SIMON FRASER UNIVERSITY
FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS

M. Alamgir Hossain
John M. Stockie

Clicker Question Bank for Numerical Analysis (Version –)

1. Introduction

Q1–1[1]. Select the best definition for “numerical analysis”:

Q1–2[1]. the study of round-off errors

Q1–3[1]. the study of algorithms for computing approximate solutions to problems from continuous mathematics

Q1–4[1]. the study of quantitative approximations to the solutions of mathematical problems including consideration of and bounds for the errors involved

Q1–5[1]. the branch of mathematics that deals with the development and use of numerical methods for solving problems

Q1–6[1]. the branch of mathematics dealing with methods for obtaining approximate numerical solutions of mathematical problems *Answer: (B). All 5 definitions are valid in some sense since they reflect some aspect of the field (most are pulled off the internet). But my favourite definition is (B) because it contains three very important keywords underlined below:*

the study of algorithms for computing approximate solutions to problems from continuous mathematics
[algorithms ⟺ computing, approximate ⟺ floating point arithmetic, continuous ⟺ solutions are smooth f’ns]

JMS

1a. Floating Point Arithmetic and Error

Q1a–1[2]. How many significant digits does the floating point number 0.03140×10^3 have?

Q1a–2[2]. 6

Q1a–3[2]. 5

Q1a–4[2]. 4

Q1a–5[2]. 3 *Answer: (C).*

Q1a–6[3]. Suppose that a hypothetical binary computer stores floating point numbers in 16-bit words as shown:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
s	exp			mantissa											

Bit 1 is used for the sign of the number, bit 2 for the sign of the exponent, bits 3-4 for the magnitude of the exponent, and the remaining twelve bits for the magnitude of the mantissa. What is machine epsilon for this computer?

Q1a–7[3]. 2^{-16}

Q1a–8[3]. 2^{-12}

Q1a–9[3]. 2^{-8}

Q1a–10^3. 2^{-4} *Answer: (B). Assume that rounding is used and recall that ε_M is essentially the same as unit round-off error $u = B^{1-t}$, where $B = 2$ is the base and t is the number of significant digits. The number of digits stored in the mantissa is $t = 12$ and so $\varepsilon_M \approx 2^{1-12} = 2^{-12}$.*

Q1a–11^4. You are working with a hypothetical binary computer that stores integers as unsigned 4-bit words. What is the largest non-negative integer that can be represented on this computer?

Q1a–12^4. 64

Q1a–13^4. 63

Q1a–14^4. 31

Q1a–15^4. 15

Q1a–16^4. 7 *Answer: (D).* $(1111)_2 = 1 \times 2^0 + 1 \times 2^1 + 1 \times 2^2 + 1 \times 2^3 = 15$.

Q1a–17^5. In 1958 the Russians developed a ternary (base-3) computer called *Setun*, after the Setun River that flows near Moscow State University where it was built. In contrast with today's binary computers, this machine used "trits" (ternary bits) whose three possible states can be represented as $\{0, 1, 2\}$. Its floating-point number system was based on 27-trit numbers, with 9 trits reserved for the exponent and 18 for the mantissa. What was the value of machine epsilon ϵ_M for the *Setun*?

Q1a–18^5. 3^{-19}

Q1a–19^5. 3^{-18}

Q1a–20^5. 3^{-9}

Q1a–21^5. $\frac{1}{3} \cdot 2^{-18}$ *Answer: (B).*
Apply the formula $\epsilon_M = B^{-t}$ from the notes, where $B = 3$ is the base and $t = 18$ is the number digits in the mantissa. You may have noticed that I didn't mention a "sign trit" for the mantissa. In actual fact, the floating-point representation on Setun *was more complicated than this and the sign of a number came from interpreting one specific trit as $\{-1, 0, +1\}$ instead.*
plus info from `http://homepage.divms.uiowa.edu/~jones/ternary/numbers.shtml`

Setun – Moscow State University ,

Q1a–22^6. In Canada, the total for any store purchase paid in cash is rounded to the nearest 5 cents, whereas no rounding is done if the payment is by credit/debit card. Suppose that when you return home after purchasing your groceries with cash, you notice that your bill was \$10.07. What is the absolute error in your actual cash payment?

Q1a–23^6. 2 cents

Q1a–24^6. 3 cents

Q1a–25^6. 4 cents

Q1a–26^6. 5 cents *Answer: (A).*

Q1a–27^7. Let $\hat{x}$ be some approximation of x. Which of the following error definitions is correct?

Q1a–28^7. absolute error $= |x - \hat{x}|$, relative error $= \frac{|x-\hat{x}|}{|x|}$

Q1a–29^7. absolute error $= \frac{|x-\hat{x}|}{|x|}$, relative error $= |x - \hat{x}|$

Q1a–30^7. absolute error $= \frac{|x-\hat{x}|}{|x|}, x \neq 0$, relative error $= |x - \hat{x}|$

Q1a–31^7. absolute error $= |x - \hat{x}|$, relative error $= \frac{|x-\hat{x}|}{|x|}, x \neq 0$ *Answer: (D).*