac{2z}{x}$$

y su solución se obtiene de:

$$z(x) = Ke^{-2\int dx/x} = \frac{K}{x^2}.$$

En este caso, una solución particular de la homogénea se puede obtener por el método de variación de constantes (o directamente):

$$z_p(x) = -\frac{x}{3}.$$

La solución general es:

$$z(x) = -\frac{x}{3} + \frac{K}{x^2}.$$

La solución general de la ecuación de Bernoulli de donde proviene nuestro ejemplo será:

$$u(x) = \frac{3x^2}{C - x^3}$$

y la de la ecuación de Riccati donde comenzamos:

$$y(x) = \frac{3x^2}{C - x^3} + \frac{1}{x}.$$

Las ecuaciones lineales de primer orden son siempre resolubles por cuadraturas cuando las funciones $a(x)$ y $b(x)$ que aparecen en ellas son integrables, en particular cuando son continuas. Como veremos más adelante cuando la ecuación es de orden superior a uno, sólo si los coeficientes son constantes o en otros casos muy específicos, es posible encontrar la solución general en forma compacta.

Nótese que para la ecuación lineal homogénea, si se conoce una solución que no sea la trivial ($y(x) = 0$), las demás se obtienen de manera algebraica, es decir:

$$\frac{y(x)}{y_1(x)} = C.$$

Para la ecuación no homogénea necesitamos dos soluciones distintas, $y_1(x)$, $y_2(x)$, de forma que $y(x) - y_1(x)$, $y(x) - y_2(x)$ son soluciones de la ecuación homogénea, con lo que:

$$\frac{y(x) - y_1(x)}{y(x) - y_2(x)} = C$$

a comparar con la ecuación de Riccati donde son necesarias tres soluciones. De esta forma, las ecuaciones lineales, homogéneas o no, se pueden considerar como casos particulares de la ecuación de Riccati.

1.5 Métodos aproximados en ecuaciones de primer orden

Dado que no podemos conocer exactamente la solución de una ED arbitraria, resulta de suma importancia disponer de algún método que nos permita saber, al menos con cierto grado de aproximación, esa solución. Los métodos aproximados son muy variados y no podemos entrar aquí en una discusión profunda. Nos limitaremos a esbozar dos de ellos. El primero por su sencillez y el segundo porque abre la puerta a métodos más completos y que permiten un mayor control sobre los resultados y los errores que de forma indefectible van unidos a este tipo de métodos.

1.5.1 Poligonal de Euler

Supongamos que queremos calcular la solución de la ecuación diferencial de primer orden $y' = f(x, y)$, que verifica la condición inicial $y(a) = y_0$. Admitamos de momento que sabemos que dicha solución existe y puede definirse en un cierto intervalo que es donde la queremos encontrar. Puesto que y' nos da la derivada de la solución construiremos una poligonal a base de segmentos que tengan como pendiente el valor dado por la función f en ciertos puntos, que se determinan de acuerdo con una discretización apropiada del intervalo donde estamos trabajando. Para fijar ideas, supongamos que lo que queremos saber es el valor de la solución en el punto $x = b > a$. Dividamos el intervalo (a, b) en n partes iguales, de paso $h = (b - a)/n$. El valor de n influye en la exactitud del resultado, pues cuanto menor sea el paso, menores serán los errores en el valor obtenido de la solución. Sin embargo otro tipo de errores, debidos a los cálculos que hay que efectuar, influyen también sobre este valor por lo que n no puede aumentarse indefinidamente. Una vez fijado n , construimos un segmento que partiendo de $(a = x_0, y_0)$ tenga como pendiente $f(x_0, y_0)$, es decir el valor de la derivada de la solución en ese punto. Calculamos sobre este segmento el valor de y para $x_1 = x_0 + h$. Tenemos así un segundo punto de la poligonal $(x_1, y_1 = y_0 + hf(x_0, y_0))$. En este punto construimos otro segmento como el anterior pero ahora de pendiente $f(x_1, y_1)$ y calculamos el valor de y en el punto $x_2 = x_1 + h$. De esta manera obtenemos una poligonal que se asemeja a la curva solución y que en el punto $x = b$ nos dará un valor aproximado. Es decir:

$$\begin{aligned} y_1 &= y_0 + hf(x_0, y_0) \\ y_2 &= y_1 + hf(x_1, y_1) \\ &\dots \\ y_n &= y_{n-1} + hf(x_{n-1}, y_{n-1}). \end{aligned}$$

Ejemplo 1.5.1 Consideremos la ecuación de Riccati cuya solución encontramos anteriormente: $y' = y^2 - 2/x^2$, y apliquemos el método de Euler al cálculo del valor en $b = 2.1$ de la solución que pasa por $(2, 0.5)$. En las tablas siguientes se dan los valores que proporciona el método con diferentes pasos:

$$n = 1, a = x_0 = 2, y_0 = 0.5$$

$$b = x_1 = 2.1, y_1 = 0.475, y_{exacto} = 0.476190, \text{error} = 0.001$$

$$n = 10, a = x_0 = 2, y_0 = 0.5$$

$$x_1 = 2.01, y_1 = 0.497500$$

$$x_2 = 2.02, y_2 = 0.495025$$

$$x_3 = 2.03, y_3 = 0.492574$$

$$x_4 = 2.04, y_4 = 0.490147$$

$$x_5 = 2.05, y_5 = 0.487743$$

$$x_6 = 2.06, y_6 = 0.487356$$

$$x_7 = 2.07, y_7 = 0.483006$$

$$x_8 = 2.08, y_8 = 0.480671$$

$$x_9 = 2.09, y_9 = 0.478359$$

$$x_{10} = 2.1, y_{10} = 0.476069, \text{error} = 0.0001$$

$$n = 100, x_0 = 2, y_0 = 0.5$$

$$x_{100} = 2.1, y_{100} = 0.476178, \text{error} = 0.00001.$$

Como otro ejemplo que usaremos más tarde, el valor en $x = 3$ de esta solución, con $n = 10$, es 0.322438, frente al valor exacto de 0.333333.

1.5.2 Método de Runge-Kutta

El método de Euler equivale en cierto sentido a un desarrollo de Taylor de primer orden. Desarrollos hasta órdenes superiores proporcionan una mejor aproximación. Sin embargo, éstos implican el cálculo de derivadas más complicadas lo que conlleva dificultades en el cálculo y sucesivos errores de redondeo que se van acumulando. El método de Runge-Kutta expuesto aquí, equivale a un desarrollo de Taylor hasta el cuarto orden pero sustituye las derivadas de estos órdenes por una media ponderada:

$$y_{i+1} = y_i + \frac{h}{6}(k_{i1} + 2k_{i2} + 2k_{i3} + k_{i4})$$

donde los coeficientes k vienen dados por:

$$k_{i1} = f(x_i, y_i)$$

$$k_{i2} = f(x_i + h/2, y_i + hk_{i1}/2)$$

$$k_{i3} = f(x_i + h/2, y_i + hk_{i2}/2)$$

$$k_{i4} = f(x_i + h, y_i + hk_{i3}).$$

Es decir, los coeficientes k_{ij} sustituyen en el punto x_i a las derivadas de la solución de orden menor o igual que 4.

Se puede probar que el error local es proporcional a h^5 , mientras que el error acumulado lo es a h . Existen métodos de Runge-Kutta a órdenes mayores, y métodos con fundamentos distintos que mejoran el cálculo y disminuyen los errores (métodos corrector-predictor, Adams-Moulton, etc).

Ejemplo 1.5.2 El caso anterior se puede tratar por el método de Runge-Kutta. Para $n = 10$, el valor obtenido para $x = 2.1$ es 0.476190 y el error cometido es menor que 10^{-11} . La mejor aproximación que proporciona este método permite obtener, por ejemplo, el valor de esta solución en $x = 8$, $n = 60$, con un error menor que 10^{-5} , $y_{60} = 0.124997$ frente al valor exacto 0.125.

Sin embargo, antes de aplicar un método numérico es necesario saber algo de la solución y de sus posibles singularidades. El siguiente ejemplo muestra lo errado de la utilización del cálculo numérico sin antes saber si la solución existe para los valores que se está intentando calcular.

Ejemplo 1.5.3 Consideremos nuevamente la ecuación de Riccati: $y' = y^2 - 2/x^2$, pero intentemos calcular ahora la solución con valor inicial: $y(2) = 1$. Usando el método de Euler con $n = 10$ podemos calcular el valor en $x = 3$, que resulta ser 3.248222 frente a 5.733333 que es el valor exacto. El error cometido es mucho mayor que antes, para la solución con dato inicial $y(2) = 0.5$. Si usamos el método de Runge-Kutta, también con $n = 10$, obtendremos 5.726704 que es un valor mucho más aproximado. Sin embargo, supongamos que queremos calcular el valor de esta solución en $x = 3.2$. El método de Euler con $n = 10$ nos da 6.093062, mientras que el Runge-Kutta con el mismo número de pasos da: 447.0796. En realidad, ninguno de los dos resultados es correcto. En este caso disponemos de la solución exacta por lo que podemos averiguar que es lo que ocurre. La solución es:

$$y(x) = \frac{1}{x} + \frac{3x^2}{32 - x^3}$$

que tiene una singularidad en $x = 3.1748$, por lo que no tiene sentido preguntarse cuál es su valor en $x = 3.2$, donde no está definida.

Nótese también, que los errores cometidos en este caso son mayores que en el anterior, debido a la variación tan fuerte de la solución.

1.6 Existencia y unicidad

Puesto que en muchas ocasiones no dispondremos de la solución explícita de una ecuación diferencial, resulta muy interesante el estudio de las propiedades de esas posibles soluciones así como de su existencia y unicidad. En general se entiende que un problema está bien planteado cuando tiene solución única y ésta depende en forma continua tanto de los parámetros de la ecuación como de los datos iniciales. En esta sección estudiaremos un teorema que permite asegurar existencia y unicidad de soluciones.

Se trata de un teorema de Cauchy, primeramente demostrado por Liouville (1838) y posteriormente por Peano y Picard que le dio la forma que estudiaremos. Es un teorema de suficiencia y daremos ejemplos de existencia aún cuando alguna de las hipótesis no se cumplan y de cómo la unicidad puede perderse si se levanta alguna de ellas sin sustituirla por algo similar. Necesitamos antes de enunciar estos teoremas, el concepto de función lipschitziana.

Definición 1.6.1 Sea $f: I \rightarrow \mathbb{R}$ una función definida en algún intervalo de la recta. Se dice que f verifica la condición de Lipschitz si:

$$|f(x) - f(y)| \leq L|x - y|, \quad \forall x, y \in I$$

donde L es la constante de Lipschitz de la función f .

Por ejemplo, la función $f(x) = x^2$ es lipschitziana en cualquier intervalo acotado de la recta (aunque no en toda la recta, pues la constante depende del intervalo). Veamos ahora como toda función derivable con derivada continua es lipschitziana localmente.

Lema 1.6.1 Sea $f: I \rightarrow \mathbb{R}$ una función derivable y con derivada acotada en I . Entonces $f(x)$ es lipschitziana y una constante de Lipschitz es el máximo de la derivada en I .

Demostración. Por el teorema del valor medio dados $x, y \in I$, existe $c \in I$ tal que $x < c < y$, verificándose:

$$|f(x) - f(y)| \leq |f'(c)| |x - y|.$$

Pero f' está acotada en I , luego se verifica el lema. QED

Hay funciones que no son derivables y sin embargo son lipschitzianas. Por ejemplo el valor absoluto en el origen. Veamos un ejemplo de una función que no es lipschitziana (y por lo tanto no es derivable, aunque sí continua): $f(x) = \sqrt{x}$ es continua en $x = 0$ aunque no derivable ni lipschitziana. En efecto, aplicando la definición:

$$f(x) - f(y) = \sqrt{x} - \sqrt{y} = \frac{x - y}{\sqrt{x} + \sqrt{y}}$$

que no se puede acotar en un entorno de 0. En lo que sigue, consideraremos funciones de varias variables y exigiremos que sean lipschitzianas en una o varias de ellas. Los comentarios y propiedades anteriormente expuestos son válidos sin más que considerar, por ejemplo, derivadas parciales en vez de totales. Así, para una función real de dos variables, se dice que es lipschitziana en un conjunto de \mathbb{R}^2 en la segunda variable, si:

$$|f(t, x_1) - f(t, x_2)| \leq L|x_1 - x_2|$$

para $(t, x_1), (t, x_2)$ en el dominio considerado. Nótese que la constante L no puede depender de t . En general si $f: I \times D \rightarrow \mathbb{R}^n$ con $D \subset \mathbb{R}^n$, se dice que es lipschitziana en $x \in D$ si:

$$\|f(t, x_1) - f(t, x_2)\| \leq L\|x_1 - x_2\|$$

donde $\|\cdot\|$ es la norma en \mathbb{R}^n y $t \in I$, que es un intervalo de \mathbb{R} . Entonces, la existencia y acotación de las derivadas parciales de f con respecto a x_i implica la lipschitzianidad en esas variables.

Teorema 1.6.1 (*Existencia y unicidad*). Sea la ecuación diferencial de primer orden: $x' = f(t, x)$ donde f es una función continua en:

$$D = \{(t, x) : |t - t_0| \leq a, |x - x_0| \leq b\}$$

y lipschitziana en la segunda variable en el mismo conjunto D (la condición de lipschitzianidad no necesita establecerse en todo D , pero no entraremos en estos detalles). Entonces existe una solución única $x(t)$ que verifica $x(t_0) = x_0$ y que está definida en un entorno de t_0 , $|t - t_0| \leq T$, donde $T = \min\{a, b/K\}$ siendo $K = \sup\{|f(t, x)| : (t, x) \in D\}$.

Demostración. Transformamos la ecuación diferencial en una integral:

$$x(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds$$

y cuando t verifica $0 \leq t - t_0 \leq T$ (para $t < t_0$ el razonamiento es similar), construimos una sucesión de funciones de la siguiente forma:

$$x_1(t) = x_0 + \int_{t_0}^t f(s, x_0) ds$$

donde hemos sustituido la función $x(s)$ en la integral por su valor en t_0 . La segunda aproximación se construye a partir de la primera por el mismo procedimiento:

$$x_2(t) = x_0 + \int_{t_0}^t f(s, x_1(s)) ds$$

y así sucesivamente, de forma que la aproximación de orden n es:

$$x_n(t) = x_0 + \int_{t_0}^t f(s, x_{n-1}(s)) ds$$

de esta forma tenemos una sucesión de funciones $x_n(t)$ definidas en el conjunto D . Veamos ahora como la sucesión $x_n(t)$ converge uniformemente en el intervalo $0 \leq t - t_0 \leq T$. En primer lugar, comprobemos que $x_n(t)$ no se sale de D :

$$|x_n(t) - x_0| \leq b$$

Esto es cierto para $n = 1$:

$$|x_1(t) - x_0| = \left| \int_{t_0}^t f(s, x_0) ds \right| \leq \int_{t_0}^t |f(s, x_0)| ds \leq K(t - t_0)$$

y por inducción se demuestra para cualquier n . Supongamos que es cierto para n :

$$|x_{n+1}(t) - x_0| = \left| \int_{t_0}^t f(s, x_n(s)) ds \right| \leq K(t - t_0) \leq b$$

pues $x_n(s)$ está en D para todo s en el intervalo $[t_0, t_0 + T]$. Usando el mismo procedimiento de inducción, podemos probar la siguiente desigualdad:

$$|x_{n+1}(t) - x_n(t)| \leq \frac{KL^n}{(n+1)!} (t - t_0)^{n+1}$$

que se verifica para $n = 1$, pues es la misma desigualdad anterior. Supuesto cierto para n ,

$$|x_{n+1}(t) - x_n(t)| = \left| \int_{t_0}^t [f(s, x_n(s)) - f(s, x_{n-1}(s))] ds \right|$$

y usando la condición de Lipschitz:

$$\leq L \int_{t_0}^t |x_n(s) - x_{n-1}(s)| ds \leq L \int_{t_0}^t \frac{KL^{n-1}}{n!} (s - t_0)^n ds = \frac{KL^n}{(n+1)!} (t - t_0)^{n+1}$$

esta desigualdad se verifica para todo $t \in [t_0, t_0 + T]$. De esta forma construimos una serie de funciones cuyos términos están mayorados por una serie numérica convergente y que por lo tanto converge uniformemente en el intervalo considerado:

$$x_0 + \sum_{n=1}^{\infty} (x_n(t) - x_{n-1}(t))$$

Las sumas parciales de esta serie son iguales a $x_n(t)$, luego, la sucesión $x_n(t)$ converge uniformemente a una función continua $x(t)$. Probaremos ahora que esa función verifica la ecuación diferencial y la condición inicial:

$$x(t) = \lim_{n \rightarrow \infty} x_n(t) = x_0 + \lim_{n \rightarrow \infty} \int_{t_0}^t f(s, x_n(s)) ds = x_0 + \int_{t_0}^t f(s, x(s)) ds$$

donde se han usado las propiedades de f adecuadamente, es decir, se ha intercambiado la integral por el límite debido a la lipschitzianidad de f y se ha tomado el valor de la función en el límite debido a la continuidad de f . Al ser f continua, $x(t)$ es diferenciable y por lo tanto verifica la ecuación diferencial. Es inmediato comprobar que también cumple la condición inicial.

Demostramos ahora la unicidad de la solución: sea $\tilde{x}(t)$ otra solución distinta de $x(t)$ definida en $[t_0 - T', t_0 + T']$ con $T' \leq T$ que verifica la ecuación y la condición inicial y además:

$$|\tilde{x}(t) - x_0| \leq b, \quad t \in [t_0 - T', t_0 + T']$$

Entonces, para $0 \leq t - t_0 \leq T$:

$$|\tilde{x}(t) - x_n(t)| \leq \int_{t_0}^t |f(s, \tilde{x}(s)) - f(s, x_{n-1}(s))| ds \leq L \int_{t_0}^t |\tilde{x}(s) - x_{n-1}(s)| ds$$

ahora bien, de nuevo por inducción, se puede probar que:

$$|\tilde{x}(t) - x_n(t)| \leq \frac{L^n b}{n!} (t - t_0)^n$$

por lo que $x_n(t)$ tiende a $\tilde{x}(t)$ cuando n tiende a infinito, lo que prueba la igualdad de $x(t)$ y $\tilde{x}(t)$ en el intervalo común de definición de ambas funciones. QED

Ejemplo 1.6.1 $x' = -2x/t + 4t$. La función f no es continua en $t = 0$. Sin embargo es posible calcular la solución que verifica $x(0) = 0$, a saber $x(t) = t^2$, que existe, es única y además continua y diferenciable en $t = 0$. Cualquier otra solución, que viene dada por $x(t) = t^2 + K/t^2$, es singular en $t = 0$. Compárese este ejemplo con $x' = x/t$. La singularidad de la función f es igual que en el caso anterior, pero las soluciones son $x(t) = Kt$, así que por el origen pasan infinitas soluciones. En estos dos ejemplos la función $f(t, x)$ no está definida en $t = 0$. Veamos ahora un caso donde la discontinuidad de f es de salto y existe una única solución al problema.

Ejemplo 1.6.2

$$x' = \begin{cases} x(1 - 2t) & t > 0 \\ x(2t - 1) & t < 0 \end{cases}$$

la función f es discontinua en $t = 0$ ($x \neq 0$). Calculemos la solución que verifica $x(1) = 1$. En el intervalo $t > 0$, la solución es: $x(t) = Ke^{t(1-t)}$, y aplicando la condición inicial, obtenemos $K = 1$. En el intervalo $t < 0$, la solución es: $x(t) = K'e^{-t(1-t)}$. Podemos empalmar estas dos funciones de forma continua en $t = 0$ calculando así el valor de $K' = 1$. La solución completa es

$$x(t) = \begin{cases} e^{t(1-t)} & t > 0 \\ e^{-t(1-t)} & t < 0 \end{cases}$$

que es continua en $t = 0$, aunque obviamente no derivable. Su derivada tiene un salto en $t = 0$, justamente el mismo que tenía la función f . Esta es la única solución de la ecuación que verifica $x(1) = 1$.

Si f es continua es posible probar que siempre existe solución. Sin embargo, si f no es lipschitziana la solución puede no ser única.

Ejemplo 1.6.3 Consideremos la función $f(x) = \sqrt{|x|}$ que como vimos antes era continua pero no lipschitziana. La ecuación diferencial $x' = f(x)$ con la condición inicial $x(0) = 0$, no tiene solución única: tanto $x(t) = 0$ como: $x(t) = t^2/4, x \geq 0$, $x(t) = -t^2/4, x \leq 0$, son soluciones de esa ecuación.

1.7 Estabilidad

Una de las cuestiones más interesantes que se pueden plantear en torno a las ecuaciones diferenciales es la de estabilidad de las soluciones, es decir, si se varían los datos iniciales una pequeña cantidad, qué se puede decir del comportamiento, para todo valor de la variable independiente, de las soluciones correspondientes. El tema no resulta sencillo y solo para algunos tipos de ecuaciones se pueden dar respuestas generales. Sin embargo, como iremos viendo a lo largo de estas notas, existen teorías y resultados muy potentes para tratar muchos casos. En esta sección nos limitaremos a la definición de estos conceptos.

Definición 1.7.1 Consideremos el problema:

$$\begin{cases} x' = f(t, x) \\ x(t_0) = x_0 \end{cases}$$

que se supone admite una solución única, $x(t)$, definida para todo $t \geq t_0$. Se dice que $x(t, x_0)$ es una solución estable (a la derecha) si: $\forall \epsilon > 0, \exists \delta > 0$ tal que, si $\tilde{x}_0 \in \mathbb{R}^n$ verifica $\|\tilde{x}_0 - x_0\| < \delta$, entonces $\exists x(t, \tilde{x}_0)$ solución con $x(t_0, \tilde{x}_0) = \tilde{x}_0$, definida en $[t_0, \infty)$ y tal que:

$$\|x(t, \tilde{x}_0) - x(t, x_0)\| < \epsilon, \quad \forall t \geq t_0$$

es decir, las soluciones que tienen puntos iniciales cerca del punto inicial de la solución a estudiar, se mantienen cerca de ésta para todo tiempo posterior. Se puede considerar una clase más particular de estabilidad, la llamada estabilidad asintótica.

Definición 1.7.2 (Estabilidad asintótica) *Se dice que $x(t, x_0)$ es una solución asintóticamente estable, si es estable y si existe $\eta > 0$ tal que, si $\|\tilde{x}_0 - x_0\| < \eta$ entonces,*

$$\lim_{t \rightarrow \infty} \|x(t, \tilde{x}_0) - x(t, x_0)\| = 0.$$

Nótese que se necesita que la solución sea previamente estable. Es decir, la condición de tender a la solución en el infinito no es suficiente para asegurar la estabilidad.

1.8 Ecuaciones autónomas

Estudiaremos aquí, como un ejemplo, las ecuaciones autónomas de primer orden, que permiten un tratamiento sencillo, y nos abren la puerta de los sistemas autónomos, que aparecerán en el tema 3.

Definición 1.8.1 *Una ecuación autónoma es: $x' = f(x)$, es decir la función f no depende de t .*

Esto hace que la ecuación (que puede ser resuelta por cuadraturas al ser de variables separadas) presente unas propiedades especiales que hacen su estudio más sencillo. Veamos en primer lugar unas propiedades elementales.

Proposición 1.8.1 *Sea $x' = f(x)$, con f derivable con continuidad en \mathbb{R} .*

1. Si $x(t)$ es una solución, $x(t + \tau)$ también es solución, para todo $\tau \in \mathbb{R}$.
2. Si $a \in \mathbb{R}$ verifica $f(a) = 0$, entonces $x(t) = a$ es una solución constante de la ecuación diferencial.
3. Toda solución es constante o estrictamente monótona.

Demostración. La parte 1) es una consecuencia inmediata del hecho de ser f función exclusiva de x . La parte 2) es igualmente trivial. En cuanto a 3), puesto que por cada punto de \mathbb{R}^2 pasa una solución y solo una, es claro que para que una solución pasara de ser creciente a decreciente o viceversa, habría de atravesar un extremo, donde f se hace cero. Pero ese punto es otra solución, constante por la parte 2) de la proposición. QED

Una propiedad que nos será muy útil para describir las soluciones de una ecuación autónoma y que es una consecuencia de un resultado elemental de análisis es la siguiente.

Teorema 1.8.1 *Todas las soluciones acotadas de una ecuación autónoma tienden a una solución constante. Aquí la acotación debe entenderse en el sentido que ya le dimos antes: $x(t)$, con $x(t_0) = x_0$ está acotada a la derecha si $\exists M > 0$ tal que $|x(t)| \leq M$ para todo $t \geq t_0$.*

Demostración. Sea $x(t)$ una solución acotada, no constante. Por ser monótona, existe el límite de $x(t)$ cuando t tiende a infinito. Sea a ese límite. La función f es continua, así que también existe el límite de f cuando t tiende a infinito, es decir el límite cuando $x \rightarrow a$, $f(a)$. Entonces:

$$\lim_{t \rightarrow \infty} x(t) = a, \quad \lim_{t \rightarrow \infty} x'(t) = f(a).$$

Un resultado de análisis elemental demuestra que $f(a) = 0$ y por tanto que $x(t) = a$ es una solución constante. En efecto, sea la sucesión:

$$a_n = x(n+1) - x(n).$$

Puesto que $x(t)$ es continua, su límite cuando $n \rightarrow \infty$ es igual a cero. Por el teorema del valor medio para cada n podemos encontrar $\xi_n \in (n, n+1)$ tal que:

$$x(n+1) - x(n) = x'(\xi_n).$$

Además, si $n \rightarrow \infty$ entonces $\xi_n \rightarrow \infty$. Por lo tanto:

$$\lim_{t \rightarrow \infty} x'(t) = \lim_{t \rightarrow \infty} x'(\xi_n) = \lim_{t \rightarrow \infty} (x(n+1) - x(n)) = 0$$

con lo que $f(a) = 0$.

Usando técnicas similares, se puede probar que las soluciones acotadas por la izquierda tienden a una solución constante cuando $t \rightarrow -\infty$, y que las soluciones comprendidas entre dos soluciones constantes están definidas para todo t . QED

En cuanto se refiere a la estabilidad de las ecuaciones autónomas, es inmediato a la vista de lo expuesto anteriormente, demostrar el siguiente teorema:

Teorema 1.8.2 Si $x(t) = a$ es una solución constante de la ecuación autónoma $x' = f(x)$ (es decir, $f(a) = 0$), entonces:

1. si $f'(a) < 0$, la solución es estable asintóticamente.
2. si $f'(a) > 0$, la solución es inestable.

En el caso $f'(a) = 0$, no se puede asegurar nada con el estudio de la primera derivada de f .

Demostración. La idea es estudiar los cambios de signo de la función f al atravesar x un cero de esa función. Si en el sentido de x crecientes, f pasa de ser negativa a positiva, $f'(a) > 0$, y la solución $x(t) = a$ es inestable. Si f pasa de ser positiva a negativa ($f'(a) < 0$) la solución es asintóticamente estable. Finalmente, si $f'(a) = 0$, f tiene un punto crítico en ese punto, y hay que analizar con derivadas superiores si existe cambio de signo o no. En el caso de que no haya cambio de signo, la solución es inestable. Si lo hay, basta referirse a la discusión anterior. QED

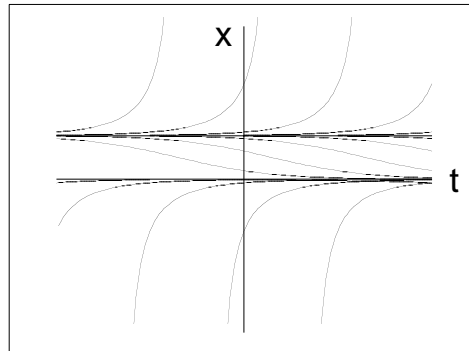


Figura 1.10

Ejemplo 1.8.1 $x' = x(x - 1)$. Esta ecuación autónoma tiene dos soluciones constantes $x(t) = 0$, $x(t) = 1$, correspondientes a los dos ceros de la función f . En $x = 0$, $f'(0) = -1$, por lo que se trata de una solución, asintóticamente estable. En $x = 1$, $f'(1) = 1$, así que esta solución es inestable (1.10). La solución explícita puede calcularse fácilmente:

$$\int \frac{dx}{x(x-1)} = \int dt \Rightarrow \log(x-1) - \log x = t + C$$

de donde uno puede despejar $x(t)$:

$$x(t) = \frac{1}{1 - Ke^{t}}$$

Si estudiamos las soluciones con condición inicial: $x(0) = x_0$, se tiene:

$x_0 > 1$: la solución no se puede prolongar hasta $t = \infty$, pues tiene una asíntota vertical en:

$$0 < t_1 = \log \frac{x_0}{x_0 - 1}$$

Su límite cuando t tiende $-\infty$ a es 1

$x_0 < 0$: la solución se puede prolongar hasta $t = \infty$. Hacia la izquierda presenta una asíntota vertical en:

$$0 > t_1 = \log \frac{x_0}{x_0 - 1}$$

Su límite cuando t tiende a ∞ es 0.

$0 < x_0 < 1$: La solución se puede prolongar indefinidamente a derecha e izquierda. Su límite cuando t tiende a $+\infty$ es 0 y cuando t tiende a $-\infty$ es 1.

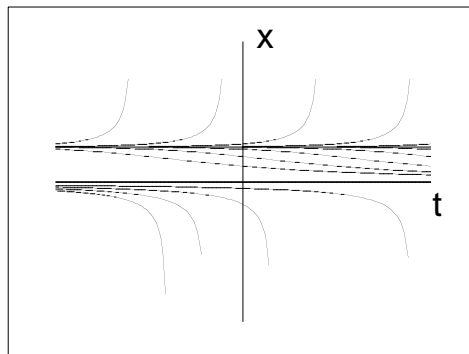


Figura 1.11

Ejemplo 1.8.2 $x' = x^2(x - 1)$. Como antes, hay dos soluciones constantes. La solución $x(t) = 0$ es ahora inestable, aunque no se pueda determinar esta característica de la derivada de f que se hace cero en $x = 0$. La otra solución constante, $x(t) = 1$ sigue siendo inestable (figura 1.11).

Ejemplo 1.8.3 $x' = x^3(x - 1)$. La solución $x(t) = 1$ es inestable. La solución $x(t) = 0$ es estable asintóticamente, aún cuando $f'(0) = 0$. En el caso anterior, $f'(0) = -2 < 0$, mientras que aquí, $f'(0) = 0$ y $f''(0) = -6 < 0$, con lo cuál uno puede estudiar los cambios de signo de la función f al atravesar $x = 0$ (figura 1.12).

Para terminar esta sección dedicada a las ecuaciones autónomas, nótese que podemos proyectar las soluciones sobre el eje x . Debido a las propiedades de estas ecuaciones, las soluciones que se obtienen unas de otras por traslación, se proyectan en los mismos puntos aunque para distinto t . Las soluciones constantes se proyectan en un solo punto (puntos críticos). Si se supone que las gráficas de las soluciones se recorren en el sentido de t creciente, se obtienen las gráficas que aparecen en la figura 1.13. Las soluciones que unen dos puntos críticos están definidas para todo t . Tienden a uno u otro punto crítico cuando t tiende a más o menos infinito. Las otras soluciones pueden no estar definidas para todo t ,

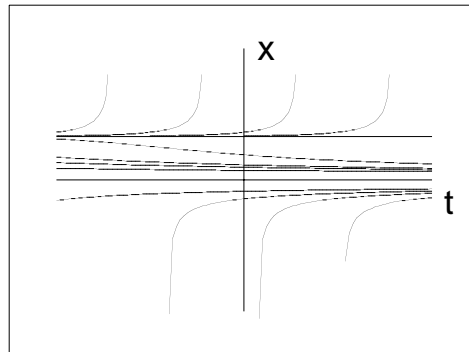


Figura 1.12

en particular las del ejemplo no lo están, llegan a infinito en un tiempo finito, y tienden a un punto crítico, bien cuando t tiende a más infinito o cuando t tiende a menos infinito. Observando las flechas se advierte con claridad la estabilidad o inestabilidad de los puntos críticos. Como hemos dicho antes, todos ellos son asintóticamente estables (focos estables) o inestables (focos inestables si las flechas se alejan del punto crítico por los dos lados, es decir ninguna solución tiende a ese punto, o puntos silla, si una de las flechas entra en el punto y la otra sale, es decir hay soluciones que tienden a ese punto y otras que se alejan). Todas estas propiedades requieren una cierta regularidad de la función $f(x)$. Es posible encontrar ejemplos con f no derivable donde dejan de ser ciertos.

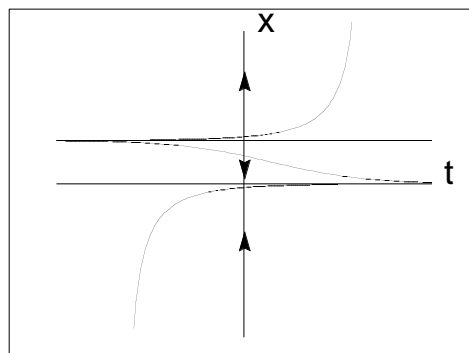


Figura 1.13

Tema 2

Sistemas y ecuaciones lineales

Toda ecuación diferencial de orden n puede ponerse como un sistema de ecuaciones de primer orden. Es obvio que habrá sistemas que no tengan esa forma tan especial, aunque en muchos casos sea posible encontrar una ecuación a la que equivalen. Conocemos condiciones suficientes bajo las cuales un sistema de ecuaciones diferenciales admite una única solución, especificadas las condiciones de Cauchy adecuadas.

Estudiaremos en este capítulo sistemas lineales, sobre todo de coeficientes constantes.

2.1 Sistemas lineales

Un sistema de ecuaciones lineales de primer orden tiene la forma:

$$x' = A(t)x + b(t)$$

donde $x(t)$ y $b(t)$ son funciones vectoriales de variable real y $A(t)$ es una función de variable real con valores en un espacio de matrices cuadradas. Tanto $A(t)$ como $b(t)$ se suponen definidas en un intervalo de la recta $[t_0, t_1]$. Los valores de A y b podrán ser complejos, aunque en general los consideraremos reales. Además podría pensarse en sistemas con matrices no cuadradas, es decir con menos ecuaciones que incógnitas o viceversa, pero nos limitaremos al caso cuadrado.

Las cuestiones de existencia y unicidad de soluciones para este tipo de problemas pueden deducirse del caso general directamente.

Teorema 2.1.1 *Sea el problema:*

$$\begin{cases} x' &= A(t)x + b(t) \\ x(t_0) &= x_0 \end{cases}$$

donde:

$$A: [t_0, t_1] \rightarrow M_n(\mathbb{R}), \quad b: [t_0, t_1] \rightarrow \mathbb{R}^n$$

son funciones continuas. Entonces existe una única solución al sistema que verifica la condición inicial dada.

Demostración. Es una aplicación de los teoremas estudiados. Veamos como la función $f(t, x) = A(t)x + b(t)$ es continua y lipschitziana. Sobre la continuidad no hay ningún problema puesto que $A(t)$ y $b(t)$ lo son y x aparece en forma lineal. En cuanto a la lipschitzianidad, aplicando la definición:

$$\|f(t, x) - f(t, y)\| = \|A(t)x - A(t)y\| \leq \|A(t)\|_s \|x - y\|$$

donde $\|A(t)\|_s$ es la norma del supremo en el espacio de matrices $M_n(\mathbb{R})$:

$$\|A(t)\|_s = \sup\left\{\frac{\|A(t)x\|}{\|x\|}, x \in \mathbb{R}^n, x \neq 0\right\}$$

por tanto, si $\|A(t)\|_s$ es una función continua en un intervalo compacto, tiene un máximo que es una constante de Lipschitz para la función $f(t, x)$. Aplicando el teorema de Cauchy-Peano, podemos asegurar, al menos localmente, la existencia y unicidad de la solución. QED

Si $A(t)$ y $b(t)$ son continuas en toda la recta, la solución se puede extender a todo \mathbb{R} . En particular, los sistemas lineales con coeficientes constantes tienen solución definida para todo t .

2.2 El espacio de soluciones de la ecuación lineal

Una vez en condiciones de asegurar la existencia y unicidad de las soluciones, estudiaremos ahora las propiedades de éstas. Para ello consideraremos en primer lugar la ecuación homogénea $x' = A(t)x$.

Teorema 2.2.1 *Una solución de un sistema lineal es un vector de n componentes y el conjunto de soluciones es un espacio vectorial real (como puede probarse fácilmente por las propiedades del producto de matrices). La dimensión de este espacio es igual a n .*

Para demostrarlo, definiremos antes la noción de matriz de soluciones y estudiaremos sus propiedades.

Definición 2.2.1 *Una matriz de soluciones es una matriz cuyos vectores columna son soluciones del sistema: $X(t) = (x^1(t), x^2(t), \dots, x^n(t))$.*

Debido a las propiedades del producto de matrices, se tiene:

$$X' = A(t)X$$

cuando $X(t)$ es una matriz de soluciones. La siguiente proposición se refiere al determinante de una matriz de soluciones. Téngase en cuenta que siempre estamos considerando a $A(t)$ como una función continua en el intervalo $I = [t_0, t_1]$ en el que trabajamos.

Proposición 2.2.1 *Sean $x^1(t), x^2(t), \dots, x^n(t)$, n soluciones de un sistema de ecuaciones lineal homogéneo y sea $X(t)$ la matriz que las tiene por columnas. Entonces: o bien $\det X(t) = 0$ para todo $t \in I$, o $\det X(t) \neq 0$ para todo $t \in I$.*

En el primer caso las soluciones son linealmente dependientes. En el segundo son linealmente independientes. El determinante de la matriz se llama wronskiano, y como asegura la proposición es siempre distinto de cero en el caso que las columnas de la matriz de soluciones sean linealmente independientes. Nótese que la condición que establece la proposición para asegurar la dependencia o independencia lineal de un conjunto de funciones vectoriales no se aplica en general, sino sólo a soluciones de un sistema lineal homogéneo.

Demostración. Basta recordar como se deriva un determinante para obtener:

$$\frac{d}{dt} \det \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \vdots & & \vdots \\ x_n^1 & \dots & x_n^n \end{pmatrix} = \det \begin{pmatrix} x_1'^1 & \dots & x_1'^n \\ \vdots & & \vdots \\ x_n^1 & \dots & x_n^n \end{pmatrix} + \dots + \det \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \vdots & & \vdots \\ x_n'^1 & \dots & x_n'^n \end{pmatrix}$$

sustituyendo las derivadas por sus expresiones respectivas y desarrollando los determinantes por la fila en la que esas derivadas aparecen, se llega a:

$$\frac{d}{dt} \det(X(t)) = \text{tr}(A(t)) \det(X(t)).$$

Se trata de una ecuación lineal de primer orden que sabemos resolver. Su solución es la fórmula de Abel-Liouville:

$$\det X(t) = C e^{\int^t \text{tr} A(s) ds}$$

y por lo tanto, el wronskiano no se anula para ningún t en el intervalo considerado, o bien (cuando $C = 0$) se anula para todo punto. QED

Para ver la relación que existe entre el wronskiano y la dependencia e independencia lineal de un conjunto de soluciones, consideremos esta otra forma de demostrar la proposición anterior. Supongamos que existe $\tau \in I$ tal que $\det X(\tau) = 0$. En este caso, existe un conjunto de constantes reales, c_1, c_2, \dots, c_n , no todas nulas, tales que:

$$\sum_{i=1}^n c_i x^i(\tau) = 0 \quad (2.1)$$

Consideremos la función:

$$x(t) = \sum_{i=1}^n c_i x^i(t) \quad (2.2)$$

que al ser una combinación lineal de soluciones es una solución del sistema homogéneo. Además verifica una condición inicial (en τ) dada por 2.1. De acuerdo con el teorema de existencia y unicidad, la solución dada en 2.2 es la solución trivial, y por tanto:

$$\sum_{i=1}^n c_i x^i(t) = 0, \quad \forall t \in I$$

es decir las funciones $x^i(t)$, $i = 1, \dots, n$ son linealmente dependientes. Obviamente, si son linealmente independientes, el wronskiano no se anula en ningún punto.

Proposición 2.2.2 *La dimensión del espacio de soluciones de un sistema de n ecuaciones lineales homogéneas es igual a n .*

Demostración. Consideremos las soluciones de los problemas:

$$\begin{cases} x' &= A(t)x \\ x(t_0) &= e_i \end{cases}$$

donde e_i es la base canónica de \mathbb{R}^n . Sean $x^i(t)$ estas soluciones. Entonces, el conjunto de soluciones $\{x^i(t), i = 1, \dots, n\}$ es una base del espacio de soluciones, que por tanto tiene dimensión n . En efecto, es inmediato probar que son linealmente independientes, pues en $t = t_0$ el wronskiano es el determinante de la matriz identidad, que es distinto de cero. Además son un sistema de generadores:

Sea $x(t)$ una solución del sistema, con $x(t_0) = x_0$. El vector x_0 tiene una cierta expresión en la base canónica:

$$x_0 = \sum_{i=1}^n x_{0i} e_i$$

Consideremos la solución del sistema construida como una combinación lineal de las soluciones $x^i(t)$ con coeficientes x_{0i} :

$$\sum_{i=1}^n x_{0i} x^i(t).$$

Esta solución es justamente $x(t)$, pues ambas verifican la misma condición inicial en $t = t_0$. QED

Definición 2.2.2 *Se llama sistema fundamental de soluciones a una base del espacio de soluciones. Se llama matriz fundamental a una matriz de soluciones cuyas columnas forman un sistema fundamental.*

De esta manera, la solución general del sistema homogéneo se puede escribir como:

$$x(t) = \Phi(t)C$$

donde $\Phi(t)$ es una matriz fundamental y C es un vector de constantes de n componentes. Si la condición inicial es $x(t_0) = x_0$, el vector C se calcula como:

$$C = \Phi(t_0)^{-1}x_0$$

(pues $\Phi(t)$ tiene una inversa) y por lo tanto la correspondiente solución es:

$$x(t) = \Phi(t)\Phi(t_0)^{-1}x_0$$

La matriz $\Psi(t) = \Phi(t)\Phi(t_0)^{-1}$ es una matriz fundamental canónica, ya que $\Psi(t_0) = I_n$, la matriz identidad:

$$x(t) = \Psi(t)x_0.$$

En cuanto al sistema completo (con $b(t)$) las soluciones forman un espacio afín, es decir:

Proposición 2.2.3 *El conjunto de soluciones del sistema lineal completo $x' = A(t)x + b(t)$, es $x_p(t) + H$, donde H es el espacio de soluciones de la ecuación homogénea y $x_p(t)$ es una solución particular del problema completo.*

Demostración. Es claro que $x_p(t) + x_h(t)$, donde $x_h(t)$ es cualquier solución del sistema homogéneo, es una solución del sistema completo, como se ve sin más que sustituir en la ecuación. Además, ésta es la solución general. En efecto, sea $y(t)$ una solución arbitraria del sistema completo. Entonces, $y(t) - x_p(t)$ es una solución del sistema homogéneo:

$$y'(t) - x_p'(t) = A(t)(y(t) - x_p(t))$$

de donde se deduce la proposición. QED

No es fácil calcular las soluciones de un sistema lineal homogéneo o completo. Sin embargo, una vez conocida una matriz fundamental, lo que proporciona todas las soluciones del problema homogéneo, se pueden hallar las soluciones del problema completo aplicando, por ejemplo el método de variación de constantes, como hicimos con las ecuaciones de primer orden.

Para ello, sea $\Phi(t)$ una matriz fundamental del problema homogéneo. La solución general de este problema es: $x(t) = \Phi(t)C$, donde C es un vector constante. Supongamos que C depende de t y veamos que condiciones debe satisfacer para que $\Phi(t)C(t)$ sea solución del sistema completo.

Sustituyendo en la ecuación:

$$x'(t) = \Phi(t)'C(t) + \Phi(t)C'(t) = A(t)\Phi(t)C(t) + b(t)$$

y como $\Phi'(t) = A(t)\Phi(t)$, se tiene:

$$\Phi(t)C'(t) = b(t)$$

es decir:

$$C'(t) = \Phi(t)^{-1}b(t)$$

de donde se puede deducir $C(t)$ por una integración. Si la solución que buscamos verifica $x(t_0) = x_0$, se puede escribir:

$$x(t) = \Phi(t)\Phi(t_0)^{-1}x_0 + \Phi(t) \int_{t_0}^t \Phi(s)^{-1}b(s)ds.$$

Sin embargo, el cálculo de $\Phi(t)$ puede ser muy complicado. Tan solo cuando los coeficientes de A no dependen de t puede calcularse una matriz fundamental fácilmente. Si los coeficientes de la matriz A son funciones periódicas de t , la teoría de Floquet proporciona indicaciones sobre la forma de las soluciones y permite en algunos casos simples calcularlas.

2.3 Sistemas lineales con coeficientes constantes

Sea el sistema lineal homogéneo con matriz A constante (sistema lineal autónomo), $x' = Ax$. De acuerdo con lo expuesto hasta ahora, existe una única solución, fijada la condición inicial, que además está definida en todo el eje real. Este tipo de sistemas aparece en muchas situaciones: en problemas de redes eléctricas o sistemas mecánicos, como aproximación de sistemas no lineales en el entorno de un punto crítico, en grupos de transformaciones uniparamétricos (transformaciones infinitesimales) etc. Como hemos dicho

antes, es posible hallar una matriz fundamental de este sistema y, como vamos a ver ahora, todo se reduce al cálculo de la exponencial de una matriz.

Recordemos la definición y propiedades elementales de la exponencial de una matriz. Sea A una matriz real (o compleja) $n \times n$. Se define $\exp(A)$:

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = I + A + \frac{A^2}{2} + \dots$$

En el espacio de matrices reales $n \times n$, $M_n(\mathbb{R})$, se puede definir una norma que lo convierte en un espacio de Banach:

$$\|A\|_s = \sup\left\{\frac{\|Ax\|}{\|x\|}, x \in \mathbb{R}^n, x \neq 0\right\}$$

donde $\|\cdot\|$ es una norma definida en \mathbb{R}^n . Usando esta estructura se puede probar que la serie definida anteriormente converge absolutamente para toda matriz de $M_n(\mathbb{R})$. Para ello se usa la desigualdad:

$$\|A^k\|_s \leq \|A\|_s^k$$

con lo cual:

$$\left\|\sum_{k=n}^m \frac{A^k}{k!}\right\|_s \leq \sum_{k=n}^m \frac{\|A\|_s^k}{k!}$$

Como la serie exponencial es absolutamente convergente en \mathbb{R} , se tiene lo mismo para la exponencial de una matriz.

La exponencial de una matriz tiene propiedades similares a la exponencial de números reales o complejos, salvo aquellas en las que interviene la conmutatividad de la multiplicación.

Entre las propiedades de la exponencial que nos resultarán más útiles señalemos las siguientes:

1. $(e^A)^{-1} = e^{-A}$.
2. $\det(e^A) = e^{\text{tr } A}$, y por tanto $\det(e^A) \neq 0$.

Consideremos ahora funciones de un intervalo de la recta real, I , en $M_n(\mathbb{R})$:

$$f: I \rightarrow M_n(\mathbb{R}).$$

Dado que el espacio $M_n(\mathbb{R})$ es un espacio de Banach, se pueden estudiar series de funciones, continuidad, derivabilidad, etc. En lo que se refiere a nuestro interés, el criterio de convergencia de Weierstrass es válido:

Proposición 2.3.1 *Si la serie de funciones $\sum_{n \geq 0} f_n(t)$, $f_n: I \rightarrow M$, donde M es un espacio de Banach, está mayorada por una serie numérica convergente, entonces converge absoluta y uniformemente en t .*

Además, si la serie es convergente, y sus términos son derivables, la serie de las derivadas converge a la derivada de la serie, si la primera es uniformemente convergente. En nuestro caso se dan esas circunstancias como veremos después.

2.3.1 Cálculo de exponenciales de matrices

De acuerdo con la definición, para calcular la exponencial de una matriz basta hallar sus potencias sucesivas. Sin embargo, en la práctica no es éste un método muy eficaz. Como veremos a continuación, debido a las propiedades de las exponenciales de matrices, basta saber calcular la exponencial de una matriz nilpotente y de múltiplos de la identidad, casos por otra parte muy sencillos de tratar.

Las siguientes propiedades permiten el cálculo de una manera sistemática.

1. Si $A = PJP^{-1}$, donde $A, J, P \in M_n(\mathbb{C})$, y P es regular, entonces:

$$e^A = Pe^J P^{-1}$$

como se puede demostrar fácilmente sin más que utilizar la definición y:

$$(PJP^{-1})^k = PJ^k P^{-1}$$

2. Si J es una matriz diagonal por bloques, la exponencial de J es otra matriz diagonal por bloques, que son las exponenciales de los bloques de J . La demostración es también evidente sin más que recordar como se multiplican matrices por bloques:

$$\exp \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix} = \begin{pmatrix} \exp(J_1) & & \\ & \ddots & \\ & & \exp(J_k) \end{pmatrix}$$

3. Si $A, B \in M_n(\mathbb{C})$ conmutan entonces:

$$e^{A+B} = e^A e^B$$

La demostración sigue los mismos pasos que en el caso real o complejo. La propiedad no es cierta cuando las matrices A y B no conmutan.

4. Si N es una matriz nilpotente de grado r , su exponencial es una suma finita:

$$e^N = I + N + \frac{N^2}{2} + \cdots + \frac{N^{r-1}}{(r-1)!}$$

puesto que las potencias de N de exponente superior o igual a r son cero.

Como toda matriz $A \in M_n(\mathbb{C})$ se puede poner en forma de Jordan, que es una matriz diagonal por bloques, y cada uno de sus bloques es suma de una matriz múltiplo de la identidad y de una matriz nilpotente (que obviamente conmutan), aplicando las propiedades anteriores uno puede calcular fácilmente la exponencial de cualquier matriz. En la práctica sin embargo, existen otros métodos que permiten obtener este resultado sin pasar por la forma de Jordan que puede ser complicada de encontrar para matrices de orden elevado, sin contar con el cálculo de la matriz de cambio de base.

Veamos ahora otra forma de enfocar este cálculo. Sea $A \in M_n(\mathbb{C})$, y $\sigma(A)$ su espectro. Dada una función $f: \mathbb{C} \rightarrow \mathbb{C}$, queremos definir $f(A)$. Supongamos que f es un polinomio $p(z)$ y sea $q(z)$ otro polinomio tal que: $p(A) = q(A)$. Entonces, $d(z) = p(z) - q(z)$ anula a A , es decir es múltiplo del polinomio mínimo de A , por lo tanto entre sus raíces están las de este polinomio mínimo, los autovalores de A , con multiplicidades en general mayores. En cualquier caso se tiene, para cualquier $\lambda \in \sigma(A)$:

$$d(\lambda) = 0, d'(\lambda) = 0, \dots, d^{(m-1)}(\lambda) = 0$$

donde m es la multiplicidad de λ como raíz del polinomio mínimo. Si f es una función arbitraria,

$$f(\lambda), f'(\lambda), \dots, f^{(m-1)}(\lambda)$$

para todo autovalor λ de A , son los valores de la función f sobre el espectro de A . En el caso anterior de los polinomios p y q , ambos toman los mismos valores sobre el espectro de A como hemos visto. Y viceversa, si dos polinomios p y q toman los mismos valores sobre el espectro de una matriz A , entonces $p(A) = q(A)$, pues la diferencia de p y q es un múltiplo del polinomio mínimo que anula a la matriz A .

Se postula para una función f suficientemente regular el mismo resultado, es decir $f(A)$ queda fijada por los valores que toma f sobre el espectro de A (en el sentido antes dicho). Se puede definir entonces $f(A)$ de la forma siguiente:

Definición 2.3.1 Sea f una función suficientemente regular, $A \in M_n(\mathbb{C})$. Se define $f(A)$ como:

$$f(A) = p(A)$$

donde $p(z)$ es un polinomio que toma los mismos valores que f sobre el espectro de A .

Evidentemente, $f(A)$ no depende del polinomio elegido por lo que hemos dicho antes, así que se puede tomar el de menor grado, que es menor que el del polinomio mínimo, y además es único. Nótese que este polinomio de menor grado es un polinomio de interpolación y se determina por:

$$p(\lambda) = f(\lambda), p'(\lambda) = f'(\lambda), \dots, p^{(m-1)}(\lambda) = f^{(m-1)}(\lambda)$$

para todo elemento del espectro de A , m multiplicidad de λ como raíz del polinomio mínimo. Estas son pues las ecuaciones que permiten calcular $p(z)$ y consiguientemente $p(A) = f(A)$. Se llama a $p(z)$ el polinomio de interpolación de Lagrange-Sylvester. Este método de cálculo está muy ligado al de proyectores ortogonales.

Ejemplo 2.3.1 Consideremos la matriz:

$$A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

que está en forma de Jordan. Por tanto su exponencial es:

$$e^A = e^{-I} e^N$$

donde I es la matriz identidad 2×2 y N es una matriz nilpotente:

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

cuya exponencial es inmediata de calcular:

$$e^N = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

pues $N^2 = 0$, y por tanto $e^N = I + N$. Por tanto, la exponencial de A es:

$$e^A = \begin{pmatrix} e^{-1} & 0 \\ 0 & e^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-1} & e^{-1} \\ 0 & e^{-1} \end{pmatrix}$$

Supongamos que queremos calcular el polinomio interpolador de Lagrange-Sylvester para esta matriz. Para ello buscamos un polinomio de grado 1 (el grado del polinomio mínimo es 2) que verifique:

$$p(-1) = e^{-1}, \quad p'(-1) = e^{-1}$$

ya que la multiplicidad de -1 como raíz del polinomio mínimo es 2. Sea $p(z) = az + b$. Las ecuaciones que determinan a, b son:

$$\begin{aligned} -a + b &= e^{-1} \\ a &= e^{-1} \end{aligned}$$

por tanto, $a = e^{-1}$, $b = 2e^{-1}$, y se obtiene el resultado anterior al poner:

$$e^A = p(A) = e^{-1}A + 2e^{-1}I.$$

Ejemplo 2.3.2 Sea la matriz:

$$A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}$$

En este caso, la matriz es diagonalizable con autovalores 2 y -1 . La matriz de cambio de base no es difícil de calcular, siendo los autovectores $(1, 0)$ y $(1, -3)$:

$$A = PJP^{-1}$$

donde:

$$P = \begin{pmatrix} 1 & 1 \\ 0 & -3 \end{pmatrix}$$

$$J = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

La exponencial de J se calcula inmediatamente con lo que la exponencial de A es:

$$e^A = \begin{pmatrix} e^2 & (e^2 - e^{-1})/3 \\ 0 & e^{-1} \end{pmatrix}.$$

Supongamos que queremos calcular el polinomio de interpolación de Lagrange–Sylvester de A , $p(z)$. Ahora tenemos dos autovalores con multiplicidad igual a 1. Por tanto el grado de $p(z)$ es 1 (el grado del polinomio mínimo es 2) y las fórmulas que determinan los coeficientes son:

$$p(z) = az + b, \quad p(2) = e^2, \quad p(-1) = e^{-1} \Rightarrow 2a + b = e^2, \quad -a + b = e^{-1}$$

es decir:

$$a = \frac{e^2 - e^{-1}}{3}, \quad b = \frac{e^2 + 2e^{-1}}{3}$$

de esta forma:

$$e^A = p(A) = aA + bI$$

y se obtiene el resultado anterior.

Ejemplo 2.3.3 Finalmente veamos un caso con matrices 3×3 .

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}.$$

Esta matriz tiene un solo autovalor, 2, y dos cajas de Jordan de dimensiones 2 y 1. En efecto el polinomio característico es:

$$(\lambda - 2)^3$$

y el polinomio mínimo es $(\lambda - 2)^2$, con lo cual tenemos dos autovectores linealmente independientes. La matriz de Jordan es:

$$J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

y la matriz de cambio de base P puede calcularse de la siguiente forma. Como $A = PJP^{-1}$, podemos escribir $P = (P_1, P_2, P_3)$, como vectores columnas y esa ecuación es:

$$\begin{aligned} AP_1 &= 2P_1 \\ AP_2 &= P_1 + 2P_2 \\ AP_3 &= 2P_3 \end{aligned}$$

donde se ve que P_1, P_3 son los autovectores de la matriz A . Estos autovectores son combinaciones lineales de $(0, 0, 1)$ y $(1, 2, 0)$. Sin embargo resulta más cómodo calcular en primer lugar el vector P_2 . Para ello escribamos las dos primeras ecuaciones como:

$$(A - 2I)P_1 = 0$$

$$(A - 2I)P_2 = P_1$$

y multiplicando la segunda ecuación por $A - 2I$ tenemos:

$$(A - 2I)^2 P_2 = 0$$

ecuación que permite calcular P_2 . En este caso resulta ser trivial, pues el polinomio mínimo es $(\lambda - 2)^2$ y anula a la matriz A . La única restricción es que P_2 no debe ser un autovector, así que elijamos $P_2 = (1, 0, 0)$. De esta forma P_1 queda fijado por la segunda ecuación:

$$P_1 = (A - 2I)P_2$$

es decir $P_1 = (-2, -4, -2)$. El tercer vector, que también es autovector, se elige linealmente independiente con P_1 , por ejemplo, $P_3 = (0, 0, 1)$. Por tanto la matriz P es:

$$P = \begin{pmatrix} -2 & 1 & 0 \\ -4 & 0 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

y la exponencial de A :

$$e^A = P \begin{pmatrix} e^{2I+N} & 0 \\ 0 & e^2 \end{pmatrix} P^{-1}$$

donde I es la matriz identidad 2×2 y N es una matriz nilpotente,

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Por tanto, la exponencial de la matriz A es:

$$e^A = e^2 \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ -2 & 1 & 1 \end{pmatrix}.$$

un cálculo laborioso. Sin embargo, construyamos el polinomio de Lagrange-Sylvester de la matriz A . Esta matriz tiene un solo autovalor y el polinomio característico tiene grado 2. Por lo tanto $p(z)$ será un polinomio de grado 1, $p(z) = az + b$, y los coeficientes a, b se calculan por:

$$p(2) = e^2, \quad p'(2) = e^2$$

con lo que $a = e^2$, $b = -e^2$.

La exponencial de la matriz A es:

$$e^A = p(A) = e^2 A - e^2 I$$

que es el resultado obtenido anteriormente.

Cuando los coeficientes son complejos los métodos de cálculo son los mismos. Sin embargo si la matriz es real, aunque los autovalores sean complejos, la exponencial es una matriz real. En este caso los autovalores complejos aparecen a pares, y siempre se puede encontrar una base en la cuál la matriz se descomponga en bloques elementales 2×2 asociados a cada pareja de estos autovalores si son simples. Si no lo son, es posible que la matriz no sea diagonalizable por esos bloques pero en cualquier caso existirá una forma similar a la de Jordan y a la postre los cálculos de la exponencial se reducirán a los de la exponencial de los bloques 2×2 . Consideremos una matriz 2×2 con coeficientes reales y autovalores complejos:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$(\operatorname{tr} A)^2 < 4 \det A$$

Sean $\lambda = \mu + i\nu$, los autovalores de esta matriz, con $\nu > 0$. Se pueden calcular los autovectores complejos $Aw = \lambda w$, $w = u + iv$, donde u, v son vectores reales linealmente independientes. La matriz A se escribe en la base $\{v, u\}$ como:

$$\begin{pmatrix} \mu & -\nu \\ \nu & \mu \end{pmatrix}$$

y la exponencial de esta matriz se calcula fácilmente sabiendo que existe un isomorfismo entre el cuerpo de los números complejos y el conjunto de matrices reales con esta forma. La exponencial de $\lambda = \mu + i\nu$ es:

$$e^{\mu+i\nu} = e^{\mu}(\cos \nu + i \operatorname{sen} \nu)$$

por lo tanto, la exponencial es:

$$e^{\mu} \begin{pmatrix} \cos \nu & -\operatorname{sen} \nu \\ \operatorname{sen} \nu & \cos \nu \end{pmatrix}.$$

Ejemplo 2.3.4

$$A = \begin{pmatrix} 3 & 1 \\ -5 & -1 \end{pmatrix}.$$

Los autovalores de esta matriz son $1 \pm i$. Los autovectores complejos son: $(1, -2 \pm i)$, es decir $(1, -2) \pm i(0, 1)$. En la base $\{(0, 1), (1, -2)\}$ la expresión de A es:

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

por lo tanto:

$$\begin{aligned} e^A &= \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} e \cos 1 & -e \operatorname{sen} 1 \\ e \operatorname{sen} 1 & e \cos 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} e(2 \operatorname{sen} 1 + \cos 1) & e \operatorname{sen} 1 \\ -5e \operatorname{sen} 1 & e(\cos 1 - 2 \operatorname{sen} 1) \end{pmatrix} \end{aligned}$$

Calculando el polinomio interpolador, la matriz A tiene dos autovalores distintos y el polinomio mínimo tiene grado igual a 2, así que $p(z)$ tiene grado 1, $p(z) = az + b$, y los coeficientes se calculan con las ecuaciones:

$$p(1+i) = e^{1+i}, \quad p(1-i) = e^{1-i}$$

es decir:

$$a(1+i) + b = e(\cos 1 + i \operatorname{sen} 1)$$

$$a(1-i) + b = e(\cos 1 - i \operatorname{sen} 1)$$

de donde $a = e \operatorname{sen} 1$, $b = e(\cos 1 - \operatorname{sen} 1)$, y por tanto la exponencial de A es:

$$e^A = Ae \operatorname{sen} 1 + Ie(\cos 1 - \operatorname{sen} 1).$$

Como se ha visto en este resumen, el cálculo de exponenciales no encierra ningún secreto y es simplemente un problema técnico. Además, como veremos a continuación, uno puede obviar en muchas ocasiones el cálculo completo de la exponencial y limitarse a calcular autovectores.

2.4 Solución general

De acuerdo con lo dicho hasta ahora, la función que asigna a cada t la exponencial de At , donde $A \in M_n(\mathbb{R})$ está bien definida y su derivada se puede calcular derivando término a término la serie que la define. Por tanto:

$$\frac{d}{dt} e^{At} = \frac{d}{dt} \sum_{n=0}^{\infty} \frac{A^n}{n!} t^n = \sum_{n=0}^{\infty} \frac{d}{dt} \left(\frac{A^n}{n!} t^n \right) = \sum_{n=0}^{\infty} \frac{A^n}{n!} n t^{n-1}$$

es decir, se tiene la relación:

$$\frac{d}{dt} e^{At} = A e^{At}$$

que nos permite demostrar lo siguiente:

Proposición 2.4.1 Una matriz fundamental del sistema homogéneo con coeficientes constantes:

$$x' = Ax$$

es $\Phi(t) = e^{At}$, que además es la única que verifica $\Phi(0) = I$.

La demostración es evidente. La solución particular que verifica $x(t_0) = x_0$ es:

$$x(t) = e^{A(t-t_0)}x_0.$$

Téngase en cuenta además que e^{At} es un grupo uniparamétrico con $t \in \mathbb{R}$, debido a las propiedades de la exponencial ya estudiadas y al hecho de que los múltiplos de A conmutan entre sí. En cuanto al sistema completo, una vez que conocemos una matriz fundamental, es inmediato escribir la solución del problema:

$$\begin{cases} x' &= Ax + b(t) \\ x(t_0) &= x_0 \end{cases}$$

como:

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^t e^{A(t-s)}b(s)ds.$$

Como hemos dicho anteriormente, sólo se necesita conocer una matriz fundamental para calcular todas las soluciones. Si escribimos A en su forma de Jordan: $A = PJP^{-1}$:

$$e^{At} = Pe^{Jt}P^{-1}$$

pero como sabemos, dada una matriz fundamental, el producto de dicha matriz por una matriz de constantes regular (a la derecha) proporciona una nueva matriz fundamental. Por tanto, para escribir la solución general del problema homogéneo podemos utilizar:

$$\Psi(t) = Pe^{Jt}$$

que es también una matriz fundamental, aunque no sea la canónica en $t = 0$. Pero es más, como ahora veremos, los autovalores de A proporcionan una base del espacio de soluciones en el caso que A sea diagonalizable, y en cualquier caso llevan a soluciones del sistema.

En efecto, si A es diagonalizable, J es una matriz diagonal y P es una matriz cuyas columnas son los autovectores del sistema:

$$\Psi(t) = (P_1, \dots, P_n) \begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{pmatrix} = (P_1 e^{\lambda_1 t}, \dots, P_n e^{\lambda_n t})$$

y por tanto una base del espacio de soluciones es:

$$\{P_1 \exp(\lambda_1 t), \dots, P_n \exp(\lambda_n t)\}$$

como por otra parte se puede comprobar directamente. La solución general es una combinación lineal de estas funciones:

$$x(t) = c_1 P_1 e^{\lambda_1 t} + \dots + c_n P_n e^{\lambda_n t}$$

aunque el cálculo de las constantes c_i implica resolver un sistema de ecuaciones lineales. Este resultado es cierto aún cuando algunos autovalores tengan multiplicidad mayor que 1. Lo esencial es que la matriz sea diagonalizable.

Cuando la matriz no es diagonalizable la situación no es tan clara. La matriz J no es diagonal y la matriz P contiene en sus columnas los autovectores de A y otros vectores necesarios para formar una base. Supongamos que P_λ es un autovector con autovalor λ . Entonces $P_\lambda e^{\lambda t}$ es una solución del sistema:

$$x'(t) = \lambda P_\lambda e^{\lambda t} = A P_\lambda e^{\lambda t}.$$

Supongamos ahora que existe una caja de Jordan de dimensión r y que P_1, \dots, P_r son los vectores de la base correspondientes a esa caja. En esta caso las soluciones son polinomios por exponenciales:

$$P_1 e^{\lambda t}, (P_1 t + P_2) e^{\lambda t}, ((1/2)P_1 t^2 + P_2 t + P_3) e^{\lambda t}, \dots, \\ ((1/(r-1)!)P_1 t^{r-1} + \dots + P_r) e^{\lambda t}$$

como se comprueba fácilmente sin más que calcular la exponencial de la caja de Jordan:

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}.$$

Ejemplo 2.4.1

$$\begin{cases} x_1' &= -x_1 + x_2 \\ x_2' &= -x_2 \end{cases}$$

La matriz es no diagonalizable. Un autovector es $(1, 0)$ y el vector que completa la base para obtener la forma de Jordan es $(0, 1)$, pues la matriz está ya escrita en la forma canónica. Por lo tanto la solución es:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-t} + c_2 \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) e^{-t}.$$

Ejemplo 2.4.2 La ecuación que describe el movimiento de caída libre de un cuerpo hacia la tierra es:

$$\frac{dv}{dt} = g - 2\omega \times v$$

donde v es la velocidad del cuerpo, g la aceleración de la gravedad y ω la velocidad de rotación de la tierra. Se trata de un sistema de ecuaciones lineales de primer orden con coeficientes constantes. La matriz del sistema es:

$$Av = -2\omega \times v$$

con lo que el sistema se escribe:

$$v' = Av + g.$$

Calculemos la exponencial de la matriz A :

$$-2 \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}$$

sus autovalores son $0, \pm 2i\Omega$, donde Ω es el módulo del vector ω . Calculando el polinomio de interpolación de la matriz At , que debe ser de grado 2, $p(z) = az^2 + bz + c$:

$$p(0) = 1, \quad p(2i\Omega t) = e^{2i\Omega t}, \quad p(-2i\Omega t) = e^{-2i\Omega t}$$

lo que proporciona a, b, c :

$$a = \frac{1 - \cos 2\Omega t}{4\Omega^2 t^2}, \quad b = \frac{\sin 2\Omega t}{2\Omega t}, \quad c = 1$$

y por lo tanto la exponencial es:

$$e^{At} = I + btA + at^2 A^2.$$

La velocidad de caída es:

$$v(t) = e^{At} v_0 + (e^{At} \int_0^t e^{-As} g = \\ v_0 + (I + \frac{1}{4\Omega^2} A^2)gt + A(v_0 - \frac{1}{4\Omega^2} Ag)bt + A(g + Av_0)at^2$$

y escribiendo la expresión de los coeficientes y la de A :

$$\begin{aligned} v(t) = & v_0 + t(g + \frac{\omega}{\Omega} \times \frac{\omega}{\Omega} \times g) + (1 - \cos 2\Omega t) \frac{\omega}{\Omega} \times (\frac{\omega}{\Omega} \times v_0 - \frac{g}{2\Omega}) \\ & - (\sin 2\Omega t) \frac{\omega}{\Omega} \times (\frac{\omega}{\Omega} \times \frac{g}{2\Omega} + v_0) \end{aligned}$$

De aquí podemos deducir la trayectoria sin más que integrar esta ecuación en t .

2.5 Ecuaciones lineales de orden n

Una ecuación lineal de orden n se puede escribir como:

$$x^{(n)} + a_{n-1}(t)x^{(n-1)} + \dots + a_1(t)x' + a_0(t)x = b(t)$$

donde los coeficientes $a_i(t)$, $b(t)$ son funciones continuas definidas en un intervalo de la recta real $[t_0, t_1]$ con valores en \mathbb{R} . Es inmediato probar que una condición suficiente para la existencia y unicidad de soluciones de esta ecuación es que los coeficientes sean continuos en el intervalo considerado. La ecuación se puede escribir como un sistema, de acuerdo con lo dicho anteriormente y se tiene:

$$\begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_{n-1}' \\ x_n' \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ -a_0(t) & -a_1(t) & \cdots & -a_{n-1}(t) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}.$$

Como se puede observar, la forma de la matriz de este sistema es muy particular, y, en general, no es necesario pasar al sistema asociado para resolver la ecuación. En cualquier caso, cuando los coeficientes no son constantes no es sencillo calcular las soluciones de este tipo de ecuaciones. Las propiedades de las soluciones se pueden deducir de las demostradas para los sistemas de ecuaciones lineales:

Proposición 2.5.1 *El conjunto de soluciones de una ecuación lineal homogénea de orden n es un espacio vectorial real de dimensión igual a n .*

Así, si el conjunto de funciones $\{x_1(t), \dots, x_n(t)\}$ es una base, la solución general de la ecuación homogénea es:

$$x(t) = c_1x_1(t) + c_2x_2(t) + \dots + c_nx_n(t)$$

donde c_1, c_2, \dots, c_n son constantes que se fijan de acuerdo con las condiciones iniciales. Debido a la independencia lineal de las soluciones, el sistema que fija las constantes c_i en función de los datos iniciales tiene solución única. Esto es consecuencia de las propiedades del wronskiano que pasamos a enunciar.

Definición 2.5.1 *Se llama wronskiano de las funciones $f_1(t), \dots, f_n(t)$ definidas en un intervalo de la recta real y derivables al menos hasta orden $n - 1$ al siguiente determinante:*

$$W[f_1, \dots, f_n; t] = \det \begin{pmatrix} f_1(t) & \cdots & f_n(t) \\ f_1'(t) & \cdots & f_n'(t) \\ \vdots & & \vdots \\ f_1^{(n-1)}(t) & \cdots & f_n^{(n-1)}(t) \end{pmatrix}.$$

Si consideramos n soluciones de la ecuación lineal homogénea, el wronskiano es el determinante de una matriz de soluciones del sistema lineal asociado. Por lo tanto verifica la ecuación de Liouville donde ahora, teniendo en cuenta la forma especial de la matriz $A(t)$, se tiene $\text{tr } A(t) = -a_{n-1}(t)$, y:

$$W' = -a_{n-1}(t)W$$

ecuación que tiene por solución:

$$W(t) = Ce^{-\int a_{n-1}(s)ds}.$$

El wronskiano tiene las mismas propiedades que en el caso de sistemas, es decir no se anula en ningún punto del intervalo, o es idénticamente cero en todo el intervalo. En el primer caso las soluciones de la ecuación son linealmente independientes y en el segundo son linealmente dependientes. Por tanto las constantes que aparecen en la solución general se fijan unívocamente en función de los datos iniciales.

La solución general de la ecuación completa es suma de la solución general de la ecuación homogénea más una solución particular de la completa. El cálculo de una solución particular se puede hacer usando el método de variación de constantes que sigue las mismas líneas que el expuesto para sistemas lineales. Sin embargo, aquí sólo hay una ecuación y por tanto es necesario imponer condiciones suplementarias. Sea entonces $x(t)$ la siguiente función, donde $\{x_1(t), \dots, x_n(t)\}$ es una base de soluciones de la ecuación homogénea:

$$x(t) = \sum_{i=1}^n c_i(t)x_i(t)$$

y calculemos las derivadas sucesivas de esta función:

$$x'(t) = \sum_{i=1}^n c'_i(t)x_i(t) + \sum_{i=1}^n c_i(t)x'_i(t).$$

Si imponemos la condición:

$$\sum_{i=1}^n c'_i(t)x_i(t) = 0$$

la derivada segunda se calcula fácilmente:

$$x''(t) = \sum_{i=1}^n c'_i(t)x'_i(t) + \sum_{i=1}^n c_i(t)x''_i(t)$$

y como antes, imponemos otra condición sobre las funciones $c_i(t)$:

$$\sum_{i=1}^n c'_i(t)x'_i(t) = 0$$

De esta forma, la derivada de orden $n-1$ será:

$$x^{(n-1)}(t) = \sum_{i=1}^n c'_i(t)x_i^{(n-2)}(t) + \sum_{i=1}^n c_i(t)x_i^{(n-1)}(t)$$

que nos proporciona una nueva condición para las $c_i(t)$:

$$\sum_{i=1}^n c'_i(t)x_i^{(n-2)}(t) = 0$$

finalmente, se calcula la derivada de orden n y se sustituye en la ecuación con lo que se obtienen las n condiciones que fijan las funciones $c_i(t)$:

$$x^{(n)}(t) = \sum_{i=1}^n c'_i(t)x_i^{(n-1)}(t) + \sum_{i=1}^n c_i(t)x_i^{(n)}(t).$$

Al sustituir en la ecuación aparece una relación entre las $c'_i(t)$ y las derivadas de orden n igualada a la función $b(t)$ más una relación entre $c_i(t)$ y las derivadas de las funciones $x_i(t)$, que no es más que la ecuación homogénea aplicada a cada una de estas últimas funciones. Como son soluciones de la ecuación homogénea, se concluye que:

$$\sum_{i=1}^n c'_i(t)x_i^{(n-1)}(t) = b(t).$$

Esta ecuación, junto con las anteriores forma un sistema de ecuaciones algebraicas lineales que permite calcular las funciones $c'_i(t)$. La matriz de coeficientes es justamente el wronskiano de la base de soluciones con las que se está trabajando, lo que asegura la existencia de solución. Una vez halladas $c'_i(t)$ se integran y multiplican por $x_i(t)$ para obtener la solución particular buscada. Veamos en el caso $n = 2$ como es esta solución.

Sea la ecuación $x'' + a_1(t)x' + a_0(t)x = b(t)$, $x_1(t)$, $x_2(t)$ una base de soluciones de la ecuación homogénea. La solución particular buscada por el método de variación de constantes es:

$$x(t) = c_1(t)x_1(t) + c_2(t)x_2(t)$$

y las ecuaciones que satisfacen las funciones $c_1(t)$, $c_2(t)$ son:

$$\begin{cases} c'_1(t)x_1(t) + c'_2(t)x_2(t) &= 0 \\ c'_1(t)x'_1(t) + c'_2(t)x'_2(t) &= b(t) \end{cases}$$

que se pueden resolver por Cramer, obteniéndose:

$$c'_1(t) = -\frac{x_2(t)b(t)}{W(t)}, \quad c'_2(t) = \frac{x_1(t)b(t)}{W(t)}$$

donde $W(t)$ es el wronskiano de las soluciones $x_1(t)$, $x_2(t)$:

$$W(t) \equiv W[x_1, x_2; t] = \det \begin{pmatrix} x_1(t) & x_2(t) \\ x'_1(t) & x'_2(t) \end{pmatrix}.$$

De esta forma la solución general del problema completo, suponiendo un punto de partida en t_0 es:

$$x(t) = a_1x_1(t) + a_2x_2(t) - x_1(t) \int_{t_0}^t \frac{x_2(s)b(s)}{W(s)} ds + x_2(t) \int_{t_0}^t \frac{x_1(s)b(s)}{W(s)} ds.$$

Si la solución particular buscada verifica $x(t_0) = x_{01}$, $x'(t_0) = x_{02}$, se tiene:

$$\begin{cases} x_{01} &= a_1x_1(t_0) + a_2x_2(t_0) \\ x_{02} &= a_1x'_1(t_0) + a_2x'_2(t_0) \end{cases}$$

como se puede comprobar fácilmente. De este sistema uno puede obtener las constantes a_1 , a_2 , pues la matriz de coeficientes es el wronskiano de $x_1(t)$, $x_2(t)$ que es distinto de cero en cualquier punto del intervalo donde se trabaja. Téngase en cuenta que siempre suponemos que los coeficientes de la ecuación son continuos en un cierto intervalo.

Ejemplo 2.5.1 Supongamos un oscilador armónico sometido a una fuerza periódica:

$$x'' + 4x = \cos 2t.$$

Como veremos en el próximo apartado, una base de soluciones del problema homogéneo está formada por las funciones $\cos 2t$, $\sin 2t$. El wronskiano de estas funciones es:

$$W(t) = \det \begin{pmatrix} \cos 2t & \sin 2t \\ -2 \sin 2t & 2 \cos 2t \end{pmatrix} = 2$$

y por lo tanto la solución particular que necesitamos es:

$$x_p(t) = -\frac{1}{2} \cos 2t \int_0^t \sin 2s \cos 2s ds + \frac{1}{2} \sin 2t \int_0^t \cos^2 2s ds$$

es decir:

$$x_p(t) = \frac{1}{2} t \sin 2t$$

La solución general es suma de la solución general de la homogénea, generada por la base $x_1(t)$, $x_2(t)$, con esta solución particular.

No es fácil calcular una base de soluciones para una ecuación lineal de orden n . Sin embargo es posible demostrar que si se conoce una solución, el orden de la ecuación se puede rebajar en una unidad. Consideremos el caso de segundo orden pues en general el argumento sigue las mismas líneas:

$$x'' + a_1(t)x' + a_0(t)x = 0$$

Sea $y(t)$ una solución de esta ecuación. Definiendo una nueva variable $v(t)$ por la relación:

$$x(t) = y(t) \int v$$

al sustituir en la ecuación se tiene:

$$[y'' + a_1(t)y' + a_0(t)y] \int v + y(t)v' + (2y'(t) + a_1(t)y(t))v = 0$$

como $y(t)$ es solución de la ecuación de partida, se tiene para $v(t)$ una ecuación lineal de primer orden:

$$y(t)v' + (2y'(t) + a_1(t)y(t))v = 0$$

Su solución es:

$$v(t) = \frac{C}{y^2(t)} e^{-\int a_1(t)}.$$

Una vez calculada $v(t)$ se sustituye en la expresión que proporciona $x(t)$, con lo que obtenemos otra solución de la ecuación, linealmente independiente con $y(t)$. En efecto, el wronskiano de las soluciones $y(t)$, $x(t)$ es (si $u(t)$ es la integral de $v(t)$):

$$W(t) = \det \begin{pmatrix} y(t) & y(t)u(t) \\ y'(t) & y'(t)u(t) + y(t)v(t) \end{pmatrix} = y^2(t)v(t) = e^{-\int a_1(t)}$$

(tomando $C = 1$) que no se anula en ningún punto.

En este caso se ha calculado una base. Para una ecuación de orden n harían falta $n - 1$ soluciones para reducirla a una de primer orden.

2.6 Ecuaciones con coeficientes constantes

Como hemos visto en el apartado anterior basta conocer una base de soluciones para calcular cualquier solución de una ecuación lineal homogénea o no. Sin embargo el cálculo de la base es en general complicado y salvo métodos como el de desarrollo de las soluciones en forma de serie, no se puede dar ningún esquema completo para cualquier ecuación. Existen algunos casos donde el cálculo de una base se puede hacer de forma sistemática. Concretamente en el caso de coeficientes constantes, este cálculo es una cuestión puramente algebraica. De hecho, pasando al sistema lineal asociado obtendríamos una matriz de constantes y por lo tanto sabemos que la solución se calcula hallando la exponencial de esa matriz. Sin embargo, dado que la matriz de los sistemas asociados a ecuaciones de orden n tiene una estructura muy peculiar, resulta interesante desarrollar la teoría sin referirse a los sistemas. Consideremos una ecuación lineal de orden n homogénea de coeficientes constantes:

$$x^{(n)} + a_{n-1}x^{(n-1)} + \dots + a_1x' + a_0x = 0$$

Definición 2.6.1 Se llama polinomio característico de esta ecuación a:

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0.$$

Suponemos que los coeficientes de la matriz son reales. El caso complejo no ofrece ninguna dificultad adicional. Los elementos básicos de esta ecuación diferencial son las raíces del polinomio característico.

No es difícil comprobar que el polinomio característico de la matriz A :

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{pmatrix}$$

es justamente $p(\lambda)$ (salvo un signo global), es decir, las raíces del polinomio característico de la ecuación son los autovalores de la matriz del sistema asociado. La siguiente proposición usa ese resultado, aunque la demostraremos desde el punto de vista de ecuaciones.

Proposición 2.6.1 *Sea λ raíz del polinomio característico de una ecuación diferencial de orden n con coeficientes constantes. Entonces, la función:*

$$x(t) = e^{\lambda t}$$

es una solución de la ecuación.

Demostración. Es muy simple pero aprovecharemos para introducir la notación de operadores que simplificará los resultados. Sea $D = d/dt$, y $L = D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0$, operador lineal diferencial de orden n con coeficientes constantes. Entonces, $p(D) = L$. La acción de D sobre las exponenciales es simplemente:

$$D(e^{\lambda t}) = \lambda e^{\lambda t}$$

y por lo tanto resulta inmediato comprobar que:

$$p(D)e^{\lambda t} = p(\lambda)e^{\lambda t}$$

de esta forma, si λ es una raíz del polinomio característico, $p(\lambda) = 0$ y de acuerdo con la expresión anterior, $L(e^{\lambda t}) = 0$. QED

La siguiente proposición establece cuándo se puede obtener una base formada solo por exponenciales:

Proposición 2.6.2 *Si las raíces del polinomio característico, $\{\lambda_1, \dots, \lambda_n\}$ son todas distintas entre sí, el conjunto de funciones:*

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}$$

es una base del espacio de soluciones de la ecuación.

Demostración. Es también muy sencilla. De acuerdo con la proposición anterior, todas estas funciones son soluciones de la ecuación. Así pues, lo único que hay que demostrar es que son linealmente independientes, ya que sabemos que la dimensión del espacio de soluciones es n . Para ello construyamos el wronskiano de estas soluciones:

$$W[x_1, x_2, \dots, x_n; t] = \begin{pmatrix} e^{\lambda_1 t} & \cdots & e^{\lambda_n t} \\ \lambda_1 e^{\lambda_1 t} & \cdots & \lambda_n e^{\lambda_n t} \\ \vdots & & \vdots \\ \lambda_1^{n-1} e^{\lambda_1 t} & \cdots & \lambda_n^{n-1} e^{\lambda_n t} \end{pmatrix}$$

calculando W en $t = 0$, se obtiene un determinante de Vandermonde que es distinto de cero al ser las raíces del polinomio característico distintas. Por lo tanto las soluciones son linealmente independientes y forman una base. QED

Téngase en cuenta que la proposición equivalente para un sistema era que la matriz A fuera diagonalizable, es decir, podían existir autovalores iguales. Sin embargo, no es difícil probar que la matriz dada

anteriormente, es decir la proveniente de una ecuación de orden n , solo es diagonalizable cuando todos sus autovalores son distintos. Por tanto, en el caso de raíces distintas, la solución general es:

$$x(t) = \sum_{i=1}^n c_i e^{\lambda_i t}.$$

Como estamos considerando ecuaciones con coeficientes reales, el espacio de soluciones es un espacio vectorial real de dimensión n . Se puede demostrar que existe una base de soluciones reales. Si alguna raíz λ es compleja, aparecerá también su compleja conjugada $\bar{\lambda}$. La pareja de soluciones:

$$\{e^{\lambda t}, e^{\bar{\lambda} t}\}$$

se puede sustituir por:

$$\{e^{\mu t} \cos \nu t, e^{\mu t} \sin \nu t\}$$

donde $\lambda = \mu + i\nu$, que son combinaciones lineales de las anteriores y son reales. En el caso de raíces complejas, la solución general, cuando todas las raíces son distintas es:

$$x(t) = \sum_{i=1}^r c_i e^{\lambda_i t} + \sum_{j=1}^s e^{\mu_j t} (b_j \cos \nu_j t + d_j \sin \nu_j t)$$

donde $r + 2s = n$, λ_i , son las raíces reales y $\mu_j \pm i\nu_j$ las complejas.

En el caso de raíces con multiplicidad mayor que 1 la situación es más complicada. Si las raíces distintas son $\lambda_1, \lambda_2, \dots, \lambda_k$, con $k < n$, las funciones:

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_k t}$$

siguen siendo soluciones de la ecuación, pero al ser n mayor que k , no es una base. Se tiene el siguiente resultado:

Proposición 2.6.3 *Sea λ_0 raíz del polinomio característico con multiplicidad r . En este caso, las funciones:*

$$e^{\lambda_0 t}, t e^{\lambda_0 t}, \dots, t^{r-1} e^{\lambda_0 t}$$

son soluciones de la ecuación y linealmente independientes.

Demostración. Para demostrar este resultado, descomponemos el polinomio $p(\lambda)$ en dos factores primos entre sí:

$$p(\lambda) = q(\lambda)(\lambda - \lambda_0)^r$$

con $q(\lambda_0) \neq 0$. El operador L se descompone de la misma forma:

$$L = q(D)(D - \lambda_0)^r$$

por tanto basta probar que $(D - \lambda_0)^r$ anula a las funciones dadas. Apliquemos una vez este operador:

$$(D - \lambda_0)t^k e^{\lambda_0 t} = kt^{k-1} e^{\lambda_0 t} + \lambda_0 t^k e^{\lambda_0 t} - \lambda_0 t^k e^{\lambda_0 t} = kt^{k-1} e^{\lambda_0 t}.$$

Es inmediato verificar que al aplicar s veces el operador se obtiene:

$$(D - \lambda_0)^s t^k e^{\lambda_0 t} = k(k-1) \dots (k-s+1) t^{k-s} e^{\lambda_0 t}$$

si s es menor o igual que k . Si s es estrictamente mayor que k , el resultado es cero. Por tanto:

$$(D - \lambda_0)^r t^k e^{\lambda_0 t} = 0, \quad k = 0, 1, \dots, r-1$$

QED

Otra forma de alcanzar este resultado es la siguiente. Considerando $\exp(\lambda t)$ como una función de las variables t y λ se puede escribir:

$$D_\lambda(e^{\lambda t}) = \frac{\partial}{\partial \lambda} e^{\lambda t} = t e^{\lambda t}$$

consideremos entonces la función $t e^{\lambda t}$ y apliquemos el operador L :

$$L(t e^{\lambda t}) = L(D_\lambda e^{\lambda t}) = D_\lambda(L e^{\lambda t}) = D_\lambda(p(\lambda) e^{\lambda t}) = p'(\lambda) e^{\lambda t} + t p(\lambda) e^{\lambda t}$$

ahora bien, si λ_0 es una raíz con multiplicidad mayor que 1, $p(\lambda_0) = 0$ y $p'(\lambda_0) = 0$, por lo que $t e^{\lambda_0 t}$ es solución de L . En el caso general la situación es la misma:

$$D_\lambda^k(e^{\lambda t}) = t^k e^{\lambda t}$$

y por tanto:

$$L(t^k e^{\lambda t}) = L(D_\lambda^k e^{\lambda t}) = D_\lambda^k(L e^{\lambda t}) = D_\lambda^k(p(\lambda) e^{\lambda t})$$

obteniéndose la exponencial por un polinomio en t cuyos coeficientes son las derivadas de p respecto a λ hasta orden k . Si λ_0 es una raíz de multiplicidad r del polinomio característico, se tiene:

$$p(\lambda_0) = 0, p'(\lambda_0) = 0, \dots, p^{(r-1)}(\lambda_0) = 0$$

por tanto las funciones $t^k e^{\lambda_0 t}$ con $k = 0, 1, \dots, r-1$ son soluciones de L .

En cuanto a la independencia lineal de este conjunto de funciones, basta calcular su wronskiano. Poniendo $t = 0$ se obtiene el determinante de una matriz triangular, pues los elementos por encima de la diagonal son potencias de t por la exponencial, y en la diagonal aparecen constantes distintas de cero (factoriales). Por tanto, el wronskiano es no nulo y las soluciones son linealmente independientes.

Podemos escribir en el caso general una base del espacio de soluciones en la forma siguiente:

Proposición 2.6.4 Sean $\lambda_1, \dots, \lambda_k$ las raíces reales del polinomio característico de multiplicidades r_1, \dots, r_k , y $\mu_1 \pm i\nu_1, \dots, \mu_m \pm i\nu_m$ las raíces complejas con multiplicidades s_1, \dots, s_m . Se tiene $r_1 + \dots + r_k + 2(s_1 + \dots + s_m) = n$. Entonces, una base del espacio de soluciones de la ecuación diferencial lineal homogénea de orden n con polinomio característico $p(\lambda)$ es:

$$\begin{aligned} & e^{\lambda_1 t}, t e^{\lambda_1 t}, \dots, t^{r_1-1} e^{\lambda_1 t} \\ & \dots \\ & e^{\lambda_k t}, t e^{\lambda_k t}, \dots, t^{r_k-1} e^{\lambda_k t} \\ & e^{\mu_1 t} \cos \nu_1 t, e^{\mu_1 t} \sin \nu_1 t, \dots, t^{s_1-1} e^{\mu_1 t} \cos \nu_1 t, t^{s_1-1} e^{\mu_1 t} \sin \nu_1 t \\ & \dots \\ & e^{\mu_m t} \cos \nu_m t, e^{\mu_m t} \sin \nu_m t, \dots, t^{s_m-1} e^{\mu_m t} \cos \nu_m t, t^{s_m-1} e^{\mu_m t} \sin \nu_m t \end{aligned}$$

Ejemplo 2.6.1 Sea la ecuación anterior, $x'' + 4x = \cos 2t$. El polinomio característico es $\lambda^2 + 4 = 0$, por lo que tenemos dos raíces complejas. Por tanto, una base real de soluciones de la ecuación homogénea es $\{\cos 2t, \sin 2t\}$, como habíamos anticipado.

2.7 Coeficientes indeterminados

En algunos casos el cálculo de una solución particular de la ecuación completa puede simplificarse debido a la forma especial de la parte no homogénea. El método, llamado de coeficientes indeterminados, puede justificarse como sigue.

Sea $b(t)$ una función que es anulada por un determinado operador diferencial lineal de orden m con coeficientes constantes L' . La ecuación de la que deseamos conocer una solución particular es:

$$Lx = b$$

aplicando L' a los dos miembros de esta ecuación, se obtiene:

$$L'Lx = L'b = 0$$

es decir, $x(t)$ es solución de una ecuación diferencial lineal de orden $n + m$ con coeficientes constantes, homogénea. Toda solución de la ecuación no homogénea de partida es solución de esta ecuación diferencial, aunque el inverso no es cierto. De cualquier forma solo necesitamos una solución particular. La idea es calcular las soluciones de esta segunda ecuación homogénea y sustituir la función obtenida en la ecuación diferencial inicial, tratando de ajustar los coeficientes de forma que se obtenga una solución. Entre las soluciones de la ecuación homogénea $L'Lx = 0$ aparecerán las soluciones de la ecuación $Lx = 0$, necesarias para el cálculo de la solución general de $Lx = b$ pero que en esta fase del desarrollo no aportan nada nuevo ya que son anuladas por L . De esta forma solo necesitamos utilizar las soluciones de $L'Lx = 0$ que no son soluciones de $Lx = 0$. Por supuesto el operador L' no es único así que habrá que buscarlo lo más sencillo posible. En el caso en que L y L' no tengan raíces comunes (sus polinomios característicos respectivos) la situación es sencilla, ya que se busca entre las soluciones de L' exclusivamente. En el caso en que haya raíces comunes habrá que tener en cuenta la multiplicidad mayor que uno que se produce al multiplicar L y L' . De acuerdo con todo lo expuesto se pueden dar los siguientes criterios para la búsqueda de soluciones particulares en este caso:

1. Si $b(t) = q(t)e^{\lambda t}$, la solución particular se busca como:

$$x(t) = t^r Q(t)e^{\lambda t}$$

donde $Q(t)$ es un polinomio de grado menor o igual que el grado de $q(t)$ y r es la multiplicidad de λ como raíz del polinomio característico de L (que puede ser cero). Sustituyendo $x(t)$ en la ecuación $Lx = b$ se calculan los coeficientes que aparecen en el polinomio $Q(t)$.

2. Si $b(t) = q(t)e^{\mu t} \cos \nu t$, o $b(t) = q(t)e^{\mu t} \sin \nu t$ se puede reducir al caso anterior. Pero se puede ver fácilmente que ambos tipos corresponden a raíces complejas $\mu \pm i\nu$ y por lo tanto la función que hay que probar es:

$$x(t) = t^r e^{\mu t} (Q_1(t) \cos \nu t + Q_2(t) \sin \nu t)$$

donde $Q_1(t)$, $Q_2(t)$ son polinomios de grado menor o igual que el grado de $q(t)$ y r es la multiplicidad de $\mu + i\nu$ como raíz de L .

3. La función $b(t)$ más general es suma de funciones de los tipos dados en 1) y 2). Sea:

$$b(t) = \sum_{k=1}^m b_k(t).$$

Una forma de resolver este problema es descomponerlo en m problemas no homogéneos, cada uno de ellos con inhomogeneidad igual a $b_k(t)$. De esta forma la solución particular buscada es la suma de las soluciones particulares halladas en cada uno de los problemas:

$$L\left(\sum_{k=1}^m x_k(t)\right) = \sum_{k=1}^m L(x_k(t)) = \sum_{k=1}^m b_k(t) = b(t)$$

Ejemplo 2.7.1 Consideremos nuevamente el oscilador armónico forzado:

$$x'' + 4x = \cos 2t.$$

El término no homogéneo es solución de una ecuación diferencial lineal de coeficientes constantes de segundo orden con raíces $\pm 2i$, que también lo son del polinomio característico de la ecuación con multiplicidad igual a 1. Por lo tanto la solución a probar es:

$$x(t) = t(a \cos 2t + c \sin 2t)$$

donde a, c son dos constantes a determinar. Sustituyendo en la ecuación:

$$-4a \operatorname{sen} 2t + 4c \operatorname{cos} 2t = \operatorname{cos} 2t$$

por tanto, $a = 0, c = 1/4$, de donde la solución particular es:

$$x(t) = \frac{t}{4} \operatorname{sen} 2t$$

como ya sabíamos.

2.8 Estabilidad de ecuaciones lineales

Consideremos un sistema lineal $x' = A(t)x + b(t)$, donde se supone que $A(t)$ y $b(t)$ son funciones continuas a partir de un cierto t_0 , con lo que las soluciones que parten de ese punto se pueden prolongar indefinidamente hacia la derecha.

Proposición 2.8.1 *Las soluciones de un sistema de ecuaciones diferenciales de primer orden lineal son estables (asintóticamente estables) si y solo si lo es la solución trivial del sistema homogéneo.*

También se tiene la misma relación entre acotación y estabilidad:

Proposición 2.8.2 *Las soluciones de un sistema lineal homogéneo son estables si y solo si son acotadas.*

Demostración. Supongamos que la ecuación es estable. Entonces la solución trivial es estable, por tanto para todo $\epsilon > 0 \exists \delta$ tal que, si $\|x_0\| < \delta$, entonces $\|x(t)\| < \epsilon$ para todo $t \geq t_0$, donde $x(t)$ es la solución del sistema con $x(t_0) = x_0$. Esto quiere decir que las soluciones cuyo dato inicial está en un entorno de 0 están acotadas (recuérdese que el concepto de acotación utilizado aquí se refiere a puntos situados a la derecha del punto inicial). Pero si $y(t)$ es una solución cualquiera con $y(t_0) = y_0$, se puede hacer una homotecia que lleve el dato inicial a ese entorno del origen, y considerar la solución con dato inicial $z(t_0) = y_0/c$, donde $\|y_0\|/c < \delta$, eligiendo adecuadamente la constante c . Como el sistema es lineal, la solución con este dato inicial es $z(t) = y(t)/c$ y por lo tanto de $\|z(t_0)\| < \delta$ se deduce $\|z(t)\| < \epsilon$, es decir, $\|y(t)\| \leq c\epsilon$ para todo $t \geq t_0$, luego toda solución está acotada.

Supongamos ahora que las soluciones de la ecuación están acotadas. Esto quiere decir que podemos construir una matriz fundamental $\Phi(t)$, acotada, de forma que dada una solución $x(t) = \Phi(t)\Phi(t_0)^{-1}x_0$ se tiene:

$$\|x(t)\| = \|\Phi(t)\Phi(t_0)^{-1}x_0\| \leq \|\Phi(t)\Phi(t_0)^{-1}\|_s \|x_0\| \leq \mu \|x_0\|$$

para todo $t \geq t_0$, pues la matriz $\Phi(t)$ está acotada por μ que en principio puede depender de t_0 (si no es así, se habla de estabilidad uniforme) pero que no depende de x_0 . Usando la desigualdad anterior se ve que basta escoger $\delta = \epsilon/\mu$ en la definición de estabilidad para asegurar que la solución trivial es estable y por ende el sistema. QED

Nótese que es necesario que todas las soluciones estén acotadas, concretamente que una base de soluciones esté acotada. En el caso de la ecuación lineal de primer orden bastaba que una solución no trivial estuviera acotada. Obviamente, pues la base allí tiene un solo elemento.

En el caso de coeficientes no constantes la situación puede complicarse pues no sabremos en general la forma explícita de las soluciones. En el caso de coeficientes constantes veremos como se pueden dar criterios muy sencillos para estudiar la estabilidad del sistema.

Proposición 2.8.3 *Sea el sistema de ecuaciones diferenciales lineal homogéneo de coeficientes constantes: $x' = Ax$. Entonces:*

1. *Si la parte real de los autovalores de A es negativa, el sistema es asintóticamente estable.*
2. *Si la parte real de los autovalores es negativa, excepto la de algunos que es cero, y estos últimos tienen cajas de Jordan asociadas de dimensión 1, entonces el sistema es estable.*

3. Si algún autovalor tiene parte real positiva o algún autovalor con parte real igual a cero tiene una caja de Jordan con dimensión mayor que 1, el sistema es inestable.

Demostración. Es trivial, basándose en la estructura de las soluciones. Recordemos que éstas son exponenciales por polinomios, así que el comportamiento cuando $t \rightarrow \infty$ viene dado por la parte real de los exponentes, que son los autovalores. Si la parte real es negativa, las soluciones están acotadas y tienden a cero cuando t tiende a infinito, luego el sistema es asintóticamente estable. Si la parte real es cero, habrá que fijarse si existe o no un polinomio (de grado mayor que 0) que multiplique a la exponencial. Si la multiplicidad geométrica del autovalor es igual a su multiplicidad aritmética (es decir, si las cajas correspondientes a ese autovalor son de dimensión 1 en la forma de Jordan correspondiente) no existe tal polinomio, y el sistema será estable. Si la caja no es diagonal, el sistema es inestable debido al polinomio. Finalmente si algún autovalor tiene parte real positiva la solución correspondiente no estará acotada y por tanto el sistema es inestable. QED

Ejemplo 2.8.1

$$\begin{cases} x' &= -5x + 4y \\ y' &= -3x + 2y \end{cases}$$

Los autovalores de la matriz A son -1 , -2 , por lo tanto el sistema es asintóticamente estable. Las soluciones tienden a cero cuando t tiende a infinito. Nótese que $\text{tr } A = -3$ y $\det A = 2$, lo que permite asegurar sin necesidad de efectuar los cálculos explícitos, que las dos raíces son negativas. Si consideramos un sistema no homogéneo, por ejemplo:

$$\begin{cases} x' &= -5x + 4y + t \\ y' &= -3x + 2y + t \end{cases}$$

las soluciones de este sistema tienden a:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} (t - 1)$$

que no está acotada (recuérdese que es un sistema no homogéneo). En efecto, la solución general del sistema es:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} e^{-2t} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-t} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} (t - 1)$$

y las dos exponenciales negativas hacen que las condiciones iniciales no influyan en los valores de la solución para t grande.

En el caso de ecuaciones diferenciales lineales de orden n con coeficientes constantes, la situación es la misma:

$$x^{(n)} + a_{n-1}x^{(n-1)} + \dots + a_1x' + a_0x = b(t)$$

se construye el polinomio característico $p(\lambda)$ y se analizan sus raíces. Si todas ellas tienen la parte real negativa, la ecuación es asintóticamente estable. Si alguna tiene la parte real igual a cero y es simple, pero todas las demás tienen parte real negativa, el sistema es estable. Si por fin, hay raíces con parte real positiva o con parte real cero pero no simples, entonces el sistema es inestable. De aquí resulta fundamental la localización de las raíces en el plano complejo.

Existen muchas técnicas para determinar la situación de las raíces. Nos limitaremos a dar uno de los criterios más sencillos, el de Routh-Hurwitz que establece condiciones necesarias y suficientes para que todas las raíces de un polinomio con los coeficientes reales estén en el semiplano izquierdo. Sin embargo, no proporciona ninguna indicación sobre la clase de inestabilidad, por lo que para obtener mayor información es necesario acudir a otro tipo de criterios como el de Nyquist, que no trataremos aquí. Todos ellos hacen uso de las técnicas de variable compleja en el estudio de los ceros de un polinomio.

2.9 El criterio de Routh-Hurwitz

Supongamos que la carga, q , de un condensador situado en un circuito eléctrico verifica la siguiente ecuación:

$$(LL_0 - M)q''' + (R_0L + RL_0)q'' + (MA/C + RR_0 + L_0/C)q' + (R_0/C)q = 0$$

donde las constantes L , L_0 , R , R_0 , C y M son características de los elementos del circuito y A es la ganancia de un cierto amplificador diferencial que se encuentra también en ese circuito. La constante M es la inductancia mutua entre las bobinas L y L_0 . Las propiedades de estabilidad de este circuito dependen de las constantes anteriormente citadas. Resulta interesante saber para qué valores de estas constantes el circuito es estable, y si no lo es, qué tipo de inestabilidad tiene. No es trivial aplicar directamente el teorema anterior pues nos veríamos obligados a calcular las raíces de un polinomio de tercer grado en función de una serie de constantes que no aparecen en forma sencilla en la ecuación. Veremos como el criterio de Routh-Hurwitz permite al menos resolver la primera cuestión.

Proposición 2.9.1 *Criterio de Routh-Hurwitz: La condición necesaria y suficiente para que un polinomio de coeficientes reales:*

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

tenga todas sus raíces en el semiplano izquierdo (parte real negativa) es que los menores principales de la siguiente matriz sean todos positivos:

$$\begin{pmatrix} a_{n-1} & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & a_1 & a_2 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & a_0 \end{pmatrix}$$

Obsérvese como en la construcción de esta matriz se han colocado los coeficientes del polinomio sobre la diagonal principal (salvo el 1 de λ^n) y a derecha e izquierda a partir de la diagonal el resto de esos coeficientes en orden creciente.

Del criterio de Routh-Hurwitz se deduce que una condición necesaria para que todas las raíces tengan parte real negativa es que todos los coeficientes del polinomio sean positivos. Sin embargo esta condición no es suficiente. La demostración del criterio de Routh-Hurwitz se basa en el concepto de índice de Cauchy pero no entraremos en ella aquí. Veamos algunos casos particulares.

Ejemplo 2.9.1 $n = 2$, $p(\lambda) = \lambda^2 + a_1\lambda + a_0$. La matriz a considerar es:

$$\begin{pmatrix} a_1 & 1 \\ 0 & a_0 \end{pmatrix}$$

por tanto las condiciones son $a_0 > 0$, $a_1 > 0$.

Ejemplo 2.9.2 $n = 4$, $p(\lambda) = \lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$. La matriz es:

$$\begin{pmatrix} a_3 & 1 & 0 & 0 \\ a_1 & a_2 & a_3 & 1 \\ 0 & a_0 & a_1 & a_2 \\ 0 & 0 & 0 & a_0 \end{pmatrix}$$

y los menores que deben ser positivos son:

$$a_3 > 0, a_2a_3 - a_1 > 0, a_1(a_2a_3 - a_1) - a_0a_3^2 > 0, a_0 > 0$$

de donde se deduce inmediatamente que todos los coeficientes son positivos además de satisfacer otras restricciones.

Ejemplo 2.9.3 Veamos ahora el circuito que se expuso al principio de este apartado. Es una ecuación de tercer orden, y el polinomio característico es:

$$p(\lambda) = \lambda^3 + \frac{b}{a}\lambda^2 + \frac{d}{a}\lambda + \frac{c}{a}$$

donde $a = LL_0 - M^2$, $b = R_0L + RL_0$, $c = R_0/C$, $d = MA/C + RR_0 + L_0/C$. Por tanto la matriz a estudiar es:

$$\begin{pmatrix} b/a & 1 & 0 \\ c/a & d/a & b/a \\ 0 & 0 & c/a \end{pmatrix}.$$

y los menores correspondientes:

$$\frac{b}{a}, \frac{db - ac}{a}, \frac{c}{a}.$$

Pero $b = R_0L + RL_0$ es positivo, así que a debe ser también positivo:

$$LL_0 \geq M^2$$

que se verifica de manera automática, pues M es el coeficiente de inductancia mutua entra las bobinas de autoinducciones L y L_0 . El coeficiente c también es positivo, $c = R_0/C$. De esta forma, solo nos queda una condición a satisfacer para que el circuito sea estable asintóticamente:

$$db - ac = \left(\frac{MA}{C} + RR_0 + \frac{L_0}{C}\right)(R_0L + RL_0) - \frac{(LL_0 - M)R_0}{C}.$$

Como se puede ver fácilmente, esta condición se satisface siempre cuando el coeficiente M es positivo (se supone A positivo). Sin embargo, si $M < 0$ existe un valor crítico de la ganancia por encima del cual el sistema es inestable:

$$A_c = \frac{RR_0C + L_0 - (LL_0 - M)R_0/(R_0L + RL_0)}{|M|}$$

El criterio de Routh-Hurwitz no permite saber que clase de inestabilidad se produce para $A > A_c$. Para ver que se trata de un foco inestable, es decir, de una situación oscilante con amplitud creciente, es necesario recurrir a métodos como el criterio de Nyquist, que estudian el movimiento de las raíces en el plano complejo según los valores de A . Nótese que cuando A alcanza el valor crítico, la ecuación es:

$$q''' + (b/a)q'' + (c/b)q' + (c/a)q = 0$$

y el menor $(db - ac)/a$ del criterio de Routh-Hurwitz es cero. Sin embargo los coeficientes del polinomio siguen siendo positivos. Puesto que las raíces varían continuamente con A , se puede pensar que en $A = A_c$, alguna de las raíces tiene su parte real igual a cero (pues para $A > A_c$ tiene su parte real positiva). Pero no puede ser cero pues el producto de las tres raíces es distinto de cero (y negativo). Por lo tanto en $A = A_c$ aparecen dos raíces imaginarias puras, que se pueden calcular. Supongamos que son $\pm i\omega$. Sustituyendo en el polinomio característico, se tiene:

$$-iab\omega^3 - b^2\omega^2 + iac\omega + cb = 0$$

e igualando a cero la parte real y la imaginaria (o sustituyendo también $-i\omega$):

$$\omega^2 = \frac{c}{b} = \frac{R_0}{(R_0L + RL_0)C}.$$

La tercera raíz es real negativa: $-b/a$. Cuando $A > A_c$, aparecen dos raíces complejas con parte real positiva y se producen oscilaciones de amplitud creciente (cuando las condiciones iniciales hacen que estas soluciones aparezcan). Esta última afirmación requiere una comprobación más cuidadosa, pues podrían presentarse otros casos de inestabilidad sin oscilación.

Ejemplo 2.9.4 Sea la ecuación diferencial lineal de coeficientes constantes cuyo polinomio característico es:

$$p(\lambda) = \lambda^3 + 5k\lambda^2 + (2k + 3)\lambda + 10.$$

Usando el criterio de Routh-Hurwitz se deduce que para $k > 1/2$ el sistema es asintóticamente estable. Un análisis más detallado de las raíces, muestra que para k menor que -0.90037 hay tres raíces reales, una de ellas negativa y dos positivas. Para $k = -0.90037$, las dos raíces positivas se convierten en complejas conjugadas con parte real positiva hasta que k vale $1/2$, donde la parte real se hace cero. A partir de ese momento, la parte real es siempre negativa. Sin embargo las raíces vuelven a ser reales (negativas) cuando $k = 46.907$. La tercera raíz es siempre negativa.

Ejemplo 2.9.5 Consideremos un circuito eléctrico en el que la intensidad verifica la siguiente ecuación:

$$I'' + 2I' + 5I = 10 \cos t.$$

El polinomio característico de la ecuación homogénea es:

$$p(\lambda) = \lambda^2 + 2\lambda + 5$$

y sus raíces son $-1 \pm 2i$, así que el sistema es asintóticamente estable, como por otra parte podría haberse deducido del criterio de Routh-Hurwitz, pues $2 > 0$ y $5 > 0$. Una solución particular de la ecuación completa es, como puede comprobarse fácilmente:

$$\sqrt{5} \cos(t + \phi)$$

con $\tan \phi = -1/2$. Por tanto la solución general de la ecuación completa es:

$$I(t) = e^{-t}(a_1 \cos t + a_2 \sin t) + \sqrt{5} \cos(t + \phi)$$

en donde se observa la suma de dos términos: el régimen transitorio que depende de las condiciones iniciales y que tiende a cero cuando t tiende a infinito (en este término aparecen las frecuencias naturales del sistema), y el régimen estacionario que es una oscilación de amplitud $\sqrt{5}$ y fase ϕ , dado por la excitación a la que se somete el circuito.

No entraremos en más detalles sobre la estabilidad de sistemas lineales. Cuando estudiemos los sistemas autónomos tendremos ocasión de volver sobre ellos.

Tema 3

Sistemas dinámicos

3.1 Introducción

En esta sección emprenderemos el estudio elemental de los sistemas autónomos, es decir, sistemas de ecuaciones diferenciales en las que no aparece de forma explícita la variable independiente que llamaremos siempre t , tiempo. Cuando t aparece en el sistema, siempre se puede escribir un sistema autónomo, considerando a t como función de un cierto parámetro $t = s$ que ya no está en las ecuaciones; el sistema tiene una ecuación y una incógnita más. La mayor parte de los resultados que estableceremos se referirán al caso de sistemas autónomos en el plano, donde la teoría es muy completa, pero lo que nos impedirá adentrarnos en el estudio de problemas más complejos, como caos, que aparecen en más dimensiones (nótese que esto excluye la posibilidad de estudiar sistemas en el plano que dependan de t , ya que el procedimiento indicado anteriormente nos llevaría a un espacio de tres dimensiones). La representación gráfica jugará un papel esencial en lo que sigue, pues, salvo en el caso de los sistemas lineales, no es fácil obtener soluciones analíticas de estos sistemas; a cambio, se pueden dar descripciones cualitativas muy sugerentes del comportamiento de las soluciones.

3.2 Definiciones y resultados elementales

Hasta que nos limitemos al caso $n = 2$, lo que sigue se aplica a cualquier número de dimensiones. Sea pues el sistema autónomo:

$$x' = f(x)$$

donde $x(t)$ es una función vectorial (con valores en \mathbb{R}^n) de variable real, y $f(x)$ una función definida en un cierto dominio (abierto conexo, por ejemplo) de \mathbb{R}^n , suficientemente regular (continua y lipschitziana, analítica las más de las veces) de forma que las cuestiones de existencia y unicidad serán automáticamente satisfechas. Una solución de esta ecuación es una función de un intervalo de la recta real $[t_0, t_1]$ en \mathbb{R}^n , es decir, $(t, x(t))$, la gráfica de la solución, que se suele llamar trayectoria, es una curva en $\mathbb{R} \times \mathbb{R}^n$. Tan importante como ella, es la proyección en el espacio \mathbb{R}^n , pues, de la independencia de f respecto a t , se deduce que estas proyecciones, que llamaremos órbitas, no se cortan en \mathbb{R}^n (en $\mathbb{R} \times \mathbb{R}^n$ las trayectorias no lo hacen debido a que se satisfacen las exigencias de los teoremas de existencia y unicidad, pero no es trivial que las proyecciones no se corten).

Definición 3.2.1 *Se dice que $x_0 \in \mathbb{R}^n$ es un punto crítico para la ecuación $x' = f(x)$ si $f(x_0) = 0$. Entonces, $x(t) = x_0$ es una solución constante del sistema.*

Ya hemos visto este tipo de situaciones para las ecuaciones autónomas ($n = 1$).

Como consecuencia del carácter autónomo del sistema, si $x(t)$ es una solución, $x(t - \tau)$, donde τ es una constante real, es otra solución, evidentemente distinta si no es constante o periódica. Estas dos soluciones tienen la misma proyección sobre el espacio \mathbb{R}^n . Sean Γ y Γ_τ las trayectorias asociadas a las soluciones $x(t)$, $x(t - \tau)$, es decir:

$$\Gamma = \{(t, x(t))\}, \quad \Gamma_\tau = \{(t, x(t - \tau))\}$$

donde t varía en un cierto intervalo; son curvas diferentes, en general, en $\mathbb{R} \times \mathbb{R}^n$. Sean γ , γ_τ las proyecciones de Γ y Γ_τ en \mathbb{R}^n :

$$\gamma = \{x(t)\}, \quad \gamma_\tau = \{x(t - \tau)\}$$

Estas dos curvas, consideradas como conjunto de puntos son las mismas en \mathbb{R}^n . Corresponden a distintas parametrizaciones del mismo lugar geométrico. Veamos como por un punto de \mathbb{R}^n , contenido en el dominio donde estamos trabajando, pasa una única órbita. Sea $x_0 \in D$ y Γ_0 la trayectoria correspondiente a la única solución, $x(t)$, que verifica $x(0) = x_0$. Sea Γ_τ la trayectoria correspondiente a la única solución, $\tilde{x}(t)$, que verifica $\tilde{x}(\tau) = x_0$. Si x_0 no es un punto crítico, las trayectorias Γ_0 y Γ_τ son distintas. Sin embargo, su proyección sobre el espacio \mathbb{R}^n es la misma, pues $x(t - \tau) = \tilde{x}(t)$ para todo t donde estén definidas. Esta última afirmación es consecuencia del hecho de estar trabajando con un sistema autónomo. Es decir, una órbita es la proyección de una familia de trayectorias que se obtienen unas de otras por traslaciones del parámetro t . En el caso x_0 punto crítico, solo hay una trayectoria que se proyecte en él, $x(t) = x_0$. Por tanto, cualquier otra trayectoria que tienda a él, lo hace cuando t tiende a más infinito o menos infinito.

Las orbitas se pueden considerar descritas en forma paramétrica, con t como parámetro. Pero también pueden ser estudiadas en forma implícita, eliminando el parámetro t . Esto equivale a considerar el siguiente sistema de ecuaciones:

Si el sistema original es:

$$x'_1 = f_1(x), \quad x'_2 = f_2(x), \dots, x'_n = f_n(x)$$

el sistema asociado que proporciona las órbitas en forma implícita, es:

$$\frac{dx_1}{f_1(x)} = \frac{dx_2}{f_2(x)} = \dots = \frac{dx_n}{f_n(x)}$$

En el caso de un sistema en el plano, este sistema asociado se reduce a una ecuación diferencial de primer orden. Las órbitas son las curvas integrales de este sistema.

Una órbita no puede cortarse a sí misma sin corresponder a una solución periódica. En efecto, supongamos que para una órbita dada se tiene:

$$x(0) = x(\tau)$$

para algún $\tau > 0$. Entonces, $x(t) = x(t + \tau)$ para todo t , de acuerdo con el teorema de existencia y unicidad, así que la solución $x(t)$ es periódica, y la órbita correspondiente es una curva de Jordan.

Se llama a \mathbb{R}^n el **espacio de fases** del sistema. Dado un punto de este espacio (supongamos que $D = \mathbb{R}^n$), existe una única órbita que pasa por él. Los puntos de la órbita se obtienen a partir de x_0 :

$$x(t) = g^t x_0$$

donde g^t es una aplicación de \mathbb{R}^n en \mathbb{R}^n , el llamado **flujo** asociado al sistema. Desde este punto de vista, el problema puede plantearse de la forma que se expone a continuación.

Definición 3.2.2 Se llama grupo uniparamétrico de transformaciones a una familia de aplicaciones de M en M (donde M es un conjunto de puntos) que dependen de un parámetro $t \in \mathbb{R}$, y que verifican:

$$g^t g^s = g^{t+s}, \quad \forall t, s \in \mathbb{R}$$

$$g^0 = \text{identidad en } M$$

Definición 3.2.3 Un flujo es el par formado por el conjunto M y un grupo uniparamétrico de transformaciones de M , $\{g^t\}$. M es el espacio de fases del flujo.

Definición 3.2.4 De acuerdo con estas definiciones, una órbita en M es la imagen de un punto de M por todas las aplicaciones del grupo $\{g^t\}$.

Es decir, como hemos visto antes, si se define la aplicación ϕ :

$$\phi: \mathbb{R} \rightarrow M$$

por: $\phi(t) = g^t x_0$, donde x_0 es un punto de M , la órbita que pasa por x_0 es la imagen de ϕ .

Definición 3.2.5 *Se dice que x_0 es un punto de equilibrio del flujo, si el grupo $\{g^t\}$ lo deja invariante.*

Entonces, la aplicación ϕ correspondiente es constante y la órbita que pasa por x_0 se reduce a ese punto.

En las aplicaciones que haremos, el conjunto M es un abierto de \mathbb{R}^n , (con más generalidad, una variedad diferenciable) y las aplicaciones del grupo son difeomorfismos (es decir, tanto ellas como sus inversas, que existen, son diferenciables). Con más precisión se tiene:

Definición 3.2.6 *El conjunto de aplicaciones $\{g^t\}$ de la variedad diferenciable M en sí misma, es un grupo uniparamétrico de difeomorfismos si:*

$$g: \mathbb{R} \times M \rightarrow M, \quad g(t, x) = g^t x$$

es una aplicación diferenciable, las aplicaciones:

$$g^t: M \rightarrow M$$

son difeomorfismos, y $\{g^t\}$ es un grupo uniparamétrico de transformaciones de M .

Otro concepto importante es el de campo de vectores.

Definición 3.2.7 *La velocidad del flujo, definida por:*

$$v(x) = \left[\frac{d}{dt} g^t x \right]_{t=0}$$

proporciona un campo de vectores en el conjunto M , asociando a cada punto x de M el vector $v(x)$ correspondiente.

Definición 3.2.8 *Un punto crítico del campo de vectores, es aquél donde la velocidad se hace cero: x_0 es un punto crítico si $v(x_0) = 0$.*

Resulta entonces que los puntos de equilibrio del flujo son puntos críticos del campo de velocidades. Y viceversa.

Dado un grupo uniparamétrico de transformaciones, hemos visto como calcular el campo de velocidades asociado a él. En general el problema es el contrario, dado un campo de vectores, calcular un grupo uniparamétrico de transformaciones que tenga como campo de velocidades el campo de vectores dado. Es decir, resolver el sistema de ecuaciones diferenciales de primer orden:

$$x' = v(x)$$

Una solución de esta ecuación es una función definida en un intervalo de la recta real con valores en M . La imagen de la aplicación que da la solución es una órbita en el espacio de las fases M . Además, esa órbita es tangente en cada uno de sus puntos al campo de vectores. Finalmente, la solución puede ponerse como:

$$x(t) = g^t x_0$$

donde $x(0) = x_0$, y $\{g^t\}$ es un grupo uniparamétrico de transformaciones en M , cuyo campo de velocidades es $v(x)$.

Llamaremos muchas veces sistemas dinámicos a este tipo de sistemas.

3.3 Sistemas lineales

El estudio de los sistemas autónomos lineales es una teoría puramente algebraica, como hemos tenido ocasión de ver, y que puede ser estudiada completamente. No ocurre así con los no lineales. Veremos, sin embargo, que en muchos casos los sistemas no lineales pueden ser estudiados a partir de las aproximaciones lineales que de ellos se hacen en entornos de puntos críticos, por lo que repasaremos los sistemas lineales desde los puntos de vista introducidos previamente.

Consideremos un sistema de ecuaciones lineales autónomo, es decir, de coeficientes constantes:

$$x' = Ax$$

donde A es una matriz $n \times n$.

El espacio de fases de este sistema es \mathbb{R}^n , y el flujo tiene como grupo uniparamétrico de transformaciones $\{e^{At}\}$, con $t \in \mathbb{R}$. En efecto, las soluciones de este sistema son:

$$x(t) = e^{At}x_0$$

y la derivada de $e^{At}x_0$ con respecto a t , calculada en $t = 0$, es justamente Ax_0 , la velocidad. Este tipo de sistemas tiene como puntos críticos las soluciones de:

$$Ax = 0$$

Si A es una matriz regular, la única solución es $x = 0$, único punto de equilibrio del flujo. En general éste será el caso que estudiemos con más detalle. Se dice que el punto crítico es elemental.

Ejemplo 3.3.1 Consideremos la ecuación del oscilador armónico amortiguado:

$$x'' + \gamma x' + kx = 0$$

donde se supone $m = 1$. Esta ecuación de segundo orden, lineal de coeficientes constantes se puede escribir como un sistema dinámico de orden dos:

$$\begin{cases} x' = y \\ y' = -kx - \gamma y \end{cases}$$

cuya solución se obtiene inmediatamente. La matriz A es:

$$\begin{pmatrix} 0 & 1 \\ -k & -\gamma \end{pmatrix}$$

y suponiendo que k y γ son números reales positivos, no nulos, con γ^2 mayor que $4k$, los autovalores λ_1 , λ_2 , son reales, distintos y negativos. Sea $\lambda_1 > \lambda_2$. La solución explícita es:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix} e^{\lambda_1 t} + c_2 \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix} e^{\lambda_2 t}$$

Existe una transformación lineal que permite diagonalizar la matriz A . En la nueva base, formada por los autovectores, la solución se escribe como:

$$\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{\lambda_1 t} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{\lambda_2 t}$$

donde $a_1 = u(0)$, $a_2 = v(0)$. En esta base, la matriz A se convierte en:

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

y por lo tanto el campo de vectores asociado es:

$$\begin{pmatrix} \lambda_1 u \\ \lambda_2 v \end{pmatrix}$$

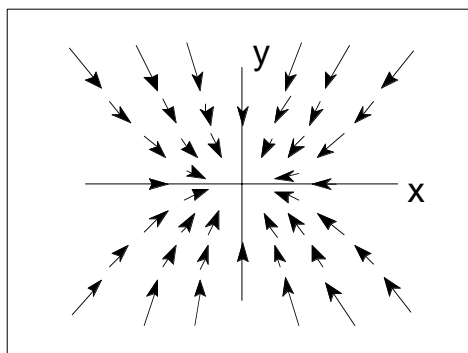


Figura 3.1

que aparece representado en la figura 3.1 (eje horizontal u , eje vertical v).

El único punto singular del campo es el origen, y los ejes resultan ser líneas del campo, es decir, tangentes al campo en todos sus puntos, por lo tanto posibles órbitas del sistema. No es cierto que los dos estén en la misma órbita, ni siquiera que uno de ellos sea una órbita. Las órbitas no se cortan y el origen es una órbita por sí solo. Por tanto cada eje contiene tres órbitas, el semieje positivo, el negativo y el origen. Puesto que el campo de vectores no tiene ninguna discontinuidad, uno cualquiera de los semiejes es una sola órbita. No puede haber más de una porque eso implicaría la existencia de nuevos puntos críticos, y como hemos dicho el único punto crítico es el origen. No resulta difícil dibujar las órbitas que aparecen en la figura 3.2.

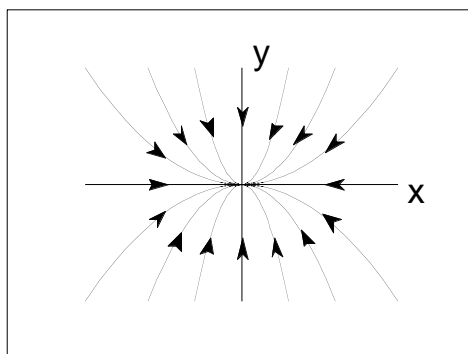


Figura 3.2

Nótese que todas tienden al origen cuando t tiende a infinito, lo que viene representado por la flecha dibujada sobre cada una de ellas. Además, todas ellas son tangentes al eje horizontal (asociado al autovalor λ_1) cuando t tiende a infinito, excepto las correspondientes al eje vertical. El campo de vectores entra en

cualquier circunferencia que se trace de centro el origen. Este punto crítico es un atractor (sumidero), es decir, la solución correspondiente es asintóticamente estable, como ya sabíamos.

La ecuación de las órbitas en implícitas, se obtiene fácilmente eliminando el tiempo:

$$|v|^{\lambda_1} = k|u|^{\lambda_2}$$

que representa todas las órbitas, con tal que se tome también $k = \infty$. Estas curvas son curvas integrales de la ecuación diferencial:

$$\frac{dv}{du} = \frac{\lambda_2 v}{\lambda_1 u}$$

que tiene un punto singular en el origen, por el que pasan infinitas soluciones. Los puntos del eje $u = 0$ son también singulares, pero por ellos no pasa ninguna solución, con lo que al escribir la ecuación asociada, cambiando los papeles de u y v , se obtiene una solución única.

En resumen, las curvas integrales de esta ecuación contienen las órbitas del sistema. Téngase en cuenta que una curva integral puede contener más de una órbita. Por ejemplo, $v = 0$ es una curva integral de esta ecuación que contiene tres órbitas del sistema.

Veamos ahora como las órbitas del sistema entran tangentes al eje horizontal (salvo las del eje vertical). De la ecuación asociada no se puede deducir, sin más, que la pendiente de todas las curvas integrales, excepto el eje vertical, se hace cero en el origen. Usando la expresión paramétrica de las soluciones se tiene:

$$\frac{dv}{du} = \frac{v'(t)}{u'(t)} = be^{(\lambda_2 - \lambda_1)t}$$

expresión que tiende a cero cuando t tiende a infinito, ya que $\lambda_1 > \lambda_2$. Solo cuando $b = \infty$, es decir, cuando estudiamos las órbitas contenidas en el eje vertical, $u = 0$, la pendiente no se hace cero al tender t a infinito.

Si ahora volvemos al sistema original, es decir, en la base canónica, la matriz del cambio de base es:

$$P = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}$$

con la que se obtiene la solución dada anteriormente. El mapa de fases se deforma al sufrir la transformación P . Sin embargo, los ejes se convierten en rectas, que son los autovectores, y con más generalidad, los elementos lineales, es decir el campo de vectores, se transforman linealmente. Por tanto, el nuevo mapa de fases es el que aparece en la figura 3.3.

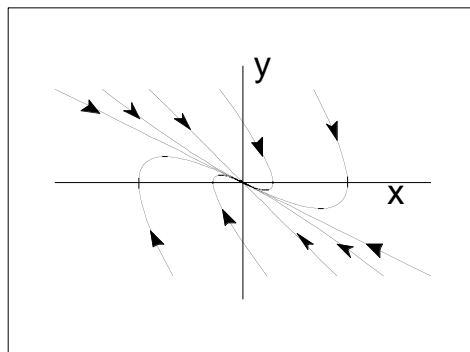


Figura 3.3

Las órbitas siguen cayendo hacia el origen cuando t tiende a ∞ , y entran en él tangentes a una de las rectas de autovectores, concretamente a la asociada al mayor de los dos autovalores, salvo por supuesto, las dos órbitas asociadas al otro autovalor, que lo hacen desde lados opuestos a la recta del autovalor mayor. En este mapa de fases, el eje horizontal es la posición del oscilador y el vertical la velocidad; cada punto representa un estado del sistema, que, como sabemos, viene dado por el par de datos posición-velocidad.

El movimiento en el espacio de configuración, que es en una recta, se puede deducir de la evolución del estado en el espacio de fases. Por ejemplo, supongamos que en $t = 0$ el estado del sistema es el punto **a** (x_0, v_0) con velocidad positiva y elongación negativa en el sector de menor ángulo entre las rectas de autovectores. El punto material con el que trabajamos, se dirige hacia el origen (el origen del espacio real) con velocidad decreciente, que tiende a cero. Cuando t tiende a ∞ tanto la posición como la velocidad se anulan, pero el punto no atraviesa el origen. Consideremos ahora el punto **b** como estado inicial (en el sector de mayor ángulo entre las rectas de autovectores) con velocidad positiva y elongación negativa, como antes. En este caso la relación entre posición y velocidad es tal que el punto llega al origen con velocidad no nula en un tiempo finito (insistimos en que se trata del origen en el espacio real, es decir el punto $x = 0$). En la órbita atravesamos el eje de velocidades. Debido a la velocidad no nula, el móvil atraviesa el origen y se desplaza hacia posiciones con x positiva, alejándose del origen hasta que su velocidad se anula (en la órbita hemos atravesado el eje de posiciones). A partir de este instante su velocidad es negativa y su posición es positiva. El móvil cae hacia el origen al que llega en un tiempo infinito. Nótese que desde el punto de vista de la ecuación de segundo orden de la que hemos partido, las soluciones tienen su gráfica en diagramas (t, x) , como se ve en la figura 3.4.

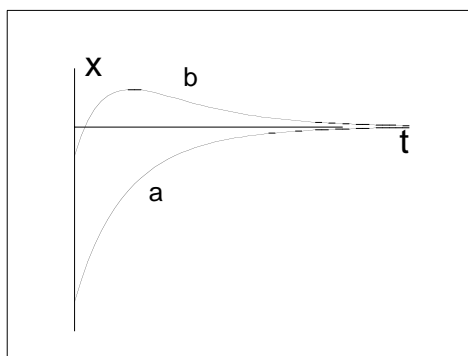


Figura 3.4

Los dibujos en el mapa de fases se pueden interpretar, desprovistos de cualquier comentario referente a las flechas que allí aparecen, como el dibujo de las curvas integrales de la ecuación de primer orden asociada, que se llama, por tanto, ecuación de las órbitas (basta recordar el método de isoclinas estudiado para las ecuaciones de primer orden para relacionar de forma clara estas dos formas de ver el problema). La relación entre un sistema de segundo orden y la ecuación de las órbitas asociada no es única (sea el sistema lineal o no, incluso sea de segundo orden o no). Dado un sistema, la ecuación de las órbitas es única, pero una ecuación de primer orden está asociada a una familia infinita de sistemas, con mapas de fases ciertamente relacionados, pero distintos en general. Si la ecuación es:

$$y' = \frac{dy}{dx} = \frac{f(x, y)}{g(x, y)}$$

podemos escribir los sistemas:

$$\begin{cases} x' &= k(x, y)g(x, y) \\ y' &= k(x, y)f(x, y) \end{cases}$$

donde $k(x, y)$ es cualquier función razonablemente suave. Todos estos sistemas tienen la misma ecuación de las órbitas. Pero dependiendo de $k(x, y)$ sus puntos singulares son distintos.

El comportamiento de las órbitas en un entorno del punto crítico, $(0, 0)$ en este ejemplo, se repite en otros muchos, y se dice que el origen es un nodo estable. Más adelante introduciremos la clasificación de los puntos singulares de los sistemas lineales. Pero antes repasaremos los conceptos de estabilidad en sistemas autónomos (lineales o no).

3.4 Estabilidad de sistemas autónomos

Consideremos el sistema:

$$x' = f(x)$$

con f definida en un abierto D de \mathbb{R}^n , continua y lipschitziana en D . Sea x_0 un punto crítico, es decir, $f(x_0) = 0$. Definimos las siguientes nociones de estabilidad. Supondremos que $x_0 = 0$, para mayor sencillez de notación.

Definición 3.4.1 *Se dice que el punto crítico 0 es estable si para todo $\epsilon > 0$ existe $\delta(\epsilon)$, tal que si $\|x_0\| < \delta$, entonces $\|x(t)\| < \epsilon$ para todo $t \geq 0$, donde $x(t)$ es la única solución con $x(0) = x_0$ (figura 3.5).*

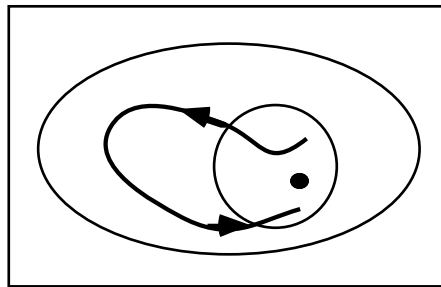


Figura 3.5

Como sabemos, en los sistemas lineales homogéneos, la estabilidad del origen equivale a la acotación de las soluciones.

Definición 3.4.2 *Si el punto crítico es estable y además, para todo $\epsilon > 0$ existe $\eta > 0$ tal que*

$$\lim_{t \rightarrow \infty} \|x(t)\| = 0,$$

donde $x(0) = x_0$, $\|x_0\| < \eta$ entonces se dice que el punto crítico es asintóticamente estable (figura 3.6).

En el caso de los sistemas lineales, sabemos que si la parte real de todos los autovalores es negativa, el sistema es asintóticamente estable.

En el caso de los sistemas lineales, $x' = Ax$, $A \in M_n(\mathbb{R})$, el origen es un punto singular. Supongamos que es el único. Se dice que $x = 0$ es un sumidero (atractor) del sistema si la parte real de todos los autovalores es negativa. Se dice también que el flujo correspondiente es una contracción. Si se elige adecuadamente la norma en el espacio de fases, todas las órbitas que atraviesan una cierta esfera de centro el origen apuntan hacia el origen. Como hemos dicho antes, las soluciones tienden hacia el origen (al ser un sistema lineal, todas las soluciones tienden a cero cuando t tiende a infinito). En el caso en

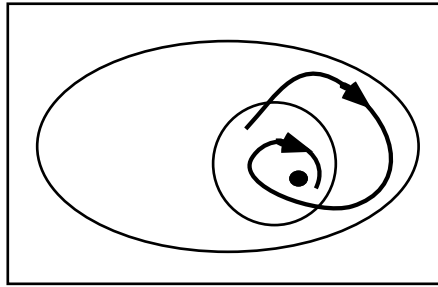


Figura 3.6

que la parte real de todos los autovalores sea positiva, se dice que el origen es una fuente (repulsor). Lo dicho anteriormente es ahora válido cambiando t por $-t$. El flujo se llama una expansión.

Los flujos más generales son aquellos en los que la matriz A tiene autovalores con parte real no nula. Se llaman flujos hiperbólicos. Se puede demostrar que todo flujo lineal puede ser aproximado arbitrariamente por un flujo hiperbólico. Los flujos lineales con autovalores de parte real cero son pues escasos (en el sentido anterior). Todo flujo lineal hiperbólico puede descomponerse en una parte correspondiente a un sumidero y otra correspondiente a una fuente. El siguiente teorema, de carácter puramente algebraico, lo afirma así:

Definición 3.4.3 Sea e^{tA} un flujo lineal hiperbólico, $A \in L(\mathbb{R}^n)$. Entonces, \mathbb{R}^n admite una descomposición en suma directa:

$$\mathbb{R}^n = E^s \oplus E^u$$

invariante bajo A , y de forma que el flujo e^{tA} sea una contracción cuando se restringe A al espacio E^s , y una expansión cuando se hace a E^u .

Esto no quiere decir, que todas las trayectorias tiendan al origen cuando t tiende a ∞ (espacio E^s) o a $-\infty$ (espacio E^u). Habrá otras que no tengan ninguno de estos dos comportamientos.

3.5 Clasificación de puntos críticos de sistemas lineales

En lo que sigue nos limitaremos a los sistemas en \mathbb{R}^2 . El caso $n > 2$ es mucho más complejo y no podemos tratarlo aquí. Estudiaremos primero los sistemas lineales y después los no lineales. Los puntos críticos serán elementales, como definiremos a continuación, aunque alguna nota se hará respecto a los no elementales.

Sea el sistema lineal autónomo en \mathbb{R}^2 :

$$\begin{cases} x' &= ax + by \\ y' &= cx + dy \end{cases}$$

donde a, b, c, d son números reales, y la matriz A :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Definición 3.5.1 Se dice que el origen, que es un punto crítico, es un punto elemental si el determinante de la matriz A es distinto de cero.

Entonces, sus autovalores son distintos de cero. El único punto singular es el origen, y esto es lo que clasificamos. El tipo de clasificación que intentamos establecer es el basado en la linealidad del sistema. Es decir, dos sistemas autónomos son linealmente equivalentes si existe una matriz 2×2 que lleva uno en otro. Por tanto:

Definición 3.5.2 Dos sistema autónomos lineales $u' = Au$, $v' = Bv$ (A, B matrices $n \times n$) son linealmente equivalentes si existe una matriz $n \times n$ regular, P , tal que:

$$B = PAP^{-1}$$

(las coordenadas se transforman como $v = Pu$).

Está claro ahora que lo único que interesa a la hora de establecer la equivalencia lineal son los autovalores y sus multiplicidades aritmética y geométrica. Teniendo en cuenta que se trata de un sistema de orden dos, las posibilidades que se presentan son las siguientes ($\det A \neq 0$).

1. Autovalores reales

(a) $(\operatorname{tr} A)^2 > 4 \det A > 0$. Es decir, raíces reales distintas del mismo signo. Según este signo, se dice que el origen es:

- i. si $\lambda_2 < \lambda_1 < 0$: NODO ESTABLE
- ii. si $0 < \lambda_1 < \lambda_2$: NODO INESTABLE

Las órbitas aparecen dibujadas en las figuras 3.7 y 3.8, donde las órbitas rectas son las direcciones de los autovectores. Nótese como en ambos casos todas las órbitas entran (nodo estable) o salen (nodo inestable) del origen tangentes a la recta de autovalor mayor (estable) o menor (inestable). Con más precisión, cuando t tiende a $+\infty$, las órbitas se aproximan al origen en el caso estable, mientras que en el caso inestable lo hacen cuando t tiende a $-\infty$. La demostración es inmediata y sigue las líneas de lo hecho en el caso del oscilador amortiguado (*mutatis mutandis*).

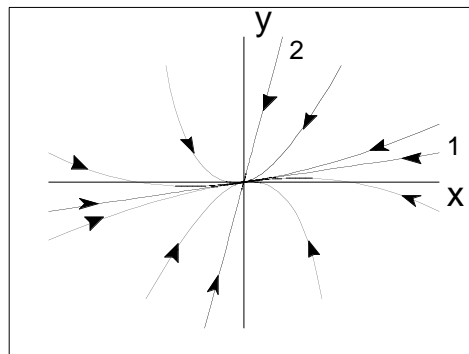


Figura 3.7

(b) $(\operatorname{tr} A)^2 = 4 \det A$. Es decir, raíces reales iguales. En este caso la matriz A es diagonal (recuérdese que es 2×2) o bien no es diagonalizable. En cualquier caso se sigue hablando de nodos.

- i. A diagonal. La razón entre $x(t)$ e $y(t)$ es constante, por lo que las órbitas son rectas que tienden al origen o se alejan de él, según que el autovalor sea negativo o positivo. Como antes, estamos en un flujo hiperbólico que es una contracción o una expansión. En este caso se suele hablar en ocasiones de un nodo estelar estable o inestable (figuras 3.9 y 3.10).
- ii. A no diagonalizable. Este caso se llama nodo de una tangente (solo hay un autovector). Hay dos órbitas que son rectas (situadas sobre una misma recta que pasa por el origen) y

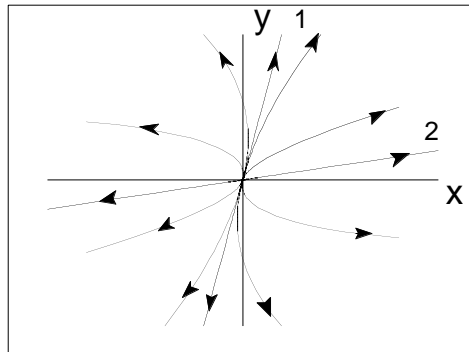


Figura 3.8

todas las demás entran (nodo estable, autovalor negativo) o salen (nodo inestable, autovalor positivo) del origen tangentes a ellas (figuras 3.11 y 3.12).

En resumen, los nodos aparecen ligados a flujos hiperbólicos que son una contracción o una expansión. Todas las órbitas acaban en el origen, bien cuando $t \rightarrow +\infty$, o cuando $t \rightarrow -\infty$.

- (c) $\det A < 0$. Es decir autovalores de distinto signo: $\lambda_1 < 0 < \lambda_2$. El flujo es aún hiperbólico, pero no es una contracción ni una expansión. De acuerdo con el teorema anterior, \mathbb{R}^2 se puede poner como suma directa de dos espacios unidimensionales, que son justamente los asociados a los dos autovectores. Sobre uno de ellos, el correspondiente al autovalor negativo, el flujo es una contracción. Sobre el otro, correspondiente al autovalor positivo, el flujo es una expansión. Hay dos órbitas que son rectas que tienden al origen cuando $t \rightarrow +\infty$. Y otras dos órbitas rectas que tienden al origen cuando $t \rightarrow -\infty$. Las otras órbitas no tienden al origen en ningún caso. Se dice que el origen es un PUNTO SILLA. En la figura 3.13 se dibujan las órbitas correspondientes.

2. Autovalores complejos $\lambda = \alpha \pm i\beta$, $\beta > 0$

- (a) $\alpha \neq 0$. Como ya hemos visto cuando estudiamos los sistemas lineales, existe una transformación lineal que permite escribir la matriz A en su forma canónica:

$$PAP^{-1} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

donde P es la matriz dada por:

$$P = \begin{pmatrix} v & u \end{pmatrix}$$

siendo $u = \Re(w)$, $v = \Im(w)$, w el autovector correspondiente a $\alpha + i\beta$. Consideremos entonces que la matriz A está en forma canónica. Si ponemos $z = x + iy$, el sistema es equivalente a la ecuación compleja:

$$z' = \lambda z$$

cuya solución se calcula inmediatamente:

$$z(t) = Ce^{\alpha t} e^{i\beta t}$$

Es decir, si $\alpha < 0$ aparecen espirales decrecientes, que tienden al origen cuando $t \rightarrow +\infty$. Tenemos nuevamente un flujo hiperbólico que corresponde a una contracción. Si $\alpha > 0$, el

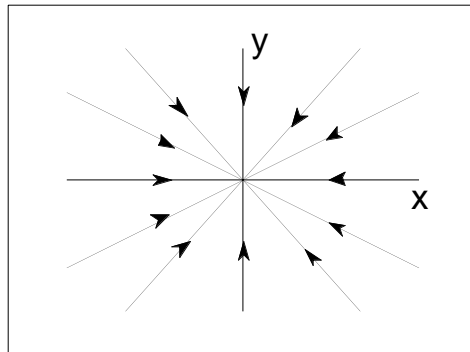


Figura 3.9

flujo hiperbólico que se obtiene es una expansión. Las órbitas son espirales que se alejan del origen cuando $t \rightarrow +\infty$. En el primer caso se tiene un FOCO ESTABLE. En el segundo se obtiene un FOCO INESTABLE (ver figuras 3.14, 3.15).

- (b) $\alpha = 0$. Estamos en el único caso en que no aparece un flujo hiperbólico. Ahora la parte real de los autovalores es cero. El sistema no es estructuralmente estable, en el sentido que una perturbación pequeña hace que la parte real deje de ser cero. Se dice que el punto crítico es un CENTRO, y es estable pero no asintóticamente estable. Las órbitas son circunferencias de centro el origen ya que el módulo de z es constante. En cualquier caso, las órbitas giran en torno al origen en sentido antihorario (positivo) (figura 3.16).

Los razonamientos anteriores se podrían haber hecho pasando a polares, donde se advierte el carácter espiral de las órbitas, en el caso de un foco, o como éstas son circunferencias en el caso de un centro. Sea:

$$z = \rho e^{i\theta}$$

Las ecuaciones que verifican el radio y el ángulo son:

$$z' = \lambda z \Rightarrow \rho' e^{i\theta} + i\rho\theta' e^{i\theta} = \rho(\alpha + i\beta)e^{i\theta}$$

de donde:

$$\begin{aligned} \rho' &= \alpha\rho \\ \theta' &= \beta \end{aligned}$$

cuya solución es inmediata:

$$\begin{aligned} \rho(t) &= Ce^{\alpha t} \\ \theta(t) &= \beta t + b \end{aligned}$$

Obtenemos pues espirales logarítmicas crecientes ($\alpha > 0$), decrecientes ($\alpha < 0$) o circunferencias ($\alpha = 0$), mientras el ángulo crece linealmente con t ($\beta > 0$). Si se elimina t entre ambas ecuaciones se obtiene la ecuación implícita de las órbitas:

$$\rho(\theta) = \rho_0 e^{\alpha\theta/\beta}$$

Nótese que $x(t)$, $y(t)$ son funciones armónicas de t , con amplitud dependiente también de t .

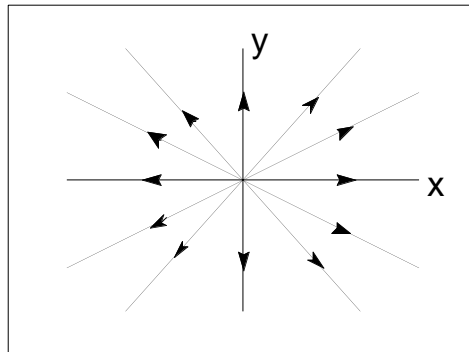


Figura 3.10

Supongamos ahora el caso general, en el que A tiene autovalores complejos pero no está en forma canónica. Al hacer la transformación P que la lleva a la forma canónica, calcular el mapa de fases y volver atrás, las órbitas se deforman, de manera que las espirales se alargan en alguna dirección y las circunferencias se transforman en elipses.

Además, el sentido de giro de las órbitas puede ser positivo o negativo. Esta última propiedad es fácilmente detectable en la matriz A . En efecto, el sentido es positivo si la transformación P conserva la orientación (ya que hemos visto que en la forma canónica el sentido era positivo). Así pues, basta calcular el determinante de la matriz P . Si es positivo, las órbitas giran en sentido positivo. Si es negativo, lo hacen en sentido negativo. Pero la matriz P es:

$$P = \begin{pmatrix} 0 & b \\ \beta & \alpha - a \end{pmatrix}$$

y su determinante es:

$$\det P = -b\beta$$

como $\beta > 0$, si b es negativo, el sentido de giro de las órbitas es positivo, y si b es positivo el sentido de giro es negativo. La constante b (el elemento 12 en la matriz A) no puede ser cero si hay autovalores complejos.

Supongamos que tenemos un centro. Los autovalores son $\lambda = \pm i\beta$, por lo que la traza de A es cero y el determinante es $\beta^2 > 0$. Sea A :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

con $bc < a^2$. Entonces $\beta^2 = -bc - a^2$. La ecuación de las órbitas puede obtenerse de la siguiente forma. Consideremos la ecuación de primer orden asociada:

$$\frac{dy}{dx} = \frac{cx - ay}{ax + by}$$

que es una ecuación homogénea (como siempre cuando el sistema es lineal), por lo que un cambio apropiado sería $v(x) = y/x$. Sin embargo, cuando la traza de la matriz A es cero, es decir cuando se trata de un centro y en algunos tipos de puntos silla, la ecuación anterior no solo es homogénea sino también exacta. Esto hace las cosas mucho más fáciles:

$$(ay - cx)dx + (ax + by)dy = 0$$

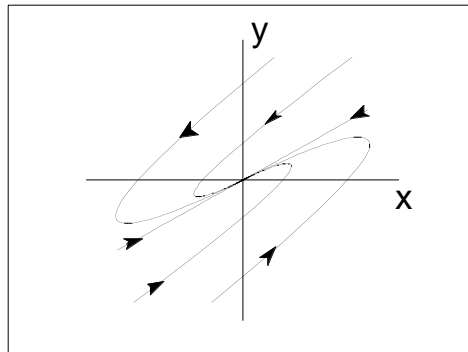


Figura 3.11

La solución viene dada por las curvas de nivel de la función:

$$U(x, y) = by^2 - cx^2 + 2axy$$

que son elipses. La matriz de la forma cuadrática que se construye es definida positiva (o definida negativa), pues, suponiendo que b sea mayor que cero, (ya sabemos que no es cero en este caso) c debe ser menor que cero (tampoco puede ser cero), y además, $-bc - a^2$ es positivo (el caso $b < 0$ es totalmente equivalente). Así pues las curvas:

$$by^2 - cx^2 + 2axy = k^2$$

son elipses de centro el origen. En cuanto al cálculo de sus ejes, comunes para toda la familia, se puede hacer de las múltiples formas en las que uno calcula los ejes de una cónica. Por ejemplo, en coordenadas polares:

$$\begin{aligned} x &= \rho \cos \theta \\ y &= \rho \sin \theta \end{aligned}$$

$$\frac{k^2}{\rho^2} = b \sin^2 \theta - c \cos^2 \theta + a \sin 2\theta$$

Los ejes unen el centro con los puntos más alejados (eje mayor) o menos (eje menor). Basta pues hallar los máximos y mínimos de la función:

$$h(\theta) = b \sin^2 \theta - c \cos^2 \theta + a \sin 2\theta$$

Derivando:

$$h'(\theta) = (b + c) \sin 2\theta + 2a \cos 2\theta$$

de donde igualando a cero:

$$\tan 2\theta = -\frac{2a}{b + c}$$

y de aquí podemos obtener la posición de los ejes.

Resumamos los resultados obtenidos (puntos críticos elementales):

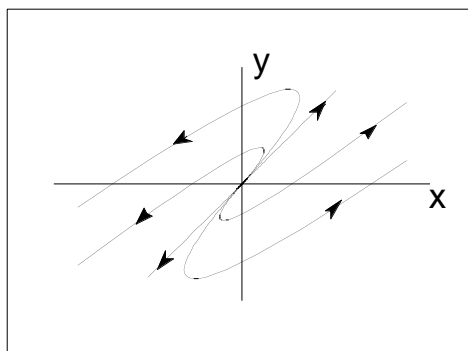


Figura 3.12

1. Autovalores reales

(a) Del mismo signo: (Positivos: Inestable; Negativos: Estable)

i. Distintos: NODO

ii. Iguales:

A. Matriz diagonalizable: Nodo estelar

B. Matriz no diagonalizable: Nodo de una tangente

(b) De distinto signo: PUNTO SILLA

2. Autovalores complejos $\lambda = \alpha \pm i\beta$, $\beta > 0$ (a) Con parte real distinta de cero: FOCO ($\alpha > 0$: Inestable; $\alpha < 0$: Estable)

(b) Con parte real igual a cero: CENTRO

En la figura 3.17, donde se han representado en el eje de abscisas el determinante de la matriz A y en el de ordenadas la traza, se hace un esquema de esta clasificación.

Ejemplo 3.5.1

$$\begin{cases} x' &= x + 2y \\ y' &= 2x - 2y \end{cases}$$

Los autovalores de la matriz A asociada son: $\lambda_1 = -3$, $\lambda_2 = 2$. Se trata por tanto de un punto silla. Las únicas órbitas rectas corresponden a los autovectores de la matriz que son:

$$\begin{aligned} \lambda_1 &\rightarrow (-1, 2) \\ \lambda_2 &\rightarrow (2, 1) \end{aligned}$$

Sobre la recta asociada al primer autovector, hay tres órbitas, el origen y dos que entran en él cuando t tiende a infinito. Sobre la otra, las dos órbitas correspondientes salen. Las otras órbitas tienden a las rectas cuando t tiende a $\pm\infty$. El hecho de que los dos autovectores sean ortogonales corresponde simplemente a que la matriz es simétrica. En general no ocurre así, como sabemos (figura 3.18).

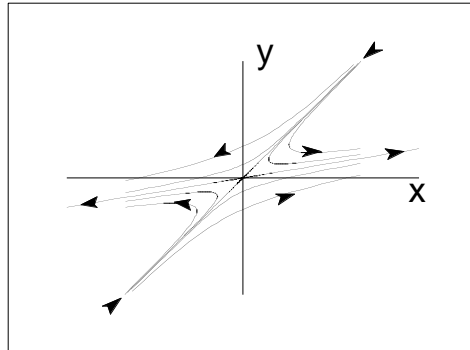


Figura 3.13

Ejemplo 3.5.2

$$\begin{cases} x' &= -5x + 4y \\ y' &= -3x + 2y \end{cases}$$

En este caso los autovalores son $\lambda_1 = -2$, $\lambda_2 = -1$, siendo el origen un nodo estable. Los autovectores correspondientes son:

$$\begin{aligned} \lambda_1 &\rightarrow (4, 3) \\ \lambda_2 &\rightarrow (1, 1) \end{aligned}$$

Todas las órbitas tienden al origen cuando t tiende a infinito, y todas entran tangentes a las órbitas correspondientes al autovalor -1 , el mayor de los dos, salvo las correspondientes al autovalor -2 , que lo hacen cada una desde un lado de la recta del autovalor -1 . El dibujo del campo de vectores en algunos puntos puede ayudar a obtener el mapa de fases (figura 3.19).

Ejemplo 3.5.3

$$\begin{cases} x' &= -6x + 4y \\ y' &= -4x + 2y \end{cases}$$

Hay un autovalor doble $\lambda = -2$. Se trata de un nodo estable de una tangente, que es el único autovector que hay: $(1, 1)$. Todas las órbitas entran en el origen cuando t tiende a infinito, tangentes a la diagonal del primer cuadrante. Si dibujamos el campo de vectores del sistema en algunos puntos se puede obtener el mapa de fases representado en la figura 3.20.

Ejemplo 3.5.4

$$\begin{cases} x' &= 3x - y \\ y' &= 2x + y \end{cases}$$

Los autovalores son ahora complejos: $\lambda = 2 \pm i$. Por lo tanto, el origen es un foco inestable. Las órbitas giran en torno al origen en sentido positivo, (pues $-1 < 0$) alejándose de él cuando t tiende a infinito (figura 3.21).

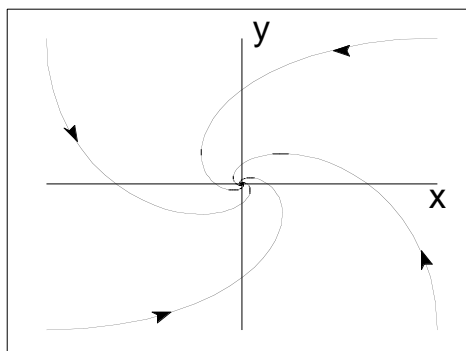


Figura 3.14

Ejemplo 3.5.5

$$\begin{cases} x' &= x + 5y \\ y' &= -x - y \end{cases}$$

Los autovalores son $\lambda = \pm 2i$. Se trata de un centro. Las órbitas son elipses que giran en torno al origen en sentido negativo ($5 > 0$). La ecuación de las órbitas es, como hemos visto:

$$5y^2 + x^2 + 2xy = C$$

donde C es una constante positiva. Los ejes de las elipses se pueden obtener de:

$$\tan 2\theta = -\frac{2a}{b+c} = -\frac{1}{2}$$

de donde:

$$\theta = -0.23, \quad \theta = 1.34$$

El primero de ellos corresponde a la localización del eje mayor, y el segundo a la del eje menor (figura 3.22).

3.6 Puntos críticos no elementales

Supongamos que uno de los dos autovalores de la matriz A es cero. Sea λ el otro autovalor no nulo y consideremos la matriz en su forma canónica:

$$A = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}$$

Los puntos del eje x , correspondiente al núcleo del operador A , son todos puntos críticos. El campo vectorial es paralelo al eje y , $(0, \lambda y)$, y se dirige hacia el eje x o en sentido contrario, según λ sea negativo o positivo. Las soluciones explícitas son:

$$\begin{aligned} x(t) &= a \\ y(t) &= be^{\lambda t} \end{aligned}$$

El mapa de fases aparece en la figura 3.23.

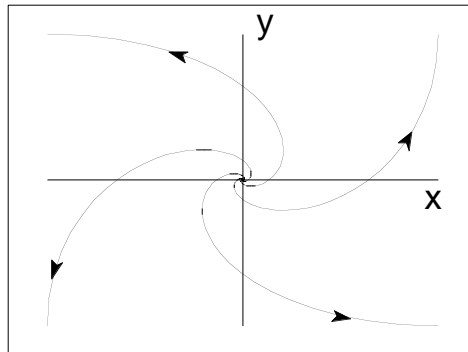


Figura 3.15

Si los dos autovalores son cero, aparte del caso trivial en el que la matriz A es cero, la forma canónica es:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Todos los puntos situados en el eje x son críticos en este caso (nuevamente los situados en el núcleo del operador A). Las soluciones explícitas son:

$$\begin{aligned} x(t) &= at + b \\ y(t) &= a \end{aligned}$$

por lo que las órbitas son paralelas al eje x y no tienden a ningún punto crítico (figura 3.24).

En \mathbb{R}^n la situación es similar, aunque el número de posibilidades sea mayor. Sin embargo, cuando el sistema es no lineal las cosas cambian. En lo que sigue estudiaremos la estabilidad de un sistema no lineal y los casos que se presentan en el plano, aún cuando algunos conceptos sean válidos para cualquier dimensión.

3.7 Sistemas autónomos no lineales planos

Cuando nos limitamos a \mathbb{R}^2 , la teoría de los sistemas autónomos con puntos elementales (a definir más adelante) se puede hacer de manera bastante completa. Veremos como en estos casos la aproximación lineal proporciona una información muy detallada sobre el mapa de fases en un entorno del punto crítico, y otros datos, como el índice, campos de vectores, integrales primeras, existencia de ciertas órbitas periódicas aisladas etc. . . , permiten hacerse una idea cualitativa del mapa completo.

Supongamos entonces un sistema autónomo en el plano:

$$\begin{cases} x' &= f(x, y) \\ y' &= g(x, y) \end{cases}$$

donde f, g son funciones analíticas en un abierto de \mathbb{R}^2 . Los puntos críticos de este sistema se obtienen resolviendo las ecuaciones:

$$\begin{cases} f(x, y) &= 0 \\ g(x, y) &= 0 \end{cases}$$

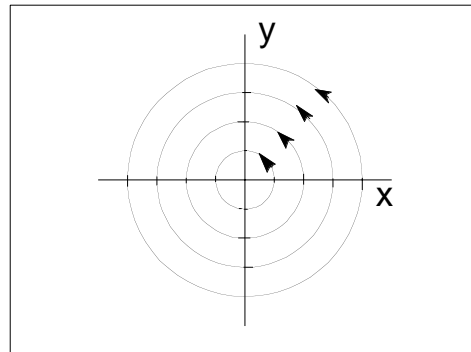


Figura 3.16

Sea (x_0, y_0) un punto crítico. Mediante una traslación podemos llevar este punto al origen. En él podemos hacer un desarrollo de Taylor hasta primer orden:

$$f(x, y) = f(0, 0) + \left[\frac{\partial f}{\partial x} \right]_0 x + \left[\frac{\partial f}{\partial y} \right]_0 y + f_2(x, y)$$

$$g(x, y) = g(0, 0) + \left[\frac{\partial g}{\partial x} \right]_0 x + \left[\frac{\partial g}{\partial y} \right]_0 y + g_2(x, y)$$

donde las funciones $f_2(x, y)$, $g_2(x, y)$ son de segundo orden en x , y , y la matriz jacobiana en el origen es:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

con $a = [\partial f / \partial x]_0$, $b = [\partial f / \partial y]_0$, $c = [\partial g / \partial x]_0$, $d = [\partial g / \partial y]_0$, las derivadas primeras calculadas en el origen.

Definición 3.7.1 *Se dice que el punto crítico es elemental si el determinante de la matriz A es diferente de cero.*

La matriz A es la matriz de la aplicación tangente, y el hecho de tener determinante distinto de cero implica ciertas consecuencias, como el que el punto crítico es aislado. El sistema anterior se puede escribir como:

$$\begin{cases} x' &= ax + by + f_2(x, y) \\ y' &= cx + dy + g_2(x, y) \end{cases}$$

Al sistema lineal autónomo cuya matriz es A se le llama sistema lineal asociado. Enunciaremos primero un teorema que establece como el sistema lineal determina el no lineal. Los puntos críticos de un sistema autónomo en el plano reciben los mismos nombres que los correspondientes de los sistemas lineales, aunque algunas características sean propias de estos últimos.

Teorema 3.7.1 *El comportamiento de las órbitas en un entorno de un punto crítico elemental es el mismo que para la primera aproximación salvo cuando los autovalores de la matriz A son imaginarios puros. En este último caso el sistema lineal tiene un centro, pero el no lineal puede tener un centro o un foco.*

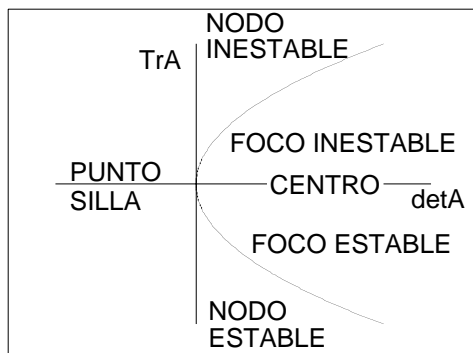


Figura 3.17

La estructura local de los mapas de fases en un entorno de un punto crítico elemental es muy sencilla y consecuencia directa del anterior teorema. La expondremos en dos teoremas concernientes a la estructura de los puntos silla y de los nodos no lineales.

Teorema 3.7.2 (Estructura local de un punto silla no lineal)

Sea $(0,0)$, punto crítico del sistema $x' = f(x,y)$, $y' = g(x,y)$, un punto silla (es decir, lo es en la aproximación lineal). Sean P y Q las rectas asociadas a los autovectores de la matriz de la aproximación lineal, P al autovalor negativo y Q al positivo. Entonces, existen dos únicas órbitas U_1 y U_2 que para $t \rightarrow \infty$ tienden al origen. Junto con el punto $(0,0)$ forman una curva continua, derivable y tangente en $(0,0)$ a la recta P . Existen también dos únicas órbitas V_1 y V_2 que tienden al origen cuando $t \rightarrow -\infty$ y que, junto con $(0,0)$ forman una curva continua, derivable y tangente en el origen a la recta Q . Las demás órbitas se comportan en un entorno de $(0,0)$ como en el caso lineal (figura 3.25).

Teorema 3.7.3 (Estructura local de un nodo no lineal)

Sea $(0,0)$ un punto crítico de un sistema $x' = f(x,y)$, $y' = g(x,y)$, que es un nodo estable para la aproximación lineal (el caso de un nodo inestable, u otro tipo de nodo, es similar). Sean P , Q las rectas asociadas a los autovectores de la matriz A de la aproximación lineal, Q con autovalor λ_1 , y P , λ_2 con $\lambda_2 < \lambda_1 < 0$. Todas las órbitas que pasan suficientemente próximas al origen tienden a él cuando t tiende a infinito y tienen una tangente en él. Solo hay dos de ellas que son tangentes a la recta P y se aproximan al origen desde lados opuestos de la recta Q . Las demás órbitas son tangentes a Q (figura 3.26). Para un nodo inestable el comportamiento es el mismo cuando t tiende a menos infinito.

Nótese que las rectas P y Q que aparecen en estos teoremas no son en general órbitas del sistema. Si el sistema lineal asociado presenta un nodo estelar o un nodo de una tangente, los resultados para el sistema no lineal son similares a los expuestos en el teorema anterior. Las órbitas pueden entrar (o salir) del origen con tangente arbitraria o una sola tangente.

Cuando el punto crítico del sistema lineal asociado es un centro, en el no lineal puede aparecer un foco o conservarse el centro, es decir las órbitas pueden ser abiertas (foco) o cerradas (centro). Para determinarlo hay que emplear otras técnicas, como el estudio del posible carácter hamiltoniano del sistema que trataremos más adelante o la existencia de ciertas simetrías del campo de vectores.

Ejemplo 3.7.1 Sea el sistema autónomo no lineal:

$$\begin{cases} x' &= -y \\ y' &= x + h(y) \end{cases}$$

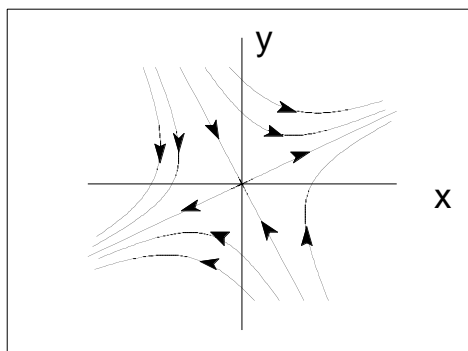


Figura 3.18

donde h es una función analítica con $h(0) = 0$ y $h'(0) = 0$. Es decir el sistema es no lineal y tiene un punto crítico en el origen. Calculando la aproximación lineal, la matriz correspondiente (siempre en el origen) es:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

es decir, tenemos un centro en la aproximación lineal.

Como veremos en la próxima sección, si h es una función par el centro se conserva, pues el segundo miembro de la primera ecuación es impar en x y el de la segunda ecuación es par en y .

Sin embargo, si la función $h(y)$ es impar, no disponemos de un criterio tan simple para determinar si el centro se conserva. En los dos ejemplos que siguen, el centro pasa a ser un foco, estable en el primer caso, inestable en el segundo:

$$\begin{cases} x' = -y \\ y' = x - y^3 \end{cases} \quad \begin{cases} x' = -y \\ y' = x + y^3 \end{cases}$$

3.8 Simetrías y campos de vectores

Una de las técnicas importantes en la solución de ecuaciones diferenciales es el estudio de las simetrías que presentan estas ecuaciones frente a transformaciones de coordenadas, cambios de variable, etc. Consideremos un abierto de \mathbb{R}^n , U , y un campo de vectores $v(x)$ definido en él. Sea $f: U \rightarrow U$ un difeomorfismo de U . El campo de vectores $v(x)$ se transforma de acuerdo con la aplicación tangente de f , que llamaremos f_* (es decir la matriz jacobiana de f):

Si $\phi(t)$ es una curva integral del campo, la curva transformada es $f(\phi(t))$ y el vector tangente en un punto $f(x_0)$ con $\phi(0) = x_0$, es:

$$f_*(v)^j(f(x_0)) = \left[\frac{d}{dt} f^j(\phi(t)) \right]_{t=0} = \sum_{i=1}^n \frac{\partial f^j}{\partial x_i}(x_0) \phi'_i(0)$$

pero $\phi'_i(0) = v(x_0)$ de donde se obtiene, escrito simbólicamente, que:

$$f_*(v) = (Df)v$$

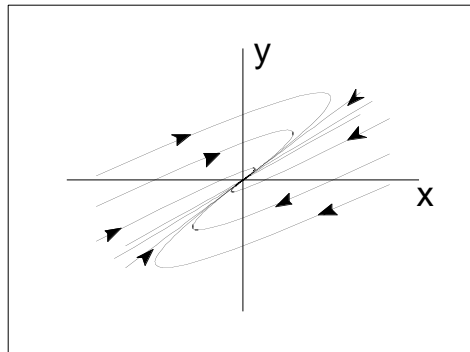


Figura 3.19

donde Df es la matriz jacobiana de la aplicación f . Evidentemente todo esto se puede escribir sin utilizar coordenadas, en un lenguaje intrínseco, pero no entraremos en este tipo de cuestiones.

La definición que usaremos de simetría es la siguiente:

Definición 3.8.1 Se dice que un difeomorfismo $g: U \rightarrow U$ es una simetría del campo de vectores $v(x)$ definido en U , si

$$v(g(x)) = g_*(v)(x), \quad \forall x \in U$$

Es decir el campo en el punto transformado es igual al campo transformado en el punto original.

Ejemplo 3.8.1 Consideremos el campo de vectores dado en el plano por:

$$v(x, y) = (-y, x)$$

y la aplicación $g(x, y)$ del plano en sí mismo dada por:

$$\begin{cases} x' &= x \cos \theta - y \operatorname{sen} \theta \\ y' &= y \operatorname{sen} \theta + x \cos \theta \end{cases}$$

La matriz jacobiana es:

$$Dg = \begin{pmatrix} \cos \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \cos \theta \end{pmatrix}$$

con lo cual tenemos:

$$v(g(x, y)) = v(x', y') = (-x \operatorname{sen} \theta - y \cos \theta, x \cos \theta - y \operatorname{sen} \theta)$$

$$g_*(v)(x, y) = (-y \cos \theta - x \operatorname{sen} \theta, -y \operatorname{sen} \theta + x \cos \theta)$$

por lo que g es una simetría del campo.

Consideremos ahora el sistema de ecuaciones diferenciales asociado al campo de vectores $v(x)$.

Definición 3.8.2 Se dice que la ecuación $x' = v(x)$ es invariante bajo el difeomorfismo g del espacio de fases U , si g es una simetría del campo de vectores. Se dice que g es una simetría de la ecuación.

Este concepto es aplicable a ecuaciones no autónomas. La consecuencia más importante de estos conceptos se contiene en el siguiente teorema.

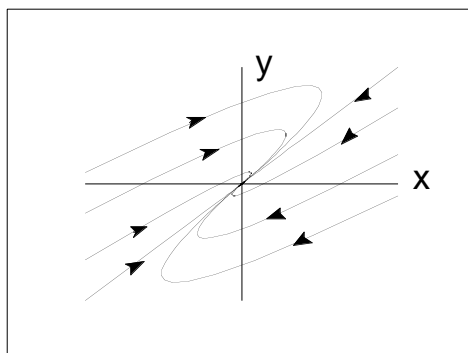


Figura 3.20

Teorema 3.8.1 Una simetría de una ecuación diferencial transforma soluciones en soluciones.

Demostración. Sea g una simetría del campo de vectores $v(x)$, y $x = \phi(t)$ una solución del sistema $x' = v(x)$. Si aplicamos el difeomorfismo g :

$$y = g(\phi(t))$$

es también solución del sistema:

$$y' = (Dg)\phi'(t) = g_*v(\phi(t)) = v(g(\phi(t))) = v(y)$$

QED

Las transformaciones de simetría de un campo de vectores forman un grupo. Cuando dependen de una manera diferenciable de uno o varios parámetros, el grupo se dice de Lie. Los grupos de Lie tienen una importancia fundamental en la teoría de ecuaciones diferenciales y en muchas otras ramas de la Matemática y de la Física en las que la simetría juega un papel.

Aunque el campo de vectores no sea invariante bajo un difeomorfismo puede intentarse una reparametrización del sistema que haga útil esa aplicación. Veamos el siguiente ejemplo.

Ejemplo 3.8.2 Sea el campo en \mathbb{R}^2 , $v(x, y) = (x + y^2, x^2 - y)$, y la reflexión respecto a la bisectriz del segundo cuadrante:

$$g(x, y) = (-y, -x)$$

Calculando $g_*(v)(x, y)$ y $v(g(x, y))$ tenemos:

$$g_*(v)(x, y) = (y - x^2, -x - y^2)$$

$$v(g(x, y)) = (-y + x^2, x + y^2)$$

por lo que:

$$g_*(v)(x, y) = -v(g(x, y))$$

De esta forma, el campo no es invariante respecto a esta reflexión, pero se transforma de manera sencilla. Si consideramos el sistema asociado:

$$\begin{cases} x' &= x + y^2 \\ y' &= x^2 - y \end{cases}$$

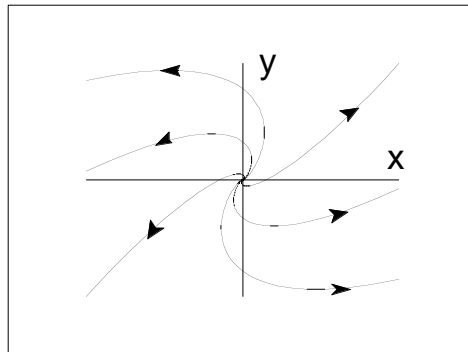


Figura 3.21

el cambio t por $-t$ nos da:

$$\begin{cases} x' &= -x - y^2 \\ y' &= -x^2 + y \end{cases}$$

donde ahora $'$ significa derivada respecto a $-t$. Está claro que el mapa de fases de ambos sistemas es el mismo, pero el sentido en el que se recorren las órbitas es opuesto. Es decir, si $(x(t), y(t))$ es una solución del primer sistema, $(x(-t), y(-t))$ lo es del segundo. De esta forma, prescindiendo del sentido en el que se recorren las órbitas, el mapa de fases es simétrico respecto de la bisectriz del segundo cuadrante. Desde un punto de vista más formal, se considera una transformación de la variable independiente $t \rightarrow h(t)$ y la transformación inducida en el sistema, ampliando de esta forma la definición de sistema invariante.

Veamos ahora como estos criterios de simetría permiten determinar en algunos casos cuando un centro se conserva. Como ya hemos visto, cuando en la aproximación lineal de un sistema autónomo en \mathbb{R}^2 aparece un centro, en el problema completo puede aparecer un centro o un foco. Si probamos, por ejemplo, que las órbitas son simétricas respecto una recta que pasa por el punto crítico, éstas serán necesariamente cerradas (al menos las próximas al punto crítico), que será un centro. Cuando la recta es uno de los ejes, y el punto crítico es el origen, el criterio de simetría se escribe de la siguiente forma:

Proposición 3.8.1 *Sea el sistema autónomo:*

$$\begin{cases} x' &= f(x, y) \\ y' &= g(x, y) \end{cases}$$

que tiene un punto crítico en el origen, centro en la aproximación lineal. Si la función f es impar en y , g par en y , o bien si f es par en x , g impar en x , entonces, el centro se conserva.

Demostración. Consideremos el primer caso:

$$\begin{aligned} f(x, -y) &= -f(x, y) \\ g(x, -y) &= g(x, y) \end{aligned}$$

Es decir, bajo una reflexión respecto al eje x , el campo $v = (f, g)$ se transforma como:

$$S: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$S(x, y) = (x, -y)$$

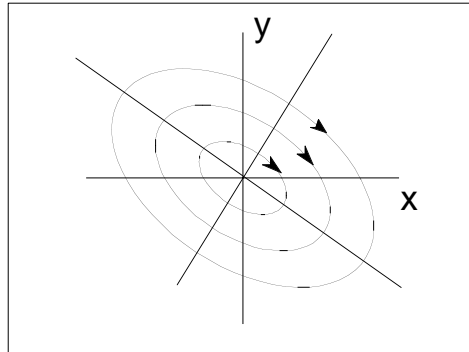


Figura 3.22

de donde:

$$S_*(v)(x, y) = (f(x, y), -g(x, y))$$

$$v(S(x, y)) = v(x, -y) = (f(x, -y), g(x, -y)) = (-f(x, y), g(x, y))$$

es decir:

$$S_*(v) = -v(S)$$

y por los mismos razonamientos que antes, las órbitas son simétricas respecto al eje x . De la misma forma se prueba el resultado para la segunda condición. En este último caso, el mapa de fases es simétrico respecto al eje y . QED

Nótese que no se indica que las órbitas son cerradas simplemente porque el mapa sea simétrico respecto a un eje. Lo que ocurre es que las órbitas próximas al origen giran en torno a él, ya que este punto es un centro o un foco; al ser simétricas son cerradas y el punto crítico es centro. Cuando la simetría no es respecto a un eje las condiciones no son tan sencillas de expresar.

Ejemplo 3.8.3 Sea el sistema:

$$\begin{cases} x' &= -y \\ y' &= x + h(y) \end{cases}$$

donde $h(y)$ es una función analítica, par y con $h(0) = h'(0) = 0$. Como ya hemos visto, el origen es un punto crítico para este sistema y es un centro en la aproximación lineal. Aplicando la proposición anterior, la función $f(x, y) = -y$, es impar en y , la función $g(x, y) = x + h(y)$ es par en y . Por lo tanto se trata de un centro.

Ejemplo 3.8.4 Sea el sistema:

$$\begin{cases} x' &= x + y^2 \\ y' &= x^2 - y \end{cases}$$

La aproximación lineal de este sistema en el punto crítico $(-1, 1)$ tiene un centro. Las funciones $f(x, y) = x + y^2$, y $g(x, y) = x^2 - y$ no verifican ninguno de los criterios de paridad dados en la proposición anterior. Sin embargo, con el difeomorfismo considerado anteriormente $S(x, y) = (-y, x)$, el sistema tiene órbitas simétricas respecto la bisectriz del segundo cuadrante, así que las próximas al punto $(-1, 1)$ son cerradas y se trata de un centro.

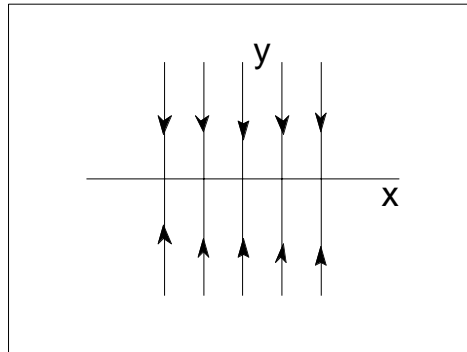


Figura 3.23

3.9 Integrales primeras

Aunque introducidas de manera informal en apartados anteriores, estudiaremos con más detalle la definición de integral primera y las propiedades que nos interesan en relación con los sistemas autónomos.

Definición 3.9.1 Dado un sistema autónomo $x' = f(x)$ con espacio de fases M , se dice que una función diferenciable $F: U \rightarrow \mathbb{R}$, U abierto de M es una integral primera si la derivada de F con respecto al campo de vectores $f(x)$ es cero.

Las propiedades que necesitamos se resumen en la siguiente proposición:

Proposición 3.9.1 Sea $F(x)$ una integral primera del sistema $x' = f(x)$

La función $F(x)$ es constante sobre cada solución $x: I \rightarrow U$ del sistema. Cada órbita del sistema pertenece a una sola curva de nivel de la función $F(x)$. Es decir, si $x: I \rightarrow U$ es una solución, la órbita correspondiente $\{x(t) : t \in I\}$ verifica: $F(x(t)) = c, \forall t \in I$.

La demostración es inmediata. Estas dos propiedades son equivalentes a la definición dada de integral primera. En cuanto concierne a los sistemas autónomos con integrales primeras, el siguiente teorema establece restricciones sobre el tipo de puntos críticos:

Teorema 3.9.1 Un sistema con integrales primeras no constantes no puede tener atractores (o repulsores), es decir, puntos críticos asintóticamente estables cuando $t \rightarrow \pm\infty$.

Ejemplo 3.9.1

$$\begin{cases} x' = x \\ y' = y \end{cases}$$

Este sistema no tiene integrales primeras (siempre supondremos no constantes). En efecto, las órbitas del sistema son las semirrectas que salen del origen. Pero sobre cada una de ellas, una posible integral primera es constante. Al tender t a menos infinito, todas tienden al origen, luego el valor de esa constante es el mismo para todas ellas, que llenan el plano, luego F es constante. Formalmente, la función:

$$F(x, y) = \frac{y}{x}$$

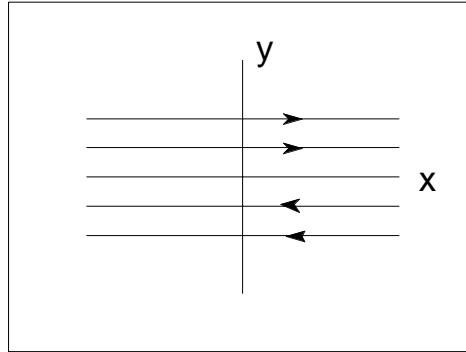


Figura 3.24

verifica que su derivada es cero a lo largo del campo de vectores:

$$L_v F(x, y) = -\frac{y}{x^2}x + \frac{y}{x} = 0$$

Pero la función F no es diferenciable, ni siquiera continua en un entorno del origen.

Ejemplo 3.9.2

$$\begin{cases} x' &= x \\ y' &= -y \end{cases}$$

El sistema admite una integral primera:

$$F(x, y) = xy$$

pues:

$$L_v F(x, y) = yx - xy = 0$$

El sistema tiene un punto silla en el origen.

Nótese que las posibles integrales primeras pueden buscarse como soluciones de la ecuación diferencial en derivadas parciales de primer orden:

$$\frac{\partial F}{\partial x} f(x, y) + \frac{\partial F}{\partial y} g(x, y) = 0$$

para un sistema con campo de vectores dado por $(f(x, y), g(x, y))$. La solución de este sistema lleva (usando el método de características que no podemos desarrollar aquí) a la ecuación de las órbitas, por lo que las integrales primeras son solución de esa ecuación. Como, al menos desde un punto de vista teórico, la órbitas pueden representarse como una función de x, y , hay que prestar atención a las condiciones de continuidad y diferenciabilidad impuestas a las integrales primeras.

Ejemplo 3.9.3 Sistemas conservativos con un grado de libertad.

Consideremos la ecuación de Newton:

$$x'' = f(x)$$

donde $f(x)$ es una función suficientemente diferenciable. Esta fuerza depende solo de la posición, luego existe una función potencial $V(x)$ tal que:

$$f(x) = -\frac{dV}{dx}$$

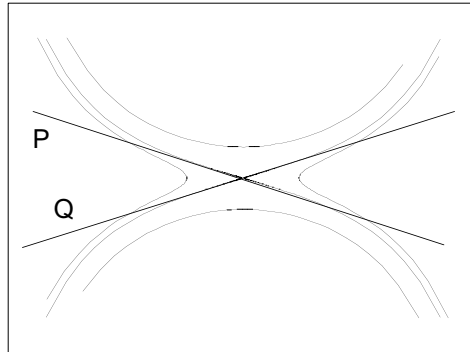


Figura 3.25

La energía total del sistema se conserva a lo largo de una trayectoria, con lo que tenemos una integral primera:

$$H(x, v) = \frac{1}{2}v^2 + V(x)$$

siendo $v = x'$ la velocidad. Los únicos tipos de puntos críticos para este sistema son centros y puntos silla (si son elementales). Además, las órbitas están contenidas en las curvas de nivel de la función $H(x, v)$. Como sabemos, un punto crítico corresponde a un mínimo del potencial:

$$\frac{dV}{dx} = 0$$

y el valor de la derivada segunda en ese punto nos da el tipo de punto crítico:

1. Si es positiva, se trata de un mínimo, es decir de un centro.
2. Si es negativa, tenemos un máximo, punto silla.
3. Si es cero, nos encontramos con un punto crítico no elemental,

una clasificación inmediata a partir, por ejemplo, del teorema de Lagrange, o del estudio de la matriz de la aproximación lineal: si el sistema es ($y = v$):

$$\begin{cases} x' &= v \\ v' &= f(x) \end{cases}$$

los puntos críticos son: $f(x_0) = 0$, $v_0 = 0$. La matriz de la aproximación lineal es:

$$\begin{pmatrix} 0 & 1 \\ f'(x_0) & 0 \end{pmatrix}$$

y los autovalores son:

$$\lambda = \pm \sqrt{f'(x_0)}$$

de donde se deduce el resultado anterior, pues $f'(x_0) = -V''(x_0)$.

Consideremos ahora dos casos particulares de la función $f(x)$:

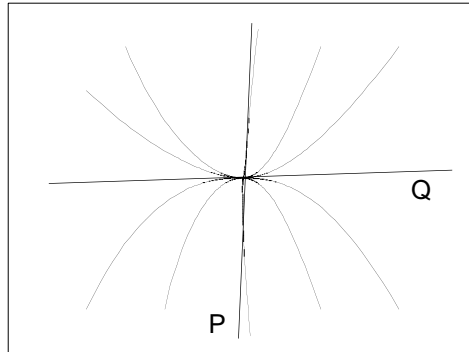


Figura 3.26

Ejemplo 3.9.4 El problema de Kepler.

La ecuación radial del problema de Kepler es:

$$x'' = -\frac{1}{x^2} + \frac{2C}{x^3}$$

con $C > 0$, de donde el potencial es:

$$V(x) = -\frac{1}{x} + \frac{C}{x^2}$$

suma del potencial coulombiano y el término centrífugo. El espacio de fase es el semiplano $x > 0$ y el único punto crítico es:

$$x = 2C, \quad v = 0$$

La segunda derivada del potencial en ese punto es:

$$\frac{d^2V}{dx^2}(2C) = -\frac{1}{8C^3}$$

y por lo tanto se trata de un punto de equilibrio estable, un centro. La energía total, que es una integral primera, es:

$$H(x, v) = \frac{1}{2}v^2 - \frac{1}{x} + \frac{C}{x^2}$$

y sus curvas de nivel $H(x, v) = E$ nos dan las órbitas del sistema (figura 3.27). No todos los valores de la energía están permitidos. Concretamente, $E \geq -1/4C$ para que la ecuación $H(x, v) = E$ tenga soluciones reales.

1. $E = -1/4C$. En este caso, el único punto que verifica $H(x, v) = E$ es $x = 1/2c$, $v = 0$, es decir el centro.
2. $-1/4C < E < 0$. Las órbitas son cerradas para estos valores de la energía, lo que se puede comprobar de manera gráfica (ver figura 3.27) o calculando explícitamente los dos puntos de retroceso:

$$H(x, 0) = E$$

$$x = -\frac{1}{2E}(1 \pm \sqrt{1 + 4CE})$$

que tiene dos soluciones reales distintas.

3. $E = 0$. La ecuación anterior solo tiene una solución $x = C$. La órbita es abierta, y la velocidad tiende a cero cuando $t \rightarrow \pm\infty$, pues en este caso, la posición tiende a infinito.
4. $E > 0$. Solo hay un punto de retroceso, dado por:

$$x = -\frac{1}{2E}(1 + \sqrt{1 + 4CE})$$

pues el otro valor de x es negativo. La velocidad, cuando $t \rightarrow \pm\infty$, tiende a $\pm\sqrt{2E}$.

El sentido en el que se recorren las órbitas viene dado por el sistema y el mapa de fases completo aparece en la figura 3.27.

El tiempo empleado en recorrer una órbita cerrada es el periodo de movimiento y puede calcularse (al menos dar una expresión compacta). Si en $t = t_0$ la partícula sometida a este potencial se encuentra en $x = x_0$, en el tiempo t se hallará en $x(t)$ dado por:

$$t - t_0 = \int_{x_0}^{x(t)} \frac{ds}{\sqrt{2(E - V(s))}}$$

donde $-1/4C < E < 0$, la energía de la partícula, es constante a lo largo de la trayectoria. Esta expresión es una consecuencia inmediata de la primera ecuación del sistema (o de la definición de velocidad). Para una órbita cerrada, el periodo es:

$$\frac{T}{2} = \int_a^b \frac{ds}{\sqrt{2(E - V(s))}}$$

donde a, b son los puntos de retroceso de la órbita. Se trata de una cantidad finita, que depende de la energía. Para una órbita abierta, el tiempo empleado en recorrerla es infinito, incluso con energía cero:

$$\int_a^\infty \frac{ds}{\sqrt{2(E - V(s))}} = \infty$$

con $0 \leq E$, y a el único punto de retroceso de la órbita.

La variable x es la distancia del planeta al centro de masas del sistema. Por tanto los puntos de retroceso son los de mínima y máxima separación. Las órbitas cerradas tienen dos (afelio y perihelio en el caso del sistema solar). Las órbitas abiertas tienen sólo uno. La órbita correspondiente al centro ($x = 2C$) es una órbita circular. La velocidad radial es cero en ella, y su distancia al centro de masas constante.

Ejemplo 3.9.5 El péndulo simple.

Consideremos ahora un sistema conservativo con el siguiente potencial:

$$V(x) = 1 - \cos x$$

donde x es una variable que podría tomarse en el intervalo $[0, 2\pi]$ con lo que el espacio de fases sería un cilindro. Sin embargo, supondremos que $x \in \mathbb{R}$. La ecuación correspondiente es:

$$x'' = -\sin x$$

y el sistema:

$$\begin{cases} x' &= v \\ v' &= -\sin x \end{cases}$$

La energía es una integral primera del sistema:

$$H(x, v) = \frac{1}{2}v^2 + 1 - \cos x$$

y las curvas de nivel, que contienen a las órbitas del sistema son:

$$\frac{1}{2}v^2 + 1 - \cos x = E$$

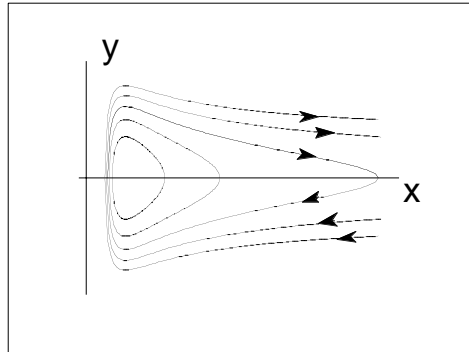


Figura 3.27

Debido a que el coseno es una función acotada, la energía es siempre positiva.

Los puntos críticos se calculan fácilmente:

$$v = 0, \quad x = n\pi, \quad n \in \mathbb{Z}$$

y corresponden a centros si n es par (mínimos del potencial) y a puntos silla si n es impar (máximos del potencial). Calculemos la matriz de la aproximación lineal en los puntos silla, para hallar los autovectores y por tanto la dirección de tangencia de las separatrices:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

por lo que los autovectores son:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

con autovalores 1 y -1 respectivamente. Con estos datos es muy sencillo dibujar el mapa de fases que aparece en la figura 3.28.

Cuando la energía es menor que 2, el movimiento es periódico (para $E = 0$ estamos en un punto de equilibrio estable). Cuando la energía es igual a 2, nos movemos por una separatriz. El movimiento no es periódico, y se tarda un tiempo infinito en recorrerla. Para $E > 2$ la órbitas son abiertas (el movimiento es periódico si se considera a x como una variable angular en $[0, 2\pi]$ y el mapa de fases sobre un cilindro).

Como antes, el periodo en las órbitas cerradas depende de la energía y se puede dar una expresión para calcularlo:

$$T = 4 \int_0^b \frac{ds}{\sqrt{2(E - 1 + \cos s)}}$$

donde b es la primera solución positiva de la ecuación:

$$\frac{1}{2}v^2 + 1 - \cos b = E$$

y $0 < E < 2$. Para una separatriz, el tiempo empleado en ir desde $x = -\pi$ hasta $x = \pi$ es infinito:

$$\int_{-\pi}^{\pi} \frac{ds}{\sqrt{2(1 + \cos s)}} = \infty$$

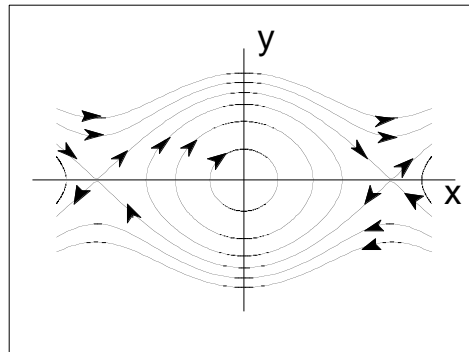


Figura 3.28

Ejemplo 3.9.6 Las ecuaciones de Lotka-Volterra.

Este sistema de ecuaciones describe las poblaciones de dos especies que comparten un mismo medio. Veamos ahora como el sistema tiene dos puntos críticos, un centro y un punto silla, y como se puede calcular una integral primera local.

$$\begin{cases} x' &= (a - by)x \\ y' &= (cx - d)y \end{cases}$$

con $a, b, c, d > 0$. Los puntos críticos son:

$$(0, 0), \quad (d/c, a/b)$$

y es inmediato comprobar que el origen es un punto silla en la aproximación lineal, que por tanto se conserva en el sistema no lineal, y que el punto $(d/c, a/b)$ es un centro en esa aproximación. Los subespacios invariantes de la matriz de la aproximación lineal para el punto silla son los ejes, que en este caso coinciden con las separatrices, como se puede ver sin más que sustituir en la ecuación. En lo que sigue, consideraremos únicamente el primer cuadrante, ya que la población de las especies es positiva (o cero). En la figura 3.29 aparece el campo de vectores del sistema.

El problema ahora es estudiar si el punto $(d/c, a/b)$ es un centro o un foco. Vamos a encontrar una integral primera, con lo cual será un centro. Para ello escribamos la ecuación que debería verificar una función $H(x, y)$ para ser integral primera del sistema:

$$L_v H(x, y) = \frac{\partial H}{\partial x}(a - by)x + \frac{\partial H}{\partial y}(cx - d)y = 0$$

Esta es una ecuación en derivadas parciales de primer orden, cuya solución general no es trivial de encontrar. Sin embargo, únicamente necesitamos una solución particular. Busquémosla usando el método de separación de variables. Sea:

$$H(x, y) = H_1(x) + H_2(y)$$

Sustituyendo en la ecuación se obtiene:

$$\frac{xH_1'(x)}{cx - d} = -\frac{yH_2'(y)}{a - by}$$

El primer miembro depende solo de x y el segundo solo de y , por lo tanto ambos son iguales a una constante λ . Esto nos proporciona dos ecuaciones diferenciales ordinarias que se pueden resolver:

$$H_1(x) = \lambda(cx - d \log x)$$

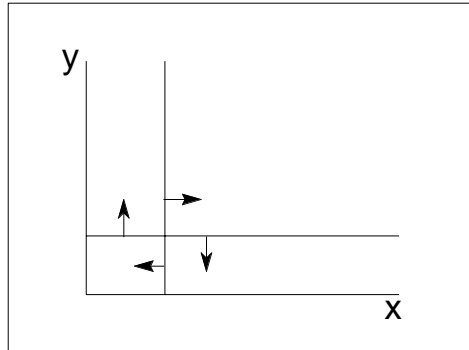


Figura 3.29

$$H_2(y) = \lambda(by - a \log y)$$

y por tanto, una posible integral primera del sistema es:

$$H(x, y) = cx + by - \log x^d y^a$$

tomando por ejemplo $\lambda = 1$. $H(x, y)$ no es una integral primera global, pues no está definida sobre los ejes. Sin embargo sí lo es en un entorno del punto crítico $(d/c, a/b)$ que es por tanto un centro.

Las órbitas próximas al centro son cerradas. Las correspondientes a los ejes son obviamente abiertas. Se puede probar que éstas son las únicas órbitas abiertas. Todas las demás son cerradas (recuérdese que estamos considerando el primer cuadrante solamente). La demostración puede hacerse probando en primer lugar que las órbitas giran en torno al centro (salvo separatrices). Este resultado no se puede deducir directamente del mapa de fases (recuérdese el caso del problema de Kepler y compárese los campos de vectores en uno y otro ejemplo). Una vez calculado el giro de las órbitas, se usa de manera adecuada el hecho de que el centro es un mínimo absoluto para la integral primera calculada antes.

De acuerdo con el mapa de fases que aparece en la figura 3.30 si en un instante dado existen las dos especies, esto es cierto en cualquier tiempo. Las poblaciones son periódicas, y salvo en el caso en el que nos encontremos en el centro, el número de individuos varía con el tiempo. Este tipo de modelos ecológicos puede mejorarse introduciendo otros factores que afectan a la evolución de las especies en competición.

3.10 Sistemas hamiltonianos

Volvamos al caso de los sistemas conservativos en una dimensión. Supongamos que existe una función $H(x, y)$, C^1 en algún dominio de \mathbb{R}^2 , tal que las ecuaciones se pueden escribir como:

$$\begin{aligned} x' &= \frac{\partial H}{\partial y} \\ y' &= -\frac{\partial H}{\partial x} \end{aligned}$$

Se dice que el sistema es hamiltoniano y que H es el hamiltoniano del sistema. Una condición necesaria para que un sistema sea hamiltoniano es la siguiente. Sea el sistema:

$$\begin{cases} x' &= f(x, y) \\ y' &= g(x, y) \end{cases}$$

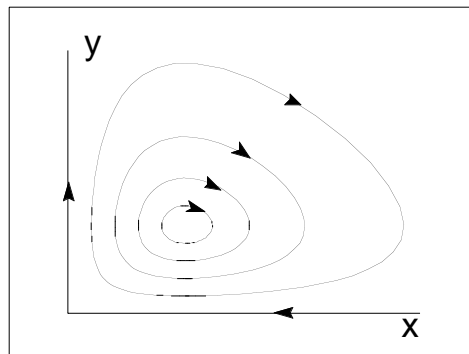


Figura 3.30

Entonces:

$$\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

que es también una condición suficiente cuando el abierto en el que trabajamos es simplemente conexo.

Supongamos pues que tenemos un sistema hamiltoniano. La primera propiedad interesante es que $H(x, y)$ es una integral primera del sistema:

$$L_v H(x, y) = \frac{\partial H}{\partial x} f(x, y) + \frac{\partial H}{\partial y} g(x, y) = 0$$

ya que:

$$\begin{aligned} f(x, y) &= \frac{\partial H}{\partial y} \\ g(x, y) &= -\frac{\partial H}{\partial x} \end{aligned}$$

Este par de ecuaciones permite calcular $H(x, y)$ si la condición establecida anteriormente se cumple. Nótese que es la misma que obtuvimos cuando se estudiaron las ecuaciones exactas. En efecto, la ecuación de las órbitas de un sistema hamiltoniano es exacta. Por eso llamaremos también a este tipo de sistemas, exactos. El cálculo de $H(x, y)$ es el de la función cuyas curvas de nivel contienen a las órbitas del sistema.

En el caso de un sistema lineal, los flujos hamiltonianos se caracterizan por la condición:

$$\text{tr } A = 0$$

donde A es la matriz del sistema. En este caso, $f(x, y) = ax + by$, $g(x, y) = cx + dy$, por lo que la condición de ser hamiltoniano es:

$$a + d = 0$$

La ecuación de autovalores es ahora:

$$\lambda^2 + \det A = 0$$

por lo que, si el determinante es positivo tendremos un centro y si es negativo un punto silla (de acuerdo con lo que sabemos de un sistema con integrales primeras, no puede tener atractores ni repulsores). Si el determinante es igual a cero los puntos críticos son no elementales.

El hamiltoniano se puede calcular:

$$\begin{aligned}\frac{\partial H}{\partial y} &= ax + by \\ -\frac{\partial H}{\partial x} &= cx - ay\end{aligned}$$

de donde:

$$H(x, y) = \frac{1}{2}(by^2 - cx^2 + 2axy)$$

y sus curvas de nivel son cónicas, elipses si b y c tienen signos opuestos (determinante positivo, centro) o hipérbolas si tienen el mismo signo (determinante negativo, punto silla).

Ejemplo 3.10.1 Los sistemas conservativos en una dimensión son obviamente hamiltonianos. Sea la ecuación:

$$x'' = x^3 - x$$

El hamiltoniano asociado es simplemente en este caso la energía cinética más la potencial:

$$H(x, y) = \frac{1}{2}y^2 + \int_0^x (s - s^3)ds = \frac{p^2}{2} + \frac{x^2}{2} - \frac{x^4}{4}$$

donde $y = x'$ y las órbitas del sistema están contenidas en las curvas de nivel de $H(x, y)$:

$$2y^2 + 2x^2 - x^4 = 4E$$

El sistema tiene tres puntos críticos correspondientes a los extremos del potencial:

$$\frac{dV}{dx} = x(1 - x^2) = 0$$

de donde $x = -1, 0, 1$. Los situados en $x = \pm 1$ son máximos, y por tanto puntos silla, y $x = 0$ es un mínimo, es decir un centro. Como siempre en estos sistemas que proceden de una ecuación de segundo orden (es decir cuya primera ecuación es $x' = y$), en los puntos críticos $y = 0$.

Las órbitas son cerradas en un entorno del centro (correspondientes a energías entre 0 y 1/4). En cada curva de nivel de la función $H(x, y)$ hay tres órbitas, una cerrada y dos abiertas (ver figura 3.31). Por encima de 1/4 las órbitas son abiertas (y hay dos órbitas en cada curva de nivel). Cuando $E = 1/4$ las órbitas que aparecen son las seis separatrices (los puntos silla tienen dos separatrices comunes) y los puntos silla mismos. La ecuación de esta curva de nivel es:

$$2y^2 + 2x^2 - x^4 = 1$$

es decir:

$$2y^2 = (x^2 - 1)^2$$

que son dos parábolas simétricas:

$$y = \pm \frac{1}{\sqrt{2}}(x^2 - 1)$$

Ejemplo 3.10.2 Volvamos de nuevo al sistema:

$$\begin{cases} x' &= x + y^2 \\ y' &= x^2 - y \end{cases}$$

Se trata de un sistema hamiltoniano (lo que nos permite asegurar sin más que el centro de la aproximación lineal se conserva). El hamiltoniano no es ahora tan trivial de escribir como en el caso anterior, pues

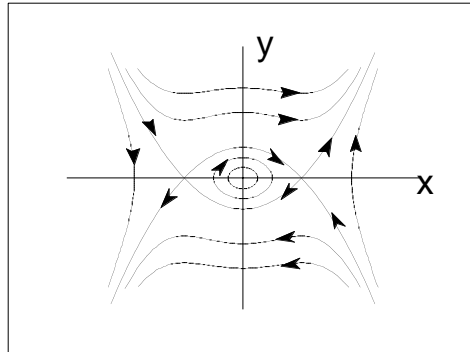


Figura 3.31

el sistema no procede directamente de una ecuación de segundo orden. En cualquier caso, resulta muy sencillo resolver:

$$\begin{aligned}\frac{\partial H}{\partial y} &= x + y^2 \\ -\frac{\partial H}{\partial x} &= x^2 - y\end{aligned}$$

obteniéndose:

$$H(x, y) = \frac{1}{3}(y^3 - x^3 + 3xy)$$

Como se ve en este caso, la segunda variable y no es la derivada de la primera x con respecto a t , impidiéndonos una interpretación tan sencilla como en el primer ejemplo (figura 3.32).

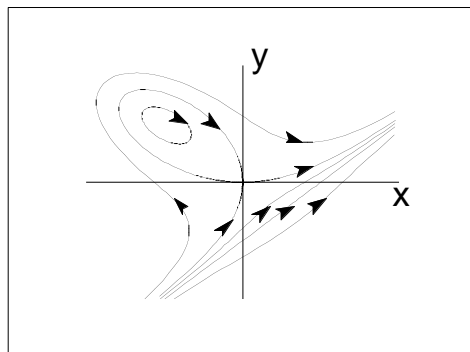


Figura 3.32

Tema 4

Soluciones en forma de serie

4.1 Introducción

Supongamos que queremos calcular las soluciones de $x'' + x = 0$ en un entorno del origen. Si éstas admiten un desarrollo en serie de Taylor que converge en un entorno de $t = 0$, podemos pensar en calcular dicho desarrollo sustituyendo una serie de potencias en la ecuación y hallando los coeficientes de esta serie. Tomemos entonces:

$$x(t) = \sum_{n=0}^{\infty} a_n t^n$$

e introduzcamos la serie en la ecuación. Para ello necesitamos que la serie se pueda derivar término a término. Si la solución es analítica en un entorno de cero, esto se puede hacer sin problemas tantas veces como queramos. Supongamos por ahora que es posible. De esta forma:

$$\sum_{n=0}^{\infty} n(n-1)a_n t^{n-2} + \sum_{n=0}^{\infty} a_n t^n = 0.$$

Podemos cambiar el índice de la primera suma, para colocar ambos términos bajo el mismo símbolo de suma:

$$\sum_{n=0}^{\infty} (n+2)(n+1)a_{n+2} t^n + \sum_{n=0}^{\infty} a_n t^n = 0$$

y por tanto:

$$\sum_{n=0}^{\infty} [(n+2)(n+1)a_{n+2} + a_n] t^n = 0.$$

Los coeficientes de esta serie, puesto que se suponen válidas todas las operaciones debido a la analiticidad de las series, deben ser cero:

$$(n+2)(n+1)a_{n+2} + a_n = 0, \quad n \geq 0.$$

Esta ecuación (en diferencias finitas) se llama relación de recurrencia entre los coeficientes de la serie que representa la solución. En este caso se puede resolver. Despejando:

$$a_{n+2} = -\frac{1}{(n+2)(n+1)} a_n, \quad n \geq 0$$

cuya solución se calcula fácilmente:

$$a_{2n} = \frac{(-1)^n}{(2n)!} a_0, \quad a_{2n+1} = \frac{(-1)^n}{(2n+1)!} a_1, \quad n \geq 0$$

donde a_0 y a_1 son arbitrarios, por lo que obtenemos dos soluciones independientes:

$$x_1(t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n}, \quad x_2(t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1}$$

la primera verificando $x_1(0) = 1$, $x_1'(0) = 0$, y la segunda $x_2(0) = 0$, $x_2'(0) = 1$. El radio de convergencia de estas series es infinito como se puede comprobar utilizando el criterio del cociente:

$$\frac{a_{2n} t^{2n}}{a_{2n+2} t^{2n+2}} = -\frac{(2n+2)!}{(2n)!} t^{-2} = -(2n+2)(2n+1)t^{-2}$$

que tiende a ∞ cuando n tiende a ∞ para cualquier t fijado (para la otra serie se obtiene lo mismo). Está claro en este caso que estas dos series representan el coseno y el seno de t respectivamente, soluciones de la ecuación como ya sabíamos. La sencillez del cálculo expuesto no debe engañar sobre la situación general. No es obvio que siempre existan soluciones desarrollables en forma de serie en el entorno de cualquier punto. Tampoco es obvio que las relaciones de recurrencia lleven a soluciones para los coeficientes de forma sencilla. Finalmente, este caso, en el que la serie converge en toda la recta, no es general. En lo que sigue trataremos de dar criterios para saber cuando existen estos desarrollos en serie y como se pueden calcular.

4.2 Ecuaciones diferenciales lineales de segundo orden

Aunque muchas de las ideas que se desarrollarán son válidas para ecuaciones de orden arbitrario, nos limitaremos al caso de orden 2. Desde el punto de vista de aplicaciones, éste es uno de los casos más interesantes y que se presenta con mayor frecuencia. Las ecuaciones no lineales no son tratables en general por estos métodos, debido a que las relaciones de recurrencia, en el caso de que existan, son demasiado complicadas. Consideremos por tanto la siguiente ecuación:

$$R(t)x'' + P(t)x' + Q(t)x = 0$$

donde $R(t)$, $P(t)$ y $Q(t)$ son funciones de t a las que se someterá a diversas restricciones. Queremos calcular las soluciones de esta ecuación mediante desarrollos en serie válidos en un entorno de un punto t_0 . De acuerdo con los teoremas de existencia y unicidad estudiados, si $R(t)$, $P(t)$ y $Q(t)$ son continuas en un entorno de t_0 y $R(t)$ no se anula en t_0 , existirá una única solución satisfaciendo unas determinadas condiciones iniciales. Si dividimos por $R(t)$ la ecuación es ahora:

$$x'' + p(t)x' + q(t)x = 0.$$

Puesta de esta forma, la condición suficiente para la existencia y unicidad de la solución es la continuidad de $p(t)$ y $q(t)$. Sin embargo, para poder hacer desarrollos en serie, exigiremos más propiedades a estas funciones. De acuerdo con ellas, los puntos se clasifican de la siguiente forma:

Definición 4.2.1 Se dice que t_0 es un punto ordinario para la ecuación:

$$x'' + p(t)x' + q(t)x = 0$$

si $p(t)$ y $q(t)$ son analíticas en un entorno de t_0 . En caso contrario, se dice que t_0 es un punto singular.

Ejemplo 4.2.1 El punto $t_0 = 0$ es un punto ordinario para la ecuación:

$$(1 - t^2)x'' - 2tx' + \alpha(\alpha + 1)x = 0$$

En cambio, los puntos $t_0 = 1$ y $t_0 = -1$ son singulares.

Ejemplo 4.2.2 En la ecuación $t^2x'' - 2tx' + 2x = 0$, $t_0 = 0$ es un punto singular. Los coeficientes $p(t) = -2/t$ y $q(t) = 2/t^2$ no son analíticos en 0. Aun así, todas las soluciones de esta ecuación son analíticas en $t_0 = 0$, e incluso en toda la recta: son combinaciones lineales de $x_1(t) = t$ y $x_2(t) = t^2$. Pero las soluciones de $4t^2x'' + 4tx' - x = 0$ no son analíticas en $t_0 = 0$: son combinaciones lineales de $t^{1/2}$ y $t^{-1/2}$.

Las ecuaciones de Euler (como estas dos) son prototipos de los casos que se presentan en general para una cierta clase de puntos singulares.

Para los puntos ordinarios se tiene el siguiente resultado:

Teorema 4.2.1 *Sea t_0 un punto ordinario para la ecuación:*

$$x'' + p(t)x' + q(t)x = 0$$

es decir, $p(t)$, $q(t)$ son funciones analíticas en un cierto entorno de t_0 . Entonces, la solución general de esta ecuación es:

$$x(t) = \sum_{n=0}^{\infty} a_n(t - t_0)^n = a_0x_1(t) + a_1x_2(t)$$

con $x_1(t)$ y $x_2(t)$ soluciones linealmente independientes, analíticas en t_0 con radio de convergencia mayor o igual que el menor de los radios de convergencia de las series que representan a $p(t)$ y $q(t)$, y que verifican las condiciones iniciales:

$$x_1(t_0) = 1, x_1'(t_0) = 0, x_2(t_0) = 0, x_2'(t_0) = 1$$

Los coeficientes a_n se fijan introduciendo la serie en la ecuación.

Es decir, en los puntos ordinarios la situación es muy similar a la estudiada en el ejemplo de la ecuación del oscilador armónico tratada anteriormente.

Los coeficientes de la serie dependen de las funciones $p(t)$ y $q(t)$. Los primeros términos del desarrollo son:

$$x_1(t) = 1 - \frac{1}{2}q(0)t^2 + \frac{1}{6}[p(0)q(0) - q'(0)]t^3 + \dots$$

$$x_2(t) = t - \frac{1}{2}p(0)t^2 + \frac{1}{6}[p(0)^2 - p'(0) - q(0)]t^3 + \dots$$

Ejemplo 4.2.3 En el estudio del oscilador armónico en mecánica cuántica se llega a la siguiente ecuación:

$$-\psi'' + t^2\psi = 2E\psi.$$

Si hacemos el cambio:

$$\psi(t) = x(t)e^{-\frac{1}{2}t^2}$$

la ecuación que verifica $x(t)$ es:

$$x'' - 2tx' + \lambda x = 0$$

con $\lambda = 2E - 1$, que es la llamada ecuación de Hermite. No es sencillo calcular las soluciones de esta ecuación. Pero el método de series permite obtener desarrollos que convergen en toda la recta (pues los coeficientes de la ecuación lo hacen). En efecto, $t_0 = 0$ es un punto ordinario, así que probemos la serie:

$$x(t) = \sum_{n=0}^{\infty} a_n t^n$$

sustituyendo en la ecuación:

$$\sum_{n=0}^{\infty} n(n-1)a_n t^{n-2} - 2t \sum_{n=0}^{\infty} n a_n t^{n-1} + \lambda \sum_{n=0}^{\infty} a_n t^n = 0$$

y desplazando los índices adecuadamente:

$$\sum_{n=0}^{\infty} [(n+2)(n+1)a_{n+2} - (2n-\lambda)]t^n = 0$$

de donde se obtiene la relación de recurrencia:

$$(n+2)(n+1)a_{n+2} - (2n-\lambda)a_n = 0$$

es decir:

$$a_{n+2} = \frac{2n-\lambda}{(n+2)(n+1)}a_n, \quad n \geq 0$$

y la solución general de la ecuación es:

$$x(t) = a_0\left(1 - \frac{\lambda}{2!}t^2 - \frac{(4-\lambda)\lambda}{4!}t^4 - \frac{(8-\lambda)(4-\lambda)\lambda}{6!}t^6 + \dots\right) + \\ a_1\left(t + \frac{2-\lambda}{3!}t^3 + \frac{(6-\lambda)(2-\lambda)}{5!}t^5 + \frac{(10-\lambda)(6-\lambda)(2-\lambda)}{7!}t^7 + \dots\right)$$

Utilizando el criterio del cociente se comprueba como estas series convergen absolutamente para todo t . Es más, se puede ver como crecen en general más rápido que $\exp(t^2/2)$, con lo que las soluciones de la ecuación inicial (la del oscilador armónico cuántico) no son de cuadrado integrable (concretamente, en lo que a nosotros nos interesa, no tienden a cero en $\pm\infty$). Sin embargo, cuando λ toma ciertos valores, una de las dos series se corta y la solución es un polinomio. Según sea $\lambda/2$ entero impar o par, es la primera solución o la segunda la que se hace un polinomio. Son los polinomios de Hermite:

$$\begin{aligned} H_0(t) &= 1 \\ H_1(t) &= 2t \\ H_2(t) &= 4t^2 - 2 \\ H_3(t) &= 8t^3 - 12t \\ &\vdots \end{aligned}$$

Obsérvese que si $\lambda = 2n$, la energía es $E = n + \frac{1}{2}$, los niveles del oscilador armónico. La siguiente expresión genera los polinomios de Hermite:

$$H_n(t) = (-1)^n e^{t^2} \frac{d^n}{dt^n} e^{-t^2}, \quad n = 0, 1, 2, \dots$$

o bien:

$$H_n(t) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k n!}{k!(n-2k)!} (2t)^{n-2k}.$$

Además, la función generatriz es:

$$w(t, z) = e^{2tz - z^2} = \sum_{n=0}^{\infty} \frac{H_n(t)}{n!} z^n.$$

Es decir, las derivadas sucesivas de $w(t, z)$ con respecto a z calculadas en $t = 0$ son los polinomios de Hermite. Para demostrarlo, se puede escribir $w(t, z)$ como:

$$w(t, z) = e^{t^2} e^{-(t-z)^2}$$

y usar:

$$\frac{\partial}{\partial z} e^{-(t-z)^2} = -\frac{\partial}{\partial t} e^{-(t-z)^2}.$$

Entre los polinomios de Hermite existen una serie de relaciones de recurrencia que relacionan unos polinomios con otros y con sus derivadas:

$$H_{n+1}(t) - 2tH_n(t) + 2nH_{n-1}(t) = 0, \quad n = 1, 2, \dots$$

$$H'_n(t) = 2nH_{n-1}(t), \quad n = 1, 2, \dots$$

Su demostración es muy sencilla a partir de la función generatriz:

$$\frac{\partial w}{\partial z} = 2(t - z)w$$

y por lo tanto, usando el desarrollo de $w(t, z)$:

$$\sum_{n=0}^{\infty} \frac{1}{n!} H_{n+1}(t) z^n = \sum_{n=0}^{\infty} \frac{2t}{n!} H_n(t) z^n - \sum_{n=0}^{\infty} \frac{2n}{n!} H_{n-1}(t) z^n$$

de donde se obtiene inmediatamente la primera relación. La segunda es aún más sencilla:

$$\frac{\partial w}{\partial t} = 2zw$$

e igual que antes:

$$\sum_{n=0}^{\infty} \frac{1}{n!} H'_n(t) z^n = \sum_{n=0}^{\infty} \frac{2n}{n!} H_{n-1}(t) z^n.$$

De estas expresiones se deduce sin dificultad que los polinomios así definidos verifican la ecuación de Hermite para $\lambda = 2n$. Es decir, uno puede partir de la definición dada más arriba de $H_n(t)$ y demostrar que verifican la ecuación de Hermite.

Los polinomios de Hermite son una familia ortogonal con peso $\exp(-t^2)$ en toda la recta, es decir:

$$\int_{-\infty}^{\infty} H_n(t) H_m(t) e^{-t^2} dt = c_n \delta_{nm}$$

expresión que se demuestra de la siguiente forma. Consideremos la ecuación inicial del oscilador armónico cuántico para dos valores $\lambda = 2n, 2m$

$$\psi''_n + (2n - 1 - t^2)\psi_n = 0$$

$$\psi''_m + (2m - 1 - t^2)\psi_m = 0$$

y multipliquemos la primera por ψ_m y la segunda por ψ_n restando ambas expresiones:

$$\frac{d}{dt}(\psi_m \psi'_n - \psi_n \psi'_m) + 2(n - m)\psi_n \psi_m = 0$$

integrando desde $-\infty$ a $+\infty$ se obtiene el resultado pedido, pues ψ y sus derivadas tienden a cero en esos puntos, recuérdese que son polinomios de Hermite por $\exp(-t^2/2)$. Cuando $n = m$ se obtiene la norma de estas funciones. Toda esta teoría se puede estudiar desde el punto de vista de los problemas de contorno, ecuaciones autoadjuntas en un cierto espacio de Hilbert con un producto escalar. No podemos entrar en ese tipo de detalles aquí que serán estudiados más adelante. Calculemos sin embargo la norma. Para ello usaremos las relaciones de recurrencia dadas anteriormente:

$$H_{n+1}(t) - 2tH_n(t) + 2nH_{n-1}(t) = 0$$

$$H_n(t) - 2tH_{n-1}(t) + 2(n-1)H_{n-2}(t) = 0$$

multiplicando la primera por $H_{n-1}(t)$ y la segunda por $H_n(t)$:

$$H_{n-1}H_{n+1} - 2tH_{n-1}H_n + 2nH_{n-1}^2 = 0$$

$$H_n^2 - 2tH_nH_{n-1} + 2(n-1)H_nH_{n-2} = 0$$

restando ambas expresiones, multiplicando por $\exp(-t^2)$ e integrando, se obtiene después de usar las propiedades de ortogonalidad:

$$\int_{-\infty}^{\infty} e^{-t^2} H_n^2(t) dt = 2n \int_{-\infty}^{\infty} e^{-t^2} H_{n-1}^2(t) dt.$$

Repitiendo este proceso, se llega a:

$$\int_{-\infty}^{\infty} e^{-t^2} H_n^2(t) dt = 2^{n-1} n! \int_{-\infty}^{\infty} e^{-t^2} H_1^2(t) dt$$

y usando la expresión de $H_1(t)$ e integrando:

$$\int_{-\infty}^{\infty} e^{-t^2} H_n^2(t) dt = 2^n n! \sqrt{\pi}.$$

Otras muchas representaciones pueden darse de los polinomios de Hermite, por ejemplo como ciertas integrales de una función (relacionada con la función generatriz). Asimismo, una clase muy amplia de funciones puede ser desarrollada en serie de polinomios de Hermite. Muchas de estas propiedades son compartidas por otras familias de polinomios ortogonales, soluciones también de ecuaciones diferenciales lineales de segundo orden: ecuaciones de Legendre, Laguerre, Tchebychev,...

Ejemplo 4.2.4 Ecuación de Legendre.

Como hemos dicho antes, el punto t_0 es un punto ordinario para la ecuación:

$$(1 - t^2)x'' - 2tx' + \alpha(\alpha + 1)x = 0$$

Por tanto podemos sustituir la serie:

$$x(t) = \sum_{n=0}^{\infty} a_n t^n$$

en la ecuación, obteniendo:

$$(1 - t^2) \sum_{n=0}^{\infty} n(n-1)a_n t^{n-2} - 2t \sum_{n=0}^{\infty} n a_n t^{n-1} + \alpha(\alpha + 1) \sum_{n=0}^{\infty} a_n t^n = 0$$

de donde se saca la relación de recurrencia:

$$(n+2)(n+1)a_{n+2} - [n(n+1) - \alpha(\alpha + 1)]a_n = 0, \quad n \geq 0$$

es decir:

$$a_{n+2} = \frac{n(n+1) - \alpha(\alpha + 1)}{(n+2)(n+1)} a_n, \quad n \geq 0$$

por tanto, la solución general del problema se puede representar en un entorno de $t = 0$ por las series siguientes:

$$a_0 \left[1 + \sum_{n=1}^{\infty} (-1)^n \frac{\alpha(\alpha-2) \dots (\alpha-2n+2)(\alpha+1)(\alpha+3) \dots (\alpha+2n-1)}{(2n)!} t^{2n} \right] +$$

$$a_1 \left[t + \sum_{n=1}^{\infty} (-1)^n \frac{(\alpha-1)(\alpha-3) \dots (\alpha-2n+1)(\alpha+2)(\alpha+4) \dots (\alpha+2n)}{(2n+1)!} t^{2n+1} \right]$$

Puesto que las funciones $p(t) = -2t/(1-t^2)$ y $q(t) = \alpha(\alpha+1)/(1-t^2)$ son analíticas en un entorno de $t = 0$ y las series que las representan tienen radio de convergencia 1, las soluciones tienen por lo menos radio de convergencia igual a 1. Así es en este caso. El radio de convergencia de las series es 1 y las dos divergen bien en 1 o en -1 . En realidad no siempre ocurre esto. En efecto, si α es un número entero positivo, una de las dos series se corta y se obtiene un polinomio: son los polinomios de Legendre. Si $\alpha = 2n$, es la primera serie, obteniéndose un polinomio par (la segunda serie diverge en -1 o 1). Si $\alpha = 2n + 1$, es la segunda serie la que da el polinomio, de grado impar. Los primeros polinomios son:

$$\begin{aligned} P_0(t) &= 1 \\ P_1(t) &= t \\ P_2(t) &= \frac{1}{2}(3t^2 - 1) \\ P_3(t) &= \frac{1}{2}(5t^3 - 3t) \\ &\vdots \end{aligned}$$

Como en los polinomios de Hermite, también podemos encontrar una fórmula que dé la expresión general (fórmula de Rodrigues):

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n, \quad n = 0, 1, 2, \dots$$

y una función generatriz:

$$w(t, z) = (1 - 2tz + z^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(t) z^n$$

válida para t suficientemente pequeño. Los polinomios de Legendre son ortogonales en el intervalo $[-1, 1]$:

$$\int_{-1}^1 P_n(t) P_m(t) dt = \frac{2}{2n+1} \delta_{nm}$$

y verifican una serie de relaciones de recurrencia:

$$(n+1)P_{n+1}(t) - (2n+1)tP_n(t) + nP_{n-1}(t) = 0, \quad n = 1, 2, \dots$$

$$P'_{n+1}(t) - P'_{n-1}(t) = (2n+1)P_n(t), \quad n = 1, 2, \dots$$

$$(1-t^2)P'_n(t) = nP_{n-1}(t) - nP_{n+1}(t), \quad n = 1, 2, \dots$$

La demostración de estas propiedades sigue las líneas de la hecha para los polinomios de Hermite.

Uno podría pensar a la vista de estos ejemplos, que la relación de recurrencia siempre es de dos términos con índices que difieren en dos unidades. El que esto es una falsa impresión, se puede ver en el siguiente caso.

Ejemplo 4.2.5 Ecuación de Airy: $x'' = tx$. El punto $t = 0$ vuelve a ser ordinario, y por lo tanto, de acuerdo con el teorema anterior, podemos sustituir la misma serie:

$$\sum_{n=0}^{\infty} n(n-1)a_n t^{n-2} = t \sum_{n=0}^{\infty} a_n t^n$$

de donde se obtiene la relación de recurrencia:

$$2a_2 = 0$$

$$(n+3)(n+2)a_{n+3} - a_n = 0, \quad n \geq 0.$$

Como vemos la relación de recurrencia incluye aún dos términos, pero la diferencia entre índices es 3. La solución de esta ecuación es:

$$a_{n+3} = \frac{1}{(n+3)(n+2)} a_n, \quad n \geq 0$$

de donde:

$$a_{3k} = \frac{1}{3k(3k-1)(3k-3)(3k-4)\dots 6.5.3.2} a_0, \quad k = 1, 2, \dots$$

$$a_{3k+1} = \frac{1}{(3k+1)(3k)(3k-2)(3k-3)\dots 7.6.4.3} a_1, \quad k = 1, 2, \dots$$

$$a_{3k+2} = 0, \quad k = 0, 1, 2, \dots$$

que se pueden escribir usando la función gamma:

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt, \quad \Re(z) > 0.$$

Esta función juega un papel fundamental en muchas de las ecuaciones que veremos a continuación, por lo que resumimos aquí sus propiedades más importantes:

1. La función anterior se puede extender por prolongación analítica a todo el plano complejo, proporcionando una función meromorfa con polos simples en $0, -1, -2, \dots$:

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} e^{-t} t^{z-1} dt, \quad z \neq 0, -1, -2, \dots$$

2. La función gamma generaliza el factorial en el siguiente sentido:

$$\Gamma(z+1) = z\Gamma(z)$$

3. Otras propiedades que nos serán útiles son:

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\operatorname{sen} \pi z}$$

$$2^{2z-1}\Gamma(z)\Gamma(z+\frac{1}{2}) = \pi^{\frac{1}{2}}\Gamma(2z)$$

Todas ellas se demuestran fácilmente a partir de la definición.

Aplicando esto al cálculo anterior:

$$3k(3k-1)(3k-3)(3k-4)\cdots 3 \cdot 2 = 3^{2k} [k(k-1)\cdots 2 \cdot 1] [(k-\frac{1}{3})(k-\frac{4}{3})\cdots \frac{5}{3} \frac{2}{3}] = 3^{2k} k! \frac{\Gamma(k+\frac{2}{3})}{\Gamma(\frac{2}{3})}$$

$$(3k+1)(3k)(3k-2)(3k-3)\cdots 4 \cdot 3 = 3^{2k} [k(k-1)\cdots 2 \cdot 1] [(k+\frac{1}{3})(k-\frac{2}{3})\cdots \frac{7}{3} \frac{4}{3}] = 3^{2k} k! \frac{\Gamma(k+\frac{4}{3})}{\Gamma(\frac{4}{3})}.$$

La solución general es:

$$x(t) = a_0 \sum_{k=0}^{\infty} \frac{\Gamma(\frac{2}{3})}{3^{2k} k! \Gamma(k+\frac{2}{3})} t^{3k} + a_1 \sum_{k=0}^{\infty} \frac{\Gamma(\frac{4}{3})}{3^{2k} k! \Gamma(k+\frac{4}{3})} t^{3k+1}.$$

Se definen las funciones de Airy como dos soluciones linealmente independientes de esta ecuación con las condiciones iniciales siguientes:

$$\operatorname{Ai}(0) = \frac{3^{-2/3}}{\Gamma(\frac{2}{3})}, \quad \operatorname{Ai}'(0) = \frac{3^{-4/3}}{\Gamma(\frac{4}{3})}$$

$$\operatorname{Bi}(0) = \frac{3^{-1/6}}{\Gamma(\frac{2}{3})}, \quad \operatorname{Bi}'(0) = \frac{3^{-5/6}}{\Gamma(\frac{4}{3})}.$$

Concretamente, $\operatorname{Ai}(t)$ es la única solución (salvo constante multiplicativa) de la ecuación de Airy que tiende a cero cuando t tiende a $+\infty$. Entonces:

$$\operatorname{Ai}(t) = \sum_{k=0}^{\infty} \frac{1}{3^{2k+2/3} k! \Gamma(k+\frac{2}{3})} t^{3k} - \sum_{k=0}^{\infty} \frac{1}{3^{2k+4/3} k! \Gamma(k+\frac{4}{3})} t^{3k+1}$$

$$\operatorname{Bi}(t) = \sqrt{3} \left[\sum_{k=0}^{\infty} \frac{1}{3^{2k+2/3} k! \Gamma(k+\frac{2}{3})} t^{3k} + \sum_{k=0}^{\infty} \frac{1}{3^{2k+4/3} k! \Gamma(k+\frac{4}{3})} t^{3k+1} \right].$$

Aunque no sea fácil de demostrar, la función $\operatorname{Ai}(t)$ es siempre mayor que cero. Cuando t es mayor que cero es decreciente y tiende a cero cuando t tiende a $+\infty$. Cuando t es menor que cero es oscilante. La función $\operatorname{Bi}(t)$ es también positiva. Cuando t es mayor que cero es creciente y tiende a infinito cuando t tiende a $+\infty$. Cuando t es menor que cero es también oscilante. Se puede probar que el wronskiano de estas dos soluciones (que no depende de t como prueba la fórmula de Abel-Liouville) es igual a $1/\pi$.

4.3 Puntos singulares

Cuando t_0 no es un punto ordinario de la ecuación, no siempre existen soluciones analíticas en un entorno de t_0 . Es más, éste es un comportamiento excepcional. En general aparecen polos, puntos de ramificación o singularidades esenciales. Solo en algunos casos se puede dar una teoría general que permite calcular la solución. Son los llamados puntos singulares regulares. En estos puntos las soluciones pueden tener singularidades que son polos o puntos de ramificación, pero no singularidades esenciales. La forma de estas soluciones es:

$$x(t) = w(t)\phi(t)$$

donde $\phi(t)$ es una función analítica ($\phi(t_0) \neq 0$) en un entorno de t_0 y $w(t)$ (que en ocasiones es simplemente 1) tiene un polo o un punto de ramificación en t_0 . Si alguna solución presenta una singularidad esencial, el punto se llama singular irregular y el estudio de las soluciones es más complicado. Tomaremos el siguiente criterio como la definición de punto singular regular:

Definición 4.3.1 Sea la ecuación lineal de segundo orden:

$$R(t)x'' + P(t)x' + Q(t)x = 0$$

y sean $p(t) = P(t)/R(t)$, $q(t) = Q(t)/R(t)$. Si en t_0 las funciones $p(t)$ y $q(t)$ no son analíticas, pero sí lo son:

$$(t - t_0)p(t), (t - t_0)^2q(t)$$

entonces se dice que t_0 es un punto singular regular.

Ejemplo 4.3.1 El origen es un punto singular regular para toda ecuación de Euler. Las soluciones como hemos visto antes, pueden ser analíticas en $t = 0$, pero en cualquier caso las únicas singularidades que presentan son polos o puntos de ramificación:

$$t^2x'' + atx' + bx = 0$$

Sustituyendo $x = t^r$ en la ecuación se obtiene el polinomio indicial: $r(r - 1) + ar + b = 0$.

Como veremos, la idea de polinomio indicial se generaliza a cualquier ecuación con puntos singulares regulares. Las raíces son r_1, r_2 . Se pueden presentar los siguientes casos:

1. Raíces distintas:

$$x(t) = c_1|t|^{r_1} + c_2|t|^{r_2}$$

2. Raíces iguales:

$$x(t) = c_1|t|^{r_1} + c_2|t|^{r_1} \log |t|$$

En el primer caso, si las raíces son complejas se puede escribir:

$$x(t) = c_1|t|^\alpha \cos(\beta \log |t|) + c_2|t|^\alpha \operatorname{sen}(\beta \log |t|)$$

con $r = \alpha + i\beta$.

Ejemplo 4.3.2 El punto $t = 0$ es singular regular para la siguiente ecuación:

$$t^2(t + 1)x'' - t^2x' + \frac{1}{4}(3t + 1)x = 0.$$

Una base de soluciones es:

$$x_1(t) = t^{1/2}, \quad x_2(t) = t^{1/2} \log(t) + t^{3/2}$$

que tienen un punto de ramificación en el origen.

Ejemplo 4.3.3 Aunque el punto $t = 0$ es también singular regular para la siguiente ecuación, existen soluciones que son analíticas en 0:

$$x'' + \frac{2}{t}x' + x = 0$$

Una base de soluciones es:

$$x_1(t) = \frac{\operatorname{sen} t}{t}, \quad x_2(t) = \frac{\cos t}{t}.$$

La primera de estas funciones es analítica en $t = 0$. La segunda tiene un polo de primer orden.

Ejemplo 4.3.4 El punto $t = 0$ es singular irregular para la ecuación:

$$t^3 x'' + t(1 + 3t)x' + x = 0.$$

Las funciones:

$$x_1(t) = 1 + \frac{1}{t}, \quad x_2(t) = e^{1/t}$$

son una base del espacio de soluciones de esta ecuación. Ninguna de ellas es analítica en $t = 0$. La primera tiene un polo y la segunda una singularidad esencial.

En los puntos singulares regulares puede ocurrir que las soluciones sean analíticas. Sin embargo en los singulares irregulares siempre hay alguna solución que presenta una singularidad esencial. Cuando el orden de la ecuación es mayor que 2, se pueden dar criterios similares para distinguir los puntos singulares regulares de los irregulares. El siguiente teorema proporciona la forma de las soluciones en un entorno de un punto singular regular (para facilitar la comprensión se toma $t_0 = 0$; para cualquier t_0 las conclusiones del teorema son fácilmente trasladables a ese punto):

Teorema 4.3.1 (Frobenius). Sea $t_0 = 0$ punto singular regular para la ecuación:

$$x'' + p(t)x' + q(t)x = 0$$

es decir $tp(t)$ y $t^2q(t)$ son analíticas en $t = 0$:

$$tp(t) = \sum_{n=0}^{\infty} p_n t^n, \quad t^2q(t) = \sum_{n=0}^{\infty} q_n t^n$$

donde p_0, q_0 o q_1 son distintos de cero. Sea ρ el menor de los radios de convergencia de estas dos series. Se define el polinomio indicial de la ecuación como:

$$r(r-1) + p_0 r + q_0$$

donde

$$p_0 = \lim_{t \rightarrow 0} tp(t), \quad q_0 = \lim_{t \rightarrow 0} t^2q(t)$$

y sus raíces son r_1 y r_2 (en el caso de raíces reales $r_1 \geq r_2$).

Entonces, en cada uno de los intervalos $(-\rho, 0)$ y $(0, \rho)$, la ecuación tiene dos soluciones linealmente independientes dadas por:

1. Si $r_1 - r_2$ no es un entero positivo o cero,

$$\begin{aligned} x_1(t) &= |t|^{r_1} \sum_{n=0}^{\infty} a_n(r_1) t^n \\ x_2(t) &= |t|^{r_2} \sum_{n=0}^{\infty} a_n(r_2) t^n \end{aligned}$$

donde $a_0(r_1) = a_0(r_2) = 1$.

2. Si $r_1 = r_2$,

$$\begin{aligned} x_1(t) &= |t|^{r_1} \sum_{n=0}^{\infty} a_n(r_1)t^n \\ x_2(t) &= x_1(t) \log |t| + |t|^{r_1} \sum_{n=1}^{\infty} b_n(r_1)t^n \end{aligned}$$

donde $a(r_0) = 1$. Nótese que la segunda serie empieza en t , para evitar repetir la primera solución.

3. Si $r_1 - r_2 = N$, entero positivo,

$$\begin{aligned} x_1(t) &= |t|^{r_1} \sum_{n=0}^{\infty} a_n(r_1)t^n \\ x_2(t) &= ax_1(t) \log |t| + |t|^{r_2} \sum_{n=0}^{\infty} c_n(r_2)t^n \end{aligned}$$

donde $a_0(r_1) = c_0(r_2) = 1$ y la constante a puede ser cero. Los coeficientes de las series, así como la constante a , se calculan introduciendo las series en la ecuación diferencial. Las series son convergentes al menos en $(-\rho, \rho)$, definiendo funciones analíticas en $t = 0$. Es decir, todas las singularidades están en t^r y en $\log t$, y son por tanto, como ya habíamos anunciado, polos y puntos de ramificación.

Aunque no vamos a demostrar este teorema, sí daremos una serie de indicaciones sobre sus resultados. Supongamos que $t = 0$ es un punto singular regular, por tanto $p(t)$ o $q(t)$ no son analíticas. Sustituyamos en la ecuación una solución de la forma:

$$\begin{aligned} x(t) &= t^r \sum_{n=0}^{\infty} a_n t^n \\ t^2 \sum_{n=0}^{\infty} (n+r)(n+r-1)a_n t^{n+r-2} + t \sum_{n=0}^{\infty} p_n t^n \sum_{m=0}^{\infty} (m+r)a_m t^{m+r-1} + \sum_{n=0}^{\infty} q_n t^n \sum_{m=0}^{\infty} a_m t^{m+r} &= 0 \end{aligned}$$

Los primeros términos del desarrollo son:

$$\begin{aligned} &(r(r-1) + rp_0 + q_0)a_0 t^r + \\ &((rp_1 + q_1)a_0 + ((r+1)r + (r+1)p_0 + q_0)a_1)t^{r+1} + \\ &((rp_2 + q_2)a_0 + ((r+1)p_1 + q_1)a_1 + ((r+2)(r+1) + (r+2)p_0 + q_0)a_2)t^{r+2} + \dots = 0 \end{aligned}$$

de donde se deduce:

$$r(r-1) + rp_0 + q_0 = 0$$

que es el polinomio indicial. Obviamente a_0 no puede ser cero, pues el comportamiento de la solución cerca de $t = 0$ no sería la potencia r de t sino $r+1$ u otra superior. Por eso se toma $a_0 = 1$ en el enunciado del teorema.

$$(rp_1 + q_1)a_0 + ((r+1)r + (r+1)p_0 + q_0)a_1 = 0$$

De esta segunda ecuación tenemos que calcular a_1 . Así que el coeficiente de a_1 no debería ser cero. Sea r_1 la mayor de las dos raíces del polinomio indicial, en el caso en que sean reales (si son complejas no hay problemas). Entonces, el coeficiente de a_1 no es cero:

$$(r_1 + 1)(p_0 + r) + q_0 = p_0 + 2r_1 = 1 + r_1 - r_2 > 0$$

puesto que r_1 es la mayor de las raíces (si son complejas, este coeficiente tampoco es cero). Aún en el caso en que las raíces sean iguales, ese coeficiente es positivo. Luego a_1 se puede calcular siempre en términos de a_0 . Consideremos ahora el término siguiente:

$$(rp_2 + q_2)a_0 + ((r+1)p_1 + q_1)a_1 + ((r+2)(r+p_0+1) + q_0)a_2 = 0$$

Si el coeficiente de a_2 es distinto de cero, se podrá calcular a_2 en función de a_0 y a_1 . Tomando $r = r_1$:

$$(r_1 + 2)(r_1 + p_0 + 1) + q_0 = 2(r_1 - r_2) + 4$$

que también es distinto de cero por las mismas razones anteriores. Veamos cual es en general el coeficiente de a_n en estas relaciones. El coeficiente de la potencia $n + r$ de t contiene a_0, a_1, \dots, a_n . El coeficiente de a_n es:

$$\begin{aligned} t^2 x'' &\rightarrow (r + n)(r + n - 1) \\ tp(t)x' &\rightarrow p_0(r + n) \\ t^2 q(t)x &\rightarrow q_0 \end{aligned}$$

por tanto, el coeficiente buscado es:

$$f(n, r) = (r + n)(r + n - 1) + (r + n)p_0 + q_0 = n(n + 2r + p_0 - 1)$$

y si se pone $r = r_1$:

$$f(n, r_1) = n(n + r_1 - r_2)$$

que es positivo para todo n . Por lo tanto, podemos ir calculando los coeficientes de la serie de una forma recursiva. Calculados a_0, a_1, \dots, a_n , se puede hallar a_{n+1} . Si la segunda raíz es distinta de la primera y no difiere de ella en un entero positivo, el coeficiente básico es ahora:

$$f(n, r_2) = n(n + r_2 - r_1)$$

que tampoco es cero. Luego existe una solución que tiene la forma indicada en el teorema. Es obvio que cuando $r_1 = r_2$ sólo hay una solución de ese tipo. Pero el caso más curioso aparece cuando las dos raíces difieren en un entero positivo: $r_1 - r_2 = N$. Ahora, el coeficiente básico es:

$$f(n, r_2) = n(n - N)$$

que se anula para $n = N$, es decir, el coeficiente a_N no se puede calcular en general a partir de los anteriores, y la segunda solución tiene la forma indicada en el teorema. Puede ocurrir sin embargo, que el coeficiente de t^{N+r_2} sea idénticamente nulo, en cuyo caso la segunda solución tiene $a = 0$.

Como hemos visto, siempre hay una solución de la forma:

$$x(t) = t^{r_1} \phi(t)$$

donde $\phi(t)$ es una función analítica en un entorno de cero, con $\phi(0)$ distinto de cero, y r_1 es la raíz mayor del polinomio indicial en el caso en que sean reales (o una cualquiera de las complejas en el caso que lo sean). La otra solución puede tener el mismo aspecto o presentar un logaritmo. Veamos que podemos decir de ella. Como sabemos una solución, podemos reducir el orden de la ecuación:

$$\begin{aligned} x(t) &= t^{r_1} \phi(t) \int v \\ tv' + (tp(t) + 2t \frac{\phi'}{\phi} + 2r_1)v &= 0. \end{aligned}$$

Esta ecuación tiene un punto singular regular en $t = 0$ (es inmediato trasladar la definición de puntos singulares regulares a ecuaciones de primer orden) pues el coeficiente de $v(t)$ es analítico. El polinomio indicial sería:

$$r + \psi(0) = 0$$

donde $\psi(t)$ es el coeficiente de $v(t)$ en la ecuación diferencial. Calculando se obtiene (como era de esperar):

$$r = -1 - r_1 + r_2$$

por lo que la solución es:

$$v(t) = t^{r_2-r_1-1}g(t)$$

donde $g(t)$ es analítica, con $g(0)$ distinto de cero. Sustituyendo para encontrar $x(t)$:

$$x(t) = t^{r_1}\phi(t) \int t^{r_2-r_1-1}g(t)dt = t^{r_1}\phi(t) \int [c_0t^{r_2-r_1-1} + c_1t^{r_2-r_1} + \dots]$$

donde c_0, c_1, \dots son los coeficientes del desarrollo en serie de $g(t)$. Está claro que, dependiendo de los exponentes r_1, r_2 , puede aparecer un logaritmo. Si $r_1 = r_2$, se obtiene un logaritmo. Si la diferencia $r_1 - r_2$ entre los exponentes es un entero positivo, puede aparecer un logaritmo, si el correspondiente coeficiente del desarrollo de $g(t)$ es distinto de cero. En la expresión anterior aparece condensado (y demostrado en buena parte) el teorema de Frobenius.

Ejemplo 4.3.5 La ecuación de Bessel de orden ν .

$$t^2x'' + tx' + (t^2 - \nu^2)x = 0, \quad \nu \geq 0$$

Supongamos que queremos calcular las soluciones de esta ecuación en un entorno de $t = 0$, con $t > 0$ (la ecuación de Bessel aparece asociada a la separación de variables en coordenadas cilíndricas, y allí t es el radio que es positivo). En cualquier caso, se podría considerar una variable t compleja e incluso ν compleja con $\Re\nu \geq 0$. Está claro que $t = 0$ es un punto singular regular. Los coeficientes $p(t)$ y $q(t)$ son:

$$p(t) = \frac{1}{t}, \quad q(t) = \frac{t^2 - \nu^2}{t^2}$$

por tanto:

$$p_0 = \lim_{t \rightarrow 0} tp(t) = 1, \quad q_0 = \lim_{t \rightarrow 0} t^2q(t) = -\nu^2$$

y el polinomio indicial:

$$r(r-1) + r - \nu^2$$

que tiene por raíces:

$$r_1 = \nu, \quad r_2 = -\nu.$$

Para cualquier valor de ν , existe una solución del tipo:

$$x(t) = t^\nu \sum_{n=0}^{\infty} a_n t^n.$$

Sustituyendo en la ecuación:

$$\sum_{n=0}^{\infty} (\nu+n)(\nu+n-1)a_n t^{n+\nu} + \sum_{n=0}^{\infty} (\nu+n)a_n t^{n+\nu} + \sum_{n=0}^{\infty} a_n t^{n+\nu+2} - \sum_{n=0}^{\infty} \nu^2 a_n t^{n+\nu} = 0$$

simplificando:

$$\sum_{n=0}^{\infty} n(n+2\nu)a_n t^{n+\nu} + \sum_{n=2}^{\infty} a_{n-2} t^{n+\nu} = 0$$

que lleva a las ecuaciones (recuérdese que $\nu \geq 0$):

$$\begin{aligned} a_1 &= 0 \\ n(n+2\nu)a_n + a_{n-2} &= 0, \quad n \geq 2 \end{aligned}$$

Puesto que la diferencia entre los índices de la relación de recurrencia es 2 y a_1 es igual a cero, está claro que esta solución es una función par, todos los coeficientes impares son cero. En cuanto a los índices pares:

$$a_{2n} = -\frac{a_{2n-2}}{2n(2n+2\nu)}, \quad n \geq 1$$

ecuación que puede ser fácilmente resuelta:

$$a_{2n} = \frac{(-1)^n}{2^{2n} n! (\nu + n)(\nu + n - 1) \dots (\nu + 1)} a_0, \quad n \geq 1$$

utilizando la función gamma, podemos escribir:

$$a_{2n} = \frac{(-1)^n \Gamma(\nu + 1)}{2^{2n} n! \Gamma(\nu + n + 1)} a_0, \quad n \geq 1.$$

Para cualquier a_0 se obtiene una solución de la ecuación de Bessel. Se define la función de Bessel de primera especie de orden ν , $J_\nu(t)$, con el siguiente valor de a_0 :

$$a_0 = \frac{1}{2^\nu \Gamma(\nu + 1)}$$

que nos da la solución:

$$J_\nu(t) = \left(\frac{t}{2}\right)^\nu \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \nu + 1)} \left(\frac{t}{2}\right)^{2n}.$$

Esta función, cuyo comportamiento en el origen es t^ν , es siempre una solución de la ecuación de Bessel, para cualquier ν . Calculemos ahora una segunda solución linealmente independiente de la primera. Ahora necesitamos distinguir varios casos, de acuerdo con el teorema. Si la diferencia entre las raíces del polinomio indicial, 2ν , no es un entero positivo o cero, la segunda solución es:

$$x_2(t) = t^{-\nu} \sum_{n=0}^{\infty} b_n t^n$$

de donde resulta evidente que los cálculos son los mismos que antes, cambiando ν por $-\nu$, así que se define la función:

$$J_{-\nu}(t) = \left(\frac{t}{2}\right)^{-\nu} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n - \nu + 1)} \left(\frac{t}{2}\right)^{2n}$$

que es linealmente independiente de la primera si 2ν no es entero positivo. Si 2ν es entero positivo, esta solución podría ser aún linealmente independiente de la primera, pues la solución presentada en el teorema contiene una constante, a , en el término logarítmico que puede ser cero. Sin embargo, veamos como en el caso $\nu = p$ entero positivo, las funciones $J_p(t)$ y $J_{-p}(t)$ son proporcionales:

$$J_{-p}(t) = \left(\frac{t}{2}\right)^{-p} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n - p + 1)} \left(\frac{t}{2}\right)^{2n}.$$

Pero la función Γ tiene polos en los enteros negativos así que la serie realmente comienza en $n = p$:

$$J_{-p}(t) = \left(\frac{t}{2}\right)^{-p} \sum_{n=p}^{\infty} \frac{(-1)^n}{n! (n - p + 1)!} \left(\frac{t}{2}\right)^{2n} = \left(\frac{t}{2}\right)^{-p} \sum_{n=0}^{\infty} \frac{(-1)^{n+p}}{(n + p)! n!} \left(\frac{t}{2}\right)^{2(n+p)}$$

de donde se obtiene inmediatamente el resultado buscado:

$$J_{-p}(t) = (-1)^p J_p(t), \quad p \text{ entero positivo.}$$

El caso en que ν no es entero pero 2ν sí lo es, no introduce logaritmos en las soluciones, es decir $J_\nu(t)$ y $J_{-\nu}(t)$ son linealmente independientes, aunque haya que demostrarlo. Pero si es entero positivo o cero (en este último caso no hay discusión, las dos raíces son iguales y por tanto una de las dos soluciones de las que habla el teorema lleva necesariamente un logaritmo) como acabamos de ver, la segunda solución lleva un logaritmo. Calculemosla en el caso $\nu = 0$:

$$x_2(t) = J_0(t) \log t + \sum_{n=1}^{\infty} b_n t^n$$

sustituyéndola en la ecuación, $tx'' + x' + tx = 0$:

$$(tJ_0'' + J_0' + tJ_0) \log t + 2J_0' + b_1 + 4b_2t + \sum_{n=2}^{\infty} [(n+1)^2 b_{n+1} + b_{n-1}] t^n = 0$$

El paréntesis que multiplica a $\log t$ es cero, pues $J_0(t)$ es solución de la ecuación. Esto ocurre siempre, el logaritmo no aparece a la hora de resolver la ecuación. Sustituyendo la expresión de $J_0(t)$:

$$J_0(t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} \left(\frac{t}{2}\right)^{2n}$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} 2n \left(\frac{t}{2}\right)^{2n-1} + b_1 + 4b_2t + \sum_{n=2}^{\infty} [(n+1)^2 b_{n+1} + b_{n-1}] t^n = 0.$$

Conviene separar la segunda serie en términos pares e impares:

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} 2n \left(\frac{t}{2}\right)^{2n-1} + b_1 + 4b_2t + \sum_{n=1}^{\infty} [(2n+1)^2 b_{2n+1} + b_{2n-1}] t^{2n} + \sum_{n=2}^{\infty} [(2n)^2 b_{2n} + b_{2n-2}] t^{2n-1} = 0$$

de donde se obtienen las siguientes relaciones de recurrencia:

$$\begin{aligned} b_1 &= 0 \\ -1 + 4b_2 &= 0 \\ (2n+1)^2 b_{2n+1} + b_{2n-1} &= 0, \quad n \geq 1 \\ \frac{(-1)^n 4n}{(n!)^2 2^{2n}} + (2n)^2 b_{2n} + b_{2n-2} &= 0, \quad n \geq 2 \end{aligned}$$

La solución de este conjunto de ecuaciones fijará los coeficientes b_n . La relación entre los coeficientes de índice impar y la primera ecuación $b_1 = 0$ implica:

$$b_{2n+1} = 0, \quad n \geq 0.$$

La ecuación que liga los términos de índice par también puede ser resuelta. Para escribirla de una forma compacta, pongamos:

$$\mathcal{H}_n = \sum_{k=1}^n \frac{1}{k}$$

con lo que la solución es:

$$b_{2n} = \frac{(-1)^{n+1}}{2^{2n} (n!)^2} \mathcal{H}_n, \quad n \geq 1$$

Una vez conocidos los coeficientes b_n , resulta inmediato escribir la segunda solución:

$$x_2(t) = J_0(t) \log t + \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \mathcal{H}_n}{2^{2n} (n!)^2} t^{2n}.$$

Las soluciones $J_0(t)$ y $x_2(t)$ son linealmente independientes, según asegura el teorema. Sin embargo, es usual utilizar en lugar de esta base otra formada por $J_0(t)$ y una combinación lineal de $J_0(t)$ y $x_2(t)$:

$$N_0(t) = \frac{2}{\pi} \left[\left(\gamma + \log \frac{t}{2} \right) J_0(t) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \mathcal{H}_n}{2^{2n} (n!)^2} t^{2n} \right]$$

que es la función de Neumann de orden cero. Esta función tiene una singularidad de tipo logarítmico en $t = 0$. De acuerdo con la notación dada antes:

$$N_0(t) = \frac{2}{\pi} x_2(t) + \frac{2}{\pi} (\gamma - \log 2) J_0(t)$$

donde γ es la constante de Euler, definida por:

$$\gamma = \lim_{n \rightarrow \infty} (\mathcal{H}_n - \log n) = 0.57721 \dots$$

El cálculo anterior puede hacerse para $\nu = p$ entero arbitrario siguiendo los mismos razonamientos, pero resulta muy complicado. Definamos en su lugar las siguientes funciones:

$$N_\nu(t) = \frac{\cos \nu\pi J_\nu(t) - J_{-\nu}(t)}{\operatorname{sen} \nu\pi}$$

directamente cuando ν no es un entero, y a través de un límite cuando ν es entero. En particular:

$$N_0(t) = \lim_{\nu \rightarrow 0} N_\nu(t).$$

Cuando $\nu = p$, entero, se puede demostrar que:

$$N_p(t) = \frac{2}{\pi}(\gamma + \log \frac{t}{2})J_p(t) - \frac{1}{\pi} \left[\left(\frac{t}{2}\right)^{-p} \sum_{n=0}^{p-1} \frac{(p-n-1)!}{n!} \left(\frac{t}{2}\right)^{2n} + \frac{\mathcal{H}_p}{p!} \left(\frac{t}{2}\right)^p + \left(\frac{t}{2}\right)^p \sum_{n=1}^{\infty} \frac{(-1)^n (\mathcal{H}_n + \mathcal{H}_{n+p})}{n!(n+p)!} \left(\frac{t}{2}\right)^{2n} \right].$$

Todas ellas tienen una singularidad logarítmica (cuando ν no es entero, no existe ese tipo de singularidad, aunque en cualquier caso no son analíticas en cero).

Las funciones de Neumann así definidas, son soluciones de la ecuación de Bessel para cada ν , y son linealmente independientes con $J_\nu(t)$. No entraremos en detalles sobre estas afirmaciones. Baste recordar aquí que, para ν dado, las funciones:

$$J_\nu(t), N_\nu(t)$$

son una base del espacio de soluciones de la ecuación de Bessel de orden ν . Si ν no es entero, las funciones:

$$J_\nu(t), J_{-\nu}(t)$$

son también una base de soluciones.

No hemos establecido que ocurre cuando 2ν es entero, pero ν no lo es. Según lo anterior, $J_\nu(t)$ y $J_{-\nu}(t)$ son una base, y no aparece ninguna singularidad logarítmica, luego la constante a del teorema debe ser cero y $J_\nu(t)$ y $N_\nu(t)$ son también una base. Calculemos J_ν y N_ν cuando $\nu = 1/2$:

$$N_{\frac{1}{2}}(t) = J_{-\frac{1}{2}}(t)$$

$$J_{\frac{1}{2}}(t) = \left(\frac{t}{2}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \frac{3}{2})} \left(\frac{t}{2}\right)^{2n}$$

usando las propiedades de la función Γ se obtiene:

$$n! \Gamma(n + \frac{3}{2}) = \left(\frac{1}{2}\right)^{2n+1} (2n+1)! \sqrt{\pi}$$

y sustituyendo en la expresión de $J_{\frac{1}{2}}$:

$$J_{\frac{1}{2}}(t) = \sqrt{\frac{2}{\pi t}} \operatorname{sen} t.$$

El mismo tipo de razonamiento permite calcular $J_{-\frac{1}{2}}$:

$$J_{-\frac{1}{2}}(t) = \sqrt{\frac{2}{\pi t}} \cos t$$

que forman una base del espacio de soluciones de la ecuación de Bessel de orden $1/2$.

Las funciones de Bessel verifican una serie de relaciones de recurrencia tales como:

$$\begin{aligned}\frac{d}{dt}[t^\nu J_\nu(t)] &= t^\nu J_{\nu-1}(t) \\ \frac{d}{dt}[t^{-\nu} J_\nu(t)] &= -t^{-\nu} J_{\nu+1}(t) \\ J_{\nu-1}(t) - J_{\nu+1}(t) &= 2J'_\nu(t) \\ J_{\nu-1}(t) + J_{\nu+1}(t) &= \frac{2\nu}{t} J_\nu(t)\end{aligned}$$

que se demuestran fácilmente usando la ecuación y la definición de $J_\nu(t)$.

Por ejemplo, la primera de ellas:

$$\frac{d}{dt}[t^\nu J_\nu(t)] = \frac{d}{dt} \left[\left(\frac{t}{2}\right)^{2\nu} \sum_{n=0}^{\infty} \frac{(-1)^n 2^\nu}{n! \Gamma(n + \nu + 1)} \left(\frac{t}{2}\right)^{2n} \right] = \sum_{n=0}^{\infty} \frac{(-1)^n 2\nu(n + \nu)}{n! \Gamma(n + \nu + 1)} \left(\frac{t}{2}\right)^{2(n+\nu)-1}$$

como $\Gamma(n + \nu + 1) = (n + \nu)\Gamma(n + \nu)$:

$$= \sum_{n=0}^{\infty} \frac{(-1)^n 2^\nu}{n! \Gamma(n + \nu)} \left(\frac{t}{2}\right)^{2(n+\nu)-1} = t^\nu \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \nu)} \left(\frac{t}{2}\right)^{2(n+\nu)-1} = t^\nu J_{\nu-1}(t)$$

La segunda se demuestra exactamente igual. La tercera se prueba a partir de las dos primeras. Multiplicando la primera por $t^{-\nu}$, la segunda por t^ν y sumando:

$$J_{\nu-1}(t) - J_{\nu+1}(t) = t^{-\nu} \frac{d}{dt}[t^\nu J_\nu(t)] + t^\nu \frac{d}{dt}[t^{-\nu} J_\nu(t)] = 2J'_\nu(t)$$

Y la cuarta, restando:

$$J_{\nu-1}(t) + J_{\nu+1}(t) = t^{-\nu} \frac{d}{dt}[t^\nu J_\nu(t)] - t^\nu \frac{d}{dt}[t^{-\nu} J_\nu(t)] = 2\nu t^{-1} J_\nu(t)$$

No podemos entrar aquí en una discusión del comportamiento de estas funciones cuando t tiende a ∞ . Pero el resultado es particularmente simple (y nada sorprendente en el caso de orden $1/2$):

$$J_\nu(t) = \sqrt{\frac{2}{\pi t}} \cos\left(t - (2\nu + 1)\frac{\pi}{4}\right) + o(t^{-3/2})$$

$$N_\nu(t) = \sqrt{\frac{2}{\pi t}} \operatorname{sen}\left(t - (2\nu + 1)\frac{\pi}{4}\right) + o(t^{-3/2})$$

es decir, tanto $J_\nu(t)$ como $N_\nu(t)$ oscilan indefinidamente, mientras la amplitud tiende a cero cuando t tiende a ∞ .

Para terminar este estudio elemental de la ecuación de Bessel, mencionemos aquí la relación entre esta ecuación y la de Bessel esférica de orden ν :

$$x'' + \frac{2}{t}x' + \left(1 - \frac{\nu(\nu + 1)}{t^2}\right)x = 0.$$

Un cambio de variable lleva esta ecuación a la de Bessel de orden $\nu + \frac{1}{2}$:

$$\begin{aligned}x &= \frac{y}{\sqrt{t}} \\ y'' + \frac{1}{t}y' + \left(1 - \frac{(\nu + \frac{1}{2})^2}{t^2}\right)y &= 0\end{aligned}$$

y por tanto una base del espacio de soluciones de la ecuación de Bessel esférica es:

$$j_\nu(t) = \sqrt{\frac{\pi}{2t}} J_{\nu+\frac{1}{2}}(t)$$

$$n_\nu(t) = \sqrt{\frac{\pi}{2t}} N_{\nu+\frac{1}{2}}(t).$$

4.4 Singularidades en el infinito

Como hemos visto, el tratamiento de los puntos ordinarios o singulares regulares no presenta ninguna dificultad. Sin embargo, en orden al estudio de los comportamientos de las soluciones de una ecuación diferencial cuando t tiende a ∞ , resulta necesario introducir una serie de métodos que permitan aplicar la teoría estudiada previamente a esta situación. La manera más usual de hacerlo es cambiar la variable, pasando de t a su inversa, con lo cual el punto del infinito pasa a ser el origen. Consideremos la ecuación:

$$R(t)x'' + P(t)x' + Q(t)x = 0$$

hagamos el cambio de variable $t = 1/s$. La nueva ecuación es:

$$\frac{dx}{dt} = -\frac{1}{t^2} \frac{dx}{ds}, \quad \frac{d^2x}{dt^2} = \frac{1}{t^4} \frac{d^2x}{ds^2} + \frac{2}{t^2} \frac{dx}{ds}$$

$$s^4 R\left(\frac{1}{s}\right) \frac{d^2x}{ds^2} + \left(2s^3 R\left(\frac{1}{s}\right) - s^2 P\left(\frac{1}{s}\right)\right) \frac{dx}{ds} + Q\left(\frac{1}{s}\right) x = 0.$$

Por lo tanto, el punto $t = \infty$ (es decir $s = 0$) es un punto ordinario de la ecuación inicial si las funciones:

$$\frac{1}{R\left(\frac{1}{s}\right)} \left(2R\left(\frac{1}{s}\right) - \frac{1}{s^2} P\left(\frac{1}{s}\right)\right), \quad \frac{1}{s^4 R\left(\frac{1}{s}\right)} Q\left(\frac{1}{s}\right)$$

son analíticas en un entorno de $s = 0$.

De forma similar, $t = \infty$ es un punto singular regular para la ecuación de partida si alguna (o las dos) de las funciones anteriores no es analítica en $s = 0$, pero sí son analíticas las funciones:

$$\frac{1}{R\left(\frac{1}{s}\right)} \left(2R\left(\frac{1}{s}\right) - \frac{1}{s} P\left(\frac{1}{s}\right)\right), \quad \frac{1}{s^2 R\left(\frac{1}{s}\right)} Q\left(\frac{1}{s}\right).$$

Ejemplo 4.4.1 La ecuación de Legendre.

$$(1 - t^2)x'' - 2tx' + \alpha(\alpha + 1)x = 0.$$

Las funciones a estudiar son:

$$\frac{2s}{s^2 - 1}, \quad \frac{\alpha(\alpha + 1)}{s^2(s^2 - 1)}.$$

La segunda no es analítica, pero al multiplicarla por s^2 se obtienen dos funciones analíticas en $s = 0$. Por lo tanto el punto $t = \infty$ es un punto singular regular para esta ecuación. Eso hace que se pueda aplicar el teorema de Frobenius a la ecuación transformada lo que da para la ecuación original una serie de potencias en t^{-1} , multiplicada por alguna potencia de t o un logaritmo.

Ejemplo 4.4.2 Ecuación de Hermite.

$$x'' - 2tx' + \lambda x = 0.$$

Haciendo el cambio, se obtienen:

$$\frac{2}{s} \left(1 + \frac{1}{s^2}\right), \quad \frac{\lambda}{s^4}$$

que tienen en $s = 0$ singularidades más fuertes de las que permite la definición de punto singular regular. Por tanto $t = \infty$ es un punto singular irregular.

Ejemplo 4.4.3 Ecuación de Airy.

$$x'' = tx.$$

En este caso las funciones a estudiar son:

$$\frac{2}{s}, \quad -\frac{1}{s^5}$$

así que $t = \infty$ es un punto singular irregular para esta ecuación. Lo mismo ocurre con la ecuación de Bessel.

El conocimiento de las singularidades de una ecuación diferencial lineal proporciona algunas veces los coeficientes de dicha ecuación y en general da ideas sobre su comportamiento en un entorno de esos puntos. Veamos como en el caso particular de la ecuación lineal de segundo orden, solo hay esencialmente una ecuación que tenga tres puntos singulares regulares con exponentes (raíces de la ecuación indicial) dados.

4.5 La ecuación hipergeométrica

La ecuación:

$$t(1-t)x'' + (\gamma - (1 + \alpha + \beta)t)x' - \alpha\beta x = 0$$

tiene puntos singulares regulares en 0, 1, ∞ . Esto hace que podamos calcular sus soluciones como series de potencias en torno a cualquiera de estos tres puntos. Los radios de convergencia de estas series no son en general infinitos, así que será necesario considerarlas en distintas situaciones según las necesidades de cada caso. Además, según los valores de las constantes, pueden aparecer logaritmos o no. Dado que no intentamos hacer una teoría completa de la ecuación hipergeométrica, nos limitaremos a los casos más sencillos.

Consideremos en primer lugar los desarrollos en serie en torno al punto $t = 0$. Los exponentes para este punto son 0 y $1 - \gamma$. Siempre existirá una solución del tipo una potencia de t por una función analítica. El exponente será 0 o $1 - \gamma$, según cuál de los dos sea mayor. Si la diferencia entre ambos no es un entero, el teorema de Frobenius permite asegurar la existencia de dos soluciones linealmente independientes de la forma dicha anteriormente. Supongamos entonces que γ no es un entero negativo o cero, con lo cual existe una solución de la forma siguiente:

$$x(t) = \sum_{n=0}^{\infty} a_n t^n$$

Sustituyendo en la ecuación:

$$t(1-t) \sum_{n=0}^{\infty} n(n-1)a_n t^{n-2} + (\gamma - (\alpha + \beta + 1)t) \sum_{n=0}^{\infty} n a_n t^{n-1} - \alpha\beta \sum_{n=0}^{\infty} a_n t^n = 0$$

multiplicando y simplificando se obtiene:

$$\sum_{n=1}^{\infty} (n(n-1) + \gamma n) a_n t^{n-1} - \sum_{n=0}^{\infty} (n(n-1) + n(\alpha + \beta + 1) + \alpha\beta) a_n t^n = 0$$

es decir:

$$\sum_{n=0}^{\infty} ((n+1)(\gamma + n) a_{n+1} - (\alpha + n)(\beta + n) a_n) t^n = 0$$

lo que nos permite obtener la siguiente relación de recurrencia:

$$a_{n+1} = \frac{(\alpha + n)(\beta + n)}{(n+1)(\gamma + n)} a_n, \quad n \geq 0$$

Si se define el símbolo:

$$(\alpha)_n = \alpha(\alpha + 1)(\alpha + 2) \dots (\alpha + n - 1)$$

se encuentra como solución, tomando $a_0 = 1$:

$$a_n = \frac{(\alpha)_n (\beta)_n}{n! (\gamma)_n}, \quad n \geq 1$$

La serie correspondiente se llama serie hipergeométrica, y converge, con las restricciones anteriormente dichas sobre γ , en el intervalo $(-1, 1)$:

$$w_{01}(t) = F(\alpha, \beta, \gamma; t) = \sum_{n=0}^{\infty} \frac{(\alpha)_n (\beta)_n}{n! (\gamma)_n} t^n.$$

La segunda solución, asociada al segundo exponente es:

$$x_2(t) = t^{1-\gamma} \sum_{n=0}^{\infty} b_n t^n$$

y los coeficientes se podrían calcular de la manera usual. Sin embargo, una de las propiedades más interesantes de la ecuación hipergeométrica es su comportamiento frente a transformaciones de la variable dependiente e independiente. Sin entrar en una discusión detallada, para este caso tenemos lo siguiente. Si:

$$y = t^{-(1-\gamma)} x$$

la nueva ecuación es:

$$t(1-t)y'' + ((2-\gamma) - (\alpha + \beta - 2\gamma + 3)t)y' - (\alpha - \gamma + 1)(\beta - \gamma + 1)y = 0$$

que es una ecuación hipergeométrica de parámetros, $\alpha - \gamma + 1$, $\beta - \gamma + 1$, $2 - \gamma$. En $t = 0$, los exponentes son 0, $\gamma - 1$, por lo que la solución asociada a 0 es:

$$y(t) = F(\alpha - \gamma + 1, \beta - \gamma + 1, 2 - \gamma; t)$$

de donde la segunda solución es:

$$w_{02}(t) = t^{1-\gamma} F(\alpha - \gamma + 1, \beta - \gamma + 1, 2 - \gamma; t).$$

Desde un punto de vista más general, se puede considerar a F como dependiente de una variable compleja z . De acuerdo con lo dicho de F , ésta está definida en el interior del círculo unidad (como siempre, γ no puede tomar valores que sean 0 o un entero negativo). Es posible demostrar que F se puede extender a una función analítica en todo el plano complejo salvo en un corte $[1, \infty)$. Esta función se puede representar por una integral, pero no iremos más lejos en esta línea.

Si queremos calcular el desarrollo en serie en el punto $t = 1$, podemos actuar como antes con series en $t - 1$. Sin embargo es más cómodo hacer un cambio de variable que nos lleve el punto 1 al 0. Sea entonces $s = 1 - t$. La ecuación transformada es:

$$s(1-s)x'' + ((\alpha + \beta - \gamma + 1) - (\alpha + \beta + 1)s)x' - \alpha\beta x = 0$$

otra vez la ecuación hipergeométrica con otros parámetros. Por tanto las soluciones correspondientes son:

$$\begin{aligned} w_{11}(t) &= F(\alpha, \beta, \alpha + \beta - \gamma + 1; 1 - t) \\ w_{12}(t) &= (1 - t)^{\gamma - \alpha - \beta} F(\gamma - \beta, \gamma - \alpha, 1 + \gamma - \alpha - \beta; 1 - t) \end{aligned}$$

Finalmente, si queremos calcular el desarrollo en $t = \infty$, hagamos el cambio $t = 1/s$. Si simultáneamente cambiamos la variable dependiente, $y = s^{-\alpha} x$, se obtiene:

$$s(1-s)y'' + (1 + \alpha - \beta - (2 + 2\alpha - \gamma)s)y' - \alpha(\alpha - \gamma + 1)y = 0$$

ecuación hipergeométrica con parámetros α , $\alpha - \gamma + 1$, $1 + \alpha - \beta$, por lo tanto:

$$\begin{aligned} w_{\infty 1}(t) &= t^{-\alpha} F(\alpha, \alpha - \gamma + 1, 1 + \alpha - \beta; t^{-1}) \\ w_{\infty 2}(t) &= t^{-\beta} F(\beta, \beta - \gamma + 1, 1 - \alpha + \beta; t^{-1}) \end{aligned}$$

Las seis soluciones $w(t)$ obtenidas convergen en los dominios adecuados, que se solapan en algunos casos. Es evidente que en esas zonas solo hay dos soluciones linealmente independientes, así que existen ciertas relaciones entre ellas. También es posible demostrar como aparecen relaciones de recurrencia entre las funciones hipergeométricas con distintos parámetros.

Elijiendo adecuadamente los parámetros, se obtienen otras ecuaciones ya tratadas individualmente. Por ejemplo: $\alpha = -n$, $\beta = n + 1$, $\gamma = 1$, $s = 1 - 2t$ nos lleva a la ecuación de Legendre:

$$(1 - s^2)x'' - 2sx' + n(n + 1)x = 0$$

por lo tanto, los polinomios de Legendre son:

$$P_n(t) = F(-n, n+1, 1; \frac{1}{2}(1-t)).$$

Como última aplicación de la ecuación hipergeométrica, estudiaremos cómo se puede obtener a partir de ella la ecuación de Kummer, y la aplicaremos a la ecuación radial que aparece al separar variables en la ecuación de Schrödinger con un potencial coulombiano.

4.6 La ecuación hipergeométrica confluyente

Consideremos la ecuación hipergeométrica:

$$t(1-t)x'' + ((\gamma - (\alpha + \beta + 1)t)x' - \alpha\beta x = 0$$

y hagamos un cambio de variable t por t/β (seguiremos llamando a la nueva variable t). La nueva ecuación es:

$$t(\beta - t)x'' + (\gamma\beta - (\alpha + \beta + 1)t)x' - \alpha\beta x = 0$$

que tiene puntos singulares en $0, \beta, \infty$. Por supuesto, la función:

$$F(\alpha, \beta, \gamma; \frac{t}{\beta})$$

es solución de esta ecuación. Dividamos por β la ecuación:

$$t(1 - \frac{t}{\beta})x'' + (\gamma - (1 + \frac{\alpha + 1}{\beta})t)x' - \alpha x = 0$$

y tomemos el límite cuando $\beta \rightarrow \infty$:

$$tx'' + (\gamma - t)x' - \alpha x = 0$$

Esta ecuación (llamada por razones obvias, ecuación hipergeométrica confluyente) tiene dos puntos singulares, en 0 , regular, y en ∞ irregular, debido a la confluencia de las dos singularidades regulares que tenía la ecuación inicial. Las soluciones se pueden calcular como límite de las soluciones de la ecuación con β finito:

$${}_1F_1(\alpha, \gamma; t) = \lim_{\beta \rightarrow \infty} F(\alpha, \beta, \gamma; \frac{t}{\beta}) = \lim_{\beta \rightarrow \infty} \sum_{n=0}^{\infty} \frac{(\alpha)_n (\beta)_n}{n! (\gamma)_n} \left(\frac{t}{\beta}\right)^n$$

es decir:

$${}_1F_1(\alpha, \gamma; t) = \sum_{n=0}^{\infty} \frac{(\alpha)_n}{n! (\gamma)_n} t^n$$

que se llama función hipergeométrica confluyente. Esta función está definida para todo t , cuando γ no es cero ni entero negativo. Es más, la función ${}_1F_1$ es analítica en todo el plano complejo, considerando a t como una variable compleja. Es posible encontrar una segunda solución para esta ecuación, linealmente independiente de la primera, la llamada función hipergeométrica confluyente de segunda clase. Ciertas ecuaciones de la Física Matemática se pueden relacionar con la ecuación hipergeométrica confluyente. Por ejemplo las ecuaciones de Hermite, Laguerre y Bessel:

Si $\alpha = n$, $\gamma = \frac{1}{2}$, y hacemos el cambio $t = s^2$, la nueva ecuación es:

$$x'' - 2sx' + 4nx = 0$$

que es la ecuación de Hermite, una de cuyas soluciones es el polinomio de Hermite de índice par, $2n$:

$$H_{2n}(t) = (-1)^n \frac{(2n)!}{n!} {}_1F_1(-n, \frac{1}{2}; t^2)$$

si se elige $\gamma = 3/2$, se obtienen los de índice impar.

Si ahora elegimos $\alpha = -n$, $\gamma = \sigma - 1$, $\sigma > -1$, la ecuación es:

$$tx'' + (\sigma + 1 - t)x' + nx = 0$$

que es la ecuación de Laguerre. Esta ecuación tiene a los polinomios de Laguerre como soluciones:

$$L_n^\sigma(t) = e^t \frac{t^{-\sigma}}{n!} \frac{d^n}{dt^n} (t^{n+\sigma} e^{-t}), \quad n = 0, 1, 2, \dots$$

Los primeros de la serie son:

$$\begin{aligned} L_0^\sigma(t) &= 1 \\ L_1^\sigma(t) &= 1 + \sigma - t \\ L_2^\sigma(t) &= \frac{1}{2}((1 + \sigma)(2 + \sigma) - 2(2 + \sigma)t + t^2) \\ &\vdots \end{aligned}$$

Estos polinomios están relacionados con los de Hermite a través de las siguientes ecuaciones:

$$\begin{aligned} L_n^{-1/2}(t) &= \frac{(-1)^n}{2^{2n} n!} H_{2n}(\sqrt{t}) \\ L_n^{1/2}(t) &= \frac{(-1)^n}{2^{2n+1} n!} t^{-1/2} H_{2n+1}(\sqrt{t}) \end{aligned}$$

La demostración de este resultado se puede hacer usando una representación integral de los polinomios de Laguerre, que escribimos seguidamente para mostrar las relaciones que existen entre las funciones especiales, y que han llevado a muchos tratamientos unificadores (como por ejemplo los que usan la teoría de los grupos de Lie):

$$L_n^\sigma(t) = \frac{1}{n!} t^{-\sigma/2} e^t \int_0^\infty s^{n+\sigma/2} e^{-s} J_\sigma(2\sqrt{st}) ds, \quad \sigma > 1, \quad n = 0, 1, 2, \dots$$

donde J_σ es la función de Bessel de orden σ . Nada de extraño tiene que la ecuación de Bessel esté también relacionada con la ecuación hipergeométrica confluyente.

La notación ${}_1F_1$ se refiere a las llamadas series hipergeométricas generalizadas:

$${}_pF_q(\alpha_r, \gamma_s; t) = \sum_{n=0}^{\infty} \frac{\prod_{r=1}^p (\alpha_r)_n}{\prod_{s=1}^q (\gamma_s)_n} \frac{t^n}{n!}$$

por lo que las series hipergeométricas que introdujimos al principio son ${}_2F_1$, y las confluentes ${}_1F_1$. Estas series convergen en $(-1, 1)$ (en el disco unidad, aunque pueden extenderse por prolongación analítica fuera de ese disco).

4.7 El átomo de hidrógeno

Cuando se considera la ecuación de Schrödinger con un potencial central, el problema es separable en coordenadas esféricas. La ecuación radial cuando el potencial es coulombiano, resulta ser:

$$\frac{d^2 u_l}{dr^2}(r) + \left(\frac{2\mu}{\hbar^2} E + \frac{2\mu Z e^2}{\hbar^2 r} - \frac{l(l+1)}{r^2} \right) u_l(r) = 0$$

con el significado usual de las constantes. Si simplificamos la ecuación cambiando adecuadamente las variables y constantes:

$$\begin{aligned} u_l(r) &= \rho^{l+1} e^{-\rho/2} v_l(\rho) \\ \rho &= \sqrt{8\mu|E|} \frac{r}{\hbar} \\ \lambda &= \sqrt{\frac{\mu}{2|E|}} \frac{Z e^2}{\hbar} \end{aligned}$$

se obtiene:

$$\rho v_l'' + (2(l+1) - \rho)v_l' - (l+1 - \lambda)v_l = 0$$

que es una ecuación hipergeométrica confluyente de parámetros:

$$\alpha = l + 1 - \lambda, \quad \gamma = 2(l + 1).$$

Su solución es una función hipergeométrica confluyente (por las condiciones de contorno que se especifican en la interpretación física del problema la solución no debe ser singular en el origen, y la función hipergeométrica confluyente de segunda clase lo es). Pero, para que la función de ondas sea de cuadrado integrable, es necesario que la serie se corte, lo que implica que α es un entero negativo:

$$l + 1 - \lambda = -n_r$$

Esta relación permite cuantificar λ , es decir la energía. La ecuación que se obtiene es, como hemos visto antes, la ecuación de Laguerre, y por lo tanto, las soluciones de la ecuación radial del átomo de hidrógeno son potencias del radio por una exponencial decreciente y por polinomios de Laguerre dependientes del nivel considerado.