Cyclic Savitzky-Golay Smoother

Zhangyi Hu

July 7, 2021

Abstract

Savitzky-Golay filter is a highly efficient low-pass filter suitable for roughly equidistant data. In this filter, a bandwidth b is specified ($b = 2r + 1$) and each point x_i has its smoothed value a result of linear transform from points within this bandwidth $\hat{x}_i = \sum_{j=i-d}^{j=i+d} a_j x_j$. This formula only handles points r steps away from the two end points ($r < i < I - r$). Two separate polynomial fits are conducted at two edges to handle those points where the key benefit of Savitzky-Golay is lost: polynomials are no-longer evaluated at 0. We introduce this cyclic Savitzky-Golay method, where edges points are handled exactly the same way as inner points. We enjoy the benefit that smoothed values are fitted polynomials evaluated at 0, while at the same time pay the cost of one additional condition: the data need to be assumed to flatten out at both ends.

1 Introduction

Savitzky-Golay is a local polynomial regression smoother parameterized by only two integers: the bandwidth b (or radius r) and the order of the polynomial d. To find the smoothed value of x_i, we first fit the data from x_{i-d} to x_{i+d} using equidistant integer coordinates from $-d$ to d. Suppose the polynomial fitted has coefficients at nth order being a_n, then the smoothed value is just $\hat{x}_i = a_0$, because the data x_i corresponds with a coordinate of 0.

Without considering the edge points, let's first derive the smoothing linear transformation. The polynomial we want to find is:

$$f(k) = \sum_{n=0}^{d} a_n k^n.$$

At point x_i the least square problem is:

$$\arg\min_{a_n} \sum_{k=-r}^{r} \left(f(k) - x_{i+k}\right)^2.$$

In matrix notation:

$$\arg\min_{a} \left(K \cdot a - x\right)^T \left(K \cdot a - x\right),$$

where

$$K = \begin{bmatrix} (-r)^0 & \cdots & (-r)^d \\ \vdots & \vdots & \vdots \\ r^0 & \cdots & r^d \end{bmatrix},$$

$$a = \begin{bmatrix} a_0 \\ \vdots \\ a_d \end{bmatrix},$$

$$x = \begin{bmatrix} x_{i-r} \\ \vdots \\ x_{i+r} \end{bmatrix}.$$

The least-square minimizer $\hat{a}$ is:

$$\hat{a} = \left(K^T K\right)^{-1} K^T x.$$

The smoothed $\hat{x}_i$ is:

$$\begin{aligned}\hat{x}_i &= f(0) = a_0 \\ &= \theta^T x,\end{aligned}$$

where θ^T is the first row of matrix Θ:

$$\Theta = \left(K^T K\right)^{-1} K^T.$$

As we can see, the matrix Θ is independent of x so it can be used for any point x_i making this method very efficient. For the smoothed value only the first row of Θ is needed, the rest rows can be used for the derivatives of the smoothed data up to order d. This paper only discuss the value (or the 0th derivative), so we will only focus on the first row θ.

2 Cyclic Savitzky-Golay

Suppose that our data flattens out to a constant c towards both ends, such that:

$$\lim_{i\to 0} x_i = \lim_{i \to I} x_i = c.$$

We can further define a cyclic version of the data y, such that:

$$\begin{gathered} y_i = \begin{cases} x_{i+I} & i < 0 \\ x_i & 0 \le i \le I \\ x_{i-I} & i > I \end{cases} \\ i \in \{-r, \ldots, 0, \ldots, I, \ldots, I+r\}. \end{gathered}$$

After that, the smoothed value for any x_i is just:

$$\hat{x}_i = \theta^T \chi_i,$$

where vector χ_i is defined as:

$$\chi_i = \begin{bmatrix} y_{i-d} \\ \vdots \\ y_i \\ \vdots \\ y_{i+d} \end{bmatrix}.$$

Both ends flatten to the same constant is too strong a condition, we can relax this condition to only requiring that both ends flatten to a constant. Let's denote the two constants at left and right as c_l and c_r, respectively. The idea is to force $c_l = c_r$ by tilting the data, and after the smoothing, the data is tilted back in the reverse direction. Formally, let's define the tilting magnitudes as:

$$t_i = \frac{i}{I}\left(c_l - c_r\right),$$

and the tilted data x'_i is:

$$x'_i = x_i + t_i.$$

Suppose the smoothed data after tilt is $\hat{x}'_i$, then the smoothed data after reverse title is:

$$\hat{x}_i = \hat{x}_i{}' - t_i.$$

3 Roughness

Sometimes we are interested at the difference between the data and its smoothed values as an indicator of the roughness of the data. The larger the difference, the rougher (or less smooth) the data is.

Let's define the cyclic version of the tilt magnitude as:

$$\begin{aligned} s_i &= \begin{cases} t_{i+I} & i < 0 \\ t_i & 0 \leq i \leq I \\ t_{i-I} & i > I \end{cases} \\ i &\in \{-r, \ldots, 0, \ \ldots, \ I, \ \ldots, \ I + r\}, \end{aligned}$$

and tilt window vector centered at ith point as:

$$\tau_i = \begin{bmatrix} s_{i-d} \\ \vdots \\ s_i \\ \vdots \\ s_{i+d} \end{bmatrix}.$$

The smoothed data for x_i is:

$$\begin{aligned} \hat{x}_i &= \theta^T (\chi_i + \tau_i) - t_i, \\ i &\in \{0, \ \ldots, \ I\}. \end{aligned}$$

As a result the difference, or error is:

$$\begin{aligned} e_i &= x_i - \hat{x}_i \\ &= x_i + t_i - \theta^T (\chi_i + \tau_i) \\ &= x_i - \theta^T \chi_i + \left(t_i - \theta^T \tau_i\right), \end{aligned}$$

Let's define:

$$\phi = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} - \theta = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} \theta_{-d} \\ \vdots \\ \theta_0 \\ \vdots \\ \theta_d \end{bmatrix},$$

then the error can be reformulated as:

$$\begin{aligned} e_i &= x_i - \theta^T \chi_i + \left(t_i - \theta^T \tau_i\right) \\ &= \phi^T \chi_i + \phi^T \tau_i \\ &= \phi^T (\chi_i + \tau_i). \end{aligned}$$

We can further take the norm of e_i to quantify the roughness of the data. The smoother is parameterized by the bandwidth b, so is the roughness. How does b control the meaning of the roughness obtained? On a high level, b can be roughly considered as a wave length threshold, waves with wave length less than b will all be considered as noise and contribute to the total error. This error measure is superior to integral of the second derivative commonly used in smoothing spline, as we have more control over how we want to set local details as noise.