KAAP686 Mathematics and Signal Processing for Biomechanics

Moving Average Filter

William Rose

2015-07-25

Different people use the term “moving average filter” differently. In textbooks of signal processing, a moving average filter is a filter whose output is a weighted sum of past inputs, where the size and even the sign of the weighting factors for a particular filter may be non-uniform. Many experimental scientists use the term moving average filter to refer to a filter whose output at each time point is the simple average of certain number of input values around that time point. This second definition of moving average filter is just a special case of the more general first definition. We will discuss both “types” of moving average filters.

*1. A general moving average filter*

The general discrete time moving average filter has an output which is a weighted sum of past inputs, and therefore it is described by an equation such as

where x(k) is the input at time k, y(k) is the output, and the bi’s (i=0..N-1) are the weighting factors. This version of a moving average filter is causal because the output only depends on current and past inputs. Moving average filters are also known as “finite impulse response” (FIR) filters, because the output is non-zero for a finite time (not an infinite time) when the input is an impulse (a single non-zero value). Some filters, which we will not discuss here, have an output that is non-zero forever, when the input is an impulse. Those are called infinite impulse response (IIR) filters.

A simple differentiator is an example of a moving average filter:

This moving average filter has N=2, b0=1, b1=-1. One could also use b0=1/T, b1=-1/T, where T=sampling interval, if the units of time for the derivative are important.

Non-causal filters use future values of the input, as well as past values, to compute the output. Non-causal moving average filters are usually “centered”, i.e. they extend equally far into the future and the past, and are described by the equation

where N is an odd number. For example, the following noncausal filter computes the second derivative:

Where N=3, b-1=1, b0=-2, b1=1. (Divide each weighting factor by (T)2 if the time units are important.)

Moving average filters (such as the second derivative example above) that are symmetric (i.e. the weighting factors are the same for positive and negative lags) have zero phase lag. This means the output does not tend to lag behind, or lead before, the input. This is a desirable property if one is interested in the relative timing of different quantities, such as EMG and muscle force, or startle response times, etc.

*2. “Simple” moving average filters*

The general “moving average filter”, described above, was not necessarily the average of part of the signal. In fact, nether of the examples given were averages. However, the simplest type of moving average filter is the one that really is an average – and this is why the “class” of such filters has the name it has. The equation for a simple causal moving average filter is

where N may be even or odd. This simple moving average filter is sometimes called a flat moving average, since a plot of the weighting coefficients (the bi’s) will be flat across the top. The equation for a flat non-causal moving average filter with zero phase lag is

where N is odd.

*3. Frequency response of a flat moving average filter*

A flat moving average filter is a low pass filter. A low pass filter passes very low frequencies through with minimal change, but it reduces the amplitude of high frequency signals, or of high frequency components in a complex signal. This makes it good for getting rid of high frequency “noise” in a recording. A flat moving average filter will cause minimal change to a sinusoidal signal whose period is long compared to the filter window length, because the filter’s window only “sees” a small and relatively constant part of the oscillating signal at each moment. Therefore the moving average will track a slow signal quite well. When the oscillation period is short compared to the window duration, the filter will attenuate the input sinusoid, because the average will include positive and negative parts of the sinusoid, which will tend to cancel out.

*3a. Theoretical response of a continuous-time flat moving average filter*

We first consider the frequency response of a continuous time flat moving average filter. This is theoretical, since a real moving average filter will be implemented as a discrete-time filter. However, the analysis of the continuous time filter is simpler. The continuous time filter is described by the following equation:

Where *Tw* is the duration of the moving average window. The frequency response H(ω) is the function describing the relationship between the input x(t) and output y(t) when the input is a sinusoid:

where X() and Y() are complex numbers representing the magnitude and phase of the input and output sinusoids. Since the filter is symmetric and centered, it has zero phase lag. We assume, without loss of generality, that *x(t)* = cos *ωt* = cos 2*πft*. , where *ω* and *f* are the frequencies in radians per second and cycles per second, respectively. Since the filter has zero phase, the output also is a cosine. Because the input and out are cosines, their phase angle is zero. Therefore X(), Y(), and H() are purely real numbers. The input and output magnitudes are equal to the values at time=0 (when cosine reaches its maximum). Then *X*(*ω*)=*x*(*t*=0)=1, and *Y*(*ω*)=*y*(*t*=0):

Therefore the magnitude of the filter’s transfer function (also known as the filter’s magnitude ratio) is

or, equivalently,

The quantity is plotted below. H(*f*)=1 at zero frequency (*f*=0; “at DC”), by l’Hopital’s rule. This means a steady signal is passed through unaltered, as expected. The filter’s pass band is the range of frequencies where the transfer function is approximately unity (i.e. the output magnitude is about equal to the input magnitude). This occurs when *f Tw* <<1, i.e. when *f* <<1/*Tw*. The “cutoff frequency”, fco, can be defined as the frequency at which the magnitude ratio is 0.707 (-3 dB). Therefore the cutoff frequency, in radians per second, is given by

We solve numerically:

Therefore, if the desired cutoff frequency is known, the moving average window duration should be

For example, if one wants a 20 Hz cutoff frequency (*fco*=20 Hz), then the moving average window width should be *Tw*=22 ms.

The filter’s frequency response is zero whenever *f* = *n*/*Tw*, where *n*=1,2,3,… The first zero is at *f*=1/*Tw*. The upper envelope of this continuous time filter’s frequency response is obtained by replacing sin (*πfTw*) with unity (since that is the maximum absolute value of the numerator):

This envelope has a slope of 20 dB/decade on a log-log plot.

*3b. Response of a discrete-time flat moving average filter*

We now consider the general discrete time moving average filter. The equation for a general causal “moving average” (or FIR) filter is

where the *bk* are the coefficients of the impulse response. The frequency response for the causal discrete-time filter is

Where *ω* is the frequency in radians per sample. For a flat moving average filter, *bk* = 1/*N*, so the frequency response is

(sum of a finite-length geometric series)

(algebra)

(Euler’s formula)

(algebra). The exponential of a purely imaginary quantity, at the right hand end of the equation above, has magnitude 1. It is the frequency-domain expression of a pure time lag of (N-1)/2. The magnitude of the filter’s frequency response is obtained by replacing the exponential with unity:

Note that absolute value is only needed for the numerator, because the denominator is always nonnegative. That is because the allowed values for *ω* are 0<*ω*<π, where π is the Nyquist frequency (in radians per sample).

The equation for a centered, non-causal filter is

where N is odd. The frequency response for the non-causal, flat, centered, moving average filter is

This can be shown by a method very similar to that shown above, for the causal filter. The causal and non-causal filters different only in their phase angle, or delay: the non-causal filter has zero delay, i.e. zero phase lag.

The envelope of the filter’s magnitude response is obtained by noting that the numerator varies between 0 and 1. We replace the numerator with 1 to get the envelope:

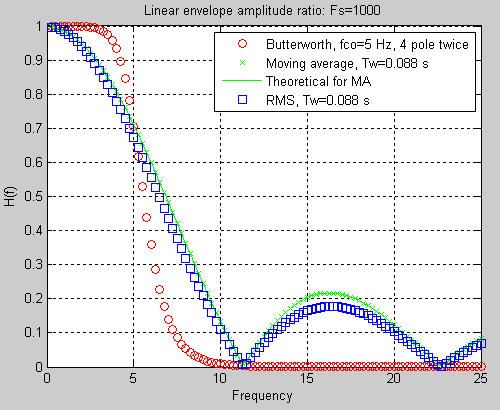
This is only valid for *ω*> 2π/N, which is the first zero of the numerator. We also note that if *ω*<<1, then we can apply the approximation sin *ω*/2 ~= *ω*/2:

And from the above equation it is obvious that the frequency response has a 20 dB/decade rolloff at higher frequencies. However, this approximation is only good at frequencies that satisfy

If N is small, there will be no frequencies that satisfy this inequality.

*4. Summary*

The flat moving average filter is a simple low pass filter with a cutoff of *fco* = 0.443/*Tw* and a 20 dB/decade rolloff. It has zero phase when implemented as a centered, non-causal filter. It is less flat in the passband, and less sharp around the cutoff frequency, and has higher sidebands, than a Butterworth low pass filter.

****

**Figure 2 Frequency response of “envelope detectors” designed for 5 Hz cutoff.**

Figure 2, like Figure 1, shows the frequency response of the envelope detection for Butterworth, moving average, and root mean square (RMS) methods. The desired final cutoff frequency for the Butterworth filtering is now 5 Hz instead of 20 Hz, and as a result a window width of 88 ms has been chosen (four times longer window to get four times lower cutoff frequency). In this case a 4th order Butterworth was used instead of a second order Butterworth. The results in the figure show that the Butterworth and moving average filters do attenuate by 0.71at 5 Hz, as expected. The 4th order Butterworth’s frequency response is closer to unity in the pass band and cuts off more sharply than the 2nd order Butterworth in Figure 1. A key point is that a wider window results in a lower cutoff frequency. (Image file emg\_fig2.jpg)

Copyright © 2015 W.C. Rose