

**M353 Hw 1** (S. Zhang) 0.2, 0.3, 1.1.

1. (0.2:4d) Convert binary number  $110.\overline{10}_2$  to base 10.

• **ans:**

$$110_2 = 2^2 + 2^1 + 0 \cdot 2^0 = 4 + 2 = 6$$

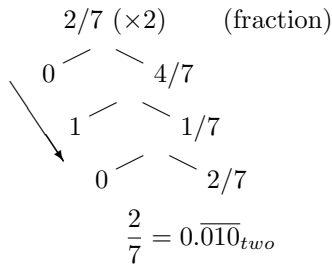
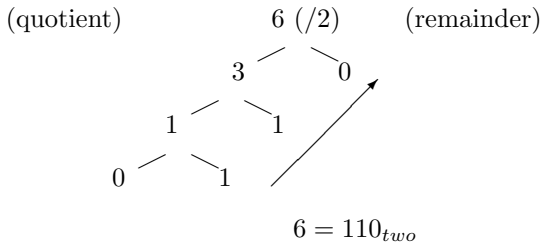
$$\begin{aligned} \overline{10}_2 &= \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \dots \\ &= \frac{1}{2}(1 + (1/4) + (1/4)^2 + \dots) \\ &= \frac{1}{2} \cdot \frac{1}{1 - (1/4)} = \frac{2}{3} \end{aligned}$$

$$110.\overline{10}_2 = 6 + \frac{2}{3} = \frac{20}{3}$$

2. (0.3:2d) Converting the following base 10 numbers to binary and express each as a floating point number  $fl(x)$  by using the Rounding to Nearest Rule:  $44/7$ .

• **ans:**

$$44/7 = 6 + \frac{2}{7}$$

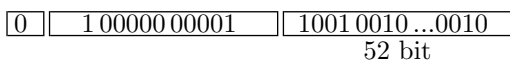


$$\frac{44}{7} = 110.\overline{010}_2 = 1.100\overline{10}_2 \times 2^2$$

$$fl\left(\frac{44}{7}\right) = \underbrace{1.10010010\dots0010}_2 \times 2^2$$

52 bit

If we write it in IEEE double:  $p = 2$ .



3. (0.3:4) Find the following sums by hand in IEEE double precision computer arithmetic, using the Rounding to Nearest Rule:

$$(a)(1 + (2^{-51} + 2^{-52} + 2^{-54})) - 1$$

$$(b)(1 + (2^{-51} + 2^{-52} + 2^{-60})) - 1$$

• **ans:** (a)

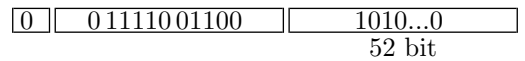
$$\begin{aligned} (1 + (2^{-51} + 2^{-52} + 2^{-54})) - 1 &= (1 + (2^{-51} + 2^{-52})) - 1 \\ &= 2^{-51} + 2^{-52} \end{aligned}$$

We do it again with details.

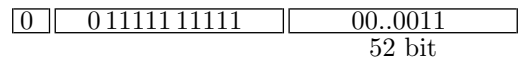
$$2^{-51} + 2^{-52} + 2^{-54} = 1.101_2 \times 2^{-51}$$

$$p = -51 = -(32 + 16 + 0(8) + 0(4) + 2 + 1) = -110011_2$$

$$1111111111_2 - 110011_2 = 1111001100_2$$



$$\begin{aligned} 1 + (2^{-51} + 2^{-52} + 2^{-54}) &= 1.00\dots001101_2 \\ &\simeq 1.\overbrace{00\dots0011}^{52 \text{ bit}} \times 2^0 \end{aligned}$$



$$\begin{aligned} (1 + (2^{-51} + 2^{-52} + 2^{-54})) - 1 &= 1.1_2 \times 2^{-51} \\ &\simeq 1.00\dots0011_2 - 1 = 1.1_2 \times 2^{-51} \end{aligned}$$

Matlab:

```
x=(1+(2^-51 +2^-52 +2^-54))-1
x*2^52
answer: 6.6613e-16 3
x=(1+(2^-51 +2^-52 +2^-60))-1
x*2^52
answer: 6.6613e-16 3
```

(b)

$$\begin{aligned} (1 + (2^{-51} + 2^{-52} + 2^{-60})) - 1 &= (1 + (2^{-51} + 2^{-52})) - 1 \\ &= 2^{-51} + 2^{-52} \end{aligned}$$

We do it again with details.

$$2^{-51} + 2^{-52} + 2^{-60} = 1.100000001_2 \times 2^{-51}$$

$$p = -51 = -(32 + 16 + 0(8) + 0(4) + 2 + 1) = -110011_2$$

$$1111111111_2 - 110011_2 = 1111001100_2$$

$$\boxed{0} \quad \boxed{01111001100} \quad \boxed{1000000010\dots0}$$

52 bit

$$\begin{aligned} & 1 + (2^{-51} + 2^{-52} + 2^{-54}) \\ & = 1.00\dots0011000000001_2 \\ & \simeq 1. \frac{\boxed{00\dots0011}}{52 \text{ bit}} \times 2^0 \end{aligned}$$

$$\boxed{0} \quad \boxed{0111111111} \quad \boxed{00\dots0011}$$

52 bit

$$\begin{aligned} & (1 + (2^{-51} + 2^{-52} + 2^{-64}) - 1) \\ & \simeq 1.00\dots0011_2 - 1 = 1.1_2 \times 2^{-51} \end{aligned}$$

1. (1.1:2a,1.1:4a) Find an interval of length one that contains a root.

$$x^5 + x = 1$$

Apply two steps of bisection method to find a root within 1/8 of the true root.

• **ans:**

$$f(x) = x^5 + x - 1$$

$$f(0) = -1, f(1) = 1$$

$$[a, b] = [0, 1]$$

After we compute the first  $c$ , we will replace either  $a$  or  $b$  by  $c$  so that the new interval can still trap a root inside it.

$$c = \frac{a + b}{2}$$

$i$	$a, f(a)$	$c, f(c)$	$b, f(b)$
0	0, [-1]	0.5, [-0.47]	1, [1]
1	0.5, [-0.47]	0.75, [-0.013]	1, [1]
2	0.5, [-0.47]		1, [1]

$$x = 0.75$$

```
f=inline('x.^5+x-1');
x=[0 1]; bis=[x f(x)];
for i=1:3
c=(x(1)+x(2))/2;
if(f(c)*f(x(1))<0) x(2)=c;
else x(1)=c; end
bis=[bis; x f(x)]
end
```

2. (1.1:a1) Find the interest rate  $I$  if 240 monthly payments of  $P = \$325$  generates an total annuity  $A = \$400,000$  by the bisection method with an initial interval  $[0.13, 0.14]$ .

• **ans:**

$$\begin{aligned} A &= P + P(1 + \frac{I}{12}) + \dots + P(1 + \frac{I}{12})^{N-1} \\ &= \frac{P}{I/12} \left( (1 + \frac{I}{12})^N - 1 \right) \\ &= \frac{325}{I/12} \left( (1 + \frac{I}{12})^{240} - 1 \right) \end{aligned}$$

Let  $f(I) = A - 400000$ .

$$c = \frac{a + b}{2}$$

$i$	$a_f(a)$	$c_f(c)$	$b_f(b)$
0	.13 <sub>-31696</sub>	0.135 <sub>-5461</sub>	.14 <sub>22878</sub>
1	0.135 <sub>-5461</sub>	0.1375 <sub>8434</sub>	.14 <sub>22878</sub>
2	0.135 <sub>-5461</sub>	0.13625 <sub>1419</sub>	0.1375 <sub>8434</sub>
3	0.135 <sub>-5461</sub>		0.13625 <sub>1419</sub>

Interest rate is

$$I = 0.13625$$