INDIAN INSTITUTE OF TECHNOLOGY TIRUPATI
Department of Mathematics and Statistics
MA633L - Numerical Analysis
Problems on Numerical Errors

Note: Usual Notations are used. Questions may have some typos or grammatical mistakes as proofreading is not done for this.

---

1. Prove or disprove the following:

   (a) $\sum_{k=0}^{n} x^k = \frac{1}{1-x} + o(x^n) \quad \text{as} \quad x \to 0$

   (b) $\sum_{k=0}^{n} a_k x^k = O(x^n), x \geq 1$

   (c) $\cos x - 1 + \frac{x^2}{2} = O(x^k) \quad \text{as} \quad x \to 0$

   (d) $\ln x = o\left(\frac{1}{x^r}\right) \quad \text{as} \quad x \to 0, x \in (0, \infty), r > 0$

   (e) $\ln x = o(x^r) \quad \text{as} \quad x \to \infty, x \in (0, \infty), r > 0$

   (f) $3 \log_8 n + \log_2(\log_2(\log_2 n)) = O(\log n)$

   (g) $0.3n + 5n^{1.5} + 2.5n^{1.75} = O(n^{1.75})$

   (h) $3n^2 + 10n \log n = O(n \log n)$

   (i) $10\sqrt{n} + \log n = O(n)$

   (j) $n^3 + 20n + 1 = O(n^3)$

   (k) $n^2 + 42n + 7 = O(n^2)$

   (l) $n^3 + 20n + 1 = \Omega(n^2)$

   (m) $n! = \omega(2^n)$

   (n) $\log(n!) = \Theta(n \log n)$

   (o) $2^{2n} = \Theta(2^n)$

   (p) $2^{n+1} = \Theta(2^n)$

   (q) $(n + a)^b = \Theta(n^b)$ for all real a and b

2. Find real numbers, a, b and c, such that

$$\ln\left(\frac{\sin x}{x}\right) = ax^2 + bx^4 + cx^6 + o(x^8) \quad \text{as} \quad x \to 0, x \in (-\pi, \pi)$$

3. Prove that

$$O(n) \subset O(\log n) \subset O(\sqrt{n}) \subset O(n) \subset O(n \log n) \subset O(n^2) \subset O(n^3) \subset O(3^n) \subset O(n^n)$$

4. Prove that for any constants $a, b, c > 0$,

$$O(a) \subset O(\log n) \subset O(n^b) \subset O(c^n)$$

5. Prove that for any $a < b$.

$$O(n^a) \subset O(n^b)$$

6. Prove that $x^a = o(x^b)$ for all nonnegative constants $a < b$

7. Prove that $logx = o(x^a)$ for all $a > 0$

8. Prove that $x^b = o(a^b)$ for any $a, b \in \mathbb{R}$ with $a > 1$

9. Verify whether the following are true or not.

   (a)

$$\frac{5}{n} + e^{-n} = O\left(\frac{1}{n}\right)$$

   (b)

$$e^{-n} = o\left(\frac{1}{n^2}\right)$$

   (c)

$$\ln 2 - \sum_{k=1}^{n-1} (-1)^{k-1} \frac{1}{k} = O\left(\frac{1}{n}\right)$$

   (d)

$$e^x - \sum_{k=0}^{n-1} x^k \frac{1}{k!} = O\left(\frac{1}{n!}\right) \qquad (|x| \leq 1)$$

10. Prove that if $f = \Omega(g)$ then f is not in $o(g)$

11. Suppose that $0 < q < p$ and that $\alpha_n = \alpha + O(n^{-p})$. Show that $\alpha_n = \alpha + O(n^{-q})$.

12. Suppose that $0 < q < p$ and that $F(h) = L + O(h^p)$. Show that $F(h) = \alpha + O(h^q)$.

13. Suppose that as $x \to 0$,

$$F_1(x) = L_1 + O(x^\alpha), \qquad F_2(x) = L_2 + O(x^\beta)$$

   Let c_1 and c_2 be nonzero constants and define

$$F(x) = c_1 F_1(x) + c_2 F_2(x), \qquad G(x) = F_1(c_1 x) + F_2(c_2)$$

   Show that if $\gamma = \min\{\alpha, \beta\}$ then

   (a) $F(x) = c_1 L_1 + c_2 L_2 + O(x^\gamma)$

   (b) $G(x) = L_1 + L_2 + O(x^\gamma)$

14. Classify $f = O(g)$ or $f = \Omega(g)$ or $f = \Theta(g)$ for the following

(a) $f(n) = n\log(n^3), g = n\log(n)$

(b) $f(n) = 2^{2n}, g(n) = 3^n$

(c) $f(n) = \sum_{i=1}^{n} \log_i, g(n) = n\log(n)$

(d)

$$f(n) = n, g(n) = \begin{cases} 1 & \text{if } n \text{is odd} \\ n^2 & \text{if } n \text{is even} \end{cases}$$

15. Prove that

$$10\ln(n) + 5(\ln(n))^3 + 7n + 3n^2 + 6n^3 = O(n^3)$$

16. Prove that $O(f).O(g) = O(fg)$

17. Prove that $O(f) + O(g) = O(\max f, g)$

18. Show that $\sin x = x + O(x^3)$ as $x \to 0$

19. Suppose $\sin x = x + O(x^3)$, prove that $\sin^2 x = x^2 + O(x^4)$ as $x \to 0$

20. Suppose

$$f(x) = \frac{1}{x} - \frac{1}{x^3} + \frac{1}{x^5}$$

then

$$f(x) = \frac{1}{x} + O(x^{-3}), x \to \infty$$

and

$$f(x) = \frac{1}{x^5} + O(x^{-3}), x \to 0$$

21. Show that $\cos(h) + \frac{h^2}{2} = 1 + O(h^4)$ as $h \to 0$

22. Show that $x^2 + 1 = O(x^2)$ as $x \to \infty$

23. Show that $x^2 + 1 = O(1)$ as $x \to 0$

24. Show that $6x^4 - 2x^2 + 5 = O(x^4)$ as $x \to \infty$

25. Show that $6x^4 - 2x^2 + 5 = O(1)$ as $x \to 1$

26. Show that $3.5n^2 + 4n + 36 = \Theta(n^2)$

27. Show that $3n^2 + 6n + 7 = \Theta(n^2)$

28. Suppose that $a > 1$ and $b \neq 0$ are constants with $|b| < a$. Prove that $a^n + b^n = \Theta(a^n)$

29. Prove or disprove $3n^2 + 6n + 7 = \Theta(n^3)$

30. Prove or disprove $3n^2 + 6n + 7 = o(n^3)$

31. Prove that $1 = o(\log n)$

32. Prove that $\log n = o(n^\varepsilon)$ for any $\varepsilon > 0$

33. Prove that $\log^k n = o(n^\varepsilon)$ for any $\varepsilon > 0, k > 0$

Order and Rate of Convergence

34. Prove that $\frac{1}{2^n}$ converges linearly with rate of convergence $\frac{1}{2}$ to 0

35. Prove that $\frac{1}{2^{2^n}}$ converges superlinearly or quadratically to 0

36. Prove that $\frac{1}{2^{n^2}}$ converges superlinearly to 0

37. Prove that $\frac{1}{n+1}$ converges logarithmically or sublinearly to 0

38. Prove that $\frac{1}{n^k}, k > 0$ converges linearly to 0

39. Prove that
$$\frac{n^\alpha}{n+1} \to \begin{cases} 0 & \text{if } \alpha < 1 \\ 1 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha > 1 \end{cases}$$

40. Prove that
$$\frac{2^{\alpha n}}{2^{n+1}} \to \begin{cases} 0 & \text{if } \alpha < 1 \\ 1/2 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha > 1 \end{cases}$$

41. Prove that $1 + \frac{1}{2^n}$ converges linearly to 1

42. Prove that $1 + \frac{1}{n^n}$ converges superlinearly to 1.

43. Prove that $1 + \frac{1}{n^{2^n}}$ converges quadratically to 1

44. Find the order of convergence of the sequence $x_n = \frac{1}{n^2}$

45. Find the order of convergence of the sequence $x_n = e^{-2^n}$

46. Find the order of convergence of the sequence $x_n = 2^{-3^n}$

47. Find the order of convergence of the sequence $x_n = 10^{-n/2}$

48. Find the order of convergence of the sequence $x_n = 10^{-2^n}$

49. Find the order of convergence of the sequence
$$x_{n+1} = \frac{x_n^3 + 6x_n}{3x_n^2 + 2}$$
where $x_0 = 1$.

50. Find the order of convergence of the sequence
$$x_{n+1} = \frac{x_n^3 + 3ax_n}{3x_n^2 + a}$$
where $x_0 = 1$.

51. Let

$$x_n = \begin{cases} \dfrac{1}{2^n} & \text{if } n \text{is even} \\ \dfrac{1}{2^n + 1} & \text{if } n \text{is odd} \end{cases}$$

Can we find the order of convergence of this sequence?

52. Let

$$x_n = \begin{cases} \dfrac{1}{\log n} & \text{if } n \text{is even} \\ \dfrac{1}{n} & \text{if } n \text{is odd} \end{cases}$$

Can we find the order of convergence of this sequence?

53. What is the order of convergence of $x_n = \dfrac{1}{n}$? What is the order of convergence of $y_n = \dfrac{1}{\log n}$? (Answer: 1), Is it true for the sequence z_n where

54. Let

$$x_n = \begin{cases} y_n & \text{if } n \text{is even} \\ x_n & \text{if } n \text{is odd} \end{cases}$$

55. Prove that if $\lim_{n\to\infty} \dfrac{x_{n+1}}{x_n^\alpha}$ is finite, then $\lim_{n\to\infty} \dfrac{x_{n+1}}{x_n} = \alpha$. Is the converse true? If so, prove it. If not, give an example.

56. Prove that if $\lim_{n\to\infty} \dfrac{x_{n+1}}{x_n} = \alpha$ then $\lim_{n\to\infty} \sqrt[n]{|\log x_n|} = \alpha$. Is the converse true? If so, prove it. If not, give an example.

57. Find the rate and order of convergence of the following sequence as $n \to \infty$

   (a) $\sin \dfrac{1}{n}$

   (b) $\sin \dfrac{1}{n^2}$

   (c) $\left(\sin \dfrac{1}{n}\right)^2$

   (d) $\ln(n+1) - \ln(n)$

58. Find the rate and order of convergence of the following sequence as $h \to 0$

   (a) $\dfrac{\sin h}{h}$

   (b) $\dfrac{1 - \cos h}{h}$

   (c) $\dfrac{\sin h - h \cos h}{h}$

   (d) $\dfrac{1 - e^h}{h}$

Floating Point

59. List all floating-point numbers that can be represented in the form

$$c = \pm(0.b_1b_2)_2 \times 2^{\pm k}$$

where $b_1, b_2, b_3, k \in \{0, 1\}$

60. List all floating-point numbers that can be represented in the form

$$c = \pm(0.b_1b_2)_2 \times 2^{\pm k}$$

where $b_1 = 1, b_2, k \in \{0, 1\}$

61. List all floating-point numbers that can be represented in the form

$$c = \pm(0.b_1b_2)_2 \times 2^{k}$$

where $b_1, b_2, \in \{0, 1\}, k \in \{-1, 0\}$

62. List all floating-point numbers that can be represented in the form

$$c = \pm(0.b_1b_2)_2 \times 2^{k}$$

where $b_1, b_2, \in \{0, 1\}, k \in \{-1, 1\}$

63. Suppose you have a register with 16 bits of which 1 bit is allotted for sign bit, 4 bits are allotted for exponents and remaining bits are allotted for fractions, what will be the machine epsilon, smallest number and largest number?

64. Suppose you have a register with 24 bits of which 1 bit is allotted for sign bit, 6 bits are allotted for exponents and remaining bits are allotted for fractions, what will be the machine epsilon, smallest number and largest number?

65. Suppose you have a register with 32 bits of which 1 bit is allotted for sign bit, 12 bits are allotted for exponents and remaining bits are allotted for fractions, what will be the machine epsilon, smallest number and largest number?

Numerical Errors

66. Suppose that you have the task of measuring the lengths of a bridge and a rivet and come up with 9999 and 9 cm, respectively. If the true values are 10,000 and 10 cm, respectively, compute (a) E_t, (b) E_{tabs} and (c) ϵ_t for each case.

67. Evaluate e^{-5} using two approaches

$$e^{-x} = \sum_{k=0}^{\infty} \frac{(-x)^k}{k!}$$

and

$$e^{-x} = \frac{1}{e^x} = 1/\sum_{k=0}^{\infty} \frac{x^k}{k!}$$

and compare with true value of 6.737947×10^{-3}. Use first 20 terms to evaluate each series. Compute E_{tabs} and ε_t.

68. Compute ε_t for the following a and $\tilde{a}$

    (a) $a = \pi, \tilde{a} = 22/7$

    (b) $a = \pi, \tilde{a} = 3.1416$

    (c) $a = e, \tilde{a} = 2.718$

    (d) $a = \sqrt{2}, \tilde{a} = 1.414$

    (e) $a = e^{10}, \tilde{a} = 22000$

    (f) $a = 10^{\pi}, \tilde{a} = 1400$

    (g) $a = 8!, \tilde{a} = 39900$

    (h) $a = 9!, \tilde{a} = \sqrt{18\pi}\left(\frac{9}{e}\right)^9$

69. Evaluate the polynomial

$$y = x^3 - 5x^2 + 6x + 0.55$$

at $x = 1.37$. Use 3-digit arithmetic with chopping. Compute ϵ_t

70. Evaluate the polynomial

$$y = x(x(x - 5) + 6) + 0.55$$

at $x = 1.37$. Use 3-digit arithmetic with round off. Compute ϵ_t

71. How can accurate values of the function $f(x) = x - \sin x$ be computed near $x = 0$?

72. How can accurate values of the function $f(x) = \sqrt{x^2 + 4} - 2$ be computed near $x = 0$?

73. How can accurate values of the function $f(x) = e^x - e^{-2x}$ be computed near $x = 0$?

74. **Range Reduction:** Using the periodicity of sin function, we need to know only values of $\sin x$ in the interval $[0, 2\pi)$ for any arbitrary x.

75. If $y = \cos^2 x - \sin^2 x$ is evaluated at $x = \pi/4$, there is a loss of significance. How can we avoid this?

76. If $y = \ln(x) - 1$ is evaluated at x near e, there is a loss of significance. How can we avoid this?

77. How can the value of the function $f(x) = \sqrt{x^4 + 4} - 2$ be computed accurately when x is small?

78. How can the value of the function $f(x) = \frac{1}{x}(\sinh x - \tanh x)$ be computed accurately when x is small?

79. How can the value of the function $f(x) = \sin x + \cos x - 1$ be computed accurately when x is near zero?

80. How can the value of the function

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

be computed accurately when $|x| < 0.5$?

81. How can the value of the function

$$f(x) = |\sin x - x|$$

be computed accurately when $|x| < 0.1$?

82. How can the value of the function $f(x) = \tan^{-1} x - x$ be computed accurately when x is near zero?

83. How can the value of the function $f(x) = \tan x - x$ be computed accurately when x is near zero?

84. How can the value of the function

$$f(x) = \frac{e^{2x} - 1}{2x}$$

be computed accurately when x is near zero?

85. For some values of x, the function $f(x) = \sqrt{x^2 + 1} - x$ cannot be computed accurately. Explain and find a way around the difficulty.

86. The hyperbolic function is defined by $\sinh x = \frac{e^x - e^{-x}}{2}$. What drawback could there be in using this formula to obtain the values of the function. How can values of $\sinh x$ be computed for floating point precision when $|x| \leq 0.5$

87. Find the first two non-zero terms in the expansion about zero for the function

$$f(x) = \frac{\tan x - \sin x}{x - \sqrt{1 + x^2}}$$

. Find the value of $f(0.0125)$

88. Find a way to calculate accurate values for

$$f(x) = \frac{\sqrt{x^2 + 1} - 1}{x^2} - \frac{x^2 \sin x}{x - \tan x}$$

. Determine $\lim_{x \to 0} f(x)$.

89. Let $f(x) = e^x - x - 1$. Calculate $f(10^{-2})$ with five significant figures.

90. What difficulty could the following assignment statement cause in computer $y = 1 - \sin x$. Circumvent it without resorting to a Taylor series if possible.

91. Without using Taylor series, how could the function

$$f(x) = \frac{\sin x}{x - \sqrt{x^2 - 1}}$$

be computed to avoid loss of significance?

92. How can values of the function $f(x) = \sqrt{x+2} - \sqrt{x}$ be computed accurately when x is large?

   (a) Discuss how to evaluate $\sin(12532.14)$ by subtracting multiples of 2π.

   (b) Show that $\sin(12532.14) = \sin(3.47)$ if we retain only two decimal digits of accuracy.

   (c) For $\sin x$, how many binary bits of significance are lost in range reduction to the interval $[0, 2\pi)$?

93. The quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

will cause problem when the quadratic equation $x^2 - 10^5 x + 1 = 0$ is solved with a machine that carries only eight decimal digits. How can you avoid this?

94. Find ways to compute these functions without serious loss of significant figures

   (a) $e^x - \sin x - \cos x$

   (b) $\ln x - 1$

   (c) $\log(x) - \log(1/x)$

   (d) $x^{-2}(\sin x - e^x + 1)$

   (e) $x - \tanh^{-1} x$

95. How will you compute $\tan(10^{100})$?

96. Prove the Loss of Precision Theorem stated in Lecture-7

97. Let x and y be two normalized binary floating point machine epsilon. Assume that

$$x = q \times 2^n, y = r \times 2^{n-1}, 0.5 \le r, q < 1, 2q - 1 \ge r$$

How much loss of significance occurs in $x - y$?

98. Let x and y be two normalized binary floating point machine epsilon. Assume that

$$x = q \times 2^n, y = r \times 2^{n-1}, 0.5 \le r, q < 1, 2q - 1 < r$$

How much loss of significance occurs in $x - y$?

99. If x is a machine epsilon on a 32-bit computer that satisfies the inequality $x > \pi \times 2^{25}$, then show that $\sin x$ will be computed with no significant digits.

100. Let x and y be two positive normalized binary floating point machine epsilon in a 32-bit system. Assume that

$$x = q \times 2^m, y = r \times 2^n, 0.5 \leq r, q < 1$$

Show that if $n = m$, then at least 1 bit of significance is lost in $x - y$.