

Assignment 1

1. In the classes, we discussed three forms of floating number representations as given below, (1) Standard/General Form, (2) Normalized Form, (3) Denormalized Form. Now, Let's take, $\beta = 2$, $m = 4$, $e_{\min} = -5$ and $e_{\max} = 2$. Based on these, answer the following:

(a) (2 marks) What are the maximum numbers that can be stored in the system by these three forms defined above? (express your answer in decimal values)

(b) (2 marks) What are the non-negative minimum numbers that can be stored in the system by the three forms defined above? (express your answer in decimal values)

(c) (2 marks) How many numbers (both non-negative and negative) can be represented in the above mentioned system using the normalized form? Explain your answer.

(d) (4 marks) Find all the decimal numbers for $e = -1$ and $e = 0$ in normalized form, plot them on a real line and prove that all the numbers are not equally spaced. Write the equally spaced sets for the number line you drew. How many equally spaced sets are possible for the mentioned system?

2. Let $\beta = 2$, $m = 7$, $e_{\min} = -4$ and $e_{\max} = 8$. Answer the following questions:

(a) (4 marks) Compute the minimum of $|x|$ for General and Denormalized form.

(b) (3 marks) Compute the Machine Epsilon value for the General and Normalized form.

(c) (3 marks) If we change the value of e_{\min} to -7 then how will it affect the value of maximum scale invariant error for the case of Denormalized form? Explain your answer.

3. Consider the quadratic equation, $5x^2 - 70x + 4 = 0$. Consider up to 5 significant figures when performing calculations.

(a) (2 marks) Calculate the roots of the quadratic equation.

(b) (4 marks) Find out where the loss of significance occurs when you calculate the roots?

(c) (4 marks) Evaluate the correct roots such that loss of significance does not occur.