Package ‘ADPF’

September 13, 2017

Title Use Least Squares Polynomial Regression and Statistical Testing to Improve Savitzky-Golay

Version 0.0.1

Maintainer Samuel Kruse <samdkruse@gmail.com>

Description This function takes a vector or matrix of data and smooths the data with an improved Savitzky Golay transform. The Savitzky-Golay method for data smoothing and differentiation calculates convolution weights using Gram polynomials that exactly reproduce the results of least-squares polynomial regression. Use of the Savitzky-Golay method requires specification of both filter length and polynomial degree to calculate convolution weights. For maximum smoothing of statistical noise in data, polynomials with low degrees are desirable, while a high polynomial degree is necessary for accurate reproduction of peaks in the data. Extension of the least-squares regression formalism with statistical testing of additional terms of polynomial degree to a heuristically chosen minimum for each data window leads to an adaptive-degree polynomial filter (ADPF). Based on noise reduction for data that consist of pure noise and on signal reproduction for data that is purely signal, ADPF performed nearly as well as the optimally chosen fixed-degree Savitzky-Golay filter and outperformed sub-optimally chosen Savitzky-Golay filters. For synthetic data consisting of noise and signal, ADPF outperformed both optimally chosen and sub-optimally chosen fixed-degree Savitzky-Golay filters. See Barak, P. (1995) <doi:10.1021/ac00113a006> for more information.

Depends R (>= 3.2.4), stats, utils

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Phillip Barak [aut], Samuel Kruse [cre, aut]
ADPF outputs a `data.frame` containing a column for the original data, the polynomial degree used to smooth it, and the requested derivative(s).

### Usage

```r
ADPF(YData, SthDeriv, MaxOrder, FilterLength, DeltaX, WriteFile)
```

### Arguments

- **YData**: a numeric `data.frame`, matrix or vector to transform
- **SthDeriv**: differentiation order
- **MaxOrder**: maximum polynomial order
- **FilterLength**: window size (must be odd)
- **DeltaX**: optional sampling interval
- **WriteFile**: a boolean that writes a `data.frame` to the working directory if true

### Details

This is a code listing of a smoothing algorithm published in 1995 and written by Phillip Barak. ADPF modifies the Savitzky-Golay algorithm with a statistical heurism that increases signal fidelity while decreasing statistical noise. Mathematically, it operates simply as a weighted sum over a given window:

\[
 f_t^{n,s} = \sum_{i=-m}^{m} h_i^{n,s,t} y_i
\]

Where \( h_i^{n,s,t} \) is the convolution weight of the \( i \)th point to the evaluate the \( s \)th derivative at point \( t \) using a polynomial of degree \( n \) on \( 2m + 1 \) data points, \( y \). These convolution weights \( h \) are calculated using Gram polynomials which are optimally selected using a \( F_{chi} \) test. This improves upon the signal fidelity of Savitzky-Golay by optimally choosing the Gram polynomial degree between zero and the max polynomial order give by the user while removing statistical noise. The sampling interval specified with the \( \text{DeltaX} \) argument is used for scaling and get numerically correct derivatives. For more details on the statistical heurism see the Barak, 1995 article. This can be found at [http://soils.wisc.edu/facstaff/barak/] under the publications section.
Author(s)

Phillip Barak
Samuel Kruse

References


Bevington, P. R. Data Reduction and Error Analysis for the Physical Sciences; McGraw-Hill Book Co.: New York, 1969; Chapter 10.


Examples

ADPF::CHROM

smooth<-ADPF(CHROM[,6],0,9,13)
numpoints=length(CHROM[,6])
plot(x=1:numpoints,y=CHROM[,6]);lines(x=1:numpoints, y=smooth[,3])

CHROM

Data frame of Chromatogram values

Description

This file contains a data frame of sample chromatography data. The 6th column is data without noise and the first five all have some gaussian noise added; these data sets showcase the advantages of ADPF over Savitzky-Golay.

Usage

data("CHROM")

Format

A data frame with 201 observations on the following 6 variables.

CHROM1  a numeric vector
CHROM2  a numeric vector
CHROM3  a numeric vector
CHROM4  a numeric vector
CHROM5  a numeric vector
CHROM6  a numeric vector

Source


Examples

ADPF::CHROM

smooth<-ADPF(CHROM[,6],0,9,13)
numpoints=length(CHROM[,6])
plot(x=1:numpoints,y=CHROM[,6]);lines(x=1:numpoints, y=smooth[,3])
Index

*Topic **datasets**
   CHROM, 4

ADPF, 2
CHROM, 4