

Savitzky-Golay Smoothing and Differentiation Filter

Neal B. Gallagher

Key words: Smoothing, differentiation, end-effects

Introduction: One of the most commonly used and frequently cited filters in chemometrics is the Savitzky-Golay smoothing and differentiation filter.[1] The “savgol” filter is often used as a preprocessing in spectroscopy and signal processing. The filter can be used to reduce high frequency noise in a signal due to its smoothing properties and reduce low frequency signal (e.g., due to offsets and slopes) using differentiation. After a brief description of the filter, the first example in this white paper will focus on the smoothing aspects of the filter and the second example on differentiation filtering.

The Savitzky-Golay (savgol) Filter: For a given signal measured at N points and a filter of width, w , savgol calculates a polynomial fit of order o in each filter window as the filter is moved across the signal. This is shown for three filter windows in the left of Figure 1 for $w = 7$. In this case, the signal is a spectrum measured at discrete points (blue line with measurements at the filled dots). The filter estimate at the center of each window is given by the polynomial fit at the center point (to make the calculation easy w is typically an odd integer). An example fit for the window [22,28] is shown in the subplot in the top right of Figure 1. The filtered signal at the center point, point 25, is given by the “X” in the subplot. The filter calculation is complete when the filter window moves the signal one-at-a-time.

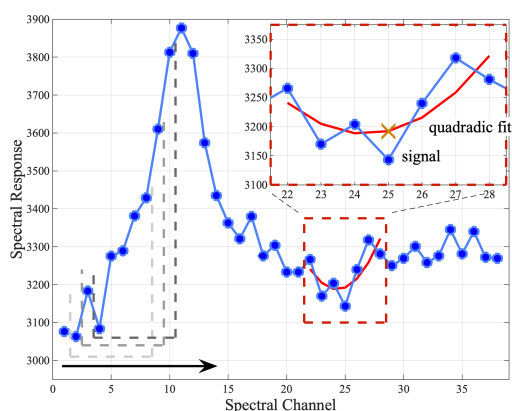


Figure 1: Spectrum measured at discrete points (blue line with dots). Filter windows, $w = 7$, are shown in the bottom left. A quadratic fit is shown in the top right for window [22,28] with corresponding filter value at point 25 given as X.

A restriction for the filter is that $w \geq (o + 1)$ i.e., the number of measurements must be \geq the number of parameters to be estimated. If $w > (o + 1)$ the filter smooths the signal and if $w = (o + 1)$ no smoothing is provided. Figure 2 shows an example savgol smoothing for a Raman spectrum (RamanDustParticles data in PLS_Toolbox [2], and script savgoldemob.m for the examples shown in this white paper). This shows that as the filter width, w , increases the smoothing increases and the peaks appear more suppressed. At higher smoothing, the peaks also acquire an artifact seen as minima on either side of the main peak (see the subplot in Figure 2). Note that the Whittaker smoother can also be used for smoothing.[3,4]

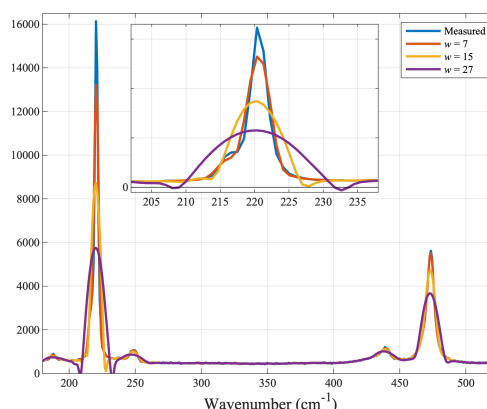


Figure 2: Measured spectrum (blue) and savgol smoothed data with filter width $w = 7, 15$ and 27 and polynomial order $o = 2$.

Derivatization and End-Effects: The derivative in each filter window can also be estimated with the derivative order given by d . Therefore, a filter can be fully defined by $\text{savgol}(w, o, d)$. Figure 3 shows an example Raman signal (top) and estimates of the first derivative (bottom). The noisy blue curve in the bottom graph is the first difference with no apparent peak signal significant above the noise. In contrast, the smooth red curve is a $\text{savgol}(15, 2, 1)$ filter. The filtered signal is significantly smoother and the minor peak at 1008 cm^{-1} is clearly evident.

Figure 4 shows another example signal in the top graph and estimates of the 1st derivative of the signal given by $\text{savgol}(7, 2, 1)$ (middle graph) and $\text{savgol}(15, 2, 1)$ (bottom graph). The ends of the signal are given as the first and last p set of points where $p = (w-1)/2$. In this

region, the filter window does not have a full set of w points for its calculation and the estimated filtered can be poor. For example, in the left end $[1, p]$ the signal (top) shows a fairly flat trend while the first derivative for $\text{savgol}(7,2,1)$ (middle) has significant positive and negative values and differs significantly from the $\text{savgol}(15,2,1)$ filter (bottom). Note that the two signals agree in the $[p+1, N-p+1]$ region with $\text{savgol}(15,2,1)$ smoother than the $\text{savgol}(7,2,1)$ filter as expected. Because of end-effects the savgol filter is typically calculated using the entire spectrum and then excluding the ends prior to modeling. It is also typical to eliminate artifacts such as cosmic ray spikes in Raman spectra prior to using the savgol filter.

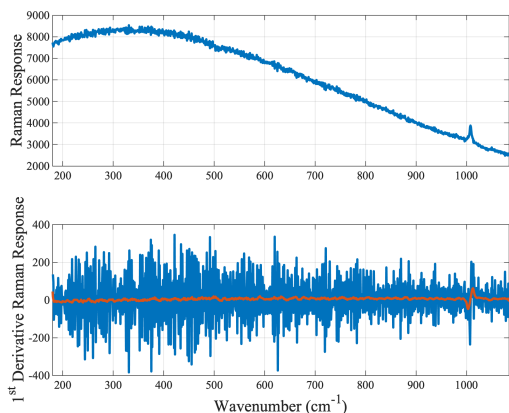


Figure 3: Measured Raman spectrum (top) and estimate 1st derivatives (bottom). The noisy curve is the first difference (blue) and the smooth curve is a $\text{savgol}(15,2,1)$ filter.

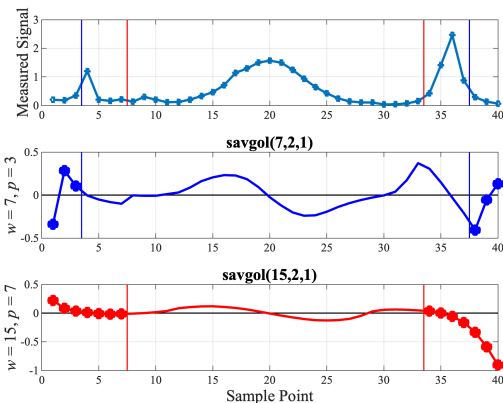


Figure 4: Measured spectrum (top) and savgol 1st derivative $\text{savgol}(7,2,1)$ (middle) and $\text{savgol}(15,2,1)$ (bottom).

Weighted Savitzky-Golay Filter: For the savgol filters shown above, the signal was fit in each window of width w to a set of polynomials using least-squares. The fit can also use weighted least-squares to allow greater control on the smoothing aspects of the filter. For example, a $1/d$ weighting can be used where d is now

defined as the distance a point is from the center of the filter window. Figure 5 shows an example measured IR spectrum (top), a $\text{savgol}(7,2,1)$ filter (middle) and a $\text{savgol}(7,2,2)$ filter (bottom). The blue curves for the filters correspond to a traditional least-squares fit and the red curves are the $1/d$ weighted least-squares fit. It should be clear that the red curves using the $1/d$ weighted fit are less smooth (this is especially evident for the peak at 905 cm^{-1}).

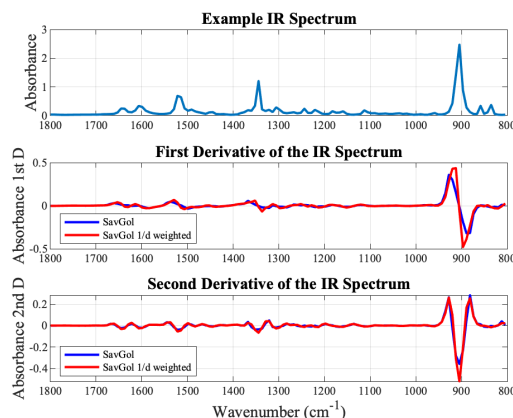


Figure 5: Measured IR spectrum (top), $\text{savgol}(7,2,1)$ filter (middle) and $\text{savgol}(7,2,2)$ filter (bottom). The blue curves for the filters correspond to the traditional least-squares fit and the red curves are the $1/d$ weighted least-squares fit.

Conclusions: The Savitzky-Golay smoothing and differentiation filter can be used to reduce high frequency noise in a signal due to its smoothing properties and reduce low frequency signal using differentiation. These properties are the reason the “ savgol ” filter is one of the most popular signal processing tools in spectroscopy and chemometrics. Additionally, weighted least-squares can provide more control over the design of the savgol filter.

References:

- [1] Savitzky, A, Golay, MJE, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.* 1964, **36**(8), 1627-1639.
- [2] PLS_Toolbox and Solo. Eigenvector Research, Inc., Manson, WA USA 98831; software available at www.eigenvector.com.
- [3] Gallagher, NB, "Whittaker Smoother," white paper Eigenvector Research, Inc., www.eigenvector.com.
- [4] Eilers, PHC, "A Perfect Smoother," *Anal. Chem.* 2003, **75**, 3631-3636.