

# AERE 361 Lab 9 Report

Yahriel Salinas-Reyes

4/5/2022

## 1 Exercise 2

### 1.1 Exercise 2.1

Table 1: Exercise 2.1				
Size	Unsigned		Signed	
	Min. Value	Max. Value	Min. Value	Max. Value
8-bit	0	255	-128	127
16-bit	0	65535	-32768	32767
32-bit	0	4294967295	-2147483648	2147483647
64-bit	0	18446744073709551615	-9223372036854775808	9223372036854775807

### 1.2 Exercise 2.2

Value	8-bit unsigned representation
88	01011000
0	00000000
1	00000001
127	01111111
255	11111111

### 1.3 Exercise 2.3

Value	8-bit 2's complement signed representation
+88	10101000
-44	00101100
-1	00000001
0	10000000
+1	10000001
-128	10000000
+127	10000001

## 1.4 Exercise 2.4

### 1.4.1 32-bit Normalized Form

- Min Positive:  $N(\min) = 1.1755E-38$
- Max Positive:  $N(\max) = 3.403E38$
- Min Negative:  $N(\min) = -1.1755E-38$
- Max Negative:  $N(\max) = -3.403E38$

### 1.4.2 32-bit Denormalized Form

- Min Positive:  $N(\min) = 1.4013E-45$
- Max Positive:  $N(\max) = 1.1755E-38$
- Min Negative:  $N(\min) = -1.4013E-45$
- Max Negative:  $N(\max) = -1.1755E-38$

### 1.4.3 64-bit Normalized Form

- Min Positive:  $N(\min) = 2.225E-308$
- Max Positive:  $N(\max) = 1.798E308$
- Min Negative:  $N(\min) = -2.225E-308$
- Max Negative:  $N(\max) = -1.798E308$

### 1.4.4 64-bit Denormalized Form

- Min Positive:  $N(\min) = 4.9407E-324$
- Max Positive:  $N(\max) = 2.225E-308$
- Min Negative:  $N(\min) = -4.9407E-324$
- Max Negative:  $N(\max) = -2.225E-308$

## 2 Exercise 4

Output using single precision:

```
5.500000
5.545455
5.590164
5.633431
5.674649
5.713329
5.749121
5.781811
5.811315
5.837664
```

Output using double precision :

```
5.500000
5.545455
5.545455
5.590164
5.633431
5.674649
5.713329
5.749121
5.781811
5.811315
5.837664
5.861078
5.883543
5.935957
6.534422
15.413043
67.472398
97.137151
99.824694
99.989540
```

Explanation of Output:

To investigate the difference between the single precision and double precision, I ran the sequence script using `./seq -d` for the double precision. In contrast to using the single precision which normally has 10 outputs, the double precision prints 20, those same 10 plus 10 more. From the lab instruction, we know that this sequence is

supposed to converge to 6 but if we use the single precision, we haven't ran enough iterations to quite prove this convergence. When ran with double precision the output shows that it gets very close to 6 after 14 iterations and then begins to increase past 6 very quickly. This means the sequence is converging to 6. The iterations 15-20 simply have overshoot the convergence and become irrelevant to the monotonically-increasing sequence.