

Dương Ngọc Sơn

Bài giảng

Phương pháp tính

Với ngôn ngữ lập trình R

30 tháng 1, 2026

Copyright ©2024–2026 by Dương Ngọc Sơn.



Tài liệu này có bản quyền theo Giấy phép Creative Commons Ghi công-Phi thương mại-Chia sẻ tương tự 4.0. Để xem bản sao của các giấy phép này, hãy truy cập <https://creativecommons.org/licenses/by-nc-sa/4.0/> hoặc gửi thư đến Creative Commons PO Box 1866, Mountain View, CA 94042, Hoa Kỳ.

Bạn có thể sử dụng, in ấn, sao chép và chia sẻ cuốn sách này bao nhiêu tùy ý. Bạn có thể dựa vào nội dung này để ghi chú của riêng mình và sử dụng lại các phần nếu giữ nguyên giấy phép. Các tác phẩm phái sinh phải được ghi rõ ràng là phái sinh.

Lời nói đầu

Tài liệu này gồm 6 chương, được biên soạn dựa trên các bài giảng môn *Phương pháp tính* tại Trường Đại học Phenikaa trong học kỳ I năm học 2022–2023, và đã được sửa đổi, bổ sung trong các năm học tiếp theo.

Với mục đích ban đầu là cung cấp cho sinh viên một tài liệu tham khảo phù hợp với chương trình học, tài liệu trình bày các nội dung theo đề cương chi tiết học phần *Phương pháp tính* (2 tín chỉ) do Khoa Khoa học cơ bản xây dựng, với thời lượng 30 tiết. Tuy nhiên, trong quá trình biên soạn, một số chủ đề như hàm ghép trơn, phương pháp bình phương bé nhất, công thức cầu phương Gauss, v.v., đã được bổ sung, giúp tài liệu có thể sử dụng cho các học phần 45 tiết.

Tài liệu được biên soạn dựa trên một số giáo trình quen thuộc ở Việt Nam như sách *Phương pháp tính* của GS. Tạ Văn Đĩnh và sách *Giải tích số* của GS. Phạm Kỳ Anh cùng với một số giáo trình nước ngoài khác như cuốn *Advanced Engineering Mathematics (Part E: Numeric Analysis)*, 10th Edition của E. Kreyszig.

Ngoài những nội dung mang tính lý thuyết thuần túy, tài liệu cũng đề cập đến công cụ tính toán là phần mềm thống kê và ngôn ngữ lập trình R. Phần mềm R đóng vai trò quan trọng trong các mục thực hành với hệ đại số máy tính (gọi tắt là CAS), được thiết kế nhằm minh họa và đánh giá các phương pháp tính toán. Ngoài ra, một số tính toán và hình vẽ trong tài liệu cũng được thực hiện bằng R.

Tài liệu được biên soạn trong thời gian ngắn, song song với việc giảng dạy trên lớp năm học 2022–2023. Mặc dù trong các năm học 2023–2024 và 2024–2025 tài liệu đã được sửa đổi và bổ sung khá nhiều, nhưng chắc chắn vẫn còn không ít thiếu sót. Tác giả hi vọng tài liệu này sẽ hữu ích cho việc giảng dạy và học tập môn *Phương pháp tính* ở bậc đại học, và rất mong nhận được những ý kiến đóng góp của bạn đọc.

Hà Nội, tháng 1 năm 2026

Dương Ngọc Sơn

Mục lục

1 Sai số	1
1.1 Khái niệm sai số, chữ số có nghĩa, chữ số chắc	2
1.1.1 Sai số thật sự, sai số tuyệt đối	2
1.1.2 Sai số tương đối	3
1.1.3 Chữ số có nghĩa, chữ số chắc	6
1.1.4 Cách viết số gần đúng	7
1.1.5 Quy tròn số gần đúng	8
1.1.6 Quy tròn số trong bất đẳng thức	10
1.1.7 Mối liên hệ giữa quy tròn số và chữ số có nghĩa	10
1.2 Sai số phép toán	11
1.2.1 Sai số của một tổng	12
1.2.2 Sai số của một tích	13
1.2.3 Sai số của một thương	13
1.2.4 Công thức tổng quát cho đánh giá sai số tích lũy	14
1.3 Phân loại sai số	14
1.3.1 Sai số phương pháp	15
1.3.2 Sai số tính toán	15
1.4 Giới thiệu ngôn ngữ lập trình R	16
1.5 Câu hỏi và bài tập	22
2 Tính gần đúng nghiệm của một phương trình	25
2.1 Nghiệm và khoảng tách nghiệm	26
2.1.1 Nghiệm của một phương trình	26
2.1.2 Nghiệm xấp xỉ và khoảng tách nghiệm	30
2.2 Phương pháp chia đôi	33
2.3 Phương pháp lặp đơn	36
2.3.1 Sự hội tụ	36
2.3.2 Đánh giá sai số	42

2.3.3	Tóm tắt phương pháp lặp đơn	43
2.4	Phương pháp Newton–Raphson	44
2.4.1	Mô tả phương pháp Newton–Raphson	44
2.4.2	Sự hội tụ	45
2.4.3	Đánh giá sai số	50
2.4.4	Tóm tắt thuật toán Newton–Raphson	53
2.5	Phương pháp dây cung và một số phương pháp khác	55
2.5.1	Phương pháp dây cung	55
2.5.2	Phương pháp Steffensen	57
2.5.3	Phương pháp vị trí sai (regula falsi)	58
2.6	Câu hỏi và bài tập	60
3	Tính gần đúng nghiệm của một hệ	63
3.1	Hệ phương trình tuyến tính. Phương pháp Gauss	63
3.1.1	Định nghĩa. Sự tồn tại và duy nhất nghiệm	64
3.1.2	Phương pháp Gauss	65
3.2	Phương pháp lặp đơn	70
3.2.1	Sự hội tụ của phương pháp lặp đơn	71
3.3	Phương pháp lặp Seidel	81
3.3.1	Sự hội tụ của phương pháp lặp Seidel	82
3.3.2	Một ví dụ: Mô hình Leontief	86
3.4	Câu hỏi và bài tập	88
4	Đa thức nội suy và xấp xỉ đa thức	91
4.1	Bài toán nội suy, đa thức nội suy	92
4.1.1	Bài toán nội suy	92
4.1.2	Sự tồn tại và duy nhất của đa thức nội suy	92
4.1.3	Tính giá trị của đa thức: Lược đồ Horner	95
4.2	Đa thức nội suy Lagrange	99
4.3	Đa thức nội suy Newton	103
4.3.1	Đa thức nội suy Newton tổng quát	103
4.3.2	Đa thức nội suy Newton với các nút nội suy cách đều	110
4.3.3	Đọc thêm: Định lý Hermite–Genocchi	114
4.4	Sai số nội suy	115
4.4.1	Đánh giá sai số trong xấp xỉ bằng đa thức nội suy	115
4.4.2	Chọn nút nội suy tối ưu: đa thức Chebyshev	118
4.5	Nội suy hàm ghép trơn	123
4.5.1	Sơ lược về hàm ghép trơn	123
4.5.2	Hàm ghép trơn bậc ba	124
4.5.3	Quy trình tính hàm ghép trơn bậc ba	128
4.6	Phương pháp bình phương bé nhất	133
4.6.1	Xấp xỉ tuyến tính	134
4.6.2	Xấp xỉ đa thức bằng phương pháp bình phương bé nhất	139

4.6.3	Xấp xỉ bằng các hàm lũy thừa	143
4.6.4	Xấp xỉ bằng hàm số mũ	146
4.7	Câu hỏi và bài tập	148
5	Tính gần đúng đạo hàm và tích phân xác định	151
5.1	Tính gần đúng tích phân xác định	151
5.1.1	Quy tắc hình thang	152
5.1.2	Quy tắc Simpson	157
5.1.3	Quy tắc hình thang hiệu chỉnh biên	164
5.1.4	Công thức cầu phương Gauss	167
5.1.5	Tích phân bằng hàm có sẵn trong R	173
5.2	Tính gần đúng đạo hàm	175
5.2.1	Xây dựng công thức tính đạo hàm bằng khai triển Taylor	175
5.2.2	Tính gần đúng đạo hàm bằng đa thức nội suy	179
5.3	Câu hỏi và bài tập	183
6	Giải gần đúng phương trình vi phân	185
6.1	Bài toán giá trị ban đầu	186
6.2	Phương pháp Euler	187
6.2.1	Mô tả phương pháp Euler	188
6.2.2	Sự hội tụ và sai số của phương pháp Euler	189
6.2.3	Phương pháp Euler ngược	194
6.3	Phương pháp Runge–Kutta và các phương pháp khác	198
6.3.1	Một số khái niệm	198
6.3.2	Khai triển tiệm cận của sai số	198
6.3.3	Phương pháp hình thang	200
6.3.4	Phương pháp Heun	201
6.3.5	Phương pháp Runge–Kutta	202
6.3.6	Sử dụng deSolve của R giải gần đúng phương trình vi phân	204
6.4	Phương trình vi phân bậc cao và hệ phương trình	206
6.4.1	Hệ phương trình vi phân	206
6.4.2	Phương trình vi phân cấp cao	208
6.5	Sơ lược về bài toán biên	209
6.6	Câu hỏi và bài tập	212
	Tài liệu tham khảo	212
	Bảng thuật ngữ	213

Chương

1

Sai số

Chương này trình bày những khái niệm cơ bản về số gần đúng, sai số của chúng, cách viết các số gần đúng, quy tròn số, sai số của quá trình quy tròn, ước lượng sai số của một biểu thức.

Nội dung

1.1	Khái niệm sai số, chữ số có nghĩa, chữ số chắc	2
1.1.1	Sai số thật sự, sai số tuyệt đối	2
1.1.2	Sai số tương đối	3
1.1.3	Chữ số có nghĩa, chữ số chắc	6
1.1.4	Cách viết số gần đúng	7
1.1.5	Quy tròn số gần đúng	8
1.1.6	Quy tròn số trong bất đẳng thức	10
1.1.7	Mối liên hệ giữa quy tròn số và chữ số có nghĩa	10
1.2	Sai số phép toán	11
1.2.1	Sai số của một tổng	12
1.2.2	Sai số của một tích	13
1.2.3	Sai số của một thương	13
1.2.4	Công thức tổng quát cho đánh giá sai số tích lũy	14
1.3	Phân loại sai số	14
1.3.1	Sai số phương pháp	15
1.3.2	Sai số tính toán	15
1.4	Giới thiệu ngôn ngữ lập trình R	16
1.5	Câu hỏi và bài tập	22

1.1 Khái niệm sai số, chữ số có nghĩa, chữ số chắc

Mục này nói về những khái niệm cơ bản nhất của giải tích số, đó là sai số của số gần đúng. Tài liệu tham khảo: T. V. Đình [6, §1.1].

Khi làm việc với những vấn đề thực tế, chúng ta thường gặp các giá trị gần đúng (cũng gọi là giá trị xấp xỉ) của các đại lượng. Đó là các giá trị mà sự sai khác so với giá trị đúng của đại lượng đang xét là nhỏ, có thể bỏ qua mà không làm ảnh hưởng tới những khía cạnh khác đáng quan tâm. Ví dụ, diện tích của một mặt bàn hình tròn bán kính 1m có thể coi một cách gần đúng là $3,14\text{m}^2$. Đó là vì chúng ta lấy π xấp xỉ bằng 3,14 (viết $\pi \approx 3,14$) và coi mặt bàn là hoàn toàn tròn. Sai lệch của nó đối với kết quả đúng trong trường hợp này là nhỏ và có thể bỏ qua trong những vấn đề đơn giản.

1.1.1 Sai số thật sự, sai số tuyệt đối

Nếu a là một số gần đúng của một đại lượng A có giá trị thật sự a^* , thì hiệu số

$$\epsilon_a := a^* - a \quad (1.1)$$

gọi là *sai số thật sự* (actual error) của a . Như vậy, *giá trị đúng bằng tổng của giá trị gần đúng và sai số*.

Chúng ta thường quan tâm nhiều hơn đến độ lớn của sai lệch giữa hai giá trị đúng và gần đúng mà bỏ qua thông tin giá trị nào lớn hơn. Độ sai lệch này cho bởi giá trị tuyệt đối $|\epsilon_a| = |a^* - a|$ của sai số, gọi là *sai số tuyệt đối* (absolute error) của a .

Thông thường, chúng ta muốn tìm một giá trị gần đúng a của A khi giá trị thật a^* không thể hoặc khó có thể tìm được. Trong những tình huống ấy, không thể tính được sai số thật sự và sai số tuyệt đối của a . Tuy nhiên, trong nhiều trường hợp, có thể tìm được số dương Δ_a sao cho

$$|a^* - a| \leq \Delta_a. \quad (1.2)$$

Số Δ_a thỏa mãn bất đẳng thức (1.2) cho một đánh giá sai số của số gần đúng a được gọi là *sai số tuyệt đối giới hạn* của số gần đúng. Như vậy, sai số tuyệt đối của một số gần đúng không vượt qua sai số tuyệt đối giới hạn của nó.

Ví dụ 1.1. Đo nồng độ NaCl trong nước biển lấy từ một vùng biển miền Trung, người ta thu được kết quả là 35g/l, với sai số không vượt quá 0,01 (g/l). Hãy chỉ ra một giá trị gần đúng của nồng độ NaCl trong nước biển nói trên cùng với một sai số tuyệt đối giới hạn của nó.

Lời giải

Trong ví dụ này, người ta đã chỉ ra rằng $a = 35$ là giá trị gần đúng của nồng độ NaCl trong nước biển với đánh giá sai số $|\epsilon_a| \leq 0,01$, có nghĩa là có thể chọn $\Delta_a = 0,01$.

Lưu ý rằng, giá trị sai số tuyệt đối giới hạn là do chúng ta chọn cho phù hợp với mục đích của chúng ta, nhưng phải thỏa mãn ước lượng (1.2). Tuy nhiên, sai số tuyệt đối giới hạn càng nhỏ thì (1.2) càng tốt. Vì vậy, trong các điều kiện phù hợp, chúng ta chọn các số $\Delta_a > 0$ bé nhất có thể thỏa mãn điều kiện (1.2).

Nếu số gần đúng a của một đại lượng A có sai số tuyệt đối giới hạn Δ_a , thì có thể viết

$$A = a \pm \Delta_a.$$

Có nghĩa là nếu a^* là giá trị thực sự của A thì

$$a^* - \Delta_a \leq a \leq a^* + \Delta_a.$$

Ví dụ, hằng số π có một giá trị gần đúng là 3,14, với một sai số tuyệt đối giới hạn có thể chọn là 0,0016. Vậy, chúng ta có thể viết

$$\pi = 3,14 \pm 0,0016.$$

Nếu không chỉ rõ sai số tuyệt đối giới hạn, chúng ta có thể viết $\pi \approx 3,14$, đọc là “ π xấp xỉ 3,14”.

1.1.2 Sai số tương đối

Để đánh giá sai số trong mối liên hệ với độ lớn của bản thân số đó, chúng ta có khái niệm sai số tương đối: Nếu a là một số gần đúng của $a^* \neq 0$ thì tỉ số

$$\delta_a := \frac{|a^* - a|}{|a^*|} \quad (1.3)$$

gọi là *sai số tương đối* (relative error) của a .

Từ đánh giá sai số tương đối, có thể đánh giá được sai số tuyệt đối. Thật vậy, bởi bất đẳng thức tam giác quen thuộc của hàm giá trị tuyệt đối, chúng ta có

$$|a^*| = |(a^* - a) + a| \leq |a^* - a| + |a|.$$

Do đó, chúng ta có

$$|a^* - a| = \delta_a |a^*| \leq \delta_a (|a^* - a| + |a|),$$

và vậy thì

$$|a^* - a|(1 - \delta_a) \leq |a|\delta_a.$$

Bất đẳng thức trên chỉ có ý nghĩa khi $\delta_a < 1$. Với giả thiết này, chúng ta có

$$|a^* - a| \leq |a| \left(\frac{\delta_a}{1 - \delta_a} \right). \quad (1.4)$$

Về phải của (1.4) cho một sai số tuyệt đối giới hạn của một số gần đúng a tính qua sai số tương đối của nó.

Ngược lại, nếu Δ_a là một sai số tuyệt đối giới hạn của a :

$$|a^* - a| \leq \Delta_a,$$

thì ta suy ra

$$\delta_a \leq \frac{\Delta_a}{|a| - \Delta_a}, \quad (1.5)$$

với điều kiện $\Delta_a < |a|$.

Ví dụ 1.2. Đo một góc A , người ta thu được kết quả là $a = 20^\circ 45'$. Biết rằng sai số tương đối của phép đo không lớn hơn 0,15%. Hãy tìm một sai số tuyệt đối giới hạn của nó.

Lời giải

Một độ tương đương với 60 phút ($'$). Như vậy, khi đổi ra đơn vị phút thì

$$a = 20^\circ 45' = 20 \times 60' + 45' = 1,245'.$$

Theo giả thiết $\delta_a \leq 0,15\% = 0,0015$. Do đó,

$$\Delta_a = |a| \left(\frac{\delta_a}{1 - \delta_a} \right) \leq 1,245 \times \left(\frac{0,0015}{1 - 0,0015} \right) \approx 1,87031 \quad (').$$

Vậy, $A = 20^\circ 45' \pm 1,87031'$.

Ví dụ 1.3. Người ta dùng hai thiết bị đo, A và B , để đo nồng độ NaCl trong một dung dịch. Thiết bị A cho kết quả 98,82 (mg/l) và thiết bị B cho kết quả 98,56 (mg/l). Biết kết quả cho bởi thiết bị A có sai số tương đối không vượt quá 0,5%.

1. Hãy ước lượng sai số tương đối của kết quả cho bởi thiết bị B .
2. Nếu một thiết bị C cho kết quả là 51,16 (mg/l) thì có thể nói gì về sai số tương đối của kết quả này?

Lời giải

Gọi $a = 98,82$ và a^* là giá trị đúng của nồng độ muối NaCl trong dung dịch đó. Giả thiết về sai số tương đối của a có thể được viết lại là

$$\delta_a = \frac{|a^* - a|}{|a^*|} \leq 0,005, \quad (1.6)$$

Chúng ta hãy xét hai trường hợp sau đây:

Trường hợp 1. $a^* \geq a$. Khi đó, từ (1.6), chúng ta thu được

$$a^* - 98,82 = |a - a^*| \leq 0,005|a^*| = 0,005a^*,$$

và từ đó kết luận được

$$a \leq a^* \leq \frac{98,82}{0,995} \leq 99,32.$$

Vậy, nếu $\tilde{a} = 98,56$ là kết quả đọc được từ B thì

$$\delta_{\tilde{a}} = \frac{|a^* - \tilde{a}|}{|a^*|} = \frac{a^* - 98,56}{a^*} \leq \frac{99,32 - 98,56}{98,82} \approx 0,00769 \leq 0,77\%.$$

Trường hợp 2. $0 < a^* < a$. Tương tự như trường hợp 1, chúng ta bỏ dấu giá trị tuyệt đối trong (1.6) và giải bất phương trình thu được để có

$$a^* \geq \frac{98,82}{1,005} \approx 98,328 \geq 98,32.$$

Chúng ta lại chia làm 2 trường hợp con:

Trường hợp 2(a): $a^* < \tilde{a}$. Khi đó,

$$|a^* - \tilde{a}| = \tilde{a} - a^* \leq 98,56 - 98,32 = 0,24.$$

Vậy,

$$\delta_{\tilde{a}} = \frac{|a^* - \tilde{a}|}{|a^*|} \leq \frac{0,24}{98,32} \approx 0,002441 < 0,25\%.$$

Trường hợp 2(b): $a^* \geq \tilde{a}$. Khi đó,

$$|a^* - \tilde{a}| = a^* - \tilde{a} \leq 98,82 - 98,56 = 0,26.$$

Từ đó suy ra

$$\delta_{\tilde{a}} = \frac{|a^* - \tilde{a}|}{|a^*|} \leq \frac{0,26}{98,56} \approx 0,00264 \leq 0,27\%.$$

Như vậy, chúng ta kết luận $\tilde{a}$ xấp xỉ a^* với sai số tương đối không vượt quá 0,77%.

Gọi $a' = 51,16$ là kết quả đọc từ C. Trong trường hợp 1 ở trên, ta có đánh giá

$$\frac{98,82 - 51,16}{99,32} \leq \delta_{a'} \leq \frac{99,32 - 51,16}{98,82}.$$

Từ đây, sau khi quy tròn, chúng ta thu được

$$48,00\% \leq \delta_{a'} \leq 48,74\%.$$

Còn trong trường hợp thứ hai, $98,32 \leq a^* \leq 98,82$, nên

$$\frac{98,32 - 51,16}{98,82} \leq \delta_{a'} \leq \frac{98,82 - 51,16}{98,32}.$$

Từ đây, sau khi quy tròn, chúng ta thu được

$$47,72\% \leq \delta_{a'} \leq 48,75\%.$$

Như vậy, chúng ta kết luận được sai số tương đối của a' không nhỏ hơn 47,72% và không vượt quá 48,75%.

1.1.3 Chữ số có nghĩa, chữ số chắc

Mỗi số thực a có một biểu diễn thập phân dạng $a_t a_{t-1} \cdots a_0, a_{-1} a_{-2} \cdots$, có nghĩa là

$$a = \sum_{s \leq t} a_s 10^s, \quad a_s \in \{0, 1, 2, \dots, 9\}. \quad (1.7)$$

Vế phải của (1.7) có thể là tổng hữu hạn hoặc là một chuỗi vô hạn hội tụ (có vô hạn a_s với s là số nguyên âm). Ví dụ, một số gần đúng của π là số $a = 3,1416$ có thể viết là

$$3,1416 = 3 \times 10^0 + 1 \times 10^{-1} + 4 \times 10^{-2} + 1 \times 10^{-3} + 6 \times 10^{-4},$$

tức là có dạng (1.7) với $a_0 = 3$, $a_{-1} = 1$, $a_{-2} = 4$, $a_{-3} = 1$, và $a_{-4} = 6$.

Trong biểu diễn thập phân của một số gần đúng, các chữ số kể từ chữ số khác không đầu tiên tính từ trái sang phải gọi là các *chữ số có nghĩa* (significant digit). Ví dụ, các số 3,14 và số 0,00314 cùng có 3 chữ số có nghĩa. Ba chữ số 0 trong biểu diễn của 0,00314 dùng để xác định vị trí dấu thập phân, nhưng không có nghĩa trong việc xác định sai số tương đối của nó.

Trong cách viết thập phân một số gần đúng a của một đại lượng A với giá trị thực sự a^* , “độ tin cậy” của một chữ số phụ thuộc vào khả năng đánh giá độ chính xác của số gần đúng ấy, tức là phụ thuộc vào một sai số tuyệt đối giới hạn đã biết. Để nói về điều này, chúng ta có khái niệm *chữ số chắc* (correct digit) của một biểu diễn thập phân của một số gần đúng.

Định nghĩa 1.1:

Cho a là một số gần đúng có biểu diễn (1.7) và sai số tuyệt đối giới hạn Δ_a .

- (i) Một chữ số a_s trong biểu diễn (1.7) là một *chữ số chắc* (theo nghĩa hẹp) nếu

$$\Delta_a \leq 0,5 \times 10^s.$$

- (ii) Một chữ số a_s trong biểu diễn (1.7) là một *chữ số chắc* (theo nghĩa rộng) nếu

$$\Delta_a \leq 10^s.$$

Nhận xét 1.1. Lưu ý rằng một chữ số là chắc hay không phụ thuộc vào một sai số tuyệt đối của số đó mà chúng ta biết được. Ví dụ, chúng ta biết rằng $\pi \approx 3,14$ với sai số tuyệt đối Δ_a có thể chọn là 0,0016. Do $\Delta_a < 0,5 \times 10^{-2}$ nên, theo Định nghĩa 1.1, chữ số 4 ở hàng phần trăm là chữ số chắc theo nghĩa hẹp. Hơn nữa, mọi chữ số có nghĩa nằm về bên trái của một chữ số chắc đều là chữ số chắc nên xấp xỉ $a = 3,14$ của π có 3 chữ số chắc.

1.1.4 Cách viết số gần đúng

Như đã nói ở trên, chúng ta có thể viết số gần đúng a với sai số tuyệt đối giới hạn Δ_a của một đại lượng A như sau:

$$A = a \pm \Delta_a.$$

Trong cách viết thập phân của một số gần đúng, chúng ta viết theo quy ước “*Mọi chữ số có nghĩa là chữ số chắc (theo nghĩa hẹp)*”. Khi đó, một số gần đúng thập phân có sai số tuyệt đối giới hạn *không lớn hơn một nửa đơn vị của chữ số cuối cùng*.

Lưu ý, với quy ước như trên, biểu diễn thập phân của một số gần đúng còn chứa đựng thông tin về sai số tuyệt đối của số đó.

Ví dụ 1.4. Chúng ta có thể viết số gần đúng của số π với 3 chữ số thập phân là chữ số chắc như sau:

$$\pi \approx 3,142.$$

Đó là vì chúng ta đã biết rằng

$$\Delta_{3,142} = |\pi - 3,142| < 0,5 \times 10^{-3}.$$

Với 4 chữ số thập phân là chữ số chắc, ta có

$$\pi \approx 3,1416.$$

Ta viết được như vậy bởi vì

$$\Delta_{3,1416} = |\pi - 3,1416| < 0,5 \times 10^{-4}.$$

Ví dụ 1.5. Tính sai số tuyệt đối giới hạn của các số gần đúng sau đây, biết rằng trong biểu diễn thập phân của chúng, tất cả chữ số có nghĩa đều là chữ số chắc theo nghĩa hẹp.

(a) $a = 0,024$

(b) $b = 0,3200$

Lời giải

- (a) Biểu diễn thập phân của số gần đúng a có 2 chữ số chắc trong đó chữ số chắc ở hàng bé nhất là chữ số 4 ở hàng 10^{-3} . Vì vậy ta có

$$\Delta_a \leq 0,5 \times 10^{-3}.$$

- (b) Vì chữ số chắc ở hàng bé nhất của biểu diễn thập phân của b là chữ số 0 ở hàng 10^{-4} nên

$$\Delta_b \leq 0,5 \times 10^{-4}.$$

1.1.5 Quy tròn số gần đúng

Mục này nói về quy tròn số (hay làm tròn số), và sai số của quá trình đó. Tài liệu tham khảo: T. V. Đĩnh [6, §1.3].

Khi làm việc với các số thập phân, chúng ta thường phải bỏ bớt các chữ số cho gọn. Những số hữu tỉ như $1/3$, $1/7$, những số vô tỉ như π , e , hoặc $\sqrt{3}$, không thể biểu diễn được dưới dạng số thập phân có hữu hạn các chữ số. Do đó, việc bỏ đi một số chữ số trong biểu diễn thập phân của chúng là điều không tránh khỏi. Việc bỏ bớt các chữ số của một số viết dưới dạng thập phân gọi là *quy tròn số*, hay *làm tròn số*, hoặc *thu gọn* (round-off). Hiển nhiên, việc quy tròn số sinh ra một sai số mới, gọi là *sai số quy tròn* (round-off error), là hiệu của số đã quy tròn và chưa quy tròn. Trị số tuyệt đối của hiệu đó gọi là *sai số quy tròn tuyệt đối*.

Khi quy tròn một số gần đúng, người ta thường chọn quy tắc quy tròn thỏa mãn điều kiện sau: “*Sai số quy tròn tuyệt đối không lớn hơn một nửa đơn vị của hàng cuối cùng của số đã quy tròn.*” Như vậy, 4,014749 quy tròn tới 4,01475. Trong trường hợp chữ số cuối là 5, ta có hai sự lựa chọn: quy tròn lên, và quy tròn xuống. Trong trường hợp này, nói chung, ta quy tròn sao cho chữ số ở hàng ngay trước đó về bên trái trở thành chữ số chẵn. Khi đó, 4,01475 quy tròn tới 4,0148. Ví dụ này cũng cho thấy rằng quá trình quy tròn số *không có tính bắc cầu*: số 4,014749 quy tròn trực tiếp thành 4,0147.

Với quy tắc này, nếu một số gần đúng a có biểu diễn thập phân có mọi chữ số có nghĩa đều là chữ số chẵn, thì mọi quy tròn của a cũng thỏa mãn điều kiện này. Ví dụ, cơ số của logarit tự nhiên e có biểu diễn thập phân vô hạn không tuần hoàn:

$$e \approx 2,7182818.$$

Quy tròn số xấp xỉ này ta có các số quy tròn là 2,718282 (bỏ bớt 1 số), 2,71828 (bỏ bớt 2 chữ số), hoặc 2,718 (bỏ bớt 4 chữ số).

Giả sử a là một số xấp xỉ của một đại lượng A với giá trị đúng a^* . Nếu $\bar{a}$ là một số quy tròn của a với sai số quy tròn tuyệt đối $|a - \bar{a}|$ thì, bởi bất đẳng thức tam giác, ta có

$$|a^* - \bar{a}| = |a^* - a + a - \bar{a}| \leq |a^* - a| + |a - \bar{a}|.$$

Như vậy, sai số tuyệt đối của số quy tròn không vượt quá tổng của sai số tuyệt đối của số ban đầu và sai số tuyệt đối quy tròn.

Ví dụ 1.6. Biết đại lượng A có giá trị gần đúng $a = 3,1388$ với sai số tương đối δ_a không vượt quá 3,5%. Quy tròn số a được một số gần đúng $\tilde{a}$ của A với hai chữ số sau dấu thập phân. Hãy tìm một sai số tuyệt đối giới hạn của số quy tròn $\tilde{a}$.

Lời giải

Trước hết, chúng ta hãy tìm một sai số tuyệt đối của a : có thể lấy

$$\Delta_a = \frac{|a|\delta_a}{1 - \delta_a} = \frac{3,1388 \times 0,035}{1 - 0,035} \leq 0,1139.$$

Quy tròn số a , thu được $\tilde{a} = 3,14$, với sai số tuyệt đối quy tròn là

$$|a - \tilde{a}| = |3,1388 - 3,14| = 0,0012.$$

Sai số tuyệt đối giới hạn của số quy tròn mà chúng ta có bằng tổng của sai số tuyệt đối của số chưa quy tròn và sai số tuyệt đối quy tròn:

$$\Delta_{\tilde{a}} = \Delta_a + |a - \tilde{a}| = 0,1139 + 0,0012 = 0,1151.$$

Ví dụ 1.7. Biết rằng số Archimedes π có một xấp xỉ

$$\pi \approx 3,1415926535.$$

Một số quy tròn π cần ít nhất bao nhiêu chữ số sau dấu thập phân để sai số tuyệt đối không vượt quá $0,3 \times 10^{-3}$?

Lời giải

Nếu xem 3,1415926535 là một số gần đúng của π thì sai số tuyệt đối của nó bé hơn 10^{-10} , không đáng kể so với yêu cầu đạt được sai số bé hơn $0,0003 = 0,3 \times 10^{-3}$. Vậy, chúng ta hãy đi tìm một số quy tròn với sai số quy tròn bé hơn 3×10^{-4} .

Rõ ràng số quy tròn phải có ít nhất 3 chữ số bên phải dấu thập phân bằng với các chữ số tương ứng của số ban đầu. Vì vậy, hãy xét số quy tròn 3,142. Rõ ràng số quy tròn này có sai số quy tròn lớn hơn 0,0004, không thỏa mãn yêu cầu sai số tuyệt đối không vượt quá $0,3 \cdot 10^{-3}$. Từ đó suy ra cần phải quy tròn với ít nhất 4 chữ số sau dấu thập phân.

Nhận thấy rằng số quy tròn 3,1416 có sai số quy tròn không vượt quá $0,73465 \cdot 10^{-5}$ thỏa mãn điều kiện sai số tuyệt đối giới hạn bé hơn $0,3 \cdot 10^{-3}$. Vậy, một biểu diễn thập phân của số gần đúng của π cần 4 chữ số sau dấu thập phân để sai số tuyệt đối giới hạn bé hơn 0,03%.

1.1.6 Quy tròn số trong bất đẳng thức

Giả sử một đại lượng A được đánh giá như sau:

$$A \leq a,$$

với a có biểu diễn thập phân có nhiều chữ số sau dấu thập phân. Chúng ta muốn quy tròn số a để thu được số $\tilde{a}$ với ít chữ số thập phân hơn sao cho bất đẳng thức trên vẫn đúng. Nếu không có thông tin gì thêm, bắt buộc chúng ta phải chọn $\tilde{a} > a$. Từ đây chúng ta có quy tắc *quy tròn lên*: Số quy tròn lên lớn hơn số ban đầu và sai số quy tròn tuyệt đối bé hơn đơn vị hàng cuối cùng.

Ví dụ 1.8. Quy tròn số gần đúng trong vế phải của bất đẳng thức sau:

$$A \leq 0,024674$$

để được số với 2 chữ số thập phân. Khi đó, chúng ta cần quy tròn lên số ở vế phải để thu được

$$A < 0,03.$$

Với bất đẳng thức chiều ngược lại, $A \geq a$, khi quy tròn a , ta cần theo quy tắc *quy tròn xuống*. Chi tiết dành cho bạn đọc.

1.1.7 Mối liên hệ giữa quy tròn số và chữ số có nghĩa

Dựa vào quy tròn số, ta có thể đưa ra khái niệm *số xấp xỉ tới k chữ số có nghĩa*. Khái niệm này thường xuất hiện trong nhiều sách khác nhau. Dưới đây, ta nêu định nghĩa cụ thể của nó, như sau:

Định nghĩa 1.2:

Giả sử số gần đúng a của số $a^* \neq 0$ có hữu hạn chữ số và $k \geq 1$. Chúng ta nói a *xấp xỉ a^* tới k chữ số có nghĩa* nếu biểu diễn thập phân của a và a^* quy tròn về cùng một số thập phân có k chữ số có nghĩa với k là số nguyên dương *lớn nhất* mà có tính chất như vậy.

Để minh họa, chúng ta tìm hiểu hai ví dụ sau đây.

Ví dụ 1.9. Xét $a = 3,14121$ và $a^* = \pi \approx 3,1415926$. Quy tròn π tới 4 chữ số có nghĩa, chúng ta thu được 3,142. Quy tròn a tới 4 chữ số có nghĩa, thu được 3,142. Quy tròn a và a^* tới 5 hoặc 6 chữ số có nghĩa, ta đều thu được hai số thập phân khác nhau. Vậy, a xấp xỉ π tới 4 chữ số có nghĩa.

Ví dụ 1.10. Xét $a = 12,2502$ và $a^* = 12,2496$. Nếu quy tròn a và a^* tới 3 chữ số có nghĩa, chúng ta lần lượt thu được các số thập phân 12,3 và 12,2. Tuy

nhiên, quy tròn a và a^* tới 5 chữ số có nghĩa, chúng ta cùng thu được 12,250. Như vậy, a xấp xỉ a^* tới 5 chữ số có nghĩa.

Nhận xét 1.2. Nếu a xấp xỉ a^* tới k chữ số có nghĩa thì có đánh giá sau về sai số. Chẳng hạn, giả sử $a^* > 0$ cho đơn giản. Khi đó, có duy nhất số nguyên r sao cho

$$a^* = a_0^* \times 10^r,$$

với $1 \leq a_0^* < 10$. Nếu a xấp xỉ a^* tới k chữ số có nghĩa thì có đánh giá sai số tuyệt đối

$$|\epsilon_a| \leq 10^{r-k+1} \quad (1.8)$$

Về phải của đánh giá này phụ thuộc vào số nguyên r . Mặt khác,

$$\delta_a = \frac{|\epsilon_a|}{|a^*|} \leq \frac{10^{r-k+1}}{|a^*|} = \frac{10^{-k+1}}{|a_0^*|} \leq 10^{-k+1}, \quad \text{do } a_0^* \geq 1.$$

Hay

$$\delta_a \leq 10^{-k+1}. \quad (1.9)$$

Đó là mối liên hệ giữa sai số tương đối và số chữ số có nghĩa.

Ví dụ 1.11. Trở lại Ví dụ 1.10, chúng ta có $a = 12,2502$ và $a^* = 12,2496$, trong khi đó, $k = 5$. Do đó, từ Nhận xét 1.2 chúng ta kết luận sai số tương đối của a phải nhỏ hơn 10^{-4} . Điều này cũng có thể thấy bằng tính toán trực tiếp,

$$\delta_a = \frac{|a - a^*|}{a^*} = 4,898119 \times 10^{-5} < 10^{-5+1},$$

như mong đợi.

1.2 Sai số phép toán

Giả sử $x_1, x_2, \dots, x_n$ là một số đại lượng mà chúng ta chỉ biết các giá trị gần đúng của chúng cùng với các sai số tương ứng và u là một đại lượng phụ thuộc vào $x_1, \dots, x_n$ theo một công thức đã biết. Chúng ta muốn tính gần đúng đại lượng u và ước lượng sai số của u theo các sai số của x_i . Nói cách khác, chúng ta phân tích *sự lan truyền của sai số* (error propagation) trong quá trình tính toán.

Để đơn giản, hãy bắt đầu với trường hợp $n = 2$ và giả sử u là một hàm số của hai biến x, y :

$$u = f(x, y),$$

trong một số trường hợp đơn giản của hàm $f(x, y)$, bao gồm tổng $x + y$ và tích xy .

1.2.1 Sai số của một tổng

Giả sử x và y là hai số xấp xỉ của đại lượng X, Y có giá trị thực là x^* và y^* , tương ứng. Giả sử đại lượng U là tổng của X và Y : $U = f(X, Y) = X + Y$. Rõ ràng giá trị thực của U là $u^* = x^* + y^*$, và một giá trị gần đúng của U là $u = x + y$. Mặt khác, ta đặt $\epsilon_x = x^* - x$ và $\epsilon_y = y^* - y$ là các sai số thực sự của x và y . Khi đó

$$\begin{aligned}\epsilon_u &:= u^* - u = (x^* + y^*) - (x + y) \\ &= (x^* - x) + (y^* - y) \\ &= \epsilon_x + \epsilon_y.\end{aligned}$$

Do đó, theo bất đẳng thức tam giác,

$$|\epsilon_u| = |\epsilon_x + \epsilon_y| \leq |\epsilon_x| + |\epsilon_y|.$$

Vậy ta có thể phát biểu thành quy tắc sau: *Sai số tuyệt đối giới hạn của một tổng hoặc hiệu bằng tổng của sai số tuyệt đối giới hạn của các số hạng*. Quy tắc này cũng đúng cho tổng của nhiều số hạng.

Nhận xét 1.3. Vì sai số tương đối của u là

$$\delta_u = \frac{|\epsilon_u|}{|u^*|} \leq \frac{|\epsilon_x| + |\epsilon_y|}{|x^* + y^*|} = \frac{|x^*|\delta_x + |y^*|\delta_y}{|x^* + y^*|}$$

nên, khi x^* và y^* cùng dấu, thì $|x^* + y^*| = |x^*| + |y^*|$ và ta có thể đánh giá

$$\delta_u \leq \frac{|x^*|\delta_x + |y^*|\delta_y}{|x^* + y^*|} < \delta_x + \delta_y. \quad (1.10)$$

Tuy nhiên, nếu x^* và y^* trái dấu, thì $|x^* + y^*| = ||x^*| - |y^*||$ và đánh giá δ_u như (1.10) không còn đúng nữa.

Để minh họa một cách chi tiết hơn, chúng ta hãy xét ví dụ sau: Giả sử chúng ta cần tính $\sqrt{2022} - \sqrt{2020}$ và viết kết quả dưới dạng số thập phân. Bằng cách tính gần đúng các căn thức bậc hai với kết quả làm tròn với 4 chữ số có nghĩa,

$$\sqrt{2022} \approx 44,97, \quad \sqrt{2020} = 44,94.$$

Sử dụng các kết quả này, chúng ta thu được xấp xỉ $\sqrt{2022} - \sqrt{2020} \approx 0,03$ với chỉ một chữ số có nghĩa.

Để có kết quả chính xác hơn, bằng cách khử căn thức ở tử số, ta có thể viết

$$\sqrt{2022} - \sqrt{2020} = \frac{2}{\sqrt{2022} + \sqrt{2020}} \approx \frac{2}{89,91} \approx 0,02224.$$

Có thể kiểm tra được kết quả cuối cùng 0,02224 có 4 chữ số có nghĩa là chữ số đáng tin. Ví dụ này minh họa một quy tắc là, trong thực hành, nếu có thể chúng ta cần tránh việc trừ các số gần nhau.

1.2.2 Sai số của một tích

Giả sử đại lượng U là tích của hai đại lượng X và Y , tức là $U = XY$. Gọi x^* và y^* tương ứng là các giá trị thực của X và Y . Cũng giả sử x và y tương ứng là hai giá trị xấp xỉ với sai số tương ứng là ϵ_x và ϵ_y . Khi đó, có thể coi $u = xy$ là một giá trị gần đúng của giá trị thực $u^* = x^*y^*$ của đại lượng U .

Để đánh giá sai số của u , ta tính toán trực tiếp

$$\begin{aligned}\epsilon_u &= x^*y^* - xy \\ &= x^*y^* - (x^* - \epsilon_x)(y^* - \epsilon_y) \\ &= x^*\epsilon_y + y^*\epsilon_x - \epsilon_x\epsilon_y.\end{aligned}$$

Theo bất đẳng thức tam giác,

$$\begin{aligned}\frac{|\epsilon_u|}{|x^*y^*|} &= \left| \frac{x^*\epsilon_y + y^*\epsilon_x - \epsilon_x\epsilon_y}{|x^*y^*|} \right| \\ &\leq \left| \frac{\epsilon_y}{y^*} \right| + \left| \frac{\epsilon_x}{x^*} \right| + \left| \frac{\epsilon_x}{x^*} \right| \cdot \left| \frac{\epsilon_y}{y^*} \right|.\end{aligned}$$

Như vậy

$$\delta_u \leq \delta_x + \delta_y + \delta_x\delta_y. \quad (1.11)$$

Nếu các xấp xỉ x và y đủ tốt, thì δ_x và δ_y là rất nhỏ ($\delta_x \ll 1$, $\delta_y \ll 1$) nên $\delta_x\delta_y$ rất nhỏ so với δ_x và δ_y . Khi đó, từ (1.11) ta có thể coi $\delta_x + \delta_y$ là một sai số tương đối của xy . Như vậy, ta có quy tắc: *Sai số tương đối giới hạn của một tích xấp xỉ bằng tổng các sai số tương đối giới hạn của các nhân tử.*

1.2.3 Sai số của một thương

Giả sử $U = Y^{-1}$, y và y^* lần lượt là một giá trị gần đúng và giá trị thật sự của đại lượng Y . Khi đó, $u = y^{-1}$ là một xấp xỉ của $u^* = (y^*)^{-1}$. Chúng ta chỉ xét trường hợp $\delta_y < 1$. Khi đó, để đánh giá độ chính xác của u , ta xét

$$\epsilon_u = u^* - u = \frac{1}{y^*} - \frac{1}{y} = \frac{y - y^*}{y y^*} = \frac{-\epsilon_y}{y^*(y^* - \epsilon_y)}.$$

Do đó,

$$\frac{|\epsilon_u|}{|u^*|} = \frac{|\epsilon_y|}{|y^* - \epsilon_y|} \leq \frac{|\epsilon_y|}{|y^*| - |\epsilon_y|} = \frac{\delta_y}{1 - \delta_y}.$$

Nếu xấp xỉ y là đủ tốt, tức là $\delta_y \ll 1$, thì ta có thể lấy $\delta_u \approx \delta_y$.

Nếu $V = X/Y$ là một thương với $v^* = x^*/y^*$ là giá trị đúng và $v = x/y$ là giá trị xấp xỉ, thì, bằng tính toán tương tự như trên, ta có đánh giá

$$\delta_v = \frac{|\epsilon_v|}{|v^*|} \leq \frac{\delta_x + \delta_y}{1 - \delta_y}. \quad (1.12)$$

Với δ_y rất nhỏ, $1 - \delta_y \approx 1$ và ta có thể lấy $\delta_v \approx \delta_x + \delta_y$. Từ đó có quy tắc: Sai số tương đối giới hạn của một thương xấp xỉ bằng tổng sai số của tử thức và mẫu thức.

1.2.4 Công thức tổng quát cho đánh giá sai số tích lũy

Giả sử U là một đại lượng phụ thuộc các đại lượng $X_1, X_2, \dots, X_n$, viết là

$$U = f(X_1, X_2, \dots, X_n),$$

với f là một mà số của n biến số. Để đánh giá sai số của đại lượng U qua các sai số của X_j , ta cần giả thiết f khả vi liên tục.

Nếu x_j^* là giá trị thực sự của X_j và x_j là giá trị gần đúng, thì theo định lý số gia giới nội, có một điểm $(c_1, c_2, \dots, c_n)$ nằm trên đoạn thẳng nối $(x_1, x_2, \dots, x_n)$ và $(x_1^*, \dots, x_n^*)$ sao cho

$$\epsilon_u = f(x_1^*, \dots, x_n^*) - f(x_1, \dots, x_n) = \sum_{j=1}^n \frac{\partial f}{\partial X_j}(c_1, c_2, \dots, c_n) \epsilon_{x_j}.$$

Từ đây, ta suy ra công thức sai số tương đối

$$\begin{aligned} \delta_u = \frac{|\epsilon_u|}{|u^*|} &\leq \sum_{j=1}^n \left| \frac{\partial f}{\partial X_j}(c_1, c_2, \dots, c_n) \right| \left| f(x_1^*, \dots, x_n^*) \right|^{-1} |\epsilon_{x_j}| \\ &\approx \sum_{j=1}^n \left| \frac{\partial \ln f}{\partial X_j}(x_1^*, \dots, x_n^*) \right| |\epsilon_{x_j}|. \end{aligned}$$

Ví dụ 1.12. Nếu $u = x^\alpha$ với $x > 0$ và hằng số $\alpha > 0$ thì

$$\delta_u \approx \frac{d \ln(x^\alpha)}{dx} |\epsilon_x| = \alpha \left| \frac{d \ln x}{dx} \right| |\epsilon_x| = \alpha \frac{|\epsilon_x|}{|x^*|} = \alpha \delta_x.$$

Nhắc lại rằng, đánh giá này chỉ tốt khi δ_x rất nhỏ: $\delta_x \ll 1$.

Nhận xét 1.4. Trong công thức tổng quát, ta không có được đánh giá sai số một cách chính xác, mà chỉ có được đánh giá gần đúng của sai số. Đó là vì đạo hàm của hàm số f tham gia trong biểu thức, nhìn chung, có thể phức tạp mà chúng ta không thể đánh giá được hết sự ảnh hưởng của nó đến kết quả tính toán.

1.3 Phân loại sai số

Trong việc quá trình tính toán gần đúng một đại lượng thực tế, có nhiều giai đoạn dẫn đến sai số của kết quả cuối cùng, như sai số từ các phép đo

trong thực nghiệm, sai số từ phương pháp giải, và sai số khi tính toán (bao gồm sai số của việc quy tròn số). Vấn đề sai số trong phép đo đạc hoặc thu thập dữ liệu không nằm trong khuôn khổ của bài giảng. Dưới đây, chúng ta trình bày sơ lược về sai số của phương pháp giải và sai số khi tính toán.

1.3.1 Sai số phương pháp

Một số đại lượng được biểu thị bằng các công thức phức tạp, không thể tính được chính xác bằng hữu hạn các phép tính. Để minh họa cho vấn đề này, ta hãy xét ví dụ sau (T. V. Đình [6, §1.5]).

$$B = \sum_{n=1}^{\infty} (-1)^n n^{-3} \quad (1.13)$$

là một giá trị được định nghĩa bởi một chuỗi hội tụ: Vế phải của (1.13) có vô hạn số hạng.

Nếu không có công thức đơn giản cho tổng vô hạn B , thì để tính B , người ta có thể dùng phương pháp chặt chuỗi vô hạn: chọn một giá trị nguyên dương $m \geq 1$ và đặt

$$B_m := \sum_{n=1}^m (-1)^{n-1} n^{-3}.$$

Bởi vì $\lim_{m \rightarrow \infty} B_m = B$ nên có thể lấy

$$B_m \approx B,$$

là một giá trị gần đúng của B .

Phương pháp này sinh ra một sai số $\Delta = |B_m - B|$, gọi là *sai số phương pháp*. Tuy nhiên, chúng ta có thể ước lượng được sai số này như sau:

$$\Delta = |B_m - B| = \left| \sum_{n=m+1}^{\infty} (-1)^{n-1} n^{-3} \right| < (m+1)^{-3}.$$

Như vậy, phương pháp chặt đối với chuỗi vô hạn trên sinh ra một sai số được đánh giá bởi $(m+1)^{-3}$. Sai số này nhỏ (tùy ý) khi ta chọn m lớn (tùy ý). Sai số đó gọi là sai số phương pháp.

1.3.2 Sai số tính toán

Tiếp tục nói về ví dụ trong mục trước, việc tính B_m để xấp xỉ B cũng gặp phải khó khăn nếu m lớn. Trong quá trình tính B_m chúng ta cũng có chỉ thu được giá trị gần đúng, và sai số trong phép tính này gọi là *sai số tính toán*.

Ví dụ, với $m = 6$, ta tính được

$$B_6 = 1 - \frac{1}{2^3} + \frac{1}{3^3} - \frac{1}{4^3} + \frac{1}{5^3} - \frac{1}{6^3} = \frac{194353}{216} \times 10^{-3}.$$

Quy tròn số, chúng ta có

$$B_6 = \frac{194353}{216} \times 10^{-3} \approx 0,90,$$

với sai số $\Delta_{0,90} = 0,5 \times 10^{-2}$. Sai số quy tròn này là một loại *sai số tính toán*.

Mặt khác, sai số phương pháp ở mục trước cho ta

$$|B - B_6| < \frac{1}{7^3} < 0,3 \times 10^{-2}.$$

Vậy,

$$|B - 0,90| < 0,4 \times 10^{-2}.$$

Với hai chữ số sau dấu thập phân là chữ số chắc, ta viết

$$B \approx 0,90,$$

với sai số $0,4 \times 10^{-2}$ là tổng hợp của sai số phương pháp và sai số tính toán.

1.4 Giới thiệu ngôn ngữ lập trình R

Phần mềm thống kê R là một công cụ mạnh mẽ được sử dụng rộng rãi trong cộng đồng nghiên cứu và phân tích dữ liệu. Tuy nhiên, R không chỉ là một phần mềm tính toán thống kê, mà còn là một môi trường tính toán mở và miễn phí, có khả năng ứng dụng trong tính toán số. Trong tài liệu này, chúng ta sẽ tìm hiểu cách sử dụng R cho mục đích trực quan hoá (như vẽ đồ thị) và thực hiện tính toán tự động theo các thuật toán giải số. Mục tiêu chính ở đây là minh hoạ cách thiết lập các chương trình máy tính đơn giản từ lý thuyết và giảm thời gian tính toán trong các ví dụ cụ thể.

Tài liệu tham khảo về sử dụng R: Bloomfield [3], Jones–Maillardet–Robinson [7].

Tải và cài đặt R, bạn đọc có thể tham khảo <https://www.r-project.org/>

Các tính toán cơ bản trong R

Phần mềm thống kê R có thể sử dụng như một máy tính cầm tay. Nó có thể thực hiện các phép tính số học (cộng, trừ, nhân, và chia), phép lũy thừa, hàm logarith, hàm lượng giác và hyperbolic, và một số hàm Bessel hay hàm Lambert (gói pracma). Cách dùng R cho các tính toán cơ bản cũng rất đơn giản. Chúng ta có thể gán vào các biến những giá trị bằng dấu <-, hoặc dấu =, và sau đó thực hiện các phép toán trên các biến đó.

CAS 1.1. Trong ví dụ sau đây, chúng ta tính xấp xỉ biểu thức

$$c = a + b = \sqrt{2} + \sqrt{100 + \sqrt{2}}$$

trong R. Kết quả như sau:

```
# Dùng dấu phẩy thập phân
options(OutDec = ",")

a <- sqrt(2)
b <- sqrt(100 + a)
c <- a + b

# In kết quả
print(c)

## [1] 11,48468
```

Đối với các phép toán có nhiều số hạng mà chúng có quy tắc tính toán chung, chúng ta có thể lập chương trình để tính toán. Hãy tính gần đúng một xấp xỉ của số e làm ví dụ. Bằng cách viết e^x dưới dạng tổng của một chuỗi số hội tụ,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots,$$

chúng ta thu được

$$e \approx e_{\text{approx}} = 1 + \frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{7!}. \quad (1.14)$$

CAS 1.2. Đoạn chương trình R đơn giản sau đây minh họa vòng lặp “while” để tính tổng ở vế phải của (1.14).

```
# Định nghĩa một hàm tính tổng 1 + 1/(2!) + ... + 1/(n!)
e_approx <- function(n) {
  partial_sum = 1
  term = 1
  count = 0

  # Bắt đầu vòng lặp while
  while (count < n) {
    count = count + 1
```

```

        term = term / count
        partial_sum = partial_sum + term
    }
    # Kết thúc vòng lặp và trả về kết quả partial_sum
    return(partial_sum)
}

# Khai báo n = 7 và chạy thủ tục e_approx() vừa định nghĩa
n <- 7
result <- e_approx(n)

# In ra kết quả
print(result)

## [1] 2,718254

```

Kết quả thu được có thể xem là một xấp xỉ của cơ số logarith tự nhiên: $e \approx 2,718254$.

Tính toán với các hàm số đặc biệt

Phần mềm thống kê và ngôn ngữ R cũng có thể tính toán với một số hàm số quan trọng, như hàm Bessel hay hàm Lambert, v.v.

Hãy lấy hàm Lambert làm ví dụ. Đối với biến thực $x \in \mathbb{R}$, hàm Lambert có hai nhánh chính là W_0 và W_{-1} . Với $x \geq 0$, $y = W_0(x)$ nếu $ye^y = x$, trong khi đó, với $e^{-1} \leq x < 0$ thì $y = W_{-1}(x)$ nếu $ye^y = x$ và $y \leq -1$. Nếu $y < -1$, $ye^y = x$ thì $y = W_{-1}(x)$.

Hàm Lambert xuất hiện trong khi giải một số phương trình đơn giản liên quan đến hàm mũ. Ví dụ, xét phương trình $x5^{x-1} = 9$. Để giải gần đúng, chúng ta đặt biến mới $u = x \ln(5)$ và viết

$$x5^{x-1} = \frac{ue^u}{5\ln 5}.$$

Từ đó, đưa phương trình về dạng

$$ue^u = 45\ln(5).$$

Vậy, $u = W(45\ln(5))$ và do đó

$$x = \frac{W(45\ln(5))}{\ln 5}.$$

Đó là biểu diễn nghiệm đúng của phương trình, thông qua hàm Lambert W .

CAS 1.3. Để tính gần đúng nghiệm của phương trình $x5^{x-1} = 9$ dưới dạng số thập phân, chúng ta có thể dùng gói `pracma`¹ của R như sau:

```
# Nạp gói pracma
library(pracma)

# Gọi hàm lambertWp()
lambertWp(45 * log(5)) / (log(5))

## [1] 1,950202
```

Tính toán với số phức

Về những khái niệm và phép toán cơ bản trên số phức, bạn đọc có thể tìm đọc N. V. Khuê–L. M. Hải–N. Đ. Sang [8, Chương 1, tập 1].

Các số phức được viết dưới dạng

$$z = a + bi.$$

Ví dụ, đoạn mã lệnh sau đưa vào hai số phức và thực hiện các phép toán cơ bản trên chúng.

```
# Khai báo các số phức z và w
z <- 3 + 4i
w <- 5 - 8i

# Tính z + w^2 và gán giá trị vào biến q
q <- z + w^2

# In ra giá trị của q
print(q)

## [1] -36-76i
```

Chúng ta thu được $q = z + w^2 = -36 - 76i$, với $z = 3 + 4i$ và $w = 5 - 8i$.

Để tính môđun (modulus) và góc cực (argument) của một số phức, có thể dùng `Mod()` và `Arg()`. Ví dụ, với $z = \sqrt{3} + i$,

```
# Khai báo z
z <- sqrt(3) + 1i

# In ra giá trị của |z|
print(Mod(z))
```

¹`pracma`: Practical Numerical Math Functions. DOI: <https://doi.org/10.32614/CRAN.package.pracma>

```
## [1] 2
# In ra giá trị Arg(z)
print(Arg(z))
## [1] 0,5235988
```

Kết quả thu được $|z| = 2$ và argument chính $\text{Arg}(z) = \pi/6 \approx 0,523599$.

Trong R, căn bậc n của một số phức được hiểu là nhánh chính (principal branch) của hàm $z^{1/n}$. Ví dụ,

```
# Khai báo số phức z = -46 + 9 i
z <- - 46 + 9i

# Tính căn bậc 3 của z và gán vào biến w
w <- z^(1/3)

# In ra giá trị của w
print(w)
## [1] 2+3i
```

Kết quả $w = 2 + 3i$ thoả mãn $w^3 = z = -46 + 9i$.

Nếu z là một số thực, chẳng hạn $z = -2$, để tính căn phức, ta viết

```
z <- -2 + 0i
w <- z^(1/3)
print(w)
## [1] 0,6299605+1,091124i
```

Kết quả thu được:

$$\begin{aligned} (-2)^{1/3} &= \exp\left(\frac{1}{3}\text{Log}(z)\right) \\ &= \exp\left(\frac{1}{3}(\ln|z| + i\text{Arg}(z))\right) \\ &= \sqrt[3]{2}(\cos(\pi/3) + i\sin(\pi/3)). \end{aligned}$$

Một cách xấp xỉ thì

$$(-2)^{1/3} \approx 0,629961 + 1,091124i.$$

Ở trên, $\text{Log}(z)$ là nhánh chính của hàm logarithm phức.

Phần mềm R cũng làm việc được với dạng mũ của số phức (chẳng hạn, công thức Euler).

Ví dụ, đoạn mã lệnh

```
exp(log(2) + (pi/3) * 1i)
## [1] 1+1,732051i
```

cho chúng ta

$$\exp(\ln(2) + i\pi/3) = 2(\cos(\pi/3) + i\sin(\pi/3)) = 1 + \sqrt{3}i.$$

Quy tròn số trong R

Trong R, chúng ta có thể đặt tùy chọn về số chữ số thập phân khi quy tròn số bằng `options(digits = n)`, trong đó n là số các chữ số còn lại sau làm tròn. Ví dụ:

```
# In ra giá trị căn bậc hai của 2
print(sqrt(2))

## [1] 1,414214

# Đặt tùy chọn hiển thị kết quả có 4 chữ số
options(digits = 4)

# In ra giá trị căn bậc hai của 2
print(sqrt(2))

## [1] 1,414
```

Bạn đọc hãy quan sát các kết quả đầu ra và đưa ra nhận xét cho riêng mình.

Để làm tròn một số gần đúng trong R tới một số chữ số thập phân nhất định, chúng ta có thể dùng `round(number, digits)`, trong đó, `digits` chỉ số các chữ số sau dấu thập phân.

Dưới đây là một số ví dụ về làm tròn số.

```
# Làm tròn 2,9451 tới 2 chữ số sau dấu thập phân
round(2.9451, 2)

## [1] 2,95

# Làm tròn sau khi làm tròn
round(round(2.9451, 3), 2)

## [1] 2,94
```

Bạn đọc hãy quan sát kết quả thu được và cho nhận xét.

Nếu muốn quy tròn số tới một số chữ số có nghĩa nhất định, chúng ta dùng `signif()`².

```
options(digits = 6)
signif(2.45301, 2)

## [1] 2,5

signif(2.45301, 6)

## [1] 2,45301

signif(2.45301, 5)

## [1] 2,453
```

1.5 Câu hỏi và bài tập

1. Khi đo khối lượng của một lít nước ở 0°C , người ta thu được $p = 999,847 \pm 0,001$ (gram). Hãy tính sai số tuyệt đối, sai số tương đối, và nêu ý nghĩa rút ra từ phép đo này.
2. Hãy xác định chữ số chắc của một số gần đúng a trong các trường hợp sau đây:
 - (a) $a = 2,56732$, $\Delta_a = 0,0005$.
 - (b) $a = 0,245123$, $\delta_a = 1,3 \times 10^{-3}$.
3. Hãy quy tròn các số đúng sau đây với 3 chữ số chắc. Tính các sai số tuyệt đối và sai số tương đối tương ứng.
 - (a) 0,123456.
 - (b) 2,022.
 - (c) 0,01501.
 - (d) $-0,00142857$.
4. Giả sử các số thập phân trong các phép toán dưới đây là các số quy tròn có tất cả các chữ số có nghĩa là chữ số chắc. Hãy tính kết quả của các phép toán cùng với các sai số tuyệt đối tương ứng:
 - (a) $3,2062 + 5,948$.
 - (b) $4,46 - 1,752$.
 - (c) $2,747 \times 8,83$.
 - (d) $10,473 \div 0,064$.

²Hàm `signif()` được lấy tên từ các chữ cái đầu của từ Tiếng Anh “significant digits”, có nghĩa là “các chữ số có nghĩa”.

5. Cho các số gần đúng $x = 0,85$, $y = 1,34$, và $z = 1,132$, với mọi chữ số có nghĩa là chữ số chắc. Hãy xác định giá trị của hàm số u cùng với sai số tuyệt đối giới hạn Δ_u và sai số tương đối giới hạn δ_u trong các trường hợp sau:

(a) $u(x) = \sqrt{1+x}$.

(c) $u(x, y) = \ln(x + y^2)$.

(b) $u(x, y) = x^3/y$.

(d) $u = (x + y^2)/z$.

6. Biết rằng hằng số Euler e có dạng

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{R_n}{(n+1)!},$$

với $1 < R_n < 3$. Tìm một biểu diễn thập phân của e với 5 chữ số chắc.

7. Độ pH của một dung dịch được tính bởi $\text{pH} = -\log_{10}[\text{H}^+]$ với $[\text{H}^+]$ là hoạt độ của ion H^+ (tính theo mol/l). Giả sử ta biết $\text{pH} = 4,25 \pm 0,02$. Tính gần đúng $[\text{H}^+]$ và sai số của nó.
8. Giả sử $E = \frac{1}{2}mv^2$. Hãy chỉ ra một cách đo các đại lượng m và v với độ chính xác như thế nào để sai số tuyệt đối giới hạn Δ_E không vượt quá $0,5 \times 10^{-3}$.

Chương

2

Tính gần đúng nghiệm của một phương trình phi tuyến

Trong chương này, chúng ta sẽ tìm hiểu các phương pháp tính nghiệm gần đúng mà có thể áp dụng để giải các phương trình như (2.1) cũng như rất nhiều phương trình khác nảy sinh trong khoa học và kĩ thuật, bao gồm phương pháp chia đôi, phương pháp lặp đơn, phương pháp Newton–Raphson cùng một số phương pháp khác.

Cụ thể hơn, nội dung chính của chương này bao gồm

- Khái niệm cơ bản về phương trình, nghiệm đúng, và nghiệm xấp xỉ.
- Khoảng tách nghiệm, một số phương pháp tìm khoảng tách nghiệm
- Các phương pháp tìm nghiệm xấp xỉ: Phương pháp lặp đơn, Newton–Raphson, chia đôi, dây cung, vị trí sai (regula falsi), và Steffensen.
- Một số đánh giá sai số của nghiệm xấp xỉ.

Tài liệu tham khảo chính ở đây là T. V. Đĩnh [6, Chương 2].

Nội dung

2.1	Nghiệm và khoảng tách nghiệm	26
2.1.1	Nghiệm của một phương trình	26
2.1.2	Nghiệm xấp xỉ và khoảng tách nghiệm	30
2.2	Phương pháp chia đôi	33
2.3	Phương pháp lặp đơn	36
2.3.1	Sự hội tụ	36
2.3.2	Đánh giá sai số	42
2.3.3	Tóm tắt phương pháp lặp đơn	43
2.4	Phương pháp Newton–Raphson	44

2.4.1	Mô tả phương pháp Newton–Raphson	44
2.4.2	Sự hội tụ	45
2.4.3	Đánh giá sai số	50
2.4.4	Tóm tắt thuật toán Newton–Raphson	53
2.5	Phương pháp dây cung và một số phương pháp khác	55
2.5.1	Phương pháp dây cung	55
2.5.2	Phương pháp Steffensen	57
2.5.3	Phương pháp vị trí sai (regula falsi)	58
2.6	Câu hỏi và bài tập	60

Một người thợ kim hoàn dùng một đồng tiền vàng hình tròn, có đường kính 2cm, mỏng và có độ dày đồng đều, để chế tác 4 chiếc nhẫn vàng. Người thợ chia đồng tiền thành 4 phần có khối lượng bằng nhau bằng cách cắt dọc theo 3 đường thẳng song song, trong đó, một đường cắt đi qua tâm hình tròn và hai đường cắt còn lại đối xứng với nhau qua đường cắt thứ nhất.

Dựa trên các kiến thức hình học, có thể thấy rằng để 4 miếng vàng thu được có khối lượng như nhau thì khoảng cách d từ tâm hình tròn đến mỗi đường cắt gần hơn phải là

$$d = \frac{a}{\sqrt{1+a} + \sqrt{1-a}},$$

trong đó a là một số dương thỏa mãn phương trình

$$x = \cos(x). \quad (2.1)$$

Vấn đề ban đầu tưởng chừng như đơn giản, nhưng mô hình toán học cho nó lại dẫn đến một phương trình phi tuyến mà các phương pháp đã học ở bậc phổ thông là không thể cho nghiệm đúng.

Mục tiêu chính của chúng ta trong chương này là các phương pháp số để tính gần đúng nghiệm của các phương trình phi tuyến như phương trình trên.

2.1 Nghiệm và khoảng tách nghiệm

Mục này, chúng ta tìm hiểu một số khái niệm cơ bản về phương trình và nghiệm của phương trình, các khoảng tách nghiệm, điều kiện có nghiệm. Tài liệu tham khảo: T. V. Đĩnh [6, §2.1, §2.2].

2.1.1 Nghiệm của một phương trình

Các phương trình bậc nhất $ax + b = 0$ xuất hiện trong những tình huống đơn giản nhất của cuộc sống, trong khi đó, các phương trình bậc hai $ax^2 + bx + c = 0$ cũng thường xuất hiện trong khoa học và kỹ thuật như trong vật lý, hoá học, hay kinh tế, và các khoa học xã hội khác. Như việc tìm thời gian chạm đất của một vật rơi tự do theo vận tốc và độ cao ban đầu cũng đưa về một phương trình bậc hai. Phương trình bậc ba có thể xuất hiện trong các tình huống của vật lý và hoá học như các phương trình

trạng thái Van der Waal về chất khí. Các phương trình lượng giác thì lại có thể xuất hiện trong các bài toán liên quan đến hình học, hoặc trong giải tích Fourier.

Nhìn chung, các phương trình xuất hiện trong khoa học kỹ thuật thường là phi tuyến và phức tạp hơn nhiều các phương trình xuất hiện trong toán học ở phổ thông trung học.

Vậy, một cách chặt chẽ toán học thì phương trình là gì? Một phương trình của một ẩn x có dạng

$$f(x) = 0, \quad (2.2)$$

trong đó, f là hàm của đối số thực x . Một nghiệm của phương trình (2.2) là một số thực α sao cho khi thay α vào vị trí của x , thì (2.2) nghiệm ra thành đúng lẽ, tức là

$$f(\alpha) = 0. \quad (2.3)$$

Ví dụ 2.1. Số thực $\alpha = 2,5$ là một nghiệm đúng của phương trình

$$f(x) := 2x^3 - 3x^2 - 3x - 5 = 0,$$

vì chúng ta dễ dàng tính được bằng giấy và bút $f(2,5) = 0$.

CAS 2.1. Trong CAS này, chúng ta minh họa việc tính toán giá trị của hàm số $f(x)$ tại một điểm $x = a$ cho trước, sử dụng cú pháp rẽ nhánh “if, else” để kiểm tra xem giá trị đó có “đủ gần” 0 hay không. Ở đây, “đủ gần” được đo bằng một giá trị sai số “tol” (tolerance) mà chúng ta xem là chấp nhận được. Từ đó đưa ra kết luận $x = a$ có phải là nghiệm hay không.

```
# Định nghĩa hàm f(x) theo công thức đã có
f <- function(x) {
  2 * x^3 - 3 * x^2 - 3 * x - 5
}

# Hàm isroot() kiểm tra xem một giá trị có phải
# là nghiệm của hàm f(x) khai báo bên trên hay không
isroot <- function(a) {

  # Sai số chấp nhận được
  tol <- 1e-9

  # Kiểm tra xem |f(a)| có nhỏ hơn sai số tol = 10-9
  if (abs(f(a)) < tol) {

    # Nếu |f(a)| < tol, thông báo a là nghiệm.
    cat(a, "là một nghiệm.\n")
  } else {
```

```

        # Nếu  $|f(a)| \geq \text{tol}$ , thông báo  $a$  không là nghiệm.
        cat(a, "không phải là nghiệm.\n")
        cat("f(", a, ") =", f(a), "\n")
    }
}

# Kiểm tra xem giá trị  $a = 2,500$  có phải là nghiệm
a <- 2.500
isroot(a)

## 2,5 là một nghiệm.

# Kiểm tra xem giá trị  $a = 2,505$  có phải là nghiệm
b <- 2.505
isroot(b)

## 2,505 không phải là nghiệm.
## f( 2,505 ) = 0,0978002

```

Lưu ý, trong CAS trên, thay vì kiểm tra $f(a) = 0$, chúng ta kiểm tra $|f(a)| < \text{tol}$ để tránh kết luận sai do sai số làm tròn. Trong đoạn mã R sau đây, chúng ta tính $\sin(\pi)$ và thu được một kết quả là một số dương rất nhỏ, do sai số tính toán. Do đó, nếu kiểm tra điều kiện $\sin(\pi) = 0$, thì chương trình sẽ cho kết luận sai!

```

# Gán  $a = \pi$ 
a <- pi

# Kiểm tra xem  $\sin(a) = 0$ 
if (sin(a) == 0) {
    cat(a, "là một nghiệm")
} else {
    cat(a, "không phải là một nghiệm của  $\sin(x) = 0$ ." )
}

## 3,14159 không phải là một nghiệm của  $\sin(x) = 0$ .

# In ra giá trị  $\sin(a)$ 
print(sin(a))

## [1] 1,22465e-16

```

Chương trình cho kết luận sai và cho giá trị $\sin(\pi) \approx 10^{-16}$!

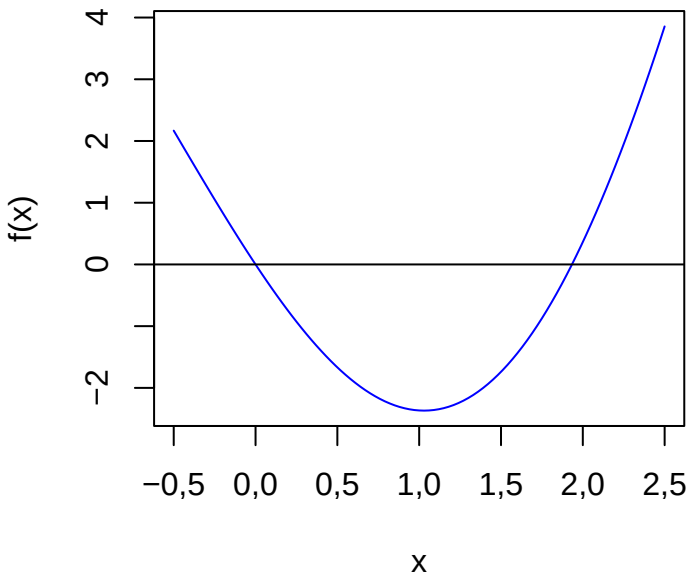
Nếu α là một nghiệm của (2.2) thì điểm $(\alpha, 0)$ trên mặt phẳng Oxy sẽ nằm trên đồ thị của hàm số $y = f(x)$. Nói cách khác, đồ thị này cắt trục hoành tại điểm có hoành độ $x = \alpha$.

CAS 2.2. Chúng ta xét phương trình $f(x) = x^2 - 4\sin(x) = 0$ với một khoảng tách nghiệm là $(1, 2)$. Đoạn mã R sau đây vẽ đồ thị của $y = f(x)$ trong khoảng $(1/2, 5/2)$. Hình vẽ thu được thể hiện rõ một nghiệm của phương trình nằm trong khoảng $(1, 5/2)$ (cùng với nghiệm $x = 0$).

```
# Định nghĩa hàm  $f(x) = x^2 - 4 * \sin(x)$ 
f <- function(x) {
  x^2 - 4 * sin(x)
}

# Vẽ đồ thị của  $y = f(x)$  trên khoảng  $[-0,5; 2,5]$ 
curve(f(x), from = -0.5, to = 2.5, col = "blue")

# Vẽ đường nằm ngang tại  $y = 0$ 
abline(h = 0)
```



Kết quả thu được trên màn hình là đồ thị của hàm số $f(x) = x^2 - 4\sin(x)$ đã cho cùng với đường nằm ngang tại vị trí $y = 0$, thể hiện rõ việc đồ thị của $f(x)$ cắt đường thẳng $y = 0$ tại hai điểm trong đó một điểm nằm giữa 1 và 2.

2.1.2 Nghiệm xấp xỉ và khoảng tách nghiệm

Nếu α là một nghiệm của phương trình (2.2) và x^* là một số xấp xỉ của α thì x^* gọi là nghiệm xấp xỉ (hoặc nghiệm gần đúng) của phương trình đó. Khi nghiệm α không tìm được, hoặc khó tìm, thì việc tìm một nghiệm xấp xỉ có ý nghĩa quan trọng. Hơn nữa, trong thực tế, các tham số của phương trình cần giải cũng chỉ biết gần đúng (vì chúng có được từ các phép đo đạc, hoặc từ các tính toán khác) nên việc giải đúng nghiệm của nó cũng không có ý nghĩa. Việc tìm một nghiệm xấp xỉ của một phương trình gọi là việc *giải gần đúng* phương trình đó.

Thông thường, chúng ta giải gần đúng (2.2) qua hai bước:

- **Bước giải sơ bộ:** Tìm một khoảng chứa nghiệm cần tính và chỉ chứa duy nhất nghiệm đó, gọi là *khoảng tách nghiệm* (Định nghĩa 2.1 sau đây). Trong những hoàn cảnh cụ thể, chúng ta chọn các khoảng tách nghiệm “bé nhất” có thể.
- **Bước giải kiện toàn:** Tìm một giá trị gần đúng của nghiệm với độ chính xác cần thiết và đánh giá sai số của nó.

Trong bước giải sơ bộ, chúng ta có thể sử dụng định lý sau đây mà phát biểu của nó đã quen thuộc trong toán học ở bậc trung học và trong môn toán cao cấp (giải tích một biến). Định lý này cho chúng ta một sự đảm bảo về sự tồn tại nghiệm trong một khoảng trong trường hợp hai giá trị của hàm số ở về trái lầy tại hai đầu mút có dấu ngược nhau.

Định lý 2.1: Bolzano–Cauchy

Giả sử $f(x)$ là một hàm liên tục trên một đoạn $[a, b]$. Nếu

$$f(a)f(b) < 0,$$

thì tồn tại một nghiệm α trong khoảng (a, b) của phương trình (2.2).

Định lý 2.1 được Bolzano¹ và Cauchy² chứng minh độc lập vào đầu thế kỷ 19. Bạn đọc quan tâm có thể tìm đọc chứng minh định lý này và các hệ quả của nó các giáo trình giải tích bậc đại học như N. V. Khuê–L. M. Hải–N. Đ. Sang [8, Chương I, §6].

Ví dụ 2.2. Phương trình $x = \cos(x)$ có thể viết lại dưới dạng tương đương

$$x - \cos(x) = 0.$$

¹ Bernhard Bolzano (1781–1848) là một nhà toán học người Bohemian.

² Augustin-Louis Cauchy (21/8/1789–23/5/1857) là một nhà toán học người Pháp.

Hàm số $f(x) = x - \cos(x)$ là hàm liên tục trên $\mathbb{R}$. Hơn nữa, $f(0) = 0 - \cos(0) = -1 < 0$, trong khi đó $f(1) = 1 - \cos(1) > 0$.

Bởi định lí Bolzano–Cauchy, tồn tại ít nhất một nghiệm $\alpha \in (0, 1)$ thỏa mãn $\alpha = \cos(\alpha)$.

Định nghĩa 2.1: Khoảng tách nghiệm

Khoảng $[a, b]$ được gọi là *khoảng tách nghiệm* của phương trình (2.2) nếu phương trình có *một và chỉ một nghiệm* trên khoảng đó.

Trong các giáo trình khác nhau, khoảng tách nghiệm còn được gọi là “khoảng phân li nghiệm” hoặc “khoảng cách li nghiệm”.

Hàm đơn điệu và tính duy nhất nghiệm

Nhắc lại rằng một hàm số $f(x)$ là *đơn điệu tăng* trên một khoảng $[a, b]$ nếu mệnh đề sau là đúng với mọi cặp x_1, x_2 thuộc $[a, b]$:

$$x_1 < x_2 \implies f(x_1) < f(x_2).$$

Tương tự, hàm số f được gọi là hàm *đơn điệu giảm* trên $[a, b]$ nếu mệnh đề sau là đúng với mọi cặp x_1, x_2 thuộc $[a, b]$:

$$x_1 < x_2 \implies f(x_1) > f(x_2).$$

Các hàm đơn điệu tăng và đơn điệu giảm được gọi chung là hàm *đơn điệu*.

Một phương pháp đơn giản để xét tính đơn điệu của các hàm số khả vi là xét dấu của đạo hàm của nó. Cụ thể được nêu trong định lí sau đây.

Định lí 2.2:

Giả sử $f(x)$ là một hàm khả vi trên đoạn $[a, b]$ với đạo hàm $f'(x)$. Nếu $f'(x) > 0$ trên $[a, b]$ thì f là hàm đơn điệu tăng trên $[a, b]$. Nếu $f'(x) < 0$ trên $[a, b]$ thì f là hàm đơn điệu giảm trên $[a, b]$.

Đây là một kết quả quen thuộc của giải tích, là cơ sở của các bài toán khảo sát hàm số trong môn toán ở giáo dục phổ thông. Người đọc có thể xem chứng minh chi tiết trong rất nhiều giáo trình toán cao cấp khác nhau về giải tích, ví dụ N. V. Khuê–L. M. Hải–N. Đ. Sang [8].

Định lí 2.3: Giá trị trung gian

Giả sử $f(x)$ là một hàm khả vi trên đoạn $[a, b]$ với đạo hàm $f'(x)$ không đổi dấu, tức là $f'(x) > 0$ với mọi $x \in (a, b)$ hoặc $f'(x) < 0$ với mọi $x \in (a, b)$. Nếu $f(a)$ và $f(b)$ trái dấu thì $[a, b]$ là một khoảng tách nghiệm của phương trình $f(x) = 0$.

Ví dụ 2.3. Xét hàm số

$$f(x) = x^3 - x - 3, \quad x \in [1, 2].$$

Hàm số $f(x)$ khả vi với

$$f'(x) = 3x^2 - 1.$$

Rõ ràng, $f'(x) > 0$ trên $[1, 2]$, $f(1) = -3$, và $f(2) = 3$. Vậy, $f(x)$ đơn điệu tăng trên $[a, b]$, hai giá trị $f(1)$ và $f(2)$ trái dấu nên theo Định lí 2.2, khoảng $[1, 2]$ là một khoảng tách nghiệm của phương trình $f(x) = x^3 - x - 3 = 0$.

Giả sử $f(x)$ có nghiệm duy nhất α trên $[a, b]$, tức là $f(\alpha) = 0$. Giả sử bằng một phương pháp nào đó chúng ta tìm được $\tilde{x} \in [a, b]$ là một xấp xỉ của α . Vấn đề đặt ra là đánh giá sai số của phép xấp xỉ $\tilde{x} \approx \alpha$. Một câu trả lời cho vấn đề này như sau:

Định lí 2.4: Ước lượng sai số

Giả sử $f(x)$ là hàm khả vi trên $[a, b]$ và đạo hàm của f thỏa mãn

$$|f'(x)| \geq m > 0, \quad x \in [a, b],$$

với m là một hằng số dương. Giả sử $\alpha \in (a, b)$ là một nghiệm của phương trình $f(x) = 0$ và $\tilde{x} \in (a, b)$ là một số xấp xỉ của α . Khi đó

$$|\tilde{x} - \alpha| \leq \frac{|f(\tilde{x})|}{m}.$$

Chứng minh. Giả sử $\tilde{x} \neq \alpha$. Theo định lí Lagrange, tồn tại c nằm giữa $\tilde{x}$ và α sao cho

$$\frac{f(\tilde{x}) - f(\alpha)}{\tilde{x} - \alpha} = f'(c).$$

Thay $f(\alpha) = 0$ và lấy giá trị tuyệt đối hai vế, ta có

$$\left| \frac{f(\tilde{x})}{\tilde{x} - \alpha} \right| = |f'(c)| \geq m.$$

Từ đây ta dễ dàng suy ra điều phải chứng minh. □

Ví dụ sau đây minh hoạ cách áp dụng định lí trên để đánh giá sai số cho một nghiệm gần đúng.

Ví dụ 2.4. Xét phương trình $f(x) = x^3 - x - 3 = 0$ có một khoảng tách nghiệm là $[1, 2]$. Giả sử $\hat{x} = 1,67$ là một nghiệm gần đúng. Tìm sai số tương đối của nó.

Lời giải

Bởi vì $f'(x) = 3x^2 - 1$ nên với mọi $x \in [1, 2]$, ta có

$$|f'(x)| \geq f'(x) \geq 3 \times 1^2 - 1 = 2.$$

Do đó, chúng ta lấy $m = 2$.

Với $\hat{x} = 1,67$, ta dễ dàng tính được

$$|f(\hat{x})| = |1,67^3 - 1,67 - 3| = 0,012537.$$

Từ đó, nếu α là nghiệm đúng của phương trình trong $[1, 2]$ thì

$$|\hat{x} - \alpha| \leq \frac{|f(\hat{x})|}{m} = \frac{0,012537}{2} = 0,006268.$$

Như vậy, nếu $\hat{x} = 1,67$ là một nghiệm gần đúng thì nó có sai số tuyệt đối $\Delta_{\hat{x}} = 0,006268$, trong khi đó, sai số tương đối của nó là

$$\delta_{\hat{x}} = \frac{\Delta_{\hat{x}}}{\hat{x} - \Delta_{\hat{x}}} \approx 0,0038.$$

Như vậy, nghiệm gần đúng $\hat{x} = 1,67$ có sai số tương đối khá nhỏ, khoảng 0,38%.

2.2 Phương pháp chia đôi

Giả sử phương trình $f(x) = 0$ có nghiệm duy nhất $x = \alpha$ trong khoảng tách nghiệm $[a, b]$ với $f(a)f(b) < 0$. Trong phương pháp chia đôi (bisection method) chúng ta bắt đầu bằng việc đặt

$$a_1 = a, \quad b_1 = b, \quad m_1 = \frac{a_1 + b_1}{2}.$$

Khi đó, điểm m_1 chia khoảng $[a_1, b_1] = [a, b]$ làm hai khoảng có độ dài bằng một nửa khoảng tách nghiệm ban đầu. Tính giá trị $f(m_1)$ và tùy theo kết quả thu được mà thực hiện bước tiếp theo. Cụ thể như sau:

- Nếu $f(m_1) = 0$ thì nghiệm $\alpha = m_1$.

- Nếu $f(m_1) \neq 0$, thì $f(m_1)$ cùng dấu với $f(a_1)$ hoặc $f(b_1)$. Khi đó, xét hai trường hợp con sau:
 - Nếu $f(m_1)$ và $f(a_1)$ cùng dấu, thì $f(m_1)$ và $f(b_1)$ trái dấu và $[m_1, b_1]$ là một khoảng tách nghiệm của mới phương trình. Khi đó, chúng ta đặt $a_2 = m_1$ và $b_2 = b_1$.
 - Nếu $f(m_1)$ và $f(b_1)$ cùng dấu, thì $f(m_1)$ và $f(a_1)$ trái dấu và $[a_1, m_1]$ là một khoảng tách nghiệm mới của phương trình. Khi đó, chúng ta đặt $a_2 = a_1$ và $b_2 = m_1$.

Như vậy, trong mọi trường hợp thì chúng ta xác định được một khoảng tách nghiệm mới với độ dài bằng một nửa khoảng tách nghiệm ban đầu:

$$b_2 - a_2 = \frac{b_1 - a_1}{2}.$$

Lặp lại quá trình trên, chúng ta thu được các khoảng tách nghiệm $[a_3, b_3]$, ..., $[a_k, b_k]$, ..., với

$$b_k - a_k = \frac{b_1 - a_1}{2^{k-1}}.$$

Khi đó, chúng ta có thể lấy $m_k = (a_k + b_k)/2$ là giá trị gần đúng của nghiệm α , với sai số

$$|m_k - \alpha| \leq \frac{b_1 - a_1}{2^k}.$$

CAS 2.3. Chúng ta hãy viết một chương trình R tính gần đúng nghiệm của phương trình phi tuyến bằng phương pháp chia đôi và áp dụng cho phương trình $x^4 - 3x + 1 = 0$ trên khoảng tách nghiệm $(1, 2)$. Chi tiết như sau:

```
# bisection_method() tìm nghiệm của một phương trình phi tuyến.
# Yêu cầu: (a, b) là một khoảng tách nghiệm với f(a)f(b) < 0.

bisection_method <- function(f, a, b, tol = 1e-6, m = 1000) {

  # Khởi tạo các giá trị ban đầu
  i <- 0
  fa <- f(a)
  fb <- f(b)

  # Vòng lặp chính
  while (abs(b - a) > tol && i < m) {
    c = (a + b) / 2
    fc <- f(c)

    # Kiểm tra f(c) và f(a) trái dấu và rẽ nhánh
  }
```

```

    if (sign(fc) * sign(fa) < 0) {
      b <- c
      fb <- fc
      i <- i + 1
    } else if (sign(fc) * sign(fb) < 0) {
      a <- c
      fa <- fc
      i <- i + 1
    } else {
      b <- c
      a <- c
      i <- i + 1
    }
  }

  # Kết thúc vòng lặp khi |a - b| < tol hoặc i >= m
  c = (a + b) / 2

  # Hàm trả về nghiệm xấp xỉ c và số vòng lặp i
  return(list(root = c, iter = i))
}

# Áp dụng bisection_method() cho  $x^4 - 3 * x + 1 = 0$ .

options(digits = 8)
f <- function(x) x^4 - 3 * x + 1
bisection_method(f, 1, 2, 1e-6)

## $root
## [1] 1,3074861
##
## $iter
## [1] 20

```

Nếu α là nghiệm đúng của phương trình thì hàm `bisection_method()` như trên cho ra một giá trị x^* thỏa mãn $|x^* - \alpha| < \text{tol}$ hoặc $f(x^*) = 0$. Kết quả chạy máy tính cho nghiệm gần đúng sau 20 bước lặp.

CAS 2.4. Phương pháp chia đôi được lập sẵn trong chức năng `bisect()` của gói `pracma` của phần mềm R. Ví dụ,

```

library(pracma)
f <- function(x) x^4 - 3 * x + 1
bisect(f, 1, 2, 20)

## $root

```

```
## [1] 1,3074865
##
## $f.root
## [1] 2,5732532e-06
##
## $iter
## [1] 20
##
## $estim.prec
## [1] 1,9073486e-06
```

Kết quả thu được từ `bisect()` của gói `pracma` với 20 bước lặp phù hợp với kết quả của chương trình R trong CAS 2.3 với sai số được yêu cầu là $\text{tol} < 10^{-6}$.

2.3 Phương pháp lặp đơn

Nội dung chính của mục này là phương pháp lặp đơn, còn gọi là phương pháp lặp điểm bất động. Tài liệu tham khảo: T. V. Đĩnh [6, §2.3], P. K. Anh [2, §VII.2], Kreyszig [9, §19.2].

Giả sử phương trình (2.2) có một nghiệm α trong khoảng tách nghiệm $[a, b]$. Phương pháp lặp đơn nói chung bắt đầu bằng việc chuyển (2.2) về dạng

$$x = \phi(x), \quad (2.4)$$

sao cho (2.4) tương đương với (2.2) trong $[a, b]$, có nghĩa là (2.4) cũng có $[a, b]$ là khoảng tách nghiệm với nghiệm α . Như vậy, bài toán tìm nghiệm của phương trình $f(x) = 0$ quy về bài toán tìm “điểm bất động” của hàm số ϕ .

Chọn một xấp xỉ đầu tiên $x_0 \in [a, b]$ và xây dựng dãy lặp theo công thức truy hồi

$$x_n = \phi(x_{n-1}), \quad n = 1, 2, \dots \quad (2.5)$$

Với một số điều kiện nhất định, giá trị x_n , khi n “đủ lớn” cho một giá trị gần đúng của nghiệm α với sai số “đủ nhỏ”. Phương pháp lặp sử dụng công thức truy hồi (2.5) gọi là phương pháp *lặp đơn*, vì các bước lặp sử dụng cùng một công thức truy hồi.

2.3.1 Sự hội tụ

Định nghĩa 2.2:

Nếu dãy $\{x_n\}_{n \geq 0}$ xác định bởi (2.5) hội tụ tới nghiệm α của (2.4) khi $n \rightarrow \infty$ thì ta nói phương pháp lặp đơn (2.5) *hội tụ*.

Nếu phương pháp lập hội tụ thì, với một sai số cho phép $\epsilon > 0$ cho trước, khi lấy n “đủ lớn”, ta có $\alpha \approx x_n$ với sai số tuyệt đối giới hạn không vượt quá ϵ . Ngược lại, nếu quá trình lập không hội tụ, thì các giá trị x_n không có ý nghĩa trong bài toán tính gần đúng nghiệm.

Định lý sau đây cho các điều kiện đủ để có được sự hội tụ trong phương pháp lập đơn.

Định lý 2.5: Định lý hội tụ

Giả sử phương trình (2.4) có một nghiệm α trong $[a, b]$ và hàm số ϕ có đạo hàm thỏa mãn hai điều kiện sau:

(i) Với mọi $x \in [a, b]$, $\phi(x) \in [a, b]$.

(ii) $|\phi'(x)| \leq q < 1$ với mọi $x \in [a, b]$.

Khi đó, x_n hội tụ tới α :

$$\lim_{n \rightarrow \infty} x_n = \alpha.$$

Định lý 2.5 là một hệ quả của định lý điểm bất động Banach³. Vì vậy, phương pháp lập đơn dựa trên định lý này thường được gọi là *phương pháp lập điểm bất động*. Trước khi chứng minh định lý trên, chúng ta tìm hiểu ví dụ cụ thể sau đây.

Ví dụ 2.5. Xét phương trình

$$f(x) = x^3 - x - 1 = 0. \quad (2.6)$$

Biết nó có một nghiệm α phân li trong đoạn $[1, 2]$. Bằng phương pháp lập đơn, hãy tính gần đúng α với sai số không vượt quá $0,5 \times 10^{-4}$.

Lời giải

Đưa phương trình đã cho về dạng phương trình điểm bất động $x = \phi(x)$, bắt đầu bằng biến đổi nó về dạng

$$x^3 = x + 1.$$

Tiếp theo, chia hai vế cho x ,

$$x^2 = 1 + \frac{1}{x}.$$

Lấy căn bậc hai hai vế,

$$x = \underbrace{\sqrt{1 + \frac{1}{x}}}_{\phi(x)}.$$

³Stefan Banach (30/3/1892 – 31/8/1945) là một nhà toán học người Ba Lan.

Dễ thấy các phép biến đổi trên là tương đương trên $[1, 2]$. Như vậy, chúng ta đã đưa phương trình đã cho về dạng $x = \phi(x)$ với

$$\phi(x) = \sqrt{1 + \frac{1}{x}}.$$

Nếu $x \in [1, 2]$ thì $0,5 \leq \frac{1}{x} \leq 1$. Do đó,

$$\sqrt{1,5} \leq \phi(x) \leq \sqrt{2}.$$

Từ đó suy ra $\phi(x) \in [1, 2]$. Như vậy, hàm $\phi(x)$ thỏa mãn điều kiện (i) của Định lí 2.5.

Mặt khác, lấy đạo hàm của $\phi(x)$, ta có

$$\phi'(x) = \frac{-1}{2x\sqrt{x+1}}.$$

Với $x \in [1, 2]$ thì $2x\sqrt{x+1} \geq 2 \times 1 \times \sqrt{1+1} = 2\sqrt{2}$. Từ đó, chúng ta suy ra

$$|\phi'(x)| = \frac{1}{2x\sqrt{x+1}} \leq \frac{1}{2\sqrt{2}} < 1.$$

Vậy, hàm $\phi(x)$ thỏa mãn điều kiện (ii) trên $[1, 2]$ với $q = \frac{1}{2\sqrt{2}}$.

Tóm lại, $\phi(x)$ thỏa mãn cả hai điều kiện của Định lí 2.5: Phương pháp lặp đơn hội tụ với mọi cách chọn $x_0 \in [1, 2]$.

Để đơn giản, ta chọn $x_0 = 1,5$ và thiết lập dãy

$$\begin{cases} x_0 = 1,5, \\ x_{i+1} = \sqrt{1 + \frac{1}{x_i}}. \end{cases}$$

Hãy tính các xấp xỉ x_i . Tại mỗi bước, ta quy tròn kết quả với 4 chữ số sau dấu thập phân, và thay vào bước tiếp theo. Kết quả được ghi trong bảng sau:

i	x_i	$ f(x_i) $	$ x_i - x_{i-1} $
0	1,5000	0,8750	-
1	1,2910	0,1393	0,2090
2	1,3321	0,0317	0,0411
3	1,3231	0,0069	0,0090
4	1,3251	0,0016	0,0020
5	1,3246	0,0005	0,0005
6	1,3247	$7,658 \cdot 10^{-5}$	0,0001

Chú ý rằng, với $x_6 = 1,3247$, giá trị $|f(x_6)|$ rất nhỏ. Ta lấy $\alpha \approx x_6$ và đánh giá sai số như sau: Với $f(x) = x^3 - x - 1$, chúng ta có

$$f'(x) = 3x^2 - 1.$$

Từ đó suy ra $|f'(x)| \geq 2$ khi $x \in [1, 2]$. Như vậy,

$$|x_6 - \alpha| \leq \frac{|f(x_6)|}{2} \leq 3,829 \cdot 10^{-5}.$$

Vậy, x_6 là một xấp xỉ thỏa mãn yêu cầu, có nghĩa là $\alpha \approx x_6 = 1,3247$.

Chú ý 2.1. Có nhiều cách đưa một phương trình dạng $f(x) = 0$ về dạng $x = \phi(x)$. Chẳng hạn, chúng ta có thể đưa (2.6) về dạng

$$x = \sqrt[3]{x+1},$$

tức là dạng $x = \phi(x)$ với $\phi(x) = \sqrt[3]{x+1}$. Khi đó,

$$\sqrt[3]{2} \leq \phi(x) \leq \sqrt[3]{3} \quad \forall x \in [1, 2].$$

Hơn nữa,

$$\phi'(x) = \frac{1}{3}(x+1)^{-2/3},$$

nên

$$|\phi'(x)| \leq \frac{1}{3\sqrt[3]{4}} < 1.$$

Do đó, với cách chọn hàm ϕ như vậy, phương pháp lập đơn như trên cũng hội tụ với mọi xấp xỉ đầu $x_0 \in [1, 2]$.

Phần tiếp theo, chúng ta chứng minh Định lí 2.5.

Mệnh đề 2.1:

Giả sử ϕ là hàm số thỏa mãn điều kiện trong định lí 2.5. Khi đó, với mọi $x, y \in [a, b]$,

$$|\phi(x) - \phi(y)| \leq q|x - y|.$$

Chứng minh. Nếu $x = y$ thì đẳng thức trên là rõ ràng. Nếu $y \neq x$, không mất tổng quát, ta giả sử $y < x$. Theo định lí Lagrange (xem N. V. Khuê-L. M. Hải-N. Đ. Sang [8, Chương II, §2]), tồn tại $z \in [y, x]$ sao cho

$$\phi'(z) = \frac{\phi(y) - \phi(x)}{y - x}.$$

Vậy

$$|\phi(x) - \phi(y)| = |\phi'(z)| \cdot |x - y| \leq q|x - y|$$

vì $|\phi'(z)| \leq q$. Ta có điều phải chứng minh. □

Chứng minh Định lí 2.5. Để chứng minh x_n hội tụ tới α , chúng ta sẽ chứng minh

$$|x_n - \alpha| \leq q^n |x_0 - \alpha|. \quad (2.7)$$

Rõ ràng, bất đẳng thức (2.7) đúng với $n = 0$. Giả sử nó đúng với $n = k$, tức là

$$|x_k - \alpha| \leq q^k |x_0 - \alpha|. \quad (2.8)$$

Vì $\alpha = \phi(\alpha)$ và $x_{k+1} = \phi(x_k)$ nên, áp dụng Định lí 2.1 và giả thiết quy nạp, ta có

$$\begin{aligned} |x_{k+1} - \alpha| &= |\phi(x_k) - \phi(\alpha)| \\ &\leq q |x_k - \alpha| \\ &\leq q(q^k |x_0 - \alpha|) \\ &= q^{k+1} |x_0 - \alpha|. \end{aligned}$$

Vậy, (2.7) đúng với $n = k + 1$. Theo nguyên lí quy nạp, nó phải đúng với mọi $n \geq 0$.

Chú ý rằng $0 < q < 1$ nên q^n hội tụ về 0 khi $n \rightarrow \infty$. Áp dụng định lí giới hạn bị kẹp (N. V. Khuê–L. M. Hải–N. Đ. Sang [8, Chương I, §4.5]) vào bất đẳng thức (2.7), ta có $|x_n - \alpha| \rightarrow 0$ hay $x_n \rightarrow \alpha$ khi $n \rightarrow \infty$. Ta kết thúc chứng minh. $\square$

Nhận xét 2.1. Trong Định lí 2.5, chúng ta có thể thay giả thiết (i), điều kiện $\phi(x) \in [a, b]$ với mọi $x \in [a, b]$, bởi điều kiện $x_n \in [a, b]$ với mọi n . Mặt khác, ví dụ sau đây cho thấy điều kiện này là cần thiết.

Ví dụ 2.6. Xét hàm số $\phi(x) = 1 - \sqrt{x}$. Phương trình $x = \phi(x)$ có dạng phương trình bậc hai của $\sqrt{x}$:

$$x + \sqrt{x} - 1 = 0. \quad (2.9)$$

Khi đó, $t = \sqrt{x}$ phải là nghiệm dương của phương trình

$$t^2 + t - 1 = 0,$$

tức là $t = \frac{1}{2}(-1 + \sqrt{5})$, và vậy thì

$$x = \left(\frac{-1 + \sqrt{5}}{2} \right)^2 = \frac{3 - \sqrt{5}}{2} \approx 0,382.$$

Gọi nghiệm này là α thì rõ ràng $\alpha \in [1/3, 1]$ và nó là một khoảng tách nghiệm của phương trình (2.9).

Mặt khác, bởi vì

$$\phi'(x) = -\frac{1}{2\sqrt{x}},$$

nên chúng ta có đánh giá

$$|\phi'(x)| \leq \frac{\sqrt{3}}{2} < 1, \quad \forall x \geq 1/3.$$

Có nghĩa là ϕ thỏa mãn điều kiện (i) của Định lí 2.5 với $q = \frac{\sqrt{3}}{2}$.

Bây giờ, chúng ta xem xét hai cách chọn xấp xỉ đầu x_0 khác nhau và chúng cho các kết quả khác nhau:

- Chọn $x_0 = 1$ và xác định dãy $x_n = \phi(x_{n-1})$ thì chúng ta có $x_1 = \phi(1) = 0$. Vậy thì x_1 nằm ngoài khoảng tách nghiệm. Tiếp tục tính, ta có $x_2 = \phi(x_1) = \phi(0) = 1, \dots$ Một cách tổng quát, chúng ta tính được

$$x_n = \begin{cases} 0, & \text{nếu } n \text{ lẻ,} \\ 1, & \text{nếu } n \text{ chẵn.} \end{cases}$$

Rõ ràng, dãy x_n không hội tụ: Dãy lặp x_n không cho ta một xấp xỉ đủ tốt của nghiệm của phương trình ban đầu. Trong trường hợp này, quá trình tính x_n là một ví dụ về *quá trình tính không ổn định*.

- Chọn $x_0 = 1/3$ và xác định $x_n = \phi(x_{n-1})$ thì ta được $x_1 = \frac{1}{3}(3 - \sqrt{3}), \dots$ Có thể chứng minh được dãy này luôn nằm trong $[1/3, 1]$ và do đó nó hội tụ tới α .

Lưu ý, bạn đọc có thể kiểm tra quá trình lặp đơn trên hội tụ với mọi cách chọn $x_0 \in (0, 1)$.

Về cách chọn xấp xỉ đầu tiên

Từ Ví dụ 2.6, nảy sinh vấn đề chọn giá trị xấp xỉ ban đầu x_0 phù hợp sao cho dãy lặp x_n hội tụ.

Giả sử ta đã có điều kiện (ii) của Định lí 2.5, $|\phi'(x)| \leq q < 1$ trên $[a, b]$. Khi đó, chúng ta chọn xấp xỉ đầu tiên dựa theo các nguyên tắc sau:

- *Trường hợp 1:* Nếu $\phi'(x) > 0$ với mọi $x \in [a, b]$. Khi đó, ta chọn x_0 tùy ý. Nếu $a \leq x_0 < \alpha$ thì dãy lặp tăng và hội tụ tới α , còn nếu $\alpha < x_0 < b$ thì dãy lặp giảm và hội tụ tới α .
- *Trường hợp 2:* Nếu $\phi'(x) < 0$, hoặc đổi dấu trên $x \in [a, b]$. Khi đó, ta cần chọn x_0 sao cho x_0 gần α hơn so với hai đầu mút a, b , có nghĩa là

$$|x_0 - \alpha| \leq \min\{|a - \alpha|, |b - \alpha|\}. \quad (2.10)$$

Nói riêng, nếu $\alpha < (a+b)/2$ thì ta có thể chọn $x_0 = a$, còn nếu $(a+b)/2 < \alpha$ thì ta có thể chọn $x_0 = b$. Trong mọi trường hợp, dãy x_n “dao động” xung quanh và hội tụ về α .

2.3.2 Đánh giá sai số

Giả sử phương trình (2.2) hoặc (2.4) có nghiệm duy nhất α trong khoảng tách nghiệm $[a, b]$ và dãy các xấp xỉ x_n xác định bởi (2.5) thỏa mãn $x_n \rightarrow \alpha$. Câu hỏi đặt ra là với giá trị n lớn bao nhiêu thì chúng ta đạt được sai số bé như mong muốn. Nói cụ thể hơn, giả sử chúng ta muốn sai số tuyệt đối $|x_n - \alpha| < \epsilon$ với ϵ cho trước, thì cần tính đến bước lặp thứ mấy?

Để trả lời câu hỏi này, chúng ta cần tìm một sai số tuyệt đối giới hạn của x_n (phụ thuộc vào n). Từ bất đẳng thức (2.7), ta thu được

$$|x_n - \alpha| < q^n |x_0 - \alpha| < q^n |b - a|.$$

Như vậy nếu chọn n sao cho $q^n |b - a| < \epsilon$, hay

$$n > \frac{\ln(\epsilon/(b-a))}{\ln q}, \quad (2.11)$$

thì giá trị x_n là một xấp xỉ của nghiệm α với sai số tuyệt đối không lớn hơn ϵ . Tuy nhiên, trong thực hành, giá trị n thỏa mãn (2.11) như trên thường quá lớn.

Thay vì đánh giá sai số $|x_n - \alpha|$ theo a, b và q , chúng ta có thể đánh giá nó qua x_n và x_{n-1} . Chúng ta bắt đầu với

$$\begin{aligned} |x_n - \alpha| &\leq q |x_{n-1} - \alpha| \\ &\leq q |x_{n-1} - x_n + x_n - \alpha| \\ &\leq q (|x_{n-1} - x_n| + |x_n - \alpha|). \end{aligned}$$

Từ đó suy ra

$$(1 - q) |x_n - \alpha| \leq q |x_{n-1} - x_n|.$$

Vì $1 - q > 0$ nên ta chia hai vế của bất đẳng thức cho $1 - q$ và thu được công thức đánh giá sai số thứ nhất:

$$|x_n - \alpha| \leq \frac{q}{1 - q} |x_{n-1} - x_n|. \quad (2.12)$$

Nhận xét 2.2. Bằng cách áp dụng Định lí 2.4 cho hàm $f(x) = x - \phi(x)$, chúng ta cũng có thể thu được công thức đánh giá sai số thứ nhất. Thật vậy, ta có

$$|f'(x)| = |1 - \phi'(x)| \geq 1 - |\phi'(x)| \geq 1 - q.$$

Vậy, ta chọn $m = 1 - q > 0$ thì $|f'(x)| \geq m$ với mọi $x \in [a, b]$. Theo định lí trên

$$\begin{aligned} |x_n - \alpha| &\leq \frac{|f(x_n)|}{m} = \frac{|x_n - \phi(x_n)|}{1 - q} \\ &= \frac{|\phi(x_{n-1}) - \phi(x_n)|}{1 - q} \\ &\leq \frac{q |x_{n-1} - x_n|}{1 - q}. \end{aligned}$$

Đó chính là công thức đánh giá sai số thứ nhất.

2.3.3 Tóm tắt phương pháp lập đơn

Phương pháp lập đơn được thực hiện qua các bước sau:

1. Cho phương trình $f(x) = 0$,
2. Ấn định sai số cho phép ϵ ,
3. Tìm khoảng tách nghiệm $[a, b]$,
4. Tìm hàm lặp hội tụ $\phi(x)$ thoả mãn điều kiện hội tụ,
5. Chọn xấp xỉ đầu x_0 ,
6. Thực hiện quá trình lặp

$$x_n = \phi(x_{n-1}),$$

cho đến khi một trong các điều kiện dừng thỏa mãn,

7. Kết quả $\alpha \approx x_n$.

CAS 2.5. Từ tóm tắt trên, dễ dàng xây dựng một chương trình máy tính để thực hiện phương pháp lập đơn. Dưới đây là một đoạn mã R minh hoạ cách dùng vòng lặp “for” trong hàm `simple_iteration()` thực hiện quá trình lập đơn với hàm lặp ϕ và bắt đầu tại `start`. Đoạn mã sẽ giúp thực hiện các tính toán trong Ví dụ 2.5 một cách nhanh chóng.

```
# Hàm simple_iteration() nhận giá trị đầu x0, hàm lặp (phi),
# sai số chấp nhận (tol) và số lần lặp tối đa (m)
simple_iteration <- function(x0, phi, tol = 1e-6, m = 100)
{
  # Giá trị xấp xỉ đầu
  x <- x0

  # Vòng lặp chính
  for (i in 1:m) {
    x1 <- phi(x)
    if (abs(x1 - x) < tol) {
      cat("Kết quả sau", i, "bước:", x1, "\n")
      break
    }
    x <- x1
  }
  return(x)
}
```

```

}

# Chạy thử simple_iteration() cho hàm phi(x) = sqrt(1 + 1/x).
phi <- function(x) sqrt(1 + 1 / x)

# Giá trị đầu x0 = 1,5
x0 <- 1.5

# Nghiệp xấp xỉ tính bằng simple_iteration()
solution <- simple_iteration(x0, phi)

## Kết quả sau 9 bước: 1,3247178

```

2.4 Phương pháp Newton–Raphson

Mục này nói về một phương pháp mang tên nhà khoa học Issac Newton và nhà toán học Joseph Raphson⁴ để xấp xỉ nghiệm của một phương trình phi tuyến. Tài liệu tham khảo: T. V. Đĩnh [6, §2.4], P. K. Anh [2, §VII.4], Kreyszig [9, §19.2].

2.4.1 Mô tả phương pháp Newton–Raphson

Cũng như phương pháp lặp đơn, *phương pháp Newton–Raphson* xây dựng một dãy $x_0, x_1, x_2, \dots$, theo một công thức truy hồi để xấp xỉ nghiệm của một phương trình dạng (2.2). Phương pháp này bắt đầu bằng việc thay thế phương trình ban đầu (2.2), thường là phương trình phi tuyến, bằng một phương trình xấp xỉ có dạng tuyến tính.

Nhắc lại rằng, nếu hàm số $f(x)$ là hàm khả vi tại x_0 thì

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Do đó, nếu chúng ta đặt

$$L_{x_0}(x) = f(x_0) + f'(x_0)(x - x_0),$$

thì có thể khai triển $f(x)$ như sau:

$$f(x) = L_{x_0}(x) + h(x) \tag{2.13}$$

với $h(x)$ thỏa mãn

$$\lim_{x \rightarrow x_0} \frac{h(x)}{x - x_0} = 0.$$

⁴Joseph Raphson (khoảng 1668 – 1715) là một nhà toán học người Anh.

Công thức (2.13) là trường hợp riêng của khai triển Taylor của hàm f tại điểm x_0 và L_{x_0} gọi là xấp xỉ tuyến tính của f tại x_0 .

Xét phương trình “xấp xỉ” tuyến tính

$$L_{x_0}(x) = 0$$

của phương trình (2.2) tại x_0 . Với điều kiện $f'(x_0) \neq 0$, nó có nghiệm

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Tiếp theo, xét phương trình tuyến tính hóa tại x_1 , phương trình $L_{x_1}(x) = 0$. Dễ thấy nếu $f'(x_1) \neq 0$ thì nghiệm của nó là

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Lặp lại quá trình này, ta thu được một dãy $x_0, x_1, x_2, \dots$

Tóm lại, chúng ta thu được dãy $\{x_n\}$ xác định bởi hai phương trình

$$\begin{cases} x_0 & \text{cho trước,} \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \end{cases} \quad (2.14)$$

gọi là *quá trình Newton* bắt đầu từ x_0 .

Nhận xét 2.3. Nếu $f'(x) \neq 0$ trên $[a, b]$ thì hàm số

$$\phi(x) := x - \frac{f(x)}{f'(x)}$$

xác định trên $[a, b]$. Khi đó, phương trình $f(x) = 0$ tương đương với $x = \phi(x)$. Vì vậy, phương pháp Newton có thể xem là một trường hợp đặc biệt của phương pháp lặp mà chúng ta đã tìm hiểu trong mục trước.

2.4.2 Sự hội tụ

Phương pháp Newton–Raphson có ý nghĩa khi dãy x_n hội tụ tới nghiệm α của phương trình $f(x) = 0$ đang xét. Sự hội tụ này có được trong trường hợp tổng quát khi một số điều kiện được thỏa mãn. Ví dụ, sự hội tụ có được khi f khả vi liên tục đến cấp 2, $f'(\alpha) \neq 0$, và xấp xỉ ban đầu x_0 chọn “đủ gần” α (Định lý 2.7). Tuy nhiên, kết quả này chỉ mang ý nghĩa lý thuyết, vì nó không nêu rõ chọn x_0 thế nào là đủ gần α . Trong thực hành, việc chọn x_0 như nào để dãy x_n xây dựng bởi thuật toán Newton–Raphson hội tụ tới nghiệm cần tìm, ta dựa trên kết quả sau đây.

Định lí 2.6: Newton–Raphson

Giả sử $[a, b]$ là một khoảng tách nghiệm của phương trình $f(x) = 0$, hàm $f(x)$ có đạo hàm tới cấp 2 với f' và f'' không đổi dấu trên (a, b) . Giả sử x_0 là a hoặc b sao cho $f(x_0)$ cùng dấu với $f''(x_0)$. Khi đó, dãy x_n xác định bởi (2.14) hội tụ tới nghiệm α khi $n \rightarrow \infty$. Cụ thể hơn, x_n là dãy đơn điệu tăng tới α nếu f' và f'' trái dấu và x_n là dãy đơn điệu giảm tới α nếu f' và f'' cùng dấu.

Chú ý 2.2. Điểm x_0 thỏa mãn điều kiện “ $f(x_0)$ và $f''(x_0)$ cùng dấu”, tức là

$$f(x_0)f''(x_0) > 0,$$

thường được gọi là *điểm Fourier*⁵.

Chứng minh Định lí 2.6. Vì f' và f'' giữ nguyên dấu trên khoảng đang xét, nên ta có thể chia làm 4 trường hợp tùy theo dấu của chúng. Tuy nhiên, ta sẽ chỉ trình bày chứng minh chi tiết trong trường hợp $f' > 0$ và $f'' < 0$ trên $[a, b]$, vì 3 trường hợp còn lại có thể được chứng minh tương tự.

Với giả thiết như vậy, $f(x)$ là hàm tăng trên $[a, b]$ nên $f(x) < 0$ trên $[a, \alpha)$ và $f(x) > 0$ trên $(\alpha, b]$. Như vậy, x_0 thỏa mãn điều kiện Fourier nếu và chỉ nếu $f(x_0) < 0$, tức là $x_0 < \alpha$.

Xét hàm số

$$\phi(x) = x - \frac{f(x)}{f'(x)}.$$

Khi đó, công thức lặp Newton–Raphson viết gọn là

$$x_n = \phi(x_{n-1}), \quad n = 1, 2, \dots$$

Ta sẽ chứng minh

$$x < \phi(x) < \alpha \tag{2.15}$$

với mọi $x \in [a, \alpha)$. Thật vậy, với $x \in [a, \alpha)$ ta có $f(x) < 0$ và $f'(x) > 0$ nên, theo định nghĩa của ϕ , ta có $\phi(x) > x$. Bất đẳng thức thứ nhất được chứng minh.

Mặt khác, bằng tính toán trực tiếp,

$$\begin{aligned} \phi'(x) &= \left(x - \frac{f(x)}{f'(x)} \right)' \\ &= 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} \\ &= \frac{f(x)f''(x)}{(f'(x))^2} > 0, \end{aligned}$$

⁵Jean-Baptiste Joseph Fourier (1768–1830) là một nhà toán học người Pháp.

với mọi $x \in [a, \alpha)$, theo giả thiết về dấu của các đạo hàm của f . Từ đó, theo định lý giá trị trung bình,

$$\phi(x) - \phi(\alpha) = (x - \alpha)\phi'(\xi) < 0,$$

do $x < \alpha$. Vậy,

$$\phi(x) < \phi(\alpha) = \alpha - \frac{f(\alpha)}{f'(\alpha)} = \alpha,$$

do $f(\alpha) = 0$. Từ đó, các bất đẳng thức trong (2.15) được chứng minh.

Bây giờ, áp dụng (2.15), chúng ta có $x_0 < \phi(x_0) = x_1 < \alpha$. Bằng quy nạp, ta suy ra

$$x_0 < x_1 < \dots < x_n < x_{n+1} < \dots < \alpha.$$

Vậy, dãy $\{x_n\}$ là dãy tăng ngặt và bị chặn trên bởi α nên nó hội tụ. Gọi ℓ là giới hạn của nó thì

$$\ell = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \phi(x_{n-1}) = \phi\left(\lim_{n \rightarrow \infty} x_n\right) = \phi(\ell),$$

có nghĩa là ℓ là nghiệm của phương trình $x = \phi(x)$, cũng là nghiệm của phương trình $f(x) = 0$ trong $[a, \alpha]$. Do đó, bắt buộc $\ell = \alpha$ và chúng ta kết thúc chứng minh. $\square$

Một trong những nhược điểm của phương pháp Newton–Raphson là cần phải tính không chỉ $f(x)$ mà cả đạo hàm $f'(x)$. Tuy nhiên, có những tình huống mà điều này lại giúp cho chúng ta đơn giản hóa việc tính toán, như trong hai ví dụ kinh điển sau đây:

Ví dụ 2.7. Cho $a > 0$. Xét phương trình

$$f(x) := a - \frac{1}{x} = 0. \quad (2.16)$$

Nó có nghiệm đúng duy nhất là $x = 1/a$.

Nếu biết một biểu diễn thập phân của a , thì chúng ta có thể tính $1/a$ bằng cách thực hiện phép chia trong hệ thập phân.

Mặt khác, $f'(x) = \frac{1}{x^2} > 0$ và $f''(x) = -2x^{-3} < 0$ (với $x \neq 0$), nên ta có thể dùng công thức truy hồi Newton–Raphson

$$x_{k+1} = 2x_k - ax_k^2, \quad (2.17)$$

để thu được một dãy hội tụ tới $1/a$, miễn là x_0 thỏa mãn điều kiện Fourier. Trong trường hợp này, x_0 phải thỏa mãn $f(x_0) < 0$ hay

$$x_0 < \frac{1}{a}.$$

Một điều thú vị nên lưu ý trong công thức truy hồi (2.17), chúng ta chỉ cần thực hiện phép cộng và nhân mà không cần đến phép chia. Nói cách khác,

chúng ta đã đơn giản hóa quá trình tính toán (thay phép chia bằng phép cộng và phép nhân), với một cái giá cần đánh đổi là kết quả thu được chỉ là gần đúng. Tuy nhiên, trong một số trường hợp, chúng ta cũng không thể biểu diễn giá trị đúng của phép chia $1/a$, ví dụ như $1/7$, vì kết quả của nó là một số thập phân vô hạn.

Để lấy ví dụ, xét $a = 7$ và chọn $x_0 = 0,1$. Khi đó, có thể dễ dàng tính bằng tay như sau:

$$x_1 = 2 \cdot 0,1 - 7 \cdot (0,1)^2 = 0,2 - 0,07 = 0,13,$$

$$x_2 = 2 \cdot 0,13 - 7 \cdot (0,13)^2 = 0,26 - 7 \cdot 0,0169 = 0,1417,$$

$$x_3 = 2 \cdot 0,1417 - 7 \cdot (0,1417)^2 = 0,2834 - 0,1405523 = 0,14284777.$$

Làm tròn kết quả cuối cùng, ta thu được $1/7 \approx 0,14285$ với 5 chữ số đáng tin.

Chú ý 2.3. Công thức truy hồi (2.17) có thể được mở rộng cho bài toán tìm nghịch đảo hoặc giả nghịch đảo của một ma trận trong thuật toán Hotelling–Bodewig (còn gọi là phương pháp Newton–Schultz). Bạn đọc quan tâm có thể tìm đọc P. K. Anh [2, §6.5] hoặc Dahlquist–Björck [5].

Ví dụ 2.8 (Thuật toán Heron⁶). Cho $a > 0$, xét phương trình bậc hai

$$f(x) := x^2 - a = 0. \quad (2.18)$$

Nghiệm dương của phương trình đó có thể viết dạng $x = \sqrt{a}$. Nhưng lời giải này chỉ có ý nghĩa về mặt lý thuyết, do căn bậc hai $\sqrt{a}$ theo định nghĩa là một số dương mà bình phương lên bằng a . Do đó, nếu cho một biểu diễn thập phân của a , thì cách viết $\sqrt{a}$ không cho chúng ta một biểu diễn thập phân của nghiệm $\sqrt{a}$.

Vấn đề của chúng ta là tính gần đúng $\sqrt{a}$ và viết kết quả dưới dạng số thập phân. Mặc dù điều này có thể dễ dàng tính được bằng máy tính, nhưng chúng ta hãy hình dung làm thế nào máy tính điện tử chỉ làm việc với các bit 0 và 1 lại có thể tính toán được những phép tính lấy căn bậc hai.

Chú ý rằng trên $(0, +\infty)$, $f'(x) = 2x > 0$ và $f''(x) = 2 > 0$. Với $x_0 > 0$ tùy ý, xét dãy số xây dựng theo phương pháp Newton (2.14):

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = \frac{1}{2} \left(x_n + \frac{a}{x_n} \right), \quad n = 0, 1, 2, \dots$$

Xảy ra các trường hợp sau:

1. $x_0 > \sqrt{a}$. Khi đó, phương trình $f(x) = 0$ có một khoảng tách nghiệm là $[0, x_0]$. Theo định lý 2.6, dãy x_n là dãy đơn điệu giảm tới $\sqrt{a}$, tức là $x_n \rightarrow \sqrt{a}$ khi $n \rightarrow \infty$.

⁶Thuật toán được mô tả bởi Hero của Alexandria vào thế kỷ thứ hai trong cuốn *Metrica* của ông.

2. $x_0 < a$. Theo bất đẳng thức Cauchy, ta có

$$x_1 = \frac{1}{2} \left(x_0 + \frac{a}{x_0} \right) \geq \sqrt{x_0 \left(\frac{a}{x_0} \right)} = \sqrt{a}.$$

Vậy, bắt đầu từ x_1 , dãy là đơn điệu giảm về $\sqrt{a}$.

3. $\sqrt{x_0} = \sqrt{a}$. Dãy là dãy dừng, tức là $x_0 = x_1 = \dots$.

Như vậy, bằng thuật toán Heron, chúng ta có thể xây dựng dãy lặp để xấp xỉ căn bậc hai của một số dương với công thức truy hồi chỉ bao gồm phép tính số học gồm phép cộng, trừ, nhân và chia.

Nhận xét 2.4. Hai phương trình (2.16) và (2.18) trong hai ví dụ trên là hai trường hợp riêng của phương trình tổng quát

$$x^p = a, \quad p = \pm 1, \pm 2, \dots \quad (2.19)$$

Dãy lặp Newton–Raphson $x_1, x_2, \dots$ cho xấp xỉ nghiệm được tính theo công thức truy hồi

$$x_{n+1} = \frac{1}{p} \left((p-1)x_n + a x_n^{1-p} \right). \quad (2.20)$$

Đó là công thức truy hồi được dùng trong một số thuật toán máy tính để tính các căn thức $\sqrt[p]{a}$ của các số a .

CAS 2.6. Để minh họa (2.20), chúng ta hãy viết một vòng lặp ngắn bằng ngôn ngữ R trong đó chỉ thực hiện các phép toán số học (cộng, trừ, nhân, chia) để tính căn bậc ba của một số thực.

*# Hàm cuberoot() nhận vào một số thực a và sai số cho phép tol,
cho ra căn bậc ba của nó.*

```
cuberoot <- function(a, tol = 1e-6) {
  if (a == 0) {
    return(0)
  } else {
    x <- 1
    while (abs(x * x * x - a) > tol) {
      x <- (2 * x + a / (x * x)) / 3
    }
    return(x)
  }
}
```

Chạy thử cuberoot() cho a = 10.

```
a <- 10
result <- cuberoot(a)
print(result)

## [1] 2,15443469
```

Kết quả chạy cuberoot(a) với tham số $a = 10$ cho $\sqrt[3]{10} \approx 2,154434692$.

2.4.3 Đánh giá sai số

Trong mục này, chúng ta trình bày hai kết quả về ước lượng sai số trong phương pháp Newton–Raphson. Kết quả thứ nhất đánh giá sai số trong bước thứ $n + 1$ thông qua sai số trong bước thứ n . Kết quả này suy ra rằng phương pháp Newton–Raphson hội tụ bậc hai. Kết quả thứ hai có nghĩa hơn về mặt thực hành, cho phép đánh giá sai số trong bước thứ $n + 1$ qua sự sai lệch giữa hai xấp xỉ thứ n và thứ $n + 1$.

Giả sử rằng hàm f có đạo hàm đến cấp hai liên tục, $f(\alpha) = 0$ và $f'(\alpha) \neq 0$. Giả sử $\{x_n\}$ là dãy xác định bởi thuật toán Newton–Raphson và $x_n \rightarrow \alpha$. Xét các sai số

$$\epsilon_n = \alpha - x_n, \quad n = 1, 2, \dots$$

Khi đó

$$\begin{aligned} \epsilon_{n+1} &= \alpha - x_{n+1} \\ &= \alpha - \left(x_n - \frac{f(x_n)}{f'(x_n)} \right) \\ &= \epsilon_n + \frac{1}{f'(x_n)} \left(f(\alpha) - f'(x_n)(\alpha - x_n) - \frac{1}{2} f''(c_n)(\alpha - x_n)^2 \right) \\ &= -\frac{f''(c_n)}{2f'(x_n)} \epsilon_n^2. \end{aligned} \quad (2.21)$$

Với mỗi n , c_n nằm giữa α và x_n . Vậy, c_n hội tụ tới α . Từ đó, ta có ước lượng

$$|\epsilon_{n+1}| \leq C |\epsilon_n|^2, \quad (2.22)$$

với C là một hằng số dương phụ thuộc vào đạo hàm cấp một và cấp hai của $f(x)$. Ví dụ, nếu $[a, b]$ là khoảng tách nghiệm của nghiệm α và $x_n \in [a, b]$ với mọi n thì ta có thể xét

$$M = \max_{x \in [a, b]} |f''(x)|, \quad m = \min_{x \in [a, b]} |f'(x)|, \quad (2.23)$$

và lấy

$$C = \frac{M}{2m}.$$

Ước lượng (2.22) có ý nghĩa về mặt lý thuyết. Nó nói rằng, với điều kiện $f' \neq 0$ và điều kiện hội tụ được thỏa mãn, phương pháp Newton–Raphson *hội tụ bậc hai* tới nghiệm α của $f(x)$.

Về mặt thực hành, ước lượng (2.22) khó áp dụng vì không cho ta đánh giá cụ thể sai số tại bước lặp thứ n . Ta có thể sử dụng đánh giá quen thuộc

$$|x_{n+1} - \alpha| \leq \frac{|f(x_{n+1})|}{m}, \quad (2.24)$$

với m xác định như trong (2.23). Mặt khác, khai triển Taylor với số dư dạng Lagrange của hàm f tại x_n cho ta

$$f(x_{n+1}) = f(x_n) + f'(x_n)(x_{n+1} - x_n) + \frac{f''(c_n)}{2}(x_{n+1} - x_n)^2.$$

Từ công thức truy hồi, ta có

$$f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0.$$

Cùng với (2.24), ta thu được

$$|x_{n+1} - \alpha| \leq \frac{M}{2m} |x_{n+1} - x_n|^2. \quad (2.25)$$

Ví dụ 2.9. Dùng phương pháp Newton–Raphson tính gần đúng nghiệm của phương trình quen thuộc sau:

$$x - \sin(x) - 0,25 = 0.$$

Các kết quả làm tròn tới 5 chữ số sau dấu phẩy thập phân (T. V. Đĩnh [6]).

Lời giải

Ta biết rằng $[\frac{\pi}{4}, \frac{\pi}{2}]$ là một khoảng tách nghiệm của phương trình đã cho. Với $f(x) := x - \sin(x) - 0,25$, ta có

$$f'(x) = 1 - \cos x, \quad f''(x) = \sin x.$$

Do trên khoảng đang xét, $f'(x)$ và $f''(x)$ đều dương nên ta chọn $x_0 = \pi/2$ là xấp xỉ đầu (thỏa mãn điều kiện Fourier) và xây dựng dãy theo quy trình Newton

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Các kết quả tính toán bằng máy tính được ghi trong bảng sau:

Để đánh giá sai số, chúng ta hãy xét đạo hàm $f'(x) = 1 - \cos(x)$. Rõ ràng, trong khoảng $x \in [\frac{\pi}{4}, \frac{\pi}{2}]$,

$$f'(x) > 1 - \frac{\sqrt{2}}{2} > 0,29288,$$

i	x_i	$f(x_i)$	$f'(x_i)$	$x_i - f(x_i)/f'(x_i)$
0	1,57080	0,32080	1	1,25000
1	1,25000	0,05102	0,68468	1,17548
2	1,17548	$0,26052 \times 10^{-2}$	0,61490	1,17124
3	1,17124	$0,63222 \times 10^{-5}$	0,61099	1,17123
4	1,17123	$0,21232 \times 10^{-6}$	...	...

Bảng 2.1: Tính toán trong Ví dụ 2.9

trong khi đó,

$$|f''(x)| = |\sin(x)| \leq 1.$$

Vậy, nếu lấy $M = 1, m = 0,29288$ thì với x_4 ,

$$|x_4 - \alpha| \leq \frac{f(x_4)}{m} \approx 1,1612 \times 10^{-7}.$$

Cũng có thể đánh giá

$$|x_4 - \alpha| \leq \frac{M}{2m} |x_4 - x_3|^2 \approx 2 \times 10^{-10}.$$

Các sai số tuyệt đối giới hạn này nhỏ hơn rất nhiều lần so với sai số quy tròn. Vậy, $\alpha \approx 1,17123$.

Định lý 2.7:

Giả sử f là hàm có đạo hàm đến cấp 2 liên tục trên $[a, b]$. Giả sử $\alpha \in (a, b)$ là một nghiệm của $f(x) = 0$, với $f'(\alpha) \neq 0$. Khi đó, tồn tại $\delta > 0$ sao cho với mọi $x_0 \in [\alpha - \delta, \alpha + \delta]$, quá trình lặp Newton–Raphson bắt đầu từ x_0 sẽ hội tụ tới α .

Chứng minh. Xét hàm số 보조

$$\phi(x) = x - \frac{f(x)}{f'(x)}.$$

Vì $f'(\alpha) \neq 0$ nên, bởi tính liên tục, $f(x) \neq 0$ trong một lân cận của α , tức là tồn tại $\delta_1 > 0$ sao cho $f'(x) \neq 0$ trên $(\alpha - \delta_1, \alpha + \delta_1)$.

Tính đạo hàm của ϕ , chúng ta thu được

$$\phi' = 1 - \frac{(f')^2 - f f''}{(f')^2} = \frac{f f''}{(f')^2}$$

xác định trên $(\alpha - \delta_1, \alpha + \delta_1)$. Hơn nữa, vì $f(\alpha) = 0$ nên $\phi'(\alpha) = 0$. Bởi tính liên tục của ϕ' trên $(\alpha - \delta_1, \alpha + \delta_1)$ (có được do đạo hàm cấp 2 của f liên tục), ta suy ra

$$\phi'(x) \approx \phi'(\alpha) = 0.$$

Có $q \in (0, 1)$ và $\delta > 0$ và nhỏ hơn δ_1 sao cho

$$|\phi'(x)| \leq q < 1, \quad \forall x \in [\alpha - \delta, \alpha + \delta].$$

Trên khoảng $[\alpha - \delta, \alpha + \delta]$, phương pháp lặp Newton–Raphson tương đương với phương pháp lặp đơn $x_n = \phi(x_{n-1})$, với hàm ϕ thỏa mãn điều kiện của Định lý 2.5. Hơn nữa, với mọi xấp xỉ đầu $x_0 \in [\alpha - \delta, \alpha + \delta]$, khoảng cách từ x_0 tới nghiệm đúng α nhỏ hơn khoảng cách bé nhất từ hai đầu mút tới α , tức là nó thỏa mãn (2.10). Do đó, áp dụng Định lý 2.5, có thể suy ra rằng phương pháp Newton–Raphson hội tụ với mọi cách chọn $x_0 \in [\alpha - \delta, \alpha + \delta]$. $\square$

Chú ý 2.4. Lưu ý rằng trong phương pháp Newton–Raphson để tìm nghiệm phương trình $f(x) = 0$, chúng ta yêu cầu hàm số $f(x)$ khả vi tại *mọi* điểm. Ví dụ sau đây là về trường hợp $f(x)$ khả vi trừ tại một điểm duy nhất trong khi đó dãy lặp thiết lập bởi phương pháp Newton–Raphson không hội tụ.

Xét hàm số $f(x) = \sqrt[3]{x}$. Rõ ràng $f(x)$ khả vi tại mọi $x \neq 0$. Phương trình $f(x) = 0$ có nghiệm duy nhất $x = 0$.

Mặt khác, với $x_n \neq 0$, ta có

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{\sqrt[3]{x_n}}{\frac{1}{3\sqrt[3]{x_n^2}}} = -2x_n.$$

Như vậy, với mọi cách chọn x_0 , dãy thiết lập bởi phương pháp Newton–Raphson có công thức truy hồi là

$$x_n = (-2)^n x_0.$$

Dãy đó hội tụ nếu và chỉ nếu $x_0 = 0$.

2.4.4 Tóm tắt thuật toán Newton–Raphson

Phương pháp Newton–Raphson thực hiện theo các bước sau:

1. Cho phương trình $f(x) = 0$.
2. Giải sơ bộ: Tìm một khoảng tách nghiệm $[a, b]$ sao cho f' và f'' không đổi dấu.
3. Ấn định sai số $\epsilon > 0$.
4. Chọn điểm x_0 thỏa mãn điều kiện Fourier: $f'(x_0)f''(x_0) > 0$.

5. Với mỗi $k = 0, 1, 2, \dots$, xác định một cách truy hồi

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

cho đến khi đạt được độ chính xác như mong muốn.

Từ tóm tắt thuật toán như trên, chúng ta dễ dàng viết được chương trình máy tính để thực hiện các phép tính. Chương trình mẫu R sau đây phát triển từ CAS 2.6 để thực hiện phương pháp Newton–Raphson cho một phương trình phi tuyến trong Ví dụ 2.9.

CAS 2.7. Xét phương trình phi tuyến

$$x - \sin(x) - 0,25 = 0$$

như trong Ví dụ 2.9. Bằng phương pháp lặp Newton–Raphson với xấp xỉ đầu $x_0 = \frac{\pi}{2}$ và điều kiện dừng vòng lặp là khi $|x_{i+1} - x_i| < 10^{-6}$ hoặc số lần lặp lớn hơn 100, chúng ta thu được kết quả như trên.

```
# Chương trình thực hiện lặp Newton - Raphson cho phương trình
# phi tuyến tính  $x - \sin(x) - 1/4 = 0$ , bắt đầu tại  $x_0 = \pi/2$ .

newton <- function(f, f_prime, start, tol = 1e-6, m = 100)
{
  x <- start
  for (i in 1:m) {
    x1 <- x - f(x) / f_prime(x)
    if (abs(x1 - x) < tol) {
      cat("Kết quả sau", i,
          "bước: ", x1, "\n")
      break
    }
    x <- x1
  }
  return(x)
}

f <- function(x) x - sin(x) - 0.25      # Hàm số  $f(x)$ 
f_prime <- function(x) 1 - cos(x)      # Đạo hàm  $f'(x)$ 
start <- pi/2                          # Bắt đầu với  $x = \pi/2$ 
newton(f, f_prime, start)

## Kết quả sau 5 bước:  1,17122965
## [1] 1,17122965
```

2.5 Phương pháp dây cung và một số phương pháp khác

Trong mục này, chúng ta tìm hiểu một số phương pháp khác để tính gần đúng nghiệm, gồm phương pháp dây cung (secant method), phương pháp vị trí sai (regular falsi), và phương pháp Steffensen. Tài liệu tham khảo: T. V. Đĩnh [6, §2.2], P. K. Anh [2, §VI.1, §VI.3].

2.5.1 Phương pháp dây cung

Phương pháp dây cung (secant method) là một trong những phương pháp “dự đoán-giới hạn” để tìm nghiệm của một hàm số thực. Phương pháp này tương đối đơn giản và hiệu quả, thường được sử dụng để tìm gần đúng các nghiệm của các hàm số không tuyến tính. Đặc biệt, phương pháp dây cung có thể áp dụng cho các hàm không có công thức đơn giản để tính toán nghiệm chính xác.

Ý tưởng chính của phương pháp dựa trên việc xấp xỉ đường thẳng qua hai điểm gần đúng của đồ thị hàm số. Giả sử ta muốn tìm nghiệm của hàm số $f(x) = 0$, biết $[a, b]$ là một khoảng tách nghiệm. Chúng ta bắt đầu bằng hai điểm $x_0 = a$ và $x_1 = b$. Sau đó, phương pháp dây cung cho nghiệm xấp xỉ mới x_2 bằng cách tính giao điểm của đường thẳng nối hai điểm $(x_0, f(x_0))$ và $(x_1, f(x_1))$ với trục hoành. Công thức tính x_2 trong phương pháp dây cung là:

$$x_2 = x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1) - f(x_0)}. \quad (2.26)$$

Sau khi tính được x_2 , lại dùng x_2 này kết hợp với x_1 để tính x_3 mới và tiếp tục quá trình lặp lại đến khi đạt được độ chính xác mong muốn hoặc số lần lặp đã đủ. Như vậy, công thức truy hồi của phương pháp này là

$$x_{n+1} = x_n - f(x_n) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}, \quad f(x_n) \neq f(x_{n-1}). \quad (2.27)$$

Phương pháp dây cung có một số ưu điểm như nó không yêu cầu tính đạo hàm, phù hợp với những hàm số phức tạp khó tính đạo hàm. Tuy nhiên, phương pháp này cũng có một số hạn chế như tốc độ hội tụ chậm hơn so với một số phương pháp khác như phương pháp Newton–Raphson. Hơn nữa, sự hội tụ phụ thuộc vào việc chọn điểm khởi tạo gần đúng ban đầu: Nếu chọn sai, quá trình lặp có thể dẫn đến sự hội tụ không đúng nghiệm hoặc không hội tụ.

CAS 2.8. Hãy viết một chương trình R đơn giản thực hiện phương pháp dây cung để xấp xỉ nghiệm của phương trình trong CAS 2.3.

```
# secant_method() tính xấp xỉ nghiệm theo phương pháp dây cung.
# Yêu cầu: Khoảng (a, b) là một khoảng tách nghiệm
# với  $f(a)f(b) < 0$ .

secant_method <- function(f, a, b, tol = 1e-6, m = 50) {
  i <- 0
  fa <- f(a)
  while (abs(b - a) > tol) {
    i <- i + 1
    if (i > m) {
      stop("Không thấy nghiệm sau ",
           m, " vòng lặp.")
    }
    fb <- f(b)
    if (fb == fa) {
      stop("Không thấy nghiệm sau ",
           i, " vòng lặp.")
    }
    x <- b - fb * ((b - a) / (fb - fa))
    a <- b
    b <- x
    fa <- fb
  }
  return(b)
}

# Áp dụng secant_method() cho phương trình  $x^4 - 3 * x + 1 = 0$ .

f <- function(x) x^4 - 3 * x + 1
secant_method(f, 1, 2, 1e-6, 50)

## [1] 1,3074861
```

Chúng ta thu được kết quả khớp với kết quả tính bằng phương pháp chia đôi trong CAS 2.3.

Chú ý 2.5. Phương pháp dây cung được cài đặt trong hàm `secant()` có sẵn trong gói `pracma` của R. Chúng ta hãy so sánh:

```
library(pracma)
f <- function(x) x^4 - 3 * x + 1
secant(f, 1, 2, 10)

## $root
## [1] 1,3074861
##
```



```
## $f.root
## [1] 6,43929354e-14
##
## $iter
## [1] 9
##
## $estim.prec
## [1] 3,79538578e-09
```

Kết quả thu được từ `secant()` của `pracma` với 9 bước lặp hoàn toàn trùng khớp với kết quả của chương trình R trong CAS 2.3 với sai số được yêu cầu là $\text{tol} < 10^{-6}$.

2.5.2 Phương pháp Steffensen

Phương pháp Steffensen⁷ là phương pháp tìm nghiệm gần đúng dựa theo công thức truy hồi sau:

$$x_{n+1} = x_n - \frac{f(x_n)}{g(x_n)}, \quad g(x_n) = \frac{f(x_n + f(x_n)) - f(x_n)}{f(x_n)}. \quad (2.28)$$

So sánh với phương pháp Newton–Raphson, phương pháp Steffensen không yêu cầu tính đạo hàm của hàm f và đặc điểm này có lợi hơn khi giải các phương trình $f(x) = 0$ trong đó hàm f có đạo hàm phức tạp. Tuy nhiên, trong mỗi bước của phương pháp Steffensen, chúng ta cần tính 2 giá trị của hàm số.

CAS 2.9. Hãy viết chương trình R dựa trên phương pháp Steffensen trong R để tìm nghiệm của phương trình $x^4 - 3x^3 + 1 = 0$ trong khoảng tách nghiệm $(1, 2)$.

```
# Hàm steffensen_method() lặp theo phương pháp Steffensen.

steffensen_method <- function(f, a, tol = 1e-6, m = 100) {
  i <- 0
  x <- a
  fx <- f(x)
  gx <- (f(x + fx) - fx) / fx
  while (i < m && abs(fx) > tol) {
    i <- i + 1
    if (gx == 0) {
      return(list(root = NA, iter = i))
    }
  }
}
```

⁷Johan Frederik Steffensen (28/02/1873 – 20/12/1961) là một nhà toán học và thống kê học người Đan Mạch.

```

else {
  x <- x - fx / gx
  fx <- f(x)
  gx <- (f(x + fx) - fx) / fx
}
}
return(list(root = x, iter = i))
}

# Áp dụng cho phương trình  $x^4 - 3 * x + 1 = 0$ .

f <- function(x) x^4 - 3 * x + 1
steffensen_method(f, 2, 1e-6, 100)

## $root
## [1] 1,3074861
##
## $iter
## [1] 60

```

Kết quả sau 60 vòng lặp, bắt đầu với giá trị $x_0 = 2$ khớp với kết quả cho trong CAS 2.3.

Chú ý 2.6. Lưu ý, cũng giống như phương pháp Newton–Raphson, có thể xảy ra các tình huống quá trình lặp Steffensen không hội tụ, hoặc hội tụ không đúng nghiệm cần tìm. Ví dụ, nếu lấy xấp xỉ đầu $x_0 = 1$ thì quá trình lặp theo phương pháp Steffensen cho dãy lặp hội tụ tới nghiệm trong $(0, 1)$ chứ không phải nghiệm trong khoảng $(1, 2)$.

```

steffensen_method(f, 1, 1e-6, 100)

## $root
## [1] 0,337666766
##
## $iter
## [1] 4

```

2.5.3 Phương pháp vị trí sai (regula falsi)

Phương pháp vị trí sai (regula falsi) là một biến thể của phương pháp dây cung trong đó, thay vì lấy dây cung đi qua hai điểm (x_n, f_n) và (x_{n-1}, f_{n-1}) để xác định x_{n+1} , người ta lấy dây cung đi qua điểm (x_n, f_n) và điểm $(x_{n'}, f_{n'})$ với $n' < n$ là chỉ số lớn nhất bé hơn n sao cho $f(x_n)$ và $f(x_{n'})$ trái dấu. Khi đó, x_{n+1} là giao điểm của đường thẳng đi qua hai điểm (x_n, f_n) và $(x_{n'}, f_{n'})$.

Cách chọn này đảm bảo sự hội tụ của dãy điểm x_n tới nghiệm của $f(x) = 0$ với điều kiện ban đầu $f(x_0)f(x_1) < 0$.

CAS 2.10. Đoạn chương trình R sau đây minh hoạ phương pháp vị trí sai áp dụng cho phương trình $x^4 - 3x + 1 = 0$.

```
# Hàm regulaFalsi() tìm nghiệm theo phương pháp vị trí sai.

regulaFalsi <- function(f, a, b, tol = 1e-6, m = 1000) {
  i <- 0
  fa <- f(a)
  fb <- f(b)

  # Vòng lặp while tiếp tục lặp khi |b - a| > tol
  # và i < m đồng thời thoả mãn
  while (abs(b - a) > tol && i < m) {
    c <- (a * fb - b * fa) / (fb - fa)
    fc <- f(c)
    if (fc == 0) {
      # Đã tìm thấy nghiệm!
      break
    } else if (sign(fa) * sign(fc) < 0) {
      b <- c
      fb <- fc
    } else {
      a <- c
      fa <- fc
    }
    i <- i + 1
  }

  return(list(root = c, iter = i))
}

# Áp dụng regulaFalsi() cho  $x^4 - 3 * x + 1 = 0$ .

f <- function(x) x^4 - 3 * x + 1
regulaFalsi(f, 1, 2, 1e-6, 1000)

## $root
## [1] 1,3074861
##
## $iter
## [1] 76
```

Sau 76 bước lặp, `regulaFalsi()` cho kết quả khớp với những kết quả thu được từ các phương pháp khác.

Lưu ý, `regulaFalsi()` cũng được lập sẵn trong gói `pracma` của phần mềm R.

2.6 Câu hỏi và bài tập

1. Cho $\phi(x) = 2 - e^{-x}$. Chứng minh rằng dãy p_n xác định bởi công thức lặp $p_n = \phi(p_{n-1})$, $n = 1, 2, \dots$, hội tụ với mọi cách chọn $p_0 \in [1, 3]$.

2. Xét phương trình $f(x) := e^x - 5x^2 + 2022x - 10 = 0$.

(a) Chứng minh $[0, 1]$ là một khoảng tách nghiệm của nó.

(b) Đặt

$$\phi(x) = \frac{5x^2 - e^x + 10}{2022}.$$

Chứng minh rằng dãy số xác định bởi $x_0 = 0$, $x_{k+1} = \phi(x_k)$, $k = 0, 1, 2, \dots$ hội tụ tới nghiệm duy nhất của phương trình $f(x) = 0$.

(c) Tính x_1, x_2 . Coi x_2 là nghiệm gần đúng của phương trình trên. Đánh giá sai số tuyệt đối của x_2 .

3. Cho phương trình $x^3 - x - 1 = 0$.

(a) Bằng khảo sát hàm số $f(x) = x^3 - x - 1$ với $x \in [1, 2]$, chỉ ra rằng phương trình $x^3 - x - 1 = 0$ có nghiệm duy nhất trên đoạn $[1, 2]$.

(b) Biết phương trình đã cho tương đương với $x = \sqrt[3]{x+1}$. Sử dụng xấp xỉ ban đầu $x_0 = 1$ và công thức $x_n = \sqrt[3]{x_{n-1}+1}$, hãy tính x_k với $1 \leq k \leq 6$.

(c) Đặt $\varphi(x) = \sqrt[3]{x+1}$. Tìm một số $q < 1$ sao cho $|\varphi'(x)| \leq q$ với mọi $x \in (1, 2)$. Từ đó, đánh giá sai số tuyệt đối cho x_6 .

(d) Cần thực hiện bao nhiêu lần lặp để thu được nghiệm gần đúng với độ chính xác $\epsilon = 10^{-5}$? Tính nghiệm gần đúng đó.

4. Cho $\phi(x) = \pi + \frac{1}{2} \cos\left(\frac{x}{2}\right)$ và $f(x) = x - \phi(x)$.

(a) Chứng minh rằng $[0, 2\pi]$ là một khoảng tách nghiệm của phương trình $f(x) = 0$.

(b) Tìm một số dương $q < 1$ sao cho $|\phi'(x)| \leq q$ với mọi $x \in [0, 2\pi]$.

(c) Chứng minh rằng phương trình $x = \phi(x)$ có nghiệm duy nhất trên $[0, 2\pi]$.

(d) Xây dựng dãy lặp $x_0 = 2\pi$, $x_{k+1} := \phi(x_k)$, $k = 0, 1, 2, \dots$. Tính các giá trị x_1, x_2 . Đánh giá sai số $|x_2 - \alpha|$, với α là nghiệm duy nhất của phương trình $x = \phi(x)$.

5. Giả sử phương trình $x = \phi(x)$ có nghiệm α trong $[a, b]$, hàm ϕ khả vi trên $[a, b]$, $a \leq \phi(x) \leq b$ với mọi $x \in [a, b]$, và đạo hàm ϕ' thỏa mãn

$$|\phi'(x)| < q,$$

với q là một hằng số dương, $q < 1$.

- (a) Chứng minh rằng với mọi $x, y \in [a, b]$ ta có

$$|x - y| \leq \frac{1}{1 - q} (|x - \phi(x)| + |y - \phi(y)|).$$

- (b) Lấy $x_0 \in [a, b]$ tùy ý và xây dựng dãy lặp $x_1 = \phi(x_0)$, $x_2 = \phi(x_1)$, ..., $x_{k+1} = \phi(x_k)$. Chứng minh rằng với mọi $j, k \geq 1$ ta có

$$|x_k - x_j| \leq \frac{q^k + q^j}{1 - q} |x_0 - x_1|.$$

- (c) Chứng minh đánh giá tiên nghiệm sau:

$$|x_k - \alpha| \leq \frac{q^k}{1 - q} |x_0 - x_1|.$$

[Cho $j \rightarrow \infty$ cùng với nhận xét rằng $q^j \rightarrow 0$ vì $0 < q < 1$.]

6. Tính gần đúng nghiệm của phương trình $x + \ln(x) = 0$ bằng phương pháp lặp đơn.

7. Xét phương trình $20x^3 + x - 22 = 0$.

- (a) Chứng minh phương trình trên có nghiệm duy nhất α trong khoảng $(1, 2)$.
 (b) Tính gần đúng nghiệm α bằng phương pháp Newton với xấp xỉ ban đầu $x_0 = 2$ và với 3 bước lặp. Đánh giá sai số của nghiệm xấp xỉ.

8. Xét phương trình $f(x) := x^4 - 3x + 1 = 0$.

- (a) Chứng minh rằng các khoảng $(0, 1)$ và $(1, 2)$ là các khoảng tách nghiệm của phương trình đó.
 (b) Dùng phương pháp Newton xuất phát từ $x_0 = 0$ giải gần đúng phương trình với độ chính xác 10^{-2} .

9. Xét phương trình $3^x - 4x = 0$.

- (a) Tính gần đúng nghiệm dương bé nhất của phương trình này bằng phương pháp Newton và viết kết quả dưới dạng thập với 6 chữ số chắc. [Trả lời: $x \approx 0,379194$].

- (b) Biết $[1, 2]$ là một khoảng tách nghiệm của phương trình này. Tính gần đúng nghiệm duy nhất trong khoảng đó với độ chính xác 10^{-5} bằng phương pháp Newton–Raphson. [*Trả lời: $x \approx 1,79369$*].
- 10.** Dùng phương pháp Newton–Raphson tính gần đúng nghiệm của phương trình $x^2 - 2 \sin(x) - 0,5 = 0$ trong khoảng $[1, 2]$ với hai chữ số thập phân là chữ số chắc.
- 11.** Xét phương trình $x^3 + 20x - 22 = 0$.
- (a) Chứng minh phương trình trên có nghiệm duy nhất α trong khoảng $(1, \frac{3}{2})$.
- (b) Tính gần đúng nghiệm α bằng phương pháp Newton–Raphson với xấp xỉ ban đầu $x_0 = 3/2$ và với 3 bước lặp. Hãy đánh giá sai số của nghiệm gần đúng thu được.

Chương

3

Tính gần đúng nghiệm của một hệ phương trình tuyến tính

Trong chương này, chúng ta bắt đầu với lí thuyết hệ phương trình tuyến tính, nhắc lại một số kết quả cơ bản về hệ phương trình tuyến tính như định lí Kronecker–Capelli, công thức Cramer, phương pháp khử Gauss và Gauss-Jordan. Tiếp theo, chúng ta sẽ tiếp tục với phương pháp số để giải gần đúng, với trọng tâm là phương pháp lặp đơn và phương pháp lặp Seidel.

Nội dung

3.1	Hệ phương trình tuyến tính. Phương pháp Gauss	63
3.1.1	Định nghĩa. Sự tồn tại và duy nhất nghiệm	64
3.1.2	Phương pháp Gauss	65
3.2	Phương pháp lặp đơn	70
3.2.1	Sự hội tụ của phương pháp lặp đơn	71
3.3	Phương pháp lặp Seidel	81
3.3.1	Sự hội tụ của phương pháp lặp Seidel	82
3.3.2	Một ví dụ: Mô hình Leontief	86
3.4	Câu hỏi và bài tập	88

3.1 Hệ phương trình tuyến tính. Phương pháp Gauss

Mục này bắt đầu với việc nhắc lại một số khái niệm và kết quả cơ bản về hệ phương trình tuyến tính. Các nội dung này trùng lặp với một số nội dung

trong học phần Toán cao cấp hoặc Đại số tuyến tính. Tiếp theo, chúng ta nói về phương pháp khử Gauss và vấn đề chọn trụ tối đại. Cuối cùng, chúng ta giới thiệu một số ví dụ về giải hệ phương trình tuyến tính bằng máy tính. Tài liệu tham khảo: T. V. Đình [6, §3.2], P. K. Anh [2, §VII.2, §VII.3], Kreyszig [9, §20.1], Dương Ngọc Sơn và đồng tác giả [10].

3.1.1 Định nghĩa. Sự tồn tại và duy nhất nghiệm

Một phương trình tuyến tính của n ẩn số $x_1, x_2, \dots, x_n$ có dạng

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b, \quad (3.1)$$

trong đó $a_1, a_2, \dots, a_n$ và b là các số thực. Với $m \geq 1$ là một số nguyên, một hệ m phương trình tuyến tính của n ẩn số $x_1, x_2, \dots, x_n$ có dạng

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{cases}, \quad (3.2)$$

trong đó a_{ij} và b_i là các hằng số cho trước. Các số a_{ij} gọi là hệ số của x_j trong phương trình thứ i còn b_i gọi là số hạng tự do của phương trình thứ i . Các hệ phương trình tuyến tính có vẻ đơn giản nhưng lại có rất nhiều ứng dụng và xuất hiện nhiều trong khoa học kỹ thuật, kinh tế, sinh học, khoa học xã hội, v.v.

Rõ ràng, hệ phương trình (3.2) được xác định nếu chúng ta biết các hệ số và số hạng tự do của nó. Chúng có thể được sắp xếp thành hai ma trận như sau:

$$A := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \bar{A} := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}. \quad (3.3)$$

Ma trận A được gọi là *ma trận hệ số*, còn ma trận $\bar{A}$ gọi là *ma trận bổ sung* của hệ của phương trình (3.2). Từ ma trận bổ sung $\bar{A}$, ta có thể xây dựng lại hệ (3.2) một cách dễ dàng.

Nghiệm của hệ (3.2) là một bộ có thứ tự $(x_1, x_2, \dots, x_n)$ (còn gọi là một vectơ trong $\mathbb{R}^n$) thỏa mãn tất cả m phương trình đó. Giải hệ phương trình (3.2) là tìm hoặc mô tả một cách “tường minh” tất cả các nghiệm của nó. Việc giải các hệ phương trình tuyến tính thường bao gồm các bước biến đổi hệ phương trình ban đầu thành các phương trình có cùng tập nghiệm, gọi là *hệ phương trình tương đương*. Các phép biến đổi giữa các hệ phương trình tương đương gọi là *biến đổi tương đương*.

Đối với một hệ phương trình tuyến tính tổng quát, có thể xảy ra một trong các tình huống sau: hệ không có nghiệm (vô nghiệm), hệ có nghiệm duy nhất, và hệ có vô số nghiệm.

Định lí 3.1: Kronecker–Capelli

Hệ phương trình (3.2) có nghiệm nếu và chỉ nếu

$$r(A) = r(\bar{A}), \quad (3.4)$$

trong đó, $r(A)$ và $r(\bar{A})$ lần lượt là hạng của ma trận hệ số A và ma trận bổ sung $\bar{A}$. Hơn nữa, hệ có nghiệm duy nhất nếu và chỉ nếu (3.4) xảy ra và hạng của A bằng số ẩn số của hệ.

Chứng minh của định lí Kronecker–Capelli bạn đọc có thể xem trong [10, Chương 1].

Khi hệ có nghiệm duy nhất thì số phương trình không ít hơn số ẩn số và hệ có thể đưa về dạng

$$Ax = b \quad (3.5)$$

với A là một ma trận vuông khả nghịch, $x = (x_1, \dots, x_n)^T$ và $b = (b_1, \dots, b_n)^T$ là các vectơ cột (n là số các ẩn số). Ở vế trái Ax là phép nhân một ma trận cỡ $n \times n$ với một vectơ cột, xem như một ma trận cỡ $n \times 1$. Hơn nữa, ta có định lí Cramer sau¹.

Định lí 3.2: Cramer

Xét hệ phương trình (3.5) với A là một ma trận vuông. Nếu $\det(A) \neq 0$ thì hệ phương trình có nghiệm duy nhất

$$x_i = \frac{\det(A \leftarrow_i b)}{\det(A)}.$$

Trong đó, $A \leftarrow_i b$ là ma trận thu được bằng cách thay cột thứ i của A bởi b .

Chứng minh. Xem [9, §7.7] hoặc [10, Chương 1]. □

3.1.2 Phương pháp Gauss

Trong mục này, chúng ta nói về hệ dạng bậc thang và phương pháp khử Gauss (Gauss elimination method). Tài liệu tham khảo: Kreyszig [9, §20.1].

¹Gabriel Cramer (1704 – 1752) là một nhà toán học người Thụy Sĩ.

Hệ tam giác, hệ bậc thang. Hệ phương trình tuyến tính dạng tam giác là hệ có dạng tổng quát (3.2) với $m = n$ (số ẩn bằng số phương trình) và, với mỗi $1 \leq j < i \leq n$, hệ số a_{ij} bằng 0, còn với mỗi k , hệ số $a_{kk} \neq 0$. Khi ấy, ma trận hệ số của hệ tam giác có dạng

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Trong trường hợp này, phương trình thứ n có dạng

$$a_{nn}x_n = b_n,$$

nên có thể giải ra được

$$x_n = \frac{b_n}{a_{nn}} \quad (a_{nn} \neq 0).$$

Nếu $n = 1$, hệ có 1 ẩn, thì ta đã giải xong. Nếu $n \geq 2$, ta thay x_n vào phương trình thứ $n - 1$, ta giải được

$$x_{n-1} = \frac{1}{a_{n-1,n-1}} \left(b_{n-1} - a_{n-1,n} \cdot \frac{b_n}{a_{nn}} \right).$$

Nếu $n = 2$, thì phương trình có hai ẩn số x_1 và x_2 và đã được giải.

Nếu $n \geq 3$, tiếp tục thay x_{n-1} và x_n vào phương trình thứ $n - 2$, ta giải được x_{n-2} . Quá trình này tiếp tục cho đến khi ta giải được x_1 .

Ví dụ 3.1. Giải hệ phương trình sau đây.

$$\begin{cases} x_1 + x_2 - x_3 = 2 \\ 2x_2 - x_3 = 2 \\ 3x_3 = 5. \end{cases}$$

Lời giải

Từ phương trình thứ 3, ta có $x_3 = \frac{5}{3}$. Thay vào phương trình thứ hai, ta có

$$2x_2 - \frac{5}{3} = 2,$$

giải được $x_2 = \frac{11}{6}$. Thay $x_2 = \frac{11}{6}$ và $x_3 = \frac{5}{3}$ vào phương trình thứ nhất

$$x_1 + \frac{11}{6} - \frac{5}{3} = 2.$$

Giải phương trình cuối, ta có $x_1 = \frac{11}{6}$. Vậy, hệ có nghiệm duy nhất

$$(x_1, x_2, x_3) = \left(\frac{11}{6}, \frac{11}{6}, \frac{5}{3} \right).$$

Phương pháp Gauss. Trên đây, ta đã thấy rằng các hệ phương trình dạng tam giác có thể được giải một cách dễ dàng từ dưới lên. Điều này là cơ sở của phương pháp Gauss² dựa trên các biến đổi sơ cấp.

Định nghĩa 3.1:

Đối với một hệ phương trình tuyến tính (3.2), các phép biến đổi sau đây gọi là các phép biến đổi sơ cấp:

1. Đổi chỗ hai phương trình của hệ,
2. Nhân hai vế của một phương trình với một số khác không,
3. Lấy hai vế của một phương trình trong hệ nhân với cùng một số khác không rồi cộng vào hai vế tương ứng của một phương trình khác.

Ý tưởng cơ bản của phương pháp Gauss là dùng các phép biến đổi sơ cấp trên các phương trình để đưa hệ về dạng tam giác trên. Quá trình này gọi là *quá trình thuận*. Từ hệ dạng tam giác trên, ta giải nghiệm như trong mục trước, ngược từ dưới lên, gọi là *quá trình ngược*.

Xét hệ phương trình

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1, \quad (3.6)$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2, \quad (3.7)$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3. \quad (3.8)$$

Các hệ số của x_1 , tức là a_{11} , a_{12} , và a_{13} , không đồng thời bằng không, vì nếu tất cả chúng đều bằng 0 thì hệ phương trình không chứa x_1 và vậy thì nó tương đương với hệ 3 phương trình của 2 ẩn x_2 và x_3 . Bằng cách đổi thứ tự các phương trình, ta có thể giả sử $a_{11} \neq 0$. Khi đó, bằng cách chia cả hai vế cho a_{11} , phương trình thứ nhất (3.6) được đưa về dạng

$$x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 = \frac{b_1}{a_{11}}. \quad (3.9)$$

Bằng cách đặt

$$a_{12}^{(1)} = \frac{a_{12}}{a_{11}}, \quad a_{13}^{(1)} = \frac{a_{13}}{a_{11}}, \quad b_1^{(1)} = \frac{b_1}{a_{11}},$$

ta viết phương trình trên là

$$x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 = b_1^{(1)}. \quad (3.10)$$

²Johann Carl Friedrich Gauß (1777 – 1855) là một nhà toán học người Đức.

Ta sẽ dùng phương trình (3.10) này để “khử” x_1 trong các phương trình thứ hai và thứ ba. Nhân (3.10) với $-a_{21}$ rồi cộng vào phương trình thứ hai, về theo về, ta thấy hệ số của x_1 trong phương trình thu được triệt tiêu. Cụ thể là

$$a_{22}^{(1)} x_2 + a_{23}^{(1)} x_3 = b_2^{(1)} \quad (3.11)$$

với

$$a_{22}^{(1)} = a_{22} - a_{12}^{(1)} a_{21}, \quad a_{23}^{(1)} = a_{23} - a_{13}^{(1)} a_{21}, \quad b_2^{(1)} = b_2 - a_{21} b_1^{(1)}.$$

Để khử x_1 ở phương trình thứ ba (3.8), ta nhân hai vế của (3.10) với $-a_{31}$ rồi cộng về theo về với (3.8), ta thấy hệ số của x_1 trong phương trình thu được triệt tiêu. Nó có dạng

$$a_{32}^{(1)} x_2 + a_{33}^{(1)} x_3 = b_3^{(1)} \quad (3.12)$$

với

$$a_{32}^{(1)} = a_{32} - a_{12}^{(1)} a_{31}, \quad a_{33}^{(1)} = a_{33} - a_{13}^{(1)} a_{31}, \quad b_3^{(1)} = b_3 - a_{31} b_1^{(1)}.$$

Đối với hệ hai phương trình của hai ẩn (3.11) và (3.12), ta có thể áp dụng quy trình tương tự để khử x_2 từ (3.12). Giả sử $a_{22}^{(1)} \neq 0$. Chia hai vế của (3.11) cho $a_{22}^{(1)}$, ta đưa nó về dạng

$$x_2 + a_{33}^{(2)} x_3 = b_3^{(2)}, \quad (3.13)$$

với

$$a_{33}^{(2)} = \frac{a_{33}^{(1)}}{a_{22}^{(1)}}, \quad b_3^{(2)} = \frac{b_3^{(1)}}{a_{22}^{(1)}}. \quad (3.14)$$

Ta dùng (3.13) để khử x_2 từ phương trình (3.12). Nhân hai vế của (3.13) với $-a_{32}^{(1)}$ rồi cộng phương trình thu được, về theo về, với (3.12), ta thu được phương trình không chứa x_2 có dạng

$$a_{33}^{(2)} x_3 = b_3^{(2)}, \quad (3.15)$$

với

$$a_{33}^{(2)} = a_{33}^{(1)} - a_{23}^{(1)} a_{32}^{(1)}, \quad b_3^{(2)} = b_3^{(1)} - b_2^{(1)} a_{32}^{(1)}.$$

Bắt buộc $a_{33}^{(2)} \neq 0$. Chia hai vế của (3.15) cho $a_{33}^{(2)}$ ta thu được

$$x_3 = b_3^{(3)},$$

với

$$b_3^{(3)} = \frac{b_3^{(2)}}{a_{33}^{(2)}}.$$

Tổng hợp các phương trình thu được, ta có hệ sau

$$\begin{aligned}x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 &= b_1^{(1)}, \\x_2 + a_{23}^{(2)}x_3 &= b_2^{(2)}, \\x_3 &= b_3^{(3)}.\end{aligned}$$

Đây là hệ phương trình dạng tam giác ta đã biết cách giải.

Chọn phần tử trụ tối đại. Các tính toán trong phương pháp khử Gauss nêu trên có thể sinh ra sai số lớn khi thực hiện phép chia cho các số có trị tuyệt đối nhỏ. Để giảm sai số, ta thực hiện khử Gauss với việc chọn phần tử trụ có trị tuyệt đối lớn nhất, gọi là phương pháp “chọn phần tử trụ tối đại” (Gauss elimination with partial pivoting). Quá trình khử Gauss cùng với việc chọn trụ tối đại được minh họa trong ví dụ sau:

Ví dụ 3.2. Giải hệ phương trình sau bằng phương pháp khử Gauss với trụ tối đại.

$$\begin{cases} 12x_2 + x_3 = 7 \\ 3x_1 + 3x_2 + x_3 = -8 \\ 10x_1 + x_2 - 2x_3 = -9 \end{cases}$$

Trong ví dụ này, hệ số của x_1 trong phương trình đầu tiên bằng 0 nên chúng ta sẽ đổi chỗ phương trình thứ nhất. Do hệ số của x_1 lớn nhất ở phương trình thứ 3 nên chúng ta chọn nó là phần tử trụ và sắp xếp lại như sau:

$$\begin{cases} 10x_1 + x_2 - 2x_3 = -9 \\ 3x_1 + 3x_2 + x_3 = -8 \\ 12x_2 + x_3 = 7. \end{cases}$$

Tiếp theo, dùng phương trình thứ nhất để khử x_1 trong phương trình thứ hai để thu được

$$\begin{cases} 10x_1 + x_2 - 2x_3 = -9 \\ 2,7x_2 + 1,6x_3 = -5,3 \\ 12x_2 + x_3 = 7. \end{cases}$$

Do hệ số của x_2 trong phương trình thứ 3 lớn hơn nên đổi chỗ phương trình thứ hai và thứ 3 để thu được

$$\begin{cases} 10x_1 + x_2 - 2x_3 = -9 \\ 12x_2 + x_3 = 7 \\ 2,7x_2 + 1,6x_3 = -5,3 \end{cases}$$

Dùng phương trình thứ hai để khử x_2 trong phương trình thứ 3, ta thu được

$$\begin{cases} 10x_1 + x_2 - 2x_3 = -9 \\ 12x_2 + x_3 = 7 \\ 1,375x_3 = 6,875 \end{cases}$$

Từ hệ phương trình cuối cùng, chúng ta giải ngược để thu được

$$x_3 = -5, x_2 = 1, x_1 = -2.$$

CAS 3.1. Phương pháp khử Gauss có sẵn trong thư viện `matlib`³ của R. Hãy thử với hệ phương trình sau:

$$\begin{cases} 4x_1 + 5x_2 - 2x_3 = -13 \\ 3x_1 + x_2 + 6x_3 = 25 \\ -3x_1 + x_2 + x_3 = 12 \end{cases} \quad (3.16)$$

3.2 Phương pháp lặp đơn

Trong mục này, chúng ta tìm hiểu phương pháp giải gần đúng một hệ phương trình tuyến tính bằng quá trình lặp. Tài liệu tham khảo: T. V. Đĩnh [6, §3.3], P. K. Anh [2, §VII.6], Kreyszig [9, §20.3].

Xét phương trình

$$Ax = b.$$

Cũng như phương pháp lặp điểm bất động cho phương trình một ẩn, ta bắt đầu bằng việc đưa phương trình về dạng

$$x = Bx + c, \quad (3.17)$$

trong đó, ma trận hệ số B và vectơ c được suy ra từ A và b . Từ đó, ta xây dựng dãy lặp

$$x^{(k+1)} = Bx^{(k)} + c, \quad (3.18)$$

$$x^{(0)} \text{ cho trước.} \quad (3.19)$$

Khi một số điều kiện được thỏa mãn, ta có thể hi vọng sau một số bước lặp, vectơ $x^{(k)}$ là một xấp xỉ của nghiệm của hệ ban đầu. Phần tiếp theo, ta sẽ nói về những điều kiện của B để bước lặp trên cho dãy các xấp xỉ của nghiệm.

³matlib: Matrix Functions for Teaching and Learning Linear Algebra and Multivariate Statistics, DOI: <https://doi.org/10.32614/CRAN.package.matlib>

3.2.1 Sự hội tụ của phương pháp lập đơn

Để có thể nói về tính hội tụ của các phương pháp lập cho hệ phương trình cho ngắn gọn, chúng ta đưa ra định nghĩa sau.

Định nghĩa 3.2:

Giả sử $\alpha = (\alpha_1, \dots, \alpha_n)$ là một nghiệm của hệ phương trình. Nếu $x_j^{(k)} \rightarrow \alpha_j$, $j = 1, 2, \dots, n$, khi $k \rightarrow \infty$ thì ta nói phương pháp lập (3.18), (3.19) hội tụ.

Chuẩn của vectơ

Để dễ dàng mô tả sự hội tụ của quá trình lập, chúng ta mô tả sự hội tụ của dãy các vectơ bằng sự hội tụ của dãy số thông qua khái niệm *chuẩn* (norm) của vectơ.

Đối với một vectơ

$$x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

chúng ta định nghĩa các chuẩn sau:

$$\|x\|_\infty := \max_{1 \leq j \leq n} |x_j|, \quad (3.20)$$

$$\|x\|_1 := |x_1| + |x_2| + \dots + |x_n|, \quad (3.21)$$

$$\|x\|_2 := \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2}. \quad (3.22)$$

Chúng được gọi là, theo thứ tự, là chuẩn ℓ_∞ , ℓ_1 , và ℓ_2 ⁴ Chúng là các trường hợp riêng của chuẩn ℓ_p ⁵) trên không gian các dãy số, hữu hạn hoặc vô hạn. Chuẩn ℓ_2 còn được gọi là chuẩn Euclidean trên $\mathbb{R}^n$ và nó được sinh ra bởi một tích vô hướng.

Ví dụ 3.3. Xác định chuẩn ℓ_∞ , ℓ_1 , và ℓ_2 của vectơ $x = (1, 2, -3) \in \mathbb{R}^3$.

Lời giải

Tính toán trực tiếp, ta có

$$\|x\|_\infty = \max\{|1|, |2|, |-3|\} = 3,$$

$$\|x\|_1 = |1| + |2| + |-3| = 6,$$

và

$$\|x\|_2 = \sqrt{|1|^2 + |2|^2 + |-3|^2} = \sqrt{14}.$$

⁴T. V. Đỉnh [6] gọi chuẩn $\|\cdot\|_\infty$ là $\|\cdot\|_0$.

⁵Chữ cái ℓ lấy trong tên của Henri Léon Lebesgue (28/6/1875 – 26/7/1941). Ông là một nhà toán học người Pháp

Sử dụng một trong các chuẩn trên, ta đưa điều kiện hội tụ trong định nghĩa 3.2 về điều kiện hội tụ của một dãy số không âm.

Định lí 3.3:

Phương pháp lặp (3.18), (3.19) hội tụ khi và chỉ khi

$$\|x^{(k)} - \alpha\|_p \rightarrow 0, \quad k \rightarrow \infty$$

với $p = \infty, 1$, hoặc 2.

Chuẩn của ma trận

Đối với ma trận cấp $m \times n$, $B \in \text{Mat}(m \times n; \mathbb{R})^6$, ta cũng định nghĩa các chuẩn của B như sau:

$$\|B\|_\infty := \max_{1 \leq i \leq m} \sum_{j=1}^n |b_{ij}|, \quad (3.23)$$

$$\|B\|_1 := \max_{1 \leq j \leq n} \sum_{i=1}^m |b_{ij}|. \quad (3.24)$$

Để tính $\|B\|_\infty$, chúng ta tính tổng các giá trị tuyệt đối các yếu tố trên mỗi hàng. Có m kết quả như vậy, tương ứng với m hàng của B . Khi đó, $\|B\|_\infty$ chính là kết quả lớn nhất trong m kết quả ấy. Vì vậy, $\|B\|_\infty$ đôi khi gọi là chuẩn hàng (cho dễ nhớ).

Chuẩn $\|B\|_1$ được tính tương tự như $\|B\|_\infty$, chỉ thay hàng bởi cột:

$$\|B\|_1 = \|B^T\|_\infty,$$

trong đó, B^T là phép chuyển vị của B . Khi đó, $\|\cdot\|_p$, với $p = \infty$ hoặc $p = 1$, thỏa mãn các tính chất của một chuẩn trong không gian vectơ.

- $\|B\|_p \geq 0$, $\|B\|_p = 0$ nếu và chỉ nếu B là ma trận 0.
- Với mọi số thực $r \in \mathbb{R}$ và ma trận B ,

$$\|rB\|_p = |r| \cdot \|B\|_p.$$

- (bất đẳng thức tam giác) Với mọi ma trận vuông B và C cùng cấp,

$$\|B + C\|_p \leq \|B\|_p + \|C\|_p.$$

⁶Nhắc lại rằng $\text{Mat}(m \times n; \mathbb{R})$ kí hiệu tập các ma trận cỡ $m \times n$.

Ví dụ 3.4. Xác định các chuẩn $\|B\|_\infty$ và $\|B\|_1$, với B là ma trận sau đây:

$$B = \begin{bmatrix} 1 & 3 & -5 \\ -2 & 4 & 9 \\ -1 & 1 & 3 \end{bmatrix}. \quad (3.25)$$

Lời giải

Để tính $\|B\|_\infty$, ta tính tổng các giá trị tuyệt đối của các yếu tố dọc theo các hàng của B . Ở hàng thứ nhất, ta có

$$\sum_{j=1}^3 b_{1j} = |1| + |3| + |-5| = 9.$$

Tính toán tương tự ở hàng thứ hai và hàng thứ ba, ta thu được

$$\sum_{j=1}^3 b_{2j} = 15, \quad \sum_{j=1}^3 b_{3j} = 5.$$

Vậy,

$$\|B\|_\infty = \max\{9, 15, 5\} = 15.$$

Để tính $\|B\|_1$, ta tính tổng các giá trị tuyệt đối của các yếu tố dọc theo các cột của B . Cụ thể như sau: Ở cột thứ nhất, ta có

$$\sum_{i=1}^3 b_{i1} = |1| + |-2| + |-1| = 4.$$

Tương tự, ta tính toán ở các cột thứ hai và thứ ba,

$$\sum_{i=1}^3 b_{i2} = 8, \quad \text{và} \quad \sum_{i=1}^3 b_{i3} = 17.$$

Vậy

$$\|B\|_1 = \max\{4, 8, 17\} = 17.$$

Chú ý 3.1. Nếu đồng nhất các vectơ với một ma trận một cột thì chuẩn của vectơ có thể xem là trường hợp riêng của chuẩn ma trận.

CAS 3.2. Trong R, chuẩn của ma trận được tính bằng `norm()`. Đoạn code mẫu sau đây tính chuẩn 1 và chuẩn vô hạn của ma trận trong ví dụ trên.

```
# Đưa vào ma trận B gồm 9 yếu tố xếp làm 3 cột.
B <- matrix(c(1, -2, -1, 3, 4, 1, -5, 9, 3), ncol = 3)

# In ra B (để kiểm tra)
print(B)

##      [,1] [,2] [,3]
## [1,]    1    3   -5
## [2,]   -2    4    9
## [3,]   -1    1    3

# In ra giá trị chuẩn 1 của B
print(norm(B, type = "O")) # Chuẩn 1 (one norm)

## [1] 17

# In ra giá trị chuẩn vô cùng của B
print(norm(B, type = "I")) # Chuẩn vô hạn (infinity norm)

## [1] 15
```

CAS 3.3. Để tính chuẩn của một vectơ, chúng ta biến nó thành một ma trận 1 cột bằng hàm `as.matrix` (lấy từ Tiếng Anh “as matrix”) và tính chuẩn ma trận của nó.

```
x <- c(1, -2, 3, 5)

# Hàm as.matrix chuyển một vectơ thành một ma trận một cột
A <- as.matrix(x)

# Tính các chuẩn của A
print(norm(A, type = "I"))

## [1] 5

print(norm(A, type = "O"))

## [1] 11
```

Một tính chất của chuẩn của ma trận $\|B\|_p$ là tính “tương thích” của chúng với chuẩn của vectơ, theo nghĩa

Mệnh đề 3.1:

Giả sử $p = \infty$ hoặc $p = 1$. Khi đó,

$$\|Bx\|_p \leq \|B\|_p \cdot \|x\|_p, \quad B \in \text{Mat}(m \times n; \mathbb{R}), x \in \mathbb{R}^n. \quad (3.26)$$

Hơn nữa, với mỗi ma trận B , $\|B\|_p$ là số dương bé nhất sao cho (3.26) đúng với mọi x , có nghĩa là

$$\|B\|_p = \min \{ C > 0 \mid \|Bx\|_p \leq C \cdot \|x\|_p \ \forall x \in \mathbb{R}^n \}. \quad (3.27)$$

Chứng minh. Chúng ta chứng minh cho trường hợp $p = 1$.

Với $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ tùy ý, gọi $Bx = y = (y_1, \dots, y_m) \in \mathbb{R}^m$, thì

$$y_k = \sum_{j=1}^n b_{kj} x_j.$$

Khi đó, dùng bất đẳng thức giá trị tuyệt đối của tổng nhỏ hơn hoặc bằng tổng các giá trị tuyệt đối,

$$\begin{aligned} \|Bx\|_1 = \|y\|_1 &= \sum_{k=1}^m |y_k| = \sum_{k=1}^m \left| \sum_{j=1}^n b_{kj} x_j \right| \\ &\leq \sum_{k=1}^m \sum_{j=1}^n |b_{kj}| \cdot |x_j| = \sum_{j=1}^n |x_j| \left(\sum_{k=1}^m |b_{kj}| \right). \end{aligned} \quad (3.28)$$

Theo định nghĩa, với mọi $j \in \{1, 2, \dots, n\}$, ta có

$$\sum_{k=1}^m |b_{kj}| \leq \max_{1 \leq \ell \leq n} \sum_{k=1}^m |b_{k\ell}| = \|B\|_1.$$

Thay vào bất đẳng thức (3.28) ở trên, ta thu được

$$\|Bx\|_1 \leq \sum_{j=1}^n |x_j| \left(\sum_{k=1}^m |b_{kj}| \right) \leq \|B\|_1 \sum_{j=1}^n |x_j| = \|B\|_1 \|x\|_1.$$

Đó chính là (3.26) trong trường hợp $p = 1$. Nói cách khác, $\|B\|_1$ thuộc vào tập hợp trong vế phải của (3.27), và vậy thì, với $p = 1$,

$$\|B\|_p \geq \min \{ C > 0 \mid \|Bx\|_p \leq C \cdot \|x\|_p \ \forall x \in \mathbb{R}^n \}. \quad (3.29)$$

Để chứng minh bất đẳng thức ngược lại, trong trường hợp $p = 1$, ta giả sử ℓ_0 là chỉ số thỏa mãn

$$\sum_{k=1}^m |b_{k\ell_0}| = \|B\|_1$$

Khi đó, lấy $x_0 = (0, \dots, 1, \dots, 0)$, với số 1 xuất hiện ở thành phần thứ ℓ_0 . Khi đó,

$$\|Bx_0\|_1 = \sum_{k=1}^m \left| \sum_{j=1}^n b_{kj} x_j \right| = \sum_{k=1}^m |b_{k\ell_0}| = \|B\|_1 = \|B\|_1 \|x_0\|_1.$$

Như vậy, nếu C tùy ý thuộc vào tập hợp trong vế phải của (3.27) thì

$$C = C \|x_0\|_1 \geq \|Bx_0\|_1 = \|B\|_1.$$

Vậy thì

$$\|B\|_1 \leq \min \{ C > 0 \mid \|Bx\|_p \leq C \cdot \|x\|_p \ \forall x \in \mathbb{R}^n \}. \quad (3.30)$$

Từ các bất đẳng thức (3.29) và (3.30), ta thu được (3.27).

Chứng minh trong trường hợp $p = \infty$ dành cho bạn đọc. $\square$

Chú ý 3.2. Một chuẩn ma trận nữa thường dùng là chuẩn 2, được định nghĩa dựa trên (3.26) như sau:

$$\|B\|_2 := \sup \{ \|Bx\|_2 : x \in \mathbb{R}^n, \|x\|_2 = 1 \}. \quad (3.31)$$

Công thức tính $\|B\|_2$ khá phức tạp, và ta sẽ không nêu cụ thể ở đây. Tuy nhiên, nếu $B = [b_{jk}]$ thì có thể chứng minh được

$$\|B\|_2 \leq \left(\sum_{j,k=1}^n |b_{jk}|^2 \right)^{\frac{1}{2}}. \quad (3.32)$$

Vế phải của (3.32) thường được gọi là *chuẩn Frobenius* của B .

Sau đây là một định lí cho điều kiện đủ để dãy lặp trong phương pháp lặp đơn là hội tụ.

Định lí 3.4:

Giả sử $p = \infty$, $p = 1$, hoặc $p = 2$. Nếu $\|B\|_p < 1$ thì phương trình (3.17) có nghiệm duy nhất, kí hiệu α . Hơn nữa, với mọi $x^{(0)} \in \mathbb{R}^n$ phương pháp lặp (3.18), (3.19) hội tụ tới α . Hơn nữa, ta có các đánh giá sai số như sau:

(i) Đánh giá hậu nghiệm:

$$\|x^{(k)} - \alpha\|_p \leq \frac{\|B\|_p}{1 - \|B\|_p} \|x^{(k)} - x^{(k-1)}\|_p.$$

(ii) Đánh giá tiên nghiệm:

$$\|x^{(k)} - \alpha\|_p \leq \frac{\|B\|_p^k}{1 - \|B\|_p} \|x^{(1)} - x^{(0)}\|_p.$$

Chứng minh. Giả sử α là một nghiệm của hệ: $\alpha = B\alpha + c$. Khi đó,

$$\begin{aligned} \|x^{(m+1)} - \alpha\|_p &= \|(Bx^{(m)} + c) - (B\alpha + c)\|_p \\ &= \|Bx^{(m)} - B\alpha\|_p \\ &= \|B(x^{(m)} - \alpha)\|_p \\ &\leq \|B\|_p \|x^{(m)} - \alpha\|_p. \end{aligned} \quad (3.33)$$

Từ đây, bằng quy nạp, ta chứng minh

$$\|x^{(k)} - \alpha\|_p \leq \|B\|_p^k \|x^{(0)} - \alpha\|_p. \quad (3.34)$$

Thật vậy, rõ ràng (3.34) đúng với $k = 0$. Giả sử nó đúng với $k = m$, tức là

$$\|x^{(m)} - \alpha\|_p \leq \|B\|_p^m \|x^{(0)} - \alpha\|_p. \quad (3.35)$$

Ta cần chứng minh nó đúng với $k = m + 1$. Áp dụng (3.33)

$$\begin{aligned} \|x^{(m+1)} - \alpha\|_p &\leq \|B\|_p \|x^{(m)} - \alpha\|_p \\ &\leq \|B\|_p (\|B\|_p^m \|x^{(0)} - \alpha\|_p) \quad (\text{giả thuyết qui nạp}) \\ &= \|B\|_p^{m+1} \|x^{(0)} - \alpha\|_p. \end{aligned}$$

Vậy, theo nguyên lý qui nạp, ta kết luận (3.34) đúng với mọi $k = 1, 2, \dots$

Cho $k \rightarrow \infty$ trong (3.34), chú ý rằng, $\|B\|_p^k \rightarrow 0$ vì $\|B\|_p < 1$, nên theo nguyên lý giới hạn bị kẹp ta thu được

$$\|x^{(k)} - \alpha\|_p \rightarrow 0 \quad \text{ khi } k \rightarrow \infty.$$

Theo Hệ quả 3.3, dãy các vectơ $x^{(k)}$ hội tụ tới α trong $\mathbb{R}^n$. Đây chính là kết luận đầu tiên của định lý.

Để chứng minh các đánh giá sai số, ta áp dụng (3.33) cho $k = m + 1$ và bất đẳng thức tam giác như sau.

$$\begin{aligned} \|x^{(k)} - \alpha\|_p &\leq \|B\|_p \|x^{(k-1)} - \alpha\|_p \\ &= \|B\|_p \|x^{(k-1)} - x^{(k)} + x^{(k)} - \alpha\|_p \\ &\leq \|B\|_p (\|x^{(k-1)} - x^{(k)}\|_p + \|x^{(k)} - \alpha\|_p). \end{aligned}$$

Từ đây, ta dễ dàng suy ra

$$(1 - \|B\|_p) \|x^{(k)} - \alpha\|_p \leq \|B\|_p (\|x^{(k-1)} - x^{(k)}\|_p).$$

Chia hai vế cho $(1 - \|B\|_p) > 0$, ta thu được đánh giá sai số hậu nghiệm.

Để chứng minh đánh giá tiên nghiệm, ta chú ý rằng với $k \geq 2$,

$$\begin{aligned}\|x^{(k)} - x^{(k-1)}\|_p &= \|(Bx^{(k-1)} + c) - (Bx^{(k-2)} + c)\|_p \\ &= \|Bx^{(k-1)} - Bx^{(k-2)}\|_p \\ &= \|B(x^{(k-1)} - x^{(k-2)})\|_p \\ &\leq \|B\|_p \|x^{(k-1)} - x^{(k-2)}\|_p.\end{aligned}$$

Từ đó, lập luận theo quy nạp như trên, ta thu được

$$\|x^{(k)} - x^{(k-1)}\|_p \leq \|B\|_p^{k-1} \|x^{(1)} - x^{(0)}\|_p.$$

Kết hợp bất đẳng thức này với đánh giá hậu nghiệm, ta thu được đánh giá tiên nghiệm.

Cuối cùng, ta chứng minh sự tồn tại và duy nhất của nghiệm của (3.17). Theo công thức nghiệm Cramer, ta chỉ cần chứng minh $I - B$ là ma trận không suy biến (I là ma trận đơn vị cấp n). Theo chứng minh ở trên, vì giới hạn của một dãy (nếu có) là duy nhất nên nghiệm α của phương trình (3.17), nếu tồn tại, là duy nhất. Đặc biệt, với $c = 0$, hệ phương trình (3.17) được đưa về dạng $x = Bx$ có nghiệm $x = 0$ và nghiệm này là duy nhất. Từ đó, ta suy ra $I - B$ phải là ma trận không suy biến, và do đó, phương trình (3.17) có nghiệm. $\square$

Ví dụ 3.5. Đưa phương trình sau đây về dạng $x = Bx + c$ sao cho B là một ma trận có chuẩn ℓ_∞ bé hơn 1, $\|B\|_\infty < 1$.

$$\begin{cases} 3x_1 - x_2 + x_3 = 1 \\ 3x_1 + 6x_2 + 2x_3 = 0 \\ 3x_1 + 3x_2 + 7x_3 = 4. \end{cases} \quad (3.36)$$

Lời giải

Từ phương trình thứ nhất, ta giải ra được

$$x_1 = \frac{1}{3}x_2 - \frac{1}{3}x_3 + \frac{1}{3}.$$

Giải các phương trình thứ hai và thứ 3, ta cũng thu được

$$x_2 = -\frac{1}{2}x_1 - \frac{1}{3}x_3$$

và

$$x_3 = -\frac{3}{7}x_1 - \frac{3}{7}x_2 + \frac{4}{7}.$$

Kết hợp 3 phương trình trên, chúng ta thu được phương trình dạng $x = Bx + c$, như sau:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 0 & -\frac{1}{3} \\ -\frac{3}{7} & -\frac{3}{7} & 0 \end{bmatrix}}_B \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{4}{7} \end{bmatrix}}_c.$$

Từ đây, dễ dàng tính được chuẩn của $\|B\|_\infty = \frac{6}{7}$.

Nhận xét 3.1. Trong Ví dụ 3.5, ta đã lần lượt giải x_i từ phương trình thứ i , với điều kiện $a_{ii} \neq 0$ để đưa phương trình về dạng $x = Bx + c$. Phương pháp này luôn cho ta ma trận B có các phần tử trên đường chéo chính bằng 0. Ma trận B thu được bằng phương pháp trên có chuẩn bé hơn 1 nếu ma trận A trong hệ ban đầu thỏa mãn điều kiện *đường chéo trội* (diagonally dominant condition). Nếu A không thỏa mãn điều kiện đường chéo trội, thì ma trận B thu được có chuẩn lớn hơn hoặc bằng 1 và có thể xảy ra tình huống phương pháp lặp không hội tụ.

Một chú ý khác là trong phương pháp lặp đơn, chúng ta không nhất thiết giải x_i từ phương trình thứ i như trong phương pháp trên. Điều này được thể hiện trong thí dụ sau.

Ví dụ 3.6. Dùng phương pháp lặp đơn với ba bước lặp để giải hệ phương trình

$$\begin{cases} 1,02x_1 - 0,05x_2 - 0,10x_3 = 1,20 \\ -0,11x_1 + 1,03x_2 - 0,05x_3 = 2,25 \\ -0,11x_1 - 0,12x_2 + 1,04x_3 = 3,75 \end{cases} \quad (3.37)$$

Đánh giá các sai số bằng các đánh giá tiên nghiệm và hậu nghiệm.

Lời giải

Đưa hệ trên về dạng $x = Bx + c$ như sau. Từ phương trình thứ nhất, ta giải được

$$x_1 = -0,02x_2 + 0,05x_3 + 1,2.$$

Tương tự, ta cũng đưa phương trình thứ hai và thứ ba về dạng

$$\begin{aligned} x_2 &= 0,11x_1 - 0,03x_2 + 0,05x_3 + 2,25, \\ x_3 &= 0,11x_1 + 0,12x_2 - 0,04x_3 + 3,75. \end{aligned}$$

Kết hợp 3 phương trình trên, chúng ta đưa hệ phương trình về dạng $x =$

$Bx + c$:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} -0,02 & 0,05 & 0,10 \\ 0,11 & -0,03 & 0,05 \\ 0,11 & 0,12 & -0,04 \end{bmatrix}}_B \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} 1,2 \\ 2,25 \\ 3,75 \end{bmatrix}}_c.$$

Ta chọn $x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ và thiết lập công thức

$$x^{(k+1)} = Bx^{(k)} + c.$$

Với $k = 0$, ta có

$$x^{(1)} = c = \begin{bmatrix} 1,2 \\ 2,25 \\ 3,75 \end{bmatrix}.$$

Tiếp theo, ta tính

$$x^{(2)} = Bx^{(1)} + c = \begin{bmatrix} -0,02 & 0,05 & 0,10 \\ 0,11 & -0,03 & 0,05 \\ 0,11 & 0,12 & -0,04 \end{bmatrix} \begin{bmatrix} 1,2 \\ 2,25 \\ 3,75 \end{bmatrix} + \begin{bmatrix} 1,2 \\ 2,25 \\ 3,75 \end{bmatrix}.$$

Cụ thể hơn, ở thành phần thứ nhất, ta có

$$\begin{aligned} x_1^{(2)} &= (-0,02; 0,05; 0,10) \cdot (1,2; 2,25; 3,75) + 1,20 \\ &= -0,02 \times 1,2 + 0,05 \times 2,25 + 0,10 \times 3,75 + 1,2 = 1,6635. \end{aligned}$$

Tương tự, ta tính được

$$x_2^{(2)} = 2,502, \quad x_3^{(2)} = 4,002.$$

Vậy,

$$x^{(2)} = \begin{bmatrix} 1,6635 \\ 2,502 \\ 4,002 \end{bmatrix}.$$

Tiếp tục tính toán, với $k = 2$, ta có

$$\begin{aligned} x^{(3)} &= Bx^{(2)} + c \\ &= \begin{bmatrix} -0,02 & 0,05 & 0,10 \\ 0,11 & -0,03 & 0,05 \\ 0,11 & 0,12 & -0,04 \end{bmatrix} \begin{bmatrix} 1,6635 \\ 2,502 \\ 4,002 \end{bmatrix} + \begin{bmatrix} 1,2 \\ 2,25 \\ 3,75 \end{bmatrix} = \begin{bmatrix} 1,6120 \\ 2,5580 \\ 4,0731 \end{bmatrix}. \end{aligned}$$

Giả sử ta xem $x^{(3)}$ là một xấp xỉ của nghiệm. Để đánh giá sai số, ta cần tính chuẩn của ma trận B . Dễ thấy

$$\|B\|_{\infty} = 0,27.$$

Vì $\|B\| < 1$ nên hệ phương trình có nghiệm duy nhất. Gọi α là nghiệm duy nhất của hệ trên. Nếu sử dụng đánh giá tiên nghiệm, ta thu được

$$\begin{aligned}\|x^{(3)} - \alpha\|_{\infty} &\leq \frac{\|B\|_{\infty}^3}{1 - \|B\|_{\infty}} \|x^{(1)} - x^{(0)}\|_{\infty} = \frac{0,27^3}{1 - 0,27} \|x^{(1)}\|_{\infty} \\ &= \frac{0,27^3}{1 - 0,27} \times 3,75 < 0,11.\end{aligned}$$

Nếu sử dụng đánh giá hậu nghiệm, ta có

$$\|x^{(3)} - \alpha\|_{\infty} \leq \frac{\|B\|_{\infty}}{1 - \|B\|_{\infty}} \|x^{(3)} - x^{(2)}\|_{\infty} = \frac{0,27}{1 - 0,27} \times 0,2032 < 0,08.$$

Như vậy, đánh giá hậu nghiệm cho một đánh giá gần hơn với thực tế so với đánh giá tiên nghiệm.

3.3 Phương pháp lặp Seidel

Mục này trình bày về một phương pháp lặp Seidel⁷ để giải các hệ phương trình tuyến tính. Tài liệu tham khảo: P. K. Anh [2, §VII.6], Kreyszig [9, §20.3].

Xét hệ phương trình dạng (3.17). Phương pháp lặp Seidel bắt đầu bằng việc phân tích ma trận B dưới dạng

$$B = L + U,$$

với L là ma trận tam giác dưới với mọi phần tử trên đường chéo chính bằng 0 và U là ma trận tam giác trên. Cụ thể hơn, nếu $B = [B_{ij}]$ thì

$$L = \begin{bmatrix} 0 & 0 & \dots & 0 \\ b_{21} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & 0 \end{bmatrix}, \quad U = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{nn} \end{bmatrix}.$$

Tiếp theo, dãy lặp được thiết lập theo công thức

$$x^{(k+1)} = Lx^{(k+1)} + Ux^{(k)} + c. \quad (3.38)$$

Mặc dù ở cả hai vế của công thức trên đều chứa $x^{(k+1)}$ nhưng do L có dạng tam giác dưới nên việc tính $x^{(k+1)}$ khi biết $x^{(k)}$ có thể được thực hiện tuần tự từ thành phần thứ nhất đến thành phần thứ n . Để mô tả kĩ hơn, ta viết

⁷Philipp Ludwig von Seidel (1821–1896) là một nhà toán học Đức.

công thức lặp theo từng thành phần như sau:

$$\begin{aligned}
 x_1^{(k+1)} &= \sum_{j=1}^n b_{1j} x_j^{(k)} + c_1, \\
 x_2^{(k+1)} &= b_{21} x_1^{(k+1)} + \sum_{j=2}^n b_{2j} x_j^{(k)} + c_2, \\
 &\dots \dots \\
 x_i^{(k+1)} &= \sum_{j=1}^{i-1} b_{ij} x_j^{(k+1)} + \sum_{j=i}^n b_{ij} x_j^{(k)} + c_i, \\
 &\dots \dots \\
 x_n^{(k+1)} &= \sum_{j=1}^{n-1} b_{nj} x_j^{(k+1)} + b_{nn} x_n^{(k)} + c_n.
 \end{aligned}$$

Phương pháp lặp Seidel khác với phương pháp lặp đơn ở chỗ các thành phần $x_j^{(k+1)}$, $j = 1, 2, \dots, i-1$ vừa tính được lại được dùng để tính $x_i^{(k+1)}$.

3.3.1 Sự hội tụ của phương pháp lặp Seidel

Để nghiên cứu tính hội tụ cũng như đánh giá sai số trong phương pháp Seidel, ta cần bổ đề sau.

Định lí 3.5:

Cho B là một ma trận vuông cỡ $n \times n$. Với mỗi $i = 1, 2, \dots, n$, ta đặt

$$r_1 = 0, \quad r_i = \sum_{j=1}^{i-1} |b_{ij}| \quad \text{với } i \geq 2,$$

và

$$s_i := \sum_{j=i}^n |b_{ij}| \quad \text{với } i \geq 1.$$

Nếu $\|B\|_\infty < 1$ thì

$$\mu := \max_{1 \leq i \leq n} \left(\frac{s_i}{1 - r_i} \right) < \|B\|_\infty.$$

Chứng minh. Vì $\|B\|_\infty < 1$ và $r_i \geq 0$ nên ta có

$$s_i + \|B\|_\infty r_i \leq s_i + r_i = \sum_{j=1}^n |b_{ij}| \leq \max_{1 \leq l \leq n} \sum_{j=1}^n |b_{lj}| = \|B\|_\infty.$$

Từ đây, ta dễ dàng thu được

$$\frac{s_i}{1 - r_i} \leq \|B\|_\infty, \quad i = 1, 2, \dots, n.$$

Từ đây và định nghĩa của μ ta dễ dàng suy ra điều phải chứng minh. $\square$

Định lí 3.6:

Giả sử $\|B\|_\infty < 1$ và μ được định nghĩa như trong bổ đề 3.5. Khi đó, dãy lặp thiết lập bởi công thức lặp Seidel (3.38) hội tụ tới nghiệm duy nhất của hệ phương trình $x = Bx + c$. Hơn nữa,

(i) (Đánh giá hậu nghiệm) Với mọi $k \geq 1$, ta có

$$\|x^{(k)} - \alpha\|_\infty \leq \frac{\mu}{1 - \mu} \|x^{(k)} - x^{(k-1)}\|_\infty. \quad (3.39)$$

(ii) (Đánh giá tiên nghiệm) Với mọi $k \geq 1$, ta có

$$\|x^{(k)} - \alpha\|_\infty \leq \frac{\mu^k}{1 - \mu} \|x^{(1)} - x^{(0)}\|_\infty. \quad (3.40)$$

Nhận xét 3.2. Nếu $\|B\|_\infty < 1$ thì $\mu \leq \|B\|_\infty$ và do đó

$$\frac{\mu^k}{1 - \mu} \leq \frac{\|B\|_\infty^k}{1 - \|B\|_\infty}.$$

Như vậy, ta có thể thay μ bởi $\|B\|_\infty$ trong các đánh giá tiên nghiệm và hậu nghiệm đã nói ở trên.

Chứng minh. Bởi định lí 3.4, phương trình (3.17) có nghiệm duy nhất, gọi là α . Vậy,

$$\alpha_i = \sum_{j=1}^n b_{ij} \alpha_j + c_i.$$

Do đó,

$$\begin{aligned} x_i^{(k+1)} - \alpha_i &= \left(\sum_{j=1}^{i-1} b_{ij} x_j^{(k+1)} + \sum_{j=i}^n b_{ij} x_j^{(k)} + c_i \right) - \left(\sum_{j=1}^n b_{ij} \alpha_j + c_i \right) \\ &= \sum_{j=1}^{i-1} b_{ij} (x_j^{(k+1)} - \alpha_j) + \sum_{j=i}^n b_{ij} (x_j^{(k)} - \alpha_j). \end{aligned} \quad (3.41)$$

Như trong Bổ đề 3.5, ta đặt

$$r_i := \sum_{j=1}^{i-1} |b_{ij}|, \quad s_i := \sum_{j=i}^n |b_{ij}|, \quad \mu := \max_{1 \leq i \leq n} \left(\frac{s_i}{1 - r_i} \right).$$

Khi đó, từ (3.41), ta dễ dàng đánh giá được

$$\begin{aligned} |x_i^{(k+1)} - \alpha_i| &\leq \sum_{j=1}^{i-1} |b_{ij}| |x_j^{(k+1)} - \alpha_j| + \sum_{j=i}^n |b_{ij}| |x_j^{(k)} - \alpha_j| \\ &\leq \sum_{j=1}^{i-1} |b_{ij}| \max_{1 \leq l \leq n} |x_l^{(k+1)} - \alpha_l| + \sum_{j=i}^n |b_{ij}| \max_{1 \leq l \leq n} |x_l^{(k)} - \alpha_l| \\ &= r_i \|x^{(k+1)} - \alpha\|_{\infty} + s_i \|x^{(k)} - \alpha\|_{\infty}. \end{aligned} \quad (3.42)$$

Mặt khác, theo định nghĩa của chuẩn $\|\cdot\|_{\infty}$, với mỗi k , có một chỉ số i_0 sao cho

$$\|x^{(k+1)} - \alpha\|_{\infty} := \max_{1 \leq i \leq n} |x_i^{(k+1)} - \alpha_i| = |x_{i_0}^{(k+1)} - \alpha_{i_0}|.$$

Vậy, với $i = i_0$, (3.42) có thể viết lại như sau:

$$\|x^{(k+1)} - \alpha\|_{\infty} \leq r_{i_0} \|x^{(k+1)} - \alpha\|_{\infty} + s_{i_0} \|x^{(k)} - \alpha\|_{\infty}.$$

Từ đây, ta dễ dàng giải được (vì $0 \leq r_{i_0} \leq \|B\|_{\infty} < 1$)

$$\|x^{(k+1)} - \alpha\|_{\infty} \leq \frac{s_{i_0}}{1 - r_{i_0}} \|x^{(k)} - \alpha\|_{\infty} \leq \mu \|x^{(k)} - \alpha\|_{\infty}. \quad (3.43)$$

Lập luận bằng quy nạp như trong chứng minh Định lí 3.4, suy ra

$$\|x^{(m)} - \alpha\|_{\infty} \leq \mu^m \|x^{(0)} - \alpha\|_{\infty}, \quad m = 0, 1, 2, \dots$$

Theo Bổ đề 3.5, $\mu \leq \|B\|_{\infty} < 1$ nên

$$\mu^m \|x^{(0)} - \alpha\|_{\infty} \rightarrow 0 \quad \text{khi } m \rightarrow \infty.$$

Áp dụng định lí giới hạn bị kẹp ta suy ra

$$\|x^{(m)} - \alpha\|_{\infty} \rightarrow 0 \quad \text{khi } m \rightarrow \infty.$$

Đó là điều phải chứng minh. □

Ví dụ 3.7. Tìm gần đúng nghiệm của hệ phương trình sau bằng phương pháp Seidel với 3 bước lặp.

$$\begin{cases} 40x_1 - 2x_2 + x_3 = 20 \\ x_1 - 40x_2 + x_3 = 0 \\ x_2 - 20x_3 = 30. \end{cases} \quad (3.44)$$

Hãy đánh giá sai số của nghiệm gần đúng thu được.

Lời giải

Dễ dàng đưa hệ về dạng

$$\begin{cases} x_1 = 0,05x_2 - 0,025x_3 + 0,5 \\ x_2 = 0,025x_1 + 0,025x_3 \\ x_3 = 0,05x_2 - 1,5. \end{cases}$$

Có thể viết hệ 3 phương trình này về dạng phương trình của vectơ $x := \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

như sau:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0,05 & -0,025 \\ 0,025 & 0 & 0,025 \\ 0 & 0,05 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0,5 \\ 0 \\ -1,5 \end{bmatrix}.$$

Gọi B là ma trận các hệ số trong vế phải, thì

$$\|B\|_{\infty} = 0,075 < 1.$$

Do $\|B\|_{\infty} < 1$ nên quá trình Seidel hội tụ.

Chọn $x^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ và tính các vectơ $x^{(k)}$, $k = 1, 2, 3$, theo công thức lặp

Seidel. Cụ thể, ta có

$$\begin{aligned} x_1^{(1)} &= 0,05x_2^{(0)} - 0,025x_3^{(0)} + c_1 = 0,5, \\ x_2^{(1)} &= 0,025x_1^{(1)} + 0,025x_3^{(0)} + c_2 = 0,0125 \\ x_3^{(1)} &= 0,05x_2^{(1)} + c_3 = -1,499375. \end{aligned}$$

Tiếp theo, ta tính $x^{(2)}$ theo công thức

$$\begin{aligned} x_1^{(2)} &= 0,05x_2^{(1)} - 0,025x_3^{(1)} + c_1 = 0,538109, \\ x_2^{(2)} &= 0,025x_1^{(2)} + 0,025x_3^{(1)} + c_2 = -0,024032, \\ x_3^{(2)} &= 0,05x_2^{(2)} + c_3 = -1,501202. \end{aligned}$$

Tiếp theo, ta tính $x^{(3)}$ theo công thức

$$\begin{aligned}x_1^{(3)} &= 0,05x_2^{(2)} - 0,025x_3^{(2)} + c_1 = 0,536328, \\x_2^{(3)} &= 0,025x_1^{(3)} + 0,025x_3^{(2)} + c_2 = -0,024122, \\x_3^{(3)} &= 0,05x_2^{(3)} + c_3 = -1,501206.\end{aligned}$$

Để đánh giá sai số, ta dùng công thức đánh giá hậu nghiệm. Theo hướng này, ta tính

$$x^{(3)} - x^{(2)} = - \begin{bmatrix} 0,001781 \\ 0,000090 \\ 0,000004 \end{bmatrix}, \quad \|x^{(3)} - x^{(2)}\|_{\infty} = 0,001781.$$

Vậy, nếu gọi $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$ là nghiệm thì

$$\begin{aligned}\|x^{(3)} - \alpha\|_{\infty} &< \frac{\|B\|_{\infty}}{1 - \|B\|_{\infty}} \|x^{(3)} - x^{(2)}\|_{\infty} = \frac{0,075}{1 - 0,075} \times 0,001781 \\ &< 0,000144.\end{aligned}$$

Đó là một sai số khá nhỏ.

Nhận xét 3.3. Trong đẳng thức (3.38), ta chuyển $Lx^{(k+1)}$ từ vế phải sang vế trái và viết

$$(I - L)x^{(k+1)} = Ux^{(k)} + c.$$

Vì L là ma trận tam giác dưới với đường chéo chính bằng 0 nên $I - L$ cũng là ma trận tam giác dưới có hạng tối đại ($= n$) vì các phần tử trên đường chéo chính bằng 1. Từ đó, nhân hai vế với ma trận nghịch đảo $(I - L)^{-1}$ về phía bên trái, ta có

$$x^{(k+1)} = (I - L)^{-1}Ux^{(k)} + (I - L)^{-1}c. \quad (3.45)$$

Phương trình (3.45) có dạng của phương pháp lặp đơn, $x^{(k+1)} = Bx^{(k)} + b$, với

$$B = (I - L)^{-1}U.$$

Như vậy, phương pháp lặp Seidel cũng là một trường hợp riêng của phương pháp lặp đơn, áp dụng cho một dạng tương đương của hệ phương trình ban đầu.

3.3.2 Một ví dụ: Mô hình Leontief

Trong một số bài toán từ thực tế, hệ phương trình cần giải lại xuất hiện ngay ở dạng $x = Bx + b$, chẳng hạn như trong mô hình kinh tế Leontief⁸. Mô

⁸Wassily Wassilyovich Leontief (1905–1999) là một nhà kinh tế gốc Xô Viết.

hình Leontief là một mô hình toán cho một nền kinh tế gồm n ngành sản xuất, được đánh số từ 1 đến n , mà chúng ta sẽ mô tả vắn tắt như sau: Gọi x_i là giá trị tổng cầu về sản phẩm của ngành thứ i (tổng giá trị sản phẩm ngành i) và b_i là giá trị cầu cuối cùng (cầu về sản phẩm ngành i từ phía các hộ tiêu dùng và các nhà xuất khẩu, v.v). Gọi x_{ik} giá trị cầu trung gian, tức là cầu về sản phẩm ngành i từ phía ngành k hay nói một cách khác là số tiền mà ngành k bỏ ra mua sản phẩm ngành i làm nguyên liệu đầu vào).

Trong trạng thái cân bằng, chúng ta có công thức tổng cầu về sản phẩm hàng hóa của ngành:

$$x_i = x_{i1} + x_{i2} + \cdots + x_{in} + b_i.$$

Tỉ phần chi phí đầu vào của ngành k đối với sản phẩm i là tỉ lệ

$$a_{ik} = \frac{x_{ik}}{x_k},$$

và nó được giả thiết là ổn định. Từ ý nghĩa kinh tế của các biến số, chúng ta thấy rằng

$$0 \leq a_{ik} \leq 1.$$

Ma trận B tạo bởi các yếu tố a_{ik} gọi là *ma trận hệ số kĩ thuật* của mô hình. Khi đó, trong trạng thái cân bằng, chúng ta thu được

$$x = Bx + b,$$

với $\|B\|_1 \leq 1$. Do đó, thay vì đưa về dạng $Ax = b$ với $A = I - B$, chúng ta có thể áp dụng trực tiếp các phương pháp lập cho phương trình cân bằng trong mô hình Leontief.

CAS 3.4. Giả sử trong một mô hình Leontief có ba ngành sản xuất với ma trận hệ số kĩ thuật của nó cho bởi

$$B = \begin{bmatrix} 0,1920 & 0,1144 & 0,0042 \\ 0,2678 & 0,2656 & 0,0124 \\ 0,0479 & 0,0741 & 0,0089 \end{bmatrix}.$$

Giả sử cầu bên ngoài là $b = [5600 \quad 1320 \quad 2430]^T$. Chúng ta tìm trạng thái cân bằng của mô hình này bằng phương pháp lập Seidel thực hiện trong R như sau:

```
# Hàm my_seidel() thực hiện lập theo phương pháp Seidel
# giải hệ phương trình tuyến tính  $x = Bx + c$ 

my_seidel <- function(B, c, tol = 1e-6, max_iter = 1000) {
  n <- length(c)
```

```
x <- numeric(n)
for (k in 1:max_iter) {
  x_old <- x
  x[1] <- c[1] + sum(B[1, 1:n] * x_old[1:n])
  for (i in 2:n) {
    sum1 <- sum(B[i, 1:i - 1] * x[1:i - 1])
    sum2 <- sum(B[i, i:n] * x_old[i:n])
    x[i] <- c[i] + sum1 + sum2
  }
  if (sqrt(sum((x - x_old)^2)) < tol) {
    cat("Hội tụ sau", k, "lần lặp.\n")
    return(x)
  }
}
warning("Không hội tụ sau số lần lặp tối đa.")
return(x)
}
```

Áp dụng my_seidel() cho ví dụ trên.

```
B <- matrix(c(0.1920, 0.1144, 0.0042,
              0.2678, 0.2656, 0.0124,
              0.0479, 0.0741, 0.0089), byrow = T, nrow = 3)
c <- c(5600, 1320, 2430)
solution <- my_seidel(B, c, 0.01, 1000)

## Hội tụ sau 14 lần lặp.

print(solution)

## [1] 7601,65476 4622,77547 3164,83394
```

Kết quả thu được là nghiệm của hệ phương trình $x = Bx + b$.

3.4 Câu hỏi và bài tập

1. Cho các vectơ sau đây:

$$x = \begin{bmatrix} 1 \\ 1,5 \\ -5 \end{bmatrix}, \quad y = \begin{bmatrix} 3 \\ 2 \\ -6 \end{bmatrix}.$$

Tính các chuẩn ℓ_1 , ℓ_∞ và ℓ_2 của các vectơ x , y , và $x - y$.

2. Cho các ma trận sau

$$A = \begin{bmatrix} 1 & 4 & -2 \\ -2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 8 & 3 \\ 2 & -0,5 & 1 \\ -0,2 & 4 & 2 \end{bmatrix}.$$

Tính $\|A\|_1, \|A\|_\infty, \|B\|_1, \|B\|_\infty$.

3. Giải hệ phương trình dạng tam giác sau đây.

$$\begin{cases} 2x_1 + x_2 - x_3 = 7 \\ x_2 - x_3 = 2 \\ 3x_3 = 6 \end{cases}$$

4. Giải hệ phương trình sau đây bằng phương pháp khử Gauss.

$$\begin{cases} 2x_1 + x_2 - x_3 = 3 \\ x_1 - x_2 + 2x_3 = 2 \\ 3x_1 - 2x_2 + x_3 = -1 \end{cases}$$

5. Xét hệ phương trình $Ax = b$ với

$$A = \begin{bmatrix} 4 & 0,24 & -0,08 \\ 0,09 & 3 & -0,15 \\ 0,04 & -0,08 & 4 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ 9 \\ 20 \end{bmatrix}.$$

(a) Biến đổi hệ này về dạng $x = Bx + c$, với b là một vectơ cột và B là một ma trận cấp 3×3 thỏa mãn $\|B\|_1 < 1$.

(b) Dùng kết quả của (a), xây dựng dãy lặp

$$x^{(k+1)} = Bx^{(k)} + c, \quad x^{(0)} = (0,0,0)^T.$$

(T kí hiệu phép chuyển vị). Tính $x^{(3)}$.

(c) Gọi α là nghiệm đúng của hệ. Hãy ước lượng sai số $\|x^{(3)} - \alpha\|_1$.

6. Cho hệ phương trình

$$\begin{cases} 10x_1 + x_2 - 5x_3 = -5 \\ x_1 - 20x_2 - x_3 = -8 \\ x_1 + x_2 + 15x_3 = 40 \end{cases}$$

(a) Đưa hệ phương trình về dạng $x = Bx + c$, với $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ sao cho B là một ma trận cấp 3×3 có chuẩn bé hơn 1: $\|B\|_\infty < 1$. Tính chuẩn $\|B\|_\infty$.

(b) Chọn $x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Tính các xấp xỉ $x^{(1)}, x^{(2)}$ theo phương pháp lặp Seidel

cho hệ phương trình $x = Bx + c$ trong phần (a).

7. Giải gần đúng hệ sau đây bằng phương pháp Seidel với sai số $\epsilon = 0,05$.

$$\begin{cases} 5x_1 - x_2 + 2x_3 = 6 \\ x_1 - 4x_2 + x_3 = -2 \\ -2x_1 - x_2 + 4x_3 = 1 \end{cases}$$

8. Cho ma trận

$$B = \begin{bmatrix} -0,6 & 0,6 \\ 0,4 & 0,2 \end{bmatrix}.$$

(a) Tính $\|B\|_\infty$, $\|B\|_1$, và $\|B\|_2$.

(b) Chứng minh rằng với mọi $x^{(0)} \in \mathbb{R}^2$, dãy lặp đơn

$$x^{(k+1)} = Bx^{(k)} + c$$

hội tụ tới nghiệm duy nhất của hệ phương trình $x = Bx + c$.

Chương

4

Đa thức nội suy và xấp xỉ đa thức

Chương này trình bày về nội suy đa thức bằng công thức Lagrange và công thức Newton, nội suy hàm ghép trơn bậc ba (cubic spline), và xấp xỉ bằng phương pháp bình phương bé nhất (least squares method).

Nội dung

4.1	Bài toán nội suy, đa thức nội suy	92
4.1.1	Bài toán nội suy	92
4.1.2	Sự tồn tại và duy nhất của đa thức nội suy	92
4.1.3	Tính giá trị của đa thức: Lược đồ Horner	95
4.2	Đa thức nội suy Lagrange	99
4.3	Đa thức nội suy Newton	103
4.3.1	Đa thức nội suy Newton tổng quát	103
4.3.2	Đa thức nội suy Newton với các nút nội suy cách đều	110
4.3.3	Đọc thêm: Định lí Hermite–Genocchi	114
4.4	Sai số nội suy	115
4.4.1	Đánh giá sai số trong xấp xỉ bằng đa thức nội suy	115
4.4.2	Chọn nút nội suy tối ưu: đa thức Chebyshev	118
4.5	Nội suy hàm ghép trơn	123
4.5.1	Sơ lược về hàm ghép trơn	123
4.5.2	Hàm ghép trơn bậc ba	124
4.5.3	Quy trình tính hàm ghép trơn bậc ba	128
4.6	Phương pháp bình phương bé nhất	133
4.6.1	Xấp xỉ tuyến tính	134
4.6.2	Xấp xỉ đa thức bằng phương pháp bình phương bé nhất	139

4.6.3	Xấp xỉ bằng các hàm lũy thừa	143
4.6.4	Xấp xỉ bằng hàm số mũ	146
4.7	Câu hỏi và bài tập	148

4.1 Bài toán nội suy, đa thức nội suy

Dưới đây, chúng ta tìm hiểu các kiến thức cơ bản về đa thức, bài toán nội suy, đa thức nội suy, các công thức nội suy, tính giá trị đa thức bằng lược đồ Horner. Tài liệu tham khảo: T. V. Đĩnh [6, §4.1], P. K. Anh [2, §II.1], Kreyszig [9, §19.3].

4.1.1 Bài toán nội suy

Giả sử chúng ta quan sát hai đại lượng x và y , với y là một hàm số của đối số $x \in [a, b]$. Tại $n + 1$ điểm $x_0, x_1, \dots, x_n$, với

$$a \leq x_0 < x_1 < \dots < x_n \leq b,$$

ta quan sát được các giá trị tương ứng là $y_0, y_1, \dots, y_n$. Vấn đề đặt ra là từ các quan sát đó có thể tìm ra được mối liên hệ giữa x và y hay không? Theo ngôn ngữ toán học, nó có nghĩa là có thể tìm ra một hàm số $f: [a, b] \rightarrow \mathbb{R}$ sao cho

$$f(x_i) = y_i, \quad i = 0, 1, 2, \dots, n \quad (4.1)$$

hay không?

Nếu không có các yêu cầu thêm nào đối với hàm f thì có vô số hàm số f thỏa mãn điều kiện như vậy. Vì vậy, người ta thường tìm f trong một lớp các hàm đã biết như các đa thức, hàm lượng giác, v.v. Trong mục này, ta sẽ tìm hiểu vấn đề xác định một đa thức P_n có bậc bé hơn hoặc bằng n sao cho

$$P_n(x_i) = y_i, \quad i = 0, 1, 2, \dots, n. \quad (4.2)$$

Vấn đề này gọi là vấn đề nội suy đa thức (polynomial interpolation).

4.1.2 Sự tồn tại và duy nhất của đa thức nội suy

Định lí sau đây cho điều kiện về sự tồn tại của các đa thức nội suy.

Định lí 4.1: Tồn tại và duy nhất

Với mọi bộ $n + 1$ giá trị $(y_0, y_1, \dots, y_n)$, tồn tại duy nhất một đa thức P_n với bậc không vượt quá n ($\deg P_n \leq n$) sao cho

$$P_n(x_i) = y_i, \quad i = 0, 1, 2, \dots, n. \quad (4.3)$$

Đa thức P_n thỏa mãn (4.3) gọi là *đa thức nội suy* của hàm số $f(x)$ tại các điểm nội suy $x_0, \dots, x_n$. Dễ dàng chứng minh được tính duy nhất của đa thức nội suy trong định lý 4.1. Thật vậy, ta giả sử $P(x)$ và $Q(x)$ cùng thỏa mãn

$$P(x_i) = y_i, \quad Q(x_i) = y_i, \quad i = 0, 1, 2, \dots, n.$$

Khi đó, $R(x) := P(x) - Q(x)$ là một đa thức với bậc không vượt quá n thỏa mãn

$$R(x_i) = 0, \quad i = 0, 1, 2, \dots, n.$$

Như vậy, $R(x)$ là đa thức bậc không vượt quá n và có $n + 1$ nghiệm nên nó phải là đa thức 0. Tức là $P(x) = Q(x)$; tính duy nhất được chứng minh.

Điều kiện tồn tại trong định lý 4.1 là hệ quả của tính duy nhất của đa thức nội suy. Thật vậy, bài toán tìm đa thức nội suy có thể đưa về bài toán tìm các hệ số $a_0, a_1, \dots, a_n$ sao cho

$$a_0 + a_1 x_j + a_n x_j^n = y_j, \quad j = 0, 1, \dots, n.$$

Chúng lập thành một hệ phương trình tuyến tính gồm $n + 1$ phương trình của $n + 1$ ẩn số $a_0, a_1, \dots, a_n$, có thể viết dưới dạng ma trận như sau:

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}. \quad (4.4)$$

Sự tồn tại và duy nhất của đa thức nội suy tương đương với sự tồn tại và duy nhất nghiệm của hệ phương trình tuyến tính (4.4).

Như đã chứng minh, hệ (4.4) không có quá một nghiệm. Nói riêng, hệ phương trình tuyến tính thuần nhất tương ứng với (4.4) chỉ có nghiệm tầm thường. Do đó, định thức của ma trận hệ số trong vế trái của (4.4) phải khác không và vậy thì hệ (4.4) có nghiệm duy nhất với mọi bộ (y_j) .

Chú ý 4.1. Ma trận hệ số trong vế trái của (4.4) gọi là ma trận Vandermonde¹, kí hiệu V . Định thức $\det(V)$ của nó được gọi là định thức Vandermonde. Bằng tính toán cụ thể,

$$\det(V) = \prod_{0 \leq i < j \leq n} (x_j - x_i).$$

Từ đó có thể thấy định thức này khác không khi các điểm x_j phân biệt.

CAS 4.1. Dựa trên các phân tích trên, chúng ta xây dựng cách tính đa thức nội suy dựa vào hệ phương trình (4.4). Với một dãy các điểm x , chúng ta lập ma trận Vandermonde tương ứng. Các hệ số của đa thức nội suy được

¹Alexandre-Théophile Vandermonde (28/2/1735 – 1/1/1796) là một nhà toán học người Pháp.

tính bằng cách vận dụng hàm `solve()` có sẵn trong R để giải hệ phương trình (4.4).

```
# vandermonde() tạo ma trận Vandermonde cỡ n tại giá trị x.

vandermonde <- function(x) {
  n = length(x)

  # Khởi tạo một ma trận 1 cột với các phần tử 1
  A = matrix(data = 1, n, 1)

  # Vòng lặp "for" xây dựng các cột tiếp theo của
  # định thức Vandermonde
  for (i in 1:(n - 1)) {
    A = cbind(A, A[, i] * x)
  }
  return(A)
}

# polycalc() tính các hệ số của đa thức nội suy.
polycalc <- function(x, y) {

  # Tính ma trận Vandermonde của dữ liệu x
  A = vandermonde(x)
  if (det(A) == 0) {
    stop("Định thức Vandermonde triệt tiêu.")
  }

  # Giải hệ  $Ac = y$  bằng solve() có sẵn
  c = solve(A, y)
  return(c)
}

# Ví dụ

x <- c(-1, 0, 2, 3)
y <- c(14, 24, 80, 342)
polycalc(x, y)

## [1] 24 -20 -12 18
```

Mặc dù chương trình R phía trên cho kết quả chính xác đối với dữ liệu trong ví dụ trên, nó có thể cho kết quả với sai số lớn khi các dữ liệu đầu vào “xấu”. Trong mục sau, chúng ta sẽ tìm hiểu các cách xây dựng hai biểu

diễn đa thức nội suy hiệu quả hơn, đó là hai công thức nội suy Lagrange² và nội suy Newton.

4.1.3 Tính giá trị của đa thức: Lược đồ Horner

Giả sử chúng ta cần tính giá trị một đa thức, chẳng hạn

$$p(x) = 18x^3 - 12x^2 - 20x + 24,$$

tại $x = 2,5$. Rõ ràng, chúng ta có thể tính các lũy thừa x^i , $i = 1, 2, 4$, của $x = 2,5$, nhân các kết quả với các hệ số tương ứng, và cộng tất cả các kết quả lại. Quá trình này có thể thực hiện trong ngôn ngữ R như sau:

```
# poly.eval() tính giá trị của đa thức.

poly.eval <- function(a, x) {
  n <- length(a)
  t <- numeric(n) + 1

  # Tính các lũy thừa của x và đưa vào t.

  for (i in 1:(n-1)) {
    t[i + 1] = t[i] * x
  }

  # Tính tổng của tích các lũy thừa và hệ số tương ứng.

  w <- sum(a * t)
  return(w)
}

# Ví dụ

a <- c(24, -20, -12, 18)
x <- 2.5
poly.eval(a, x)

## [1] 180,25
```

Kết quả thu được $p(2,5) = 180,25$ sau một số bước tính toán với 5 phép nhân. Cách tính trên nhìn chung không hiệu quả do cần khá nhiều phép tính, đặc biệt là phép nhân. Dưới đây, chúng ta nói về lược đồ Horner để tính giá trị của đa thức.

²Joseph-Louis Lagrange (1736 – 1813) là nhà toán học người Italy.

Lược đồ Horner

Trong mục này, chúng ta nói về *lược đồ Horner*³ (Horner's Scheme) để tăng hiệu quả tính toán trong vấn đề tính giá trị đa thức.

Ý tưởng của lược đồ Horner trong tính toán giá trị của các đa thức có thể được minh hoạ trong sơ đồ sau:

$$p(x_0) = 24 + x_0(-20 + x_0(\underbrace{-12 + 18x_0}_{w_2})).$$

$$\underbrace{\underbrace{-20 + w_2 x_0}_{w_1}}_{w_0 = 24 + w_1 x_0}$$

Người ta gọi dạng của đa thức trong vế phải của đẳng thức trên là “dạng lồng nhau” (nested form) của đa thức $p(x)$. Khi được viết ở dạng này, việc tính toán được thực hiện từ trong ra ngoài theo quy tắc sau:

$$\begin{aligned} w_3 &= 18, \\ w_2 &= -12 + w_3 x_0, \\ w_1 &= -20 + w_2 x_0, \\ w_0 &= 24 + w_1 x_0. \end{aligned}$$

Thay số $x_0 = 2,5$ và tính toán cụ thể, chúng ta thu được

$$\begin{aligned} w_3 &= 18, \\ w_2 &= -12 + 18 \times 2,5 = 33, \\ w_1 &= -20 + 33 \times 2,5 = 62,5, \\ w_0 &= 24 + 62,5 \times 2,5 = 180,25. \end{aligned}$$

Khi đó,

$$p(x_0) = w_0 = 180,25$$

là giá trị cần tính (như trên), thu được sau 3 phép nhân và 3 phép cộng.

Tổng quát, đối với một đa thức

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$

chúng ta sẽ tính một dãy $w_0, w_1, \dots, w_n$, bắt đầu từ w_n và đi ngược về w_0 như sau:

$$\begin{aligned} w_n &= a_n, \\ w_{n-1} &= a_{n-1} + w_n x_0, \\ &\dots = \dots \\ w_1 &= a_1 + w_2 x_0, \\ w_0 &= a_0 + w_1 x_0. \end{aligned}$$

³William George Horner (9/6/1786 – 22/9/1837) là một nhà toán học người Anh.

Khi đó, từ các giá trị w_i , chúng ta tính được thương và phần dư của phép chia đa thức $p(x)$ cho nhị thức bậc nhất $x - x_0$ như sau:

$$p(x) = w_0 + (x - x_0) \sum_{i=1}^n w_i x^{i-1}.$$

Nói riêng, giá trị của $p(x)$ tại $x = x_0$ là

$$p(x_0) = w_0.$$

Khi tính toán bằng tay, lược đồ Horner thường được ghi dưới dạng một bảng số với 3 hàng như sau:

a_n	a_{n-1}	a_{n-2}	$\dots$	a_0
$x_0 w_n$	$x_0 w_{n-1}$	$\dots$	$x_0 w_1$	
$w_n = a_n$	w_{n-1}	w_{n-2}	$\dots$	$w_0 = p(x_0)$

Ở hàng trên cùng, chúng ta viết các hệ số của đa thức. Ở hàng thứ 3, chúng ta bắt đầu với $w_n = a_n$, sau đó tính $x_0 w_n$ và viết kết quả vào hàng thứ hai, cột bên phải của cột w_n . Cộng hai giá trị của cột này, ta thu được w_{n-1} . Tiếp tục làm như vậy cho cột $w_{n-1}, \dots$

Ví dụ 4.1. Lập lược đồ Horner tính giá trị tại $x_0 = 2,5$ của đa thức bậc 4 sau đây:

$$p(x) = x^4 + 20x^3 - 11x^2 - 20x + 22.$$

Lời giải

Ta lập bảng cho lược đồ Horner như sau:

1	20	-11	-20	22
	2,5	56,25	113,125	232,8125
$w_4 = 1$	22,5	45,25	93,125	$254,8125 = w_0 = p(2,5)$

Như vậy, $p(2,5) = 254,8125$. Hơn nữa,

$$p(x) = 254,8125 + (x - 2,5)(93,125 + 45,25x + 22,5x^2 + x^3).$$

CAS 4.2. Hãy viết chương trình máy tính để kiểm tra các tính toán trong Ví dụ 4.1. Lưu ý, trong lược đồ Horner như trên, nếu đặt

$$w_{n+1} = 0,$$

thì, quá trình tính w_k trong lược đồ có công thức truy hồi

$$w_k = a_k + w_{k+1}x_0, \quad k = n, n-1, \dots, 0.$$

Do đó, sẽ chỉ cần tạo một vòng lặp để thực hiện công thức truy hồi trên. Trong R, chúng ta có thể viết như sau:

```
# Hàm horner() thực hiện quá trình tính theo lược đồ Horner.

horner <- function(x, a) {

  # Khởi tạo các giá trị ban đầu trong lược đồ Horner
  n <- length(a)
  w <- numeric(n)
  w[n] <- a[n]

  # Vòng lặp chính trong lược đồ Horner
  while (n > 1) {
    n <- n - 1
    w[n] <- a[n] + w[n + 1] * x
  }

  # Trả về các giá trị trong lược đồ Horner
  return(w)
}

# Áp dụng horner() cho dữ liệu trong ví dụ trên.

a <- c(22, -20, -11, 20, 1)
x <- 2.5
horner(x, a)

## [1] 254,8125  93,1250  45,2500  22,5000  1,0000
```

Kết quả của `horner(x, a)` cho dãy các giá trị w_i đúng như kết quả đã tính trong Ví dụ 4.1, với giá trị đầu tiên là giá trị của $p(x_0)$.

Các đa thức có thể xuất hiện xấp xỉ các hàm số siêu việt bằng một đa thức Taylor của nó. Sau đây là một ví dụ.

CAS 4.3. Hãy áp dụng lược đồ Horner cho các tổng riêng của các chuỗi lũy thừa, lấy khai triển Taylor của hàm $y = \sin(x)$ làm ví dụ. Nhắc lại rằng, các đa thức Taylor của hàm số $\sin$ có dạng

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots$$

Chương trình R sau đây tính giá trị của các đa thức Taylor thứ 11 của hàm $\sin(x)$ tại các giá trị $x_1 = 0,1$; $x_2 = 0,2$; ...; $x_5 = 0,5$. Lưu ý, trong đoạn mã sau, hàm `horner2()` đơn giản hơn đoạn mã phía trên do chúng ta chỉ giữ lại giá trị cuối cùng trong lược đồ Horner.

```

# Hàm horner2() tính giá trị của đa thức bằng lược đồ Horner.

horner2 <- function(x, a) {
  n <- length(a)
  w <- 0
  while (n > 0) {
    w <- a[n] + w * x
    n <- n - 1
  }
  return(w)
}

# Tính hệ số của đa thức Taylor của hàm sin(x)

taylor_series_sine <- function(truncation) {
  a <- numeric(truncation)
  i <- 2
  a[i] <- 1
  while (i < truncation - 2) {
    a[i + 2] <- - a[i] / (i * (i + 1))
    i <- i + 2
  }
  return(a)
}

# Tính giá trị của đa thức Taylor thứ 11 của hàm sin(x)

n <- 12
coeffs <- taylor_series_sine(n)
x <- seq(0.1, 0.4, by = 0.1)
horner2(x, coeffs)

## [1] 0,0998334166 0,1986693308 0,2955202067 0,3894183423

```

Kết quả là các giá trị gần đúng của $\sin(x_i)$, $i = 1, 2, 3, 4$ chính xác tới ít nhất 10 chữ số thập phân.

4.2 Đa thức nội suy Lagrange

Trong mục này, chúng ta nói về công thức nội suy Lagrange. Tài liệu tham khảo: T. V. Đĩnh [6, §4.1], P. K. Anh [2, §II.2], Kreyszig [9, §19.3].

Nghiệm của bài toán nội suy (4.3) với vế phải $y_0, y_1, \dots, y_n$ tổng quát có thể được xây dựng bằng cách “chồng chất” các nghiệm của bài toán với vế phải đơn giản hơn. Cụ thể hơn, gọi $\ell_i(x)$ là đa thức bậc không vượt quá n

sao cho

$$\ell_i(x_j) = \begin{cases} 1 & \text{nếu } j = i, \\ 0 & \text{nếu } j \neq i. \end{cases} \quad (4.5)$$

Nói cách khác, ℓ_i là nghiệm của bài toán (4.1) với $y_i = 1$ và $y_j = 0$ với $j \neq i$.

Các đa thức $\ell_i(x)$ được gọi là các **đa thức nội suy cơ bản** bậc n ứng với các nút nội suy $x_0, x_1, \dots, x_n$.

Định lí 4.2: Công thức nội suy Lagrange

Đặt $P(x)$

$$P(x) = \sum_{i=0}^n y_i \ell_i(x) = y_0 \ell_0(x) + y_1 \ell_1(x) + \dots + y_n \ell_n(x).$$

Khi đó, $P(x)$ là một đa thức nội suy các giá trị $y_0, y_1, \dots, y_n$, có nghĩa là $P(x)$ là đa thức bậc không vượt quá n thỏa mãn $P(x_i) = y_i$ với $i = 0, 1, \dots, n$.

Chứng minh. Rõ ràng $P(x)$ có bậc không vượt quá n , vì mỗi số hạng là một bội số của một đa thức ℓ_k , $0 \leq k \leq n$, có bậc không vượt quá n .

Mặt khác, với mỗi k , ta có $\ell_i(x_k) = 0$ trừ khi $i = k$. Do đó,

$$P(x_k) = y_0 \ell_0(x_k) + y_1 \ell_1(x_k) + \dots + y_n \ell_n(x_k) = y_k \ell_k(x_k) = y_k.$$

Từ đó ta suy ra $P(x)$ như trên là đa thức thỏa mãn yêu cầu $P(x_k) = y_k$ với mọi $k = 0, 1, 2, \dots, n$, tức là thỏa mãn (4.3). $\square$

Vậy, chúng ta cần biết công thức cụ thể cho các **đa thức nội suy cơ bản** ℓ_i , $i = 0, 1, 2, \dots, n$. Chúng được cho bởi công thức (4.6) trong định lí sau đây.

Định lí 4.3:

Đa thức duy nhất ℓ_i thỏa mãn (4.5) được cho bởi công thức

$$\ell_i(x) = \prod_{0 \leq k \leq n, k \neq i} \frac{x - x_k}{x_i - x_k}. \quad (4.6)$$

Chứng minh. Vế phải của (4.6) là tích của n nhân tử bậc nhất nên ℓ_i là đa thức bậc n . Lấy j tùy ý. Nếu $j \neq i$ thì có

$$\ell_i(x_j) = \prod_{0 \leq k \leq n, k \neq i} \frac{x_j - x_k}{x_i - x_k} = 0,$$

vì về phải là một tích trong đó có một nhân tử bằng 0 (nhân tử tương ứng với chỉ số $k = j$). Còn nếu $j = i$ thì rõ ràng

$$\ell_i(x_i) = \prod_{0 \leq k \leq n, k \neq i} \frac{x_i - x_k}{x_i - x_k} = 1,$$

vì mọi nhân tử ở về phải bằng 1. Vậy ℓ_i thỏa mãn (4.5). $\square$

Ví dụ 4.2 (Nội suy bậc nhất). Với $n = 1$, ta có $n + 1 = 2$ nút nội suy x_0 và x_1 . Hai đa thức Lagrange cơ bản là

$$\ell_0(x) = \frac{x - x_1}{x_0 - x_1}, \quad \ell_1(x) = \frac{x - x_0}{x_1 - x_0}.$$

Đa thức nội suy bậc nhất tại các nút x_0, x_1 với các giá trị tương ứng y_0, y_1 là

$$P_1(x) = y_0 \ell_0(x) + y_1 \ell_1(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0}.$$

Ví dụ 4.3 (Nội suy bậc hai). Với $n = 2$, ta có $n + 1 = 3$ nút nội suy x_0, x_1, x_2 . Với các giá trị tương ứng y_0, y_1, y_2 tại các nút này, đa thức nội suy là

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

Ví dụ 4.4. Cho $n = 3$ và $n + 1 = 4$ nút nội suy x_0, x_2, x_2 , và x_3 . Viết công thức tường minh của đa thức nội suy cơ bản $\ell_2(x)$.

Lời giải

Đa thức nội suy cơ bản $\ell_2(x)$ là tích của các nhân tử $(x - x_j)/(x_2 - x_j)$, với $j = 0, 1, \dots, 3, j \neq 2$. Vậy,

$$\ell_2(x) = \frac{x - x_0}{x_2 - x_0} \cdot \frac{x - x_1}{x_2 - x_1} \cdot \frac{x - x_3}{x_2 - x_3}.$$

Ví dụ 4.5. Tìm đa thức bậc hai $P_2(x)$ có giá trị trùng với hàm $y = 5^x$ tại các điểm $x_0 = -1, x_1 = 0$, và $x_2 = 1$.

Lời giải

Ta có 3 điểm nội suy phân biệt nên đa thức bậc hai $P_2(x)$ được xác định duy nhất và có thể biểu diễn qua 3 đa thức nội suy Lagrange cơ bản

$$\begin{aligned} \ell_0(x) &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} = \frac{(x - 0)(x - 1)}{(-1) \times (-2)} = \frac{1}{2}x^2 - \frac{1}{2}x, \\ \ell_1(x) &= \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{(x - (-1))(x - 1)}{(0 - (-1)) \times (0 - 1)} = -x^2 + 1, \\ \ell_2(x) &= \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{(x - (-1))(x - 0)}{(1 - (-1)) \times (1 - 0)} = \frac{1}{2}x^2 + \frac{1}{2}. \end{aligned}$$

Vì $P_2(x)$ có giá trị trùng với $y = 5^x$ tại các điểm x_0, x_1 và x_2 , nên ta có

$$\begin{aligned} P_2(x) &= y_0 \ell_0(x) + y_1 \ell_1(x) + y_2 \ell_2(x) \\ &= \frac{1}{5} \left(\frac{1}{2} x^2 - \frac{1}{2} x \right) + (-x^2 + 1) + 5 \left(\frac{1}{2} x^2 + \frac{1}{2} \right) \\ &= \frac{8}{5} x^2 + \frac{12}{5} x + 1. \end{aligned}$$

Đó là tam thức bậc hai cần tìm.

CAS 4.4. Dưới đây, chúng ta viết chương trình R minh họa quá trình tính toán giá trị của đa thức nội suy Lagrange và áp dụng vào bộ dữ liệu sau.

x	-1	0	1	2
y	-1	4	15	50

Quá trình tính được chia làm 2 bước: Đầu tiên, chúng ta tính giá trị của các đa thức Lagrange cơ sở. Sau đó, chúng ta tính giá trị của đa thức nội suy tùy ý bằng công thức nội suy Lagrange.

*# lagrange_base() tính giá trị của các đa thức nội suy Lagrange
cơ bản đối với các nút nội suy x tại một điểm xi cho trước.*

```
lagrange_base <- function(x, xi) {
  n = length(x)
  t = numeric(n) + 1

  # Hai vòng lặp lồng nhau tính giá trị của đa thức
  # nội suy Lagrange cơ sở
  for (i in 1:n) {
    for (j in 1:n) {
      if (j != i) {
        p = (xi - x[j]) / (x[i] - x[j])
        t[i] = t[i] * p
      }
    }
  }
  return(t)
}
```

lagrange.eval() tính giá trị của đa thức nội suy tổng quát.

```
lagrange.eval <- function(x, y, xi) {
  if (length(x) != length(y)) {
    stop("Lỗi! Số nút và số giá trị khác nhau.")
  }
}
```

```

l = lagrange.base(x, xi)

# Trả về tích vô hướng của y và l
return(sum(l * y))
}

# Ví dụ
x <- c(-1, 0, 1, 2)
y <- c(-1, 4, 15, 50)
xi <- 1.2
lagrange.eval(x, y, xi)

## [1] 19,504

```

Lưu ý, nếu l và y là hai vectơ trong R có độ dài bằng nhau thì $l * y$ là vectơ thu được bằng cách nhân các thành phần tương ứng của l và y , trong khi đó $\text{sum}(l * y)$ sẽ lấy tổng các thành phần của $l * y$.

Chú ý 4.2. Chúng ta có thể dùng `poly.calc()` có sẵn trong gói `polynom`⁴ của phần mềm R để tính toán các đa thức nội suy.

4.3 Đa thức nội suy Newton

Dưới đây, chúng ta tìm hiểu về công thức nội suy Newton cho đa thức nội suy. Trong trường hợp các mốc nội suy tùy ý, công thức nội suy Newton dựa trên các *tỉ sai phân*. Khi các mốc nội suy cách đều, chúng ta sẽ có công thức đẹp hơn, dựa trên các *sai phân*.

4.3.1 Đa thức nội suy Newton tổng quát

Nếu như trong công thức nội suy Lagrange, đa thức nội suy được viết dưới dạng một “tổ hợp tuyến tính” của các đa thức Lagrange cơ bản ℓ_j (trong (4.6)), thì trong công thức Newton, đa thức nội suy được viết dưới dạng một tổ hợp tuyến tính của các đa thức $1, (x - x_0), (x - x_0)(x - x_1), \dots$, và $(x - x_0)(x - x_1) \cdots (x - x_{n-1})$. Nói cách khác, công thức nội suy Newton sẽ có dạng

$$\begin{aligned}
 P_n(x) = & A_0 + A_1(x - x_0) + A_2(x - x_0)(x - x_1) \\
 & + \cdots + A_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}),
 \end{aligned} \tag{4.7}$$

⁴polynom: A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations, DOI: 10.32614/CRAN.package.polynom

với $A_0, A_1, \dots, A_n$ là các hệ số, chính là toạ độ của $P_n(x)$ trong một cơ sở của không gian vectơ các đa thức bậc $\leq n$.

Sau đây, chúng ta sẽ tính các hệ số $A_0, A_1, \dots, A_n$, dựa vào các giá trị đã biết y_i của P_n tại các giá trị x_i .

Trường hợp có ít điểm nội suy, ta có thể tính các hệ số này trực tiếp mà không gặp khó khăn gì. Ví dụ, bắt đầu bằng việc thay $x = x_0$ vào (4.16), ta thu được

$$y_0 = P_n(x_0) = A_0.$$

Tiếp tục, thay $x = x_1$ vào (4.16), ta thu được

$$y_1 = P_n(x_1) = A_0 + A_1(x_1 - x_0) = y_0 + A_1(x_1 - x_0).$$

Từ đây, ta tính được

$$A_1 = \frac{y_1 - y_0}{x_1 - x_0}. \quad (4.8)$$

Tương tự như trên, thay $x = x_2$ và các giá trị đã tính được của A_0 và A_1 vào (4.16), ta thu được

$$y_2 = P_n(x_2) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x_2 - x_0) + A_2(x_2 - x_0)(x_2 - x_1).$$

Vì x_0, x_1 và x_2 là khác nhau từng đôi nên $(x_2 - x_0)(x_2 - x_1) \neq 0$ và do đó, ta có thể giải được giá trị A_2 . Cụ thể hơn thì

$$\begin{aligned} y_2 &= y_0 + \frac{y_1 - y_0}{x_1 - x_0}[(x_2 - x_1) + (x_1 - x_0)] + A_2(x_2 - x_0)(x_2 - x_1) \\ &= y_1 + \frac{y_1 - y_0}{x_1 - x_0}(x_2 - x_1) + A_2(x_2 - x_0)(x_2 - x_1). \end{aligned}$$

Đẳng thức này tương đương với

$$y_2 - y_1 - \frac{y_1 - y_0}{x_1 - x_0}(x_2 - x_1) = A_2(x_2 - x_0)(x_2 - x_1)$$

Chia hai vế cho $(x_2 - x_0)(x_2 - x_1)$, ta thu được

$$\frac{y_2 - y_1}{(x_2 - x_0)(x_2 - x_1)} - \frac{y_1 - y_0}{(x_1 - x_0)(x_2 - x_0)} = A_2.$$

Đẳng thức cuối cùng có thể viết lại là

$$A_2 = \frac{1}{x_2 - x_0} \left(\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0} \right). \quad (4.9)$$

Quá trình này có thể tiếp tục để tính A_3, A_4 , v.v.

Nếu hệ các điểm $\{x_0, x_1, \dots, x_n\}$ là khác nhau từng đôi thì ta luôn giải được bộ giá trị $A_0, A_1, \dots, A_{n-1}$. Đó là bởi vì, theo công thức nội suy Lagrange, đa thức nội suy $P_n(x)$ luôn tồn tại và thuộc vào không gian vectơ các đa thức bậc không vượt quá n còn hệ các đa thức

$$N_n := \{1, x - x_0, (x - x_0)(x - x_1), \dots, (x - x_0)(x - x_1) \cdots (x - x_{n-1})\}$$

lập thành một hệ độc lập tuyến tính gồm $n + 1$ vectơ trong không gian $n + 1$ chiều chứa đa thức nội suy (không gian các đa thức của biến x với bậc không vượt quá n). Các đa thức trong N_n gọi là các *đa thức nội suy Newton*. Điều này có nghĩa là bộ các hệ số $(A_0, A_1, \dots, A_n)$ chính là toạ độ của $P_n(x)$ trong cơ sở N_n . Do đó, theo một kết quả cơ bản của đại số tuyến tính, bộ số này tồn tại và duy nhất. Vấn đề còn lại là tìm một thuật toán để tính các giá trị $A_3, A_4, \dots$ một cách hiệu quả.

Sau đây, chúng ta xây dựng quá trình tính các hệ số $A_k, k = 3, 4, \dots, n$. Nếu $y = y(x)$ thì *tỉ hiệu*⁵ của y tại các điểm x_i và x_j được định nghĩa là

$$y[x_i, x_j] := \frac{y_i - y_j}{x_i - x_j}.$$

Tỉ hiệu cấp hai tại các điểm x_i, x_j và x_k (theo thứ tự đó) là

$$y[x_i, x_j, x_k] := \frac{y[x_i, x_j] - y[x_j, x_k]}{x_i - x_k}, \quad (4.10)$$

và tỉ hiệu cấp ba tại các điểm x_i, x_j, x_k , và x_l là

$$y[x_i, x_j, x_k, x_l] := \frac{y[x_i, x_j, x_k] - y[x_j, x_k, x_l]}{x_i - x_l}. \quad (4.11)$$

Tỉ hiệu cấp cao hơn được xác định tương tự theo truy hồi.

Khái niệm tỉ hiệu cho ta thấy, từ các phương trình (4.8), (4.9), các hệ số A_1, A_2 chính là các tỉ hiệu cấp 1 và cấp 2:

$$A_1 = y[x_0, x_1], \quad A_2 = y[x_0, x_1, x_2].$$

Tiếp theo, ta sẽ chứng minh rằng các hệ số A_k được cho bởi tỉ hiệu cấp k , $k \geq 3$.

Bây giờ, xét trường hợp $y(x)$ là hàm đa thức bậc n : $y(x) = P_n(x)$. Tỉ hiệu của $P_n(x)$ tại điểm (x, x_0) , xem như là hàm của đối số x , là một đa thức bậc $n - 1$. Thật vậy, đa thức $Q(x) := P_n(x) - P_n(x_0)$ thỏa mãn $Q(x_0) = 0$, nên theo định lí Bézout⁶, $Q(x)$ chia hết cho $x - x_0$. Do đó, sau khi giản ước nhân tử $(x - x_0)$ ở tử thức và mẫu thức của tỉ hiệu

$$P_n[x, x_0] = \frac{P_n(x) - P_n(x_0)}{x - x_0},$$

⁵Tỉ hiệu (divided difference), một số sách còn gọi là *tỉ sai phân*.

⁶Étienne Bézout (1730–1783) là nhà toán học người Pháp.

chúng ta thu được một đa thức có bậc giảm đi 1, tức là có bậc $n - 1$.

Cũng như vậy, tỉ hiệu cấp hai của P_n là đa thức bậc $n - 2, \dots$. Tiếp tục như vậy, tỉ hiệu cấp n của P_n là đa thức bậc 0, tức là một hằng số, và tỉ hiệu cấp $n + 1$ của P_n bằng 0.

Ví dụ 4.6. Giả sử $P_n(x)$ là một đa thức bậc n , có dạng tổng quát

$$P_n(x) = a_0 + a_1x + \dots + a_nx^n.$$

Nếu $n = 0$ thì $P_n(x)$ là đa thức hằng số: $P_n(x) = a_0$, nên tỉ hiệu

$$P_n[x, x_0] = 0.$$

Nếu $n = 1$ thì $P_1(x) = a_0 + a_1x$. Do đó

$$P_1(x) - P_1(x_0) = (a_0 + a_1x) - (a_0 + a_1x_0) = a_1(x - x_0).$$

Vậy,

$$P_1[x, x_0] = \frac{P_1(x) - P_1(x_0)}{x - x_0} = \frac{a_1(x - x_0)}{x - x_0} = a_1.$$

Với $n = 2$, P_2 có dạng

$$P_2(x) = a_0 + a_1x + a_2x^2.$$

Khi đó,

$$\begin{aligned} P_2[x, x_0] &= \frac{(a_0 + a_1x + a_2x^2) - (a_0 + a_1x_0 + a_2x_0^2)}{x - x_0} \\ &= a_2x + a_2x_0 + a_1. \end{aligned}$$

Tiếp tục tính, chúng ta thu được

$$P_2[x, x_0, x_1] = a_2.$$

Với $n = 1, 2, 3$, tỉ hiệu cấp n , $P_n[x, x_0, \dots, x_{n-1}]$, của một đa thức cấp n là hằng số, không phụ thuộc x . Bằng quy nạp, có thể dễ dàng chứng minh được điều này đúng với mọi $n \geq 1$. Chi tiết dành cho bạn đọc.

Từ định nghĩa của tỉ hiệu, chúng ta suy ra

$$\begin{aligned} P_n(x) &= P_n(x_0) + (x - x_0)P_n[x, x_0], \\ P_n[x, x_0] &= P_n[x_0, x_1] + (x - x_1)P_n[x, x_0, x_1], \\ &\dots = \dots \\ P_n[x, x_0, \dots, x_{n-1}] &= P_n[x_0, x_1, \dots, x_n] \end{aligned}$$

vì $P_n[x, x_0, x_1, \dots, x_{n-1}]$ là một hằng số. Thay liên tiếp các đẳng thức này từ dưới lên, chúng ta thu được kết quả sau:

Định lí 4.4: Công thức nội suy Newton 1

Nếu $P_n(x)$ là một đa thức bậc n và $x_0, x_1, \dots, x_n$ là $n+1$ điểm phân biệt thì

$$P_n(x) = P_n(x_0) + \sum_{i=1}^n P_n[x_0, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j). \quad (4.12)$$

Rõ ràng, các thành phần $P_n[x_0, \dots, x_i]$ trong vế phải của (4.13) có thể tính chỉ dựa vào các giá trị $P_n(x_j)$, $j = 0, 1, \dots, i$. Từ đó, chúng ta suy ra công thức nội suy Newton như sau:

Định lí 4.5: Công thức nội suy Newton 2

Nếu $x_0, x_1, \dots, x_n$ là $n+1$ điểm phân biệt và $y_0 = y(x_0), y_1 = y(x_1), \dots, y_n = y(x_n)$ là một bộ $n+1$ giá trị tương ứng của y thì

$$p_n(x) := y_0 + \sum_{i=1}^n y[x_0, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j) \quad (4.13)$$

là đa thức nội suy của y tại các nút x_i .

Ví dụ 4.7. Giả sử ta biết giá trị của hàm số $y = f(x)$ tại một số giá trị của đối số x như sau.

x	1	2	4	5
y	25	36	106	177

Tính đa thức nội suy $p_3(x)$ của y tại các nút nội suy bằng cách dùng công thức nội suy Newton tiến.

Lời giải

Kí hiệu $x_0, \dots, x_3$ là các điểm nội suy và $y_0, \dots, y_3$ là các giá trị của y tại các điểm nội suy đó. Ta tính các tỉ hiệu và viết vào bảng sau:

Nút	x	y	Tỉ hiệu 1	Tỉ hiệu 2	Tỉ hiệu 3
0	1	25	11	8	1
1	2	36	35	12	...
2	4	106	71	...	...
3	5	177	...	...	...

Ở hàng thứ hai từ cột thứ 3, chúng ta thu được $y_0, y[x_0, x_1], y[x_0, x_1, x_2], \dots$. Vậy, đa thức nội suy Newton của hàm số y tại các nút nội suy trên là

$$p_3(x) = (x-4)(x-2)(x-1) + 8(x-2)(x-1) + 11(x-1) + 25. \quad (4.14)$$

Lưu ý, chúng ta có thể đơn giản hóa biểu thức ở vế phải, chúng ta thu được

$$p_3(x) = x^3 + x^2 + x + 22.$$

Tuy nhiên, quá trình này không cần thiết cho việc tính giá trị của đa thức nội suy.

CAS 4.5. Hãy viết một chương trình R tính các giá trị của bảng tỉ hiệu trong phương pháp nội suy Newton và áp dụng vào các dữ liệu trong Ví dụ 4.7.

Kết quả thu được là bảng tỉ hiệu khớp với những tính toán bằng tay trong Ví dụ 4.7.

```
# Chương trình tính các tỉ hiệu của x và y
newton_interpolation <- function(x, y) {
  n <- length(x)
  if (n == 1) {
    stop("Lỗi! Cần ít nhất hai điểm dữ liệu.")
  }
  else {
    # Tạo một ma trận A để lưu trữ các tỉ hiệu

    A <- matrix(data = 0, n, n - 1)

    # Tính các tỉ hiệu và gán vào các yếu tố của A

    for (j in 1:(n - 1))
      for (i in 1:(n - j)) {
        if (j == 1) {
          A[i, j] = (y[i + 1] - y[i]) / (x[i + 1] - x[i])
        } else {
          dy = A[(i + 1), (j - 1)] - A[(i), (j - 1)]
          dx = x[i + j] - x[i]
          A[i, j] = dy / dx
        }
      }
  }

  # Đặt tên cho các cột tỉ hiệu

  colnames(A) <- as.list(seq(1, n - 1, by = 1))
  return(cbind(x, y, A))
}
```

```

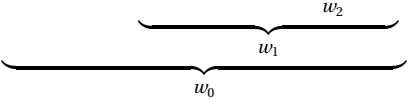
    }
}
# Chạy thử newton_interpolation() với dữ liệu trong ví dụ trên.

x <- c(1, 2, 4, 5)
y <- c(25, 36, 106, 177)
newton_interpolation(x, y)

##      x    y    1    2    3
## [1,] 1    25   11    8    1
## [2,] 2    36   35   12    0
## [3,] 4   106   71    0    0
## [4,] 5   177    0    0    0

```

Chú ý 4.3. Thay vì viết kết quả của nội suy Newton về dạng “tiêu chuẩn”, tức là dạng tổng các đơn thức của x , sắp xếp theo thứ tự tăng hoặc giảm của bậc của x , ta có thể viết nó ở dạng “lồng nhau” (nested form). Khi đó, việc tính giá trị của đa thức rất hiệu quả khi sử dụng lược đồ Horner (sửa đổi). Chẳng hạn đối với đa thức $p_3(x)$ trong (4.14),

$$p_3(x) = 25 + (x-1)(11 + (x-2)(8 + (x-4)))$$


Việc tính giá trị của đa thức sẽ được thực hiện tương tự như lược đồ Horner. Ví dụ, hãy tính $p_3(2,5)$ như sau:

$$\begin{aligned}
 w_3 &= 1 \\
 w_2 &= 8 + w_3(2,5-4) \\
 w_1 &= 11 + w_2(2,5-2) \\
 w_0 &= 25 + w_1(2,5-1).
 \end{aligned}$$

Tính toán cụ thể,

$$w_2 = 6,5, \quad w_1 = 14,25, \quad p_3(2,5) = w_0 = 46,375.$$

Chúng ta sẽ lập chương trình tự động tính các kết quả trên trong CAS 4.7.

Chú ý 4.4. Biểu diễn đa thức nội suy theo về phải của (4.13) gọi là đa thức nội suy Newton tiến. Bằng cách hoán đổi vai trò của các nút nội suy, ta có thể đưa ra các biểu diễn khác nhau. Nói riêng, nếu ta có *đa thức nội suy Newton lùi* như sau:

$$p_n(x) := y_n + \sum_{i=1}^n y[x_n, \dots, x_{n-i}] \prod_{j=0}^{i-1} (x - x_{n-j}). \quad (4.15)$$

4.3.2 Đa thức nội suy Newton với các nút nội suy cách đều

Khi các nút nội suy *cách đều* nhau, ta có thể biểu diễn đa thức nội suy Newton một cách đơn giản hơn. Bằng cách đặt

$$x_j = x_0 + jh, \quad j = 0, 1, 2, \dots,$$

với h là khoảng cách giữa hai nút nội suy liên tiếp, và

$$x = x_0 + th,$$

ta có

$$x - x_j = (x_0 + th) - (x_0 + jh) = h(t - j).$$

Vậy,

$$\prod_{j=0}^{i-1} (x - x_j) = h^i \prod_{j=0}^{i-1} (t - j).$$

Tiếp theo, ta biểu diễn các tỉ hiệu qua các sai phân như sau: *Sai phân tiến* (forward difference) của dãy y_j là

$$\Delta y_j = y_{j+1} - y_j, \quad j = 0, 1, 2, \dots, n-1.$$

Sai phân tiến cấp 2 của là

$$\Delta^2 y_j = \Delta(\Delta y_j) = \Delta y_{j+1} - \Delta y_j, \quad j = 0, 1, 2, \dots, n-2.$$

Rõ ràng

$$\Delta^2 y_j = (y_{j+2} - y_{j+1}) - (y_{j+1} - y_j) = y_{j+2} - 2y_{j+1} + y_j.$$

Bằng quy nạp, ta có thể chứng minh được

$$\Delta^k y_j = \sum_{i=0}^j (-1)^i \binom{k}{i} y_{j+k-i}, \quad j = 0, 1, \dots, n-k.$$

Ở đây, $\binom{k}{i}$ là tổ hợp chập i của k phần tử: $\binom{k}{i} = \frac{k!}{i!(k-i)!}$. Các tỉ hiệu được biểu diễn qua sai phân như sau

$$y[x_0, x_1, \dots, x_k] = \frac{\Delta^k y_0}{(k!)h^k}.$$

Thay các biểu thức của tỉ hiệu này vào công thức nội suy Newton tiến, ta thu được định lí sau:

Định lí 4.6: Nội suy Newton với mốc cách đều

Nếu $x_0, x_1, \dots, x_n$ là $n + 1$ điểm phân biệt cách đều nhau:

$$x_j = x_0 + jh, \quad j = 0, 1, 2, \dots,$$

và $y_0 = y(x_0), y_1 = y(x_1), \dots, y_n = y(x_n)$ là một bộ $n + 1$ giá trị tương ứng của y thì, với $x = x_0 + th$, ta có

$$p_n(x) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \dots + \frac{t(t-1)\dots(t-n+1)}{n!}\Delta^n y_0 \quad (4.16)$$

là đa thức nội suy của y tại các nút x_i .

Công thức (4.16) còn được gọi là công thức Newton–Gregory.⁷

Ví dụ 4.8. Giả sử ta biết giá trị của hàm số $y = f(x)$ tại một số giá trị của đối số x như sau.

x	1	2	3	4	5
y	2018	1992	1902	1682	1242

Tính đa thức nội suy của y tại các nút nội suy bằng cách dùng công thức nội suy Newton tiến với các nút cách đều.

Lời giải

Kí hiệu $x_0, \dots, x_3$ là các điểm nội suy và $y_0, \dots, y_3$ là các giá trị của y tại các điểm nội suy đó. Ta tính các tỉ hiệu và viết vào bảng sau:

i	x_i	y_i	Sai phân 1	Sai phân 2	Sai phân 3	Sai phân 4
0	1	2018	−26	−64	−66	−24
1	2	1992	−90	−130	−90	...
2	3	1902	−220	−220	...	...
3	4	1682	−440	...	...	...
4	5	1242	...	...	...	...

Ở hàng thứ hai từ cột thứ 3, ta thu được các sai phân hữu hạn $\Delta y_0, \Delta^2 y_0$,

⁷James Gregory FRS (tháng 11 năm 1638 – tháng 10 năm 1675) là một nhà toán học và thiên văn học người Scotland.

.... Vậy, đa thức nội suy Newton của hàm số y tại các nút nội suy trên là

$$p_4(1+t) = 2018 + t(-26) + (-64)\frac{t(t-1)}{2} + (-66)\frac{t(t-1)(t-2)}{6} - 24\frac{t(t-1)(t-2)(t-3)}{24}.$$

Đó là đa thức nội suy Newton cần tính, viết dưới dạng biến $t = x - 1$. Lưu ý, cũng có thể viết lại đa thức nội suy là

$$p_4(1+t) = -t^4 - 5t^3 - 10t^2 - 10t + 2018,$$

và đổi biến $x = 1 + t$ hay $t = x - 1$, để thu được

$$p_4(x) = -x^4 - x^3 - x^2 - x + 2022.$$

Tuy nhiên, việc này không cần thiết để tính giá trị của đa thức nội suy.

CAS 4.6. Chúng ta lập hàm `newton_coefficients()` nhận vào một dãy (sắp thứ tự) các giá trị $y = (y_i)$, là những giá trị của hàm số cần nội suy tại những mốc nội suy cách đều. Kết quả trả về là dãy các sai phân trong công thức nội suy Newton.

```
# Hàm newton_coefficients() tính các hệ số trong đa thức
# nội suy Newton với mốc cách đều: y là dãy các giá trị
# nội suy với độ dài ít nhất 2.

newton_coefficients <- function(y) {
  n <- length(y)
  t <- numeric(n)
  t[1] <- y[1]
  for (i in 2:n) {
    y <- diff(y)      # Sử dụng diff() có sẵn
    t[i] <- y[1]
  }
  return(t)
}

# Chạy thử newton_coefficients() với dữ liệu trong ví dụ trên.

y <- c(2018, 1992, 1902, 1682, 1242)
print(newton_coefficients(y))

## [1] 2018 -26 -64 -66 -24
```

Kết quả thu được là dãy các hệ số như trong Ví dụ 4.8.

Mục CAS sau đây kết hợp phương pháp nội suy Newton và lược đồ Horner (mở rộng) để tính giá trị của đa thức nội suy Newton.

CAS 4.7. Hãy tính các hệ số của đa thức nội suy Newton (hàng đầu tiên trong bảng tỉ sai phân) của dữ liệu trong Ví dụ 4.7 và tính giá trị của đa thức nội suy thu được tại $x = a = 2.5$.

```
# Hàm newton_interpolation() tính các tỉ hiệu của hai dãy x, y.
# Tham số x, y là hai dãy có cùng độ dài,
# a là điểm cần tính giá trị.

newton_interpolation <- function(x, y, a) {
  n <- length(x)

  # Dãy t lưu trữ các tỉ hiệu trên hàng đầu tiên.

  t <- numeric(n)
  t[1] <- y[1]
  for (i in 2:n) {
    for (j in 1:(n - i + 1)) {
      z <- diff(y)
      y[j] <- z[j] / (x[j + i - 1] - x[j])
    }
    t[i] <- y[1]
  }

  # Tính giá trị của đa thức tại x = a.

  w <- t[n]
  while (n > 1) {
    n <- n - 1
    w <- t[n] + w * (a - x[n])
  }
  return(list(coefficients = t, value = w))
}

# Chạy thử newton_interpolation()
x <- c(1, 2, 4, 5)
y <- c(25, 36, 106, 177)
a <- 2.5
newton_interpolation(x, y, a)

## $coefficients
## [1] 25 11 8 1
##
## $value
```

[1] 46, 375

Các kết quả thu được trùng khớp với các kết quả đã tính bằng tay.

4.3.3 Đọc thêm: Định lí Hermite–Gennochi

Trong tiểu mục này, ta nói về định lí Hermite–Gennochi⁸ cho một biểu diễn tích phân của các tỉ sai phân. Biểu diễn này cho phép ta mở rộng khái niệm tỉ sai phân của các hàm số khả vi khi các điểm nội suy trùng nhau.

Định lí 4.7: Hermite–Gennochi

Cho $x_0, x_1, \dots, x_n \in (a, b)$ và hàm số $f(x)$ khả vi n lần trên (a, b) . Khi đó,

$$f[x_0, x_1, \dots, x_n] = \int \cdots \int_T f^{(n)}(t_0 x_0 + \cdots + t_n x_n) dt_1 \dots dt_n, \quad (4.17)$$

trong đó, miền lấy tích phân bội là

$$T = \left\{ (t_1, \dots, t_n): t_i \geq 0, \sum_{i=1}^n t_i \leq 1 \right\} \subset \mathbb{R}^n,$$

$$\text{và } t_0 = 1 - \sum_{i=1}^n t_i.$$

Trường hợp $n = 1$, công thức Hermite–Gennochi (4.17) có thể dễ dàng suy ra từ công thức tích phân Newton–Leibniz: Với $x_1 - x_0 \neq 0$, đổi biến $u = (1 - t_1)x_0 + t_1 x_1$, $du = \frac{1}{x_1 - x_0} dt_1$, ta có

$$\begin{aligned} \int_T f'(t_0 x_0 + t_1 x_1) dt_1 &= \int_0^1 f'((1 - t_1)x_0 + t_1 x_1) dt_1, \\ &= \frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f'(u) du \\ &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\ &= f[x_0, x_1]. \end{aligned}$$

Với $n = 2$, công thức có thể suy ra từ định lí Fubini cho tích phân kép cùng

⁸Angelo Gennochi (5/3/1817-7/3/1889) là một nhà toán học người Italy. Công thức Hermite–Gennochi được cho là xuất hiện trong công bố “Relation entre la différence et la dérivée d’un même ordre quelconque, *Archiv Math. Phys.* (I) 49:342-345, 1869” của ông.

với công thức Newton–Leibniz. Về chứng minh công thức này trong trường hợp tổng quát, bạn đọc có thể xem Atkinson [1], trang 145-146.

Trong định nghĩa tỉ sai phân, người ta yêu cầu các điểm $x_0, x_1, \dots, x_n$ là phân biệt. Tuy nhiên, về phía của công thức Hermite–Genocchi (4.17) hoàn toàn xác định mà không cần điều kiện này. Do đó, công thức này cho phép chúng ta mở rộng tỉ sai phân cho các hàm số khả vi n lần.

Một hệ quả trực tiếp của công thức (4.17) là khẳng định sau: Nếu f là một đa thức bậc nhỏ hơn n thì

$$f[x_0, x_1, \dots, x_n] = 0.$$

4.4 Sai số nội suy

Trong mục này, chúng ta tìm hiểu vấn đề đánh giá sai số trong nội suy, vấn đề chọn mốc nội suy “tối ưu”, và một ví dụ về hiện tượng Runge. Tài liệu tham khảo: T. V. Đình [6, §4.1.7], P. K. Anh [2, §II.3–II.4].

4.4.1 Đánh giá sai số trong xấp xỉ bằng đa thức nội suy

Trong mục này, ta trình bày phương pháp tính gần đúng giá trị của hàm số bằng đa thức nội suy. Phương pháp này được mô tả như sau: Giả sử $f(x)$ là một hàm số mà ta muốn tính giá trị của nó tại một số điểm trong đoạn $[a, b]$. Khi công thức của $f(x)$ quá phức tạp (như đối với các hàm siêu việt $\sin x, \cos x$, hoặc a^x), ta có thể xấp xỉ giá trị $f(x)$ bằng giá trị $P_n(x)$ với P_n là một đa thức nội suy của f tại một số điểm $x_0, x_1, \dots, x_n$, sao cho các giá trị tại các nút đó có thể được tính dễ dàng. Do đó, nảy sinh vấn đề đánh giá sai số của phép nội suy, tức là đánh giá phần dư

$$r_n(x) := f(x) - P_n(x). \quad (4.18)$$

Định lý 4.8: Đánh giá sai số nội suy

Nếu hàm số $f(x)$ liên tục trên $[a, b]$ và có đạo hàm đến cấp $n + 1$ trong (a, b) , thì

$$r_n(x) = f^{(n+1)}(c) \frac{\omega(x)}{(n+1)!}, \quad (4.19)$$

với $c \in [a, b]$ là một điểm phụ thuộc vào x và

$$\omega(x) = \prod_{i=0}^n (x - x_i) = (x - x_0)(x - x_1) \cdots (x - x_n). \quad (4.20)$$

Định lí 4.8 cho phép ta đánh giá sai số tuyệt đối của phép tính gần đúng $f(x)$ bằng giá trị của đa thức nội suy $P_n(x)$. Cụ thể hơn, giả sử $f^{(n+1)}(x)$ bị chặn trên $[a, b]$:

$$|f^{(n+1)}(x)| \leq M, \quad a < x < b.$$

Khi đó,

$$|r_n(x)| \leq \frac{M}{(n+1)!} |\omega(x)|. \quad (4.21)$$

Như vậy, sai số tuyệt đối được đánh giá qua vế phải của (4.21) và được phân tích làm 2 phần. Phần thứ nhất, hằng số M , chỉ phụ thuộc vào hàm f , và phần thứ hai, $|\omega(x)|$, chỉ phụ thuộc vào các nút nội suy.

Chứng minh Định lí 4.8. Cố định x , $x \neq x_j$ với mọi $j \in \{0, 1, \dots, n\}$, xét một biến t mới, và thiết lập một hàm phụ như sau:

$$G_x(t) = f(t) - P_n(t) - R(x)\omega(t).$$

Trong đó, $R(x)$ không phụ thuộc vào t và được chọn sao cho $G_x(x) = 0$. Nói cách khác,

$$R(x) = \frac{f(x) - P_n(x)}{\omega(x)}.$$

Rõ ràng, $G_x(t)$ khả vi tới cấp $n+1$. Mặt khác $G_x(x_j) = 0$, với $j = 0, 1, \dots, n$, và $G_x(x) = 0$. Bằng cách áp dụng liên tiếp định lí Rolle, ta có $G'_x(t)$ triệt tiêu tại $n+1$ điểm phân biệt, $G''_x(t)$ triệt tiêu tại n điểm phân biệt, $\dots$, $G_x^{(n+1)}(c) = 0$ với $c \in (a, b)$.

Mặt khác, bằng cách tính trực tiếp, ta có

$$G_x^{(n+1)}(t) = f^{(n+1)}(t) - R(x)(n+1)! \quad \forall t \in [a, b].$$

Thay $t = c$, ta có

$$0 = G_x^{(n+1)}(c) = f^{(n+1)}(c) - R(x)(n+1)!.$$

Từ đây, ta giải được

$$R(x) = \frac{f^{(n+1)}(c)}{(n+1)!}.$$

Thay vào công thức của $G_x(t)$, và thay $t = x$, nhớ rằng $G_x(x) = 0$, ta dễ dàng thu được (4.19).

Trường hợp $x = x_j$ nào đó thì có thể chọn c bất kì; (4.19) đúng vì hai vế đều bằng 0. □

Ví dụ 4.9. Giả sử đã tính được gần đúng hàm số $y = \sin(x)$ tại một số giá trị của x như sau (xem CAS 4.3):

x_i	0,1	0,2	0,3	0,4
$y_i = \sin(x_i)$	0,09983	0,19867	0,29552	0,38942

- (a) Lập đa thức nội suy bậc 3, kí hiệu $P_3(x)$, của $\sin(x)$ tại các nút nội suy nêu trên.
- (b) Tính $P_3(0,25)$ và so sánh với $\sin(0,25)$.

Lời giải

- (a) Ta lập đa thức nội suy $P_3(x)$ bằng công thức Newton tiến, bắt đầu từ $x_0 = 0,1$ và các nút cách đều với khoảng cách $h = 0,1$. Dễ dàng tính được các sai phân

i	x_i	y_i	Sai phân 1	Sai phân 2	Sai phân 3
0	1	0,09983	0,09884	-0,00199	-0,00096
1	2	0,19867	0,09685	-0,00295	...
2	3	0,29552	0,09390	...	...
3	4	0,38942	...	...	...

Từ công thức Newton, ta suy ra, với $x = 0,1 + 0,1t$ hay $t = 10x - 1$,

$$P_3(x) = 0,09983 + 0,09884t + \frac{t(t-1)}{2}(-0,00199) + \frac{t(t-1)(t-2)}{6}(-0,00096).$$

- (b) Để tính $P_3(x)$ tại $x = 0,25$, ta thay $t = 10x - 1 = 1,5$, ta có

$$\begin{aligned} P_3(0,25) &= 0,09983 + 0,09884 \times 1,5 - \frac{1,5(1,5-1)}{2} \times 0,00199 \\ &\quad - \frac{1,5(1,5-1)(1,5-2)}{6} \times 0,00096 \\ &= 0,24740375. \end{aligned}$$

Vì P_3 là đa thức nội suy của hàm $y = \sin(x)$ nên ta có thể coi

$$\sin(0,25) \approx P_3(0,25) = 0,24740375.$$

Để đánh giá sai số, ta xét

$$\omega(x) = (x-0,1)(x-0,2)(x-0,3)(x-0,4).$$

Rõ ràng,

$$\omega(0,25) = 0,00005625.$$

Do hàm $y = \sin(x)$ có đạo hàm mọi cấp nhận giá trị trong đoạn $[-1, 1]$, nên ta có thể ước lượng

$$|\sin(0,25) - P_3(0,25)| \leq |r_3(0,25)| \leq \frac{1}{4!} \omega(0,25) < 2,35 \times 10^{-6}.$$

Như vậy, ta tính được $\sin(0,25)$ chính xác tới vài phần triệu.

4.4.2 Chọn nút nội suy tối ưu: đa thức Chebyshev

Từ Định lý 4.8, nảy sinh ra vấn đề chọn các điểm nội suy x_i sao cho $\max|\omega(x)|$ bé nhất. Để tìm hiểu vấn đề này, để đơn giản, ta xét trường hợp riêng khi đoạn $[a, b]$ là đoạn $[-1, 1]$. Trong trường hợp này, các nút nội suy “tối ưu” này là các nghiệm của đa thức Chebyshev⁹. Các đa thức Chebyshev loại 1 bậc m , kí hiệu T_m , là đa thức duy nhất thỏa mãn

$$T_m(\cos y) = \cos(my), \quad y \in \mathbb{R}.$$

Như vậy, nếu $x \in [-1, 1]$ thì

$$T_m(x) = \cos(m \arccos(x)).$$

Ví dụ, với $m = 1$ thì $T_1(x)$ thỏa mãn

$$T_1(\cos(y)) = \cos(y),$$

với mọi $y \in \mathbb{R}$, nên phải có

$$T_1(x) = x.$$

Với $m = 2$, $T_2(x)$ phải thỏa mãn

$$T_2(\cos(y)) = \cos(2y) = 2\cos^2(y) - 1,$$

với mọi $y \in \mathbb{R}$, và vậy thì

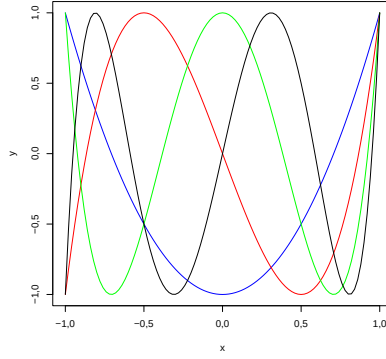
$$T_2(x) = 2x^2 - 1.$$

Với $m \geq 3$ tùy ý, công thức tường minh T_m thu được từ việc khai triển $\cos(my)$ theo $\cos(y)$.

Nghiệm của đa thức Chebyshev $T_m(x)$ có dạng $\cos(y_i)$, với y_i là nghiệm của phương trình lượng giác $\cos(my) = 0$. Từ đó ta thấy rằng các nghiệm này là

$$x_i = \cos\left(\frac{2i+1}{m} \cdot \frac{\pi}{2}\right),$$

⁹Pafnuty Lvovich Chebyshev (1821 – 1894) là một nhà toán học người Nga.



Hình 4.1: Các đa thức Chebyshev bậc 2, 3, 4 và 5

với $i = 0, 1, 2, \dots, m-1$. Khi đó, với

$$\omega_m(x) = (x - x_0)(x - x_1) \cdots (x - x_{m-1}), \quad (4.22)$$

thì $\omega_m(x)$ và $T_m(x)$ sai khác một nhân tử. Do $\omega_m(x)$ có hệ số đầu tiên bằng 1 nên nhân tử đó chính là hệ số đầu tiên của $T_m(x)$, tức là 2^{m-1} . Từ đó ta có thể thấy rằng

$$\omega_m(x) = \frac{1}{2^{m-1}} T_m(x).$$

Theo trên, rõ ràng $|T_m(x)| \leq 1$ với mọi $x \in [-1, 1]$ nên ta có

$$|\omega_m(x)| \leq \frac{1}{2^{m-1}}. \quad (4.23)$$

Dấu “=” xảy ra khi và chỉ khi $x = \cos(j\pi/m)$, $j \in \{1, 2, \dots, m-1\}$.

Định lí 4.9: Chebyshev

Giả sử $P(x)$ là một đa thức bậc m với hệ số đầu bằng 1:

$$P(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0.$$

Khi đó, xảy ra bất đẳng thức

$$\max_{x \in [-1, 1]} |P(x)| \geq 2^{-m+1}. \quad (4.24)$$

Chứng minh. Xét đa thức

$$R(x) = P(x) - \omega_m(x).$$

Nếu $R(x) = 0$, có nghĩa là $P(x) = \omega_m(x)$, thì dễ thấy dấu bằng xảy ra trong (4.24). Giả sử $R(x) \neq 0$. Do $P(x)$ và $\omega_m(x)$ cùng có hệ số đầu (hệ số của x^m) bằng 1, nên $R(x)$ có bậc không vượt quá $m-1$.

Tại các điểm

$$c_j = \cos\left(\frac{j\pi}{m}\right), \quad j = 0, 1, \dots, m-1,$$

ta có thể tính toán trực tiếp

$$T_m(c_j) = \cos(m \arccos(c_j)) = \cos(j\pi) = (-1)^j.$$

Vậy, ta có

$$\omega(c_j) = 2^{-m+1} T_m(c_j) = (-1)^j 2^{-m+1}.$$

Ta sẽ chứng minh bất đẳng thức sau (mạnh hơn (4.24))

$$\max_{j \in \{0, 1, \dots, m-1\}} |P(c_j)| \geq 2^{-m+1}. \quad (4.25)$$

Thật vậy, ta thấy rằng đoạn $[-1, 1]$ được chia thành m đoạn liên tiếp bởi các điểm chia

$$-1 = c_m < c_{m-1} < c_{m-2} < \dots < c_1 < c_0 = 1.$$

Do $R(x)$ có bậc không vượt quá $m-1$, $R(x) \neq 0$, nên $R(x)$ có nhiều nhất $m-1$ nghiệm trong $[-1, 1]$. Từ đó suy ra tồn tại một khoảng $[c_k, c_{k-1})$ sao cho $R(x)$ không có nghiệm trên đó và vậy thì $R(x)$ giữ nguyên dấu trên $[c_k, c_{k-1})$. Xảy ra hai trường hợp:

- *Trường hợp 1:* $R(x) > 0$ trên $[c_k, c_{k-1})$. Từ đó suy ra $P(x) \geq \omega_m(x)$ trên $[c_k, c_{k-1}]$. Nói riêng

$$\begin{aligned} \max\{|P(c_k)|, |P(c_{k-1})|\} &\geq \max\{P(c_k), P(c_{k-1})\} \\ &\geq \max\{\omega_m(c_k), \omega_m(c_{k-1})\} \\ &= \max\{2^{-m+1}(-1)^k, 2^{-m+1}(-1)^{k-1}\} \\ &= 2^{-m+1}. \end{aligned}$$

- *Trường hợp 2:* $R(x) < 0$ trên $[c_k, c_{k-1})$. Từ đó suy ra $P(x) \leq \omega_m(x)$ trên $[c_k, c_{k-1}]$. Nói riêng,

$$\begin{aligned} \min\{P(c_k), P(c_{k-1})\} &\leq \min\{\omega_m(c_k), \omega_m(c_{k-1})\} \\ &= \min\{2^{-m+1}(-1)^k, 2^{-m+1}(-1)^{k-1}\} \\ &= -2^{-m+1}. \end{aligned}$$

Từ đó, dễ dàng suy ra

$$\max\{|P(c_k)|, |P(c_{k-1})|\} \geq 2^{-m+1}.$$

Vậy, trong cả hai trường hợp, ta có

$$\max_{x \in [-1, 1]} |P(x)| \geq \max_{j \in \{0, 1, \dots, m-1\}} |P(c_j)| \geq \max\{|P(c_k)|, |P(c_{k-1})|\} \geq 2^{-m+1},$$

và từ đó suy ra (4.25). Ta hoàn thành chứng minh. $\square$

Định lí 4.10: Chebyshev

Giả sử $P(x)$ là một đa thức bậc m với hệ số đầu bằng 1:

$$P(x) = x^m + a_{m-1}x^{m-1} + \dots + a_1x + a_0.$$

Giả sử

$$\max_{x \in [-1, 1]} |P(x)| = 2^{-m+1}. \quad (4.26)$$

Thì $P(x) = \omega_m(x)$.

Định lí này cơ bản nói rằng, trong tập hợp các đa thức bậc m với hệ số đầu bằng 1, đa thức $2^{-m+1} T_m(x)$ có chuẩn L^∞ bé nhất. Ngoài ra, có thể chứng minh được¹⁰ nó là đa thức duy nhất có trong đó có tính chất như vậy. Ta không đi sâu vào chứng minh này.

Xét bài toán nội suy các hàm số xác định trên $[-1, 1]$. Nếu chọn $n+1$ mốc nội suy là các không điểm của đa thức Chebyshev $T_{n+1}(x)$ thì ta có đánh giá sai số “tối ưu” sau đây:

Định lí 4.11: Sai số nội suy

Nếu hàm số $f(x)$ liên tục trên $[-1, 1]$ và có đạo hàm đến cấp $n+1$ liên tục trong $(-1, 1)$, và nếu P_n là đa thức nội suy của f tại các không điểm của đa thức Chebyshev $T_{n+1}(x)$ thì

$$|f(x) - P_n(x)| \leq 2^{-n} \max_{y \in [-1, 1]} |f^{(n+1)}(y)| \quad (4.27)$$

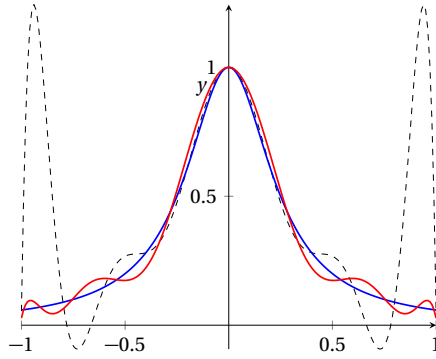
Chú ý 4.5. Trong trường hợp tổng quát $[a, b]$ không phải là đoạn $[-1, 1]$ thì phép đổi biến

$$\tilde{x} = \frac{1}{b-a}(2x - a - b)$$

đưa vấn đề đang xét về đoạn $[-1, 1]$.

Ta kết thúc mục này bằng một ví dụ về một tình huống xấp xỉ một hàm số (hữu tỉ) bởi đa thức nội suy với mốc cách đều và với mốc nội

¹⁰Đó là một kết quả của Borel, 1905.



Hình 4.2: Đồ thị hàm số $y = (1 + 16x^2)^{-1}$ và các đa thức nội suy của nó

suy Chebyshev. Ví dụ cho thấy, nhìn chung, nội suy với các mốc nội suy Chebyshev cho “sai số đều” nhỏ hơn. Nó cũng minh họa một tình huống mà xấp xỉ hàm số bằng nội suy có thể cho sai số rất lớn.

Ví dụ 4.10 (Hiện tượng Runge). Xấp xỉ hàm hữu tỉ

$$f(x) = \frac{1}{1 + 16x^2}$$

trên đoạn $[-1, 1]$ bằng đa thức nội suy với các nút nội suy cho bởi

- (a) các điểm cách đều $x_0 = -1, x_1 = -0,8, x_2 = -0,6, \dots, x_9 = 0,8, x_{10} = 1$;
- (b) các không điểm của đa thức Chebyshev $T_{11}(x)$.

Trong tình huống (a), ta có được đa thức nội suy xấp xỉ bằng

$$P_{10}(x) = 1 - 12,7196x^2 + 83,4829x^4 - 247,244x^6 + 314,581x^8 - 139,041x^{10}.$$

Kết quả này có thể thu được bằng tính toán trực tiếp, hoặc tính toán dưới sự trợ giúp của phần mềm. Mặt khác, trong tình huống (b), nội suy bằng các không điểm của đa thức Chebyshev, ta thu được

$$P_{10}^{\text{Cheb}}(x) = 1 - 10,1903x^2 + 46,8493x^4 - 98,8976x^6 + 95,0685x^8 - 33,8021x^{10}.$$

Đồ thị của $P_{10}^{\text{Cheb}}(x)$ (màu đỏ) và $P_{10}(x)$ (màu xanh) và đồ thị của hàm $1/(1 + 16x^2)$ được thể hiện trong Hình 4.2.

4.5 Nội suy hàm ghép tron

Trong mục này, chúng ta tìm hiểu về nội suy hàm ghép tron với trọng tâm là hàm ghép tron bậc ba (cubic spline), ứng dụng của hàm ghép tron để tính xấp xỉ giá trị của hàm số. Tài liệu tham khảo: P. K. Anh [2, §II.9], Kreyszig [9, §19.3].

4.5.1 Sơ lược về hàm ghép tron

Giả sử y phụ thuộc vào x theo một hàm đa thức bậc n . Chúng ta đã biết trong bài toán nội suy đa thức, nếu xác định được chính xác giá trị $y_j = y(x_j)$, $j = 1, 2, \dots, m$, tại m điểm phân biệt $x_1, x_2, \dots, x_m$, với $m \geq n + 1$, thì chúng ta có thể xác định được công thức của y theo x . Tuy nhiên, do các giá trị y_j có thể chỉ là giá trị gần đúng, nên kết quả của phép nội suy cũng chỉ là gần đúng. Hơn nữa, nếu $m > n + 1$ thì những sai số của y_j dẫn đến kết quả thu được của quá trình nội suy có thể chứa các số hạng với số mũ lớn hơn n , dẫn đến những sai số không chấp nhận được. Điều đó cho thấy khi số dữ liệu đã lớn hơn bậc của đa thức thì việc tăng thêm số lượng các dữ kiện đầu vào không làm kết quả tốt hơn.

Sau đây, chúng ta nói về một phương pháp nội suy trên các khoảng nhỏ và ghép các kết quả lại một cách tron tru, được giới thiệu lần đầu bởi I. J. Schoenberg vào năm 1946¹¹. Trong phương pháp này, người ta hạn chế bậc của các đa thức tham gia vào quá trình nội suy bằng cách chỉ thực hiện phép nội suy Hermite trên hai điểm nội suy liên tiếp. Sau đó, các kết quả được ghép lại được một cách “tron tru” để thu được *hàm ghép tron* (spline).

Để mô tả kĩ hơn, giả sử miền xác định của x là $[a, b]$, được chia thành nhiều đoạn nhỏ

$$a = x_0 < x_1 < \dots < x_n = b.$$

Sau đó, trên mỗi đoạn nhỏ $[x_{k-1}, x_k]$, chúng ta tìm một hàm số S_k sao cho

$$S_k(x_{k-1}) = y_{k-1}, \quad S_k(x_k) = y_k.$$

Khi đó, hàm số

$$S(x) = \begin{cases} S_1(x), & x \in [x_0, x_1), \\ S_2(x) & x \in [x_1, x_2) \\ \dots \\ S_n(x), & x \in [x_{n-1}, b]. \end{cases}$$

là một hàm số liên tục và thỏa mãn $y_k = S(x_k)$ với mọi $k \in \{0, 1, 2, \dots, n\}$.

¹¹I. J. Schoenberg, *Contribution to the Problem of Approximation of Equidistant Data by Analytic Functions*, Quarterly of Applied Mathematics, Vol. 4, No. 1, 1946, pp. 45-99, 112-141.

Bây giờ, vấn đề đặt ra là chúng ta sẽ xác định các hàm S_k như thế nào? Về vấn đề này, chúng ta có một số phân tích sau:

- Nếu trên mỗi khoảng $[x_{k-1}, x_k]$, chúng ta lấy hàm nội suy bậc nhất, tức là hàm tuyến tính với đồ thị đi qua hai điểm $(x_k, y_k), (x_{k-1}, y_{k-1})$, thì chúng ta có một hàm *tuyến tính từng khúc* (piecewise linear) trên $[a, b]$. Lưu ý rằng các hàm tuyến tính từng khúc nhìn chung không khả vi tại các điểm x_j .
- Nếu trên mỗi khoảng $[x_{k-1}, x_k]$, chúng ta xác định một hàm nội suy bậc hai, tức là tìm một hàm sao cho đồ thị của nó là một parabol đi qua hai điểm $(x_k, y_k), (x_{k-1}, y_{k-1})$ thì chúng ta gặp phải vấn đề là có vô hạn các hàm như vậy. Chúng ta sẽ cần thêm điều kiện để xác định duy nhất các hàm bậc hai đó, chẳng hạn như điều kiện về hệ số góc của tiếp tuyến của parabol tại 1 điểm đầu mút.
- Nếu trên mỗi khoảng $[x_{k-1}, x_k]$, chúng ta xác định các hàm bậc ba với đồ thị đi qua hai điểm $(x_k, y_k), (x_{k-1}, y_{k-1})$, thì chúng ta cần thêm điều kiện xác định duy nhất. Thông thường, chúng ta cũng sẽ đặt các điều kiện về hệ số góc của đồ thị tại hai điểm đầu mút.

Từ phân tích trên, chúng ta thấy các hàm bậc nhất là quá “cứng”, dẫn đến hàm ghép trơn $S(x)$ thường không trơn tại các điểm nút, trong khi đó các hàm bậc hai cũng chưa đủ “mềm dẻo”: Các hàm bậc hai với đồ thị đi qua hai điểm cố định được mô tả bởi 1 tham số. Mặt khác, họ các hàm bậc ba với đồ thị đi qua hai điểm cố định là họ với hai tham số tự do. Điều này làm cho việc sử dụng các hàm bậc ba cho bài toán ghép trơn (spline) gọi là ghép trơn bậc ba (cubic spline) trở nên thuận tiện hơn.

4.5.2 Hàm ghép trơn bậc ba

Bây giờ, chúng ta tìm hiểu kĩ hơn về ghép trơn bậc ba. Bắt đầu bằng hai trường hợp đơn giản sau: $n = 1$ và $n = 2$. Tìm hiểu kĩ hai trường hợp riêng sẽ làm lộ rõ ý tưởng chúng minh trường hợp tổng quát.

Chúng ta bắt đầu xét với trường hợp $n = 1$. Điều thú vị là trường hợp này chúng ta không ghép các hàm số. Thay vào đó, chúng ta chỉ xác định một hàm bậc ba $p(x)$ với điều kiện

$$p(x_0) = y_0, \quad p(x_1) = y_1.$$

Tất nhiên, có vô hạn hàm số bậc ba thỏa mãn hai điều kiện trên. Do đó, cần thêm các điều kiện nữa, đó là

$$p'(x_0) = k_0, \quad p'(x_1) = k_1.$$

Khi đó, chúng ta được một trường hợp đơn giản của *bài toán nội suy Hermite*, tức là nội suy các giá trị của hàm số cùng với các đạo hàm của nó tại một số nút nội suy. Các hệ số của hàm bậc ba $p(x)$ được xác định bởi hệ phương trình

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 \\ 0 & 1 & 2x_0 & 3x_0^2 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 0 & 1 & 2x_1 & 3x_1^2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} y_0 \\ k_0 \\ y_1 \\ k_1 \end{bmatrix}. \quad (4.28)$$

Nếu gọi A là ma trận cỡ 4×4 trong vế trái của (4.28) thì, bằng tính toán trực tiếp, bằng tay hoặc bằng máy tính (chẳng hạn Python với SymPy), có thể thấy rằng $\det(A) = (x_1 - x_0)^4$. Do đó, hệ (4.28) có nghiệm duy nhất, theo định lý Cramer. Từ đó, thiết lập được các công thức tính p_i , $i = 1, 2, 3, 4$.

Như vậy, bài toán tìm hàm đa thức bậc ≤ 3 với giá trị của hàm số và các giá trị của đạo hàm tại hai điểm cho trước bằng những giá trị cho trước là giải được và có nghiệm duy nhất.

Chú ý 4.6. Lưu ý rằng trường hợp $n = 1$ không minh họa sự “ghép tron” đặc trưng của phương pháp này, khi người ta cần ghép hai hay nhiều hơn hàm bậc ba thành một hàm nội suy đa thức từng khúc (piecewise polynomial).

Trường hợp thú vị hơn là khi $n = 2$, lấy $a = x_0 < x_1 < x_2 = b$ và đi tìm hai hàm số bậc ba

$$\begin{aligned} p(x) &= p_0 + p_1 x + p_2 x^2 + p_3 x^3, \\ q(x) &= q_0 + q_1 x + q_2 x^2 + q_3 x^3, \end{aligned}$$

sao cho

$$p(x_0) = y_0, \quad p(x_1) = q(x_1) = y_1, \quad q(x_2) = y_2.$$

Từ đây dẫn đến 4 phương trình tuyến tính đối với 8 ẩn số p_i, q_i , $i = 1, \dots, 4$.

Các điều kiện tiếp theo là:

- Tính khả vi của hàm ghép tại x_1 dẫn đến

$$p'(x_1) = q'(x_1),$$

hay

$$p_1 + 2x_1 p_2 + 3x_1^2 p_3 = q_1 + 2x_1 q_2 + 3x_1^2 q_3.$$

- Tính khả vi hai lần của hàm ghép tại x_1 dẫn đến

$$p''(x_1) = q''(x_1),$$

hay, sau khi chia hai vế của phương trình cho 2,

$$p_2 + 3x_1 p_3 = q_2 + 3x_1 q_3.$$

Đến đây, chúng ta thu được thêm 2 phương trình tuyến tính nữa đối với 8 ẩn số trên, tổng cộng là 6 phương trình. Để có thể xác định được toàn bộ 8 ẩn số, cần đặt thêm hai điều kiện nữa đối với hàm spline. Có một số lựa chọn sau:

Hàm ghép trơn bị kẹp (clamped spline) là hàm ghép trơn với điều kiện cố định các hướng tiếp xúc của hàm ghép trơn tại hai đầu mút (x_0 và x_2), có nghĩa là

$$p'(x_0) = k_0, \quad q'(x_2) = k_2,$$

với k_0 và k_2 cho trước. Điều kiện này cho chúng ta hai phương trình

$$p_1 + 2x_0p_2 + 3x_0^2p_3 = k_0$$

$$q_1 + 2x_2q_2 + 3x_2^2q_3 = k_2.$$

Tổng hợp các phương trình trên, chúng ta có hệ phương trình sau đối với các hệ số của các hàm ghép trơn:

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & 0 & 0 & 0 & 0 \\ 1 & x_1 & x_1^2 & x_1^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_1 & x_1^2 & x_1^3 \\ 0 & 0 & 0 & 0 & 1 & x_2 & x_2^2 & x_2^3 \\ 0 & 1 & 2x_1 & 3x_1^2 & 0 & -1 & -2x_1 & -3x_1^2 \\ 0 & 0 & 1 & 3x_1 & 0 & 0 & -1 & -3x_1 \\ 0 & 1 & 2x_0 & 3x_0^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2x_2 & 3x_2^2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_1 \\ y_2 \\ 0 \\ 0 \\ k_0 \\ k_2 \end{bmatrix}. \quad (4.29)$$

Gọi M là ma trận cỡ 8×8 trong vế trái của (4.29). Khi đó, có thể tính được (bằng tay hoặc bằng máy tính, chẳng hạn bằng SymPy trong Python)

$$\det(M) = -2(x_0 - x_1)^3(x_0 - x_2)(x_1 - x_2)^3. \quad (4.30)$$

Như vậy, $\det(M) \neq 0$ bởi vì các điểm x_0, x_1 và x_2 là đôi một khác nhau. Do đó, theo định lý Cramer, với mọi bộ giá trị ở vế phải của (4.29), chúng ta đều có thể tìm được duy nhất cặp các đa thức bậc ba p và q thỏa mãn các điều kiện nội suy ghép trơn.

Bổ đề sau cho phép chúng ta tính hướng tiếp tuyến của đồ thị hàm ghép trơn tại điểm giữa khi biết hướng tiếp tuyến tại hai điểm hai bên.

Mệnh đề 4.1:

Giả sử các hệ số p_i và q_i , $i = 0, 1, 2, 3$, của các đa thức bậc ba $p(x)$ và $q(x)$ là nghiệm của hệ phương trình tuyến tính (4.29). Đặt $k_0 = p'(x_0)$,

$k_1 = p'(x_1)$, $k_2 = q'(x_2)$. Khi đó, $k_1 = q'(x_1)$ và

$$\frac{k_0}{x_1 - x_0} + \frac{2(x_2 - x_0)k_1}{(x_2 - x_1)(x_1 - x_0)} + \frac{k_2}{x_2 - x_1} = 3 \left(\frac{y_1 - y_0}{(x_1 - x_0)^2} + \frac{y_2 - y_1}{(x_2 - x_1)^2} \right). \quad (4.31)$$

Ngược lại, nếu k_0, k_1, k_2 thoả mãn (4.31) thì các hàm số bậc ba $p(x)$ và $q(x)$ giải các bài toán nội suy Hermite

$$p(x_0) = y_0, \quad p(x_1) = y_1, \quad p'(x_0) = k_0, \quad p'(x_1) = k_1,$$

và

$$q(x_1) = y_1, \quad q(x_2) = y_2, \quad q'(x_1) = k_1, \quad q'(x_2) = k_2$$

thoả mãn điều kiện

$$p''(x_1) = q''(x_1).$$

Sơ lược của chứng minh. Chứng minh bổ đề này dựa hoàn toàn vào tính toán, cụ thể là giải được các giá trị p_0, p_1, p_2 , và p_3 . Từ đó tính được k_1 theo các tham số còn lại. Thay biểu thức của k_1 vào vế trái của (4.31) và đơn giản hoá biểu thức thu được, chúng ta có thể đưa vế trái về vế phải. Các tính toán này khá cồng kềnh nhưng có thể kiểm tra được bằng các phần mềm tính toán kí hiệu. $\square$

Hàm ghép tron bậc ba tự nhiên (natural cubic spline) là hàm ghép tron với điều kiện biên là

$$p''(x_0) = q''(x_2) = 0.$$

Lúc này, hệ phương trình để xác định p_i, q_i có dạng

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & 0 & 0 & 0 & 0 \\ 1 & x_1 & x_1^2 & x_1^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_1 & x_1^2 & x_1^3 \\ 0 & 0 & 0 & 0 & 1 & x_2 & x_2^2 & x_2^3 \\ 0 & 1 & 2x_1 & 3x_1^2 & 0 & -1 & -2x_1 & -3x_1^2 \\ 0 & 0 & 1 & 3x_1 & 0 & 0 & -1 & -3x_1 \\ 0 & 0 & 1 & 3x_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3x_2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_1 \\ y_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (4.32)$$

Ma trận hệ số của hệ phương trình (4.32) có định thức là

$$D = 6(x_0 - x_1)^2(x_0 - x_2)(x_1 - x_2)^2.$$

Kết quả này có thể kiểm tra lại dễ dàng bằng các phần mềm tính toán, như Python (SymPy). Do định thức của ma trận hệ số khác không, bài toán hàm ghép tron tự nhiên cũng giải được duy nhất nghiệm.

Trong trường hợp tổng quát, chúng ta có định lí sau đây khẳng định tính tồn tại và duy nhất nghiệm của bài toán hàm ghép trơn bậc ba.

Định lí 4.12: Hàm ghép trơn bậc 3

Tồn tại duy nhất hàm nội suy ghép trơn bậc ba cho các dữ kiện $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ cho trước, $x_0 < x_1 < \dots < x_n$, trong đó $n \geq 2$, và thỏa mãn điều kiện biên $S'(0) = k_0$ và $S'(x_n) = k_n$ với k_0 và k_n cho trước. Khẳng định tương tự cũng đúng với điều kiện biên $S''(x_0) = S''(x_n) = 0$.

Chứng minh. Bước đầu tiên, chúng ta chỉ ra rằng tồn tại các hệ số góc $k_1, \dots, k_{n-1}$ thỏa mãn các hệ thức như trong bổ đề trên. Điều này dễ thấy vì chúng là nghiệm của một hệ phương trình tuyến tính Cramer với 3 đường chéo và đường chéo chính trội. Tiếp theo, các đa thức bậc ba nội suy Hermite tương ứng cho chúng ta kết quả của hàm ghép trơn bậc ba. $\square$

4.5.3 Quy trình tính hàm ghép trơn bậc ba

Bài toán tìm hàm ghép trơn có thể được tiếp cận theo hai hướng sau đây. Thứ nhất, chúng ta có thể viết ra hệ phương trình tuyến tính đối với tất cả các hệ số của tất cả các đa thức bậc ba cần tìm. Nếu có $n + 1$ nút nội suy, chúng ta có một hệ phương trình tuyến tính của $4n$ ẩn số. Một điều có thể nhận thấy ngay là ma trận hệ số của hệ phương trình khá “thưa” với định thức phụ thuộc vào phân bố các nút nội suy. Một số thuật toán giải hệ tuyến tính hiệu quả có thể cho ngay kết quả cuối cùng.

Ở hướng tiếp cận thứ hai, chúng ta sẽ giải một bài toán phụ và dùng kết quả của nó để tìm kết quả của bài toán chính. Theo hướng này, bài toán được giải qua hai bước sau:

1. Xác định các hệ số góc k_i tại các nút x_i , tức là các đạo hàm $S'(x_i)$, với $i = 1, 2, \dots, n - 1$.
2. Trên mỗi khoảng x_{i-1}, x_i , xác định hàm bậc ba nội suy Hermite các điều kiện

$$S(x_{i-1}) = y_{i-1}, \quad S(x_i) = y_i, \quad S'(x_{i-1}) = k_{i-1}, \quad S'(x_i) = k_i,$$

với k_{i-1} và k_i đã được xác định từ Bước 1.

Từ đó, có thể thấy rằng, các đa thức bậc ba trong ghép trơn có thể được xác định nếu xác định được các hệ số góc của chúng tại tất cả các nút nằm giữa x_0 và x_n (hệ số góc tại hai nút đầu và cuối x_0 và x_n được cho trước).

Ví dụ 4.11. Cho bộ dữ liệu $(-1, 2), (0, 1), (1, 3)$. Giả sử $S(x)$ là hàm nội suy ghép trơn bậc ba tương ứng thỏa mãn $k_0 = S'(-1) = 1$, $k_2 = S'(1) = -1/2$.

- (a) Hãy tính $k_1 = S'(0)$.
 (b) Hãy tính hàm ghép trơn bậc ba $S(x)$.

Lời giải

Theo các kí hiệu trên, chúng ta có $x_0 = -1$, $x_1 = 0$ và $x_2 = 1$, trong khi đó, $k_0 = 1$, $k_2 = -1/2$ và cần tính $k_1 = S'(0)$. Theo bổ đề trên

$$k_0 + 4k_1 + k_2 = \frac{3}{h}(y_2 - y_0) = 3.$$

Từ đó suy ra $S'(0) = k_1 = 5/8$.

Đa thức nội suy Hermite bậc ba cho bộ số liệu $(-1, 2, 1)$ và $(0, 1, 5/8)$, tức là đa thức bậc ba $p(x) = p_0 + p_1x + p_2x^2 + p_3x^3$ thỏa mãn

$$p(-1) = 2, \quad p'(-1) = 1, \quad p(0) = 1, \quad p'(0) = 5/8,$$

có các hệ số là nghiệm của

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 5/8 \end{bmatrix}.$$

Giải hệ này, ta thu được $p_0 = 1$, $p_1 = 5/8$, $p_2 = 21/4$, và $p_3 = 29/8$.

Trên đoạn $[0, 1]$, hàm ghép trơn $S(x) = q(x)$, trong đó $q(x)$ là đa thức bậc ba Hermite nội suy bộ số liệu $(0, 1, 5/8)$ và $(1, 3, -1/2)$. Các hệ số của q thỏa mãn

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5/8 \\ 3 \\ -1/2 \end{bmatrix}.$$

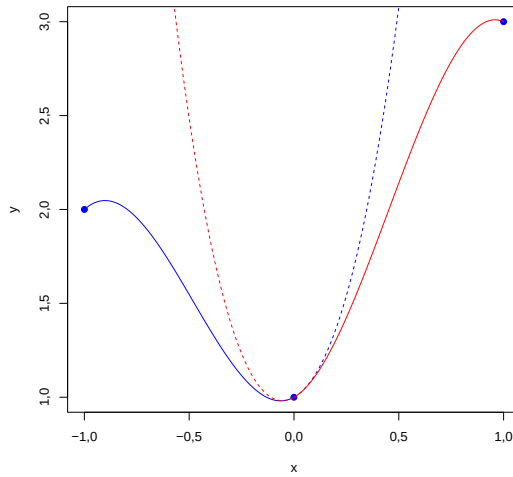
Giải hệ phương trình trên, chúng ta thu được

$$q_0 = 1, \quad q_1 = 5/8, \quad q_2 = 21/4, \quad q_3 = -31/8.$$

Tổng hợp các kết quả, chúng ta có

$$S(x) = \begin{cases} 1 + \frac{5}{8}x + \frac{21}{4}x^2 + \frac{29}{8}x^3 & \text{nếu } -1 \leq x < 0, \\ 1 + \frac{5}{8}x + \frac{21}{4}x^2 - \frac{31}{8}x^3 & \text{nếu } 0 \leq x \leq 1. \end{cases}$$

Đó là hàm ghép trơn bậc ba với các điều kiện bị kẹp.



Hình 4.3: Các điểm dữ liệu trong Ví dụ 4.11 và hàm ghép trơn bậc ba của chúng

CAS 4.8. Giả sử chúng ta có bảng số liệu sau:

x	1	2	4	5	8	10
y	2	1,5	1,25	1,2	1,125	1,1

Gọi $S(x)$ là hàm ghép trơn bậc ba *tự do* của bộ số liệu trên. Đoạn code mẫu sau tính giá trị của S tại một điểm nào đó, chẳng hạn $S(6,46)$.

```
# Nạp gói pracma
library(pracma)

# Nạp các dữ liệu x và y
x <- c(1, 2, 4, 5, 8, 10)
y <- c(2, 1.5, 1.25, 1.2, 1.125, 1.1)

# cubicspline tự do tính giá trị của hàm ghép
# trơn tại xi = 6,46

result <- cubicspline(x, y, xi = 6.46, endp2nd = FALSE)

# In ra kết quả
print(result)

## [1] 1,14868973
```

Kết quả thu được là $S(6,46) \approx 1,1487$, đúng như mong đợi.

CAS 4.9. Để lấy một ví dụ về hàm ghép tron bậc ba bị kẹp, xét bảng số liệu sau đây:

x	1	2	3	4
y	1	0,5	1/3	0,25

Gọi $S(x)$ là hàm ghép tron bậc ba *bị kẹp* bởi hai điều kiện $S'(1) = -7/12$ và $S'(4) = -1/12$. Đoạn mã lệnh R sau tính $S(1,45)$.

```
library(pracma)

# Các dữ kiện x và y
x <- c(1, 2, 3, 4)
y <- c(1, 0.5, 1/3, 0.25)

# Hai giá trị của đạo hàm tại hai đầu mút
k <- c(-7/12, -1/12)

# cubicspline của pracma với điều kiện tại hai đầu mút
cubicspline(x, y, xi = 1.45, endp2nd = TRUE, der = k)

## [1] 0,74509375
```

Kết quả chạy máy tính, $S(1,45) \approx 0,7450937$, trong khi đó, công thức tường minh của $S(x)$ trên đoạn $[1, 2]$ có thể tính bằng tay là

$$S(x) = \frac{1}{12}x^3 - \frac{1}{4}x^2 - \frac{1}{3}x + \frac{3}{2}, \quad 1 \leq x \leq 2.$$

Từ đó, bạn đọc có thể tự so sánh kết quả tính tự động bởi R và kết quả tính bằng tay.

Kết thúc mục này là định lí sau đây nêu ra một tính chất thú vị của hàm ghép tron bậc ba: tính cực tiểu hoá trung bình bình phương của đạo hàm cấp hai.

Định lý 4.13:

Cho $f(x)$ là một hàm khả vi liên tục hai lần trên $[a, b]$ và $S(x)$ là một hàm ghép trơn bậc ba nội suy $f(x)$ tại một số nút nội suy trong $[a, b]$. Giả sử $S(x)$ thỏa mãn *một trong hai* điều kiện sau:

- (i) Điều kiện nội suy tự nhiên: $S''(a) = S''(b) = 0$, hoặc
- (ii) Điều kiện nội suy bị kẹp: $S'(a) = f'(a)$, $S'(b) = f'(b)$.

Khi đó, xảy ra bất đẳng thức sau:

$$\int_a^b |S''(x)|^2 dx \leq \int_a^b |f''(x)|^2 dx. \quad (4.33)$$

Chứng minh. Đặt $R(x) = f(x) - S(x)$ hay $f(x) = S(x) + R(x)$. Hơn nữa, $R(x)$ cũng khả vi liên tục hai lần. Chúng ta có

$$\begin{aligned} \int_a^b |f''(x)|^2 dx &= \int_a^b |S''(x)|^2 dx + \int_a^b |R''(x)|^2 dx \\ &\quad + 2 \int_a^b R''(x)S''(x) dx. \end{aligned} \quad (4.34)$$

Lưu ý rằng đạo hàm cấp ba $S'''(x)$ xác định trên $[a, b]$ trừ các nút nội suy và $a = x_0 < \dots < x_n = b$. Do đó, chúng ta có thể sử dụng công thức tích phân từng phần như sau: Trên mỗi đoạn $[x_{i-1}, x_i]$ chúng ta có

$$\int_{x_{i-1}}^{x_i} R''(x)S''(x) dx = S''(x)R'(x) \Big|_{x_{i-1}}^{x_i} - \int_{x_{i-1}}^{x_i} R'(x)S_i'''(x) dx,$$

trong đó $S_i(x)$ là hàm bậc ba, $S_i(x) = S(x)$ trên $[x_{i-1}, x_i]$. Nhưng $S_i'''(x)$ là hằng số, nên

$$\int_{x_{i-1}}^{x_i} R'(x)S_i'''(x) dx = S_i'''(x_i)R(x) \Big|_{x_{i-1}}^{x_i} = S_i'''(x_i)(R(x_i) - R(x_{i-1})) = 0.$$

Do đó,

$$\int_{x_{i-1}}^{x_i} R''(x)S''(x) dx = S''(x)R'(x) \Big|_{x_{i-1}}^{x_i} = S''(x_i)R'(x_i) - S''(x_{i-1})R'(x_{i-1}).$$

Lấy tổng khi $i = 1, 2, \dots, n$, lưu ý rằng $S''(x)$ liên tục, chúng ta có

$$\begin{aligned} \int_a^b R''(x)S''(x)dx &= \sum_{i=1}^n S''(x_i)R'(x_i) - S''(x_{i-1})R'(x_{i-1}) \\ &= S''(b)R'(b) - S''(a)R'(a). \end{aligned} \quad (4.35)$$

Xét hai trường hợp: (i) Nếu S là hàm ghép trơn nội suy tự nhiên, $S''(a) = S''(b) = 0$, thì vế phải của (4.35) bằng 0. (ii) Nếu S là hàm ghép trơn bị kẹp, $S'(a) = f'(a)$, $S'(b) = f'(b)$, thì

$$R'(a) = f'(a) - S'(a) = 0, \quad R'(b) = f'(b) - S'(b) = 0.$$

Từ đó, vế phải của (4.35) bằng 0.

Như vậy, trong cả hai trường hợp, vế trái của (4.35) bằng 0. Do đó, kết hợp với (4.34), chúng ta thu được

$$\int_a^b |f''(x)|^2 dx = \int_a^b |S''(x)|^2 dx + \int_a^b |R''(x)|^2 dx \geq \int_a^b |S''(x)|^2 dx.$$

Đó là điều phải chứng minh. $\square$

Áp dụng định lý trên, bạn đọc có thể dễ dàng suy luận để thu được hệ quả sau:

Định lý 4.14:

Cho $f(x)$ là một hàm khả vi liên tục hai lần trên $[a, b]$, $S_n(x)$ là một hàm ghép trơn bậc ba tự nhiên nội suy $f(x)$ tại một số nút nội suy trong $[a, b]$ và $S_c(x)$ là hàm ghép trơn bậc ba bị kẹp với điều kiện $S'_c(a) = f'(a)$, $S'_c(b) = f'(b)$. Khi đó, xảy ra bất đẳng thức sau:

$$\int_a^b |S''_n(x)|^2 dx \leq \int_a^b |S''_c(x)|^2 dx \leq \int_a^b |f''(x)|^2 dx. \quad (4.36)$$

4.6 Phương pháp bình phương bé nhất

Sau đây, chúng ta tìm hiểu phương pháp bình phương bé nhất (ordinary least squares method) để làm khớp dữ liệu (data fitting). Đó là một phương pháp quan trọng thường được dùng để tìm các đường hồi quy (regression curve) trong thống kê học. Tuy nhiên, trong khuôn khổ của tài liệu này chúng ta chỉ tập trung vào vấn đề tính toán các đường hồi quy mà không nói về những khía cạnh thống kê như vấn đề xây dựng khoảng tin cậy của các hệ số hồi quy. Tài liệu tham khảo: T. V. Đĩnh [6, §4.2], Kreyszig [9, §20.5, §25.9].

4.6.1 Xấp xỉ tuyến tính

Định luật Hooke¹² về lò xo nói rằng, trong giới hạn đàn hồi, độ lớn lực đàn hồi F tỉ lệ thuận với độ biến dạng của lò xo:

$$F = ks.$$

Hằng số k được gọi là *độ cứng* (hệ số đàn hồi) của lò xo. Để xác định k bằng thực nghiệm, người ta cho tác dụng một lực F vào một đầu của lò xo và đo độ dãn/nén s . Kết quả thu được là một bộ dữ liệu (s_i, F_i) , $i = 1, 2, \dots, n$, với n là số lần đo.

Nếu các phép đo đặc là hoàn chính xác (và lò xo tuân theo định luật Hooke một cách hoàn hảo) thì $k = F_i/s_i$ không phụ thuộc vào chỉ số i . Nhưng các phép đo chỉ cho kết quả gần đúng và chúng ta có các giá trị $k_i = F_i/s_i$ không nhất thiết bằng nhau. Từ đây nảy sinh vấn đề tìm một giá trị k duy nhất từ các dữ liệu trên. Vấn đề này được giải quyết bằng phương pháp bình phương bé nhất mà chúng ta sẽ tìm hiểu ngay sau đây.

Trong phương pháp bình phương bé nhất, ta tìm k sao cho tổng bình phương các sai số đối với F là bé nhất. Tức là tìm một giá trị k^* bởi

$$k^* = \arg \min \sum_{i=1}^n |F_i - ks_i|^2.$$

Theo hướng này, ta xét hàm số

$$S(k) := \sum_{i=1}^n |F_i - ks_i|^2.$$

Rõ ràng, $S(k)$ là hàm bậc hai đối với k . Cực tiểu duy nhất của nó là nghiệm của

$$S'(k) = 0.$$

Từ đó, ta có

$$k^* = \frac{\sum_{i=1}^n F_i s_i}{\sum_{i=1}^n s_i^2}.$$

Ví dụ 4.12. Để xác định độ cứng k của một lò xo, người ta tiến hành treo các vật nặng vào một đầu của lò xo tạo ra các trọng lực F_i lần lượt là 2, 4, và 6 Newton tác dụng vào lò xo. Khi đó, các li độ s_i tương ứng là đo được là 1,5, 2,9, và 4,4 (mm). Nếu các phép đo này là chính xác hoàn toàn thì định luật Hooke nói rằng

$$2 = 1,5k, \quad 4 = 2,9k, \quad 6 = 4,4k.$$

¹²Robert Hooke FRS (1635 – 1703) là một nhà toán học và vật lý học người Anh.

Tuy nhiên, 3 phương trình trên lập thành một hệ phương trình quá xác định (overdetermined system) không có nghiệm. Nếu dùng phương pháp bình phương bé nhất mô tả ở trên, ta thu được

$$k \approx \frac{2 \times 1,5 + 4 \times 2,9 + 6 \times 4,4}{(1,5^2 + 2,9^2 + 4,4^2)} = \frac{41}{30,02} \approx 1,366 \text{ (N/mm)}.$$

Trong thực tế, ngoài lực đàn hồi, chúng ta còn có thể có một ngoại lực F_b (như lực hấp dẫn tác dụng lên một đầu lò xo). Khi đó, tổng hợp lực đo được có dạng

$$F = ks + F_b.$$

Khi đó, ta cần xác định hai tham số k và F_b dựa trên các giá trị đo đạc của s và F . Nó là một trường hợp đặc biệt của xấp xỉ tuyến tính bằng phương pháp bình phương bé nhất, được mô tả như sau. Giả sử ta đã biết hai đại lượng x và y phụ thuộc nhau qua biểu thức dạng

$$y = ax + b,$$

nhưng các hệ số a, b chưa xác định. Giả sử rằng ta đã biết giá trị $y_i, i = 1, 2, \dots, n$ của hàm số tại các điểm x_i . Ta sẽ tìm các giá trị của tham số a và b sao cho hàm này xấp xỉ các giá trị đã cho với tổng bình phương các sai số tại các điểm (x_i, y_i) là bé nhất. Nói cách khác, ta cần tìm

$$(a^*, b^*) = \arg \min S(a, b),$$

với

$$S(a, b) := \frac{1}{2} \sum_{i=1}^n |y_i - (ax_i + b)|^2.$$

Rõ ràng $S(a, b)$ là hàm khả vi của hai biến a và b . Để tìm điểm cực tiểu, ta bắt đầu với việc tìm *điểm tới hạn* (critical point) của nó, tức là tìm (a, b) sao cho các đạo hàm riêng của S triệt tiêu:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0. \quad (4.37)$$

Đặt

$$S_i := \frac{1}{2} |y_i - (ax_i + b)|^2 = \frac{1}{2} y_i^2 + \frac{1}{2} a^2 x_i^2 + \frac{1}{2} b^2 - x_i y_i a - y_i b + a b x_i,$$

Khi đó,

$$\begin{aligned} \frac{\partial S_i}{\partial a} &= a x_i^2 - x_i y_i + b x_i \\ \frac{\partial S_i}{\partial b} &= b - y_i + a x_i \end{aligned}$$

Do $S = S_1 + S_2 + \dots + S_n$ nên

$$\begin{aligned}\frac{\partial S}{\partial a} &= \sum_{i=1}^n \frac{\partial S_i}{\partial a} = a \left(\sum_{i=1}^n x_i^2 \right) - \sum_{i=1}^n x_i y_i + b \left(\sum_{i=1}^n x_i \right) \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^n \frac{\partial S_i}{\partial b} = n b - \sum_{i=1}^n y_i + a \left(\sum_{i=1}^n x_i \right).\end{aligned}$$

Khi đó, điều kiện (4.37) có dạng

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}.$$

Hệ trên có nghiệm duy nhất được tính theo công thức Cramer. Thật vậy, định thức của ma trận hệ số cấp 2×2 là

$$D := \begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{vmatrix} = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2.$$

Theo bất đẳng thức Cauchy–Bunyakovsky–Schwarz, $D > 0$ nếu có ít nhất hai giá trị $x_i \neq x_j$, $1 \leq i < j \leq n$. Trong trường hợp này, hệ có nghiệm duy nhất là a^* , b^* . Để tính nghiệm, ta tính

$$\begin{aligned}D_a &:= \begin{vmatrix} \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i & n \end{vmatrix} \\ &= n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right), \\ D_b &:= \begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \end{vmatrix} \\ &= \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i y_i \right).\end{aligned}$$

Từ đó, ta thu được kết quả sau:

Định lí 4.15:

Xấp xỉ tuyến tính với sai số bình phương bé nhất của bộ n số liệu (x_i, y_i) , $i = 1, 2, \dots, n$ được cho bởi $y = ax + b$ với

$$a = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2},$$

$$b = \frac{\left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}.$$

Chú ý 4.7. Nếu ta đã tính được a , thì b có thể được tính qua a như sau:

$$b = \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i \right) / n.$$

Ví dụ 4.13. Tìm hàm số bậc nhất $ax + b$ xấp xỉ bộ dữ liệu (x_i, y_i) , $i = 1, 2, 3, 4$ cho bởi bảng sau:

x_i	-1,2	-0,1	0,2	1,4
y_i	0,105	1,101	0,890	1,925

Lời giải

Ta cần tính a và b theo công thức cho bởi Định lí 4.15. Như vậy, ta cần tính x_i^2 và $x_i y_i$ với $i = 1, 2, 3, 4$. Các kết quả được ghi trong bảng sau. Dòng cuối cùng là tổng các kết quả ở phía trên.

x_i	y_i	x_i^2	$x_i y_i$	$y = 0,68135x + 0,95415$
-1,2	0,105	1,44	-0,126	0,13653
-0,1	1,101	0,01	-0,1101	0,88602
0,2	0,890	0,04	0,178	1,09042
1,4	1,925	1,96	2,695	1,90804
0,3	4,021	3,45	2,6369	$S(a, b) = 0,087669$

Ta thu được hàm bậc nhất $y = 0,68135x + 0,95415$ với tổng bình phương các sai số là 0,087669.

Ví dụ 4.14. Giả sử đại lượng $P(t)$ phụ thuộc tuyến tính vào t , $P(t) = at + b$. Tại một số giá trị của t , có các giá trị gần đúng của P như sau:

t_i	3	4	5	6	7	8	9	10
P_i	3,45	3,60	5,89	6,31	6,90	8,52	9,62	11,30

Tìm gần đúng $P(t)$ bằng phương pháp bình phương nhỏ nhất và minh họa kết quả bằng đồ thị.

Trả lời. Hàm $P(t)$ với các hệ số tính đến 4 chữ số sau dấu thập phân là $P(t) \approx 1,1135t - 0,2887$. Đồ thị của hàm số $P(t)$ và các điểm dữ liệu được cho trong CAS 4.10 □

CAS 4.10. Hãy giải Ví dụ 4.14 bằng R.

```
# Khai báo các vectơ P và t
t <- c(3, 4, 5, 6, 7, 8, 9, 10)
P <- c(3.45, 3.60, 5.89, 6.31, 6.90, 8.52, 9.62, 11.30)

# Khai báo mô hình model bằng hàm lm()
model <- lm(P ~ t)

# Các biến a và b lưu trữ các hệ số hồi quy
a <- model$coefficients[2]
b <- model$coefficients[1]

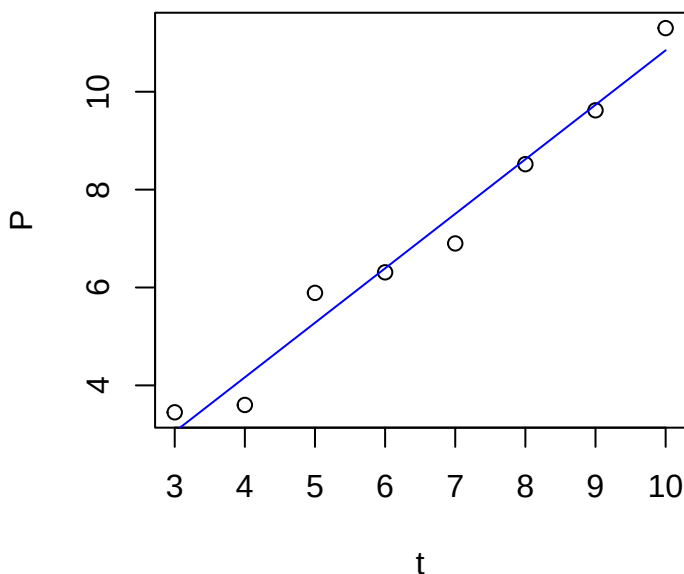
# Công thức xấp xỉ tuyến tính
f <- function(x) a * x + b

# In ra các tham số của mô hình
print(model)

##
## Call:
## lm(formula = P ~ t)
##
## Coefficients:
## (Intercept)          t
##    -0,28869      1,11345

# Vẽ các điểm dữ liệu
plot(P ~ t)

# Vẽ đường thẳng xấp xỉ tuyến tính
curve(f, from = 3, to = 10, col = "blue", add = TRUE)
```



Kết quả thu được khớp với kết quả tính bằng tay.

Lưu ý, rất nhiều máy tính khoa học cầm tay (như Casio FX-580VN) có sẵn chức năng tính các hệ số a và b trong mô hình tuyến tính như trên.

4.6.2 Xấp xỉ đa thức bằng phương pháp bình phương bé nhất

Để mở rộng bài toán xấp xỉ tuyến tính cho các mối quan hệ ở dạng tổng quát hơn, chúng ta sẽ sử dụng một số khái niệm và kết quả của đại số tuyến tính.

Xét các ma trận và vectơ sau đây:

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad c = \begin{bmatrix} b \\ a \end{bmatrix}.$$

Thế thì hàm $S(a, b)$ có thể viết dưới dạng

$$S(a, b) = \frac{1}{2} \|Ac - y\|_2^2.$$

Vế phải là chuẩn ℓ^2 (chuẩn Euclid) của các vectơ trong $\mathbb{R}^n$ (xem §3.2.1).

Chú ý rằng Ac thuộc vào không gian ảnh (image) của ánh xạ tuyến tính $L_A: c \mapsto Ac$, tức là không gian sinh bởi các cột của A :

$$Ac \in R(A).$$

Do đó, bài toán tìm $c = (a, b)^T$ sao cho $S(a, b)$ bé nhất có thể được giải quyết thông qua việc tìm hình chiếu trực giao của y lên không gian ảnh $R(A)$. Ở đây,

$$R(A) := \{w \in \mathbb{R}^n \mid w = Av, v \in \mathbb{R}^2\}.$$

Nếu y_0 là hình chiếu trực giao của y lên $R(A)$ thì $y - y_0$ thuộc phần bù trực giao của $R(A)$. Tức là $y - y_0 \in R(A)^\perp$. Theo một kết quả cổ điển của đại số tuyến tính thì $R(A)^\perp = N(A^T)$. Ở đây $N(A^T)$ là không gian zero của A^T , tức là

$$N(A^T) = \{d \in \mathbb{R}^2 \mid Ad = 0\}.$$

Vậy, ta suy ra $y - y_0 \in N(A^T)$. Tức là

$$0 = A^T(y - y_0) = A^T y - A^T y_0.$$

Do đó, nếu c là nghiệm của bài toán xấp xỉ bình phương bé nhất thì $Ac = y_0$. Từ đó suy ra,

$$A^T Ac = A^T y. \quad (4.38)$$

Nếu $A^T A$ là khả nghịch thì c được xác định duy nhất theo phương trình (4.38). Cụ thể là

$$c = (A^T A)^{-1} A^T y. \quad (4.39)$$

Đây là công thức dạng đóng cho nghiệm của bài toán bình phương bé nhất.

Những lập luận trên cũng đúng nếu A là ma trận cỡ $n \times k$ và c là vectơ có k thành phần, với $n \geq 1$ và $k \geq 1$ tùy ý. Các ma trận A như vậy nảy sinh trong bài toán bình phương bé nhất khi các mối quan hệ ta cần tìm ở dạng

$$y = F(x; c_1, c_2, \dots, c_k),$$

với F là hàm phụ thuộc một cách tuyến tính vào k tham số $c_1, c_2, \dots, c_k$, và điều kiện xấp xỉ hàm được lấy tại n bộ dữ kiện. Nếu A có hạng bằng k , thì có thể chứng minh được $A^T A$ là ma trận khả nghịch cỡ $k \times k$ và khi đó c xác định duy nhất.

Ví dụ 4.15. Biết đại lượng y là một tam thức bậc hai của x và tại một số giá trị của x , ta biết xấp xỉ các giá trị của y như sau:

x_i	0,5	1,6	2,3	3,1	3,8	4,2
y_i	2,5	1,2	1,1	2,2	4,2	6,8

Hãy lập công thức xấp xỉ biểu diễn y qua x bằng phương pháp bình phương bé nhất.

Giả sử $y = c_0 + c_1 x + c_2 x^2$. Đặt $c = (c_0, c_1, c_2)^T$ và xét ma trận A cỡ 5×3 như sau:

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \\ 1 & x_6 & x_6^2 \end{bmatrix} = \begin{bmatrix} 1 & 0,5 & 0,25 \\ 1 & 1,6 & 2,56 \\ 1 & 2,3 & 5,29 \\ 1 & 3,1 & 9,61 \\ 1 & 3,8 & 14,44 \\ 1 & 4,2 & 17,64 \end{bmatrix}.$$

Khi đó,

$$A^T A \approx \begin{bmatrix} 6 & 15,5 & 49,79 \\ 15,5 & 49,79 & 175,139 \\ 49,79 & 175,139 & 646,6355 \end{bmatrix}.$$

Gọi $y = (y_1, y_2, \dots, y_6)^T$ thì

$$b = A^T y = \begin{bmatrix} 18 \\ 57,04 \\ 18 \end{bmatrix}.$$

Theo (4.38), các giá trị $c = (c_0, c_1, c_2)^T$ cần tìm là nghiệm của hệ phương trình Cramer của 3 ẩn:

$$A^T A c = b,$$

có nghĩa là

$$\begin{bmatrix} 6 & 15,5 & 49,79 \\ 15,5 & 49,79 & 175,139 \\ 49,79 & 175,139 & 646,6355 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 18 \\ 57,04 \\ 18 \end{bmatrix}.$$

Giải hệ phương trình trên, chúng ta thu được nghiệm xấp xỉ (làm tròn tới 4 chữ số thập phân) là $c = (4,2641; -3,7247; 1,0072)$. Công thức xấp xỉ y biểu diễn qua x là

$$y = 4,2641 - 3,7247 x + 1,0072 x^2. \quad (4.40)$$

Trong CAS 4.11 sau đây, chúng ta sẽ thực hành tính toán cho Ví dụ 4.15 trong R, vận dụng thuật toán `qr.solve()` có sẵn.

CAS 4.11. Các tính toán trong Ví dụ 4.15 được tính tự động bằng R như sau:

```
# Dữ liệu ban đầu
x <- c(0.5, 1.6, 2.3, 3.1, 3.8, 4.2)
y <- c(2.5, 1.2, 1.1, 2.2, 4.2, 6.8)
```

```
# Tạo ma trận A với các cột 1, x, x^2 bằng cbind()
A <- cbind(1, x, x^2)

# Véc tơ b = y A
b <- y %*% A

# Tìm nghiệm với chuẩn bé nhất của hệ dưới xác định Ay = b,
# làm tròn tới 4 chữ số thập phân.
model <- round(qr.solve(A, y), 4)

# In ra kết quả
print(model)

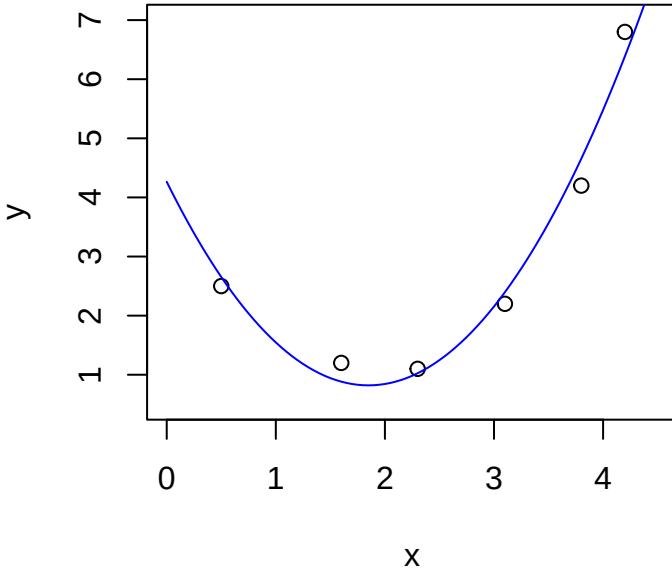
##                x
## 4,2641 -3,7247 1,0072
```

Kết quả của `print(model)` là các hệ số cần tìm. Đoạn mã lệnh R tiếp sau đây, ta vẽ đồ thị của dữ liệu và đường xấp xỉ để có một minh hoạ trực quan về lời giải.

```
# Vẽ hình minh hoạ các điểm mẫu (x, y)
plot(y ~ x, xlim = c(0, 4.5), ylim = c(0.5, 7))

# Định nghĩa hàm số f(x) thu được từ đoạn mã phía trên
f <- function(x) {
  model[3] * x^2 + model[2] * x + model[1]
}

# Vẽ đồ thị hàm f(x)
curve(f, from = 0, to = 5, col = "blue", lty = 1, add = TRUE)
```



4.6.3 Xấp xỉ bằng các hàm lũy thừa

Năm 1924, khi nghiên cứu tương quan giữa khối lượng của càng lớn của một loài cua, Huxley¹³ nhận thấy rằng, nếu y là khối lượng của càng lớn của một cá thể cua, w là khối lượng của toàn bộ cá thể đó, và $x = w - y$ là khối lượng của phần cơ thể còn lại không bao gồm càng của, thì mối quan hệ giữa x và y là một mối quan hệ phi tuyến.

Tuy nhiên, khi biểu diễn các dữ liệu đo đạc được trên *thang log-log*, tức là thang với trục hoành biểu diễn $\ln(x)$ và trục tung biểu diễn $\ln(y)$, Huxley nhận thấy mối quan hệ giữa $\ln(x)$ và $\ln(y)$ “khá gần” với một quan hệ tuyến tính. Từ đó, bằng cách áp dụng phương pháp ở mục trước cho các đại lượng $\ln(y)$ và $\ln(x)$, ta có thể tìm được mối liên hệ dạng

$$\ln(y) = k \ln(x) + \ln(B), \quad k > 1, B < 1. \quad (4.41)$$

Mối liên hệ (4.41) có thể viết lại dưới dạng hàm lũy thừa

$$y = Bx^k. \quad (4.42)$$

¹³HUXLEY, J. *Constant Differential Growth-Ratios and their Significance*. Nature 114, 895–896 (1924).

Trong thực tế, mối quan hệ giữa x và y như trong (4.42) xuất hiện phổ biến trong sinh học, thường được gọi là *tương quan sinh trưởng* (allometric relation) hoặc *luật mũ* (power law)¹⁴. Những mối quan hệ tương tự như vậy cũng xuất hiện trong những ngành khoa học khác như trong Kinh tế học (như hàm sản xuất Cobb–Douglas $Y = AL^\beta K^\alpha$) hay trong Vật lí, v.v.

Ví dụ 4.16. Áp suất P (kg cm^{-2}) của một khối chất khí tương ứng với thể tích V (cm^3) của nó được đo đạc và ghi chép lại như sau:

V	50	60	70	90	100
P	64,7	51,3	40,5	25,9	7,8

Định luật khí lí tưởng nói rằng $PV^\gamma = C$, trong đó, γ và C là hằng số. Hãy ước tính từ dữ liệu trên các giá trị γ và C .

Lời giải

Mối quan hệ $PV^\gamma = C$ có thể viết lại dưới dạng $\ln(P) = -\gamma \ln(V) + \ln(C)$. Do đó, để ước tính γ và C , chúng ta sử dụng phương pháp xấp xỉ tuyến tính trên thang log-log.

CAS 4.12. Hãy viết chương trình R để thực hiện các tính toán trong Ví dụ 4.16.

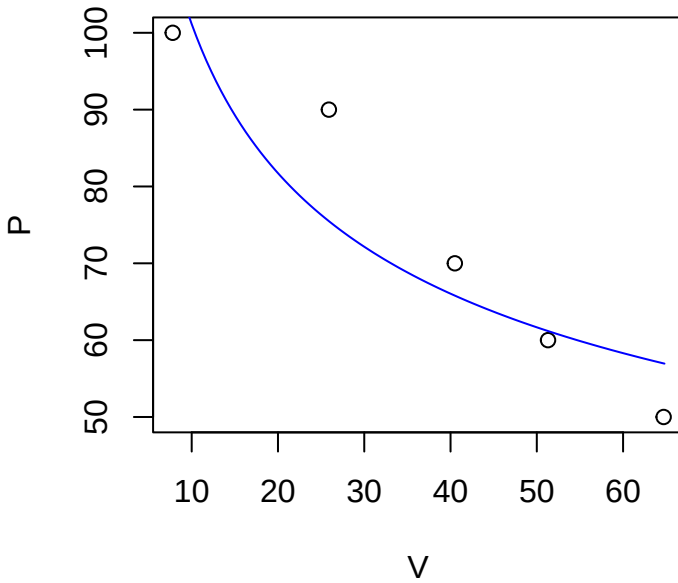
```
P <- c(50, 60, 70, 90, 100)
V <- c(64.7, 51.3, 40.5, 25.9, 7.8)
P_log <- log(P)
V_log <- log(V)
model <- lm(P_log ~ V_log)
print(model)

##
## Call:
## lm(formula = P_log ~ V_log)
##
## Coefficients:
## (Intercept)      V_log
##    5,324181    -0,307357
```

Tiếp theo, chúng ta biểu diễn các dữ liệu vẽ đồ thị của đường cong xấp xỉ như sau:

¹⁴Shingleton, A. (2010) *Allometry: The Study of Biological Scaling*. Nature Education Knowledge 3(10):2


```
plot(P ~ V)
f <- function(x) {
  exp(model$coefficients[1]) * x^(model$coefficients[2])
}
curve(f, from = 7.7, to = 64.8, col = "blue", add = TRUE)
```



Nhận xét 4.1. Mặc dù hai mối quan hệ (4.41) và (4.42) là tương đương, nhưng khi áp dụng phương pháp bình phương bé nhất cho hai mối quan hệ này thì ta thu được hai bài toán không tương đương. Đối với quan hệ (4.41) trên thang log-log, ta đi tìm cực tiểu của

$$S_{\log}(k, t) := \frac{1}{2} \sum_{i=1}^n (\log(y_i) - k \log(x_i) - t)^2,$$

với $t = \log(b)$ và n là số các điểm dữ liệu. Đó là hàm bậc hai đối với k và t . Khi áp dụng phương pháp bình phương bé nhất cho quan hệ (4.42), ta đi tìm cực tiểu của hàm

$$S(k, b) = \frac{1}{2} \sum_{i=1}^n (y_i - b x_i^k)^2.$$

$S(k, b)$ là một hàm số “siêu việt” của k và b . Có thể thấy hai hàm trên đạt cực tiểu tại các giá trị khác nhau của k và $t = \log(b)$.

4.6.4 Xấp xỉ bằng hàm số mũ

Một số đại lượng trong khoa học kỹ thuật, kinh tế, sinh học, khoa học xã hội có mối quan hệ có thể được mô hình hoá bởi các hàm số dạng $y = Be^{ax}$. Từ đó, nảy sinh vấn đề tìm các giá trị a và B sao cho mô hình $y = Be^{ax}$ phù hợp với một dữ liệu đo đạc cho trước.

Hàm lỗi không có trọng

Trong phương pháp này (giả sử $B > 0$ cho đơn giản), chúng ta lấy logarith hai vế,

$$\ln(y) = \ln(B) + ax = b + ax,$$

trong đó $b := \ln(B)$. Bài toán đưa về bài toán xấp xỉ tuyến tính các giá trị a và $b = \ln(B)$ theo các dữ liệu $(x_i, \ln(y_i))$. Lưu ý, phương pháp này có một đặc điểm là giá trị của a và B phụ thuộc nhiều vào các giá trị y_i gần 0 so với các giá trị y_i lớn hơn.

Hãy xem một ví dụ thực tế sau.

CAS 4.13. Dữ liệu từ một nghiên cứu với 2 871 tai nạn xe hơi gồm nguy cơ tương đối y và nồng độ cồn trong máu x của người lái xe như sau:

x	0	0,01	0,03	0,05	0,07	0,09	0,11	0,13	0,15	0,17	0,19	0,21
y	1	1,03	1,06	1,38	2,09	3,54	6,41	12,6	22,1	39,05	65,32	99,78

Hãy xây dựng mô hình $y = Be^{ax}$ cho dữ liệu trên.

Phương pháp ở đây là tìm xấp xỉ tuyến tính bằng bình phương bé nhất đối với dữ liệu $(x_i, \ln(y_i))$. Để nhanh chóng, chúng ta dùng hàm có sẵn `lm()` của R.

```
# Tạo vectơ x bắt đầu từ 0, 0,01, và các điểm cách đều
# với khoảng cách 0,02.
x <- c(0, seq(0.01, 0.21, by = 0.02))

# Vectơ y từ giả thiết
y <- c(1, 1.03, 1.06, 1.38, 2.09, 3.54, 6.41, 12.6, 22.1,
      39.05, 65.32, 99.78)

# Xấp xỉ tuyến tính cho các dữ liệu x và ln(y)
y_log <- log(y)
model <- lm(y_log ~ x)
```

```
# In ra màn hình kết quả
```

```
print(model$coefficients)
```

```
## (Intercept)          x
```

```
## -0,539485261 23,817576640
```

Kết quả thu được từ đoạn mã trên như sau:

$$\ln(y) = -0,5395 + 23,8176x$$

Biểu đồ phân tán (scatter plot) của dữ liệu cùng với đồ thị của mô hình hàm mũ được thể hiện trong Hình 4.4 (đường màu đỏ). Chú ý quan sát thấy đường cong màu đỏ đi khá xa so với các dữ liệu khi y_i lớn.

Hàm lỗi có trọng

Thay vì xét hàm mục tiêu $S(a, b)$ như trên, chúng ta xét hàm mục tiêu có “trọng” sau:

$$S_w(a, b) := \frac{1}{2} \sum_{i=1}^n y_i (\ln(y_i) - b - ax_i)^2. \quad (4.43)$$

Điểm khác biệt so với phương pháp ở mục trước là nhân tử y_i được thêm vào trước bình phương của giá trị “lỗi” tại x_i . Điều này làm tăng ảnh hưởng của lỗi khi các giá trị y_i lớn và giảm nhẹ ảnh hưởng của lỗi khi y_i bé.

Điểm tới hạn của hàm $S_w(a, b)$ là nghiệm của hệ phương trình

$$\frac{\partial S_w(a, b)}{\partial a} = \frac{\partial S_w(a, b)}{\partial b} = 0, \quad (4.44)$$

trong đó, các đạo hàm riêng này là

$$\frac{\partial S_w(a, b)}{\partial a} = a \sum_{i=1}^n x_i^2 y_i + b \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i y_i \log(y_i) \quad (4.45)$$

$$\frac{\partial S_w(a, b)}{\partial b} = a \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n y_i - \sum_{i=1}^n y_i \log(y_i). \quad (4.46)$$

Thay vào (4.44), chúng ta thu được hệ phương trình tuyến tính không suy biến đối với hai biến a và b . Lưu ý, do $y_i > 0$ nên cũng theo BĐT Buniakovski–Cauchy–Schwarz, chúng ta có định thức của ma trận hệ số là

$$\Delta = \left(\sum_{i=1}^n x_i^2 y_i \right) \left(\sum_{i=1}^n y_i \right) - \left(\sum_{i=1}^n x_i y_i \right)^2 \geq 0,$$

trong khi dấu bằng chỉ xảy ra khi các giá trị x_i như nhau. Từ đó, công thức Cramer cho chúng ta

$$a = \left(\sum_{i=1}^n x_i y_i \ln(y_i) \right) \left(\sum_{i=1}^n y_i \right) - \left(\sum_{i=1}^n y_i \ln(y_i) \right) \left(\sum_{i=1}^n x_i y_i \right) / \Delta.$$

Khi đã tính được a , chúng ta có thể thay vào phương trình thứ hai để tính

$$b = \left\{ \left(\sum_{i=1}^n y_i \ln(y_i) \right) - a \left(\sum_{i=1}^n x_i y_i \right) \right\} / \left(\sum_{i=1}^n y_i \right).$$

Tính toán cụ thể (CAS 4.14) chúng ta thu được các hệ số a và b . Đồ thị của hàm số $y = \exp(b + ax)$ là đường màu xanh trong Hình 4.4.

CAS 4.14. Chúng ta viết chương trình R vẽ hình minh họa các dữ liệu và đường xấp xỉ bằng hàm mũ trong hai trường hợp có trọng và không có trọng. Kết quả thu được là các hệ số

$$a = 25,69034, \quad b = -0,758393,$$

cùng với Hình 4.4.

Nhận xét 4.2. Trong Hình 4.4, hai đường đồ thị thể hiện hai mô hình khá gần nhau khi các giá trị y nhỏ. Hai đường đồ thị rời xa nhau với y lớn hơn, đường màu xanh đi sát các điểm dữ liệu hơn. Kết quả này phù hợp với cách xây dựng “hàm lỗi” (hàm $S_w(a, b)$ trong (4.43)) đặt trọng nhiều hơn vào các sai số khi y lớn.

4.7 Câu hỏi và bài tập

1. Giả sử hàm số $y = f(x)$ có giá trị tại một số điểm cho bởi bảng sau:

x	-1	0	1	2
y	-1	4	15	50

Hãy xây dựng đa thức nội suy Lagrange $P(x)$ của $f(x)$ và tính gần đúng giá trị $f(1,2)$ bằng cách lấy $f(1,2) \approx P(1,2)$.

[Trả lời: $P(x) = 4 + 5x + 3x^2 + 3x^3$, $f(1,2) \approx P(1,2) \approx 19,504$]

2. Giả sử hàm số $y = f(x)$ có giá trị tại một số điểm cho bởi bảng sau:

x	0	2	4	5
y	-2	8	115	150

- (a) Hãy xây dựng đa thức nội suy Lagrange $P(x)$ của $f(x)$ và tính gần đúng giá trị $f(1)$ bằng cách lấy $f(1) \approx P(1)$.
- (b) Giả sử ta biết rằng hàm $f(x)$ có đạo hàm cấp bốn bị chặn bởi 10, tức là $|f^{(4)}(x)| \leq 10$ trên đoạn $[0,5]$. Ước lượng sai số của phép xấp xỉ $f(1)$ bởi $P(1)$.
3. Hãy xây dựng đa thức nội suy Newton tiến $P_3(x)$ tại 4 nút nội suy $x_0 = -2, x_1 = -1, x_2 = 0$, và $x_3 = 1$ của hàm số $y = f(x) = 2^x$. Từ đó tính xấp xỉ giá trị $\sqrt{2}$ bằng $P(0,5)$.
- Thêm nút nội suy $x_4 = 2$. Hãy xây dựng đa thức nội suy Newton tiến $P_4(x)$ bậc 4 tương ứng. Từ đó xấp xỉ giá trị $\sqrt{8}$ bằng $P_4(1,5)$.
4. Giả sử $\ell_i, i = 0, 1, \dots, n$ là các đa thức nội suy cơ bản của các điểm nội suy $x_0, x_1, \dots, x_n$, cho bởi Định lý 4.3. Chứng minh rằng

$$\sum_{i=0}^n \ell_i(x) = 1,$$

với mọi x .

5. Trong một thí nghiệm đo độ cứng của lò xo, người ta thu được bảng số liệu về độ lớn của lực đàn hồi và độ biến dạng của lò xo như sau:

F_i (N)	0,49	0,98	1,47	1,96	2,45
s_i (cm)	1,5	3,1	4,6	6,3	7,9

Giả thiết rằng độ lớn của lực đàn hồi tỉ lệ thuận với độ biến dạng của lò xo, $F = ks$ (định luật Hooke). Hãy xác định giá trị của k (đơn vị N/m) từ bảng số liệu trên theo phương pháp bình phương bé nhất. Viết kết quả với 4 chữ số chắc.

Trả lời: $k \approx 0,3126$ (N/cm).

6. Thu thập dữ liệu thực tế, người ta thấy rằng chiều cao trung bình c của trẻ em theo từng lứa tuổi t được cho bởi bảng sau:

c	75	92	108	121	130	142	155
t	1	3	5	7	9	11	13

Hãy xây dựng công thức biểu diễn c qua t dạng tuyến tính bằng phương pháp xấp xỉ bình phương bé nhất.

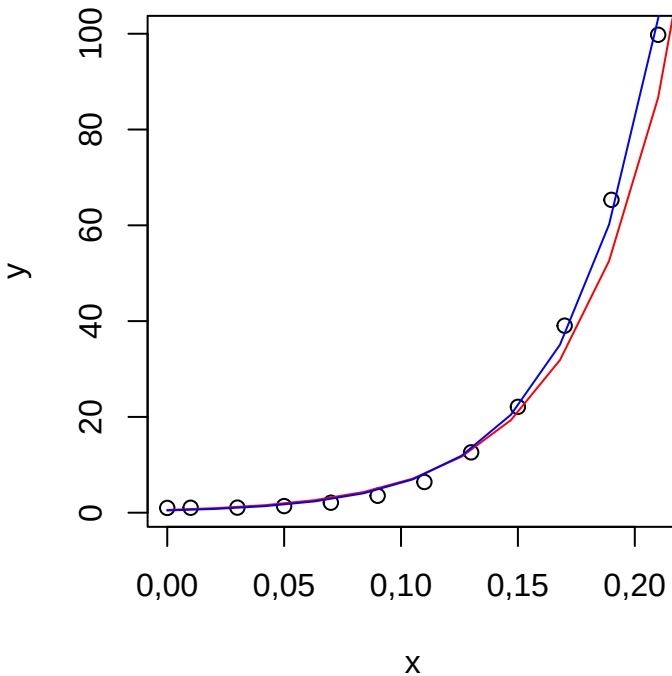
Trả lời: Nếu $c = at + b$ là hàm xấp xỉ bình phương tối thiểu dữ liệu trên thì

$$\begin{bmatrix} 7 & 49 \\ 49 & 455 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 823 \\ 6485 \end{bmatrix}.$$

Từ đó giải được $b \approx 72,3$ và $a \approx 6,46$.

```
plot(y ~ x)
f <- function(x) {
  exp(model$coefficients[1] + model$coefficients[2] * x)
}
curve(f, from = 0, to = 2.1, col = "red", add = TRUE)
ylny <- y * log(y)
a <- (sum(x * ylny) * sum(y) - sum(ylny) * sum(x * y)) /
      (sum(x * x * y) * sum(y) - (sum(x * y))^2)
b <- (sum(ylny) - a * sum(x * y)) / sum(y)

# Mô hình  $y = g(x)$  cực tiểu hoá hàm lỗi có trọng.
g <- function(x) exp(b + a * x)
curve(g, from = 0, to = 2.1, col = "blue", add = TRUE)
```



Hình 4.4: Dữ liệu và mô hình trong CAS 4.13

Chương 5

Tính gần đúng đạo hàm và tích phân xác định

Trong chương này, chúng ta tìm hiểu một số phương pháp đạo hàm và tính gần đúng tích phân xác định, bao gồm phương pháp tính đạo hàm bằng khai triển Taylor, công thức quy tắc hình thang, quy tắc Simpson, và công thức cầu phương Gauss.

Nội dung

5.1	Tính gần đúng tích phân xác định	151
5.1.1	Quy tắc hình thang	152
5.1.2	Quy tắc Simpson	157
5.1.3	Quy tắc hình thang hiệu chỉnh biên	164
5.1.4	Công thức cầu phương Gauss	167
5.1.5	Tính tích phân bằng hàm có sẵn trong R	173
5.2	Tính gần đúng đạo hàm	175
5.2.1	Xây dựng công thức tính đạo hàm bằng khai triển Taylor	175
5.2.2	Tính gần đúng đạo hàm bằng đa thức nội suy	179
5.3	Câu hỏi và bài tập	183

5.1 Tính gần đúng tích phân xác định

Trong mục này, chúng ta tìm hiểu một số phương pháp tính gần đúng tích phân xác định, bao gồm công thức hình thang, công thức Simpson, công

thức cầu phương Gauss. Tài liệu tham khảo: T. V. Đĩnh [6, §5.2], P. K. Anh [2, §V.2], Kreyszig [9, §19.5].

Chúng ta hãy bắt đầu với một tích phân xác định

$$I = \int_a^b f(x) dx. \quad (5.1)$$

Nếu $f(x)$ khả tích Riemann trên $[a, b]$ và $F(x)$ là một nguyên hàm của $f(x)$ trên (a, b) , liên tục trên $[a, b]$ thì, theo định lý cơ bản của giải tích (còn gọi là công thức Newton–Leibniz), ta có

$$I = \int_a^b f(x) dx = F(b) - F(a).$$

Tuy nhiên, trong thực tiễn, có rất nhiều tích phân xác định dạng (5.1) mà nguyên hàm của hàm dưới dấu tích phân $f(x)$ không có công thức đơn giản và tường minh. Một trong những tích phân như vậy là tích phân Gauss

$$\int_0^1 e^{-x^2} dx, \quad (5.2)$$

xuất hiện rất nhiều trong các bài toán thống kê liên quan đến phân phối xác suất chuẩn tắc (normal distribution). Trong những trường hợp như vậy, người ta không thể dùng công thức Newton–Leibniz để tính I . Từ đây, nảy sinh vấn đề tính gần đúng các tích phân xác định.

Trong các mục tiếp theo của chương này, chúng ta tìm hiểu một số phương pháp tính gần đúng tích phân xác định và cách thiết lập chúng trong máy tính.

5.1.1 Quy tắc hình thang

Xét một tích phân xác định

$$I = \int_a^b f(x) dx,$$

với $a < b$ và $f(x)$ là hàm liên tục trên $[a, b]$. Chúng ta chia đoạn $[a, b]$ thành n đoạn có độ dài bằng nhau bởi $n + 1$ điểm chia

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b.$$

Khi đó, ta có mỗi đoạn có độ dài

$$h = \frac{b - a}{n},$$

còn các điểm chia được cho bởi

$$x_i = a + ih, \quad i = 0, 1, 2, \dots, n.$$

Nếu với mỗi $i \in \{1, 2, \dots, n\}$, ta chọn một điểm $c_i \in [x_{i-1}, x_i]$ và thiết lập tổng

$$\sigma(h, \{c_j\}) := \sum_{i=1}^n f(c_i)h,$$

thì $\sigma(h, \{c_j\})$ là một tổng Riemann của tích phân I . Nếu f khả tích Riemann thì tổng Riemann $\sigma(h, \{c_j\})$ hội tụ tới I khi $h \rightarrow \infty$ mà không phụ thuộc vào cách chọn các điểm c_j . Đặc biệt, các tổng Riemann trái σ_T và tổng Riemann phải σ_P , cho bởi,

$$\sigma_T := \sum_{i=1}^n f(x_{i-1})h, \quad \sigma_P := \sum_{i=1}^n f(x_i)h$$

đều hội tụ tới I khi $h \rightarrow 0$. Với h đủ nhỏ, trung bình cộng của σ_T và σ_P , gọi là I_T là một xấp xỉ của I .

$$\begin{aligned} I \approx I_T &:= \frac{\sigma_T + \sigma_P}{2} = \frac{1}{2} \sum_{i=1}^n (f(x_{i-1}) + f(x_i))h \\ &= \frac{h}{2} (f(x_0) + 2f(x_1) + \dots + 2f(x_{n-1}) + f(x_n)). \end{aligned}$$

Như vậy, chúng ta có thể lấy I_T là một xấp xỉ của I . Sai số của phép xấp xỉ này có thể được đánh giá thông qua đạo hàm cấp hai f'' của f như trong định lí sau đây:

Định lí 5.1:

Giả sử f có đạo hàm cấp hai f'' trên $[a, b]$ và $|f''(x)| \leq M$ (M là hằng số) trên $[a, b]$. Khi đó,

$$|I - I_T| \leq \frac{M}{12} h^2 (b - a). \quad (5.3)$$

Chứng minh. Gọi $L_i(x)$ là hàm tuyến tính nội suy hàm $f(x)$ tại hai nút x_i và x_{i+1} . Khi đó, dễ dàng thấy rằng

$$\int_{x_i}^{x_{i+1}} L_i(x) dx = \frac{1}{2} (f(x_i) + f(x_{i+1}))h.$$

(Trong trường hợp $f > 0$, đẳng thức trên chính là công thức tính diện tích hình thang vuông giới hạn bởi trục hoành, các đường $x = x_i$, $x = x_{i+1}$, và đồ thị của L_i .) Từ đó suy ra

$$I_T = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} L_i(x) dx.$$

Mặt khác, theo công thức đánh giá sai số nội suy:

$$f(x) - L_i(x) = \frac{f''(c)}{2}(x - x_i)(x - x_{i+1})$$

với $c \in [x_i, x_{i+1}]$. Từ đó, sử dụng giả thiết $|f''(x)| \leq M$, ta suy ra

$$|f(x) - L_i(x)| \leq \frac{M}{2} |(x - x_i)(x - x_{i+1})| = \frac{M}{2} (x - x_i)(x_{i+1} - x).$$

Lấy tích phân hai vế,

$$\begin{aligned} \int_{x_i}^{x_{i+1}} |f(x) - L_i(x)| dx &\leq \frac{M}{2} \int_{x_i}^{x_{i+1}} (x - x_i)(x_{i+1} - x) dx \\ &= \frac{M}{2} \int_0^h t(h - t) dt \\ &= \frac{M}{2} \left(\frac{ht}{2} - \frac{t^3}{3} \right) \Big|_{t=0}^h \\ &= \frac{Mh^3}{12}. \end{aligned}$$

Ở trên, ta đã đổi biến $x = x_i + t$, $h = x_{i+1} - x_i$. Từ đó suy ra

$$\begin{aligned} \left| \int_{x_i}^{x_{i+1}} f(x) dx - \int_{x_i}^{x_{i+1}} L_i(x) dx \right| &= \left| \int_{x_i}^{x_{i+1}} (f(x) - L_i(x)) dx \right| \\ &\leq \int_{x_i}^{x_{i+1}} |f(x) - L_i(x)| dx \\ &\leq \frac{Mh^3}{12}. \end{aligned}$$

Từ đó suy ra

$$\begin{aligned}
 |I - I_T| &= \left| \int_a^b f(x) dx - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} L_i(x) dx \right| \\
 &= \left| \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x) dx - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} L_i(x) dx \right| \\
 &\leq \sum_{i=0}^{n-1} \left| \int_{x_i}^{x_{i+1}} f(x) dx - \int_{x_i}^{x_{i+1}} L_i(x) dx \right| \\
 &= \sum_{i=0}^{n-1} \frac{M h^3}{12} \\
 &= \frac{M h^2}{12} (b-a), \quad \text{bởi vì } n = h/(b-a).
 \end{aligned}$$

Ta được đánh giá cần chứng minh. $\square$

Ví dụ 5.1. Bằng quy tắc hình thang với 3 điểm chia, tính gần đúng tích phân

$$I = \int_0^1 \frac{4 dx}{1+x^2}.$$

Lời giải

Với $n = 2$, ta có $h = (b-a)/2 = 1/2$, chúng ta lấy các điểm chia $x_0 = 0$, $x_1 = 1/2$, và $x_2 = 1$. Khi đó

$$I_T = \frac{1}{4} (f(0) + 2f(1/2) + f(1)) = \frac{1}{4} \left(4 + 2 \left(\frac{16}{5} \right) + 2 \right) = \frac{31}{10}.$$

Chú ý rằng,

$$I = \int_0^1 \frac{4 dx}{1+x^2} = 4 \arctan(x) \Big|_0^1 = \pi.$$

Do đó, với $\pi \approx 3,1415926$, ta thu được

$$|I - I_T| = \pi - 3,1 \leq 0,0416.$$

Mặt khác, ta có thể dùng đánh giá sai số trong Định lý 5.1 như sau. Với $f(x) = 4/(1+x^2)$, ta có

$$f''(x) = \frac{32x^2}{(1+x^2)^3} - \frac{8}{(1+x^2)^2}.$$

Ta chọn $M = 8$ là hằng số dương bé nhất sao cho $|f''(x)| \leq M$ trên $[0, 1]$. Định lý đó cho ta đánh giá sai số

$$|I - I_T| \leq \frac{M}{12}(b-a)h^2 = \left(\frac{8}{12}\right)\left(\frac{1}{2}\right)^2 = 0.17.$$

Trong Ví dụ 5.1, ta cũng có thể tính tích phân đã cho bằng cách lấy nguyên hàm và dễ dàng so sánh hai phương pháp và đánh giá sai số. Các phương pháp tính gần đúng sẽ hữu ích hơn trong tình huống mà tích phân không tính được bằng nguyên hàm, như trong ví dụ sau đây:

Ví dụ 5.2. Tính gần đúng tích phân $\int_0^1 e^{-x^2} dx$ bằng quy tắc hình thang với $n = 10$.

Lời giải

Ta có khoảng lấy tích phân là $[a, b] = [0, 1]$. Với $n = 10$, ta có $h = 0.1$. Do số điểm chia khá lớn, để thuận tiện ta sẽ lập bảng các giá trị cần tính. Từ Bảng

i	x_i	x_i^2	$e^{-x_i^2}$
0	0	0	1,00000
1	0,1	0,01	0,990050
2	0,2	0,04	0,960789
3	0,3	0,09	0,913931
4	0,4	0,16	0,852144
5	0,5	0,25	0,778801
6	0,6	0,36	0,697676
7	0,7	0,49	0,612626
8	0,8	0,64	0,527292
9	0,9	0,81	0,444858
10	1,0	1,00	0,367879
Tổng			1,367879 6,778167

Bảng 5.1: Tính toán trong Ví dụ 5.2

5.1, ta có

$$I_T \approx \frac{0,1}{2}(1,367879 + 2(6,778167)) = 0,746211.$$

CAS 5.1. Đo cường độ dòng điện $I(t)$ qua một mạch điện tại các thời điểm cách nhau 1s trong 6s, người ta thu được các kết quả ghi trong bảng sau:

t	0	1	2	3	4	5	6
$I(t)$	80,00	92,61	106,48	121,67	138,24	156,25	175,76

Biết rằng điện lượng Q chuyển qua đoạn mạch trong khoảng thời gian từ t_1 đến t_2 cho bởi

$$Q(t_1; t_2) = \int_{t_1}^{t_2} I(t) dt.$$

Sau đây là một chương trình R để tính gần đúng $Q(0; 6)$ theo quy tắc hình thang. Lưu ý, với những thông tin có được, không thể có đánh giá sai số của kết quả này.

```
# Hàm trapz() nhận vào một dãy giá trị y và bước nhảy h,
# cho ra xấp xỉ của tích phân theo phương pháp hình thang.

trapz = function(y, h) {
  n <- length(y)
  s1 = (y[1] + y[n]) / 2
  s2 = sum(y[2:(n - 1)])
  return(h * (s1 + s2))
}

# Chạy thử trapezoid() với một dữ liệu cụ thể.

y <- c(80, 92.61, 106.48, 121.67, 138.24, 156.25, 175.76)
h <- 1
trapz(y, h)

## [1] 743,13
```

Bằng phương pháp này, chúng ta thu được xấp xỉ $Q(0; 6) \approx 743,13$.

5.1.2 Quy tắc Simpson

Nội dung của mục này là quy tắc Simpson¹ (còn gọi là quy tắc parabol) tính gần đúng một tích phân xác định.

Giả sử $[a, b]$ được chia thành $2n$ đoạn có độ dài bằng nhau bởi các điểm chia x_i , $i = 0, 1, \dots, 2n$, với

$$a = x_0 < x_1 < \dots < x_{2n-1} < x_{2n} = b.$$

Với mỗi i , gọi Q_i là đa thức bậc hai nội suy hàm $f(x)$ tại các nút $x_{2i}, x_{2i+1},$

¹Quy tắc Simpson để tính gần đúng các tích phân xác định được Thomas Simpson công bố lần đầu tiên vào năm 1743. Mặc dù được đặt theo tên của ông, Thomas Simpson không thực sự là người đầu tiên phát hiện ra nó. Bonaventura Cavalieri đã phát hiện ra phiên bản ban đầu của quy tắc này vào năm 1639 và James Gregory đã công bố biến thể này vào năm 1668 cùng với một số phương pháp số khác để tính gần đúng các tích phân xác định.

và x_{2i+2} . Trước hết, ta xét

$$I_i := \int_{x_{2i}}^{x_{2i+2}} Q_i(x) dx,$$

và xem nó là một xấp xỉ của tích phân của $f(x)$ trên đoạn $[x_{2i}, x_{2i+2}]$:

$$\int_{x_{2i}}^{x_{2i+2}} f(x) dx \approx I_i.$$

Bây giờ, ta tính I_i . Theo công thức nội suy Lagrange, ta có

$$Q_i(x) = f(x_{2i})\ell_0(x) + f(x_{2i+1})\ell_1(x) + f(x_{2i+2})\ell_2(x),$$

với $\ell_0(x)$, $\ell_1(x)$, và $\ell_2(x)$ là các đa thức nội suy Lagrange cơ bản. Cụ thể hơn, ta có

$$\ell_0(x) = \frac{(x - x_{2i+1})(x - x_{2i+2})}{(x_{2i} - x_{2i+1})(x_{2i} - x_{2i+2})}.$$

Với phép đổi biến $x = x_{2i} + ht$, $dx = h dt$, ta dễ dàng tính được

$$\int_{x_{2i}}^{x_{2i+2}} \ell_0(x) dx = \frac{h}{2} \int_0^2 (t-1)(t-2) dt = \frac{h}{3}.$$

Tương tự như vậy, với

$$\ell_1(x) = \frac{(x - x_{2i})(x - x_{2i+2})}{(x_{2i+1} - x_{2i})(x_{2i+1} - x_{2i+2})}$$

ta có

$$\int_{x_{2i}}^{x_{2i+2}} \ell_1(x) dx = -h \int_0^2 t(t-2) dt = \frac{4h}{3}.$$

Cuối cùng, với

$$\ell_2(x) = \frac{(x - x_{2i})(x - x_{2i+1})}{(x_{2i+2} - x_{2i})(x_{2i+2} - x_{2i+1})},$$

ta tính được

$$\int_{x_{2i}}^{x_{2i+2}} \ell_2(x) dx = \frac{h}{2} \int_0^2 t(t-1) dt = \frac{h}{3}.$$

Tổng hợp các tính toán trên, ta thu được hệ quả

$$\begin{aligned}
 I_i &= \int_{x_{2i}}^{x_{2i+2}} Q_i(x) dx \\
 &= f(x_{2i}) \int_{x_{2i}}^{x_{2i+2}} \ell_0(x) dx + f(x_{2i+1}) \int_{x_{2i}}^{x_{2i+2}} \ell_1(x) dx \\
 &\quad + f(x_{2i+2}) \int_{x_{2i}}^{x_{2i+2}} \ell_2(x) dx \\
 &= \frac{h}{3} (f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})).
 \end{aligned}$$

Để xấp xỉ tích phân của f trên toàn bộ đoạn $[a, b]$, ta lấy tổng các I_i và gọi nó là I_S :

$$\begin{aligned}
 I_S := \sum_{i=0}^{n-1} I_i &= \frac{h}{3} \left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots \right. \\
 &\quad \left. \dots + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n}) \right). \quad (5.4)
 \end{aligned}$$

Khi đó I_S là một xấp xỉ của I . Sau đây, chúng ta cần đánh giá sai số của phép xấp xỉ I bởi I_S .

Định lí 5.2: Sai số toàn cục

Giả sử f có đạo hàm cấp bốn $f^{(4)}$ trên $[a, b]$, $|f^{(4)}(x)| \leq M$ (M là hằng số) trên $[a, b]$. Khi đó

$$|I - I_S| \leq \frac{M}{180} h^4 (b - a). \quad (5.5)$$

Để chứng minh định lí này, ta cần bổ đề sau đây về đánh giá sai số địa phương của phương pháp Simpson, tức là đánh giá sai số trên mỗi đoạn nhỏ.

Định lí 5.3: Sai số địa phương

Giả sử f có đạo hàm cấp 4. Khi đó,

$$\int_{x_{i-1}}^{x_{i+1}} f(x) dx = \frac{h}{3} [f(x_{i-1}) + 4f(x_i) + f(x_{i+1})] - \frac{h^5}{90} f^{(4)}(c), \quad (5.6)$$

với $c \in [x_{i-1}, x_{i+1}]$.

Chứng minh. Với mỗi t , ta gọi phần dư $R(t)$ cho bởi

$$R(t) = \int_{x_i-t}^{x_i+t} f(x) dx - \frac{t}{3} [f(x_i-t) + 4f(x_i) + f(x_i+t)]$$

Ta cần chứng minh

$$R(h) = -\frac{h^5}{90} f^{(4)}(c),$$

với $c \in [x_{i-1}, x_{i+1}]$. Thật vậy, bởi công thức đạo hàm tích phân phụ thuộc tham số, ta có

$$\frac{d}{dt} \int_{x_i-t}^{x_i+t} f(x) dx = f(x_i-t) + f(x_i+t).$$

Từ đó, ta tiếp tục tính

$$R'(t) = \frac{2}{3} [f(x_i-t) + f(x_i+t)] - \frac{4}{3} f(x_i) - \frac{t}{3} [-f'(x_i-t) + f'(x_i+t)].$$

Nói riêng,

$$R'(0) = 0.$$

Tiếp tục lấy đạo hàm, ta có

$$\begin{aligned} R''(t) &= \frac{2}{3} [-f'(x_i-t) + f'(x_i+t)] - \frac{1}{3} [-f'(x_i-t) + f'(x_i+t)] \\ &\quad - \frac{t}{3} [f''(x_i-t) + f''(x_i+t)] \\ &= \frac{1}{3} [-f'(x_i-t) + f'(x_i+t)] - \frac{t}{3} [f''(x_i-t) + f''(x_i+t)]. \end{aligned}$$

Nói riêng, ta có

$$R''(0) = 0.$$

Tiếp tục lấy đạo hàm hai vế, ta có

$$R'''(t) = \frac{t}{3} [-f'''(x_i-t) + f'''(x_i+t)].$$

Xét hàm phụ sau:

$$\varphi(t) := R(t) - t^5 R(h)/(h^5).$$

Từ tính toán ở trên, ta dễ dàng thấy rằng

$$\varphi(0) = \varphi'(0) = \varphi''(0) = \varphi'''(0) = 0.$$

Hơn nữa,

$$\varphi(h) = 0.$$

Áp dụng định lí Rolle cho hàm φ trên đoạn $[0, h]$, ta nhận được một giá trị $c_1 \in [0, h]$ sao cho

$$\varphi'(c_1) = 0.$$

Tiếp theo, áp dụng định lí Rolle cho hàm $\varphi'(t)$ trên đoạn $[0, c_1]$, ta có một giá trị c_2 sao cho $\varphi''(c_2) = 0$. Tiếp tục áp dụng định lí Rolle, ta suy ra tồn tại $c \in (0, c_2)$ sao cho

$$\varphi'''(c_3) = 0.$$

Theo định nghĩa của φ , ta có

$$0 = \varphi'''(c_3) = R'''(c_3) - 60c_3^2 R(h)/(h^5).$$

Từ đây suy ra,

$$60c_3^2 R(h)/(h^5) = R'''(c_3) = \frac{c_3}{3} [-f'''(x_i - c_3) + f'''(x_i + c_3)] = \frac{2c_3^2}{3} f^{(4)}(c),$$

với $c \in (x_i - c_3, x_i + c_3)$. Ở bước cuối cùng, ta đã áp dụng định lí Rolle một lần nữa cho hàm f''' . Từ đây, ta dễ dàng suy ra

$$R(h) = -\frac{h^5}{90} f^{(4)}(c).$$

Đó là điều phải chứng minh. □

Chứng minh Định lí 5.2. Với mỗi i , gọi $R_i(h)$ là sai số địa phương trong bố đề trên. Khi đó,

$$|R_i(h)| \leq h^5 M/90.$$

Lấy tổng với i chạy qua các số lẻ từ 1 đến $2n-1$, ta có ngay

$$\begin{aligned} |I - I_S| &\leq \sum_{i=1, i \text{ lẻ}}^{2n-1} |R_i(h)| \\ &\leq \sum_{i=1, i \text{ lẻ}}^{2n-1} h^5 M/90 \\ &= nh^5 M/90 \\ &= Mh^4(b-a)/180. \end{aligned}$$

Đó là điều phải chứng minh. □

Nhận xét 5.1. Từ Định lí 5.2, ta thấy rằng nếu $f(x)$ là một hàm đa thức với bậc bằng 3 hoặc bé hơn thì quy tắc Simpson cho một giá trị đúng vì đạo hàm cấp 4 của f đồng nhất bằng 0.

Trong ví dụ sau đây, ta dùng phương pháp Simpson để tính gần đúng giá trị của một hàm siêu việt bằng các phép tính số học thông qua việc viết giá trị cần tính dưới dạng một tích phân xác định.

Ví dụ 5.3. Dùng quy tắc Simpson để tính gần đúng $\ln(2)$ với 4 chữ số chắc theo nghĩa rộng.

Lời giải

Từ kiến thức cơ bản của giải tích, chúng ta có thể viết $\ln(2)$ dưới dạng một tích phân. Cụ thể như sau:

$$\ln(2) = \ln(1+x) \Big|_0^1 = \int_0^1 \frac{dx}{1+x}.$$

Vậy, bài toán quy về tính gần đúng tích phân trong vế phải của đẳng thức trên. Bằng quy tắc Simpson với $n=8$, ta có

$$I_S = \frac{1}{24} \left(1 + 4 \cdot \frac{8}{9} + 2 \cdot \frac{4}{5} + 4 \cdot \frac{8}{11} + 2 \cdot \frac{2}{3} + 4 \cdot \frac{8}{13} + 2 \cdot \frac{4}{7} + 4 \cdot \frac{8}{15} + \frac{1}{2} \right).$$

Tính và quy tròn I_S tới 6 chữ số chắc (theo nghĩa hẹp) ta thu được

$$I_S = \frac{1498711}{2162160} \approx 0,693155,$$

với sai số quy tròn không vượt quá $0,5 \cdot 10^{-4}$. Mặt khác, hàm dưới dấu tích phân là $f(x) = (1+x)^{-1}$ có đạo hàm cấp 4 là

$$f^{(4)}(x) = \frac{24}{(1+x)^5}.$$

Do đó, trên đoạn $[0, 1]$, ta có đánh giá $|f^{(4)}(x)| \leq 24$. Vậy,

$$|\ln(2) - I_S| = |I - I_S| \leq 24 \cdot \frac{(0,125)^4}{180} \approx 0,4 \times 10^{-4}.$$

Từ đó, ta có xấp xỉ $\ln(2) \approx 0,693155$, với sai số không vượt quá tổng của sai số quy tròn và sai số phương pháp, tức là không vượt quá $(0,4 + 0,5) \times 10^{-4} < 0,9 \cdot 10^{-4}$, thỏa mãn yêu cầu đặt ra.

Ví dụ 5.4. Tính gần đúng tích phân $I = \int_0^1 x^x dx$ bằng quy tắc Simpson với $n=4$.

Lời giải

Với $n = 4$, ta chia đoạn $[0, 1]$ thành 4 đoạn nhỏ có độ dài bằng nhau và bằng $\frac{1}{4}$. Tính giá trị của hàm $f(x) = x^x$ tại các điểm chia này, ta thu được bảng sau:

x_i	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$y_i = x_i^{x_i}$	1	$\sqrt{\frac{1}{2}}$	$\sqrt{\frac{1}{2}}$	$\frac{1}{2}\sqrt{\frac{3\sqrt{3}}{2}}$	1

Ở trên, ta đã đặt $f(0) := \lim_{x \rightarrow 0+} x^x = 1$ để mở rộng liên tục $f(x)$ lên $[0, 1]$. Trong quy tắc Simpson, ta có $h = \frac{1}{4}$, nên

$$\begin{aligned}
 I_S &= \frac{1}{12} \left(f(0) + 4f\left(\frac{1}{4}\right) + 2f\left(\frac{1}{2}\right) + 4f\left(\frac{3}{4}\right) + f(1) \right) \\
 &= \frac{1}{12} \left(1 + 4\left(\sqrt{\frac{1}{2}}\right) + 2\left(\sqrt{\frac{1}{2}}\right) + 4\left(\frac{1}{2}\sqrt{\frac{3\sqrt{3}}{2}}\right) + 1 \right) \\
 &= \frac{1}{12} \left(2 + 3\sqrt{2} + \sqrt{6\sqrt{3}} \right).
 \end{aligned}$$

Biểu diễn gần đúng các căn thức bằng số thập phân, chúng ta có

$$\int_0^1 x^x dx \approx I_S \approx 0,7888625.$$

Ví dụ 5.5. Giả sử hàm $f(x)$ liên tục trên $[0; 0,4]$ và có các giá trị tại một số điểm được tính xấp xỉ như sau:

x_i	0,0	0,1	0,2	0,3	0,4
$f(x_i)$	1,000	0,998	0,990	0,978	0,860

Hãy sử dụng dữ kiện này tính xấp xỉ tích phân $I = \int_0^{0,4} f(x) dx$ bằng quy tắc Simpson.

Lời giải

Từ dữ kiện đã cho, ta suy ra $n = 4$, $h = 0,1$. Do đó

$$I_S = \frac{0,1}{3} (f(0,0) + 4f(0,1) + 2f(0,2) + 4f(0,3) + f(0,4)).$$

Thay các giá trị của f tại các điểm đã cho từ bảng trên, ta có

$$I_S \approx \frac{0,1}{3} (1,000 + 4(0,998) + 2(0,990) + 4f(0,978) + 0,860) = 0,3914\bar{6}.$$

CAS 5.2. Hàm `simpson()` trong đoạn chương trình R sau đây nhận vào một dãy các giá trị y và độ dài khoảng chia h và trả về giá trị xấp xỉ của tích phân theo quy tắc Simpson.

```
# Hàm simpson() nhận vào một dãy giá trị y và bước đi h,
# cho ra xấp xỉ của tích phân theo công thức Simpson.

simpson <- function(y, h) {
  n = length(y)
  if (n %% 2 != 1 | n == 1) {
    stop("Lỗi! Số điểm chia không đúng.")
  }
  s = y[1] + 4 * y[n - 1] + y[n]
  i = 2
  while (i < n - 2) {
    s = s + 4 * y[i] + 2 * y[i + 1]
    i = i + 2
  }
  approx = s * h / 3
  return(approx)
}

# Ví dụ áp dụng

y = c(1, 0.998, 0.990, 0.978, 0.860)
h = 0.1
simpson(y, h)

## [1] 0.391466667
```

Kết quả thu được hoàn toàn khớp với kết quả đã tính bằng tay trong Ví dụ 5.5.

5.1.3 Quy tắc hình thang hiệu chỉnh biên

Mục này nói về quy tắc hình thang hiệu chỉnh biên (trapezoidal rule with end-point correction) để tính gần đúng tích phân. Trong quy tắc này, chúng ta cần biết giá trị của hàm số tại các điểm lưới và giá trị của đạo hàm tại hai đầu mút.

Quy tắc này phát biểu như sau:

$$\int_a^b f(x) dx \approx I_T + \frac{h^2}{12} (f'(a) - f'(b)), \quad (5.7)$$

trong đó I_T là xấp xỉ của tích phân theo công thức hình thang,

$$I_T = \frac{h}{2} \left(f(a) + f(b) + 2 \sum_{i=1}^{n-1} f(a + ih) \right), \quad h = (b - a)/n.$$

Một điều thú vị là chúng ta có thể thu được công thức xấp xỉ này bằng cách thay hàm $f(x)$ dưới dấu tích phân bằng đa thức nội suy Hermite bậc ba của nó.

Định lí 5.4:

Giả sử $g(x)$ là một hàm đa thức bậc không vượt quá 3. Khi đó,

$$\int_a^b g(x) dx = \frac{b-a}{2} (g(a) + g(b)) + \frac{(b-a)^2}{12} (g'(a) - g'(b)). \quad (5.8)$$

Chứng minh. Một hàm đa thức bậc không vượt quá 3 có dạng tổng quát là

$$g(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3,$$

trong đó $c_0, \dots, c_3$ là các hằng số, tức là một “tổ hợp tuyến tính” của các đơn thức $1, x, x^2, x^3$. Mặt khác, hai vế của (5.8) là tuyến tính theo g (tức là các ánh xạ tuyến tính từ không gian vectơ các đa thức bậc không vượt quá 3 vào $\mathbb{R}^1$). Do đó, chỉ cần chứng minh (5.8) cho các đơn thức $1, x, x^2, x^3$. Chẳng hạn, với $g(x) = x^3$, vế trái có thể tính như sau:

$$\begin{aligned} \int_a^b g(x) dx &= \int_a^b x^3 dx \\ &= \frac{1}{4} x^4 \Big|_a^b = \frac{1}{4} (b^4 - a^4) \\ &= \frac{1}{4} (b-a)(b^3 + b^2 a + b a^2 + a^3). \end{aligned}$$

Mặt khác, vế phải có thể biến đổi như sau:

$$\begin{aligned} \frac{b-a}{2} (g(a) + g(b)) + \frac{(b-a)^2}{12} (g'(a) - g'(b)) \\ &= \frac{b-a}{2} (a^3 + b^3) + \frac{(b-a)^2}{12} (3a^2 - 3b^2) \\ &= \frac{1}{4} (b-a)(b^3 + b^2 a + b a^2 + a^3). \end{aligned}$$

So sánh hai tính toán trên, chúng ta kết luận được (5.8) đúng khi $g(x) = x^3$.

Các trường hợp còn lại đơn giản hơn. Chi tiết dành cho bạn đọc. $\square$

Từ mệnh đề trên, có thể thấy rằng công thức xấp xỉ hình thang thu được từ việc xấp xỉ hàm số dưới dấu tích phân bằng các hàm ghép trơn bậc ba bị kẹp. Lưu ý, công thức hình thang hiệu chỉnh còn là một trường hợp đặc biệt của các công thức xấp xỉ tích phân Euler–Maclaurin.

CAS 5.3. Hãy lập chương trình R tính gần đúng tích phân bằng công thức hình thang với hiệu chỉnh tại các đầu mút, áp dụng vào tích phân

$$\int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4}$$

Kết quả thu được cho một xấp xỉ của $\pi/4$.

```
# Hàm trapezoid_endpoint_correction() tính xấp xỉ tích phân theo
# công thức hình thang hiệu chỉnh.
# Tham số: y là dãy các giá trị, k0, k1 là các đạo hàm tại hai
# đầu mút, n là số khoảng chia và h là độ dài mỗi khoảng chia.

trapezoid_endpoint_correction = function(y, k0, k1, h) {
  n = length(y)
  s1 = (y[1] + y[n]) / 2
  s2 = sum(y[2:(n - 1)])
  cor = h * (k0 - k1) / 12
  return(h * (s1 + s2 + cor))
}

# Định nghĩa hàm số f(x)
f <- function(x) {
  1 / (1 + x^2)
}

# Đạo hàm f'(x)
f_prime <- function(x) {
  - 2 * x / ((1 + x^2)^2)
}

# Số điểm chia n và độ dài khoảng chia h
n <- 10
h <- 1 / 10

# Lưới các điểm cần tính giá trị
x <- seq(from = 0, to = 1, length.out = (n + 1))

# Dãy các giá trị y tại các giá trị x
y <- f(x)
```

```
# Các đạo hàm tại hai đầu mút
k0 <- f_prime(0)
k1 <- f_prime(1)

# Tính giá trị xấp xỉ bằng công thức hình thang hiệu chỉnh
result <- trapezoid_endpoint_correction(y, k0, k1, h)
print(result)

## [1] 0,785398164
```

Bạn đọc có thể so sánh giá trị tính được và giá trị xấp xỉ của $\pi/4$ và tự rút ra nhận xét cho mình.

5.1.4 Công thức cầu phương Gauss

Các quy tắc hình thang và quy tắc Simpson ta đã tìm hiểu trong mục trước là hai trường hợp riêng của *công thức Newton–Cotes đóng*. Chúng có dạng

$$\int_a^b f(x) dx \approx c_0 f(x_0) + c_1 f(x_1) + \cdots + c_n f(x_n),$$

trong đó, các điểm x_j cách đều, $x_0 = a$, $x_n = b$, còn các hệ số c_j được tính dựa trên việc lấy tích phân các đa thức nội suy Lagrange bậc d , $n \geq 1$. quy tắc hình thang tương ứng với trường hợp $d = 1$, còn quy tắc Simpson tương ứng với trường hợp $d = 2$.

Lưu ý là đối với các hàm đa thức bậc không vượt quá 3, công thức Simpson cho giá trị đúng của tích phân từ giá trị của hàm số tại 3 nút. Một câu hỏi tự nhiên đặt ra là có cách chọn các nút $x_0, x_1, \dots, x_n$ và một công thức xấp xỉ sao cho kết quả là đúng đối với các đa thức bậc khá lớn?

Liên quan đến câu hỏi trên, chúng ta giới thiệu phương pháp *cầu phương Gauss* (Gaussian quadrature rule), một công thức tính gần đúng tích phân, được thiết lập trong nhiều gói phần mềm tính toán hiện đại như gaussquad của phần mềm R và Matlab. Trong phương pháp này, các điểm nút x_j , $j \in \{0, 1, \dots, n\}$, không cách đều nhau và được chọn sao cho công thức xấp xỉ trở thành công thức đúng nếu hàm dưới dấu tích phân là đa thức có bậc không vượt quá $2n + 1$. Nói riêng, với 3 nút nội suy ($n = 2$), công thức cầu phương Gauss cho kết quả đúng nếu hàm dưới dấu tích phân là đa thức có bậc không vượt quá 5, tốt hơn so với quy tắc Simpson (chỉ đúng khi đa thức đó có bậc không vượt quá 3).

Phương pháp này được xây dựng dựa trên các *đa thức trực giao Legendre*² (Legendre orthogonal polynomial). Chi tiết như sau: Gọi P_k là đa thức

²Adrien-Marie Legendre (1752–1833) là một nhà toán học người Pháp.

Legendre bậc k . Với $k = 0, 1, 2$, công thức tường minh của chúng là

$$P_0(x) = 1, P_1(x) = x, P_2(x) = \frac{1}{2}(3x^2 - 1), \dots$$

Các đa thức Legendre bậc cao hơn có thể được tính theo công thức truy hồi Bonnet³

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad n \geq 1. \quad (5.9)$$

Ví dụ, với $n = 2$, thay vào (5.9), ta thu được

$$3P_3(x) = 5xP_2(x) - 2P_1(x) = 5x\left(\frac{1}{2}(3x^2 - 1)\right) - 2x.$$

Sau khi rút gọn về phải và chia hai vế cho 3, ta thu được

$$P_3(x) = \frac{1}{2}(5x^3 - 3x).$$

Đó là đa thức Legendre bậc 3.

Một tính chất quan trọng của các đa thức Legendre mà chúng ta cần để xây dựng công thức cầu phương Gauss đó là tính trực giao:

$$\int_{-1}^1 Q(x)P_k(x)dx = 0,$$

đúng với mọi đa thức $Q(x)$ có bậc bé hơn k . Từ tính chất này, ta có thể chứng minh kết quả sau.

Định lý 5.5:

Giả sử rằng $x_0, x_1, x_2, \dots, x_n$ là các nghiệm của đa thức Legendre thứ $n+1$, $P_{n+1}(x)$. Gọi ℓ_k là đa thức nội suy Lagrange cơ bản tương ứng với các nút nội suy $x_0, x_1, \dots, x_n$. Đặt

$$c_k = \int_{-1}^1 \ell_k(x)dx.$$

Khi đó, với mọi đa thức $P(x)$ có bậc không vượt quá $2n+1$,

$$\int_{-1}^1 P(x)dx = c_0P(x_0) + c_1P(x_1) + \dots + c_nP(x_n).$$

³Ossian Bonnet (1819–1892) là một nhà toán học người Pháp có nhiều đóng góp cho ngành hình học vi phân (một nhánh của toán học).

Chứng minh. Chia đa thức P cho P_{n+1} , ta được thương là Q và dư là R . Tức là

$$P(x) = P_{n+1}(x)Q(x) + R(x).$$

Do bậc của P không vượt quá $2n+1$, bậc của Q không vượt quá n còn bậc của R không vượt quá bậc của P_{n+1} . Do tính trực giao của đa thức Legendre, ta có

$$\int_{-1}^1 P(x) dx = \int_{-1}^1 Q(x)P_{n+1}(x) dx + \int_{-1}^1 R(x) dx = \int_{-1}^1 R(x) dx.$$

Theo công thức nội suy Lagrange

$$R(x) = R(x_0)\ell_0(x) + R(x_1)\ell_1(x) + \cdots + R(x_n)\ell_n(x)$$

và vậy thì

$$\begin{aligned} \int_{-1}^1 R(x) dx &= R(x_0) \int_{-1}^1 \ell_0(x) dx + R(x_1) \int_{-1}^1 \ell_1(x) dx \\ &\quad + \cdots + R(x_n) \int_{-1}^1 \ell_n(x) dx. \end{aligned} \quad (5.10)$$

Mặt khác, do x_j là nghiệm của đa thức Legendre $P_{n+1}(x)$

$$P(x_j) = P_{n+1}(x_j)Q(x_j) + R(x_j) = R(x_j).$$

Thay $R(x_j)$ và công thức c_j vào (5.10) ta thu được điều phải chứng minh. $\square$

Định nghĩa 5.1:

Giả sử $x_0, x_1, x_2, \dots, x_n$ nghiệm của đa thức Legendre thứ $n+1$, $P_{n+1}(x)$. Công thức cầu phương Gauss thứ n là công thức xấp xỉ

$$\int_{-1}^1 f(x) dx \approx c_0 f(x_0) + c_1 f(x_1) + \cdots + c_n f(x_n) \quad (5.11)$$

với c_j , $j = 0, 2, \dots, n$, được cho trong định lí 5.5.

Như vậy, để áp dụng công thức cầu phương Gauss n điểm, ta cần tìm các không điểm của đa thức Legendre cấp n và xác định các trọng số c_j . Dưới đây là xác định các giá trị này trong 2 trường hợp n bé.

Cầu phương Gauss 2 điểm

Cho $n = 2$, các điểm cầu phương Gauss là nghiệm của đa thức Legendre với $n = 2$, tức là nghiệm của

$$\frac{1}{2}(3x^2 - 1) = 0.$$

Giải phương trình, ta thu được

$$x_0 = \frac{\sqrt{3}}{3}, \quad x_1 = -\frac{\sqrt{3}}{3}$$

Đa thức nội suy Lagrange cơ bản tương ứng với x_0 là $\ell_0(x) = (x - x_1)/(x_0 - x_1)$. Do vậy,

$$c_0 = \int_{-1}^1 \ell_0(x) dx = \int_{-1}^1 \frac{\sqrt{3}}{2} \left(x + \frac{\sqrt{3}}{3} \right) dx = 1.$$

Tương tự, ta cũng tính được $c_1 = 1$. Do đó, công thức cầu phương Gauss đối với $n = 2$ là:

$$\int_{-1}^1 f(x) dx \approx f\left(\frac{\sqrt{3}}{3}\right) + f\left(-\frac{\sqrt{3}}{3}\right). \quad (5.12)$$

Ví dụ, lấy $f(x) = 1/(3+x)$ với

$$\int_{-1}^1 f(x) dx = \int_{-1}^2 \frac{dx}{3+x} = \ln(3+x) \Big|_{-1}^1 = \ln(4) - \ln(2) = \ln(2),$$

và áp dụng công thức cầu phương Gauss 2 điểm, ta thu được

$$\int_{-1}^1 f(x) dx \approx \frac{1}{3 - \frac{\sqrt{3}}{3}} + \frac{1}{3 + \frac{\sqrt{3}}{3}} = \frac{9}{13} \approx 0,6923077 \dots$$

Giá trị gần đúng thu được chính xác tới 2 chữ số sau dấu thập phân (0,69).

Cầu phương Gauss 3 điểm

Cho $n = 3$, các điểm cầu phương Gauss là nghiệm của đa thức Legendre với $n = 3$, tức là nghiệm của

$$\frac{1}{2}(5x^3 - 3x) = 0.$$

Ta được các điểm cầu phương

$$x_0 = -\sqrt{\frac{3}{5}}, \quad x_1 = 0, \quad x_2 = \sqrt{\frac{3}{5}}.$$

Bây giờ, ta cần xác định các trọng số c_0, c_1, c_2 trong công thức

$$\int_{-1}^1 f(x) dx = c_0 f\left(-\sqrt{\frac{3}{5}}\right) + c_1 f(0) + c_2 f\left(\sqrt{\frac{3}{5}}\right).$$

Các trọng số tương ứng có thể được xác định thông qua việc tính tích phân các đa thức nội suy Lagrange cơ bản tương ứng. Tuy nhiên, ta sẽ trình bày một phương pháp sơ cấp hơn, sử dụng tính chất công thức cầu phương Gauss n điểm sẽ trở thành công thức đúng khi áp dụng cho các hàm đa thức bậc $\leq 2n$ (Định lý 5.5). Nói riêng, với $f(x) = 1$, ta có

$$2 = \int_{-1}^1 dx = c_0 + c_1 + c_2.$$

Với $f(x) = x$, ta có

$$0 = \int_{-1}^1 x dx = c_0 \cdot \sqrt{\frac{3}{5}} + c_1 \cdot 0 + c_2 \cdot \left(-\sqrt{\frac{3}{5}}\right).$$

Với $f(x) = x^2$, ta có

$$\frac{2}{3} = \int_{-1}^1 x^2 dx = c_0 \cdot \frac{3}{5} + c_1 \cdot 0 + c_2 \cdot \frac{3}{5}.$$

Từ hai phương trình sau, ta giải được

$$c_0 = c_2 = \frac{5}{9}.$$

Từ đây, thay vào phương trình đầu tiên, ta thu được

$$c_1 = \frac{8}{9}.$$

Như vậy, ta có công thức cầu phương Gauss 3 điểm:

$$\int_{-1}^1 f(x) dx \approx \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right). \quad (5.13)$$

Ví dụ 5.6. Để lấy ví dụ, chúng ta trở lại vấn đề xấp xỉ $\ln(2)$. Lấy hàm dưới dấu tích phân là $f(x) = 1/(3+x)$. Khi đó, ta có

$$\begin{aligned} \ln(2) &= \int_{-1}^1 \frac{dx}{3+x} \\ &\approx \frac{5}{9} \left(\frac{1}{3-\sqrt{3/5}} + \frac{1}{3+\sqrt{3/5}} \right) + \frac{8}{9} \left(\frac{1}{3+0} \right) \\ &= \frac{25}{63} + \frac{8}{27} = \frac{131}{189} \\ &\approx 0,6931217 \dots \end{aligned}$$

Kết quả này chính xác tới 4 chữ số sau dấu thập phân (0,6931).

Chú ý 5.1. Một ưu điểm của phương pháp cầu phương Gauss là nó không dùng đến giá trị của hàm số lấy tích phân tại điểm đầu mút của đoạn lấy tích phân. Điều này có lợi khi tính các tích phân suy rộng hội tụ, nhưng hàm dưới dấu tích phân tiến ra vô hạn tại một hoặc hai đầu mút. Chúng ta hãy xét ví dụ sau:

Ví dụ 5.7. Tính gần đúng tích phân suy rộng sau bằng phương pháp cầu phương Gauss⁴

$$\int_0^1 \frac{dx}{\sqrt{\sin(x)}}.$$

Đổi biến $t = 2x - 1$ hay $x = \frac{1}{2}(t + 1)$, $dx = \frac{1}{2} dt$, ta có

$$I = \int_0^1 \frac{dx}{\sqrt{\sin(x)}} = \frac{1}{2} \int_{-1}^1 \frac{dt}{\sqrt{\sin(\frac{1}{2}(t + 1))}}.$$

Công thức cầu phương Gauss 3 điểm:

$$I \approx \frac{5}{9} \sin^{-\frac{1}{2}}\left(\frac{1 - \sqrt{0,6}}{2}\right) + \frac{8}{9} \sin^{-\frac{1}{2}}\left(\frac{1}{2}\right) + \frac{5}{9} \sin^{-\frac{1}{2}}\left(\frac{1 + \sqrt{0,6}}{2}\right).$$

Tính toán cụ thể, ta có

$$I \approx \frac{5}{9} \sin^{-\frac{1}{2}}(0,1127) + \frac{8}{9} \sin^{-\frac{1}{2}}(0,5) + \frac{5}{9} \sin^{-\frac{1}{2}}(0,8873) \approx 1,8.$$

Nhận xét 5.2. Trong Ví dụ 5.7, chúng ta tính tích phân suy rộng (loại 1), do hàm số dưới dấu tích phân, hàm $f(x) = 1/\sqrt{\sin(x)}$ tiến ra vô hạn khi $x \rightarrow 0$. Điều này cho thấy một ưu điểm của công thức cầu phương Gauss so với các công thức hình thang và công thức Simpson khi chúng ta không phải tính giá trị của hàm số tại hai đầu mút của đoạn lấy tích phân.

Mặt khác, kết quả $I \approx 1,8$ có độ chính xác thấp. Đó là vì chúng ta đã sử dụng cầu phương Gauss với chỉ 3 điểm. Để tăng độ chính xác, có thể xem xét áp dụng công thức cầu phương Gauss với nhiều điểm hơn.

CAS 5.4. Trong phương pháp cầu phương Gauss, chúng ta xấp xỉ tích phân dưới dạng

$$\int_{-1}^1 f(x) dx \approx \sum_{j=0}^n c_j f(x_j)$$

⁴Tính hội tụ của tích phân suy rộng này dễ dàng được suy ra từ đánh giá bằng nguyên lí so sánh, dựa trên giới hạn $\sin(x)/x \rightarrow 1$ khi $x \rightarrow 0$ và tính hội tụ của tích phân suy rộng $\int_0^1 x^{-\frac{1}{2}} dx = 2$.

trong đó, x_j là các điểm chia, được chọn là nghiệm của đa thức Legendre thứ $n + 1$, còn c_j là các hệ số tương ứng. Ví dụ, với $n + 1 = 4$ chúng ta có các điểm chia và hệ số cho bởi bảng sau:

```
library(gaussquad)
legendre.quadrature.rules(4)[4]

## [[1]]
##           x           w
## 1  0,861136312 0,347854845
## 2  0,339981044 0,652145155
## 3 -0,339981044 0,652145155
## 4 -0,861136312 0,347854845
```

Trong kết quả thu được, cột x cho các điểm chia và cột w cho các hệ số tương ứng.

Ví dụ sau đây minh họa công thức cầu phương Gauss bốn điểm để tính tích phân hàm số $f(x) = 1/(3 + x)$ trên đoạn $[-1, 1]$ với kết quả là $\ln(2)$:

$$\int_{-1}^1 \frac{dx}{3+x} = \ln(2).$$

```
library(gaussquad)
Lq <- legendre.quadrature.rules(4)[[4]]
nodes <- Lq$x
coeffs <- Lq$w
f <- function(x) 1 / (3 + x)
result <- sum(coeffs * f(nodes))
print(result)

## [1] 0,693146417
```

Bạn đọc hãy so sánh kết quả này với giá trị $\ln(2)$ tính bằng các phương pháp khác.

5.1.5 Tính tích phân bằng hàm có sẵn trong R

Với các phương pháp số đã được cài đặt trong máy tính, tích phân Gauss trên có thể tính được gần đúng một cách nhanh chóng. Trong tiểu mục này, chúng ta sẽ tìm hiểu hai ví dụ minh họa sau.

CAS 5.5. Trong R, chức năng `integrate()` của `pracma` tính gần đúng tích phân của một hàm số một biến. Ví dụ, đoạn mã sau cho kết quả gần đúng

0,7468241 của tích phân

$$\int_0^1 e^{-x^2} dx.$$

```
library(pracma)

# Khai báo hàm f(x) = e^(-x^2)
f <- function(x) {
  exp(-x * x)
}

# Tính tích phân của f(x) theo x trên [0, 1]
integrate(f, 0, 1)

## 0,746824133 with absolute error < 8,3e-15
```

CAS 5.6. Tích phân bội được tính bằng `integral2()` của `pracma` trong R. Ví dụ sau tính

$$I = \iint_D \sin(y^2) dA,$$

trong đó $D = \{(x, y) : 0 \leq x \leq 1, x \leq y \leq 1\}$. Viết lại tích phân kép dưới dạng tích phân lặp:

$$I = \int_0^1 dx \int_x^1 \sin(y^2) dy.$$

Từ đó, áp dụng `integral2()` với các cận của x là 0 và 1 còn cận của y gồm hàm số $p(x) = x$ và $q(x) = 1$.

```
library(pracma)

# Hàm số dưới dấu tích phân
f <- function(x, y) sin(y^2)

# Hai cận của x: a = 0, b = 1
a <- 0; b <- 1

# Hai cận của y: p(x) = x, q(x) = 1
p <- function(x) x; q <- 1

# Tích phân kép
integral2(f, a, b, p, q)

## $Q
```

```
## [1] 0,229848847
##
## $error
## [1] 2,67485514e-10
```

Kết quả thu được so với giá trị chính xác $(1 - \cos(1))/2$ có sai số rất nhỏ.

5.2 Tính gần đúng đạo hàm

Nội dung của mục này là về tính gần đúng đạo hàm của hàm số bằng các phương pháp khai triển Taylor, công thức đạo hàm 3 điểm, 5 điểm, tính gần đúng đạo hàm bằng đa thức nội suy và hàm ghép trơn. Tài liệu tham khảo: T. V. Đình [6, §5.1], P. K. Anh [2, §V.1], Kreyszig [9, §19.5].

5.2.1 Xây dựng công thức tính đạo hàm bằng khai triển Taylor

Nhắc lại rằng một hàm số $f: (a, b) \rightarrow \mathbb{R}$ là khả vi tại $x_0 \in (a, b)$ nếu giới hạn

$$\ell = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

tồn tại và hữu hạn. Khi đó, đạo hàm của f tại x_0 được định nghĩa là

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Từ đó, chúng ta thấy một phương pháp khá đơn giản để tính gần đúng đạo hàm $f'(x_0)$ bằng cách tính tỉ số giữa số gia hàm số và số gia đối số h , khi h nhỏ.

CAS 5.7. Ví dụ, chúng ta hãy thực nghiệm tính đạo hàm của hàm số

$$f(x) = x^3 \cos(x^2) \log(\sqrt{x}), \quad x > 0,$$

tại $x = 1$ bằng cách tính giá trị của $f(1)$ và $f(1 + h)$ với $h = 10^{-5}$. Bằng R, chúng ta tính như sau:

```
# Tùy chọn số chữ số = 9
options(digits = 9)

# Định nghĩa hàm số f(x) cần tính đạo hàm
f <- function(x) {
  x^3 * cos(x^2) * log(sqrt(x))
}
```

```
# Tính đạo hàm tại x = a với h = 1e-5
a <- 1; h <- 1e-5

# Công thức xấp xỉ đạo hàm 2 điểm
(f(a + h) - f(a)) / h

## [1] 0,270149492
```

Để so sánh, lưu ý rằng bằng những quy tắc tính đạo hàm quen thuộc trong giải tích toán học, chúng ta có

$$f'(1) = \frac{\cos(1)}{2} \approx 0,270151.$$

So sánh hai kết quả ở dòng 4 và 6 chúng ta thấy sự sai khác bé hơn 2×10^{-6} .

Chú ý 5.2. Khi $h > 0$, chúng ta gọi công thức ở dòng 3 là “đạo hàm tiến” (forward derivative). Ngược lại, khi $h < 0$, chúng ta gọi là đạo hàm lùi (backward derivative).

Sai số

Bây giờ, để ước tính sai số của phương pháp tính gần đúng trên, chúng ta sẽ cần một số giả thiết về đạo hàm cấp hai của f . Điều này khá tự nhiên, vì đạo hàm cấp 2 cho biết tốc độ biến đổi tức thời của đạo hàm cấp một theo đối số.

Giả sử $f(x)$ khả vi tới cấp hai. Khi đó, khai triển Taylor của nó có dạng

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}h^2 f''(c),$$

với c nằm giữa x_0 và $x_0 + h$. Do đó, nếu đạo hàm cấp hai của f bị chặn bởi một hằng số nào đó, $|f''(x)| \leq M$, thì có đánh giá sai số:

$$\left| f'(x_0) - \frac{f(x_0 + h) - f(x_0)}{h} \right| = \left| \frac{1}{2}h f''(c) \right| \leq \frac{Mh}{2}.$$

Nói cách khác,

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0)}{h},$$

với sai số tuyệt đối không vượt quá $Mh/2$. Với ước lượng sai số này, chúng ta cần tính toán với $|h|$ khá nhỏ. Nhưng khi ấy sai số tính toán trở nên lớn hơn khi trong phép tính chúng ta chia cho h . Do đó, chúng ta muốn các công thức gần đúng với đánh giá sai số dạng Mh^α , với $\alpha > 1$.

Công thức 3 điểm, công thức 5 điểm

Một ý tưởng khá đơn giản để thiết lập một công thức với sai số có bậc cao hơn h^2 khi $h \rightarrow 0$ như sau: Giả sử hàm số f khả vi tới cấp ba. Khai triển Taylor tại x_0 cho chúng ta

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f'''(c)h^3, \quad (5.14)$$

với c là một điểm phụ thuộc h nào đó trong khoảng đang xét. Tiếp theo, thay h bằng $-h$, chúng ta thu được:

$$f(x_0 - h) = f(x_0) + f'(x_0)(-h) + \frac{1}{2}f''(x_0)(-h)^2 + \frac{1}{6}f'''(\tilde{c})(-h)^3, \quad (5.15)$$

với $\tilde{c}$ phụ thuộc $-h$. Tiếp theo, trừ hai đẳng thức (5.14) và (5.15) trên theo vế, chúng ta thu được

$$f(x_0 + h) - f(x_0 - h) = 2f'(x_0)h + \frac{1}{6}(f'''(c) + f'''(\tilde{c}))h^3.$$

Từ đó, chúng ta có

$$\frac{f(x_0 + h) - f(x_0 - h)}{2h} - f'(x_0) = \frac{1}{12}(f'''(c) + f'''(\tilde{c}))h^2.$$

Như vậy, nếu đạo hàm cấp ba của f bị chặn: $|f'''(x)| \leq M$ với mọi $x \in (x_0 - h, x_0 + h)$ thì có phép xấp xỉ

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0 - h)}{2h}, \quad (5.16)$$

với sai số tuyệt đối không vượt quá $Mh^2/6$. Công thức (5.16) gọi là công thức xấp xỉ đạo hàm 3 điểm.

CAS 5.8. Chúng ta sẽ tiến hành thực nghiệm trong R với bài toán tính gần đúng đạo hàm của hàm số dựa theo công thức (5.16). Tương tự như CAS 5.7, mã lệnh chi tiết như sau:

```
f <- function(x) x^3 * cos(x^2) * log(sqrt(x))
a <- 1; h <- 1e-5
(f(a + h) - f(a - h)) / (2 * h)

## [1] 0,270151153
```

Rõ ràng, cùng với $h = 10^{-5}$, kết quả thu được ở đây tốt hơn kết quả thu được dựa trên công thức đạo hàm tiến.

Ở trên, chúng ta đã áp dụng công thức khai triển Taylor cho các giá trị khác nhau của h nhằm khử các số hạng có bậc thấp của h trong công thức đánh giá sai số. Ý tưởng này có thể mở rộng để có được các công thức xấp xỉ đạo hàm khác như công thức 5 điểm. Ví dụ, công thức Taylor có thể viết

$$f(x_0 + 2h) = f(x_0) + f'(x_0)(2h) + \frac{1}{2}f''(x_0)(2h)^2 + \frac{1}{6}f'''(\hat{c})(2h)^3. \quad (5.17)$$

Kết hợp (5.17) với (5.14), chúng ta có thể khử số hạng của h^2 . Cụ thể như sau: Nhân hai vế của (5.14) với -4 rồi cộng vào (5.17), chúng ta thu được

$$\begin{aligned} -4f(x_0 + h) + f(x_0 + 2h) &= -3f(x_0) - 2f'(x_0)h \\ &\quad + \left[-\frac{2}{3}f'''(c) + \frac{4}{3}f'''(\hat{c}) \right] h^3. \end{aligned} \quad (5.18)$$

Từ đây, chúng ta thu được

$$f'(x_0) \approx \frac{-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h)}{2h}, \quad (5.19)$$

với sai số tuyệt đối không vượt quá Mh^2 nếu $|f'''(x)| \leq M$ với mọi $x \in (x_0, x_0 + 2h)$.

CAS 5.9. Thực nghiệm trên máy tính với $h = 10^{-4}$, chúng ta thu được kết quả sau:

```
f <- function(x) x^3 * cos(x^2) * log(sqrt(x))
a <- 1; h <- 1e-4
(-3 * f(a) + 4 * f(a + h) - f(a + 2 * h)) / (2 * h)

## [1] 0,270151204
```

Bằng các tính toán tương tự, chúng ta cũng có thể thu được công thức nhiều điểm hơn như công thức 5 điểm

$$f'(x_0) \approx \frac{f(x_0 - 2h) - 8f(x_0 - h) + 8f(x_0 + h) - f(x_0 + 2h)}{12h}, \quad (5.20)$$

với sai số không vượt quá một hằng số nhân với chặn trên của đạo hàm cấp 5 của f . Chi tiết dành cho bạn đọc.

CAS 5.10. Thực nghiệm công thức 5 điểm trên phần mềm thống kê R như sau:

```
f <- function(x) x^3 * cos(x^2) * log(sqrt(x))
a <- 1; h <- 1e-2
(f(a - 2 * h) - 8 * f(a - h) + 8 * f(a + h) - f(a + 2 * h)) / (12 * h)

## [1] 0,270151195
```

Đối với công thức 5 điểm, chúng ta không cần chọn h quá bé: Với $h = 0,01$, kết quả thu được chính xác tới 7 chữ số đáng tin.

5.2.2 Tính gần đúng đạo hàm bằng đa thức nội suy

Chúng ta vẫn xét bài toán tính gần đúng $f'(x)$ với $f: [a, b] \rightarrow \mathbb{R}$ là một hàm số giá trị thực trong tình huống giá trị của nó chỉ được biết tại một số điểm $x_0, x_1, \dots, x_n \in [a, b]$. Một ý tưởng tự nhiên là chúng ta sẽ xấp xỉ đạo hàm $f'(x)$ bởi $P'_n(x)$, trong đó $P_n(x)$ là đa thức nội suy của f trên các điểm $x_0, x_1, \dots, x_n$.

Nhắc lại rằng đa thức nội suy của $f(x)$ tại các mốc $x_0, x_1, \dots, x_n$ là

$$P_n(x) = \sum_{k=0}^n f(x_k) \ell_k(x),$$

trong đó, $\ell_k(x)$ là các đa thức Lagrange cơ bản tương ứng với các mốc nội suy đã cho. Từ đó, chúng ta có

$$f' \approx P'_n(x) = \sum_{k=0}^n f(x_k) \ell'_k(x).$$

Rõ ràng, $\ell'_k(x)$ hoàn toàn được xác định khi biết các điểm nội suy x_k .

Vấn đề nảy sinh bây giờ là đánh giá sai số

$$\epsilon(x) = f'(x) - P'_n(x). \quad (5.21)$$

Định lý 5.6:

Gọi $P_n(x)$ là đa thức nội suy của hàm $f(x)$ tại $n+1$ điểm nội suy $\{x_0, x_1, \dots, x_n\} \subset (a, b)$. Giả sử $f(x)$ có đạo hàm đến cấp $n+1$ bị chặn: $|f^{(n+1)}(x)| \leq M$ với mọi $x \in (a, b)$. Khi đó, $f'(x_j) \approx P'_n(x_j)$, với sai số tuyệt đối thỏa mãn

$$|\epsilon(x_j)| \leq \frac{M}{(n+1)!} \prod_{k \neq j} |x_j - x_k|. \quad (5.22)$$

Trong định lý trên, có một điều thú vị là mặc dù ta đánh giá sai số (5.21) trong tính toán đạo hàm cấp một của hàm số nhưng lại đặt giả thiết lên đạo hàm cấp $n+1$ của nó.

Ví dụ 5.8. Xét hàm số $f(x) = \sqrt[3]{x^8}$. Dễ dàng thấy rằng f khả vi liên tục đến cấp hai:

$$f'(x) = \frac{8}{3} \sqrt[3]{|x|^5} \operatorname{sign}(x), \quad f''(x) = \frac{40}{9} \sqrt[3]{x^2}.$$

Trong khi đạo hàm cấp 3 của nó là

$$f^{(3)}(x) = \frac{80}{81 \sqrt[3]{|x|}}$$

với $x \neq 0$, trong khi đó đạo hàm cấp ba không tồn tại tại $x = 0$.

Xét $h > 0$ và các điểm $x_0 = -h, x_1 = 0, x_2 = h$. Khi đó, đa thức nội suy của $f(x)$ trên tập nội suy này là

$$P_h(x) = x^2 \sqrt[3]{h^2}.$$

Do đó, ta có

$$P'_h(0) = f'(0) = 0.$$

Tuy nhiên, do f không khả vi đến cấp 3, ta không thể áp dụng để đánh giá sai số trong phép xấp xỉ $f'(0) \approx P'_h(0)$.

Trong mệnh đề sau đây, chúng ta sẽ ước lượng sai số (5.21) chỉ với điều kiện về đạo hàm cấp hai của f , cùng với một điều kiện bị chặn của một hàm phụ khác. Từ kết quả này có thể dễ dàng suy ra được định lý 5.6.

Mệnh đề 5.1:

Cho $f(x)$ là một hàm khả vi hai lần với đạo hàm cấp hai $f''(x)$ bị chặn trên $[a, b]$ bởi $-K$ và K :

$$|f''(x)| \leq K, \quad a \leq x \leq b.$$

Gọi $P_n(x)$ là đa thức nội suy của $f(x)$ tại các điểm nội suy phân biệt $\{x_0, x_1, \dots, x_n\} \subset (a, b)$, $r_n(x) = f(x) - P_n(x)$, và $s(x), x \in [a, b] \setminus \{x_0, x_1, \dots, x_n\}$, là hàm số xác định bởi

$$r_n(x) = s(x) \prod_{k=0}^n (x - x_k). \quad (5.23)$$

Giả sử $|s(x)| \leq L$ với mọi $x \in [a, b] \setminus \{x_0, x_1, \dots, x_n\}$. Khi đó,

$$|f'(x_j) - P'_n(x_j)| \leq L \left| \prod_{k \neq j} (x_j - x_k) \right|. \quad (5.24)$$

Chứng minh. Trước hết, nhận xét rằng vì $f(x)$ khả vi tới cấp hai còn $P_n(x)$ là một đa thức nên $r_n(x)$ khả vi tới cấp hai. Hơn nữa, (5.29) có thể được viết lại là

$$r_n(x) = \begin{cases} s(x) \prod_{k=0}^n (x - x_k) & x \notin \{x_0, \dots, x_n\} \\ 0 & x \in \{x_0, \dots, x_n\} \end{cases}. \quad (5.25)$$

Ấn định $j \in \{0, 1, 2, \dots, n\}$ tùy ý và khai triển Taylor hàm $r_n(x)$ trong lân cận

của x_j , chúng ta có

$$r_n(x) = r_n(x_j) + r'_n(x_j)(x - x_j) + \frac{r''_n(c)}{2!}(x - x_j)^2, \quad (5.26)$$

với c là một điểm phụ thuộc x và nằm giữa x_j và x . Lưu ý rằng $r_n(x_j) = 0$ nên, cùng các đẳng thức (5.25) và (5.26), chúng ta suy ra

$$r'_n(x_j)(x - x_j) + \frac{r''_n(c)}{2!}(x - x_j)^2 = s(x) \prod_{k=0}^n (x - x_k), \quad x \notin \{x_0, x_1, \dots, x_n\}.$$

Chia hai vế cho $x - x_j \neq 0$ và chuyển một số hạng sang vế phải, chúng ta thu được

$$r'_n(x_j) = -\frac{r''_n(c)}{2!}(x - x_j) + s(x) \prod_{k \neq j} (x - x_k). \quad (5.27)$$

Theo giả thiết, $f''(x)$ bị chặn trên $[a, b]$ nên $r''_n(x) = f''(x) - P''_n(x)$ cũng bị chặn trên $[a, b]$: Với mọi $x \in [a, b]$,

$$|r''_n(x)| \leq \tilde{K}.$$

Từ đó, cùng với bất đẳng thức tam giác, chúng ta thu được

$$|r'_n(x_j)| \leq \frac{\tilde{K}}{2}|x - x_j| + L \prod_{k \neq j} |x - x_k|, \quad x \notin \{x_0, x_1, \dots, x_n\}.$$

Cho $x \rightarrow x_j$ và lấy giới hạn hai vế, chúng ta thu được

$$|r'_n(x_j)| \leq L \prod_{k \neq j} |x_j - x_k|. \quad (5.28)$$

Chúng ta kết thúc chứng minh. $\square$

Chứng minh Định lí 5.6. Nếu f có đạo hàm đến cấp $n + 1$ thì chúng ta có công thức sai số dạng

$$f(x) = P_n(x) + s(x) \prod_{k=0}^n (x - x_k). \quad (5.29)$$

Trong đó,

$$s(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}, \quad (5.30)$$

trong đó $\xi(x)$ tồn tại và phụ thuộc vào x .

Từ giả thiết về đạo hàm cấp $n + 1$ của f , chúng ta suy ra

$$|s(x)| \leq \frac{M}{(n+1)!}$$

và từ đó suy ra kết luận của định lý bằng cách áp dụng mệnh đề 5.1 với $L = M/(n+1)!$. $\square$

Chú ý 5.3. Một số sách đưa ra những cách chứng minh ngắn hơn cho định lý 5.6, nhưng một số lập luận trong chúng không chặt chẽ. Ví dụ, trong [4], hai tác giả Burden và Faires lấy đạo hàm hai vế của (5.29) để thu được

$$f'(x_j) = P'_n(x_j) + f^{(n+1)}(\xi(x_j)) \left(\frac{1}{(n+1)!} \prod_{k=0, k \neq j}^n (x_j - x_k) \right).$$

Tuy nhiên, đây là một phương trình vô nghĩa, do $\xi(x_j)$ không xác định!

CAS 5.11. Người ta thả một vật rơi trong không khí từ một độ cao 20m và đo khoảng cách từ vật đó tới vị trí ban đầu sau mỗi khoảng thời gian 1s. Kết quả được ghi trong bảng sau:

Thời gian (s)	0	1	2	3	4
Quãng đường (m)	0	1,811	6,254	12,293	19,301

Dùng phương pháp nội suy hãy tính gần đúng vận tốc của vật tại thời điểm $t = 2s$ (có thể sử dụng máy tính).

Gọi $P_4(t)$ là đa thức nội suy của bộ dữ liệu trên. Chúng ta có thể coi $P'_4(2)$ là một giá trị xấp xỉ của vận tốc tại $t = 2$. Thực hành trên máy tính như sau:

```
# Tính gần đúng đạo hàm bằng đa thức nội suy.
# Thư viện polynom có sẵn trong R.

library(polynom)

# Các dữ liệu ban đầu
t <- c(0, 1, 2, 3, 4)
s <- c(0, 1.811, 6.254, 12.293, 19.301)

# poly.calc trong gói polynom tính đa thức nội suy

p <- poly.calc(t, s)

# deriv() tính đạo hàm của đa thức p
p_prime <- deriv(p)
```

```
# as.function() chuyển một đa thức về một hàm số đa thức

p_prime_func <- as.function(p_prime)
print(p_prime_func(2))

## [1] 5,37958333
```

Trong chương trình R phía trên, chúng ta tính đạo hàm của đa thức nội suy tại $t = 2$ và thu được $v(2) \approx 5,379583$ (m/s), là giá trị xấp xỉ của vận tốc của vật rơi tại $t = 2$ giây.

Nhận xét 5.3. Trong ví dụ trên, vật rơi do tác dụng của trọng lực và chịu lực cản của không khí. Quỹ đạo đi được của vật, tính từ vị trí ban đầu, được cho bởi một công thức dạng

$$s(t) = s_0 + \bar{v}t + \frac{\bar{v}}{k}(\exp(-kt) - 1),$$

trong đó, $\bar{v}$ là vận tốc pha cuối (terminal velocity), và $k = b/m$, với b là hệ số lực cản không khí (phụ thuộc vào kích thước của vật và không khí) và m là khối lượng của vật.

Từ bảng trên, người ta có thể xác định được các hệ số s_0 , $\bar{v}$, và k . Các giá trị của tham số cụ thể là $s_0 = 0$, $\bar{v} = 8,5$ và $k = 0,5$. Như vậy,

$$s(t) = 8,5t + 17(\exp(-0,5t) - 1).$$

Do đó, vận tốc của nó cho bởi

$$v(t) = 8,5(1 - \exp(-0,5t)).$$

Từ công thức cuối cùng, chúng ta tính được $v(2) \approx 5,373025$. Bạn đọc hãy so sánh hai kết quả và rút ra nhận xét.

5.3 Câu hỏi và bài tập

- Một vật bắt đầu chuyển động thẳng với vận tốc thay đổi theo thời gian được tính bởi công thức $v(t) = 1/(5 + \sqrt{t})$ (m/s).
 - Sử dụng quy tắc hình thang với số khoảng chia bằng 4, tính gần đúng quãng đường vật đi được trong khoảng thời gian từ $t = 1$ đến $t = 5$ giây.
 - Sử dụng quy tắc Simpson với số khoảng chia bằng 4, tính gần đúng đến 3 chữ số sau dấu thập phân quãng đường vật đi được trong khoảng thời gian từ $t = 1$ đến $t = 5$ giây [Trả lời: 0,599 (m).]
- Tính gần đúng các tích phân sau đây bằng quy tắc hình thang và quy tắc Simpson với và đánh giá sai số của các kết quả gần đúng thu được.

(a) $\int_1^2 x^2 dx,$

(c) $\int_1^5 \frac{dx}{x},$

(b) $\int_1^2 x^{-2} dx,$

(d) $\int_1^{\frac{\pi}{2}} \sqrt{4 - \sin^2(x)} dx.$

3. Tính gần đúng tích phân

$$\int_0^{\frac{\pi}{2}} \frac{\cos(x)}{1+x} dx$$

với 4 chữ số đáng tin.

Chương

6

Giải gần đúng phương trình vi phân

Chương này trình bày một số phương pháp giải bài toán Cauchy đối với phương trình và hệ phương trình vi phân thường, bao gồm phương pháp Euler, phương pháp Euler ngược, phương pháp Runge–Kutta, và một số phương pháp khác.

Nội dung

6.1	Bài toán giá trị ban đầu	186
6.2	Phương pháp Euler	187
6.2.1	Mô tả phương pháp Euler	188
6.2.2	Sự hội tụ và sai số của phương pháp Euler	189
6.2.3	Phương pháp Euler ngược	194
6.3	Phương pháp Runge–Kutta và các phương pháp khác	198
6.3.1	Một số khái niệm	198
6.3.2	Khai triển tiệm cận của sai số	198
6.3.3	Phương pháp hình thang	200
6.3.4	Phương pháp Heun	201
6.3.5	Phương pháp Runge–Kutta	202
6.3.6	Sử dụng deSolve của R giải gần đúng phương trình vi phân	204
6.4	Phương trình vi phân bậc cao và hệ phương trình	206
6.4.1	Hệ phương trình vi phân	206
6.4.2	Phương trình vi phân cấp cao	208
6.5	Sơ lược về bài toán biên	209
6.6	Câu hỏi và bài tập	212

6.1 Bài toán giá trị ban đầu

Trong mục này, chúng ta bắt đầu với bài toán cơ bản của lý thuyết phương trình vi phân, đó là bài toán giá trị ban đầu, hay còn gọi là bài toán Cauchy. Trong các mục tiếp theo, chúng ta tìm hiểu một số phương pháp tính gần đúng nghiệm của phương trình vi phân của bài toán giá trị ban đầu.

Bài toán tìm hàm $y = y(x)$ trên đoạn $[a, b]$ sao cho

$$\begin{cases} y' = f(x, y), & x \in [a, b], \\ y(a) = \alpha, \end{cases} \quad (6.1)$$

với α là một giá trị cho trước, gọi là *bài toán giá trị ban đầu* (initial value problem–IVP) của phương trình vi phân $y' = f(x, y)$. Điều kiện $y(a) = \alpha$ thường được gọi là điều kiện giá trị ban đầu (hay điều kiện Cauchy).

Bài toán giá trị ban đầu có thể có nghiệm duy nhất, vô nghiệm, hoặc có nhiều nghiệm. Hơn nữa, nghiệm của bài toán giá trị ban đầu có thể chỉ tồn tại trong một khoảng nào đó của điểm ban đầu.

Ví dụ 6.1. Xét bài toán giá trị ban đầu

$$\begin{cases} y' = \sqrt{y}, \\ y(0) = 0. \end{cases}$$

Dùng phương pháp tách biến đã biết trong môn giải tích, chúng ta đưa phương trình về dạng

$$\frac{dy}{\sqrt{y}} = dx.$$

Lấy tích phân hai vế,

$$\int \frac{dy}{\sqrt{y}} = \int dx = x + C.$$

Từ đó, ta suy ra

$$y(x) = \frac{(x + C)^2}{4}.$$

Với $C = 0$, chúng ta thu được nghiệm $y = x^2/4$ thỏa mãn bài toán Cauchy. Mặt khác, $y = 0$ cũng thỏa mãn bài toán đó. Như vậy, bài toán giá trị ban đầu trên có *nhiều hơn một* nghiệm.

Ví dụ 6.2. Xét bài toán giá trị ban đầu

$$\begin{cases} y' = y^2, \\ y(0) = 1. \end{cases}$$

Dùng phương pháp tách biến, chúng ta có thể giải được

$$y(x) = \frac{-1}{x + C}.$$

Thay $x = 0, y = 1$, ta thu được phương trình của C :

$$1 = \frac{-1}{C},$$

và do đó, $C = -1$, nên

$$y = \frac{1}{1 - x}.$$

Như vậy, nghiệm của bài toán không xác định tại điểm $x = 1$.

Trong các phương pháp số để giải các bài toán giá trị ban đầu, chúng ta yêu cầu các bài toán đó là *có nghiệm* và nghiệm là *duy nhất*. Để có được hai điều này, về lí thuyết, chúng ta có thể dựa vào định lí sau:

Định lí 6.1:

Giả sử $f(x, y)$ là một hàm hai biến xác định và liên tục trên hình chữ nhật $R = [x_0 - a, x_0 + a] \times [\alpha - \beta, \alpha + \beta]$, $a > 0, \beta > 0$, sao cho $f(x, y)$ và đạo hàm riêng $f_y(x, y)$ là bị chặn trên R : Với các hằng số $K > 0, M > 0$ nào đó,

$$|f(x, y)| \leq K, \quad |f_y(x, y)| \leq M, \quad (x, y) \in R. \quad (6.2)$$

Khi đó, bài toán giá trị ban đầu

$$y' = f(x, y), \quad y(x_0) = \alpha \quad (6.3)$$

có nghiệm duy nhất $y(x)$ trên một khoảng $x_0 - \epsilon < x < x_0 + \epsilon$, với $\epsilon > 0$ nào đó.

6.2 Phương pháp Euler

Chúng ta bắt đầu với phương pháp cơ bản nhất để giải gần đúng phương trình vi phân, đó là *phương pháp Euler*¹. Mặc dù phương pháp này không phải là phương pháp cho kết quả chính xác nhất, nhưng nó lại rất đơn giản

¹Leonhard Euler (1707 – 1783) là một nhà toán học người Thụy Sĩ.

và có thể dễ dàng mở rộng cho các tình huống phức tạp hơn, như việc giải các phương trình vi phân bậc cao, các hệ phương trình lớn, hay phương trình vi phân ngẫu nhiên, v.v. Tài liệu tham khảo: T. V. Đĩnh [6, §6.1].

6.2.1 Mô tả phương pháp Euler

Trong *phương pháp Euler* (còn gọi là phương pháp Euler–Cauchy) giải gần đúng phương trình (6.1), chúng ta bắt đầu bằng việc chia đoạn $[a, b]$ thành n đoạn nhỏ bởi các điểm chia

$$a = x_0 < x_1 < \dots < x_n = b$$

và tính gần đúng các giá trị của nghiệm duy nhất $y(x)$ tại các điểm này, tức là chúng ta sẽ xấp các giá trị $y(x_i)$, $i = 0, 1, 2, \dots, n$. Như vậy, chúng ta sẽ xây dựng dãy $u_0, u_1, \dots, u_i, \dots$ sao cho u_i xấp xỉ $y(x_i)$:

$$u_i \approx y(x_i).$$

Rõ ràng, với $i = 0$, chúng ta có $y(x_0) = a$, theo điều kiện ban đầu nên cần chọn

$$u_0 = a.$$

Giả sử ta xây dựng được u_i xấp xỉ giá trị (đúng) $y(x_i)$, khai triển Taylor của y tại x_i có dạng

$$y(x) = y(x_i) + y'(x_i)(x - x_i) + \frac{y''(c_i)}{2}(x - x_i)^2$$

với c_i là một điểm nằm giữa x và x_i . Vì y là nghiệm của phương trình vi phân (6.1) nên

$$y'(x_i) = f(x_i, y(x_i)).$$

Thay giá trị này vào khai triển Taylor ở trên, chúng ta thu được

$$y(x) = y(x_i) + f(x_i, y(x_i))(x - x_i) + \frac{y''(c_i)}{2}(x - x_i)^2. \quad (6.4)$$

Hai số hạng đầu tiên ở vế phải hoàn đã biết, nếu chúng ta đã xác định được $y(x_i)$. Số hạng cuối cùng rất bé khi $x \approx x_i$, có thể bỏ qua trong phép tính gần đúng.

Đặc biệt, nếu các điểm chia x_j , $j = 0, 2, \dots, n$ là cách đều,

$$x_j = x_0 + jh,$$

thì với $x = x_{i+1}$ chúng ta có

$$y(x_{i+1}) = y(x_i) + f(x_i, y(x_i))h + \frac{y''(c_i)}{2}h^2. \quad (6.5)$$

với c_i phụ thuộc h . Từ đây, ta xác định

$$u_{i+1} := u_i + f(x_i, u_i)h. \quad (6.6)$$

Công thức này cho phép chúng ta xác định theo quy nạp dãy u_i khi biết $u_0 = \alpha$. Phương pháp tính u_i theo công thức truy hồi (6.6) gọi là *phương pháp Euler*.

6.2.2 Sự hội tụ và sai số của phương pháp Euler

Sai số của phương pháp Euler là sai số thực sự của các xấp xỉ $y(x_i)$ bởi u_i . Chúng ta gọi các sai số này là ϵ_i :

$$\epsilon_i := y(x_i) - u_i. \quad (6.7)$$

Định lí dưới đây cho một ước lượng của sai số trong phương pháp này.

Định lí 6.2:

Giả sử hàm $f(x, y)$ trong phương trình (6.1) thỏa mãn

$$\left| \frac{\partial f}{\partial y} \right| \leq L, \quad (6.8)$$

và nghiệm của nó thỏa mãn

$$|y''| \leq K, \quad (6.9)$$

trong đó K và L là các hằng số dương. Khi đó

$$|\epsilon_i| \leq \frac{K}{2L} h (e^{L(x_i - x_0)} - 1) + |\epsilon_0| e^{L(x_i - x_0)}. \quad (6.10)$$

Như vậy, nếu sai số ban đầu bằng 0, $\epsilon_0 = 0$, thì sai số tại điểm x_i hội tụ về 0 khi $h \rightarrow 0$.

Chứng minh. Từ các phương trình (6.5) và (6.6) ta thu được

$$\begin{aligned} \epsilon_{i+1} &= u_{i+1} - y(x_{i+1}) \\ &= (u_i + f(x_i, u_i)h) - \left(y(x_i) + f(x_i, y(x_i))h + \frac{y''(c_i)}{2} h^2 \right) \\ &= (u_i - y(x_i)) + h(f(x_i, u_i) - f(x_i, y(x_i))) - \frac{y''(c_i)}{2} h^2 \\ &= \epsilon_i + h(f(x_i, u_i) - f(x_i, y(x_i))) - \frac{y''(c_i)}{2} h^2. \end{aligned}$$

Từ công thức Lagrange, ta có

$$f(x_i, u_i) - f(x_i, y(x_i)) = \frac{\partial f}{\partial y}(x_i, d_i)(u_i - y(x_i)).$$

Kết hợp với điều kiện $|\partial f / \partial y| \leq L$, ta thu được

$$|f(x_i, u_i) - f(x_i, y(x_i))| \leq L|e_i|.$$

Vậy

$$|e_{i+1}| \leq |e_i|(1 + Lh) + Kh^2/2.$$

Bất đẳng thức này đúng với $i = 0, 1, 2, \dots$ với các hằng số K, L , và h không phụ thuộc i . Bằng quy nạp, ta dễ dàng suy ra từ đó bất đẳng thức

$$|e_k| \leq (1 + Lh)^k |e_0| + \frac{Kh^2}{2} \sum_{i=0}^{k-1} (1 + Lh)^i$$

Mặt khác, ta có (với $Lh > 0$)

$$\sum_{i=0}^{k-1} (1 + Lh)^i = \frac{(1 + Lh)^k - 1}{(1 + Lh) - 1} = \frac{(1 + Lh)^k}{Lh} - \frac{1}{Lh} \leq \frac{1}{Lh} (e^{kLh} - 1).$$

Ở trên, ta đã áp dụng bất đẳng thức $1 + Lh < e^{Lh}$ (Bất đẳng thức này có thể được chứng minh như sau: Khai triển Taylor hàm mũ tại 0, ta có $e^{Lh} = 1 + Lh + \frac{1}{2}e^c(Lh)^2 > 1 + Lh$, ở đây c là một giá trị nằm giữa 0 và Lh). Ta thu được

$$|e_k| \leq \frac{K}{2L} h (e^{Lkh} - 1) + e^{Lkh} |e_0|. \quad \square$$

Trong công thức đánh giá sai số trong định lí trên, cần giả thiết rằng trong mỗi phép tính trung gian, các kết quả trung gian thu được là hoàn toàn chính xác. Giả thiết này không còn phù hợp với thực tế, đặc biệt là khi chúng ta cần chọn h nhỏ, bởi vì khi đó số các phép tính tăng lên và ảnh hưởng của sai số tính toán đến kết quả cuối cùng là đáng kể.

Cụ thể hơn, nếu $\tilde{u}_j$ là xấp xỉ của u_j , với sai số tuyệt đối không vượt quá $\delta > 0$ thì ta có thể chứng minh được

$$|e_k| \leq \frac{1}{L} \left(\frac{hK}{2} + \frac{\delta}{h} \right) h (e^{Lkh} - 1) + e^{Lkh} |e_0|. \quad (6.11)$$

Do đó, bước đi h “tốt nhất” là bước đi sao cho $\frac{hK}{2} + \frac{\delta}{h}$ bé nhất. Vì K và δ là các số dương nên có đánh giá

$$\frac{hK}{2} + \frac{\delta}{h} \geq 2\sqrt{\frac{hK}{2} \cdot \frac{\delta}{h}} = \sqrt{2K\delta}.$$

Dấu bằng xảy ra khi và chỉ khi

$$\frac{hK}{2} = \frac{\delta}{h},$$

hay

$$h = \sqrt{2\delta/K}.$$

Đó là ngưỡng mà khi lấy bước đi h bé hơn nó thì sai số tính toán sẽ ảnh hưởng lớn hơn sai số lí thuyết. Vì vậy, trong thực hành, người ta thường lấy $h > \sqrt{2\delta/K}$.

Ví dụ 6.3. Sử dụng phương pháp Euler để giải gần đúng bài toán giá trị ban đầu sau đây

$$\begin{cases} y' = xe^{2x} - 2y, & 0 \leq x \leq 1, \\ y(0) = 0, \end{cases}$$

với $h = \frac{1}{2}$.

Lời giải

Phương pháp Euler với $h = \frac{1}{2}$ cho phép tính gần đúng giá trị của nghiệm tại các nút $x_0 = 0$, $x_1 = \frac{1}{2}$, và $x_2 = 1$. Gọi các giá trị gần đúng này là u_0 , u_1 , và u_2 . Khi đó, với $f(x, y) = xe^{2x} - 2y$, ta có

$$u_0 = y(0) = 0,$$

$$u_1 = u_0 + hf(x_0, u_0) = 0 + \frac{1}{2} \times (0 \times e^{2 \times 0} - 2 \times 0) = 0,$$

$$u_2 = u_1 + hf(x_1, u_1) = 0 + \frac{1}{2} \times \left(\frac{1}{2} \times e^{2 \times \frac{1}{2}} - 2 \times 0 \right) = \frac{e}{4}.$$

Nói riêng $y(1) \approx e/4$. Trong lời giải trên, chúng ta đã chọn $h = 0,5$ để minh hoạ phương pháp Euler. Khi đó, kết quả cho sai số khá lớn. Thật vậy, bài toán giá trị ban đầu trong Ví dụ 6.3 trên có thể giải theo công thức

$$y = \frac{1}{16} (e^{2x}(4x - 1) + e^{-2x}). \quad (6.12)$$

Từ đó, tính gần đúng được $y(1) \approx 1,393906$.

Để có các xấp xỉ tốt hơn, chúng ta thường phải chọn h bé hơn và số bước lặp nhiều hơn. Tuy nhiên, lưu ý rằng khi bước đi h nhỏ hơn thì quá trình tính toán phức tạp hơn và sai số tính toán trở nên lớn hơn. Tuy nhiên, với sự trợ giúp của máy tính, chúng ta có thể thực hiện phương pháp Euler với hàng triệu bước đi.

CAS 6.1. Chương trình R sau đây sẽ thực hiện các phép tính trong phương pháp Euler cho phương trình trong Ví dụ 6.3, với khoảng hai triệu bước đi và cho kết quả chính xác tới phần triệu.

```
# Phương pháp Euler giải phương trình vi phân cấp một.

euler_method <- function(f, a, b, n, alpha) {
  h <- (b - a) / n # n = số bước đi
  x <- a # Điểm bắt đầu
  y <- alpha # Giá trị ban đầu

  # Bắt đầu vòng lặp của phương pháp Euler
  for (i in 1:n) {
    y <- y + h * f(x, y)
    x <- x + h
  }
  return(y)
}

# Chạy thử euler_method()

f <- function(x, y) x * exp(2 * x) - 2 * y # Vé phải
result <- euler_method(f, 0, 1, 1e6, 0)
print(result)

## [1] 1,39390508
```

Lúc đó, với bước đi $h = 10^{-6}$, tức là số bước đi 10^6 , chúng ta cũng có thể thu được $y(1) \approx 1,39390508$. Lưu ý rằng nếu tính giá trị của nghiệm theo công thức giải tích (6.12) bằng R, chúng ta cũng thu được kết quả như vậy.

Dưới đây, chúng ta xây dựng công thức gần đúng để đánh giá sai số trong phương pháp Euler một cách thiết thực hơn. Cơ sở nó là *công thức khai triển tiệm cận của sai số*, được phát biểu trong định lý sau đây:

Định lý 6.3:

Gọi $y(x, h)$ là nghiệm gần đúng được tính bởi phương pháp Euler đối với bài toán Cauchy với bước nhảy $h > 0$. Khi đó,

$$y(x, h) = y(x) + c_1(x)h + c_2(x)h^2 + \cdots + c_p(x)h^p + O(h^{p+1}), \quad (6.13)$$

với $c_j(x)$ là các hệ số không phụ thuộc h .

Chứng minh của Định lý 6.3 nằm ngoài khuôn khổ của tài liệu này. Bạn đọc có thể tham khảo Hans Stetter [11].

Nói riêng, với $p = 1$ thì (6.13) trở thành

$$y(x_i, h) = y(x_i) + c_1(x)h + O(h^2), \quad (6.14)$$

$$y\left(x_i, \frac{h}{2}\right) = y(x_i) + c_1(x)\frac{h}{2} + O(h^2). \quad (6.15)$$

Dùng (6.15) để khử số hạng tuyến tính đối với h ở vế phải của (6.14), ta thu được

$$y(x_i, h) - 2y\left(x_i, \frac{h}{2}\right) = -y(x_i) + O(h^2). \quad (6.16)$$

Phương trình này có thể viết lại như sau

$$\epsilon\left(x_i, \frac{h}{2}\right) = y(x_i) - y\left(x_i, \frac{h}{2}\right) = y\left(x_i, \frac{h}{2}\right) - y(x_i, h) + O(h^2). \quad (6.17)$$

Trong vế phải của phương trình (6.17), số hạng $O(h^2)$ bé so với h khi h bé. Hai số hạng còn lại có thể tính được cụ thể. Từ đó, chúng ta thu được đánh giá gần đúng sai số của nghiệm của lược đồ Euler với bước đi $h/2$ như sau:

$$\epsilon\left(x_i, \frac{h}{2}\right) \approx y\left(x_i, \frac{h}{2}\right) - y(x_i, h). \quad (6.18)$$

Nhận xét 6.1. Chương trình R trong CAS 6.1 không lưu lại các giá trị trung gian mà chỉ giữ lại giá trị cuối cùng, là một xấp xỉ của $y(1)$. Chúng ta có thể dễ dàng sửa đổi chương trình trên để kết quả trả về là một dãy các giá trị của y_i tại các giá trị $x_i = x_0 + ih$.

CAS 6.2. Xét bài toán Cauchy

$$\begin{cases} y' = y + x^2, \\ y(0) = 1. \end{cases} \quad (6.19)$$

Chương trình R sau đây tính gần đúng giá trị của nghiệm trên các điểm lưới với bước đi $h = 0,1$ và biểu diễn chúng trên mặt phẳng tọa độ.

```
# Phương pháp Euler giải phương trình vi phân cấp một.
```

```
euler_method <- function(f, a, b, n, alpha) {  
  h <- (b - a) / n  
  x <- seq(a, b, by = h)  
  y <- numeric(n + 1)  
  y[1] <- alpha
```

```
# Vòng lặp theo phương pháp Euler
```

```

for (i in 1:n) {
  y[i + 1] <- y[i] + h * f(x[i], y[i])
}
return(data.frame(x = x, y = y))
}

```

Tiếp theo, chúng ta plot các giá trị tính bằng phương pháp Euler cùng với đồ thị của nghiệm giải tích. Kết quả thu được như sau là Hình 6.1, trong đó, đường đồ thị nét liền biểu diễn hàm số

$$y = -2 + 3e^x - 2x - x^2,$$

là công thức giải tích của nghiệm của bài toán Cauchy (6.19), và các chấm tròn biểu diễn các điểm (x_i, y_i) với $y_i \approx y(x_i)$ được tính từ phương pháp Euler.

6.2.3 Phương pháp Euler ngược

Phương pháp Euler ngược (backward Euler method) là một phương pháp giải số được sử dụng để giải phương trình vi phân thường (ODE), đặc biệt hữu ích khi cần đảm bảo tính ổn định của nghiệm trong các bước thời gian dài. Đây là phương pháp ẩn, có nghĩa là ở mỗi bước thời gian, chúng ta cần giải một phương trình để tìm giá trị tiếp theo.

Đối với bài toán Cauchy

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = \alpha, \quad (6.20)$$

dãy xấp xỉ của nghiệm dọc theo các điểm $x_i = x_0 + ih$, $i = 1, 2, \dots$ được xác định theo công thức

$$u_{i+1} = u_i + f(x_{i+1}, u_{i+1}), \quad (6.21)$$

với giá trị ban đầu $u_0 = \alpha$. Như đã nói, chúng ta cần giải (6.21) để thu được công thức truy hồi cho dãy u_i .

Ví dụ 6.4. Xét bài toán giá trị ban đầu

$$\frac{dy}{dx} = -2y + x, \quad y(0) = 1.$$

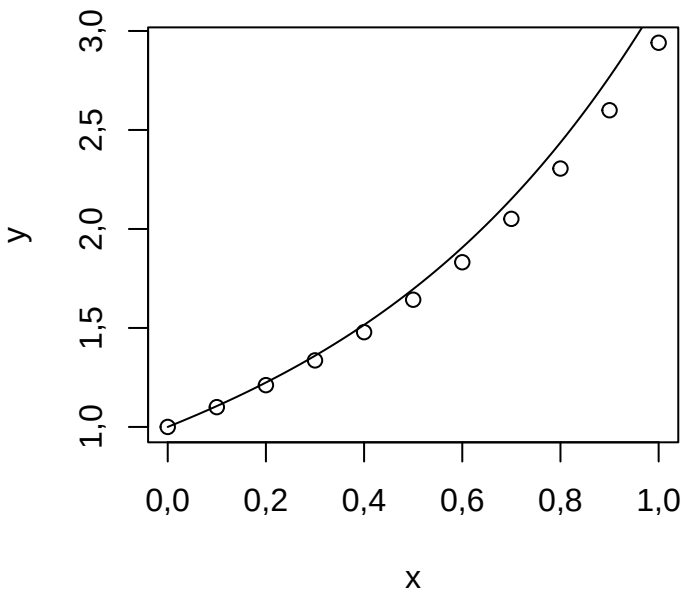
Chúng ta sẽ sử dụng phương pháp Euler ngược với bước đi $h = 0,1$ để xấp xỉ nghiệm của phương trình này.

Sử dụng công thức (6.21) trong đó, hàm $f(x, y) = -2y + x$, vì vậy

$$u_{i+1} = u_i + h \cdot (-2u_{i+1} + x_{i+1})$$

```
# Ví dụ
```

```
f <- function(x, y) y + x^2 # Vế phải của phương trình  
solution <- euler_method(f, 0, 1, 10, 1)  
plot(solution$y ~ solution$x, xlab = "x", ylab = "y")  
exact_soln <- function(x) -2 + 3 * exp(x) - 2 * x - x^2  
curve(exact_soln, from = 0, to = 1, add = TRUE)
```



Hình 6.1: Đồ thị của nghiệm giải tích và các giá trị của nghiệm tính bằng phương pháp Euler

Từ đó ta có công thức truy hồi là

$$u_{i+1} = \frac{u_i + h x_{i+1}}{1 + 2h} \quad (6.22)$$

Từ điều kiện ban đầu, ta có $x_0 = 0$, $h = 0,1$, và $u_0 = 1$. Việc tính toán u_i có thể được thực hiện một cách tuần tự.

Ví dụ, với $i = 0$, ta có

$$u_1 = \frac{u_0 + h x_1}{1 + 2h} = \frac{1 + 0,1 \times (0 + 0,1)}{1 + 2 \times 0,1} \approx 0,8417,$$

và

$$u_2 = \frac{u_1 + h x_2}{1 + 2h} = \frac{0,8417 + 0,1 \times (0 + 2 \times 0,1)}{1 + 2 \times 0,1} \approx 0,7181,$$

Các giá trị tiếp theo được tính tương tự.

CAS 6.3. Trong CAS này, chúng ta tính các giá trị tiếp theo của u_i đối với bài toán trong Ví dụ 6.4 và biểu diễn trên mặt phẳng cùng với đồ thị của nghiệm giải tích của nó

$$y(x) = -0,25 + 0,5x + 1,25e^{-2x}.$$

Chi tiết như sau:

```
# Phương pháp Euler ngược cho phương trình y' = -2y + x

backward_euler_method <- function(x0, y0, h, n) {
  x <- seq(x0, x0 + n * h, by = h)
  y <- numeric(n + 1)
  y[1] <- y0
  for (i in 1:n) {
    y[i + 1] <- (y[i] + h * x[i+1]) / (1 + 2 * h)
  }
  return(data.frame(x = x, y = y))
}

soln <- backward_euler_method(0, 1, 0.1, 10)
print(soln)

##      x      y
## 1 0,0 1,000000000
## 2 0,1 0,841666667
## 3 0,2 0,718055556
## 4 0,3 0,623379630
```

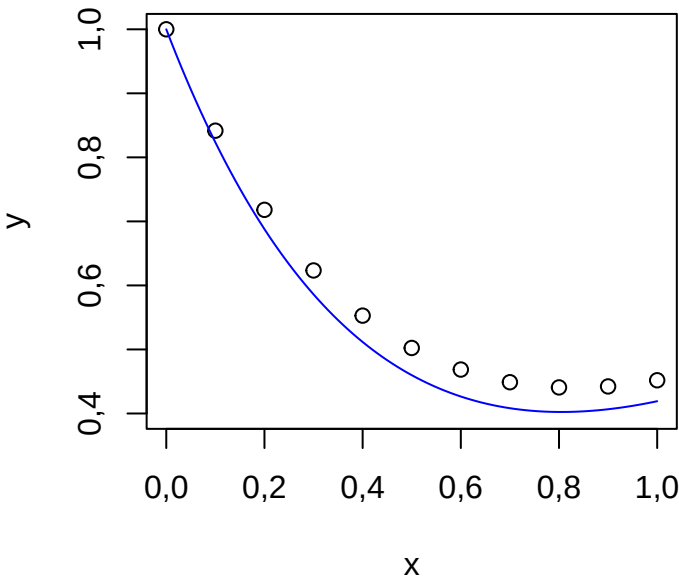
```
## 5 0,4 0,552816358
## 6 0,5 0,502346965
## 7 0,6 0,468622471
## 8 0,7 0,448852059
## 9 0,8 0,440710049
## 10 0,9 0,442258374
## 11 1,0 0,451881979
```

Biểu diễn nghiệm và nghiệm giải tích trên mặt phẳng tọa độ.

```
plot(soln$y ~ soln$x, ylim = c(0.4, 1), xlab = "x", ylab = "y")
```

Công thức giải tích của nghiệm đúng và đồ thị của nó.

```
exact_soln <- function(x) (-1 + 5 * exp(-2 * x) + 2 * x) / 4
curve(exact_soln, from = 0, to = 1, col = "blue", add = TRUE)
```



Kết quả thu được là dãy các giá trị u_i thể hiện ở kết quả của `print(soln)` trong khi đó kết quả của hàm `plot()` và `curve()` được thể hiện ở hình trên.

Chú ý 6.1. Với hàm $f(x, y) = -2y + x$ như trong Ví dụ 6.4, phương trình ẩn (6.21) giải được rất dễ dàng. Trong những tình huống tổng quát hơn, việc

giải ra u_{i+1} từ (6.21) có thể không đơn giản. Trong trường hợp ấy, người ta có thể kết hợp với phương pháp Newton–Raphson để giải nghiệm.

6.3 Phương pháp Runge–Kutta và các phương pháp khác

Trong mục này ta giới thiệu một số phương pháp cải tiến tăng độ chính xác của phương pháp Euler, bao gồm phương pháp sai số tiệm cận, phương pháp hình thang, phương pháp Heun (dự báo và điều chỉnh), và cuối cùng là phương pháp Runge–Kutta bậc 4. Tài liệu tham khảo: T. V. Đĩnh [6, §6.5].

6.3.1 Một số khái niệm

Để nói chi tiết hơn về “độ chính xác” của một phương pháp, ta cần khái niệm vô cùng bé (infinitesimal). Xét các đại lượng $\phi(h)$ phụ thuộc h và $\phi(h) \rightarrow 0$ khi $h \rightarrow 0$. Nói cách khác, $\phi(h)$ là một vô cùng bé khi $h \rightarrow 0$. Nếu tồn tại $M > 0$ và k sao cho

$$|\phi(h)| < Mh^k,$$

thì ta viết $\phi(h) = O(h^k)$, với O viết hoa. Ta cũng nói $\phi(h)$ là một vô cùng bé có cỡ $O(h^k)$.

Nếu

$$\lim_{h \rightarrow 0} \frac{\phi(h)}{h^k} = 0,$$

thì ta viết $\phi(h) = o(h^k)$ (với chữ o viết thường, và cũng nói $\phi(h)$ là vô cùng bé có bậc lớn hơn h^k). Hai cách viết quy ước này được dùng rất phổ biến trong toán học. Lưu ý, cần thận trọng để tránh nhầm lẫn các kí hiệu O lớn và o nhỏ.

Nhắc lại rằng trong việc giải gần đúng bài toán giá trị ban đầu (6.1), chúng ta kí hiệu u_i là giá trị gần đúng của nghiệm y tính tại điểm x_i và sai số của nó được kí hiệu là

$$\epsilon_i := y(x_i) - u_i.$$

Nếu, trong một phương pháp giải gần đúng, chúng ta có đánh giá sai số

$$|\epsilon_i| = O(h^k), \quad k > 0, \quad (6.23)$$

khi $h \rightarrow 0$, thì chúng ta nói rằng phương pháp có độ chính xác cấp k . Ví dụ, phương pháp Euler có độ chính xác $k = 1$, như đã biết ở mục trước. Nếu $k \geq 2$, chúng ta nói phương pháp đó có độ chính xác cao.

6.3.2 Khai triển tiệm cận của sai số

Từ Định lí 6.3, chúng ta có thể xây dựng một phương pháp có độ chính xác cấp hai. Ý tưởng ở đây, gợi ý bởi phương pháp ngoại suy Richardson, là

dùng hai bước đi, h và $h/2$, để khử số hạng bậc nhất của h trong khai triển tiệm cận sai số trong Định lý 6.3.

Giả sử phương pháp Euler với bước h cho xấp xỉ u_i của $y(x_i)$, viết một cách chi tiết là $u(x_i; h)$ để chỉ rõ bước nhảy nào được dùng để tính u_i . Tương tự như vậy, bước đi $\tilde{h} = \frac{h}{2}$ chúng ta thu được xấp xỉ $u(x_i; h/2)$. Cuối cùng, đặt

$$v(x_i; h) := 2u\left(x_i; \frac{h}{2}\right) - u(x_i; h). \quad (6.24)$$

Khi đó, từ (6.16), chúng ta thấy dãy $v(x_i; h)$ xấp xỉ giá trị $y(x_i)$ với sai số

$$|v(x_i; h) - y(x_i)| = O(h^2). \quad (6.25)$$

Phương pháp mô tả ở trên là một cải tiến của phương pháp Euler với độ chính xác cấp hai.

Ví dụ 6.5. Xét bài toán Cauchy trong Ví dụ 6.3. Hãy dùng phương pháp Euler gần đúng bài toán với $h = \frac{1}{4}$.

Lời giải

Với bước đi $h = \frac{1}{4}$, ta có các nút $x_0 = 0, x_1 = \frac{1}{4}, x_2 = \frac{1}{2}, x_3 = \frac{3}{4}$, và $x_4 = 1$. Với $f(x, y) = xe^{2x} - 2y$, chúng ta có

$$\begin{aligned} u_0 &= 0, \\ u_1 &= u_0 + hf(x_0, u_0) = 0, \\ u_2 &= u_1 + hf(x_1, u_1) = 0 + \frac{1}{4} \left(\frac{1}{4} \times e^{2 \times \frac{1}{4}} - 2 \times 0 \right) = \frac{\sqrt{e}}{16}, \\ u_3 &= u_2 + hf(x_2, u_2) = \dots = \frac{e}{8} + \frac{\sqrt{e}}{32}, \\ u_4 &= u_3 + hf(x_3, u_3) = \dots = \frac{3e\sqrt{e}}{16} + \frac{e}{16} + \frac{\sqrt{e}}{64}. \end{aligned}$$

Vậy,

$$y(1) \approx u_4 \approx 1,03597.$$

Kết hợp với Ví dụ 6.3, ta có

$$\begin{aligned} v\left(1; \frac{1}{2}\right) &= 2u\left(1; \frac{1}{4}\right) - u\left(1; \frac{1}{2}\right) \\ &= 2\left(\frac{3e\sqrt{e}}{16} + \frac{e}{16} + \frac{\sqrt{e}}{64}\right) - \frac{e}{4} \\ &= \frac{3e\sqrt{e}}{8} - \frac{e}{8} + \frac{\sqrt{e}}{64} \\ &\approx 1,39237. \end{aligned}$$

Mặt khác, từ nghiệm đúng của phương trình vi phân, ta có

$$y(1) = \frac{3}{16}e^2 + \frac{1}{16}e^{-2} \approx 1,393906.$$

Ta thấy sai số của $v(1; \frac{1}{2})$ bé hơn 0,0016. Như vậy, trong ví dụ này phương pháp đang xét có độ chính xác cao hơn phương pháp Euler với $h = \frac{1}{4}$.

6.3.3 Phương pháp hình thang

Phương pháp hình thang là một cải tiến của phương pháp Euler dựa trên một ý tưởng tương tự như phương pháp hình thang trong bài toán tính tích phân xác định đó người ta xây dựng dãy u_i như sau:

$$u_0 = \alpha, \quad (6.26)$$

$$u_{i+1} = u_i + \frac{h}{2} [f(x_i, u_i) + f(x_{i+1}, u_{i+1})]. \quad (6.27)$$

Có thể chứng minh được rằng phương pháp hình thang là một phương pháp có độ chính xác cấp 2.

Một nhược điểm của phương pháp hình thang xuất phát từ đặc điểm u_{i+1} xuất hiện ở cả hai vế của (6.27). Do đó, khi tính u_{i+1} , ta phải giải một phương trình của u_{i+1} . Quá trình giải sẽ cần khá nhiều năng lực tính toán nếu phương trình cần giải là phi tuyến.

Để tránh việc phải giải nghiệm của (6.27), người ta có thể cải tiến phương pháp này bằng cách thêm vào một quá trình lặp. Trước hết, chúng ta xét ví dụ sau trong đó phương trình cần giải là tuyến tính

Ví dụ 6.6. Xét bài toán Cauchy

$$\begin{cases} y' = y + x, \\ y(1) = 0. \end{cases}$$

Tính xấp xỉ giá trị $y(1,2)$ bằng phương pháp hình thang với bước đi $h = 0,2$. So sánh với giá trị tính bởi nghiệm đúng $y(x) = 2e^{x-1} - x - 1$.

Lời giải

Với bước đi $h = 0,2$, chúng ta chỉ tính 1 bước để xấp xỉ $y(1,2)$, cụ thể: $f(x, y) = y + x$, $u_0 = 0$. Theo công thức hình thang,

$$u_1 = u_0 + 0,1 (f(1,0) + f(1,2; u_1)) = 0,1(1 + u_1 + 1,2)$$

Đây là một phương trình của u_1 . Giải phương trình (trong trường hợp này, nó là một phương trình bậc nhất), ta thu được

$$u_1 = \frac{0,22}{0,9} \approx 0,24.$$

So sánh với công thức nghiệm giải tích $y(x) = 2e^{x-1} - x - 1$,

$$y(1,2) \approx 0,242806,$$

Chúng ta thấy rằng xấp xỉ $u_1 = 0,24$ có 2 chữ số đáng tin.

6.3.4 Phương pháp Heun

Trong tiểu mục này, ta nói về **Phương pháp Heun**², còn gọi là phương pháp “dự báo và điều chỉnh” (predictor–corrector method) hoặc phương pháp “hình thang hiện”.

Bắt đầu với $u_0 = \alpha$, để tính giá trị u_{i+1} từ u_i , ta tính một giá trị phụ, gọi là giá trị “dự báo” (predictor):

$$u_{i+1}^* = u_i + hf(x_i, u_i).$$

Tiếp theo, giá trị dự báo này được dùng để tính giá trị tiếp theo, gọi là giá trị “điều chỉnh” (corrector),

$$\begin{aligned} u_{i+1} &= u_i + \frac{h}{2} [f(x_i, u_i) + f(x_{i+1}, u_{i+1}^*)] \\ &= u_i + \frac{h}{2} [f(x_i, u_i) + f(x_{i+1}, u_i + hf(x_i, u_i))]. \end{aligned} \quad (6.28)$$

Dãy $\{u_i\}$ được dùng để xấp xỉ giá trị của nghiệm $y(x)$ của phương trình tại các điểm x_i .

Lưu ý, phương pháp Heun cũng có độ chính xác cấp 2, còn được gọi là phương pháp Euler cải tiến, hoặc Runge–Kutta 2.

Bài tập 6.1. Giải bài toán Cauchy

$$\begin{cases} y' = y + x, \\ y(1) = 0, \end{cases} \quad (6.29)$$

bằng phương pháp Heun với bước đi $h = 0,1$ và $N = 6$ và phương pháp Euler. So sánh sai số của hai phương pháp.

CAS 6.4. Phương pháp Heun có sẵn trong gói `pracma` của phần mềm R. Trong CAS này, ta sử dụng chức năng có sẵn để giải ví dụ trên.

```
# Số các chữ số trong các kết quả bằng 4
options(digits = 4)

# Nạp thư viện pracma của R
```

²Karl Heun (1859–1929) là một nhà toán học người Đức.

```
library(pracma)

# Hàm f(t, y) của vế phải của phương trình.
f <- function(t, y) t + y

# Gọi euler_heun() có sẵn trong pracma.
euler_heun(f, 1, 2, 0, 6, improved = TRUE)

## $t
## [1] 1,000 1,167 1,333 1,500 1,667 1,833 2,000
##
## $y
## [1] 0,0000 0,1944 0,4541 0,7907 1,2182 1,7530 2,4144
```

Kết quả là dãy các giá trị t_i và các giá trị xấp xỉ của $y(t_i)$.

6.3.5 Phương pháp Runge–Kutta

Trong tiểu mục này, ta nói về một phương pháp rất mạnh mẽ, được hai nhà toán học người Đức là Carl Runge³ và Wilhelm Kutta⁴ xây dựng và phát triển trong những năm 1900. Ngày nay, nó được gọi là **phương pháp Runge–Kutta 4 (RK4)**.

Phương pháp này thực hiện như sau: Đặt

$$u_0 = a \quad \text{là giá trị ban đầu.}$$

Giả sử ta đã tính được u_i , tính u_{i+1} qua các bước phụ như sau:

$$\begin{aligned} k_1 &= hf(x_i, u_i), \\ k_2 &= hf\left(x_i + \frac{h}{2}, u_i + \frac{k_1}{2}\right), \\ k_3 &= hf\left(x_i + \frac{h}{2}, u_i + \frac{k_2}{2}\right), \\ k_4 &= hf(x_i + h, u_i + k_3). \end{aligned}$$

Giá trị u_{i+1} được tính qua các giá trị ở trên như sau:

$$u_{i+1} = u_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \quad (6.30)$$

Phương pháp này có độ chính xác bậc 4.

³C. Runge, *Über die numerische Auflösung von Differentialgleichungen*. Math. Ann., 46(2):167–178, 1895.

⁴W. Kutta, *Beitrag zur näherungsweise Integration totaler Differentialgleichungen*. Z. Math. Phys., 46:435–453, 1901.

Ví dụ 6.7. Giải gần đúng bài toán giá trị ban đầu

$$y' = x - 2y + 3, \quad y(0) = 1, \quad (6.31)$$

bằng phương pháp Runge-Kutta bậc 4 để tính xấp xỉ $y(0,2)$ và $y(0,4)$, với bước đi $h = 0,2$.

Lời giải

Gọi $f(x, y) = x - 2y + 3$. Ta bắt đầu với

$$x_0 = 0, \quad u_0 = 1$$

là các giá trị ban đầu. Tiếp theo,

$$\begin{aligned} k_1 &= hf(x_0, u_0) = 0,2 \times f(0, 1) = 0,2, \\ k_2 &= hf\left(x_0 + \frac{h}{2}, u_0 + \frac{k_1}{2}\right) = 0,2 \times f(0,1, 1,1) = 0,18, \\ k_3 &= hf\left(x_0 + \frac{h}{2}, u_0 + \frac{k_2}{2}\right) = 0,2 \times f(0,1, 1,09) = 0,184, \\ k_4 &= hf(x_0 + h, u_0 + k_3) = 0,2 \times f(0,2, 1,184) = 0,1664. \end{aligned}$$

Thay vào công thức Runge-Kutta 4, ta có

$$\begin{aligned} u_1 &= u_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ &= 1 + \frac{1}{6}(0,2 + 2 \times 0,18 + 2 \times 0,184 + 0,1664) \\ &= 1,1824. \end{aligned}$$

Vậy, nghiệm xấp xỉ $y(0,2) \approx 1,1824$. Giá trị u_2 được tính tương tự.

Nhận xét 6.2. Công thức giải tích của nghiệm của bài toán giá trị ban đầu (6.31) là

$$y = \frac{1}{4}(5 + 2x - e^{-2x})$$

Từ đây có thể tính xấp xỉ giá trị $y(0,2) \approx 1,18242$. So sánh với giá trị tính bởi phương pháp Runge-Kutta, chúng ta thấy sai số của phương pháp RK4, trong trường hợp này, không vượt quá $0,3 \times 10^{-4}$.

CAS 6.5. Trong gói pracma của R, phương pháp Runge-Kutta 4 nằm trong `rk4()`, cách sử dụng như trong ví dụ sau:

```
library(pracma)

# Định nghĩa hàm số trong vế phải của phương trình
f <- function(x, y) x - 2 * y + 3

# Chạy rk4() cho thuật toán Runge-Kutta
rk4(f, 0, 0.4, 1, 2)

## $x
## [1] 0,0 0,2 0,4
##
## $y
## [1] 1,000 1,182 1,338
```

6.3.6 Sử dụng deSolve của R giải gần đúng phương trình vi phân

Trong mục này, chúng ta tìm hiểu gói deSolve⁵ có sẵn trong R để giải số bài toán Cauchy:

$$\begin{cases} \frac{dy}{dt} = f(t, y), \\ y(0) = \alpha. \end{cases}$$

Ở đây, t là biến độc lập. Tài liệu tham khảo: Bloomfield [3, Chapter 8].

CAS 6.6. Viết chương trình R, sử dụng chức năng ode() có sẵn trong gói deSolve để giải số bài toán giá trị ban đầu

$$\frac{dy}{dt} = t \exp(2t) - py,$$

trong trường hợp $p = 1$ và $p = 2$, trên đoạn $t \in [0, 1]$, với điều kiện ban đầu $y(0) = 0$.

Ta bắt đầu bằng việc đưa vào R hàm số ở vế phải của phương trình. Ở đây chỉ có 1 phương trình, nên vế phải sẽ là một danh sách (list) chỉ gồm 1 mục.

```
# Sử dụng gói deSolve
library(deSolve)

# Định nghĩa hàm func cho vế phải của phương trình.
func <- function(t, y, p){
  return(list(c(t * exp(2 * t) - p * y)))
}
```

⁵deSolve: Solvers for Initial Value Problems of Differential Equations ('ODE', 'DAE', 'DDE'). DOI: <https://doi.org/10.32614/CRAN.package.deSolve>

Tiếp theo, đưa vào R điều kiện ban đầu y_0 và dãy các điểm t_i mà ta cần tính giá trị $y(t_i)$. Đó là 11 điểm trên đoạn $[0, 1]$ cách đều với khoảng cách giữa hai điểm liên tiếp là $h = 0,1$.

```
# Điều kiện ban đầu
initial_value <- c(y_0 = 0)

# Dãy giá trị t cần tính
t_values <- seq(0, 1, 0.1)
```

Giá trị của tham số $p = 2$.

```
# Tham số p = 2
p <- 2
```

Bước cuối cùng, gọi hàm `ode()` của thư viện `deSolve` theo mẫu sau:

```
# Giải phương trình bằng ode() trong gói deSolve

result <- ode(initial_value, t_values, func, p)
```

Tổng hợp các bước được mô tả như trên, chúng ta thu được chương trình R như sau:

```
library(deSolve)
f <- function(t,y,p){
  return(list(c(t * exp(2 * t) - p * y)))
}
initial_value <- c(y_values = 0)
t_values <- seq(0, 1, 0.1)
p <- 2
result <- ode(initial_value, t_values, f, p)
print(result)

##      time y_values
## 1    0,0 0,000000
## 2    0,1 0,005369
## 3    0,2 0,023248
## 4    0,3 0,057078
## 5    0,4 0,111542
## 6    0,5 0,192887
## 7    0,6 0,309336
## 8    0,7 0,471624
## 9    0,8 0,693662
## 10   0,9 0,993401
## 11   1,0 1,393908
```

Kết quả là 11 giá trị gần đúng $y(t_i)$ tại các điểm t_i của nghiệm $y(t)$ của bài toán Cauchy trong trường hợp tham số $p = 2$. Ví dụ, có thể đọc từ đầu ra của chương trình kết quả $y(0,6) \approx 0,309336$.

Để giải gần đúng cho các phương trình vi phân khác, chúng ta chỉ cần sửa đổi mã lệnh trên cho phù hợp.

6.4 Phương trình vi phân bậc cao và hệ phương trình

Dưới đây, chúng ta tìm hiểu một số mở rộng của phương pháp Euler để giải gần đúng hệ phương trình vi phân và phương trình vi phân bậc cao. Tài liệu tham khảo: T. V. Đĩnh [6, §6.6, 6.7].

6.4.1 Hệ phương trình vi phân

Xét một hệ phương trình vi phân

$$\frac{dY}{dx} = F(x, Y),$$

với $F = (f_1, f_2, \dots, f_n)$ là hàm giá trị vectơ với n thành phần của $n + 1$ biến và $Y(x) = (y_1(x), y_2(x), \dots, y_n(x))$. Trong bài toán Cauchy, chúng ta thêm vào một điều kiện dạng

$$Y(x_0) = \alpha \in \mathbb{R}^n,$$

trong đó α và một vectơ cho trước.

Với xấp xỉ đầu tiên, ta chọn

$$U_0 = \alpha.$$

Với bước đi h , ta viết

$$x_j = x_0 + jh.$$

Trong phương pháp Euler, U_j được tính theo công thức truy hồi

$$U_{j+1} = U_j + hF(x_j, U_j).$$

Ví dụ 6.8. Giải xấp xỉ phương trình

$$\begin{aligned} y_1' &= x + y_1 y_2, \\ y_2' &= x^2 - y_1^2, \end{aligned}$$

với điều kiện ban đầu

$$\begin{aligned} y_1(0) &= 1, \\ y_2(0) &= 0, \end{aligned}$$

bằng phương pháp Euler với bước đi $h = 0,1$.

Lời giải

Chúng ta xấp xỉ nghiệm $Y = (y_1, y_2)$ tại các giá trị $x_j = 0 + jh = 0, 1$ j bởi các vectơ $U_j = (u_{1,j}, u_{2,j})$, tính theo phương pháp Euler với các tham số như sau:

$$h = 0,1, \quad U_0 = (u_{1,0}, u_{2,0}) = (1,0) \quad (\text{điều kiện ban đầu}),$$

và

$$\begin{cases} u_{1,j+1} = u_{1,j} + h(x_j + u_{1,j}u_{2,j}), \\ u_{2,j+1} = u_{2,j} + h(x_j^2 - u_{1,j}^2). \end{cases}$$

Từ đó, tính cụ thể ta được

$$\begin{aligned} u_{1,1} &= 1 + 0,1(0 + 1 \times 0) = 1, \\ u_{2,1} &= 0 + 0,1(0^2 - 1^2) = -0,1. \end{aligned}$$

Các giá trị tiếp theo được tính hoàn toàn tương tự.

CAS 6.7. Viết code R, sử dụng `ode()` của `deSolve`, để giải xấp xỉ hệ phương trình

$$\begin{aligned} \frac{dx}{dt} &= pt + xy, \\ \frac{dy}{dt} &= t^2 - x^2, \end{aligned}$$

trên đoạn $[0, 1]$, với điều kiện ban đầu

$$\begin{aligned} x(0) &= 1, \\ y(0) &= 0, \end{aligned}$$

với bước đi $h = 0,1$, và tham số $p = 2$.

Tương tự như ví dụ trước, chi tiết chương trình của chúng ta như sau:

```
library(deSolve)

# Định nghĩa hàm ở vế phải của phương trình
f <- function(t, y, p){
  return(list(c(p * t + y[1] * y[2], t^2 - y[1]^2)))
}

# Giá trị ban đầu
initial_value <- c(x_values = 1, y_values = 0)

# Dãy các giá trị của t tại đó cần tính xấp xỉ nghiệm
t_values <- seq(0,1,0.1)
```

```
# Tham số p = 2
p <- 2

# Giải phương trình bằng ode() và hiển thị kết quả
result <- ode(initial_value, t_values, f, p)
print(result)

##      time x_values y_values
## 1    0,0      1,000    0,0000
## 2    0,1      1,005   -0,1000
## 3    0,2      1,020   -0,2000
## 4    0,3      1,044   -0,3000
## 5    0,4      1,077   -0,4000
## 6    0,5      1,118   -0,4999
## 7    0,6      1,165   -0,5997
## 8    0,7      1,217   -0,6992
## 9    0,8      1,274   -0,7980
## 10   0,9      1,334   -0,8956
## 11   1,0      1,395   -0,9914
```

6.4.2 Phương trình vi phân cấp cao

Một phương trình vi phân cấp cao có thể đưa về hệ phương trình vi phân cấp một bằng cách đặt các đạo hàm bậc thấp là các biến mới. Từ đó, chúng ta có thể áp dụng các phương pháp giải hệ phương trình vi phân cấp một để giải.

CAS 6.8. Xét phương trình nổi tiếng của van der Pol⁶ sau đây,

$$z'' - \mu(1 - z^2)z' + z = 0.$$

Phương trình này có thể được đưa về hệ phương trình vi phân cấp một bằng cách đặt $y = z'$ and $x = z$). Khi đó,

$$y' = z'',$$

và vậy thì ($' = d/dt$)

$$\begin{cases} x' = y, \\ y' = \mu(1 - x^2)y - x = 0. \end{cases} \quad .$$

Giả sử chúng ta cần giải phương trình này với điều kiện ban đầu

$$x(0) = 2, \quad y(0) = 0.$$

⁶B. van der Pol and J. van der Mark. *Frequency de multiplication*. Nature, 120:363–364, 1927.

Hệ phương trình có thể được giải bằng R như sau:

```
library(deSolve)

# Vế phải của phương trình van der Pol

vdpol <- function(t, y, mu){
  return(list(c(y[2], mu * (1 - y[1]^2) * y[2] - y[1])))
}

# Giá trị ban đầu, giá trị cần giải.

initial_value <- c(x_values = 2, y_values = 0)
t_values <- seq(0, 1, 0.1)

# Tham số mu

mu <- 1000

# Giải phương trình bằng ode() của deSolve

result <- ode(initial_value, t_values, vdpol, mu)
print(result)
```

##	time	x_values	y_values
## 1	0,0	2,000	0,0000000
## 2	0,1	2,000	-0,0006667
## 3	0,2	2,000	-0,0006667
## 4	0,3	2,000	-0,0006668
## 5	0,4	2,000	-0,0006668
## 6	0,5	2,000	-0,0006669
## 7	0,6	2,000	-0,0006669
## 8	0,7	2,000	-0,0006669
## 9	0,8	1,999	-0,0006670
## 10	0,9	1,999	-0,0006670
## 11	1,0	1,999	-0,0006670

6.5 Sơ lược về bài toán biên

Những bài toán trong đó điều kiện cho nghiệm của một phương trình vi phân phải thỏa mãn được đặt ra tại cả hai đầu mút của một đoạn $[a, b]$ được gọi là *bài toán biên* (boundary-value problem). Những vấn đề như vậy thường phức tạp hơn vấn đề giá trị ban đầu mà chúng ta đã nói đến trong các mục trước.

Chúng ta hãy bắt đầu với bài toán tuyến tính

$$y'' + P(x)y' + Q(x)y = F(x), \quad a < x < b \quad (6.32)$$

với các điều kiện biên $y(a) = A$, $y(b) = B$. Sự tồn tại và duy nhất nghiệm của bài toán biên là những vấn đề lí thuyết khá phức tạp, ngay cả với trường hợp phương trình tuyến tính (6.32). Ví dụ, xét phương trình “dao động điều hoà”

$$y'' + \lambda y = 0, \quad \lambda > 0, \quad (6.33)$$

trên khoảng $(0, \pi)$ với “điều kiện biên Dirichlet”

$$y(0) = y(\pi) = 0. \quad (6.34)$$

Nghiệm tổng quát của phương trình (6.33) là

$$y(x) = C_1 \sin(\sqrt{\lambda} x) + C_2 \cos(\sqrt{\lambda} x).$$

Thay điều kiện biên, chúng ta có các ràng buộc đối với C_1 và C_2 như sau:

$$C_2 = 0, \quad C_1 \sin(\sqrt{\lambda} \pi) + C_2 \cos(\sqrt{\lambda} \pi) = 0.$$

Xảy ra hai trường hợp sau:

- Nếu $\sin(\sqrt{\lambda} \pi) = 0$, hay $\sqrt{\lambda}$ là một số nguyên⁷, thì C_1 có thể chọn tuỳ ý. Nghiệm của bài toán biên có dạng

$$y(x) = C_1 \sin(kx), \quad k^2 = \lambda.$$

- Nếu $\sin(\sqrt{\lambda} \pi) \neq 0$ thì C_1 bắt buộc phải bằng 0; nghiệm duy nhất của bài toán biên là $y(x) = 0$.

Ở ví dụ tiếp theo, chúng ta xét

$$y'' + \lambda y = f(x), \quad \lambda > 0, \quad (6.35)$$

Khi đó, nếu $\sqrt{\lambda}$ không phải là số nguyên thì (6.35) có không nhiều hơn một nghiệm. Trong trường hợp còn lại, $\sqrt{\lambda}$ là một số nguyên và nếu $y_p(x)$ là một nghiệm của (6.35) thì (6.35) có vô số nghiệm, dạng

$$y(x) = C \sin(\sqrt{\lambda} x) + y_p(x).$$

Mặt khác, cũng dễ dàng chỉ ra một số trường hợp nghiệm không tồn tại. Một điều kiện cần để (6.35) có nghiệm đó là vế phải $f(x)$ phải “trực giao”

⁷ Các giá trị $\lambda = k^2$ chính là các giá trị riêng của toán tử Laplace một chiều trên $[0, \pi]$ với điều kiện biên Dirichlet

với không gian nghiệm của phương trình thuần nhất (6.33). Cụ thể, giả sử tồn tại một nghiệm y_p của (6.35),

$$f(x) = y_p''(x) + \lambda y_p(x).$$

Trước hết, chúng ta áp dụng công thức tích phân từng phần

$$\begin{aligned} \int_0^\pi y_p''(x) \sin(\sqrt{\lambda} x) dx &= y_p'(x) \sin(\sqrt{\lambda} x) \Big|_0^\pi - \int_0^\pi y_p'(x) \frac{d \sin(\sqrt{\lambda} x)}{dx} dx \\ &= -\sqrt{\lambda} \int_0^\pi y_p'(x) \cos(\sqrt{\lambda} x) dx \\ &= y_p(x) \cos(\sqrt{\lambda} x) \Big|_0^\pi - \lambda \int_0^\pi y_p(x) \sin(\sqrt{\lambda} x) dx \\ &= -\lambda \int_0^\pi y_p(x) \sin(\sqrt{\lambda} x) dx. \end{aligned}$$

Từ đó,

$$\begin{aligned} \int_0^\pi f(x) \sin(\sqrt{\lambda} x) dx \\ = \int_0^\pi y_p''(x) \sin(\sqrt{\lambda} x) dx + \lambda \int_0^\pi y_p(x) \sin(\sqrt{\lambda} x) dx = 0. \end{aligned} \quad (6.36)$$

Điều kiện (6.36) là điều kiện cần để bài toán biên (6.35) có nghiệm.

Trở lại với phương trình (6.32). Ý tưởng ở đây là tìm nghiệm của nó dưới dạng

$$y(x) = u(x) + c w(x),$$

trong đó $u(x)$ là nghiệm của phương trình (6.32) với giá trị ban đầu

$$u(a) = A, \quad u'(a) = 0,$$

trong khi đó $v(x)$ là nghiệm của phương trình thuần nhất

$$z'' + P(x)z' + Q(x)z = 0.$$

Hằng số c sẽ được xác định sao cho $u(b) + c w(b) = B$.

Để tìm u , ta đặt $v = u'$ và đưa phương trình ban đầu về hệ hai phương trình

$$\begin{cases} v' = -P(x)v - Q(x)u + F(x), \\ u' = v. \end{cases}$$

với điều kiện ban đầu $(u(a), v(a)) = (A, 0)$.

Tương tự, $z(x)$ được xác định qua hệ phương trình vi phân cấp 1

$$\begin{cases} z' = w, \\ w' = -P(x)w - Q(x)z, \end{cases}$$

với điều kiện ban đầu $(z(a), w(a)) = (0, 1)$.

6.6 Câu hỏi và bài tập

1. Sử dụng phương pháp Euler với bước đi $h = 0,1$ để giải gần đúng các bài toán Cauchy sau:

(a) $y' + 0,5y = 0, 0 \leq x \leq 1, y(0) = 2.$

(b) $2y' - xy = 0, 0 \leq x \leq 1, y(0) = 1.$

(c) $y' = x^2 + y^2, 0 \leq x \leq 1, y(0) = 1.$

(d) $y' = (y + x)^2, 0 \leq x \leq 1, y(0) = 0.$

2. Sử dụng phương pháp Heun (dự báo và điều chỉnh) với bước đi $h = 0,1$ để giải gần đúng các bài toán Cauchy sau:

(a) $2y' - xy = 0, 0 \leq x \leq 1, y(0) = 1.$

(b) $y' = x^2 + y^2, 0 \leq x \leq 1, y(0) = 1.$

3. Dùng phương pháp Heun (dự báo và điều chỉnh) với bước đi $h = 0,1$ giải gần đúng bài toán Cauchy trong mô hình “dân số logistic” sau:

$$y' = y - y^2, y(0) = 0,2.$$

4. Cho bài toán giá trị ban đầu

$$y' = \frac{y - x}{y + x}, y(0) = 1.$$

(a) Bằng phương pháp Euler với bước đi $h = 0,1$, tìm nghiệm gần đúng của bài toán trên tại điểm $x = 0,4$.

(b) Bằng phương pháp Runge–Kutta 4 với bước đi $h = 0,1$, tìm nghiệm gần đúng của bài toán trên tại điểm $x = 0,4$.

5. Sử dụng phương pháp Runge–Kutta 4 với bước đi $h = 0,1$, tìm gần đúng $y(0,5)$ với y là nghiệm của bài toán Cauchy sau:

(a) $y' = xy^2, y(0) = 1.$

(b) $y' = 1 + y^2, y(0) = 0.$

(c) $y' + y \tan(x) = \sin(2x), y(0) = 1.$

(d) $y' = 4x^3 y^2, y(0) = \frac{1}{2}.$

Tài liệu tham khảo

- [1] K. E. Atkinson. *An Introduction to Numerical Analysis*. 2nd Edition. John Wiley & Sons, Inc. 1989.
- [2] P. K. Anh. *Giải tích số*. In lần thứ 8. NXB Đại học Quốc Gia Hà Nội, Hà Nội, 2008.
- [3] V. A. Bloomfield. *Using R for Numerical Analysis in Science and Engineering*. CRC Press, Taylor & Francis Group, Boca Raton 2014.
- [4] R. L. Burden, J. D. Faires. *Numerical Analysis*. 9th edition. Brooks/Cole, Cengage Learning, 2011.
- [5] G. Dahlquist, A. Björck. *Numerical Methods*. Prentice-Hall, Inc., 1974.
- [6] T. V. Đình. *Phương pháp tính: Dùng cho các trường đại học kỹ thuật*. Tái bản lần thứ 19. NXB Giáo dục, Hà Nội, 2017.
- [7] O. Jones, R. Maillardet, A. Robinson. *Introduction to Scientific Programming and Simulation Using R*. 2nd edition. Chapman & Hall, 2014.
- [8] N. V. Khuê, P. N. Thao, L.M. Hải, N. Đ. Sang. *Toán cao cấp: Tập 1-3*. NXB Giáo dục, Hà Nội 1997.
- [9] E. Kreyszig. *Advanced Engineering Mathematics, 10th Edition*. John Wiley & Sons, 2011.
- [10] D. N. Sơn, B. X. Quang, P. Q. Sáng, Đ. Q. Tuấn. *Giáo trình đại số tuyến tính*. NXB Đại học Quốc gia Hà Nội, 2025.
- [11] H. J. Stetter. *Analysis of discretization methods for ordinary differential equations*, volume 23. Springer, 1973.
- [12] Е. А. Волков. Численные Методы. Москва «Наука», Главная Редакция, Физико-Математической Литературы, 1987.

Bảng thuật ngữ

bài toán
 giá trị ban đầu, 186
 biên, 209
 Cauchy, 186

chữ số
 chắc, 6
 có nghĩa, 6

cầu phương Gauss, 167

hàm ghép trơn, 123
hàm Lambert, 18
hàm nội suy cơ bản, 100

khoảng tách nghiệm, 31

lược đồ Horner, 96

phương pháp
 chia đôi, 33
 dây cung, 55
 Euler, 188
 Euler ngược, 194
 Heun, 198
 khử Gauss, 65
 Newton–Raphson, 44
 Runge–Kutta 4, 202
 Steffensen, 57
 vị trí sai, 58

quy tắc
 hình thang hiệu chỉnh, 164

sai phân

lùi, 110
 tiến, 110
sai số
 quy tròn, 8
 thật sự, 2
 tuyệt đối, 2
 tuyệt đối giới hạn, 2
 tương đối, 3

tỉ hiệu, 105
tỉ sai phân, 105

vô cùng bé, 198

đa thức
 nội suy, 92
 nội suy Lagrange, 99
 nội suy Newton, 105
đa thức trực giao, 167
đơn điệu
 giảm, 31
 tăng, 31
định lí
 Bolzano–Cauchy, 30