**19.1.9 Algorithms (Peak Analyzer)**

## Baseline Detection Algorithm

There are four methods used in Origin to detect the baseline automatically in the data: **User Defined** for general purpose baseline model, **XPS** mode for X-Ray photoemission spectrum, **End Points Weighted** and **Straight Line** (Not available for **Create Baseline** Goal). You can also define a constant baseline value or provide existing baseline points through the datasheet.

### **User Defined**

Four methods are provided for anchor points detection, but only the following methods are used in "auto baseline detection".

#### **2nd Derivative (zeroes)**

This method is based on the fact that the baseline area has a smaller curvature than the peak area. The curvature of a curve is defined as

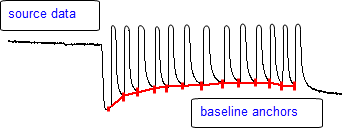
\kappa=\frac{y''}{(1+y'^2)^{ \frac{3}{2}}}

where y' and y'' are the first and second derivatives of the curve, respectively. After [the **Adjecent-Averaging** smoothing](https://www.originlab.com/doc/en/Origin-Help/Smooth-Algorithm#The_adjacent-averaging_method) is computed, the second derivative at each data point is calculated. Next, all data points whose second derivative approaches to 0 (under the tolerance) are used to make a second-order polynomial fit. With the fitted baseline in hand, we can adopt the points which lie closest to the fitted line as anchor points.

#### **2nd Derivative (peaks)**

This method is useful when the baseline is constructed from connecting the negative peaks of the pulse.

After [the **Savitzky-Golay** smoothing](https://www.originlab.com/doc/en/Origin-Help/Smooth-Algorithm#The_Savitzky-Golay_method) is computed, the 2nd-order derivative of data point is calculated. Next, all peaks of the 2nd-order derivative curve are found by [the **Local Maximum** method](https://www.originlab.com/doc/Origin-Help/PA-Algorithm#Local_Maximum). With the peaks of 2nd-order derivative in hand, we can adopt the points which lie closest to the peaks as anchor points.



#### **1st Derivative and 2nd Derivative**

In this method, we implement the **[Savitzky-Golay](https://www.originlab.com/doc/en/Origin-Help/Smooth-Algorithm" \l "The_Savitzky-Golay_method" \o "UserGuide:Algorithm (Smooth))** smoothing algorithm. Besides the second derivative threshold, this method also selects the points passing the first-derivative threshold. Since usually a smaller first derivative means a smaller change in the original data set.

This method is more powerful when the baseline is approximately constant. In this case, both the first and second derivatives of the baseline are approximately zero.

### **XPS**

This mode is designed especially for X-Ray photoemission spectrum analysis. Two options are supported: **Shirley** and **Tougaard**.

#### **Shirley**

The **Shirley** algorithm is an attempt to use information about the spectrum to construct a background sensitive to changes in the data. The essential feature of the **Shirley** algorithm is the iterative determination of a background using the area of the peak to compute the background intensity B(E) at energy E.

B_n (E)=k_n \int_{E}^{E_{max}}dE'[I(E')-I_{max}-B_{n-1}(E')],

where I_{max} is the end point intensity at the upper bound of the energy bin. In the dialog this parameter is called the **Final Height**. The iterative value of the scattering factor is given by

k_n=\frac{I_{min}-I_{max}}{\int_{E_{min}}^{E_{max}}dE'[I(E')-I_{max}-B_{n-1}(E')]}

The **Shirley** baseline is set to 0 outside the specified range [E_{min}, E_{max}].

#### **Tougaard**

![B_n (E)=B_{n-1}+k \int_{E}^{E_{max}}\frac{dE'[I(E' )-I_{max}-B_{n-1}(E' )](E'-E)}{[1643+(E'-E)^2]^2}](data:image/png;base64,),

where k depends on the option set previously,

![k=\frac{I_{min}-I_{max}}{\int_{E_{min}}^{E_{max}} \frac{dE'[I(E')-I_{max}-B_{n-1}(E')](E'-E)}{[1643+(E'-E)^2]^2}}](data:image/png;base64,),

for **Final Height** option and k equals to **Adjustable Parameter** when the user selects the **Adjustable Parameter** option.

##### Reference

* Shirley D.A. High-resolution X-ray photoemission spectrum of valence bands of gold. Phys. Rev. B 1972; 5(12):4709–14.
* Tougaard S. Practical Algorithm for Background Subtraction. Surface Science 1989; 216(3): 343-360.
* Tougaard S., Jansson C. Comparison of validity and consistency of methods for quantitative XPS peak analysis. Surface Interface Anal. 20, 1993; 1013-1046.

### **End points weighted**

This method is designed for the special case when you want to create a baseline based on the end points, both start and end.

You can choose a specific fraction of points as end points to detect the baseline. Then, the adjacent average smoothing method is used to reduce the noise. The default window for smoothing is 6 percent of the total selected points. Since these points are presumed as a baseline, a simple linear interpolation is used to generate the baseline.

**Note:** This method depends heavily on the selection of end points. You should select the end points fraction very carefully.

### **Asymmetric least squares smoothing(Pro)**

Asymmetric least squares smoothing (**ALS**) method is used to find the baseline so that:

1. Baseline is smooth.
2. Baseline is faithful to the original curve.

It is implemented by minimizing the sum of two terms: distances between points and baseline, 2nd derivatives of baseline. The sum can be expressed as:

S=\sum_{i=1}^n w_i(y_i-{y_b}_i)^2 + \lambda \sum_{i=2}^{n-1} [ {(y_b}_{i+1}-{y_b}_i) - ({y_b}_i-{y_b}_{i-1}) ]^2

where y is original data, y_b is calculated baseline, w_i is weight for each point, \lambda is a factor to balance the residual and the 2nd derivative, and smoothing factor in X-Function is the log of this value.

Iteration procedures are introduced as below:

1. In the first iteration, w_i=1 is used, and once baseline is calculated, asymmetric factor p will be applied on points above baseline (for positive peaks) as the weight, and weight for remaining points is 1-p.
2. In the next iteration, use calculated weights in the last iteration to calculate the new baseline and update weight.
3. Repeat the above procedure until the specified number of iterations reaches.

##### Reference

* P.H.C. Eilers and H.F.M. Boelens. Baseline correction with asymmetric least squares smoothing, Leiden University Medical Centre Report, 2005.

## **Peak Finding Algorithm**

There are five methods used in Origin to automatically detect peaks in the data: **Local Maximum**, **Window Search**, **First Derivative**, **Second Derivative**, and **Residual After First Derivative**. The first three methods are designed for normal peak finding in data, while the last two are designed for hidden peak detection.

### **Local Maximum**

The local maximum method is a brute force searching algorithm which finds the local maximum in a moving window. The window size is determined by a predefined a number of local points.

Initially, an n-point window is placed at the start point of data stream. The maximum in this window, as well as its index, is recorded. Then the window is moved one step further. If the new maximun is greater than the saved maximum, update both the maximum value and index value and then move forward. If the maximum moves out of the window, i.e. all points in the window are less than the maximum, a peak is found an the whole window configuration is reconstructed for the next peak.

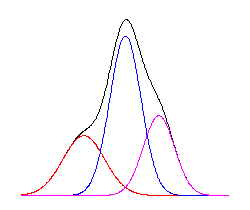
### **Window Search**

The window search method only differs from the local maximum method by the searching criteria used. This method uses a fixed window size, height, and width as its criteria, while the local maximum method searches in a fixed number of points.

### **First Derivative**

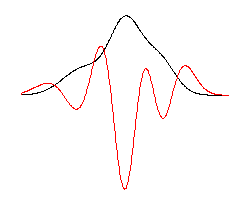
First derivative methods make use of the fact that the first derivative of a function at a local extreme point is equal to zero. There are two options in the method used to determine whether to smooth the original data. These methods include: **None** and **Savitzky-Golay**.

The three methods described above are used to find obvious peaks in the data. However, sometimes there may exist hidden peaks in data (See graph below). Origin provides two methods to detect hidden peaks in your data.



### **Second Derivative**

Since the second derivative can amplify the signal in the original data, we can use the second derivative to detect hidden peaks in data. The second derivative (red solid line) of the data with hidden peaks (black solid line) is sketched in the graph shown below.

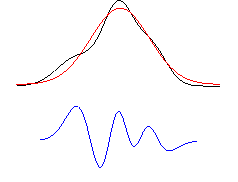


From the graph above, we find that the signal of the hidden peak is amplified, which makes it possible to detect the hidden peaks.

Origin provides four methods to smooth the derivative these include: **FFT Filter**, **Savitzky-Golay**, **Adjacent Averaging**, and **Quadratic Savitzky-Golay**. Please refer to the [smooth algorithm](https://www.originlab.com/doc/en/Origin-Help/Smooth-Algorithm) page for a detailed description.

### **Residual after First Derivative**

In the first step, Origin uses the first derivative method to find the visible peaks. Then a series of Gaussian peak functions are used to produce the local maximum in the data stream. A hidden peak is defined as one which fails to produce this local maximum. Origin then uses the first derivative method again to find peaks in the residual data.



### **Fourier Self-Deconvolution**

The Fourier Self-Deconvolution (FSD) method is used to find overlapping peaks in the spectrum.

Firstly FSD is calculated on the spectrum. Then the **Local Maximum** method is used to find peaks from the FSD result. Origin then uses the found peak centers to calculate peak heights from the original spectrum data, and further check whether the heights meet the constraints specified in the Peak Filter option.