

CYBR 304
Spring 2024
Review for First Examination

1. A floating point number of the type UNK has the form $\pm b_0.b_1b_2b_3b_4 \times 2^e$, where b_0 through b_4 are bits and the exponent e is an integer such that $-4 \leq e \leq 3$. The bit b_0 is not stored—it defaults to 1.
 - (a) Find the *largest* positive number of the type UNK.
 - (b) Find the *smallest positive* number of the type UNK.

- 5 2. The standard deviation σ of a list of two or more real numbers x_1, x_2, \dots, x_n is

$$\sigma = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n}},$$

where $\bar{x} = \frac{\sum_{k=1}^n x_k}{n}$. An alternative formula for σ is

$$\sigma = \sqrt{\frac{\sum_{k=1}^n x_k^2 - (\sum_{k=1}^n x_k)^2 / n}{n}}.$$

For floating point evaluation, which formula should you prefer? Explain.

- 5 3. Let a and b be numbers and let \oplus be floating point addition. Show that

$$|(a \oplus b) - (a + b)| \leq \varepsilon_m (|a| + |b| + |a + b|) + \varepsilon_m^2 (|a| + |b|).$$

- 5 4. The sequence $k \mapsto 8 + \frac{1}{8^k}$ converges to 5. Show that the sequence converges to 5 *linearly*.

5. Define a sequence F , by $F_{k+1} = \begin{cases} 1 & \text{if } k = -1 \\ 1 - \frac{1}{2}F_k & \text{otherwise} \end{cases}$.

- 5 (a) Assuming F converges, show that F converges to $2/3$.

- 5 (b) Again assuming F converges, show that F converges linearly.

- 5 6. Explain how given a term in the Newton sequence, how is the next term determined? Draw a labeled graph that helps explain.

7. The function $G(x) = (x^2 - 3x + 4)/2$ has a fixed point at 1. For an initial value that is close to 1, do you expect the fixed point sequence for G to converge or to diverge? Why.
8. The function $H(x) = x^2 - 4x + 4$ has a fixed point at 1. For an initial value that is close to 1, do you expect the fixed point sequence for H to converge or to diverge? Why.

9. For real numbers x and y , you need to compute $\sqrt{\sqrt{y^2 + x^2} + x}$. An alternative formula for this expression is $\frac{|y|}{\sqrt{\sqrt{y^2 + x^2} - x}}$.

- (a) When $x < 0$, from the point of view of accuracy, which expression should you use? Why?
- (b) When $x > 0$, from the point of view of accuracy, which expression should you use? Why?

10. Understand the function fl (See page 36) and its property

$$|\text{fl}(x) - x| \leq |x|\varepsilon_m$$

11. Know what we mean by subtractive cancellation. You should be able to give an example of subtractive cancellation.

Example: Suppose $x = 3.14 \pm 0.01$ and $y = 3.12 \pm 0.01$. Thus x and y are known accurate to two decimal places. But the difference of these numbers $x - y$ is a number in the interval $(0, 0.2)$. In \pm language, we have $x - y = 0.1 \pm 0.1$. Thus although x and y are known to be accurate to two decimal places, the difference $x - y$ is known accurately to zero digits. This is known as subtractive cancellation.

12. Know the rule that for floating point addition of two floats that $x \oplus y = (x + y)(1 + \varepsilon)$, where $|\varepsilon| \leq \varepsilon_m$. Using this rule, be able to show that for floats a and b , we have

$$|a \oplus b - (a + b)| \leq \varepsilon_m |a + b|.$$

13. Understand the derivation and the significance of the result

$$\left| \frac{x + y - (\text{fl}(x) + \text{fl}(y))}{x + y} \right| \leq \varepsilon_m \frac{|x| + |y|}{|x + y|}.$$

14. Know how the Newton sequence is defined. Know how to determine its next term from the previous term.

15. Know what it means for a function to have a fixed point.

16. Know how the fixed point sequence is defined.

17. Know how to prove that if a fixed point sequence converges, that it converges to a fixed point of the function.

18. Know what it means for a sequence to converge linearly (page 54).

19. Know what it means for a sequence to converge quadratically (page 54).

20. Find the *fixed points* of the function $x \mapsto \frac{10(1-x)(2-x)}{3}$.

21. The sequence $k \mapsto 5 + \frac{1}{8^k}$ converges to 5. Show that the sequence converges to 5 *linearly*.
22. Given that a sequence F converges to 1, what does it mean to say that F converges to 1 *linearly*?
23. Show that the sequence $k \mapsto 1 + \left(\frac{1}{2}\right)^k$ converges to 1 *linearly*.
24. Although $\sqrt{1+x} - 1 = \frac{x}{\sqrt{1+x} + 1}$ is an identity, for floating point evaluation one formula should be preferred over the other when x is near zero. Which formula is better when x is near zero and why?