1 Lecture Outline

Reading: Chapter 4 FIR Filtering and Convolution

- FIR convolution (direct form)
- Transient and steady-state behavior
- Overlap-Add Block Convolution Method

2 FIR Convolution Direct Form (4.1.2)

Consider the convolution of a causal FIR filter of length $M+1$

$$\mathbf{h} = [h_0,\ h_1,\ \ldots,\ h_M]^T \tag{1}$$

with an input sequence of length L

$$\mathbf{x} = [x_0,\ x_1,\ \ldots,\ x_{L-1}]^T. \tag{2}$$

Using the direct form of the convolution we have

$$y(n) = \sum_{m=-\infty}^{\infty} h(m)x(n-m). \tag{3}$$

We must determine: 1) the range of values of the output index n and 2) the range of the summation index m.

1) The index of $h(m)$ must be within the range of indices for which it is defined over, i.e.

$$0 \le m \le M. \tag{4}$$

The index of $x(n-m)$ must be within the range of indices for which it is also defined over, i.e.

$$0 \le n-m \le L-1 \tag{5}$$

or

$$m \le n \le L-1+m. \tag{6}$$

Since m (the index on $\mathbf{h}$) can range from $0 \le m \le M$ we have

$$0 \le m \le n \le L-1+m \le L-1+M \tag{7}$$

or

$$0 \le n \le L-1+M. \tag{8}$$

The output sequence is therefore given by

$$\mathbf{y} = [y_0,\ y_1,\ \ldots,\ y_{L-1+M}] \tag{9}$$

Note that in the convolution

$$\text{length}(\mathbf{y}) = \text{length}(\mathbf{x}) + \text{length}(\mathbf{h}) - 1 = L + (M+1) - 1 = L + M. \tag{10}$$

2) The summation range over m in the convolution equation can be determined in the following way. For $h(m)$ we must have

$$0 \le m \le M. \tag{11}$$

For $x(n-m)$ we must have

$$0 \leq n-m \leq L-1 \tag{12}$$

which is equivalent to

$$1-L \leq m-n \leq 0 \tag{13}$$

or

$$n+1-L \leq m \leq n. \tag{14}$$

So m must satisfy

$$0 \leq m \leq M \quad \text{and} \quad n+1-L \leq m \leq n. \tag{15}$$

The range for the summation index m is then

$$\max(0, n-L+1) \leq m \leq \min(n, M). \tag{16}$$

Therefore the convolution of a causal FIR filter, $\mathbf{h}$ of length $M+1$ and an input sequence, $\mathbf{x}$ of length L is given by

$$y(n) \quad = \quad \sum_{m=\max(0,n-L+1)}^{\min(n,M)} h(m)x(n-m). \tag{17}$$

for $0 \leq n \leq L-1+M$.

3 Transient and Steady-State Behavior (4.1.7)

As shown earlier, for an order M (length $M+1$) filter and length L input sequence the output range will be

$$0 \leq n \leq L-1+M. \tag{18}$$

The output sequence can be partitioned into subranges Obviously, the values of M and L effect the duration

Table 1: Filtering timeline

$0 \leq n \leq M$,	input-on transients
$M \leq n \leq L-1$,	steady state
$L-1 < n \leq L-1+M$,	input-off transients

of these subranges. If the input sequence is shorter in length then the filter, there is no steady state.

Figure 1: Orfanidis p. 133 Fig. 4.1.5

4 Overlap-Add Block Convolution Method (4.1.10)

As described in Chapter 3, convolution can be carried out on an input block of samples to produce an output block of samples

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{19}$$

where $\mathbf{H}$ is the convolution matrix. This approach is not feasible in applications where the input is extremely long due to memory requirements. A practical approach is to divide the long input into contiguous non-overlapping blocks of manageable length, say L samples, then filter each block and piece the output blocks together to obtain the overall output, as shown in Fig. 4.1.6. Thus, processing is carried out block-by-block.

Figure 2: Orfanidis p. 143 Fig. 4.1.6

This is known as the *overlap-add* method of block convolution. Each of the input sub-blocks $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, ..., is convolved with the order-M filter $\mathbf{h}$ producing the outputs blocks:

$$\mathbf{y}_0 = \mathbf{h} \star \mathbf{x}_0$$

$$\mathbf{y}_1 = \mathbf{h} \star \mathbf{x}_1$$

$$\mathbf{y}_2 = \mathbf{h} \star \mathbf{x}_2$$

and so on. The resulting blocks are pieced together according to their absolute timing. Block $\mathbf{y}_0$ starts at absolute time $n = 0$; block $\mathbf{y}_1$ starts at $n = L$ because the corresponding input block $\mathbf{x}_1$ starts then; block $\mathbf{y}_2$ starts at $n = 2L$, and so forth.

Because each output block is longer than the corresponding input block by M samples, the last M samples of each output block will overlap with the first M outputs of the next block. Note that only the next sub-block will be involved if we assume that $2L > L + M$, or, $L > M$. To get the correct output points, the overlapped portions must be added together (hence the name, overlap-add).

Example 4.1.10, p. 143: $\mathbf{x} = [1, 1, 2, 1, 2, 2, 1, 1, 0]$ and $\mathbf{h} = [1, 2, -1, 1]$.