Experiment 8

M Naveen Kumar
16EE230

1 Aim of the experiment

To familiarise with the concept of Finite Precision effects on the Filter Coefficients.

2 Results/Graphs

2.1 Question 1

User defined function which takes the input X and convert it to the quantized coefficients with Xm the scaling factor and BitLength the total bit length including the sign bit (We take the scaling factor to be the maximum absolute value in X).

Code : fp.m

The following code was successfully written and used for Question 3. The observations for the step responses are further below in Question 3.

```
function X1 = fp(X,B)
        Xm = max(abs(X));
        X /=Xm;
        B -=1;
        x1 = round(X/2^-B)*2^-B;
        X1 = x1*Xm;
end
```

Observations

- When given a suitable input, scaling factor and bit precision specification, this function performs all the mentioned quantization applications mentioned successfully. This function has been utilized in the succeeding questions.
- For example fp([0.126,0.3,2],4) returns [0.25000 0.25000 2.00000] in which all elements are a multiple of 0.125 or $\frac{1}{2^3}$ which shows that it is quantised.

Conclusion

- Quantization, in mathematics and digital signal processing, is the process of mapping input values from a large set (often a continuous set) to output values in a (countable) smaller set, often with a finite number of elements.
- Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form ordinarily involves rounding.
- Quantization also forms the core of essentially all lossy compression algorithms.

2.2 Question 2

Generate a digital band-pass elliptical filter.

Code : task2.m

The following code was successfully written and used for Question 3. The observations for the step responses are further below in Question 3.

```
pkg load signal
 N=5;
[B,A]=ellip(N,-20*log10(.99),20,[.3  .4]);
```

2.3 Question 3

Code : task3.m

a) Graph

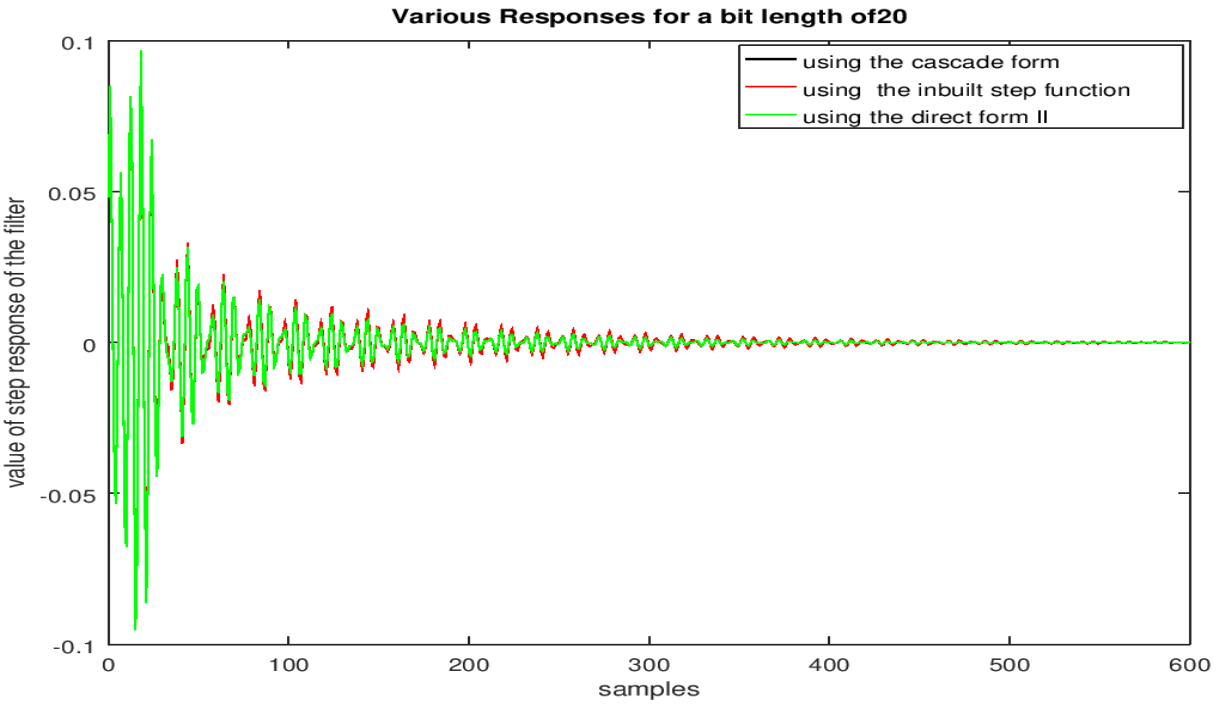

Figure 1: Various step responses using a Bit length of 20

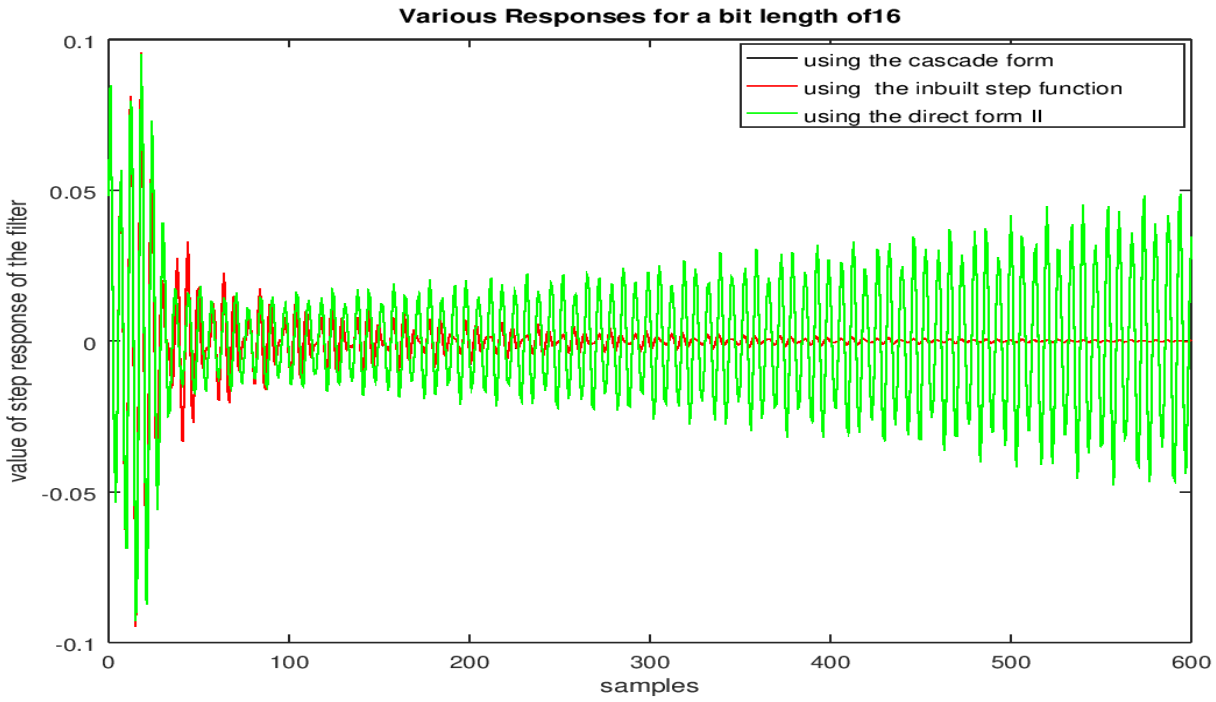

Figure 2: Various step responses using a Bit length of 16

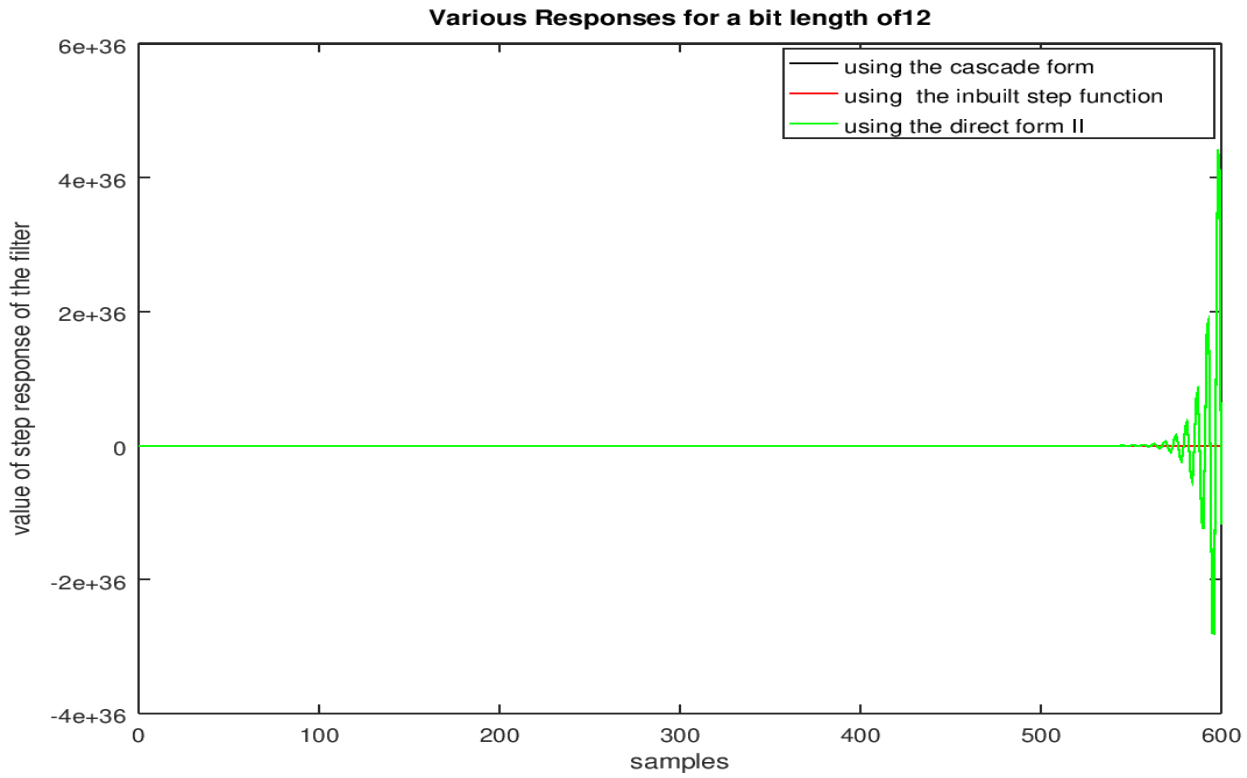

Figure 3: Various step responses using a Bit length of 12

Observations

Filter	Error for Bit length 20	Error for Bit length 16	Error for Bit length 12
Quantized Direct II IIR Filter	7.0303e-04	0.014953	4.4535e+34
Quantized Cascade Filter	1.2138e-07	2.4645e-06	1.7480e-05

- As the bit length was decreased, the response varied greatly from the original response. The mean error increased.
- However, the cascade filter output wasnt affected much when compared to DirectIIR.
- The mean absolute error in quantized and non-quantized Cascaded filter response is 1.2138e-07, which is nearly 500 times smaller error.
- When bit length is decreased the resolution decreases, hence the quantization error introduced also increases.
- This introduces an error in the filter response.
- The DirectIIR response of quantized coefficients tends towards infinity as bitlength decreases (As seen for Bitlength 12).

Conclusion

- The finite precision coefficients affect the DirectIIR filter and the inbuilt step response.
- Cascaded filter is not affected much as the error doesnt get compounded in many layers unlike in DirectIIR.
- Each time the error gets affected in only second order systems, and then gets cascaded to another second order.
- This reduces the overall error for it.

- It does not have any visible effect on Cascaded filter. As the bit length decreases, the mean error introduced increases visibly and the DirectIIR response tends towards infinity and blows out of proportion.