Reto 1: Interpolación

Asignatura:
Análisis Numérico

Docente:
Eddy Herrera Daza

Integrantes:

Andrés Felipe Vásquez Rendón
Fabián Andrés Olarte Vargas
Johan Mateo Rosero Quenguan
David Santiago Suarez
Heyling Burgos Algarin

Octubre de 2021

Índice

1. Introduction

El siguiente documento tiene el fin de poder desarrollar y determinar un instante en el clima, en este caso se hace una análisis del nordeste brasilero, esto para poder hacer uso del método de interpolación entre estaciones de acuerdo a los datos brindados.

2. Contexto

El problema planteado es dado un conjunto de valores asociados a variables climátias, indexados en el tiempo y en el espacio, se solicita determinar numéricamente primero los valores de la variable y cada media hora en una estación de monitoreo seleccionada, y asimismo tambien los valores de la variable y cada hora en una estación de monitoreo utilizando los datos dados de una estación cercana. Para la solución de este ejercicio se hace uso de la interpolación.

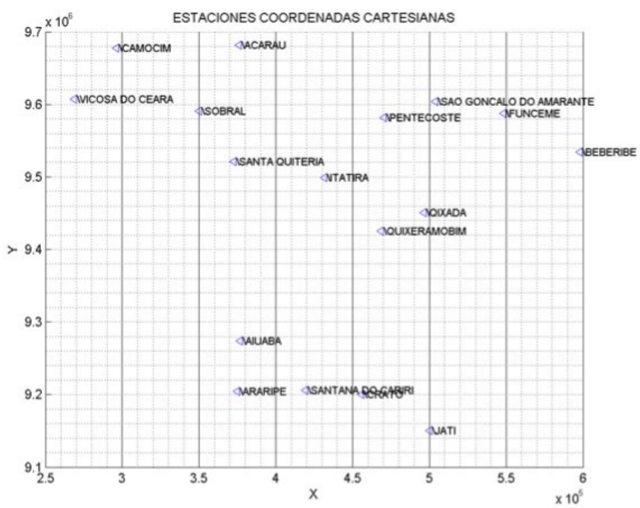

Figura 1: Resultado de 3 peticiones a la vez.

En primera instancia en la figura 1, se presenta el mapa de estaciones climáticas cercanas a la de zona de Fortaleza en Brasil. En donde, el grupo escogió la estación de Itatira y Quixada para el desarrollo del problema, ya que ambas estaciones estan relativamente cerca, además de poseer la misma cantidad de datos para su análizis.

Para el desarrollo de este reto se acudió al uso del método de interpolación para de esta forma poder determinar nuevos puntos haciendo uso de datos dados. Todo esto permite la determinación de nuevos valores para la resolución de la problemática planteada en el reto.

2.1. Conceptos

En el desarrollo del ejercicio propuesto es necesario conocer los conceptos y algoritmos con el fin de poder usar de forma correcta.

• Error Relativo: está determinado por el cociente entre el error absoluto y el valor exacto.

• Interpolación cuadrático: Mediante una serie de puntos tomando de a tres pares es posible obtener un polinomio de grado 2 que determine la parábola correspondiente dela suavidad de la curva.

$$P_2(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) = y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

Figura 2: Ecuación interpolación cuadrática.

• Interpolación Splines: Interpolación determinada a polinomios que específicamente presentan oscilaciones, con el fin de determinar la función con base a los puntos de las oscilaciones.

$$s(x) = \begin{cases} s_0(x) & si \quad x \in [x_0, x_1] \\ s_1(x) & si \quad x \in [x_1, x_2] \\ \vdots & \\ s_{n-1}(x) & si \quad x \in [x_{n-1}, x_n] \end{cases}$$

Figura 3: Ecuación interpolación cúbica.

A continuación, se encuentran contenidos en dos archivos de tipo *.xls*, En el archivo de Excel, podemos encontrar la información según fecha y lugar, además de las condiciones de la zona. Los datos encontrados dentro del archivo, y que corresponden a las columnas del archivo son:

• Año.
• Dia Juliano.
• Hora.
• Temperatura Interna.
• Presión Atmosférica.
• temperatira del aire.
• Humedad.
• Velocidad, dirección, Velocidad y dirección máximas del Viento.
• Precipitación.
• Radiación Solar Total.
• Nivel del Agua.
• Temperatura del Suelo a 10 cm, 20 cm y 50 cm.
• Humedad del Suelo.
• Flujo de Calor en el suelo.
• Datos Relacionados con panel solar utilizado (Corriente y Batería).

Además, es necesario aclarar que el documento presenta diferentes hojas, en donde se hace referencia a los datos por cada ciudad de donde se obtuvo la información. Podemos encontrar los siguientes datos:

• Coordenada en X
• Coordenada en Y
• Estaciones

2.2. Diagrama de flujo

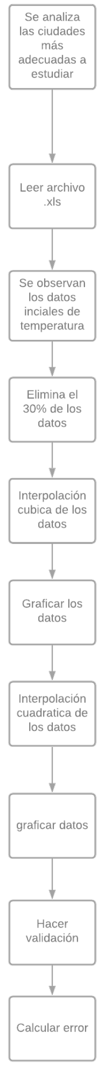

Figura 4: Diagrama de Flujo.

La intención del diagrama de flujo es poder mostrar como fue ese paso a paso en el desarrollo de nuestro taller.

2.3. Código fuente

En el Listing 1, se puede ver la implementación de interpolación cúbica y cuadrática, a su vez los diferentes métodos para encontrar los errores máximos y mínimos para la solución de los problemas planteados, esto programado en python.

```
import pandas as pd
import numpy as np
import math
```

```
import matplotlib.pyplot as plt
import scipy.interpolate as spi
from scipy.spatial import distance as jc

def clasificarDatos(nombreArchivo, nombreHoja, col1, col2, col3):
    dF = pd.read_excel(nombreArchivo, sheet_name=nombreHoja, usecols=[col1, col2,
    col3])
    Rcol1 = dF.get(col1)
    Rcol2 = dF.get(col2)
    Rcol3 = dF.get(col3)
    Rcol1 = list(Rcol1)
    Rcol2 = list(Rcol2)
    Rcol3 = list(Rcol3)
    return Rcol1, Rcol2, Rcol3

def eliminarPorcentajeDatos(porcentaje, indices, columna1, columna2, columna3):
    datosEliminar = int(porcentaje*len(columna1))
    restantes = []

    for i in range(10, datosEliminar):
        A = np.random.randint(10, len(columna1)-10)
        restantes.append((columna1[i], columna2[i]))
        indices = np.delete(indices, A, 0)
        columna1 = np.delete(columna1, A, 0)
        columna2 = np.delete(columna2, A, 0)
        columna3 = np.delete(columna3, A, 0)
    return indices, columna1, columna2, columna3, restantes

def errorMed(original, interpolado):
    emc=0
    for i in range (len(original)):
        emc += pow(interpolado[i]-original[i],2)

    return math.sqrt(emc/2)

def errorMinMax(original, interpolado):

    error=[]
    erroraux=[]
    contador=0
    for i in range(len(original)):
        resta = abs(interpolado[i]-original[i])
        if resta > 0:
            error.append(resta)
            erroraux.append(resta)

        contador+=1
        if contador == 100:
            print("error en el rango ",(i+1)-contador,"-",i+1, "maximo: ",round(max(
    erroraux),2),"y minimo: ",round(min(erroraux),2))
            contador=0
            erroraux.clear()
    print("Error maximo: {} y minimo: {} general".format(round(max(error),2),round(
    min(error),2) ))

    media=sum(error)/len(original)
    errorAbs=0
    for i in error:
        errorAbs+=abs(media-i)
    print("El error absouluto es: ",round(errorAbs/len(error),2))
```

```
#Extraer datos
DC1P, DC2P, DC3P = clasificarDatos("Datos.xls", "Itatira", "Dia Juliano", "Temp. Interna ( C )", "Hora")

#Guardar datos originales
xP = np.arange(0, len(DC2P),1) #Indices
DatosOr1P = DC1P.copy()
DatosOr2P = DC2P.copy()
DatosOr3P = DC3P.copy()

#Eliminar porcentaje de datos
p = 0.3 #Porcentaje
indices, columna1AP, columna2AP, columna3AP, eliminadoP = eliminarPorcentajeDatos(p, xP, DC1P,  DC2P, DC3P)

#Interpolaci nes
#Interpolacion cubica
s_pen = spi.splrep(indices, columna2AP)
xn_pen = xP
yn_pen = spi.splev(xn_pen, s_pen)
yn_pen=np.round(yn_pen,2)
#Interpolacion cuadratica
f = spi.interpolate.interp1d(indices, columna2AP, kind = 'quadratic')
xnew = xP
ynew = f(xnew)
ynew=np.round(ynew,2)
```

3. Solución primer problema

Para la solución del primer problema se toman los 720 datos que se tenian de la temperatura interna de la base Itatira y seguidamente realizar los métodos de interpolación cuadratíca y spline cúbicos.

3.1. Gráficas

A continuación se presenta el resultado de ambos métodos de inerpolación en una misma gráfica de la estación de Itatira.

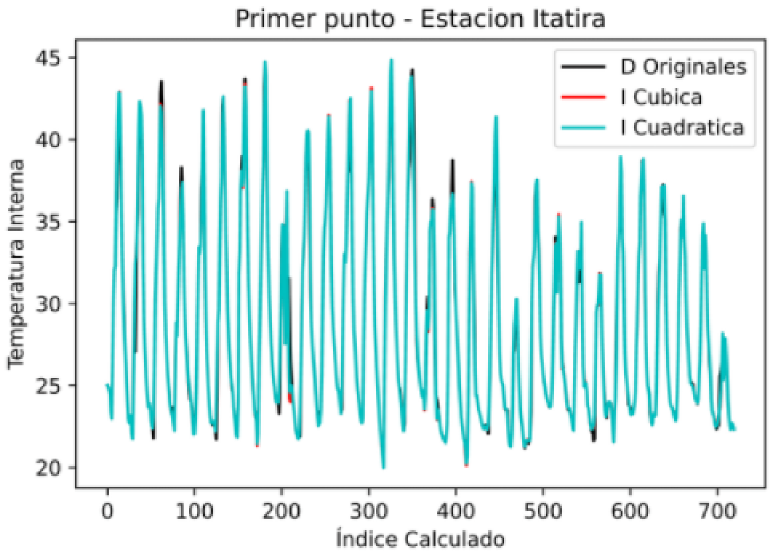

Figura 5: Estación itátira con interpolación cúbica y cuadrática.

Cómo se puede observar en la figura 4, ambos métodos de interpolación no alcanzan a cubrir la totalidad de los datos originales, además en ciertos puntos estas interpolaciones se desafan o estan desplazadas hacia el lado derecho.

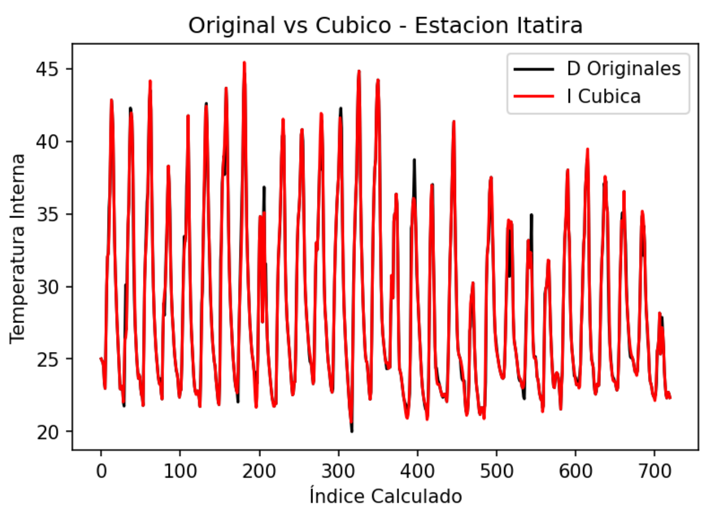

Figura 6: Interpolación cúbica vs datos originales de la estación Itatira.

En la anterior imagen, figura 5, se observa de manera más detallada la comparación entre los datos originales y la interpolación cúbica. Asimismo, podemos observar como la interpolación cúbica no logra llegar a cubrir algunos puntos originales, además que se logra desplazar en algunas zonas hacia la derecha.

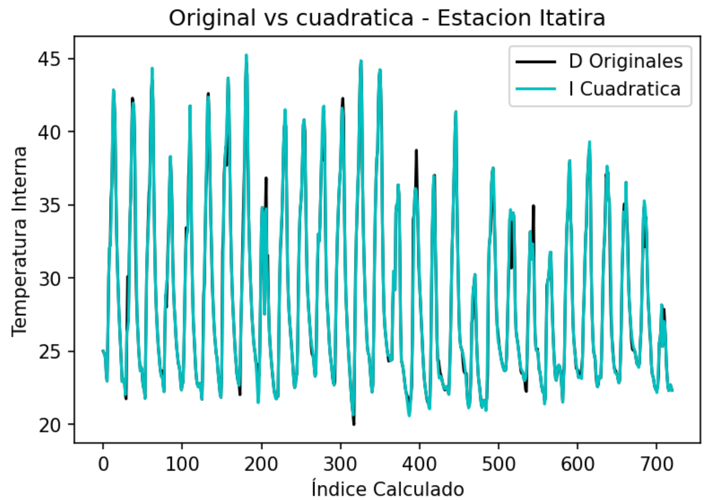

Figura 7: Interpolación cuadrática vs datos originales de la estación Itatira.

En la imagen anterior, figura 6, se logra observar con más detalle la comparación entre los datos originales y la interpolación cuadrática, se puede observar como en algunos puntos no logra cubrir a los datos originales, además de que se logra desplazar y desbordar en otros puntos.

3.2. Cálculo de errores

Para obtener los errores de las interpolaciones, se emplearon diversas métricas que nos permitieron obtener mayor informacion sobre las interpolaciones realizadas y su exactitud. Estas métricas fueron: Índice de Jaccard [1], error medio cuadrático, errores máximos y mínimos, error medio y error absoluto.

Conjuntos	Índice de Jaccard
Datos originales - interpolación cubica	0,283333333
Datos originales - interpolación cuadrática	0,280555556
Interpolación cubica - interpolación cuadrática	0,263888889

Tabla 1: Resultados obtenidos para el índice de Jaccard.

El índice de Jaccard nos indica que, de las dos interpolaciones empleadas, la mejor interpolación es la interpolación cubica, ya que esta, aunque tienen un índice lejano a uno, es mejor que la cuadrática por decimales. Esto también nos demuestra por qué gráficamente ambas interpolaciones se comportan similarmente.

Errores interpolación cubica	Valor
Error medio cuadrático	11,01
Error máximo	3,85
Error mínimo	0,01
Error absoluto	0,589

Tabla 2: Resultados obtenidos para los errores interpolación cubica.

[1] El índice de Jaccard indica en un rango de 0 a 1 el valor de similitud que hay en dos conjuntos

Rango	Máximo	Mínimo
0-100	3,63	0,03
100-200	3,51	0,01
200-300	3,85	0,01
300-400	3,53	0,02
400-500	2,28	0,02
500-600	3,55	0,01
600-700	2,8	0,01

Tabla 3: Resultados obtenidos de máximos y mínimos para la interpolación cúbica.

Errores interpolación cuadrática	Valor
Error medio cuadrático	10,89
Error máximo	3,61
Error mínimo	0,01
Error absoluto	0,59

Tabla 4: Resultados obtenidos para los errores interpolación cuadrática.

Rango	Máximo	Mínimo
0-100	3,61	0,01
100-200	3,49	0,04
200-300	3,56	0,02
300-400	3,57	0,04
400-500	2,32	0,05
500-600	3,52	0,01
600-700	2,77	0,01

Tabla 5: Resultados obtenidos de máximos y mínimos para la interpolación cuadrática.

4. Solución segundo problema

Para el desarrollo del segundo punto se analizaron las estaciones que estuvieran cercanas y que al mismo tiempo tuvieran la misma cantidad de datos para obtener una interpolación mas exacta sin datos faltantes. Por ende, las estaciones escogidas fueron Itatira y Quixada ya que se encuentran cercanas y al mismo tiempo contienen 720 datos cada una.

Similar al punto anterior se usaron dos métodos númericos para el calculo de las gráficas los cuales son la interpolación cuadrática y los spline cúbicos.

4.1. Gráficas

En las siguientes gráficas se evidenciarán la forma de las curvas de interpolación respecto a los datos originales, en este caso se usaron los datos de la estación Quixada para lograr hacer la interpolación de la estación Itatira.

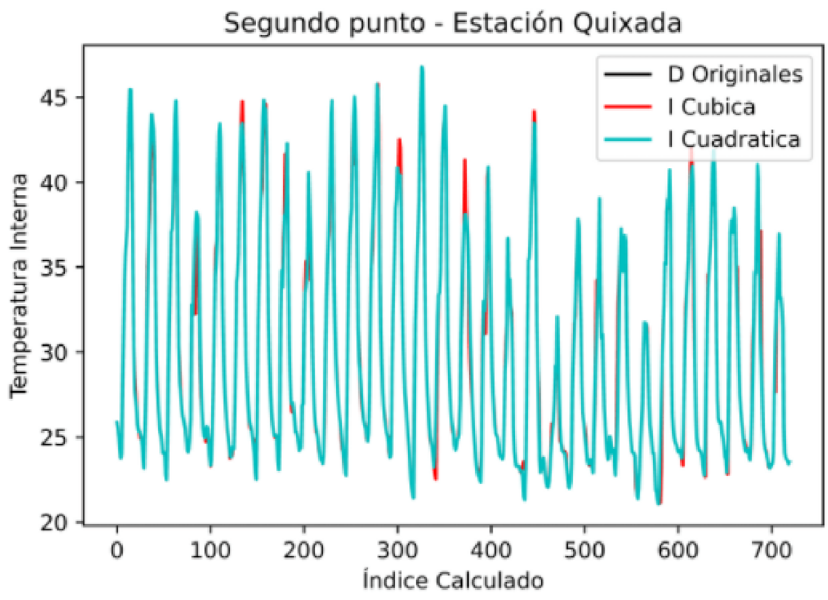

Figura 8: Estación itátira con los datos de Quixada usando interpolación cúbica y cuadrática.

Cómo se puede observar en la anterior figura, alguno de los spline logra obtener un acercamiento acertado a los datos originales ya que no es posible evidenciar la linea que representa los datos originales, no obstante, si existen diferencias en los maximos y minimos de la interpolación cuadratica con la cúbica ya que en algunos puntos hay desfases de la gráfica.

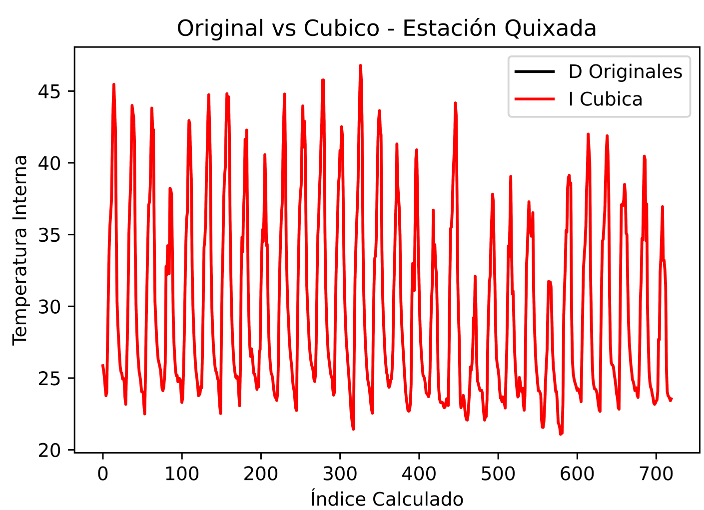

Figura 9: Interpolación cúbica vs datos originales de la estación Itatira con los datos de Quixada.

En esta gráfica de la figura 8, podemos claramente analizar que la verdadera interpolación que se asemeja a los datos originales es la interpolación cúbica ya que la presencia de la linea negra, es casi nula y no hay un cambio significativo en los picos de los máximos y minimos.

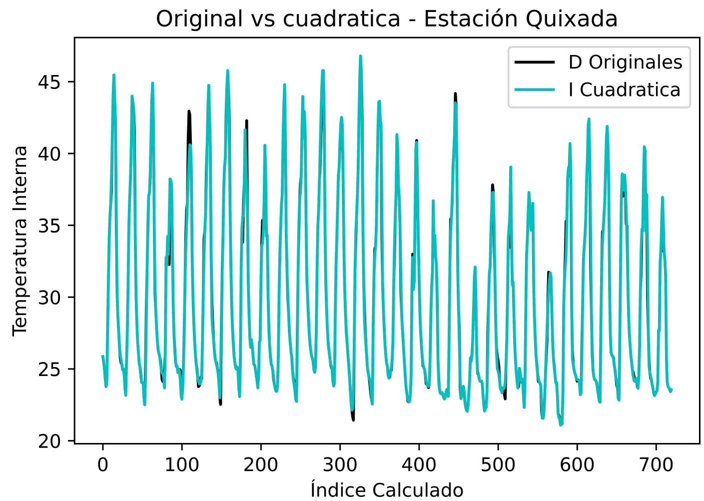

Figura 10: Interpolación cuadrática vs datos originales de la estación Itatira con los datos de Quixada.

Finalmente, en esta última gráfica, comparamos la interpolación cuadrática con los datos originales donde si existe la presencia de la gráfica de los datos originales especialmente en algunos picos, donde se ve que contiene inconsistencias y desfases.

4.2. Cálculo de errores

Se presenta a continuación los errores obtenidos en el punto dos, en los cuales se emplean las mismas métricas que en la sección anterior.

Conjuntos	Indice de Jaccard
Datos originales - interpolación cubica	0,00
Datos originales - interpolación cuadrática	0,277777778
Interpolación cubica - interpolación cuadrática	0,277777778

Tabla 6: Resultados obtenidos para el indice de Jaccard.

Errores interpolación cubica	Valor
Error medio cuadrático	0
Error máximo	10,71
Error mínimo	0,01
Error absoluto	2,21

Tabla 7: Resultados obtenidos para los errores interpolación cubica.

Rango	Máximo	Mínimo
0-100	5,64	0,01
100-200	6,44	0,01
200-300	9,01	0,08
300-400	7,88	0,06
400-500	7,00	0,02
500-600	8,36	0,02
600-700	8,32	0,02

Tabla 8: Resultados obtenidos de máximos y mínimos para la interpolación cubica.

Errores interpolación cubica	Valor
Error medio cuadrático	10,03
Error máximo	10,71
Error mínimo	0,01
Error absoluto	1,24

Tabla 9: Resultados obtenidos para los errores interpolación cuadrática.

Rango	Máximo	Mínimo
0-100	5,64	0,01
100-200	6,44	0,01
200-300	9,01	0,08
300-400	7,88	0,06
400-500	7,00	0,02
500-600	8,36	0,02
600-700	8,32	0,02

Tabla 10: Resultados obtenidos de máximos y mínimos para la interpolación cuadrática.

5. Conclusiones

Después de haber realizado los problemas propuestos para este reto el grupo ha concluido que:

- El uso de interpolación se puede aplicar para situaciones de la vida real, dando una mejor apropación al análisis de este tipo de problemas.
- Cómo se puede evidenciar en los errores, se puede concluir que el método de spline cúbico da una precisión un poco mayor a comparación de la interpolación cuadrática en este problema. No se puede evidenciar con claridad a nivel grafico pero con ayuda del método de jaccard se evidencia esta mejora en el spline cúbico aunque por poco.
- El uso de cálculos matemáticos para el análisis de gráficas genera una buena práctica al momento de hacer uso de metodologías tales como la interpolación.
- Gracias a la interpolación y al ajuste de curvas es posible determinar futuros resultados de una variable de acuerdo a la función obtenida.
- La exactitud de una interpolación con los datos reales dependen en gran medida del polinomio con el que se este trabajando para una especifica cantidad de datos, por tal motivo, el método de interpolación lineal no es tan preciso en este caso por la gran cantidad de datos que posee.

Referencias

[1] Mathonline *Quadratic Polynomial Interpolation*, Recuperado de: http://mathonline.wikidot.com/deleted:quadratic-polynomial-interpolation

[2] Wladimiro Diaz Villanueva (1998) *7.3 Splines cúbicos*, Recuperado de: https://www.uv.es/~diaz/mn/node40.html

[3] Julio Setién (2005) *Splines cúbicos*, Recuperado de: http://artico.lma.fi.upm.es/numerico/asigs/c_numerico/cuadernos/grupo34t.pdf