MS211 - Turma Z - Teste 1 - Entrega: 15/09/2022

Nome: RA:

Escreva a resolução deste teste à mão, exceto os itens 1 (e),(f) e 2 (c),(f) e entregue-a na sala de aula CB03. As resoluções dos itens 1 (e),(f) e 2 (c),(f) devem ser entregues pelo Google Classroom.

1. O formato de ponto flutuante da IEEE chamado binary64 é um formato com 64 bits que usa a base 2.

   - 1 bit é reservado para o sinal;
   - 11 bits correspondem ao expoente no formato "biased";
   - 53 bits correspondem à mantissa. De fato, somente 52 bits precisam ser armazendados explicitamente porque sabemos que o primeiro dígito a_1 da mantissa é 1 se o número x representado difere de 0 ($x = 0$ é tratado como um caso especial).

   Este formato corresponde a um formato de ponto flutuante que conheçemos da aula, mas com algumas pequenas modificações devido a certas considerações técnicas.

   (a) Para obter o formato "biased"do expoente, faça o seguinte: Use os 11 bits disponíveis para representar um intervalo da forma $[0, L]$ de números inteiros. Calcule $c = \frac{L-1}{2}$ e detemine o intervalo $[-c + 2, c + 1] \subset \mathbb{Z}$. Este intervalo é o intervalo dos expoentes permissiveis.

   (b) Determine o maior número M que pode ser representado no formato binary64.

   (c) Determine o menor número positivo $N > M$ que não pode ser mais representado no formato binary64. Justifique a sua resposta.

   (d) Determine o menor número positivo m que pode ser representado no formato binary64 segundo as informações providenciadas acima. (A representação "subnormal"permite representar números menores ainda - porém, com perda de precisão.)

   (e) Confira usando Matlab, Octave ou Python que o número M obtido no item (b) pode ser e N não pode ser representado. Envie um printout no Google Classroom.

   (f) Confira usando Matlab, Octave ou Python que o número m obtido no item (c) pode ser representado e que é possivel representar um número menor ainda devido à representação subnormal. Envie um printout no Google Classroom.

Veja a Questão 2 na próxima página!

2. Consider a seguinte matriz A:

$$A = \begin{pmatrix} 0.5 & 1.1 & 3.1 \\ 1 & 4.5 & 0.36 \\ 5 & 0.96 & 6.5 \end{pmatrix}.$$

(a) Calcule a fatoração LU SEM pivoteamento de A numa maquina que trabalha *no sistema* $F(10, 3, [-9, 9])$ *com arredondamento*. Para tanto, escreve as matrizes $R^{(i)}$ para $i = 1, 2$. Sejam L_S e U_S as matrizes triangulares inferior e superior, respectivamente, calculadas neste processo. Exibe estas matrizes L_S e U_S.

(b) Gire a folha por 90 graus e utilize uma folha inteira para exibir o esquema geral utilizado para executar uma subtração $x - y$ na aritmética de ponto flutuante. Utilize este esquema para exibir os passos necessárias nas subtrações utilizadas para obter $R^{(1)}$ e $R^{(2)}$. (Veja a próxima página.)

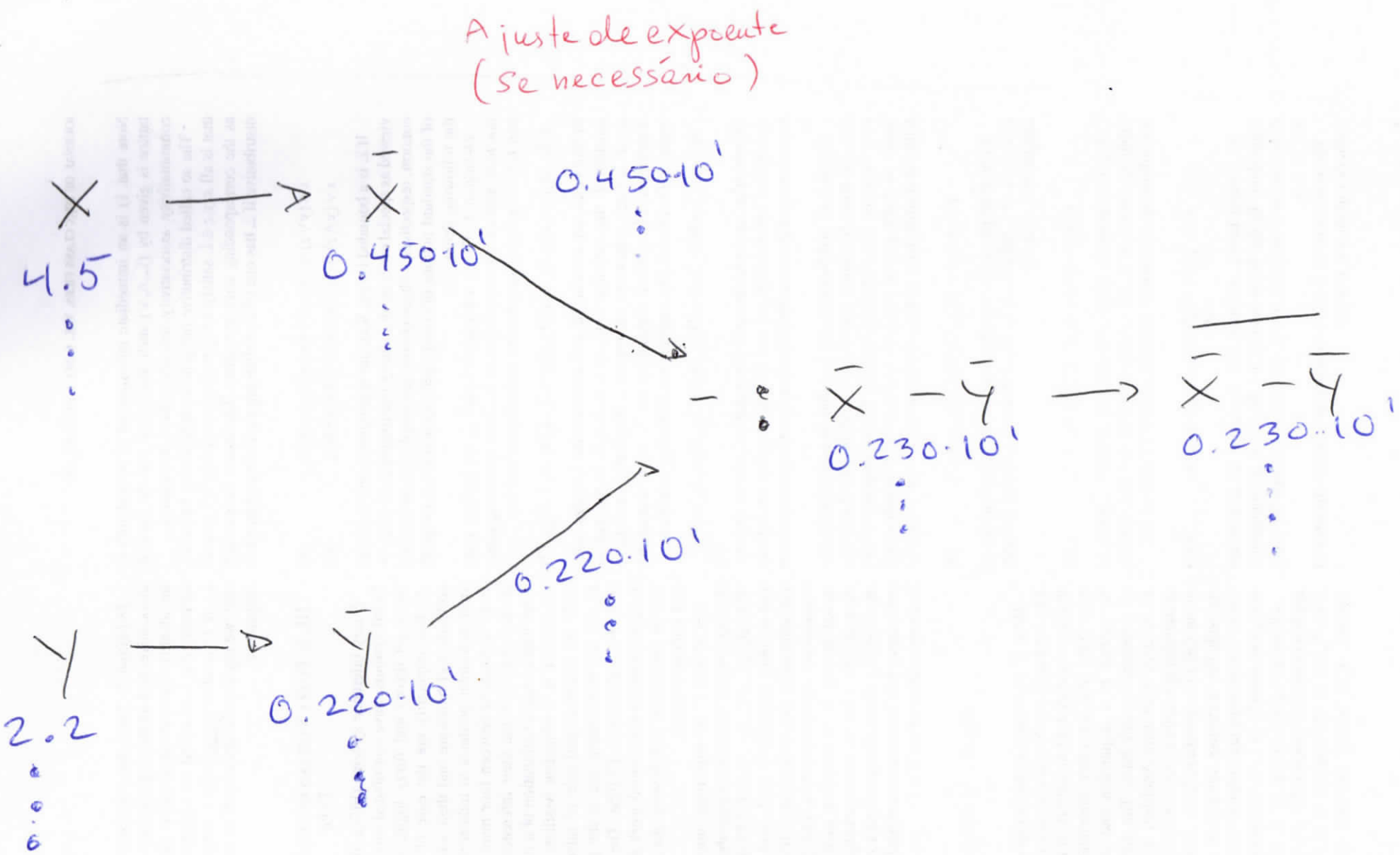
(c) Considere as matrizes L_S e U_S determinadas no item (a). Utilize um software da sua escolha para calcular $L_S \cdot U_S$ e a distância entre A e $L_S \cdot U_S$, dada pela norma infinidade de $A - L_S \cdot U_S$ exatamente. Por exemplo, em Matlab/Octave a norma infinidade de $A - B$ é dada pelo comando "norm(A-B,inf)".

(d) Calcule a fatoração LU COM pivoteamento de A numa maquina que trabalha *no sistema* $F(10, 3, [-9, 9])$ *com arredondamento*. Escreve as matrizes $R^{(i)}$ e - se for necessário - $R^{(1)'}$ além dos vetores p. Sejam L_C e U_C as matrizes triangulares inferior e superior, respectivamente, calculadas neste processo. No final, exibe as matrizes L_C, U_C e P obtidos.

(e) Gire a folha por 90 graus e utilize uma folha inteira para exibir o esquema geral utilizado para executar uma subtração $x - y$ na aritmética de ponto flutuante. Utilize este esquema para exibir os passos necessárias nas subtrações utilizadas.

(f) Considere L_C, U_C e P determinadas no item (d). Utilize um software da sua escolha para calcular $L_C U_C$ e a distância entre PA e $L_C U_C$, dada pela norma infinidade de $PA - L_C U_C$ exatamente.

(g) Compare as distâncias entre A e $L_S U_S$ e entre PA e $L_C U_C$ calculadas nos itens (c) e (f). Explique o que aconteceu.


Ajuste de expoente
(se necessário)
X
4.5
$\bar{X}$
$0.450 \cdot 10^1$
$0.450 \cdot 10^1$
Y
2.2
$\bar{Y}$
$0.220 \cdot 10^1$
$0.220 \cdot 10^1$
$\bar{X} - \bar{Y}$
$0.230 \cdot 10^1$
$\overline{\bar{X} - \bar{Y}}$
$0.230 \cdot 10^1$