

# Phương pháp tính MAT1099

Lê Phê Đô  
dolp@vnu.edu.vn

07/09/2020





# Mục lục

<b>1</b>	<b>Giải tích sai số</b>	<b>5</b>
1.1	Giới thiệu về phép tính gần đúng . . . . .	5
1.1.1	Một số ví dụ về tính toán khoa học và phương pháp tính	5
1.2	Số gần đúng, sai số tuyệt đối và tương đối . . . . .	6
1.2.1	Sai số tuyệt đối . . . . .	6
1.2.2	Sai số tương đối . . . . .	7
1.2.3	Các loại sai số khác . . . . .	7
1.3	Sai số tích lũy và các bài toán sai số . . . . .	8
1.3.1	Sai số hàm một biến . . . . .	8
1.3.2	Sai số qua các phép toán số học . . . . .	9
1.3.3	Sai số hàm nhiều biến . . . . .	10
1.4	Sai số quy tròn, quan hệ sai số & số chữ số chắc . . . . .	12
1.4.1	Chữ số có nghĩa . . . . .	12
1.4.2	Chữ số chắc . . . . .	12
1.4.3	Số thu gọn . . . . .	13
1.4.4	Dấu phẩy động . . . . .	14



# Chương 1

## Giải tích sai số

### 1.1 Giới thiệu về phép tính gần đúng

#### 1.1.1 Một số ví dụ về tính toán khoa học và phương pháp tính

Lý do nghiên cứu Phương pháp tính

- Làm việc với các số gần đúng.
- Giải gần đúng các phương trình và hệ phương trình.
- Xấp xỉ hàm số: Phương pháp nội suy, phương pháp xấp xỉ hàm số, chuỗi Taylor hoặc chuỗi Maclaurin.
- Số học IEEE.

Các nhiệm vụ

- Tìm hiểu và ứng dụng các thuật toán.
- Thể hiện các thuật toán bằng các chương trình.
- Tìm các bài toán thực tiễn.

Trong thực tế chúng ta thường phải xử lý, tính toán với các đại lượng gần đúng như các số đo vật lý, các dữ liệu ban đầu, các số làm tròn...với sai số nào đó, tức là các số gần đúng. Việc ước lượng sai số hợp lý cho phép ta đánh giá được chất lượng của quá trình tính toán, quyết định số chữ số giữ lại trong các phép tính trung gian và trong kết quả. Vì vậy, trước tiên ta cần nghiên cứu về các phép tính gần đúng và sai số.

## 1.2 Số gần đúng, sai số tuyệt đối và tương đối

### 1.2.1 Sai số tuyệt đối

Nếu số gần đúng  $a$  có giá trị đúng là  $a_0$  thì ta nói  $a$  xấp xỉ  $a_0$  hay  $a$  là số gần đúng của  $a_0$ . Khi đó sai số của  $a$  là:

$$E_a = a - a_0 \quad (1.1)$$

Nhưng giá trị này nói chúng ta không biết được mà chỉ ước lượng được cận trên của trị tuyệt đối của nó.

**Định nghĩa 1.** *Giá trị ước lượng  $\Delta a$  sao cho*

$$|a - a_0| \leq \Delta a \quad (1.2)$$

*được gọi là sai số tuyệt đối của số gần đúng  $a$ .*

Sai số tuyệt đối nhỏ nhất có thể biết được gọi là sai số tuyệt đối giới hạn của  $a$ . Thông thường ước lượng sai số tuyệt đối giới hạn là khó và nhiều khi không cần thiết nên người ta chỉ cần ước lượng sai số tuyệt đối đủ nhỏ và dùng từ 1 đến 3 chữ số có nghĩa (là số chữ số bắt đầu từ chữ số khác không đầu tiên từ trái sang phải - xem mục 2.1) để biểu diễn sai số tuyệt đối của số gần đúng.

Thay cho 1.2 người ta còn dùng cách biểu diễn sau để chỉ sai số tuyệt đối của  $a$ :

$$a_0 = a \pm \Delta a \quad (1.3)$$

Trong thực tế thì sai số  $E_a$  không thể biết được nên khi không có sự hiểu lầm người ta còn dùng từ *sai số* để chỉ sai số tuyệt đối  $E_a$ .

**Ví dụ 1.1.** *Căn phòng có chiều dài  $d = 5,45$  m và chiều rộng  $r = 3,94$  m với sai số 1 cm.*

*Khi đó ta hiểu là:*

$$\Delta d = 0,01 \text{ m hay } d = (5,45 \pm 0,01) \text{ m}$$

$$\Delta r = 0,01 \text{ m hay } r = (3,94 \pm 0,01) \text{ m}$$

*Như vậy diện tích của phòng được ước lượng bởi:*

$$S = d \cdot r = 5,45 \cdot 3,94 = 21,473 \text{ m}^2$$

*với cận trên và cận dưới của  $S$  là:*

$$(5,45 - 0,01)(3,94 - 0,01) = 21,3792 \leq S \leq (5,45 + 0,01)(3,94 + 0,01) = 21,567$$

*Vậy ta có ước lượng sai số tuyệt đối của  $S$  là:*

$$|S - S_0| \leq 0,094 \text{ m}^2$$

### 1.2.2 Sai số tương đối

Hai số gần đúng có cùng sai số tuyệt đối sẽ có "mức độ chính xác" khác nhau nếu độ lớn của chúng khác nhau, số bé hơn sẽ có độ chính xác kém hơn. Để biểu diễn độ chính xác này người ta dùng sai số sai số tương đối.

**Định nghĩa 2.** Sai số tương đối của số gần đúng  $a$  là tỷ số giữa sai số tuyệt đối và giá trị tuyệt đối của nó, được ký hiệu là  $\delta a$ .

$$\delta a = \frac{\delta a}{|a|} \quad (1.4)$$

Thường sai số tương đối được biểu diễn dưới dạng phần trăm với 2 hoặc 3 chữ số.

Từ 1.4 ta thấy nếu biết  $\delta a$  thì:

$$\Delta a = |a|\delta a \quad (1.5)$$

nên ta chỉ cần biết một trong hai loại sai số của nó là được.

**Ví dụ 1.2.** Nếu  $a = 57$  và  $\Delta a = 0,5$  thì  $\delta a = 0,0087719$  hoặc  $0,88\%$  (gọn hơn là  $0,9\%$ ).

### 1.2.3 Các loại sai số khác

Để hình dung các loại sai số khác ta xét ví dụ sau:

**Ví dụ 1.3.** Một vật thể rơi từ độ cao  $H_0$  với vận tốc ban đầu  $v_0$  (được đo nhờ thiết bị nào đó). Tính độ cao  $H(t)$  của vật thể sau thời gian  $t$ . Bài toán có thể giải như sau:

Nếu gọi ngoại lực tác động vào vật thể là  $F(t)$  (gồm lực hút trọng trường và lực cản), khối lượng vật thể là  $m$  thì  $H(t)$  là nghiệm của phương trình vi phân cấp hai

$$H''(x) = \frac{-F(t)}{m} \quad (1.6)$$

với điều kiện ban đầu  $H(0) = H_0$  và  $H'(0) = -v_0$ .

Ta chọn một phương pháp gần đúng để giải phương trình này, chẳng hạn nếu giả thiết  $\frac{F(t)}{m}$  không đổi thì

$$H(t) = H_0 - g\frac{t^2}{2} - v_0 t$$

Qua ví dụ trên ta thấy sai số của kết quả nhận được chịu ảnh hưởng của:

- các số đo  $H_0, v_0$
- cách lập luận để xác định  $F(t)$
- phương pháp giải phương trình 1.6



- và các yếu tố ngẫu nhiên khác

Theo các yếu tố ảnh hưởng tới kết quả tính toán ta phân ra các loại sai số sau:

- *Sai số dữ liệu*: Còn gọi là sai số của số liệu ban đầu. Trong thí dụ trên là sai số khi đo  $H_0$  và  $v_0$ .
- *Sai số giả thiết*: Sai số này gặp phải khi ta đơn giản hoá bài toán thực tiễn để thiết lập mô hình toán học có thể giải được. Trong thí dụ trên có thể giả thiết ngoại lực chỉ là trọng lực.
- *Sai số phương pháp*: Là sai số của phương pháp giải gần đúng bài toán theo mô hình được lập. Trong thí dụ trên là phương pháp giải phương trình vi phân 1.6.
- *Sai số tính toán*: Là sai số tích lũy trong quá trình tính toán theo phương pháp được chọn.
- *Sai số làm tròn*: Khi tính toán ta thường phải làm tròn các số nên ảnh hưởng tới kết quả nhiều khi rất đáng kể.
- *Sai số ngẫu nhiên*: Là sai số chịu các quy luật chi phối ngẫu nhiên không tránh được.

Về sau ta quan tâm tới sai số tính toán và sai số phương pháp.

### 1.3 Sai số tích lũy và các bài toán sai số

#### 1.3.1 Sai số hàm một biến

Cho hàm số  $y = f(x)$  và  $x$  là số gần đúng của  $x_0$ . Ký hiệu  $\Delta x$  và  $\Delta y$  là sai số tuyệt đối tương ứng của đối số và hàm số. Ta sẽ xét các bài toán ước lượng sai số của hàm hoặc của đối số khi biết một trong hai sai số.

##### Bài toán thuận

Bài toán này ta ước lượng  $\Delta y$  khi biết  $x$  và  $\Delta x$ .

Theo công thức số gia hữu hạn ta có:

$$|y - y_0| = |f'(c)||x - x_0|$$

ở đây  $y_0$  là giá trị đúng của  $y$  và  $c$  là điểm thuộc miền  $(x, x_0)$  nếu  $x < x_0$  và thuộc  $(x_0, x)$  nếu  $x_0 < x$ .

Khi  $\Delta x$  bé,  $x$  gần  $x_0$  ta có ước lượng:

$$\begin{aligned} \Delta y &\approx |f'(x)||x - x_0| \\ \text{hay } \Delta y &\leq |f'(x)|\Delta x \end{aligned} \tag{1.7}$$

**Ví dụ 1.4.** Cho  $y = \ln x$  ta có ước lượng:

$$\Delta(\ln x) = \frac{1}{x}\Delta x = \delta x$$

**Bài toán ngược**

Trong bài toán này, ta biết giá trị gần đúng  $x$ , ta cần xác định phải tính  $x$  với  $\Delta x$  là bao nhiêu để đảm bảo  $\Delta y \leq \Delta$ . Với giá trị  $\Delta$  cho trước, từ công thức 1.7 ta thấy nếu

$$\Delta x \leq \frac{\Delta}{|f'(x)|} \quad (1.8)$$

thì đủ để  $\Delta y \leq \Delta$ .

**Ví dụ 1.5.**  $y = e^x$  với  $x \approx 3$  để có  $\Delta y \leq 0,01$  ta tính  $x$  với  $\Delta x \leq \frac{0,01}{e^3}$  là đủ.

**1.3.2 Sai số qua các phép toán số học**

Khi tính toán với các số gần đúng thì sai số sẽ tích lũy qua các phép toán cơ bản. Sau đây ta ước lượng sai số khi cộng trừ, nhân chia các số gần đúng.

**Sai số của tổng hoặc hiệu**

**Mệnh đề.** Sai số tuyệt đối của một tổng hoặc hiệu bằng tổng các sai số tuyệt đối thành phần.

*Chứng minh.* Để đơn giản ta xét  $u = a \pm b$  với các số  $a, b$  có giá trị đúng  $a_0, b_0$  và sai số tuyệt đối  $\Delta a, \Delta b$  tương ứng. Trường hợp có nhiều số hạng được xét tương tự.

Khi đó ta có:

$$\begin{cases} a_0 - \Delta a \leq a \leq a_0 + \Delta a \\ b_0 - \Delta b \leq b \leq b_0 + \Delta b \end{cases}$$

Do đó ta có:

$$\begin{cases} a_0 + b_0 - (\Delta a + \Delta b) \leq a + b \leq a_0 + b_0 + (\Delta a + \Delta b) \\ a_0 - b_0 - (\Delta a + \Delta b) \leq a - b \leq a_0 - b_0 + (\Delta a + \Delta b) \end{cases}$$

Nên

$$a_0 \pm b_0 - (\Delta a + \Delta b) \leq a \pm b \leq a_0 \pm b_0 + (\Delta a + \Delta b)$$

đpcm.

**Ví dụ 1.6.** Cho  $a = 50,5$ ,  $b = 50,9$  với  $\Delta a = \Delta b = 0,05$  và  $u = a - b$ .

Ta có  $u = 0,4$  với  $\Delta u = 0,05 + 0,05 = 0,1$ .

$$\text{Vậy } \delta u = \frac{0,1}{0,4} = 25\%.$$

**Sai số của tích hoặc thương**

**Mệnh đề.** Sai số tương đối của tích hoặc thương bằng tổng các sai số tương đối thành phần.

Chứng minh. Xét

$$u = \frac{x_1 \cdot \dots \cdot x_m}{y_1 \cdot \dots \cdot y_n}$$

Ta có thể giả thiết các  $x_i$  và  $y_j$  đều dương. Khi đó ta có:

$$\ln u = \sum_{i=1}^m \ln x_i + \sum_{j=1}^n \ln y_j$$

Theo mệnh đề 1.3.2 ở trên, ta suy ra:

$$\delta u = \sum_{i=1}^m \delta x_i + \sum_{j=1}^n \delta y_j$$

đpcm.

**Ví dụ 1.7.** Xét  $S = d \cdot r$  như ở ví dụ 1.1  $d = 5,45$ ,  $r = 3,94$ ,  $\Delta d = \Delta r = 0,01$ .  
Ta có:

$$\delta d = 0,001\,835$$

$$\delta r = 0,002\,538$$

$$\delta S = 0,004\,373 \text{ nên } \Delta S = 0,094$$

**1.3.3 Sai số hàm nhiều biến**

Ta xét hàm nhiều biến  $u = f(x_1, \dots, x_n)$  với giá trị gần đúng  $x_1, \dots, x_n$  và  $y$  đã biết.

**Bài toán thuận**

Trong bài toán này, ta cần ước lượng sai số  $\Delta y$  khi biết  $\delta x_i \forall i \leq n$ .

Tương tự hàm một biến, sử dụng công thức số gia hữu hạn ta có ước lượng:

$$\Delta u = \sum_{i=1}^n |f'_i(x_1, \dots, x_n)| \Delta x_i \quad (1.9)$$

với  $f'_i$ ) là đạo hàm riêng của  $u$  theo biến  $x_i$ .

**Ví dụ 1.8.** Xét  $u = a^2 b$  với  $a = 2,0$ ,  $b = 25,0$ ,  $\Delta a = \Delta b = 0,1$ .

Ta có:

$$u = 100$$

$$\Delta u = 2ab\Delta a + a^2\Delta b$$

$$= 2 \cdot 2,0 \cdot 25,0 \cdot 0,1 + 2,0^2 \cdot 0,1$$

$$= 10,4$$

**Bài toán ngược**

Bây giờ ta đã biết các số gần đúng  $x_i$ , ta phải tính chúng với sai số tuyệt đối như thế nào để có  $\Delta y \leq \Delta$ ; ở đây  $\Delta$  là số cho trước.

Các phương pháp xử lý bài toán này đều dựa trên công thức 1.9 một cách linh hoạt. Sau đây ta xét hai phương pháp thông dụng.

1. Sai số của đối số như nhau. Ta xét khi:

$$\Delta x_k = \Delta x \forall k \leq n$$

Từ 1.8 ta có:

$$\Delta u = \sum_{i=1}^n |f'_i(x_1, \dots, x_n)| \Delta x$$

Vậy để cho  $\Delta u \leq \Delta$  thì chỉ cần:

$$\Delta x \leq \frac{\Delta}{\sum_{i=1}^n |f'_i(x_1, \dots, x_n)|} \quad (1.10)$$

là đủ.

2. Phân bố đều sai số. Bây giờ ta xét khi

$$|f'_i(x_1, \dots, x_n)| \Delta x_i = |f'_k(x_1, \dots, x_n)| \Delta x_k \forall i, k \leq n$$

Khi đó  $\forall j \leq n$ , từ 1.9 ta có:

$$\Delta u = n |f'_i(x_1, \dots, x_n)| \Delta x_j$$

Vậy để cho  $\Delta u \leq \Delta$  thì chỉ cần tính:

$$\Delta x_j \leq \frac{\Delta}{n |f'_i(x_1, \dots, x_n)|} \quad \forall j = 1, \dots, n \quad (1.11)$$

**Ví dụ 1.9.** *Mảnh vườn có cạnh  $d \approx 45,0$  m và  $r \approx 20,0$  m. Cần tính  $d$  và  $r$  với  $\Delta d, \Delta r$  như thế nào để  $\Delta S \leq 0,1$  m<sup>2</sup>.*

Cách 1. Xét  $\Delta d = \Delta r = \Delta x$ , ta áp dụng 1.10:

$$\Delta x \leq \frac{0,1}{45 + 20} = 0,0015 \text{ m}$$

Cách 2. Khi đo chiều dài thường có sai số lớn hơn chiều rộng nên ta có thể dùng 1.11.

$$\Delta d \leq \frac{0,1}{2,20} = 0,0025 \text{ m} \quad \Delta r \leq \frac{0,1}{2,45} = 0,0010 \text{ m}$$

là đủ để  $\Delta S \leq 0,1$  m<sup>2</sup>

## 1.4 Sai số qui tròn, quan hệ giữa sai số và số chữ số đáng tin

Trong mục này ta xét các số được biểu diễn dưới dạng thập phân. Khi các số là gần đúng, vấn đề đặt ra là nên biểu diễn chúng với bao nhiêu chữ số? Thu gọn chúng như thế nào?

### 1.4.1 Chữ số có nghĩa

**Định nghĩa 3.** Trong biểu diễn theo cơ số  $b$  (trường hợp riêng là biểu diễn thập phân):

- các chữ số kể từ chữ số khác 0 đầu tiên tính từ trái sang phải gọi là các chữ số có nghĩa,
- các chữ số 0 bên trái là không có nghĩa

Nếu  $a$  được viết dưới dạng

$$a = \sum_{k=p}^n a_k 10^k \quad (1.12)$$

thì các chữ số 0 bên trái không có ở biểu diễn này, ý nghĩa của các chữ số 0 bên phải liên quan tới cách biểu diễn số gần đúng sẽ xét dưới đây.

**Ví dụ 1.10.** Số  $a = 03.4050$  thì chữ số 0 đầu không có nghĩa (người ta có thể điền để tránh viết thêm) còn các chữ số 3, 4, 0, 5, 0 là có nghĩa.

Số  $b = 0.034$  thì các chữ số 3, 4 là có nghĩa, hai chữ số 0 bên trái không có nghĩa vì nếu biểu diễn theo dạng 1.12 thì các chữ số này không cần đến.

### 1.4.2 Chữ số chắc

**Định nghĩa 4.** Xét  $a$  có biểu diễn 1.12 với sai số  $\Delta a$ .

- Nếu  $\Delta a \leq 0,5 \cdot 10^m$  thì  $a_k$  là chữ số chắc (đáng tin)  $\forall k \geq m$  (theo nghĩa hẹp dùng trong tính toán).
- Nếu  $0,5 \cdot 10^m \leq \Delta a \leq 10^m$  thì  $a_m$  là chắc theo nghĩa rộng.

**Ví dụ 1.11.**  $a = 21,473$  và  $\Delta a = 0,094 = 0,94 \cdot 10^{-1}$  thì:

- Các chữ số 2, 1 là chắc theo nghĩa hẹp.
- Chữ số 4 là chắc theo nghĩa rộng.
- Các chữ số 7, 3 là không đáng tin hay không chắc.

Khi cho số gần đúng ta có thể cho theo hai cách:

- Cách 1: Viết kèm với sai số tuyệt đối.

- *Cách 2:* Chỉ viết các chữ số chắc. Nếu ta có số gần đúng mà không cho sai số thì luôn ngầm hiểu các chữ số có nghĩa là các chữ số chắc. Như vậy các chữ số 0 ở bên phải cho ta biết nó là chữ số chắc.

Trong quá trình tính toán, người ta thường để lại vài chữ số không chắc và trong kết quả thì giữ lại các chữ số chắc theo nghĩa rộng.

### 1.4.3 Số thu gọn

Khi số  $a$  có nhiều chữ số không chắc hoặc có quá nhiều chữ số có nghĩa thì người ta thường thu gọn thành số  $\bar{a}$  có ít chữ số có nghĩa hơn. Nếu  $a$  có biểu diễn 1.12 và số thu gọn được giữ lại đến  $a_m$  ( $m > p$ ) thì  $\bar{a}$  có biểu diễn

$$\bar{a} = \sum_{k=m}^n b_k 10^k \quad (1.13)$$

nhờ bỏ đi các chữ số  $a_k$  ( $k < m$ ) theo quy tắc sau:

**Quy tắc.** *Quy tắc chữ số chẵn:* Giả sử  $a > 0$  và phần bỏ đi là  $\mu$ .

- Nếu  $\mu < 0,5 \cdot 10^m$  thì

$$\bar{a} = \sum_{k=m}^n a_k 10^k \quad (1.14)$$

*nghĩa là ta giữ nguyên các chữ số đến hàng  $m$  tính từ trái sang phải.*

- Nếu  $\mu > 0,5 \cdot 10^m$  thì

$$\bar{a} = \sum_{k=m}^n a_k 10^k + 10^m \quad (1.15)$$

- Nếu  $\mu = 0,5 \cdot 10^m$ , ta xét tiếp:
  - Nếu  $a_m$  chẵn, làm theo 1.14.
  - Nếu  $a_m$  lẻ, làm theo 1.15.

*Khi  $a < 0$  ta thu gọn giá trị tuyệt đối và giữ nguyên dấu.*

Khi thu gọn  $a$  thành  $\bar{a}$  ta có sai số thu gọn  $\Gamma_a \leq 0,5 \cdot 10^m$ . Để nó ít ảnh hưởng tới sai số tuyệt đối, ta thu gọn số và giữ lại một hoặc hai chữ số không chắc.

Nếu  $a$  có biểu diễn 1.12 và  $a_k$  chắc với  $k \geq m$  thì  $\Delta a \leq 10^m$  nên

$$\delta_a = \frac{\Delta a}{|a|} \leq \frac{10^m}{\sum_{k=m}^n a_k 10^k} = \frac{1}{\sum_{k=0}^n a_{k+m} 10^k}$$

Như vậy sai số tương đối của số gần đúng có thể ước lượng bởi nghịch đảo của số gồm các chữ số chắc của  $a$  không có dấu phẩy.

#### 1.4.4 Dấu phẩy động

Chúng ta biết rằng trong biểu diễn thập phân, mọi số thực được biểu diễn bởi một hữu hạn hoặc một dãy vô hạn các chữ số thập phân.

Bây giờ hầu hết các máy tính có hai cách biểu diễn số, được gọi là dấu phẩy tĩnh và dấu phẩy động.

Trong một biểu diễn *dấu phẩy tĩnh* tất cả các số được đưa ra với một số cố định các số thập phân sau dấu thập phân; ví dụ, số được đưa ra với 3 số thập phân là 62,358, 0,014, 1,000. Trong một văn bản chúng ta sẽ viết, nói, biểu diễn 3 số thập phân là biểu diễn *3D*.

Biểu diễn dấu phẩy tĩnh có ưu thế:

- Thuận tiện trong tính toán hàng ngày,
- và trong tính toán với các số gần đơn vị

Trong hệ thống *dấu phẩy động*, chúng ta viết, ví dụ,

$$0,6247 \cdot 10^3; 0,1735 \cdot 10^{-13}; -0,2000 \cdot 10^{-1}$$

hay đôi khi ta biểu diễn:

$$6,247 \cdot 10^2; 1,735 \cdot 10^{-14}; -2,000 \cdot 10^{-2}$$

Chúng ta thấy rằng trong biểu diễn này số các chữ số có nghĩa được giữ cố định, trong khi dấu phẩy là "động". Ở đây, một chữ số có nghĩa của một số  $c$  là chữ số bất kỳ của  $c$ , ngoại trừ chữ số 0 nằm ở bên trái chữ số khác 0 đầu tiên; các chữ số 0 này chỉ để xác định vị trí của dấu phẩy (như vậy, bất kỳ chữ 0 khác đều là chữ số có nghĩa của  $c$ ).

##### Ví dụ 1.12.

$$13\,600; 1,3600; 0,001\,360\,0$$

tất cả đều có 5 chữ số có nghĩa.

Trong văn bản chúng ta nói rút gọn một số đến 5 chữ số có nghĩa là *5S*.

Việc sử dụng số mũ cho phép chúng ta biểu diễn số rất lớn và rất nhỏ. Thật vậy, về mặt lý thuyết số khác 0 bất kỳ  $a$  có thể được viết như sau:

$$a = \pm m 10^n \mid 0.1 \leq |m| < 1, n \in \mathbb{Z}$$

**Quy tắc.** Để làm tròn số  $x$  đến  $k$  chữ số sau dấu phẩy, ta cộng vào  $x$  lượng  $0,5 \cdot 10^{-(k+1)}$  và bỏ đi từ chữ số thứ  $k+1$  sau dấu phẩy trở đi.

#### Số học IEEE

Trong máy tính hiện đại người ta dùng các số nhị phân, ở đây  $m$  được giới hạn bởi  $k$  chữ số nhị phân (ví dụ,  $k = 8$ ) và  $n$  cũng được giới hạn, ta có biểu diễn sau (chỉ biểu diễn được một số hữu hạn số):

$$\bar{a} = \bar{m} \cdot 2^n \mid \bar{m} = 0, d_1 d_2 \dots d_k, d_1 > 0 \quad (1.16)$$

Các số  $\bar{a}$  ở đây được gọi là số máy nhị phân  $k$  chữ số. Phần sau dấu phẩy  $m$  (được gọi là *mantissa*), biểu diễn các chữ số có nghĩa của  $\bar{a}$ ,  $n$  được gọi là lũy thừa của  $\bar{a}$ .