Analysis of CO_2 Levels: Imputation and Derivative Approximation Using Lagrange Polynomials

1. Introduction

The dataset analyzed in this project contains annual global CO_2 levels recorded over multiple decades. However, the dataset includes missing values for certain years, which necessitated imputation techniques to estimate these values. In this study, I employed polynomial interpolation using Lagrange polynomials of degrees 1 (linear), 2 (quadratic), and 3 (cubic) to fill in missing values. Additionally, I approximated the derivatives of the dataset to study trends and rate of change in CO_2 levels over time.

2. Data Overview and Missing Values

The dataset consists of yearly recorded CO_2 levels from 1940 to 2024. However, certain years, such as 1945, 1955, 1965, 1975, 1985, 1995, 2005, and 2015, were missing from the dataset. Since CO_2 levels tend to follow a continuous trend, we utilized interpolation methods to estimate the missing values based on existing data points.

3. Imputation Using Lagrange Polynomial Interpolation

To impute missing values, I used **Lagrange polynomial interpolation**, which constructs a polynomial that passes through given data points. We applied three different degrees of

$$p_n(x) = \sum_{j=0}^{n} y_j L_j^n(x)$$

where Lagrange Polynomials (LPs) are:

$$L_j^n(x) = \prod_{i=0, i \neq j}^{n} \frac{x - x_i}{x_j - x_i}$$

polynomial interpolation:

1. **Linear Interpolation (Degree 1)**
    - Linear interpolation estimates missing values by constructing a straight line between the two nearest known data points.
    - This method assumes a constant rate of change between points.

    - While simple, it may not capture the fluctuations in CO_2 levels accurately.
2. **Quadratic Interpolation (Degree 2)**
    - This method uses three neighboring data points to construct a quadratic polynomial.
    - It provides a smoother approximation compared to linear interpolation.
    - Quadratic interpolation accounts for acceleration or deceleration trends in CO_2 levels.
3. **Cubic Interpolation (Degree 3)**
    - Using four data points, cubic interpolation captures more complex trends in CO_2 level fluctuations.
    - This method is expected to provide a more accurate representation, especially when the data follows a nonlinear trend.

Each of these methods was applied to estimate missing CO_2 levels in the dataset, and comparisons were made to assess their effectiveness.

4. Derivative Approximation for Trend Analysis

After imputing the missing values, I approximated the **derivatives of CO_2 levels** to analyze how the rate of change evolved over time. The derivative approximation was conducted as follows:

- **Linear Approximation (Secant Line Method)**:
    - Using the basic finite difference formula:

      $F'(a) = (y2 - y1) / (x2 - x1)$

    - This provides an estimate of the rate of change between consecutive years.
- **Quadratic and Cubic Approximation of Derivatives**:
    - By differentiating the quadratic and cubic polynomials derived in the imputation process, I obtained smoothed approximations of the rate of change.
    - These higher-order approximations helped me in identifying trends such as periods of rapid CO_2 level growth and stabilization.

5. Observations and Results

- The linear interpolation provided a simplistic approach but did not fully capture fluctuations in CO_2 levels.
- Quadratic and cubic approximations offered smoother and more reliable estimates of missing values.
- The derivative analysis revealed periods of accelerated CO_2 growth, particularly in later decades, indicating increasing emissions.
- Differences among the interpolation methods became more noticeable in periods with significant variations in CO_2 levels.

6. Conclusion

This project demonstrated the effectiveness of polynomial interpolation in estimating missing values and using derivatives to analyze trends in CO_2 levels. Among the interpolation methods, **cubic interpolation** provided the most accurate estimations due to its ability to capture nonlinear trends. Additionally, derivative approximations helped identify periods of significant growth in CO_2 levels, highlighting the increasing emissions trend over the decades.