https://github.com/JuanEstevezMoreno/GOPH-419-F2025-M2

Question 1 [15 marks total]:

a) **[3 marks]** Using the 11 data points for annual mean CO_2 for 2010-2020 (inclusive) form the coefficient matrix to fit cubic splines for interpolating CO_2 concentration. Recall from Lab #2 that the coefficient matrix for cubic splines with "not-a-knot" boundary conditions has the form

$$\begin{bmatrix} -\Delta x_{1,2} & (\Delta x_{0,1}+\Delta x_{1,2}) & -\Delta x_{0,1} & & & \\ \Delta x_{0,1} & 2(\Delta x_{0,1}+\Delta x_{1,2}) & \Delta x_{1,2} & & & \\ & \Delta x_{1,2} & 2(\Delta x_{1,2}+\Delta x_{2,3}) & \Delta x_{2,3} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \Delta x_{N-2,N-1} & 2(\Delta x_{N-2,N-1}+\Delta x_{N-1,N}) & \Delta x_{N-1,N} \\ & & & -\Delta x_{N-1,N} & (\Delta x_{N-2,N-1}+\Delta x_{N-1,N}) & -\Delta x_{N-2,N-1} \end{bmatrix} \begin{Bmatrix} c_0^{(3)} \\ c_1^{(3)} \\ c_2^{(3)} \\ \vdots \\ c_{N-1}^{(3)} \\ c_N^{(3)} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 3(f[x_1,x_2]-f[x_0,x_1]) \\ 3(f[x_2,x_3]-f[x_1,x_2]) \\ \vdots \\ 3(f[x_{N-1},x_N]-f[x_{N-2},x_{N-1}]) \\ 0 \end{Bmatrix}$$

Clearly identify what the $\Delta x_{i,i+1}$ values and the $f[x_i, x_{i+1}]$ values represent in terms of this data.

Screen shot below shows x_i as the year, y_i as the CO_2 data. Δx_i is the interval, which in this case is 1 year. $f[x_i, x_{i+1}]$ is the average rate of change of the CO_2 data between each year. $f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$

The RHS column is given by the formula: $r_i = 3(f_i - f_{i-1})$

i	Year x_i	CO2 y_i	Δx_i	f[xi,xi+1]	RHS
0	2010	390.10	1	1.75	0.00
1	2011	391.85	1	2.21	1.38
2	2012	394.06	1	2.68	1.41
3	2013	396.74	1	2.07	-1.83
4	2014	398.81	1	2.20	0.39
5	2015	401.01	1	3.40	3.60
6	2016	404.41	1	2.35	-3.15
7	2017	406.76	1	1.96	-1.17
8	2018	408.72	1	2.93	2.91
9	2019	411.65	1	2.56	-1.11
10	2020	414.21			

For the cubic spline the unknowns are $[c_0, c_1, c_2, \dots c_{10}]$

The interior spline equation equation is:

$$\Delta x_{i-1}, c_{i-1} + 2(\Delta x_{i-1} + \Delta x_{i+1})c_i + \Delta x_{i+1} c_{i+1} = 3(f_i - f_{i-1})$$

$\Delta x = 1$, so this simplifies to:

$$c_{i-1} + 4c_i + c_{i+1} = 3(f_i - f_{i-1})$$

Left boundary: $c_0 - 2c_1 + c_2 = 0$

Right boundary: $c_8 - 2c_9 + c_{10}$

Coefficient Matrix

1	-2	1	0	0	0	0	0	0	0	0
1	4	1	0	0	0	0	0	0	0	0
0	1	4	1	0	0	0	0	0	0	0
0	0	1	4	1	0	0	0	0	0	0
0	0	0	1	4	1	0	0	0	0	0
0	0	0	0	1	4	1	0	0	0	0
0	0	0	0	0	1	4	1	0	0	0
0	0	0	0	0	0	1	4	1	0	0
0	0	0	0	0	0	0	1	4	1	0
0	0	0	0	0	0	0	0	1	4	1
0	0	0	0	0	0	0	0	1	-2	1

b) **[2 marks]** What are the requirements for the iterative Gauss-Seidel and Jacobi iteration techniques to converge? Does the system constructed in a) satisfy this condition? Is it *guaranteed* to converge? Is it *possible* that it could converge? Be as specific as possible.

The requirement for an iterative Gauss-Seidel / Jacobi iteration to converge is for the matrix to be diagonally dominant or for the matrix to be symmetric positive.

Our matrix has 4's for most of the diagonal entries except for the first and last rows. Since it is not diagonally dominant it is not guaranteed to converge but it's still possible.

c) **[5 marks]** Solve the system in a) using either Gauss-Seidel or Jacobi iteration. Check your result either using a known working solution technique or by checking whether the solution satisfies the original equation.

Below are screenshots from excel and python. I modified my code from lab 2. midterm2-qc1.py is attached. The residuals are to the order of 10^{-13} meaning the solution satisfies the original equation.

```
(.venv) juanestevez@Mac GOPH-419-F2025-M2 % "/Users/juanestevez/Documents/Repos/Courses/GOPH 419/GOPH-419-F2025-M2/.ven
v/bin/python" "/Users/juanestevez/Documents/Repos/Courses/GOPH 419/GOPH-419-F2025-M2/tests/midterm2-qc1.py"

=== Q1(c): Not-a-knot cubic spline system ===
Points used: 2010 to 2020 (n=11)

Solution c (c0..c10):
c00 = 0.027119385125
c01 = 0.230000000000
c02 = 0.432880614875
c03 = -0.551522459499
c04 = -0.056790776878
c05 = 1.168685567010
c06 = -1.017951491164
c07 = -0.246879602356
c08 = 0.835469900589
c09 = -0.185000000000
c10 = -1.205469900589

Residual checks:
||A c - rhs||_2   = 7.020e-13
||A c - rhs||_inf = 6.983e-13
(.venv) juanestevez@Mac GOPH-419-F2025-M2 %
```

Solution
0.027
0.230
0.433
-0.552
-0.057
1.169
-1.018
-0.247
0.835
-0.185
-1.205

d) **[3 marks]** Solve for the remaining coefficients in the cubic splines using the following formulae

$$d_i^{(3)} = \frac{c_{i+1}^{(3)} - c_i^{(3)}}{3\Delta x_{i,i+1}} \qquad b_i^{(3)} = f\left[x_i, x_{i+1}\right] - c_i^{(3)}\Delta x_{i,i+1} - d_i^{(3)}\Delta x_{i,i+1}^2 \qquad a_i^{(3)} = y_i$$

and use these to interpolate the value of CO_2 concentration in March of 2015 (i.e. time = 2015.25 years).

See midterm2-q1d.py for code.

Results were 401.81 ppm for March of 2015

```
(.venv) juanestevez@Mac GOPH-419-F2025-M2 % "/Users/juanestevez/Documents/Repos/Courses/GOPH 419/GOPH-419-F2025-M2/.ven
v/bin/python" "/Users/juanestevez/Documents/Repos/Courses/GOPH 419/GOPH-419-F2025-M2/tests/midterm2-q1d.py"

=== Q1(d): spline coefficients and interpolation ===
Interval used: [2015, 2016] (i=5)
t = x - x_i = 0.25
a_i = 401.010000
b_i = 2.960193452381
c_i = 1.168685567010
d_i = -0.728879019391

CO2(2015.25) = 401.811702476355 ppm
(.venv) juanestevez@Mac GOPH-419-F2025-M2 %
```

e) **[2 marks]** Do you think that your result in d) is an accurate representation of CO_2 concentration in March of 2015? Why or why not?

The results are not an accurate representation of CO_2 concentration in March because the dataset is annual mean and does not include monthly or seasonal data. It is an estimate of what the annual-mean trend curve could look like

Question 2 [15 marks total]:

a) **[4 marks]** Show the form of Newton's interpolating polynomials for 0- through 3rd-order. Clearly define all terms in the expression, in general and in terms of the CO_2 concentration data. Which is the *independent variable* and which is the *dependent variable*? In which order should data points be incorporated in this expression to minimize truncation error?

Let x = time (years) = independent

y = CO_2 concentration (ppm) = dependent

Divided differences

$$f(x_i) = y_i$$

$$f(x_i, x_{i+1}) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$

$$f(x_i, x_{i+1}, x_{i+2}) = \frac{f(x_{i+1}, x_{i+2}) - f(x_i, x_{i+1})}{x_{i+2} - x_i}$$

Newton Polynomails

0th order: $P_0(x) = f(x_0)$

1st order: $P_1(x) = f(x_0) + f(x_0, x_1)(x - x_0)$

2nd order: $P_2(x) = P_1(x) + f(x_0, x_1, x_2)(x - x_0)(x - x_1)$

3rd order: $P_3(x) = P_2(x) + f(x_0, x_1, x_2, x_3)(x - x_0)(x - x_1)(x - x_2)$

To minimize error you would use the centered method. Start by incorporated points closest to 2015.25.

Order to use: 2015, 2016, 2014, 2017, 2013, 2018...

b) **[3 marks]** What are the assumptions involved in using polynomial interpolation as a curve fitting technique? Compare and contrast polynomial interpolation with other curve fitting approaches? Do you think that polynomial interpolation would be a good approach for extrapolating to predict future CO_2 concentration? Why or why not?

Assumption in polynomial interpolation:

- Data is exact, meaing no errors and the interpolation passes through every point
- The underlying function is smooth and continuous.

Other Curve Fitting approaches:

Polynomail interpolation: passes exactly through chosen points

Splines: Piecewise low-order polynomials, usually smoother & more stable. Avoids Ruge's phenomenon.

Least-Squares: Finds curve of best fit, it doesn't pass through all of the points. It is good for noisy data

Polynomial interpolation is generally not good for extrapolation. They tend to diverge outside of the data range, especially at higher orders. Using it to predict CO_2 would be unreliable

c) **[4 marks]** Use the Mauna Loa annual mean CO_2 data to estimate the concentration in March of 2015 (i.e. time = 2015.25) using Newton's interpolating polynomials until at least 4 significant digits have converged. What order of polynomial was required and how many data points were incorporated into the estimate?

See excel sheet for calculations.
A 4^{th} order polynomial was required and 5^{th} data points, years 2013-2017 were incorporated into the estimate.

d) **[4 marks]** Compare your result in part c) with your result from Question 1? Are they consistent? Considering the uncertainty estimates provided in the data set, do you think that 4 significant digits of precision is needed from the polynomial interpolation? How many significant digits of precision would be consistent with the data uncertainty? What order of polynomial interpolation would be sufficient for that target?

The result from Q1 was 401.812 ppm. My result from from part c was 401.8 ppm. A difference of 0.02-0.03 ppm making them consistent. The uncertainty for the data is ±0.12 ppm, meaning it's only meaningful to the nearest 0.1 ppm. Reporting 401.8 ppm is appropriate, 401.8120 ppm implies false percision.

To get a stable 0.1ppm, you need the 3rd order result which use 4 points.

Question 3 [15 marks total]:

a) **[5 marks]** Use *2nd-order* and *4th-order* centred differences to estimate the *first derivative* of CO_2 concentration in ppm/yr over the period 1959-2024. Compare and contrast these formulae. What order of interpolating polynomial would have been used as a basis in each case? Plot the data and assess whether the results are consistent.

https://github.com/JuanEstevezMoreno/GOPH-419-F2025-M2

See excel sheet for work

2nd-order centered difference

$$f'(t_i) \approx \frac{f_{i+1} - f_{i-1}}{2h} + O(\Delta x^2)$$

4th-order centered difference

$$f'(t_i) \approx \frac{f_{i-2} - 8f_{i-1} + 8f_{i+1} - f_{i+2}}{12h} + O(\Delta x^4)$$

The 2nd order centered difference is the derivative of a 2nd order interpolating polynomial. It uses 3 points.

The 4th order centered difference is the derivative of a 4th order interpolating polynomial. It uses 5 points.

b) **[4 marks]** Use *2nd-order* and *4th-order* <u>centred</u> differences to estimate the *second derivative* of CO_2 concentration in ppm/yr^2 over the period 1959-2024. Compare and contrast these formulae. Plot the data and assess whether the results are consistent.

See excel sheet for work

2nd order centered difference

$$f''(t_i) \approx \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + O(\Delta x^2)$$

4th order centered difference

$$f''(t_i) \approx \frac{-f_{i+2} + 16f_{i+1} - 30f_i + 16f_{i-1} - f_{i-2}}{12h^2} + \frac{f^{(16)}(5^6)}{90} (\Delta x^4)$$

The 2nd order centered difference is the 2nd derivative of a 2nd order interpolating polynomial. It uses 3 points.

The 4th order centered difference is the 2nd derivative of a 4th order interpolating polynomial It uses 5 points

c) **[5 marks]** Based on your results from parts a) and b), comment on the overall trend in CO_2 concentration. Is the rate of change slowing down, approximately constant, or increasing? How certain are you of this? [*Hint:* Consider the graphs of both $d[CO_2]/dt$ and $d^2[CO_2]/dt^2$ and also the mean and standard deviation of the $d^2[CO_2]/dt^2$ data.]

See excel sheet for work

The first derivative $d[CO_2]/dt$ is alway positive which means there is an increase in CO_2 every year. The plot of the first derivative also shows a general uptrend meaning the rate of increase has increased overall.

The mean of the second derivative (both 2nd and 4th order) is positive meaning the rate is slowly increasing. The standard deviation however shows theres a lot of fluctuation year to year.

Overall concentration is seen to be increasing and the rate it's increasing seems to also be increasing.

d) **[BONUS, 5 marks]** Using your results from b), provide a 95% *confidence interval* on the rate of change of the rate of increase in CO_2 concentration. That is, what range of values is the true value of $d^2[CO_2]/dt^2$ 95% likely to fall between, based on this data?

See excel sheet for math

The confidence interval are:

2nd order 2nd derivative: [-0.1146, 0.1959]

4th order 2nd derivative: [-0.1646, 0.2226]