Interpolacja

Sprawozdanie z laboratorium 3

Jakub Grześ

22.03.2024

1 Treść zadań

Zadanie 1. Populacja Stanów Zjednoczonych na przestrzeni lat przedstawiała się następująco:

Rok	Populacja
1900	76 212 168
1910	92 228 496
1920	106 021 537
1930	123 202 624
1940	132 164 569
1950	151 325 798
1960	179 323 175
1970	203 302 031
1980	226 542 199

Istnieje dokładnie jeden wielomian ósmego stopnia, który interpoluje powyższe dziewięć punktów, natomiast sam wielomian może być reprezentowany na różne sposoby. Rozważamy następujące zbiory funkcji bazowych $\varphi_j(t)$, $j = 1; \ldots ; 9$:

- $\varphi_j(t) = t^{j-1}$
- $\varphi_j(t) = (t - 1900)^{j-1}$
- $\varphi_j(t) = (t - 1940)^{j-1}$
- $\varphi_j(t) = \left(\frac{t-1940}{40}\right)^{j-1}$

(a) Dla każdego z czterech zbiorów funkcji bazowych utwórz macierz Vandermonde'a.

(b) Oblicz współczynnik uwarunkowania każdej z powyższych macierzy, używając funkcji `numpy.linalg.cond`.

(c) Używając najlepiej uwarunkowanej bazy wielomianów, znajdź współczynniki wielomianu interpolacyjnego dla danych z zadania. Narysuj wielomian interpolacyjny. W tym celu użyj schematu Hornera i oblicz na przedziale [1900,1990] wartości wielomianu w odstępach jednorocznych. Na wykresie umieść także węzły interpolacji.

(d) Dokonaj ekstrapolacji wielomianu do roku 1990. Porównaj otrzymaną wartość z prawdziwą wartością dla roku 1990, wynoszącą 248 709 873. Ile wynosi błąd względny ekstrapolacji dla roku 1990?

(e) Wyznacz wielomian interpolacyjny Lagrange'a na podstawie 9 węzłów interpolacji podanych w zadaniu. Oblicz wartości wielomianu w odstępach jednorocznych.

(f) Wyznacz wielomian interpolacyjny Newtona na podstawie tych samych węzłów interpolacji i oblicz wartości wielomianu w odstępach jednorocznych.

(g) Zaokrąglij dane podane w tabeli do jednego miliona. Na podstawie takich danych wyznacz wielomian interpolacyjny ósmego stopnia, używając najlepiej uwarunkowanej bazy z podpunktu (c). Porównaj wyznaczone współczynniki z współczynnikami obliczonymi w podpunkcie (c). Wyjaśnij otrzymany wynik.

2 Metoda rozwiązania problemu

2.1 Utworzenie macierzy Vandermonde'a

Dla każdej z zadanych funkcji bazowych zbudowano macierz Vandemonde'a postaci:

$$\begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \varphi_3(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \varphi_3(x_2) & \cdots & \varphi_n(x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \varphi_3(x_n) & \cdots & \varphi_n(x_n) \end{bmatrix}$$

Dzięki niej, współczynniki $a_1, ..., a_n$ wielomianu znajdziemy rozwiązując równanie:

$$\begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \cdots & \varphi_n(x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

gdzie punkty (x_i, y_i) są węzłami interpolacji.

Dla każdej ze zbudowanych macierzy obliczono współczynnik uwarunkowania z użyciem funkcji linalg.cond. Najlepiej uwarunkowaną macierzą była ta skonstruowana dla funkcji bazowej $\varphi_j(t) = \left(\frac{t-1940}{40}\right)^{j-1}$ z wynikiem 1605.44.

2.2 Obliczenie współczynników wielomianu interpolacyjnego

Współczynniki wielomianu obliczono, używając równania z 2.1. oraz funkcji linalg.solve rozwiązującej równanie macierzowe $Ax = b$ ze względu na x.

```
coef = np.linalg.solve(V_1940_div, population)
```

Dysponując współczynnikami, można było wizualizować wielomian interpolacyjny. Obliczono jego wartość w odstępach jednorocznych, korzystając ze

schematu Hornera. Pozwala on przedstawić wielomian w postaci:
$P(x) = a_0 + x(a_1 + x(a_2 + ... + x(a_{n-1} + xa_n)...))$ ograniczając w ten sposób liczbę mnożeń do n gdzie n jest stopniem wielomianu.

```
def horner(x, coef):
    x = (x - 1940) / 40 #skalowanie argumentu
    n = len(coef)
    result = coef[n - 1]
    for i in range(n - 2, -1, -1):
        result = result * x + coef[i]
    return result
```

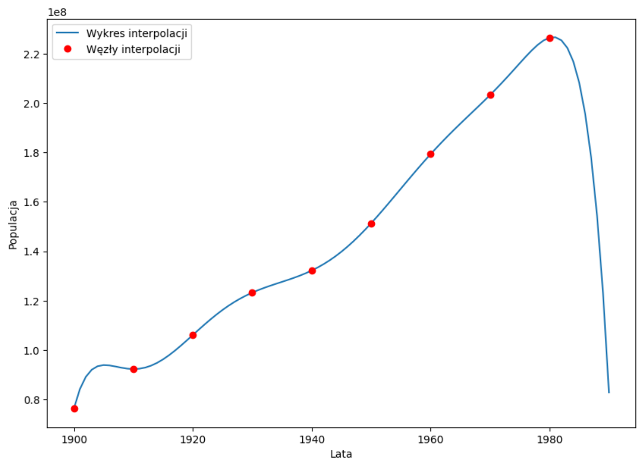

Figure 1: Wielomian interpolacyjny

Dokonano także ekstrapolacji, czyli obliczenia wartości znajdującej się poza zakresem posiadanych danych. Oszacowano populację Stanów Zjednoczonych w roku 1990, używając znalezionych współczynników, a następnie porównano tę wartość z prawdziwą.

Interpolowana wartość wynosi: 82,749,141. Różni się ona od rzeczywistej o 165,960,731, co oznacza błąd względny na poziomie 66.73%.

Błąd jest bardzo wysoki. Brak danych w pobliżu poszukiwanej wartości może wpływać na precyzję otrzymanych danych.

2.3 Wielomian interpolacyjny Lagrange'a

Obliczono także wielomian interpolacyjny Lagrange'a postaci:

$L(x) = \sum_{i=1}^{n} y_i \cdot L_i(x)$, gdzie L_i jest funkcją:
$L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}$
Poniższy kod wykonano, jako argument podając punkty z treści zadania i kolejne lata. Następnie wyniki przedstawiono na wykresie.

```
def calculate_lagrange(x, points):
    res = 0
    n = len(points)
    for i, (x_i, y_i) in enumerate(points):
        l_i = 1

        for j in range(n):
            if j != i:
                x_j, _ = points[j]
                l_i *= (x - x_j) / (x_i - x_j)
        res += y_i * l_i
    return res
```

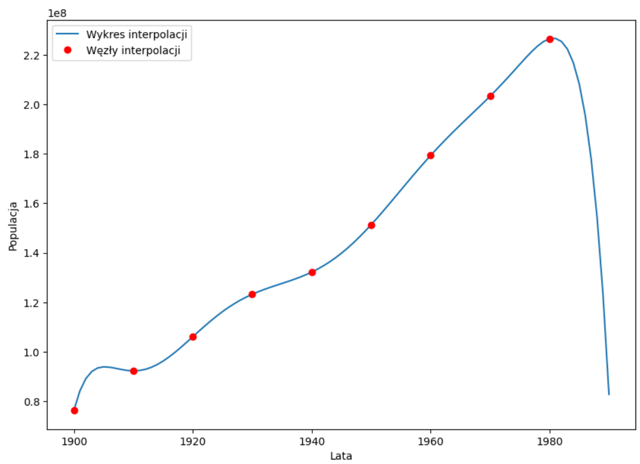

Figure 2: Wielomian interpolacyjny Lagrange'a

2.4 Wielomian interpolacyjny Newtona

Dla danych z zadania obliczono także wielomian interpolacyjny Newtona. Zaletą tej metody jest to, że jeśli chcielibyśmy włączyć do danych nowy punkt, nie trzeba przeliczać wszystkich współczynników i wystarczy rozszerzyć istniejący wielomian. Wielomian interpolacyjny Newtona ma postać:

$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \ldots + a_n(x - x_0)(x - x_1)\ldots(x - x_{n-1})$

gdzie $a_0, a_1, \ldots, a_n$ są ilorazami różnicowymi, obliczanymi według wzoru:

$$a_i = f[x_0, x_1, \ldots, x_i]$$

Ilorazy różnicowe są definiowane jako:

$$f[x_i] = f(x_i)$$

$$f[x_i, x_{i+1}, \ldots, x_j] = \frac{f[x_{i+1}, \ldots, x_j] - f[x_i, \ldots, x_{j-1}]}{x_j - x_i}$$

```python
def newton_diff(points): #Obliczanie ilorazow roznicowych
    n = len(points)
    diff = np.zeros((n, n))
    for i in range(n):
        diff[i][0] = points[i][1]

    for j in range(1, n):
        for i in range(n-j):
            diff[i][j] = (diff[i + 1][j - 1] - diff[i][j - 1])
            / (points[i + j][0] - points[i][0])

    return diff[0]

def calculate_newton(x, points): # Obliczanie wartosci wielomianu
    coef = newton_diff(points)
    n = len(points)
    res = coef[n - 1]

    for i in range(n - 2, -1, -1):
        res = res * (x - points[i][0]) + coef[i]

    return res
```

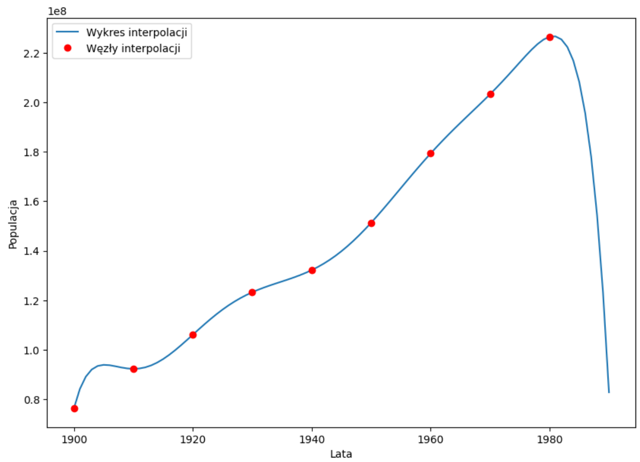

Figure 3: Wielomian interpolacyjny Newtona

2.5 Wykonanie obliczeń dla zaokrąglonych danych wejściowych

Sporządzono zestaw danych, w którym dane dotyczące populacji z danych wejściowych zaokrąglono do jednego miliona. Następnie korzystając z macierzy ustalonej w 2.1., obliczono współczynniki wielomianu ósmego stopnia, a następnie porównano wyniki z tymi otrzymanymi dla niezmienionych danych.

Największy błąd względny wynosi 6.66%, najmniejszy 0.12%, a średni 2.62%.

Mimo znacznego zmniejszenia precyzji danych wyniki nie odbiegają zbytnio od tych otrzymanych wcześniej. Dzięki dużej liczbie punktów, wymuszającej wysoki stopień wielomianu, nie jest on bardzo wrażliwy na zaokrąglanie danych i wciąż pozwala znaleźć ogólną tendencję.
Dla znalezionych współczynników dokonano także ekstrapolacji populacji dla roku 1990.

Interpolowana wartość wynosi: 109000000. Różni się ona od rzeczywistej o 139709872, oznacza to błąd względny na poziomie 56.17%.

Dla zaokrąglonych współczynników błąd ekstrapolacji wciąż był wysoki, ale istotnie niższy niż dla współczynników obliczanych na podstawie prawdziwych danych. Może wynikać to z tego, że zaokrąglanie upraszcza model, zachowując ogólny trend co może prowadzić do poprawy jego jakości np. w zakresie przewidywań wartości będących poza zakresem danych.

3 Wnioski

Wybór właściwej funkcji bazowej okazał się kluczowym elementem zapewniającym numeryczną stabilność procesu interpolacji. Dostosowanie skali danych wejściowych pozwoliło znacząco poprawić uwarunkowanie macierzy. Zaobserwowano również, że interpolacja wielomianowa może nie być najlepszym rozwiązaniem w przypadku ekstrapolacji. W takich okolicznościach konieczne jest ostrożne interpretowanie wyników ekstrapolacji, zwłaszcza przy braku dodatkowych informacji. Wyniki ekstrapolacji okazały się lepsze po zaokrągleniu danych dotyczących liczby ludności do pełnych milionów, mogło to być spowodowane tym, że wysoki stopień wielomianu jest wrażliwy na niewielkie zmiany wartości, utrącając zdolność do reprezentowania ogólnej tendencji. Choć analizowany zestaw danych nie był obszerny, zastosowanie schematu Hornera przyspiesza obliczanie wartości wielomianu, redukując potrzebne operacje mnożenia do $O(n)$ w porównaniu do $O(n^2)$ dla metody naiwnej. Inną istotną obserwacją jest przewaga metody Newtona nad metodą Lagrange'a, głównie ze względu na możliwość dodawania nowych punktów danych bez konieczności ponownego przeliczania wszystkich współczynników wielomianu. Dodatkowo, przeprowadzona analiza wykazała, że metoda interpolacji wykazuje odporność na zaokrąglenia danych wejściowych. Jest to szczególnie istotne w sytuacjach, gdy dane mogą charakteryzować się niedokładnością. W przyszłych pracach warto zwrócić szczególną uwagę na wybór metody interpolacyjnej najlepiej dopasowanej do specyfiki danego problemu, jak również na zachowanie ostrożności przy interpretowaniu wyników ekstrapolacji, ewentualnie modyfikując dane tak aby przystosować model do tego zadania.

Bibliografia

[1] Marcin Kuta, *Interpolation*

[2] Katarzyna Rycerz, *Wykład 2, Interpolacja*