# ỨNG DỤNG PHƯƠNG PHÁP NỘI SUY NEWTON XẤP XỈ MỘT SỐ DỮ LIỆU THỰC TẾ

**Lê Thị Thu**[(1)]
*(1) Trường Đại học Thủ Dầu Một*
*Ngày nhận bài 10/9/2024; Chấp nhận đăng 01/10/2024*
*Liên hệ email: thult.khtn@tdmu.edu.vn*

## *Tóm tắt*

*Trong bài viết này, chúng tôi sử dụng phương pháp nội suy Newton để xây dựng hàm gần đúng cho một số dữ liệu thực tế như dữ liệu về thu nhập và chi tiêu bình quân đầu người Việt Nam, cũng như dữ liệu về diện tích và sản lượng lúa cả năm của Việt Nam. Các dữ liệu được dùng trong bài báo này được chúng tôi trích xuất từ website của Tổng cục Thống kê Việt Nam. Chúng tôi cũng sử dụng phần mềm Matlab để lập trình tìm đa thức nội suy, cũng như tính giá trị của hàm nội suy tại những giá trị cụ thể. Từ hàm nội suy thu được, chúng tôi đã dự đoán được giá trị còn thiếu trong bảng dữ liệu về thu nhập và chi tiêu bình quân đầu người tại Việt Nam ở các năm 2019, 2021. Ngoài ra, sản lượng lúa cả năm cũng được chúng tôi ước lượng bằng kết quả của đa thức nội suy. Sau khi tính toán cho thấy độ chênh lệch về sản lượng trung bình giữa dữ liệu thực tế và dữ liệu ước lượng từ hàm nội suy là khá nhỏ.*

***Từ khóa:*** *dữ liệu, nội suy Newton, ứng dụng*

***Abstract***

***APPLICATION OF THE NEWTON INTERPOLATION METHOD TO APPROXIMATE SOME REAL DATA***

*In this paper, we use the Newton polynomial interpolation method to build approximate functions for some real data in Vietnam such as data on expenditure and income per capita, data on rice productivity and rice cultivation area. The data used in this paper are extracted from the General Statistics Office of Vietnam. We also develop some Matlab programs to find the coefficients of interpolation functions and calculate the values of the interpolation functions at specific points. Using the obtained interpolation function, some missing values in the data on income and expenditure per capita of Vietnam in 2019 and 2021 are estimated. We also estimated the rice productivity using the results of the interpolation polynomial. We found that the difference in average rice productivity between the calculated data from the interpolation function is quite small.*

## 1. Đặt vấn đề

Trong toán học chúng ta thường gặp các bài toán liên quan đến khảo sát và tính giá trị các hàm $y = f(x)$ nào đó. Tuy nhiên, trong thực tế có nhiều trường hợp chúng ta không xác định được biểu thức của hàm $f(x)$ mà chỉ nhận được các giá trị rời rạc $y_i$ tại các điểm tương ứng $x_i \in [a, b]$. Từ những dữ liệu đó, chúng ta cần ước lượng các dữ liệu tại những điểm khác thuộc đoạn $[a, b]$. Để giải quyết vấn đề này thì mô hình Toán học thường được sử dụng là phương pháp nội suy. Phương pháp nội suy giúp chúng ta xây dựng hàm gần đúng cho bộ dữ liệu $(x_i, y_i)$ và từ đó chúng ta có thể dễ dàng ước lượng được giá trị của $f(x)$ tại bất kỳ $x \in [a, b]$ (Anh, 2008; Chương và nnk., 2009; Đĩnh, 2009; Phương & Nhân, 2022).

Đã có một số bài viết về ứng dụng của nội suy đa thức. Nguyễn Thu Thủy và Lê Thị Hà (2022) đã ứng dụng đa thức nội suy Lagrange để giải quyết các bài toán ứng dụng trong nông nghiệp và y tế. Senol Celik (2018) cũng dùng đa thức nội suy Lagrange cho dữ liệu ứng dụng trong chăn nuôi. Một vài tác giả khác sử dụng kết hợp đa thức nội suy Newton và nội suy Lagrange để ứng dụng giải phương trình vi phân cấp một (Apichat Neamvonk & Boonyong Sriponpaew, 2023). Tuy nhiên, theo tìm hiểu của chúng tôi, chưa có bài viết về ứng dụng của đa thức nội suy Newton để ước lượng cho các dữ liệu về thu nhập và chi tiêu của người dân Việt Nam, cũng như ước lượng cho diện tích và sản lượng cây trồng.

Trong bài viết này, chúng tôi sử dụng đa thức nội suy Newton để xây dựng hàm gần đúng cho dữ liệu về chi tiêu và thu nhập bình quân đầu người ở Việt Nam, cũng như dữ liệu về diện tích và sản lượng lúa cả năm của Việt Nam. Các dữ liệu này được chúng tôi trích xuất từ website của Tổng cục Thống kê Việt Nam. Ngoài ra, chúng tôi cũng sử dụng phần mềm Matlab để lập trình tìm các hệ số của đa thức nội suy Newton. Việc sử dụng phần mềm để lập trình thay cho việc tính toán thủ công là cần thiết để giảm thời gian tính toán, cũng như hạn chế những sai sót khi tính toán thủ công (Bảo & Hiếu, 2014; Bảo & Việt, 2016).

## 2. Cơ sở khoa học và phương pháp nghiên cứu

Bài viết được thực hiện dựa trên việc phân tích các nghiên cứu và tài liệu liên quan, thu thập số liệu thực tế từ Tổng cục Thống kê Việt Nam và sử dụng công cụ là phần mềm Matlab để lập trình hỗ trợ cho việc tính toán.

Bài viết sử dụng các kiến thức liên quan về đa thức nội suy Newton như dưới đây.

**Tỷ sai phân:**

Xét hàm số $y = f(x)$ xác định trên đoạn $[a, b]$ và cho trước $n + 1$ giá trị $(x_i, y_i), i = \overline{0, n}$. Các tỷ sai phân các cấp của hàm $f$ được xác định như dưới đây (Anh, 2008):

Tỷ sai phân cấp 1:

$$f[x_i, x_{i-1}] = \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}, i = \overline{1, n}$$

Tỷ sai phân cấp 2:

$$f[x_{i+1}, x_i, x_{i-1}] = \frac{f[x_{i+1}, x_i] - f[x_i, x_{i-1}]}{x_{i+1} - x_{i-1}}, i = \overline{1, n-1}$$

Tương tự, tỷ sai phân cấp $n$:

$$f[x_n, x_{n-1}, \dots, x_1, x_0] = \frac{f[x_n, x_{n-1}, \dots, x_1] - f[x_{n-1}, \dots, x_1, x_0]}{x_n - x_0}$$

Tỷ sai phân có các tính chất: tính chất tuyến tính; tính chất đối xứng; tỷ sai phân của hằng số luôn bằng 0.

**Đa thức nội suy Newton:**

Cho hàm số $y = f(x)$ xác định trên đoạn $[a, b]$ và một phân hoạch $\Delta = \{a = x_0 < x_1 < \cdots < x_n = b\}$. Ta xây dựng đa thức $P_n(x)$ có bậc không quá $n$ sao cho $P_n(x_i) = f(x_i), i = \overline{0, n}$.

Đa thức nội suy Newton tiến xuất phát từ điểm $x_0$ được xác định theo công thức dưới đây (Anh, 2008):

$$P_n(x) = f(x_0) + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2] + \cdots$$
$$+ (x - x_0)(x - x_1) \dots (x - x_{n-1})f[x_0, x_1, \dots, x_n] \quad (1)$$

## 3. Kết quả và thảo luận

### *3.1 Xây dựng chương trình tìm đa thức nội suy Newton*

Rõ ràng rằng, việc tính toán thủ công để tìm đa thức nội suy với dữ liệu lớn hoặc dữ liệu là các số thập phân là khá mất thời gian, và dễ gặp sai sót khi tính toán. Trong phần này, chúng tôi sử dụng phần mềm Matlab để lập trình chương trình hỗ trợ cho bài viết này. Chương trình được xây dựng để tìm đa thức nội suy Newton với dữ liệu đầu vào là hai vector X, Y chứa các giá trị $x_i, i = \overline{0, n}$ và $y_i, i = \overline{0, n}$ tương ứng. Kết quả mà chương trình trả về là vector P chứa các hệ số của đa thức nội suy Newton nội suy tập dữ liệu (X, Y).

Code của chương trình tìm đa thức nội suy Newton như sau:

```
function P = newton(X, Y)
format rat;
% Tim cac ty sai phan, TSP(i) = f[x1, x2,..,xi]
TSP = Y;
n = length(X);
for i=2:n
 for j=n:-1:i
 TSP(j) = (TSP(j)-TSP(j-1))/(X(j)-X(j-i+1));
 end
end
syms x;
P = Y(1);
for i=2:n
 G = 1;
 for j=1:i-1
 G = G*(x - X(j));
 end
 P = P + TSP(i)*G;
end
P = sym2poly(P);
```

### *3.2 Ứng dụng đa thức nội suy Newton để xấp xỉ một số dữ liệu thực tế*

*3.2.1 Ứng dụng đa thức nội suy Newton để xấp xỉ dữ liệu về thu nhập và chi tiêu bình quân đầu người Việt Nam*

Trong phần này, chúng tôi sử dụng phương pháp nội suy Newton để xây dựng đa thức nội suy xấp xỉ cho dữ liệu gồm thu nhập (X1) và chi tiêu (Y1) bình quân đầu người Việt Nam (đơn vị tính: triệu đồng/tháng). Các số liệu này được trích từ website của Tổng cục Thống kê Việt Nam. Bảng số liệu được thể hiện ở bảng 1. Trong dữ liệu này, chúng tôi nhận thấy rằng số liệu về thu nhập bình quân đầu người năm 2019 và 2021 được thu thập, tuy nhiên số liệu về chi tiêu bình quân đầu người ở các năm này bị khuyết.

**Bảng 1.** *Thu nhập và chi tiêu bình quân đầu người Việt Nam*

| Năm | 2008 | 2010 | 2012 | 2014 | 2016 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|---|---|---|
| Thu nhập (X1) | 0,995 | 1,387 | 2,000 | 2,637 | 3,098 | 3,874 | 4,295 | 4,250 | 4,205 | 4,673 |
| Chi tiêu (Y1) | 0,792 | 1,211 | 1,603 | 1,888 | 2,157 | 2,546 | | 2,892 | | 2,795 |

Từ dữ liệu ở bảng 1, chúng tôi sử dụng công thức (1) để xây dựng đa thức nội suy Newton xấp xỉ cho dữ liệu gồm thu nhập (X1) và chi tiêu (Y1) của các năm 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022. Sử dụng chương trình 1 mà chúng tôi đã trình bày ở mục 2.2, với các câu lệnh:

```
>> X1 = [0.995, 1.387, 2.000, 2.637, 3.098, 3.874, 4.250, 4.673];
>> Y1 = [0.792, 1.211, 1.603, 1.888, 2.157, 2.546, 2.892, 2.795];
>> newton(X1, Y1)
```

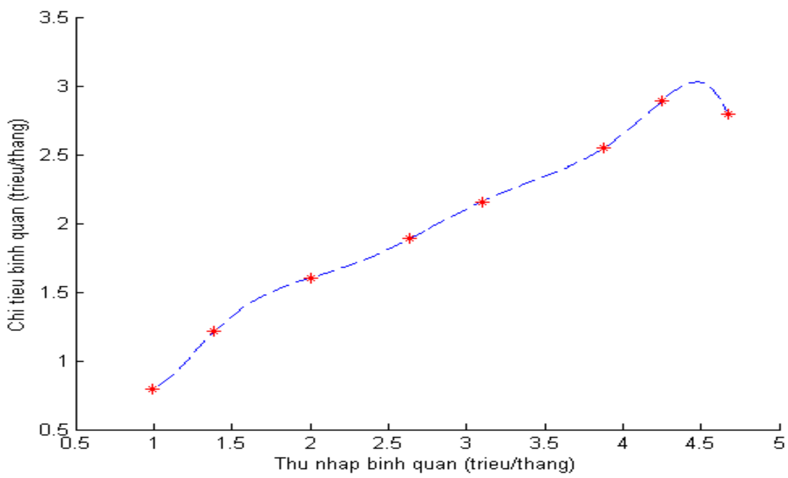
chúng tôi thu được kết quả sau:

```
ans = -443/9800 281/325 -4548/665 20118/697 -4953/71 7756/81
-17018/251 7212/365
```

Như vậy, đa thức nội suy Newton cho tập dữ liệu (X1, Y1) là:

$$P_1(x) = -\frac{443}{9800}x^7 + \frac{281}{325}x^6 - \frac{4548}{665}x^5 + \frac{20118}{697}x^4 - \frac{4953}{71}x^3 + \frac{7756}{81}x^2 - \frac{17018}{251}x + \frac{7212}{365}$$

Đồ thị của đa thức nội suy $P_1(x)$ cho dữ liệu về thu nhập và chi tiêu bình quân đầu người của Việt Nam được thể hiện ở hình 1. Các dấu * trong đồ thị chính là vị trí các điểm $(x_i, y_i)$ của bảng dữ liệu. Dễ thấy rằng, đồ thị của đa thức nội suy thu được luôn đi qua các điểm dữ liệu.




***Hình 1**. Đồ thị của đa thức nội suy Newton cho dữ liệu về thu nhập và chi tiêu*

Từ đa thức nội suy thu được, chúng ta có thể ước lượng mức chi tiêu bình quân đầu người mỗi tháng dựa vào thu nhập bình quân của họ. Chẳng hạn, với thu nhập bình quân $x_0$ triệu/tháng thì mức chi tiêu bình quân tính theo đa thức nội suy sẽ là $P_1(x_0)$ triệu/tháng. Ở đây, chúng tôi ước lượng các giá trị còn thiếu ở bảng 1. Năm 2019, thu nhập bình quân đầu người là 4,295 triệu/tháng nên ước lượng cho số liệu về chi tiêu bình quân năm 2019 là $P_1(4{,}295) \approx 2{,}9332$ triệu/tháng. Tương tự, chúng tôi tìm được ước lượng cho chi tiêu bình quân đầu người năm 2021 là $P_1(4{,}205) \approx 2{,}8481$ triệu/tháng.

Các giá trị của đa thức nội suy có thể tìm bởi hàm polyval trong Matlab như các câu lệnh bên dưới đây:

```
>> P = newton(X1, Y1);
>> Z1 = [4.295, 4.205];
>> F = polyval(P, Z1)
```

*3.2.2. Ứng dụng đa thức nội suy Newton để xấp xỉ dữ liệu về diện tích và sản lượng lúa cả năm của Việt Nam*

Trong mục này, chúng tôi sử dụng phương pháp nội suy Newton để tìm đa thức nội suy xấp xỉ cho dữ liệu gồm diện tích trồng lúa (X2) và sản lượng lúa (Y2) cả năm của Việt Nam. Các số liệu này được trích từ website của Tổng cục Thống kê Việt Nam. Bảng số liệu được thể hiện ở bảng 2.

***Bảng 2.*** *Diện tích và sản lượng lúa cả năm của Việt Nam*

| Năm | 2018 | 2019 | 2020 | 2021 | 2022 | Sơ bộ 2023 |
|---|---|---|---|---|---|---|
| Diện tích (triệu ha) (X2) | 7,5709 | 7,4699 | 7,2789 | 7,2389 | 7,1089 | 7,1193 |
| Sản lượng (triệu tấn) (Y2) | 44,046 | 43,4954 | 42,7648 | 43,8526 | 42,6608 | 43,4977 |

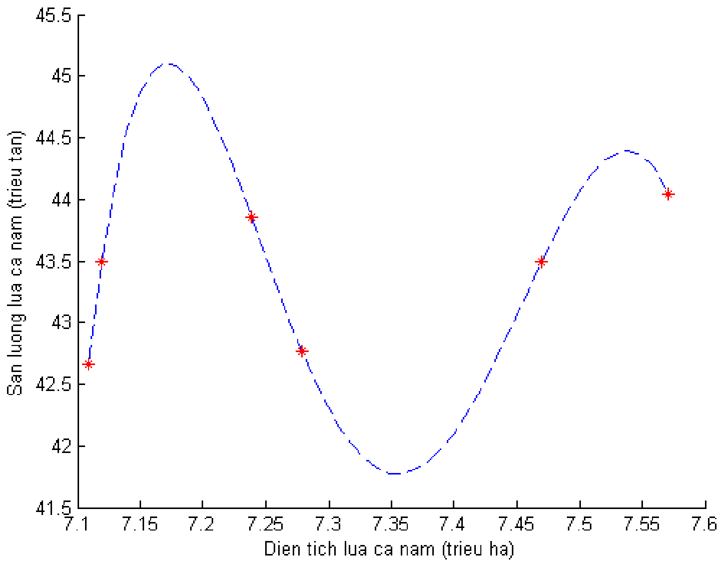
Chúng tôi nhận thấy rằng, dữ liệu về diện tích ở bảng 2 không theo thứ tự tăng dần. Vì vậy, để sử dụng phương pháp nội suy Newton, chúng ta cần sắp xếp lại dữ liệu của diện tích theo thứ tự tăng dần. Sau khi sắp xếp lại dữ liệu, sử dụng chương trình 1 mà chúng tôi đã trình bày ở mục 2.2, với các câu lệnh:

```
>> X2 = [7.1089, 7.1193, 7.2389, 7.2789, 7.4699, 7.5709];
>> Y2 = [42.6608, 43.4977, 43.8526, 42.7648, 43.4954, 44.046];
>> newton(X2, Y2)
```

chúng tôi thu được đa thức nội suy Newton cho tập dữ liệu (X2, Y2) như sau:

$$P_2(x) = 2555{,}51162352\,x^5 - 96631{,}06324599\,x^4 + 1460309{,}96046021\,x^3 - \\ -11025197{,}14226459\,x^2 + 41586593{,}80971701\,x - 62697099{,}99204298$$

Hình 2 thể hiện đồ thị của đa thức nội suy $P_2(x)$ cho dữ liệu về diện tích và sản lượng lúa cả năm của Việt Nam. Các dấu * trong đồ thị chính là vị trí các điểm $(x_i, y_i)$ của bảng dữ liệu. Chúng ta có thể dễ dàng nhận thấy rằng, đa thức nội suy luôn đi qua các mốc nội suy $(x_i, y_i)$.




***Hình 2.*** *Đồ thị của đa thức nội suy Newton cho dữ liệu về diện tích và sản lượng lúa cả năm*

Từ đa thức nội suy thu được, chúng ta có thể ước lượng sản lượng lúa bình quân dựa vào diện tích ứng với các giá trị trong đoạn [7,1089; 7,5709]. Với diện tích $x_0$ thì sản lượng lúa tính theo đa thức nội suy sẽ là $P_2(x_0)$. Trong phần này, chúng tôi tính sản lượng lúa tương ứng với các diện tích 7,15; 7,2; 7,25; 7,3; 7,4; 7,45; 7,5. Chúng tôi sử dụng các câu lệnh sau:

```
>> Z2 = [7.15, 7.2, 7.25, 7.3, 7.4, 7.45, 7.5];
>> P = newton(X2, Y2)
>> F = polyval(P, Z2)
```

Kết quả thu được sau khi làm tròn tới 4 số sau dấu phẩy được thể hiện ở bảng 3.

***Bảng 3.*** *Ước lượng sản lượng lúa theo dện tích gieo trồng bằng kết quả nội suy Newton*

| Diện tích (triệu ha) | 7,15 | 7,2 | 7,25 | 7,3 | 7,4 | 7,45 | 7,5 |
|---|---|---|---|---|---|---|---|
| Sản lượng (triệu tấn) | 44,8764 | 44,8161 | 43,5384 | 42,3043 | 42,1113 | 43,0621 | 44,0680 |

Từ số liệu ở bảng 3, chúng tôi tính được năng suất lúa trung bình là 5,94686 tấn/ha. Trong khi đó, giá trị trung bình từ dữ liệu thực tế ở bảng 2 là 5,94511 tấn/ha. Như vậy, độ chênh lệch về năng suất lúa trung bình giữa dữ liệu thực tế và dữ liệu của các giá trị ước lượng từ đa thức nội suy là 0,00175 tấn/ha.

## 4. Kết luận

Bài viết đã sử dụng phương pháp nội suy Newton để xây dựng hàm gần đúng cho dữ liệu về thu nhập và chi tiêu bình quân đầu người của Việt Nam theo số liệu trích từ Tổng cục Thống kê Việt Nam từ năm 2008 đến 2022, cũng như dữ liệu về sản lượng lúa và diện tích trồng lúa ở Việt Nam từ năm 2018 đến 2023. Từ hàm nội suy thu được, chúng tôi ước lượng các giá trị còn khuyết trong bảng dữ liệu về thu nhập và chi tiêu bình quân đầu người của Việt Nam ở các năm 2019 và 2021. Kết quả chỉ ra rằng, chi tiêu bình quân đầu người năm 2019 là 2,9332 triệu/tháng và năm 2021 là 2,8481 triệu/tháng. Ngoài ra, chúng tôi cũng đã đưa ra ước lượng cho sản lượng lúa theo một số giá trị của diện tích trồng lúa. Khi tính trung bình của dữ liệu được ước lượng bằng hàm nội suy, chúng tôi thu được năng suất lúa trung bình ở Việt Nam là 5,94686 tấn/ha, xấp xỉ với giá trị trung bình tính từ dữ liệu thực tế là 5,94511 tấn/ha. Kết quả này cũng phù hợp với kết quả mà Thủy và Hà (2022) đã ứng dụng đa thức nội suy Lagrange để ước lượng cho dữ liệu về diện tích và sản lượng lúa ở Đồng bằng sông Hồng và Đồng bằng sông Cửu Long.

Thêm nữa, chúng tôi cũng đã sử dụng phần mềm Matlab để viết chương trình tìm các hệ số của đa thức nội suy, cũng như tính giá trị của đa thức nội suy tại những điểm cụ thể. Việc xây dựng các chương trình hỗ trợ sẽ giúp tránh khỏi những sai sót khi tính toán, đặc biệt hữu ích khi dữ liệu lớn hoặc dữ liệu là các số thập phân. Chương trình này có thể được sử dụng cho dữ liệu trong các nghiên cứu ở những lĩnh vực khác nhau như kinh tế, nông nghiệp,… nhằm dự đoán các giá trị chưa biết, để từ đó đưa ra các đề xuất, điều chỉnh phù hợp để ứng dụng trong thực tế.

## TÀI LIỆU THAM KHẢO

[1] Apichat Neamvonk, Boonyong Sriponpaew (2023). Solving the First Order Differential Equations using Newton's Interpolation and Lagrange Polynomial. *European journal of pure and applied mathematics, 16*(2), 965-974.

[2] Nguyễn Minh Chương (ch.b) (2009). *Giải tích số*. NXB Giáo dục.

[3] Nguyễn Thu Thủy và Lê Thị Hà (2022). Một số ứng dụng của đa thức nội suy Lagrange trong thực tế. *Tạp chí Khoa học trường Đại học Sư phạm Hà Nội*, *67*(1), 10-18.

[4] Phạm Kỳ Anh (2008). *Giải tích số*. NXB Đại học Quốc gia Hà Nội.

[5] Phạm Thế Bảo và Huỳnh Trung Hiếu (2014). *Tính toán số và Matlab (Phần cơ bản)*. NXB Đại học Quốc gia Thành phố Hồ Chí Minh.

[6] Phạm Thế Bảo và Ngô Quốc Việt (2016). *Tính toán số và Matlab (Phần nâng cao)*. NXB Đại học Quốc gia Thành phố Hồ Chí Minh.

[7] Senol Celik (2018). Using Lagrange interpolation to determine the milk production amount by the number of milked animals. *American Journal of Engineering Research*, *7*(8), 264-271.

[8] Tạ Văn Đĩnh (2009). *Phương pháp tính*. NXB Giáo dục.

[9] Tổng cục Thống kê (2023). *Niên giám thống kê*. NXB Thống kê.

[10] Trần Minh Phương và Nguyễn Thành Nhân (2022). *Giải tích số và ứng dụng (Phần cơ bản)*. NXB Đại học Sư phạm Thành phố Hồ Chí Minh.