ELECTRONOTES

WEBNOTE 18 2/15/2014

ENWN-18

Savitzky-Golay = Maximally Flat?

Very recently we posted an app note, AN-404 [1] on Savitzki-Golay (S-G) filters, that was based on polynomial fitting, and noted that these low-pass filters had (apparently) a maximally flat response at zero frequency. That is, like analog Butterworth, and digital filters derived from analog Butterworth (in particular, Bilinear z-Transform), they have a maximally flat passband. The S-G filters are of course FIR. The derivation of the S-G data from a maximally flat starting criterion is clearly "out there" although perhaps it is not obvious. As a stopgap here, we have simply decided to examine our favorite S-G example from AN-404 to see if it is at least apparently maximally flat. Here we will use two approaches.

First Approach:

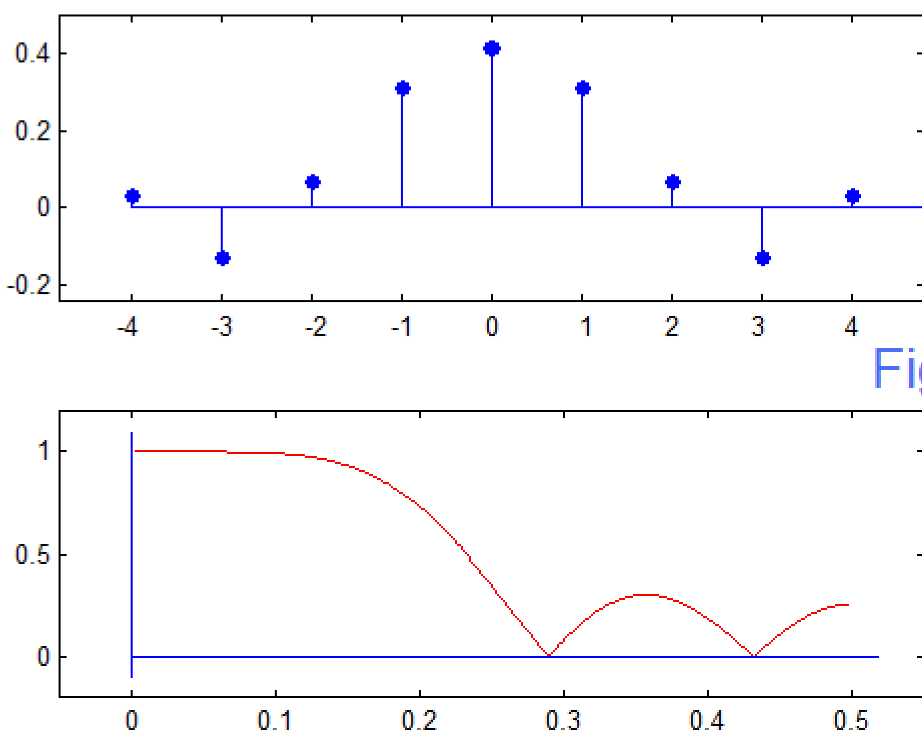


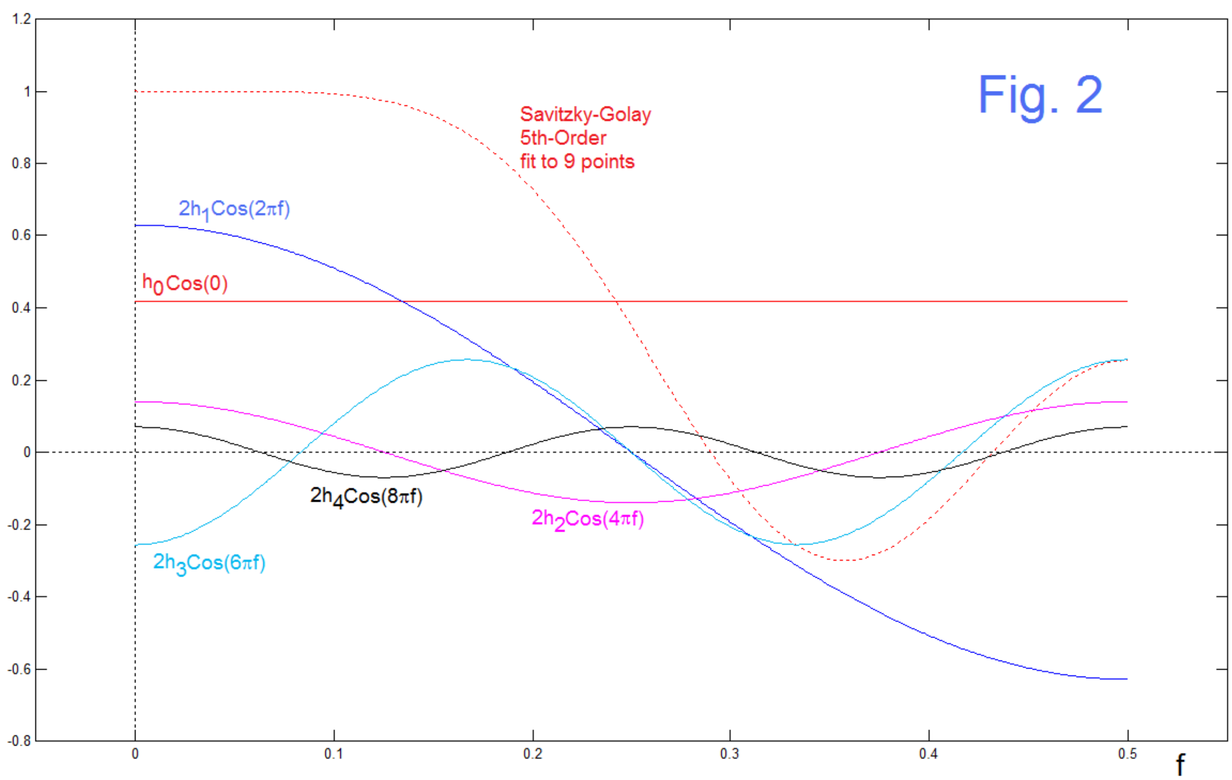
Fig 1

Starting with a S-G design from polynomial fitting (5^{th}-order to 9 points), we can calculate (from the computer data) the slope at DC. Well, that doesn't prove much! We already know (by eye) that the slope is small, and when we calculate a number, how would we know if it approached zero closely enough, or if something better could perhaps be found? Accordingly what we propose to do is compare the slope with the slopes of the cosine functions that are summed to obtain it. That is, we are looking for evidence that a delicate balance has been achieved. The slope can be computed from a numerical estimate at a very small frequency increment (2π/1000) away from zero.

Fig. 1 shows the output figures of the ***sg*** program from AN-404. The filter's impulse response is computed as:

$$
\begin{aligned}
h(-4) &= 0.03496503496504 \\
h(-3) &= -0.12820512820513 \\
h(-2) &= 0.06993006993007 \\
h(-1) &= 0.31468531468532 \\
h(0) &= 0.41724941724942 \\
h(1) &= 0.31468531468531 \\
h(2) &= 0.06993006993007 \\
h(3) &= -0.12820512820513 \\
h(4) &= 0.03496503496503
\end{aligned}
\qquad (1)
$$

From these we easily compute the components of the frequency response – a Fourier Series in the frequency domain [2 – if you have forgotten this]**:**



$$H(f) = h_0 + 2h_1 Cos(2\pi f) + 2h_2 Cos(4\pi f) + 2h_3 Cos(6\pi f) + 2h_4 Cos(8\pi f) \quad (2)$$

These five cosines, counting Cos(0), are plotted in Fig. 2 along with the sum (dotted red) which is the same response as the lower panel in Fig. 1 (which shows magnitude). Note that we start with the constant term (solid red line) with amplitude 0.417. We add to this the dark blue cosine of amplitude 0.619, so we are already over 1. Likewise the magenta and black cosines add even more. But there is the light blue negative cosine that restores everything back to 1. In fact, if we sum the h's in equation 1 we get exactly 1. Nothing above is new.

We now want to approximate the derivatives of the functions in Fig. 2 at f=0. Because we computed some 501 points of each of these curves for plotting purposes, it is convenient to take the difference between the first point (f=0) and the second point (f=2π/1000) and divide by the frequency separation (2π/1000) to get a good estimate of the derivatives. The derivative of the S-G final result is:

$$dsg = -1.501928599579994e{-}012 \quad (3)$$

which is clearly a number that approximates zero, but still is negative. But in order to see how small this might be, we calculate the derivatives of the cosine components:

$$\begin{aligned} d1 &= -0.00197721964082 \\ d2 &= -0.00175751122355 \\ d3 &= 0.00724961454430 \\ d4 &= -0.00351488368141 \end{aligned} \quad (4)$$

These are very very much larger than dsg, so we surmise how they must have fought with each other to arrive at a tiny final slope. The total of the four slopes d1 to d4 is indeed:

$$dtot = -1.486468066691504e{-}012 \quad (5)$$

which is not only tiny but nearly identical to dsg as it should have been.

We have derived nothing. All we have is one piece of very solid evidence that it is not an accident that the slope is maximally flat.

Second Approach:

Approach 1 was just a confidence-builder and we ultimately would appreciate a derivation and/or actual proof. I tried – not too hard. Soon enough one runs off to a text for a hint. All I found convinced me (1) that it was not going to be simple, but (2) that there was a second test that was very easy to do, involving an interesting trick. What I found was part of a "text" [3] which it was hard to read on screen, but which gave me a hint I couldn't resist. What the authors suggested was to start with a filter that is high-pass**:** the desired S-G subtracted from "one". This we easily get by subtracting 1 from the h(0) term of equation (1). Now, if indeed there are five derivatives equal to zero in the S-G at f=0 and an amplitude value of 1 (the start of the passband) this same slopes will now be in the high-pass at f=0 at an amplitude 0. Hence, there should be a fifth-order zero at z=1 for the high-pass. Matlab gave the seven zeros of this high-pass as:

```
zeros =   -1.7676
          -0.5657
           1.0025
           1.0013+ 0.0022i
           1.0013- 0.0022i
           0.9987+ 0.0023i
           0.9987- 0.0023i
           0.9974
```

Close enough. Encouraging for more work later?

REFERENCES:

[1] B. Hutchins, "Savitzki-Golay Smoothing", Electronotes Application Note No. 404, Feb. 13, 2013.
http://electronotes.netfirms.com/AN404.pdf

[2] See our digital filter design series in EN#197, pg. 11, etc.
http://electronotes.netfirms.com/EN197.pdf

[3] Google savitszky golay maximally flat which leads to:
http://books.google.com/books?id=sNEodegXy2MC&pg=PT293&lpg=PT293&dq=savitzky+golay+maximally+flat&source=bl&ots=6eMXzGHFCl&sig=c4s5LRCjhsSa50oGzpZ55rKp5c0&hl=en&sa=X&ei=_af-Uq6jGKnMsQTdq4CYCg&ved=0CCwQ6AEwAA#v=onepage&q=savitzky%20golay%20maximally%20flat&f=false
That's quite a URL. This apparently being part of a Digital Signal Processing Handbook on CD-ROM by V. Madisetti and D. Williams (1999). The part relevant to S-G is online, but hard to read.