

Métodos Numéricos 2019 - Obligatorio 2

Bruno Figares (4391788-8), Adrián Gioda (4954044-5),
Daniel Martinez (4462694-5), Adriana Soucoff (3190794-8)

*Instituto de Matemática y Estadística
Facultad de Ingeniería. Universidad de la República
Montevideo, Uruguay*

Abstract

Este informe presenta resolución de problemas utilizando distintos métodos y comparando cada uno de ellos, está organizado según tres temas:

- Problema de mínimos cuadrados no lineal (PMCNL) utilizando los métodos de Ecuaciones Normales, descomposición QR y Gauss-Newton luego se discute las ventajas y desventajas evaluando cual es el mejor método para este caso.
- Problema de resolver ecuaciones diferenciales ordinarias (EDO) utilizando los métodos de Euler hacia atrás, Euler hacia adelante y Runge-Kutta, se discute ventajas y desventajas.
- Problema de Interpolación utilizando los métodos de interpolación lineal a trozos y Splines Cúbicos, se discute ventajas y desventajas.

Keywords: PMCL, PMCNL, Ecuaciones Normales, Descomposición QR, Gauss-Newton, EDO, Euler, Estabilidad, Runge-Kutta, Interpolación, Lineal a trozos, Splines Cúbicos.

1 Introducción

Se presentan dos tablas de mediciones para las funciones no lineales g_1 y g_2
se definen :

$$g_1(x) = cx^{-p} \text{ y } g_2(x) = dx^{-q}$$

se convierten los problemas a PMCL y se resuelven a través del método de Ecuaciones Normales y método de descomposición QR, al igual que se resuelve de manera directa el PMCNL con el método de Gauss-Newton.

Se busca resolver la ecuación diferencial:

$$(PVI) : \begin{cases} y'(x) = -g_1(x)y + g_2(x) \\ y(1/2) = 0 \end{cases}$$

para ello se resuelve de forma analítica, con métodos de Euler (hacia adelante y hacia atrás) y método de Runge-Kutta.

Se realiza interpolación utilizando método Lineal a Trozos y Splines Cúbicos.

2 Metodología

2.1 PMCNL - Transformación a PMCL

El modelo de ambas funciones tiene la forma $y = cx^{-p}$. Esta puede transformarse en una relación lineal aplicando un logaritmo a ambos lados de la expresión de esta forma, $\log(y) = \log(cx^{-p}) = \log(c) + (-p)\log(x)$. Aplicando cambios de variable apropiados, se obtiene la relación lineal $Y = c_2 + c_1X$.

Se busca el vector C de coeficientes que minimiza $\|AC - Y\|_2^2$, con $A = \begin{pmatrix} X_1 & 1 \\ \vdots & \vdots \\ X_m & 1 \end{pmatrix}$.

Para esto se deben resolver las ecuaciones normales $A^tAC = A^tY$.

Resolución de las ecuaciones normales: desarrollamos para el caso de g_1 es análogo para g_2 sustituyendo c por d y p por q

$Y = \log(y)$, $X = \log(x)$, $c_1 = \log(c)$, $c_2 = -p$
siendo n el largo del vector X tenemos las ecuaciones normales

$$\begin{cases} \sum Y = n * c_1 + c_2 * \sum X \\ \sum X * Y = c_1 * \sum X + c_2 * \sum X^2 \end{cases} \quad (1)$$

Siendo

$$C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Resolvemos : $C = (A^t * A) \setminus (A^t * Y)$

Finalmente obtenemos: $c = \exp(c_1)$, $p = -c_2$

Aplicación de descomposición QR: Debido a que el calculo de A^tA esta mal

condicionado, se aplica la descomposición QR de A . Teniendo que $A \in \mathcal{M}_{m \times n}$, con $m > n$ tiene rango completo (sus columnas son LI), por el teorema 4.3.1 de los apuntes se tiene que existen matrices $Q \in \mathcal{M}_{m \times m}$, $R \in \mathcal{M}_{m \times n}$ tales que $A = QR$. Con Q ortogonal y R triangular superior de la forma

$$\begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$

con $R_1 \in \mathcal{M}_{n \times n}$.

Sustituyendo A por QR , el problema de minimización se torna $\min \|QRC - Y\|_2^2$. Separando a Q en dos partes $[Q_1 Q_2]$ tal que $Q_1 \in \mathcal{M}_{m \times n}$ y $Q_2 \in \mathcal{M}_{m \times (m-n)}$ y operando, se llega al problema de minimización $\min \|R_1 C - Q_1^t Y\|_2^2$ cuya solución proviene del sistema de ecuaciones $R_1 C = Q_1^t Y$, que como R_1 es triangular superior se resuelve con sustitución hacia atrás.

2.2 PMCNL - Resolución por Gauss-Newton

Otra manera de resolver el problema sin tener que utilizar logaritmos es a través del uso de los polinomios de Taylor lineales. Para esto, para ajustar los puntos X_i, Y_i , con $i \in \{1, \dots, n\}$

$$\begin{aligned} F(c, p) &= \|X_{c,p} - Y\|_2^2 \\ X_{c,p} &= c * X_i^p \forall i \in \{1, \dots, n\} \end{aligned}$$

Con este problema nuevo, asumimos que estamos cerca de una solución y linealizamos el problema a través del polinomio de Taylor de orden 1, con la posibilidad de obtener una solución con un error acotado por el cuadrado de la distancia de la solución inicial k

$$F_k = \|F(k) + \Delta F(k) \Delta_{c,p} - Y + o(\Delta_{c,p}^2)\|_2^2$$

Haciendo los siguientes cambios de variable obtenemos una secuencia de PMCL que podemos resolver y encontrar soluciones progresivamente mejores

$$\begin{aligned} F_{k_i} &\equiv \|F(k_i) + \Delta F(k_i) \Delta_{c,p} - Y\|_2^2 \\ Y_{k_i} &= Y - F(k_i) \\ k_{i+1} &= k_i + \min_{\Delta_{c,p}} \|\Delta F(k_i) \Delta_{c,p} - Y_{k_i}\|_2^2 \end{aligned}$$

2.3 EDO - Resolución Analítica

Se resuelve analíticamente en primer lugar a fin de poder evaluar la solución provista por distintos métodos numéricos. En la parte 2 se vio que $g_1(x) = cx^{-2}$ y $g_2(x) = dx^{-3}$ con $c, d \in \mathbb{R}$. La ecuación diferencial es entonces:

$$y' + \frac{c}{x^2}y = \frac{d}{x^3} \quad (2)$$

Solución de la homogénea:

$$\begin{cases} y' + \frac{c}{x^2}y = 0 \\ y_h = \exp(-\int cx^{-2}dx) \\ y_h = ke^{c/x} \end{cases}$$

Variación de constantes: Se escribe la función y como $y = k(x)e^{c/x}$. La idea es obtener la expresión de $k(x)$. Se deriva y : $y' = k'(x)e^{c/x} + k(x)\left(-\frac{c}{x^2}e^{c/x}\right)$. Sustituyendo en la EDO, se cancelan dos términos y se despeja $k'(x)$. Integrando la expresión resultante se obtiene $k(x)$.

$$k(x) = \int \frac{d}{x^3}e^{-c/x}dx \quad (3)$$

Aplicando integración por partes con $u = \frac{d}{x}$, $du = -\frac{d}{x^2}$, $v = e^{-c/x}$ y $dv = \frac{c}{x^2}e^{c/x}$ se llega a:

$$k(x) = \frac{1}{c} \left(\frac{d}{x} e^{-c/x} + C_1 + \frac{d}{c} \int e^{-c/x} \frac{c}{x^2} dx \right) \quad (4)$$

Aplicando integración por sustitución con la misma v y dv se tiene que:

$$k(x) = \frac{1}{c} \left(\frac{d}{x} e^{-c/x} + C_1 + \frac{d}{c} e^{-c/x} + C_2 \right) = \frac{d}{c} e^{-c/x} \left(\frac{1}{x} + \frac{1}{c} \right) + k \quad (5)$$

Finalmente se llega a la expresión de y :

$$y(x) = \frac{d}{c^2} + \frac{d}{cx} + ke^{c/x} \quad (6)$$

Usando la condición inicial $y(1/2) = 0$ se halla que $k = -\frac{d}{c} \left(\frac{1}{c} + 2 \right) e^{-2c}$.

2.4 EDO - Métodos de Euler

En el método de Euler hacia adelante se aproxima la derivada en el punto x_k mediante el cociente entre la diferencia de los pasos consecutivos y_{k+1} e y_k y el paso h .

Entonces se tiene la iteración:

$$\begin{cases} y_{k+1} = y_k + hf(x_k, y_k) \\ y_0 = y_{inicial} \end{cases}$$

con $x_k = x_0 + kh$ con $h > 0$ fijo

Para nuestro caso tenemos:

$$\begin{aligned} y_{k+1} &= y_k + hf(x_k, y_k) \\ &= y_k + h \left(\frac{d}{x_k^3} - \frac{c}{x_k^2} y_k \right) \end{aligned}$$

El método de Euler hacia atrás estima la derivada en el punto x_{n+1} imponiendo $y'(x_{n+1}) = \frac{y_{n+1} - y_n}{h} = f(x_{n+1}, y_{n+1})$. Entonces se tiene la iteración:

$$\begin{cases} y_{k+1} = y_k + hf(x_{k+1}, y_{k+1}) \\ y_0 = y_{inicial} \end{cases}$$

Debido a que y_{k+1} aparece en ambos lados de la ecuación, este es un método implícito. Sin embargo, si f es lineal en y , se puede obtener una formulación explícita.

En este caso, la EDO a resolver es $y' = f(x, y) = \frac{d}{x^3} - \frac{c}{x^2}y$. Dado que f es lineal en y , se desarrolla la formulación explícita para y_{k+1} .

$$\begin{aligned} y_{k+1} &= y_k + hf(x_{k+1}, y_{k+1}) \\ &= y_k + h \left(\frac{d}{x_{k+1}^3} - \frac{c}{x_{k+1}^2} y_{k+1} \right) \end{aligned}$$

Despejando y_{k+1} se obtiene:

$$y_{k+1} = \frac{y_k + \frac{hd}{x_{k+1}^3}}{1 + \frac{hc}{x_{k+1}^2}} \quad (7)$$

Estabilidad numérica

Un método es numéricamente estable si la diferencia entre la solución y la solución numérica se mantiene acotada al crecer las iteraciones.

2.5 EDO - Implementación de los Métodos de Euler

2.6 EDO - Método de Runge-Kutta

Los métodos de Runge-Kutta calculan el valor de un punto y_{n+1} a partir del punto anterior y_n y de un promedio ponderado de las derivadas en distintos puntos en el intervalo $[x_n, x_{n+1}]$. El usado para esta sección es el implementado por la función `ode45` de Octave. Este es un método de 6 etapas de orden 5 y funciona de la siguiente manera: Empezando en el punto (x_n, y_n) con pendiente $s_1 = y' = f(x_n, y_n)$ y un paso h de tamaño apropiado, se busca computar y_{n+1} en el punto $x_{n+1} + h$. Usando "Euler hacia adelante" con la pendiente s_1 y una fracción del paso h se obtiene un segundo punto con el cual se calcula la pendiente s_2 . Este proceso se repite hasta conseguir s_6 . Mediante un promedio ponderado entre las 6 pendientes se calcula y_{n+1} . Se calcula una última pendiente s_7 en (x_{n+1}, y_{n+1}) que servirá para calcular una estimación del error que será usado para decidir el paso h para el cálculo del siguiente punto.

La función `ode45` produce una sucesión de puntos (x_n, y_n) que parecen ser demasiado espaciados como se vera en las gráficas de la sección siguiente. Debido a esto, puede necesitarse un interpolador para poder crear una gráfica mas suave en Octave. Esto se presentará en Interpolación.

2.7 Interpolación Lineal a Trozos

2.8 Interpolación con Splines Cúbicos

3 Conclusions

3.1 Comparación - PMCNL

Los métodos todos llegan a una solución similar, si bien en este caso la solución funciona bien, en otros casos puede que la transformación cambie sustancialmente la precisión del error ya que el logaritmo afecta de manera más profunda a ciertos puntos que a otros. Si los puntos de la gráfica ajustan perfectamente a una curva con los valores, podemos llegar al mismo número con ambas soluciones, pero sino pasa esto (todos los casos que realmente nos importan donde estamos encontrando una mejor aproximación) hay que comparar la fórmula de los errores de las derivadas parciales de las coordenadas de Y frente a la función de ajuste. Se utiliza una simplificación de la fórmula del error asumiendo que la función de ajuste se mantiene estática con el mejor ajuste previo para el punto para mantener el análisis simple.

$$\begin{aligned}Err(Y_i) &= (X_i - Y_i)^2 \\Err_{log}(Y_i) &= (\log(X_i) - \log(Y_i))^2\end{aligned}$$

entonces

$$\begin{aligned}\frac{\delta Err}{\delta Y_i} &= 2Y_i - 2X_i \\ \frac{\delta Err_{log}}{\delta Y_i} &= 2 \frac{\log(Y_i) - \log(X_i)}{Y_i}\end{aligned}$$

Conforme los errores sean mayores, el PMCL simplificado va a dar soluciones que valoren menos el error puntual en proporción en el caso de los logaritmos. Por otro lado, es mucho más fácil de calcular con logaritmos. Una cosa que se podría hacer para un problema genérico del estilo es usar primero el problema simplificado con logaritmos y después usar Newton Raphson para llegar a una solución al problema inicial.

3.2 Comparación - EDO

En la figura ??, se ven las graficas de los metodos de Euler y Runge-Kutta junto con la solucion analitica. Se puede observar que para el paso elegido, los metodos de

Euler hacia adelante y hacia atras envuelven la analitica, acercandose mas conforme se achica este. En cambio, la solucion dada por Runge-Kutta se aproxima mucho mas a la analitica con menos pasos, dado que ode45, siendo un metodo de orden 5, tiene un error proporcionalmente mucho menor en cada paso.

En general, Euler hacia adelante tiene la ventaja de usar una expresion explicita para la iteracion, pero con la desventaja de tener una region de estabilidad pequeña. Por otro lado, Euler hacia atras usa una expresion implicita para la iteracion con una region de estabilidad mas grande. En el caso tratado, se pierde la desventaja dado que se puede reformular la expresion implicita como una explicita, dando como resultado que sea mas apto Euler hacia atras que Euler hacia adelante

3.3 Comparación - Interpolación

Como se ve en las graficas de la figura ??, la interpolacion con splines cúbicos queda por sobre la curva de la solucion analitica, en cambio la interpolacion a trozos muestra un error mayor en los puntos intermedios. En lo que se refiere a ventajas y desventajas, en la interpolacion a trozos, es mucho mas facil actualizar la funcion interpolante luego de agregar puntos nuevos, pero no da una funcion derivable en todo punto, mientras que en splines cubicos si. Se podrian combinar propiedades de ambos enfoques realizando, para obtener el valor de la funcion en un punto intermedio de un segmento, un spline cubico usando una serie de puntos dato cercanos al segmento.