

Measurement Science and Technology



PAPER

OPEN ACCESS

RECEIVED
1 October 2025

REVISED
9 June 2026

ACCEPTED FOR PUBLICATION
22 June 2026

PUBLISHED
18 August 2026

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Resolving features and derivatives in noisy data using weighted Whittaker–Henderson smoothing

Bert Mulder^{1,2,*} , Ad Lagendijk^{1,2} and Willem L Vos^{1,2}

¹ Complex Photonic Systems (COPS) Chair, Faculty of Science and Technology (S&T), University of Twente (UT), PO Box 217, 7500 AE Enschede, The Netherlands

² Complex Photonic Systems (COPS) Group, Photonic and Semiconductor Nanostructures (PSN) Chair, Department of Applied Physics and Science Education (APSE), Eindhoven University of Technology (TU/e), PO Box 513, 5600 MB Eindhoven, The Netherlands

* Author to whom any correspondence should be addressed.

E-mail: l.mulder1@tue.nl and w.l.vos@tue.nl

Keywords: noisy data, signal processing, regularization method, regression analysis, nonlinear filter, error analysis, optical spectroscopy

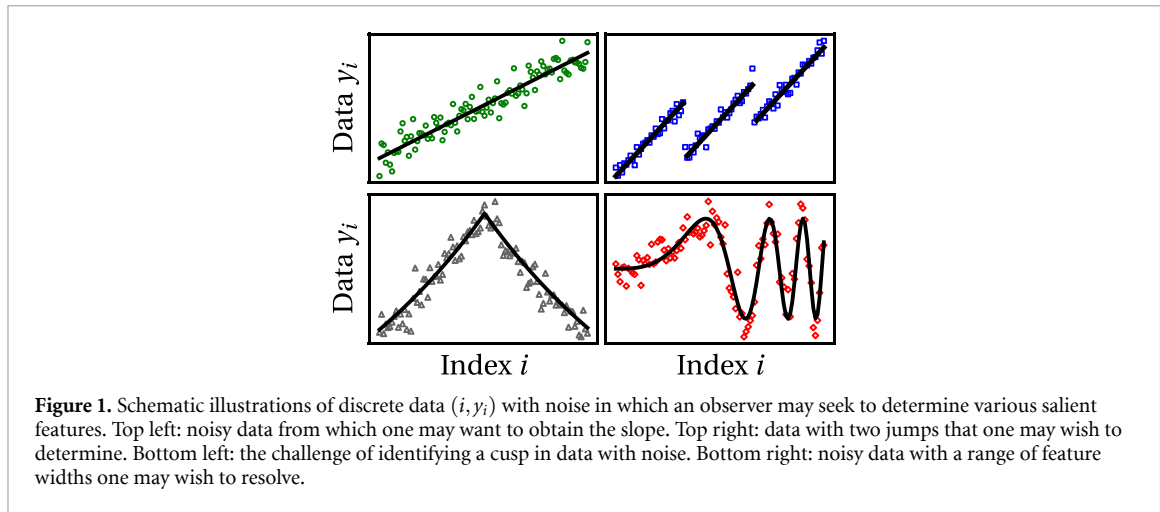
Abstract

A frequently occurring challenge in experimental and numerical observations is how to resolve features, such as spectral peaks—with center, width, height—and derivatives from measured data with unavoidable noise. Although many smoothing procedures exist, most are ineffective at reducing noise when the widths of the features varies strongly. Therefore, we modify the Whittaker–Henderson smoothing procedure to locally balance the spectral features and the noise. The central contribution of our procedure is that we introduce adjustable weights that are optimized using cross-validation. Using the measurement errors, a straightforward error analysis of the smoothed results is feasible. To illustrate the effectiveness of our smoothing algorithm, we derive for an optical Bragg reflector nanostructure the chirp of an optical pulse (group delay dispersion) using synthetic phase data with noise. The smoother faithfully reconstructs the group delay dispersion, reducing noise by more than a factor 40, allowing to identify details that otherwise remain buried in noise. Finding the optimal weights using the limited-memory BFGS algorithm for $N = 1000$ complex valued reflectivity data points takes on average less than two seconds on a typical computer. To further illustrate the power of our smoother, we introduce a general framework to solve commonly occurring difficulties in data and data analysis; how to properly smoothen unequally sampled data, how to identify and quantify discontinuities, including discontinuous derivatives or kinks, how to properly smooth data in the vicinity of boundaries to the data domains, and multidimensional smoothing.

1. Introduction

A basic feature of natural sciences is the collection of sequences of observations [1, 2], where the observations inevitably suffer from various imperfections [3, 4]. Many different kinds of such sequential observations are made in our field of research in optics, varying from spectral measurements where one records intensities, *e.g.*, reflectivity, scattering and Raman, as a dependent variable as a function of a control variable like wavelength or optical frequency, angle or momentum, position, or time. As a robust primary interpretation of the observations y_i as a function of some index i , one may wish to determine characteristic features, such as a slope, a jump, a peak, or a resonance frequency, see figure 1.

In all interpretations of observational data, also outside optics, one has to properly deal with imperfections in the data that hamper interpretation. Examples include noise in the observation due to intrinsic effects and imperfect apparatus, or unequal or even random sampling of the data as illustrated in figure 1. Therefore, one needs an educated filtering of the data to emphasize the property one seeks [5, 6].



The basic challenge with filtering is that noise and intrinsic features are treated the same. One may reduce noise by filtering high frequencies, but this leads to an underestimation of peaks and valleys in the data. Another issue is that it is hard to do a proper error estimation of the filtered data [7].

Filtering procedures are distinguished into two classes, namely causal filters, and non-causal filters that are also known as smoothers. Notable examples of causal filters are the Wiener filter [8, 9] and Kalman filter [5]. The Wiener filter is the first proposed filter that is designed using a statistical approach, as opposed to the usual frequency response approach. Similarly, Kalman filters combine previous measurements and a mathematical model of the system under study to estimate the optimal next value. Therefore, this method is the best causal linear filter for tracking the state of a dynamic system. For many kinds of data such as optical spectra, however, the use of a causal filter is detrimental. The position and shape of features change due to unavoidable (frequency-dependent) lag of causal filters, making these filters unsuitable for applications where phase preservation is critical.

The class of non-causal filters include local regression [10, 11] and smoothing splines [12, 13]. Local regression procedures such as LOESS [14], LOWESS [15] and the famous Savitzky–Golay filter [16] use a moving window over the data. The local smooth value is found using regression analysis within the window. A limiting factor to the smoothness is the size of the window. A larger window allows for smoother curves but also leads to more problems in interpreting data at the boundaries of the domain. The smoothing spline procedure splits the data into many parts at positions called knots. Higher-order derivatives at these knots are discontinuous, which may lead to artifacts when we are interested in these derivatives. Furthermore, the locations of the knots determine for a part the smoothed result, which leads to some ambiguity in the smoothed result.

Our method is based on Whittaker–Henderson smoothing, that was introduced in 1922 [17, 18], but seems to have been overlooked in literature [19, 20]. The Whittaker–Henderson smoothing procedure, from now on called in shorthand Whittaker–Henderson smoother, does not require any additional assumptions about the data, as it is not based on a model of the system under study, but on a general notion of roughness. Whittaker–Henderson smoothers thus offer the liberty to apply to many different types of data, in particular those for which no underlying model or analytic theory is known. A second favorable property of Whittaker–Henderson smoothers is that all data is smoothed together, meaning there is no moving window and there are no knots. The lack of a moving window leads to increased high-frequency noise suppression, which is beneficial for extracting features such as (higher-order) derivatives. A third favorable property is that Whittaker–Henderson smoothers do not use a set of basis functions, making it a simpler and therefore computationally cheaper method.

Whittaker–Henderson smoothing is an ℓ^2 -regularization technique and can be directly used as an alternative for, *e.g.*, local-regression and smoothing-spline procedures. Our main contribution to Whittaker–Henderson smoothing is to introduce adjustable weights that are optimized to locally balance the spectral features and the noise. This allows to construct the most likely and maximally smooth curve that tracks noisy data (see figure 1) with credible-interval estimation. We present a practical method to find the optimal weights using a cross-validation metric. This method is highly beneficial for a wide range of domains, ranging from, *e.g.*, statistics, signal processing, spectroscopy, to econometrics. In our research in optics and nanophotonics [21, 22], our method allows to identify resonances (with main properties center, width, amplitude), slopes (dispersion) and slow light, or even higher-order slopes (group velocity dispersion), see, *e.g.*, references [23–26].

Our article is organized as follows: section 2 sets the general framework of Whittaker–Henderson smoothing. It introduces the underlying principles, explains the roughness measure and how to do error analysis. Section 3 describes assumptions on noise and our numerical implementation of Whittaker–Henderson smoothing. Section 4 extends Whittaker–Henderson smoothing by introducing adjustable weights w_i for each respective data point i , and how to use cross-validation to find the optimal weights w_i . Section 5 has the main example of this manuscript. We show quantitatively how to optimally reconstruct the optical group delay dispersion from noisy phase data with sharp features. Sections 6–9 provide qualitative overview on applications of the same general framework for many common use-cases and pathologies, namely unequal sampling, jumps, cusps, boundaries, and multidimensional smoothing.

2. Whittaker–Henderson smoothing

To begin with, we possess a 1D-array of measured data including noise $\mathbf{y}$ that is the input of our procedure. The smoothed output data array is $\mathbf{z}$. The Whittaker–Henderson smoothing method relies on minimizing a functional $Q[\mathbf{z}]$, that is defined as

$$Q[\mathbf{z}] \equiv \|\mathbf{z} - \mathbf{y}\|^2 + \alpha \|\mathbf{D}\mathbf{z}\|^2. \quad (1)$$

Here the first term $\|\mathbf{z} - \mathbf{y}\|^2$ is the squared distance (using the ℓ^2 -norm) between the smoothed result $\mathbf{z}$ and the data $\mathbf{y}$. The second term $\alpha \|\mathbf{D}\mathbf{z}\|^2$ is a measure of the ‘roughness’ of the smoothed result, where $\mathbf{D}$ is a matrix operating on $\mathbf{z}$ and the smoothness parameter $\alpha \geq 0$ controls how strong the roughness is penalized. Hence, Whittaker–Henderson smoothing is a special case of ridge regression [27], also known as Tikhonov regularization [28] or penalized least-squares. While there is a strong similarity to compressed sensing [29, 30], compressed sensing uses ℓ^1 -norm penalties, however, leading to sparse models [31].

Minimizing $Q[\mathbf{z}]$ corresponds to balancing the smoothness of the resulting output $\mathbf{z}$ versus the difference between the data $\mathbf{y}$ and the result $\mathbf{z}$. With $\alpha = 0$, the functional $Q[\mathbf{z}]$ is only minimized when $\mathbf{z} = \mathbf{y}$, hence no smoothing is applied. The larger α becomes, the smoother the result $\mathbf{z}$ becomes, and ultimately the output is so smoothed that all features have been removed. Consequently, it is important to find the smoothness parameter α that gives the best compromise between smoothness and accuracy. The optimal smoothness parameter α is usually found using one of the many available statistical tests.

Using the variational principle, we find that $\mathbf{z}$ minimizing equation (1) is found by solving

$$(\mathbf{I} + \alpha \mathbf{D}^T \mathbf{D}) \mathbf{z} = \mathbf{y}, \quad (2)$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{D}^T$ the matrix transpose of $\mathbf{D}$. Generally, this linear system consists of sparse matrices, scales linearly with the number of data points, and is thus very efficient to solve.

2.1. Understanding the roughness measure

The roughness measure is often taken to be a derivative of $\mathbf{z}$ [32], with the matrix $\mathbf{D}$ acting as a finite difference operator. In this paper we will assume $\mathbf{D}$ is always some linear combination of finite difference operators. Other kinds of roughness measures can be used as well and can be tailored depending on the type of data. Care has to be taken when picking a roughness measure, as an improper measure may exaggerate specific features, leading to undesirable artifacts.

Historically, three kinds of roughness measures are commonly used. Bohlmann smoothers [33], introduced two decades before Whittaker–Henderson smoothers, use a first difference measure. This measure can be understood as penalizing the length of the curve $\mathbf{z}$, thus, a shorter curve is favorable. Hodrick–Prescott smoothers [34], which are used in macroeconomics to remove the cyclical component from a time-series, use a second difference measure. This can be understood as penalizing bending of the curve $\mathbf{z}$, thus, a flatter curve is favorable. Whittaker–Henderson smoothers originally used a third difference measure. This penalizes a change in the bending of the curve $\mathbf{z}$.

For every roughness measure $\mathbf{D}$ there exist null vectors $\mathbf{v}$ that fulfill the relation

$$\|\mathbf{D}\mathbf{v}\|^2 \equiv 0. \quad (3)$$

Since these curves have no roughness penalty, any data $\mathbf{y}$ will converge to a linear combination of these null vectors with sufficiently strong smoothing. This means that for an n^{th} derivative roughness penalty, the smoothed result will converge to an $(n - 1)^{\text{th}}$ -order polynomial.

2.2. Error analysis

Smoothing reduces the number of effective degrees of freedom d_{eff} . We calculate the degrees of freedom of the smoothed result using the smoother (or influence) matrix $\mathbf{H}$ [35], that is defined implicitly as

$$\mathbf{H}\mathbf{y} \equiv \mathbf{z}. \quad (4)$$

It is called the smoother matrix, because it transforms the data $\mathbf{y}$ into the smoothed result $\mathbf{z}$. This definition also holds for modified Whittaker–Henderson smoothers. When strong smoothing is applied, the diagonal of $\mathbf{H}$ tends to zero, while in the case of no smoothing the diagonal of $\mathbf{H}$ will be equal to unity, *i.e.*, $\mathbf{H}$ will then be equal to the identity matrix. The effective number of degrees of freedom d_{eff} is defined as [10]

$$d_{\text{eff}} \equiv \text{Tr}\{\mathbf{H}\}. \quad (5)$$

In the absence of smoothing the number of degrees of freedom is equal to the number of data points $d_{\text{eff}} = N$. With smoothing the number of degrees of freedom reduces and tends to near zero for strong smoothing.

Smoothing introduces a bias to the error estimation, since the number of effective degrees of freedom is reduced. A better estimate of the error is found by using the smoother matrix as defined in equation (4). Assuming we found the smooth curve $\mathbf{z}$ such that the residual $(\mathbf{y} - \mathbf{z})$ is independent and normally distributed, we then use the result of Marra and Wood [36] to estimate an unbiased credibility interval σ_z for $\mathbf{z}$, where the credibility interval for each data point i is given by

$$\sigma_{z,i} = \sigma_{y,i} \sqrt{H_{ii}}, \quad (6)$$

where σ_y is a known measurement error. This is equivalent to rescaling the error by the square root of the degrees of freedom of each data point.

Another way of estimating the error is using the covariance matrix [35]. Assuming that the errors in the data are uncorrelated, we construct the covariance matrix Σ_z as

$$\Sigma_z = \mathbf{H} \text{diag}\{\sigma_y^2\} \mathbf{H}^T. \quad (7)$$

The credibility interval is then given by

$$\sigma_{z,i} = \sqrt{\Sigma_{z,ii}}. \quad (8)$$

We find that the credibility intervals calculated using equations (6) and (8) are in very good mutual agreement. Therefore, we choose to calculate the credibility interval using equation (8), as this requires fewer assumptions.

2.3. Continuous analytical frequency response

To better understand the smoothing properties of the Whittaker–Henderson smoother, we approximate the *discrete* data y_i with a *continuous* function $y(x)$. After this excursion to the continuous case, we return to discrete smoothers. The generalization to continuous functions allows us to study the frequency response of the smoother analytically. This approximates the case of smoothing densely sampled data far from the boundaries of the data domain. For the continuous case we have to find the output function $z(x)$ that minimizes a functional $Q_c[z(x)]$ defined as

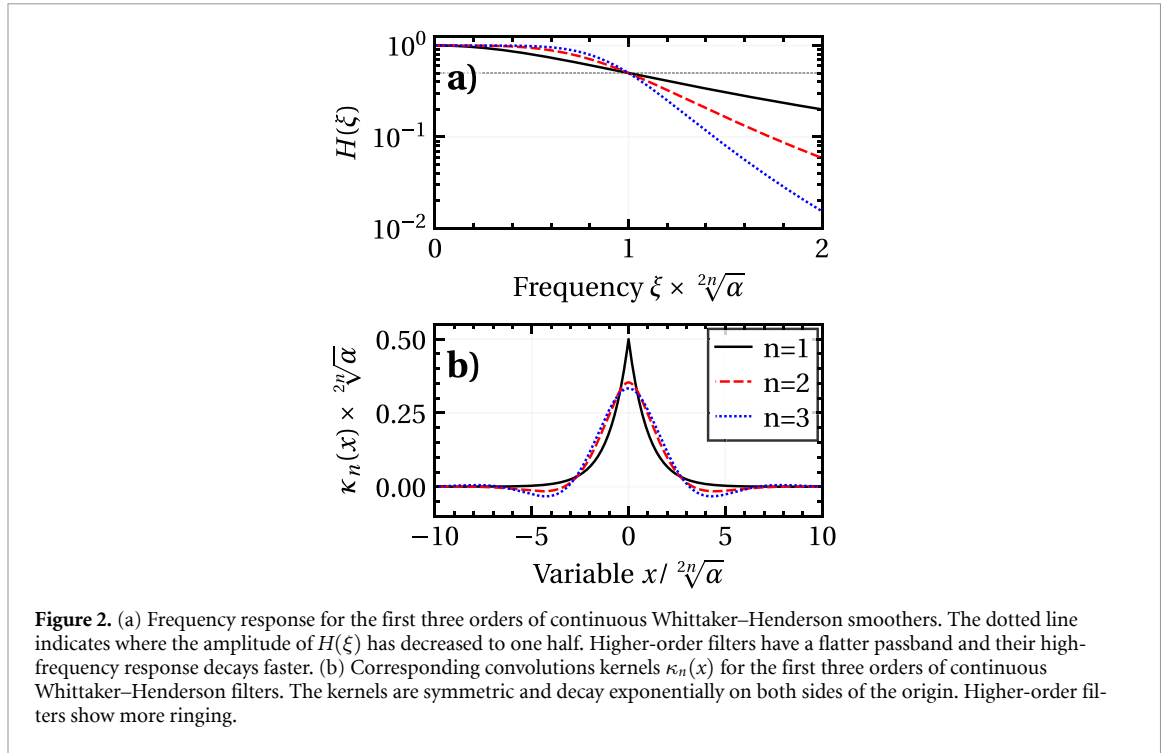
$$Q_c[z(x)] \equiv \int_{-\infty}^{\infty} |z(x) - y(x)|^2 dx + \alpha \int_{-\infty}^{\infty} |\hat{D}z(x)|^2 dx, \quad (9)$$

with $\hat{D}$ a linear differential operator as roughness measure and $y(x)$ the input function to be smoothed. Assuming that $y(x)$ vanishes as $x \rightarrow \pm\infty$, then the function $z(x)$ that minimizes $Q_c[z(x)]$ is found by solving the differential equation

$$z(x) + \alpha \hat{D}^\dagger \hat{D}z(x) = y(x), \quad (10)$$

where the Hermitian conjugate of a derivative is given by

$$\left[\left(\frac{d}{dx} \right)^n \right]^\dagger = \left(-\frac{d}{dx} \right)^n. \quad (11)$$



The linear differential equation is readily solved in the frequency-domain using a Fourier transform. We find that the frequency response $H(\xi)$ of the filter is

$$H(\xi) \equiv \frac{\tilde{z}(\xi)}{\tilde{y}(\xi)} = \frac{1}{1 + \alpha |\tilde{D}(\xi)|^2}, \quad (12)$$

where $\tilde{D}(\xi)$ is the linear differential operator $\hat{D}$ in the frequency-domain. The transformation of a differential operator to the frequency-domain is

$$\left(\frac{d}{dx} \right)^n \xrightarrow{\mathcal{F}} (i\xi)^n. \quad (13)$$

The result in equation (12) has three important implications. Firstly, we find that the frequency response is always real and positive, meaning it is a zero-lag filter. Secondly, if the differential operator $\hat{D}$ is of order n , then we find that the amplitude of the frequency response will decay as ξ^{-2n} for high frequencies. Thirdly, we find that a higher-order differential operator leads to a flatter passband. The frequency response $H(\xi)$ for the first three orders of Whittaker–Henderson smoothers are shown in figure 2(a). We observe that all but the first-order ($n=1$) smoother have a flat passband. As the order n increases the passband becomes flatter and high frequencies become increasingly suppressed.

The solution $z(x)$ of equation (10) can be written as a convolution of a kernel $\kappa(x)$ with the input function $y(x)$ as

$$z(x) = \int_{-\infty}^{\infty} y(x') \kappa(x - x') dx'. \quad (14)$$

The first-order ($n=1$) Whittaker–Henderson smoother is equivalent to smoothing with the (exponential) Poisson window function [37]

$$\kappa_1(x) = \frac{1}{2\sqrt{\alpha}} e^{-|x|/\sqrt{\alpha}}, \quad (15)$$

and is shown in figure 2(b). Due to the absolute value, the kernel has a cusp at $x=0$. The second- ($n=2$) and third- ($n=3$) order kernels $\kappa_n(x)$ are also illustrated in figure 2(b). The kernel function for the second-order case is similar to the basis function of a cubic regression spline smoother [38]. Higher-order convolution kernels become increasingly complex. The expressions for the convolution kernels

$\kappa_n(x)$ of the second-, third- and fourth-order Whittaker–Henderson smoothers are

$$\kappa_2(x) = \frac{1}{2\sqrt[4]{\alpha}} \sin\left(\frac{|x|}{\sqrt[4]{4\alpha}} + \frac{\pi}{4}\right) e^{-|x|/\sqrt[4]{4\alpha}}, \quad (16)$$

$$\kappa_3(x) = \frac{1}{3\sqrt[6]{\alpha}} \left[\frac{1}{2} e^{-|x|/2\sqrt[6]{\alpha}} + \sin\left(\frac{\sqrt{3}|x|}{2\sqrt[6]{\alpha}} + \frac{\pi}{6}\right) \right] e^{-|x|/2\sqrt[6]{\alpha}}, \quad (17)$$

$$\kappa_4(x) = \frac{1}{4\sqrt[8]{\alpha}} \left[e^{-\frac{|x|}{\sqrt[8]{\alpha}} \cos(\frac{\pi}{8})} \cos\left(\frac{|x|}{\sqrt[8]{\alpha}} \sin\left(\frac{\pi}{8}\right) - \frac{\pi}{8}\right) + e^{-\frac{|x|}{\sqrt[8]{\alpha}} \sin(\frac{\pi}{8})} \sin\left(\frac{|x|}{\sqrt[8]{\alpha}} \cos\left(\frac{\pi}{8}\right) + \frac{\pi}{8}\right) \right]. \quad (18)$$

All Whittaker–Henderson convolution kernels $\kappa_n(x)$ are symmetric, decay exponentially on both sides of the origin, and have a discontinuity at $x = 0$ due to the absolute value.

3. Methods

3.1. Assumptions on noise

When trying to remove noise from measured data, it is important to know the statistical properties of the noise. In this article we always assume that noise is independent, homoscedastic, *i.e.*, when all random variables have the same variance, normally distributed and stationary. Our assumptions represent the optimal case for linear (penalized) least-squares smoothers, which does not necessarily hold for real-world data. However, even if these assumptions are not exactly valid, Whittaker–Henderson smoothing can still be successfully applied with some limitations.

3.2. Numerical implementation

All our results are calculated using Python scripts running on a standard Lenovo ThinkPad laptop with a 13th Gen Intel(R) Core(TM) i7-13700H processor (2.40GHz base, 5.00GHz boost). To model noise, we use a pseudo-random noise generator as implemented by the ‘normal’ function from the NumPy random module. For reproducibility, the random number generator is initialized with a seed. Sparse linear systems are solved using the SciPy sparse module. Initializing the sparse matrices of the linear system in equation (2) to smooth $N = 10^5$ data points takes on average about 13 ms. Solving this linear system then takes on average about 21 ms. The numerical complexity is unknown to us, but appears linear for large N . Different and potentially faster methods are available in literature [39]. We find the optimal smoothing parameter α , or any other single filtering parameter, such as moving-window widths, using the generalized cross-validation method [40, 41]. All Python code is available as per data availability statement.

4. Weighted Whittaker–Henderson smoothing

4.1. Preliminaries

A modification of Whittaker–Henderson smoothing is by adding weights w_i for each data point [19, 20, 42]. We have to minimize a functional $Q_w[\mathbf{z}]$ defined as

$$Q_w[\mathbf{z}] \equiv \sum_{i=1}^N w_i |y_i - z_i|^2 + \alpha \sum_{i=1}^N |(Dz)_i|^2. \quad (19)$$

We find that $\mathbf{z}$ minimizing equation (19) is found by solving

$$(\mathbf{W} + \alpha \mathbf{D}^T \mathbf{D}) \mathbf{z} = \mathbf{W} \mathbf{y}, \quad (20)$$

with $\mathbf{W}$ a diagonal matrix consisting of the weights w_i for all data points, defined as

$$\mathbf{W} \equiv \text{diag}\{w_1, w_2, \dots, w_N\}. \quad (21)$$

A high weight w_i causes a high penalty to the least-squares difference, so that locally $\mathbf{z}$ closely tracks the data $\mathbf{y}$. With a weight set to zero ($w_i = 0$), the corresponding data point y_i will be completely ignored and smoothly interpolated instead. This feature is useful for incomplete data sets, where data points are missing, or to exclude outliers. Weighted Whittaker–Henderson smoothing is normally used whenever there are unequal variances in observations, for example when noise is heteroscedastic or non-stationary.

The weights w_i are then proportional to the inverse noise variances. We use the weights to apply different levels of smoothing to different parts of the data, which is vital to optimally resolve complex features in noisy data.

4.2. Find the optimal weights by cross-validation

For given input data $\mathbf{y}$, we want to find the optimal weight-to-smoothness ratio w_i/α such that the smoothed output $\mathbf{z}$ is *maximally smooth* while the *residual* $(\mathbf{y} - \mathbf{z})$ *remains random*. Previous approaches include the balancing of the trade-off between bias and variance [43], iterative procedures using Gaussian noise statistics [44], or local and non-local sample correlations to remove noise [45]. In contrast, we find the optimal weights directly using a statistical cross-validation method.

Cross-validation is a statistical tool to evaluate how well the smoothed data correlates with the noisy data, by using a subset of the noisy data to predict the rest of the smoothed data [35]. The cross-validation value is the RMS difference between the predicted smooth output data and the real noisy input data. In the case of weighted Whittaker–Henderson smoothing, a data point i is predicted by setting its respective weight w_i to zero. A smaller cross-validation value means a better prediction, indicating a better matching fit between the smoothed data and the noisy data, and *vice versa* a large cross-validation value indicates a worse fit. By minimizing the cross-validation metric we thus construct the most likely smooth curve that tracks our data.

Usual leave-one-out cross-validation [32] is extremely computationally expensive as it requires as many smooths as there are data points to calculate a single cross-validation value. The minimization process simply takes much too long. Therefore, we developed our own method where we split the data into two parts. This way we only require two smoothing operations to calculate a single cross-validation value, which makes the minimization process computationally feasible. Our method is similar to k -fold cross-validation with $k = 2$. However, in our case we assume a strong serial correlation in the ground truth and no serial correlation in the noise. Furthermore, random partitioning adds some variability, making minimization of the cross-validation metric much harder. Therefore, we use odd–even splitting of the data as opposed to the random partitioning used in k -fold cross-validation. In the case that there is some serial correlation in the noise, it is necessary to split the data into more than two parts to break the serial correlation.

For the first part of our method, we set even indexed $i = 2, 4, 6, \dots$ of w_i to zero to interpolate the even indexed $i = 2, 4, 6, \dots$ of $\mathbf{z}_e$ using the odd indexed $i = 1, 3, 5, \dots$ of the data. For the second part, we set odd indexed $i = 1, 3, 5, \dots$ of w_i to zero to interpolate the odd indexed $i = 1, 3, 5, \dots$ of $\mathbf{z}_o$ using the even indexed $i = 2, 4, 6, \dots$ of the data. We use both the even and odd indexed part of the data to make sure that we do not skip important features. Our cross-validation metric is then

$$S_{cv} \equiv \sqrt{\frac{1}{N} \sum_{i=1}^N \begin{cases} |y_i - z_{e,i}|^2 & \text{for even } i, \\ |y_i - z_{o,i}|^2 & \text{for odd } i. \end{cases}} \quad (22)$$

In case of complex valued data, the real and imaginary components are taken jointly using the absolute difference.

One might be tempted to try to find the weights w_i directly using a multivariate minimization algorithm based on our cross-validation metric. This works up to a point but often noise in the data prevents us from finding the weights w_i directly. To get around this limitation we have to reduce the effects of noise by strongly reducing the number of degrees of freedom of the weights. We do this by approximating the weights w_i with a cubic b-spline. The locations of the knots for the b-spline are found using the knot-vector constructor from the FORTRAN FITPACK package on the measured data. The knot-vector constructor uses the least-squares criterion to find the best locations and the required number of knots. We found that less knots also suffice and therefore we always pick about half the number of knots that the least-squares criterion gives. The optimum weights w_i are relatively insensitive to the exact number of knots. In case of complex valued data, we find the knot locations for the real and imaginary part of the data separately and then combine the locations into a single knot-vector.

The optimum b-spline coefficients are found by minimizing our cross-validation metric using the multivariate bounded limited-memory BFGS algorithm as implemented by the ‘minimize’ function of the SciPy optimize module. The L-BFGS-B algorithm is a quasi-Newton method that iteratively refines an initial estimate until a termination condition is reached. In our case the termination condition is reached when all the components of the projected gradient are smaller than 10^{-6} .

5. Extracting derivatives from noisy data: Optical group delay dispersion

The original motivation for our work is the study of the speed of light in advanced photonic nanostructures [23–25], but the methods described in this section are universally applicable. In our research we are interested in the group delay and the group delay dispersion that effectively correspond to taking first and second derivatives with optical frequency of spectral data. Simply taking the numerical derivative of the measured data as finite differences will greatly amplify the effects of noise [46]. Therefore, a filtering method is required to increase the signal to noise ratio.

A practical example where smoothing is important is when determining the group delay dispersion from a widely-used distributed Bragg reflector of light [47–49]. These reflectors are made of alternating layers of high- and low-index materials. When the reflections of light from the different layers are all exactly in phase in the backward direction, then the reflected waves add up constructively, meaning that many small reflections add up to one large reflection. Bragg reflectors made from many layers of lossless materials reach significantly higher reflectivities than metal reflectors, making them the logical choice when any amount of light loss would be detrimental. Examples of applications include (laser) cavities, gravitational wave detectors [50] and X-ray mirrors [51] for EUV lithography and synchrotrons.

Group delay dispersion adds a chirp to a light pulse [52–55]. A positive chirp ($D_2 > 0$), also known as normal dispersion, means that waves with a lower optical frequency (longer wavelength) arrive before waves with a higher optical frequency (shorter wavelength). Conversely, a negative chirp ($D_2 < 0$), also known as anomalous dispersion, means that waves with a higher optical frequency (shorter wavelength) arrive before waves with a lower optical frequency (longer wavelength). In both cases the pulse duration increases and the peak power decreases, leading to less efficient nonlinear optical interactions and reduced time resolution in pump-probe experiments. Understanding and compensating dispersion is therefore vital in any optical device or application employing nonlinear interactions or ultrafast processes.

The group delay dispersion $D_2(\omega)$ is defined as the second optical angular-frequency derivative of the optical phase [56, 57],

$$D_2(\omega) \equiv \frac{d^2\phi(\omega)}{d\omega^2}, \quad (23)$$

where ω is the optical angular frequency and $\phi(\omega)$ is the measured optical frequency-dependent phase shift of light detected in a reflection experiment. The second derivative is extremely sensitive to noise, even small amounts of noise in the measured signal will completely obscure the calculated dispersion. Generally, two methods are used to get around this problem. For simple samples, such as homogeneous glass slabs with weak dispersion, the measured phase is often approximated by a high-order polynomial [58]. The group delay dispersion is then calculated by differentiating the polynomial. For more complex nanophotonic samples, however, where the phase shift is not readily described by a polynomial, the phase data is often filtered using, e.g., a Savitzky–Golay filter [16, 59, 60] before it is differentiated.

As an alternative method to the optical angular-frequency derivative of the phase in equation (23), we obtain the dispersion D_2 from the optical angular-frequency derivatives of the complex reflectivity $r(\omega)$ as

$$D_2(\omega) = \text{Im} \left\{ \frac{1}{r(\omega)} \frac{d^2 r(\omega)}{d\omega^2} - \left(\frac{1}{r(\omega)} \frac{dr(\omega)}{d\omega} \right)^2 \right\}. \quad (24)$$

Using our method to find the weights, described in section 4.2, we estimate the most likely smooth curve from our complex reflectivity data. By finding the optimal weights w_i for all data points we eliminate noise, while preserving true features. From the smooth curve we calculate the derivatives $\partial_\omega r(\omega)$, and $\partial_\omega^2 r(\omega)$, and then use equation (24) to calculate the group delay dispersion. This is important, because the noise in real complex reflectivity data is often very nearly normally distributed. In contrast, noise in the phase data is not normally distributed, which violates the assumptions made while deriving the method. A second reason for using equation (24) is that the complex reflectivity is expected to be more or less continuous and smooth in physically realistic systems, while the group delay dispersion is not. A diverging group delay dispersion is associated with large jumps in the phase; this would otherwise be much harder to properly reconstruct using a smoothing algorithm.

To obtain synthetic data on the complex reflectivity of a Bragg reflector, we use the transfer-matrix method [61] for normally incident light. As a realistic sample we use a Bragg reflector made from 30 pairs of GaAs/AlAs $\lambda/4$ -layers stacked on a GaAs wafer in air. For GaAs we assume a relative permittivity of $\varepsilon_{\text{GaAs}} = 11.6$ and for AlAs we assume a relative permittivity of $\varepsilon_{\text{AlAs}} = 8.44$. The GaAs layer

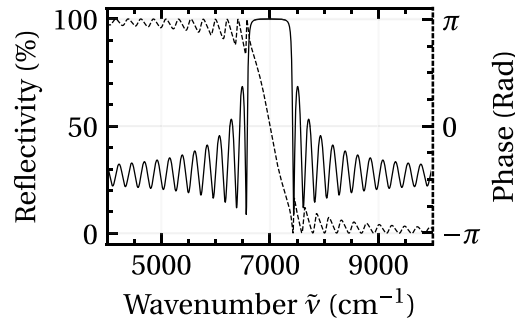


Figure 3. Intensity reflectivity (solid line) and reflection phase shift (dashed line) from a 30 pair GaAs/AlAs $\lambda/4$ -layer Bragg reflector with a center wavenumber $\tilde{\nu}_c = 7000 \text{ cm}^{-1}$ ($\lambda_c = 1429 \text{ nm}$). Data synthesized using transfer-matrix method [61].

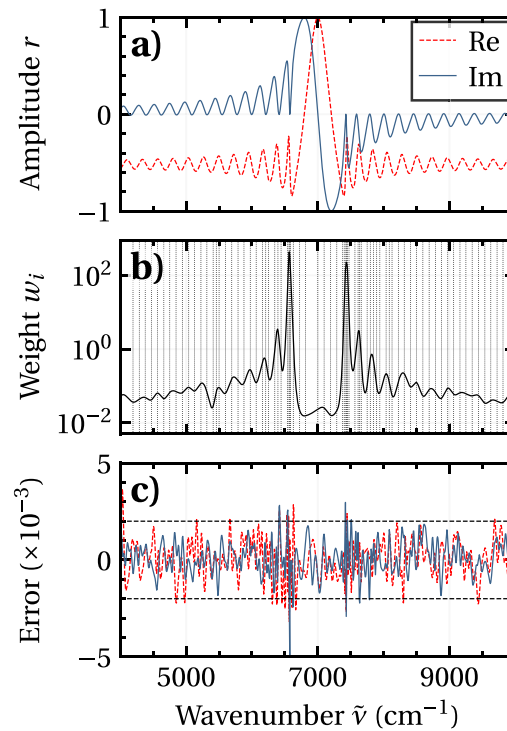
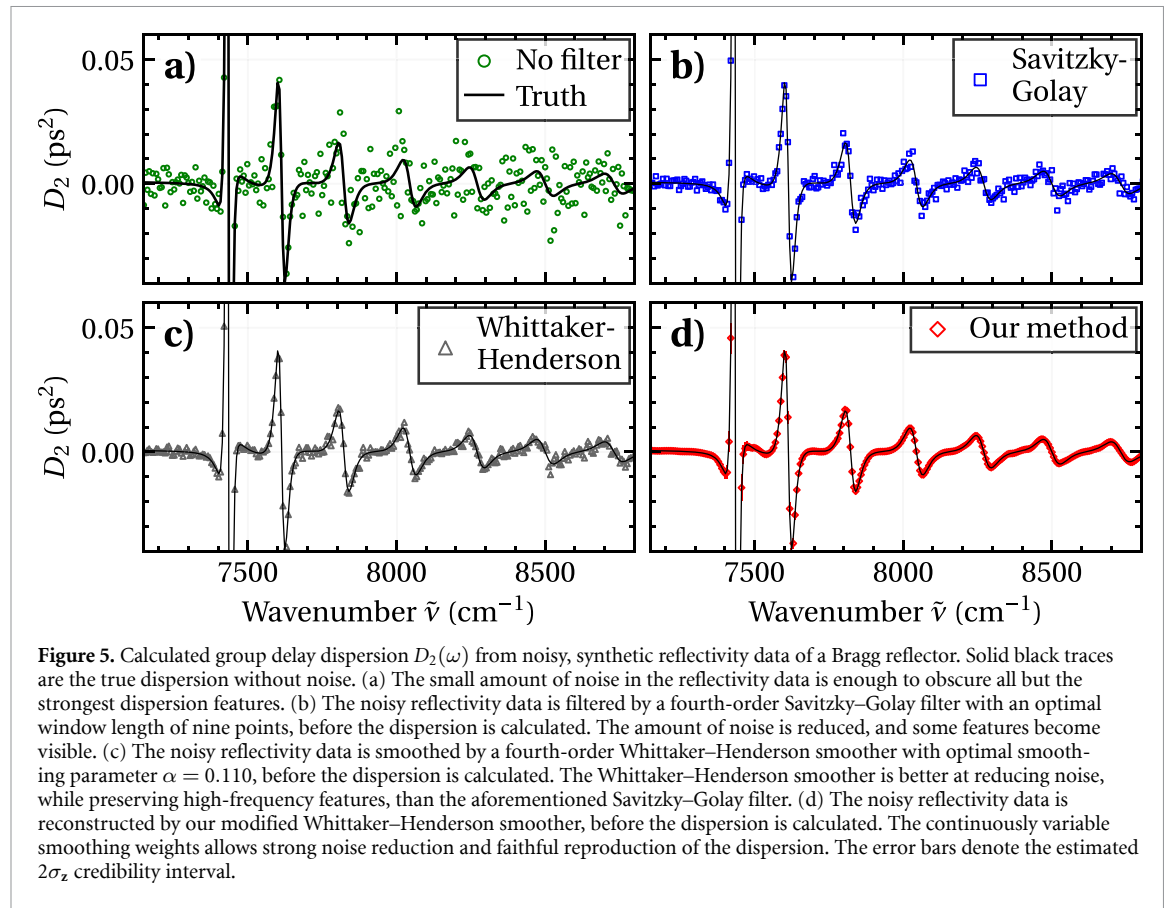


Figure 4. (a) Synthetic complex reflectivity data with noise. (b) The resulting weights w_i to optimally smooth the complex reflectivity data. The domain is divided in 80 knot locations (dotted lines), the weights in between the knots are interpolated using a cubic b-spline. The smoothness parameter is fixed to $\alpha = 1$. (c) Difference between smoothed synthetic data and the ground truth. The error is mostly smaller than the standard deviation (dashed lines) of the added noise.

has a thickness of about $L_{\text{GaAs}} = 104.9 \text{ nm}$ and the AlAs layer a thickness of $L_{\text{AlAs}} = 122.9 \text{ nm}$. The reflectivity spectrum is calculated from wavenumber $\tilde{\nu} = 4000 \text{ cm}^{-1}$ to 10000 cm^{-1} with step size $\delta\tilde{\nu} = 6 \text{ cm}^{-1}$. The reflectivity and reflection phase shift are shown in figure 3, and reveal a broad photonic stopband with a relative bandwidth $\Delta\tilde{\nu}/\tilde{\nu}_c = 11.9\%$ centered at $\tilde{\nu}_c = 7000 \text{ cm}^{-1}$ ($\lambda_c = 1429 \text{ nm}$). In the photonic stopband, the phase is seen to decrease from $+\pi$ to $-\pi$. To synthesize noisy data, we add normally distributed random numbers with a standard deviation $\sigma = 0.002$ to *both* the real and imaginary parts of the synthetic complex reflectivity amplitude.

The real and imaginary part of the complex reflectivity, as shown in figure 4(a), are smoothed together. This is important for finding the weights w_i using our cross-validation metric, because the Whittaker–Henderson smoother itself is a linear operation on the data. As a roughness measure, we use a linear combination of a fourth and second finite difference as



$$\mathbf{D} = \Delta_{\nu}^{(4)} - \frac{1}{2} \Delta_{\nu}^{(2)}. \quad (25)$$

Our choice of roughness measure is largely empirical and based on intuition, it should therefore not be considered as a universal recommendation. A simple second or third difference roughness measure might be sufficient in case no derivative is taken. Here, we take the second derivative of noisy data, therefore we decided to have at least a fourth difference roughness measure so that high-frequency noise is sufficiently suppressed. We found that this tends to slightly oversmooth the data, to combat this we added a second difference to the roughness measure. The second difference is dominant for low frequency components, compared to the fourth difference, reducing oversmoothing. When combining roughness measures, we must take care not to exaggerate specific features, as discussed in section 2.1. In case of combining a fourth and second difference, a resonance can appear in the frequency response of the smoother. To avoid this, we require the sign of the second difference term to be negative. Furthermore, we found that adding a factor one half to the second difference gives better results, leading to the empirically found roughness measure in equation (25).

Figure 4(b) shows the optimal weights w_i that minimize the cross-validation metric. The smoothness parameter is fixed to $\alpha = 1$. The weights w_i are divided into 80 knots, where the spline coefficient of each knot is found using a simple multivariate minimization method, as described in section 4.2. The resulting weights w_i span a range of more than four orders of magnitudes, clearly showing that a single smoothing parameter is not sufficient for smoothing the data. As expected, we find that the weights are highest where the data behaves more erratically. A high weight is needed for the smoothed result $\mathbf{z}$ to more closely follow the input data $\mathbf{y}$. The difference between the smoothed result and the ground truth is shown in figure 4(c). The error is mostly smaller than the standard deviation of the added noise, showing that no oversmoothing is occurring. The amount of high-frequency noise is strongly reduced in the smoothed data, which is important when calculating derivatives from data.

The four panels in figure 5 compare four different ways to determine the group delay dispersion $D_2(\omega)$ from synthetic data with noise presented in figure 4, and thus form a central result of this section. The four panels in appendix figure 14 show the respective residuals between the calculated group delay dispersion and the ground-truth dispersion that are shown in figure 5. Figure 5(a) shows the group delay dispersion that is calculated from noisy synthetic reflectivity data without smoothing. The mean absolute error (MAE) of the dispersion without smoothing compared to the ground truth is

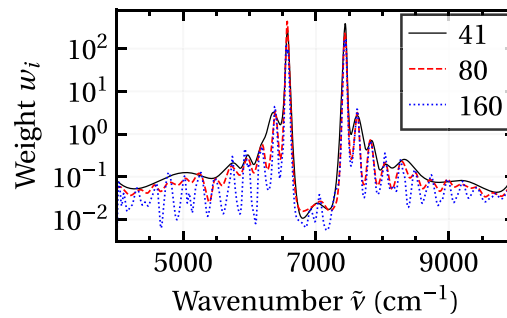


Figure 6. Comparison of the resulting optimal weights w_i for 41, 80 and 160 knots respectively. The smoothness parameter is fixed to $\alpha = 1$.

5259 fs². The large amount of noise makes it impossible to discern all but the strongest features. The first peak at 7432 cm⁻¹ is visible, as is the second at 7600 cm⁻¹ and faintly the third at 7800 cm⁻¹, but all subsequent features are obscured by the noise.

In figure 5(b), the noisy reflectivity data is first filtered by a fourth-order Savitzky–Golay filter with an optimal window length of nine points, before the group delay dispersion is determined. The filter is able to reproduce all features, except the sharpest $D_2(\omega)$ feature at 7432 cm⁻¹, which is underestimated by 14%. However, a single smoothing parameter, the kernel width, does not allow to optimally reduce noise and reproduce sharp features. Furthermore, Savitzky–Golay filters are not good at reducing high-frequency noise [60], which severely hampers the effectiveness of the filter when calculating derivatives from noisy data.

To better assess the performance of the Savitzky–Golay filter, we compare the result of a third-, fourth-, fifth- and sixth-order filter. The optimal window length of each filter is found using generalized cross-validation. All filters give visually very similar results and therefore only the result from the fourth-order filter is plotted. For the third-order filter we find an optimal window length of five points, which gives an MAE of 1982 fs². For the fourth-order filter we find an optimal window length of nine points, which gives an MAE of 1608 fs². For the fifth-order filter we find an optimal window length of nine points, which gives an MAE of 1612 fs². For the sixth-order filter we find an optimal window length of thirteen points, which gives an MAE of 1425 fs². Even higher order filters barely improve the MAE metric and increase ringing and are therefore inadequate.

In figure 5(c), the noisy reflectivity data is first filtered by a normal Whittaker–Henderson smoother with a fourth-order roughness penalty and optimal smoothness parameter $\alpha = 0.110$, before the group delay dispersion is determined. The Whittaker–Henderson smoother is much better at suppressing high-frequency noise than the Savitzky–Golay filter and is therefore better suited when calculating derivatives from noisy data. The resulting smoothed group delay dispersion has much less noise and more accurately follows the peaks and valleys compared to the Savitzky–Golay filter. The filter is also able to reproduce all features, except the sharpest $D_2(\omega)$ feature at 7432 cm⁻¹, which is underestimated by 11%.

To better assess the performance of the Whittaker–Henderson filter, we compare the result of a third-, fourth-, fifth- and sixth-order roughness measure. The optimal smoothing parameter α of each filter is found using generalized cross-validation. All smoothers give visually very similar results and therefore only the result from the smoother with a fourth-order roughness measure is plotted. For the third-order roughness measure we find an optimal smoothing parameter $\alpha \approx 0.118$, which gives an MAE of 1569 fs². For the fourth-order roughness measure we find an optimal smoothing parameter $\alpha \approx 0.110$, which gives an MAE of 1240 fs². For the fifth-order roughness measure we find an optimal smoothing parameter $\alpha \approx 0.085$, which gives an MAE of 1162 fs². For the sixth-order roughness measure we find an optimal smoothing parameter $\alpha \approx 0.061$, which gives an MAE of 1146 fs². For the same order filter, we find that the Whittaker–Henderson filter has significantly better performance than the Savitzky–Golay filter. Even higher order roughness measures barely improve the MAE metric and increase ringing and are therefore inadequate.

In figure 5(d), the noisy reflectivity data is first reconstructed by our weighted Whittaker–Henderson smoothing method, before the group delay dispersion is determined. The continuously variable smoothing weights w_i allow for a good balance between noise reduction and faithful reproduction of the group delay dispersion for both sharp and broad peaks and valleys. To study the effects of the number of knots we compare the results of 41, 80 and 160 knots respectively. The resulting optimal weights w_i for each number of knots are shown in figure 6. The results are visually very similar and therefore only the result of the smoother with 80 knots is plotted.

When using 41 knots we find an MAE of only 278 fs^2 , significantly better than the single parameter Savitzky–Golay filter and Whittaker–Henderson smoother. The root-mean-squared error (RMSE) is 869 fs^2 . All features are accurately reconstructed, however due to the limited degrees of freedom of the weights w_i some undersmoothing occurs near sharp features. The optimization process takes on average about 1.2 s. When using 80 knots we find a slightly better MAE of only 254 fs^2 and an RMSE of 893 fs^2 . This is more than 20 times smaller than the MAE of the dispersion without smoothing. The optimization process for 80 knots takes on average about 1.9 s. When further increasing to 160 knots we find an MAE of 242 fs^2 and an RMSE of 924 fs^2 . The weights w_i now have too many degrees of freedom, leading to oversmoothing, as is apparent from the strongly increased RMSE. The optimization process for 160 knots takes on average about 3.7 s.

Using the noise variances, we do an error analysis of the smoothed dispersion. By calculating the covariance matrix from equation (7), we estimate the unbiased credibility interval of the smoothed complex reflectivity r . We then use error propagation to estimate the credibility interval σ_z of the smoothed group delay dispersion. When using 80 knots, the estimated credibility interval σ_z of the smoothed dispersion for all data points is on average about 41 times smaller than the error of the dispersion without smoothing. We find that 68.9% of all smoothed dispersion data points agree with the ground-truth dispersion within the $1\sigma_z$ credibility interval. Within the $2\sigma_z$ credibility interval we find no less than 95.8% agreement, and for the $3\sigma_z$ credibility interval we find a strikingly high 99.8% agreement. All smoothed dispersion values are within $4\sigma_z$ of the ground-truth dispersion. The agreement percentages are close to the expected values for normally distributed errors, showing a faithful reconstruction from noisy data to a smooth curve, and thus illustrating the power of our method for our original goal.

6. Smoothing unequally spaced data

An important source of difficulties in interpreting data arises when the data points are unequally spaced, as is the case for a significant part of real-world data. Smoothing unequally spaced data often involves changing window widths over the data or resampling data points, which is messy, and prone to artifacts [6]. These issues do not exist for Whittaker–Henderson smoothing, but we have to take two considerations into account. Firstly, it is vital to properly calculate the finite difference coefficients of the matrix $\mathbf{D}$ for each data point in $\mathbf{y}$. Here it is necessary to know the (relative) location of each data point. Secondly, in case that the feature sizes do not scale with the sampling density, we have to make sure that each part of the smooth curve $\mathbf{z}$ contributes equally to the roughness penalty. Therefore, we define a scaled Whittaker–Henderson smoother that is based on minimizing a functional $Q_s[\mathbf{z}]$ defined as

$$Q_s[\mathbf{z}] \equiv \sum_{i=1}^N w_i |y_i - z_i|^2 \Delta_i + \alpha \sum_{i=1}^N |(Dz)_i|^2 \Delta_i, \quad (26)$$

where Δ_i is the step size for each respective data point i . This can be understood as balancing the average squared difference $|y_i - z_i|^2$ against the average roughness $|(Dz)_i|^2$. The smooth result $\mathbf{z}$ minimizing this functional is found by solving

$$(\mathbf{S}\mathbf{W} + \alpha \mathbf{D}^T \mathbf{S} \mathbf{D}) \mathbf{z} = \mathbf{S}\mathbf{W}\mathbf{y}, \quad (27)$$

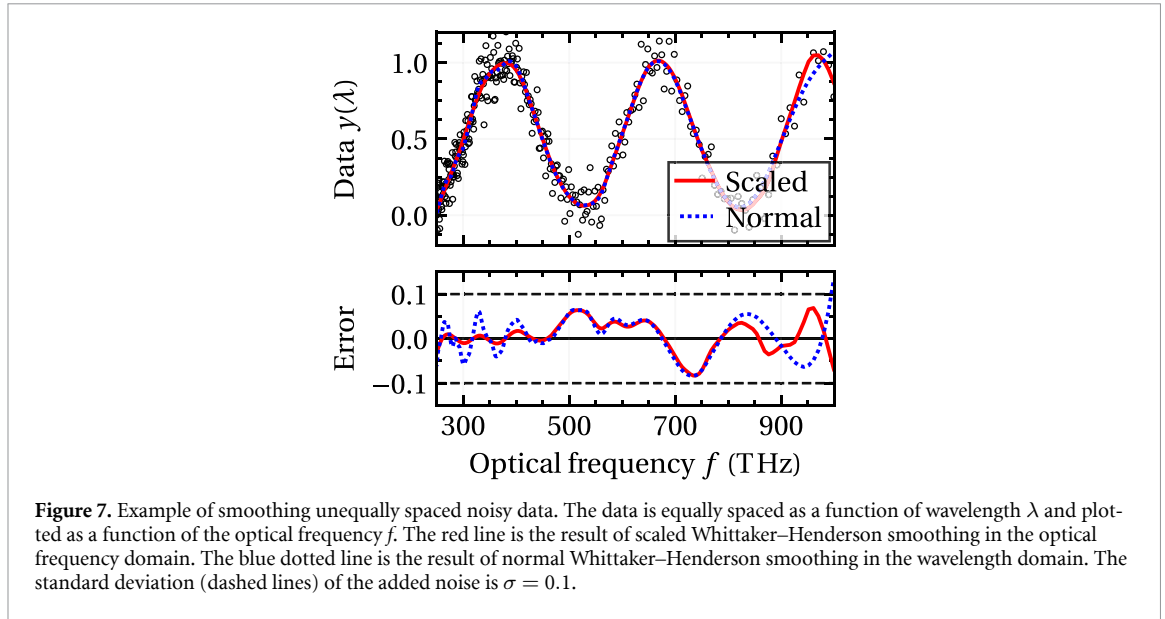
with $\mathbf{S}$ a diagonal scaling matrix consisting of the step sizes Δ_i for all data points,

$$\mathbf{S} \equiv \text{diag}\{\Delta_1, \Delta_2, \dots, \Delta_N\}, \quad (28)$$

and $\mathbf{W}$ the diagonal weight matrix as defined in equation (21). In case all step sizes Δ_i are equal, *i.e.*, for equally spaced data, this reduces back to weighted Whittaker–Henderson smoothing as defined in equation (20). This linear system consists of sparse matrices, like normal Whittaker–Henderson smoothing, and is thus also very efficient to solve. The scaled Whittaker–Henderson smoother can also be used to smoothly interpolate the unequally spaced data onto an equally spaced grid using the weights w_i .

6.1. Mapped unequally spaced data

The first type of unequally spaced data we consider is the situation where the data is equally spaced in one x_c -domain, but where we want to know the features as a function of the unequally spaced x_u -domain. An example is data that are equally sampled in wavelength whereas we want to obtain



information in the optical frequency domain. When the mapping from x_e to x_u is given by the known and smooth function m as

$$x_u = m(x_e), \quad (29)$$

then it is possible to approximate the scaling matrix $\mathbf{S}$ as

$$\mathbf{S} = \text{diag}\{m'(x_{e,1}), m'(x_{e,2}), \dots, m'(x_{e,N})\}. \quad (30)$$

The roughness matrix $\mathbf{D}$ has to be calculated using a roughness penalty in the unequally spaced x_u -domain. For example, a second-order roughness penalty in x_u -domain is expressed as

$$\frac{d^2 z(x_e)}{dx_u^2} = \frac{1}{m'(x_e)^2} \frac{d^2 z(x_e)}{dx_e^2} - \frac{m''(x_e)}{m'(x_e)^3} \frac{dz(x_e)}{dx_e}, \quad (31)$$

using derivatives in the equally spaced x_e -domain.

An example of a real-world application is when we use a spectrometer that measures an optical spectrum in units of nanometers, but we want to know the spectrum in optical frequency units. We then have a mapping from wavelength λ to optical frequency f given by

$$f(\lambda) = c/\lambda. \quad (32)$$

The roughness penalty from equation (31) becomes

$$\frac{d^2 z(\lambda)}{df^2} = \frac{\lambda^4}{c^2} \frac{d^2 z(\lambda)}{d\lambda^2} + \frac{\lambda^3}{c^2} \frac{dz(\lambda)}{d\lambda}, \quad (33)$$

which is approximated using finite differences.

An example is shown in figure 7, where the synthetic data is equally spaced as a function of wavelength, but plotted as a function of optical frequency. The synthetic data is a sine wave with a period of 300 THz, sampled from wavelength $\lambda = 300$ nm to 1200 nm at an interval of $\delta\lambda = 3$ nm. To simulate noisy data, we add normally distributed random numbers with a standard deviation of $\sigma = 0.1$ to the ground-truth data. Both smoothers use a second finite difference as roughness measure, in wavelength units and optical frequency units respectively. When smoothing in wavelength units, using the normal Whittaker–Henderson smoother, we find an optimal smoothing parameter $\alpha_\lambda \approx 1.47 \times 10^4$ and an MAE of 0.030. When smoothing in optical frequency units, using the scaled Whittaker–Henderson smoother, we find an optimal smoothing parameter $\alpha_f \approx 4.24 \times 10^4$ and a much smaller MAE of 0.019. The result from the scaled smoother is equally smooth over the entire optical frequency axis, while the result from the normal smoother shows both over- and undersmoothing. Oversmoothing occurs on the right, sparsely sampled, side of the data and causes the smoothed result to no longer accurately follow the trends of the data. Undersmoothing occurs on the left, densely sampled, side of the data and causes the smoothed result to fluctuate too much due to insufficient noise suppression.

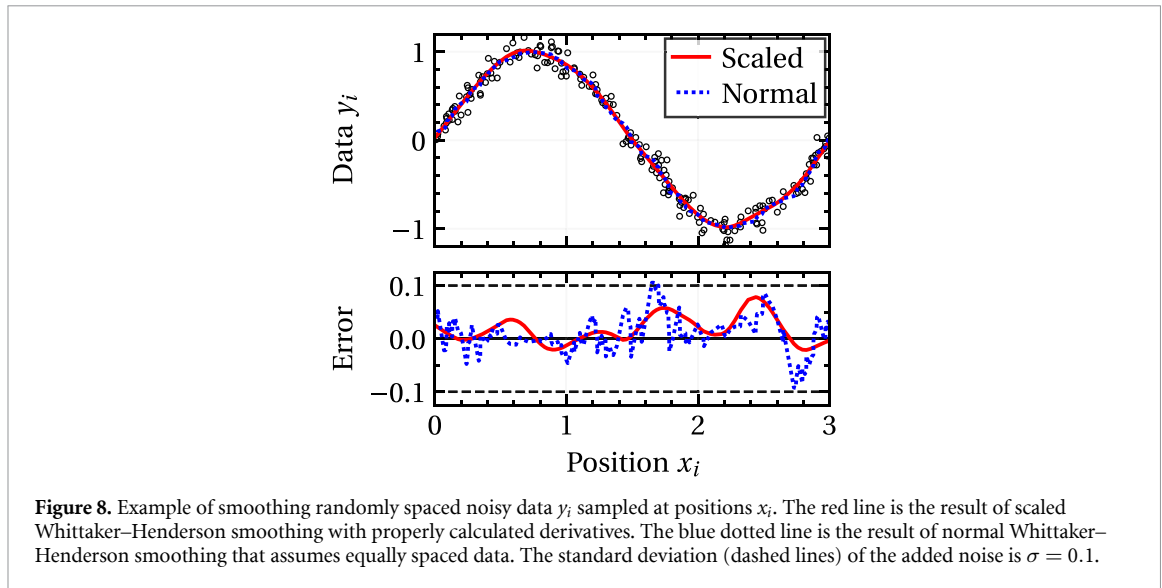


Figure 8. Example of smoothing randomly spaced noisy data y_i sampled at positions x_i . The red line is the result of scaled Whittaker–Henderson smoothing with properly calculated derivatives. The blue dotted line is the result of normal Whittaker–Henderson smoothing that assumes equally spaced data. The standard deviation (dashed lines) of the added noise is $\sigma = 0.1$.

6.2. Randomly spaced data

A second type of unequally spaced data we consider is where the data is sampled at random positions in the x -domain. As a consequence, we have to estimate the finite difference coefficients at each position x_i separately. An approximate form of a first finite difference for arbitrarily spaced data points is

$$z'_i \approx \frac{z_i - z_{i-1}}{x_i - x_{i-1}}, \quad (34)$$

and an approximate second difference form is

$$z''_i \approx \frac{2}{x_{i+1} - x_{i-1}} \left[\frac{z_{i+1} - z_i}{x_{i+1} - x_i} - \frac{z_i - z_{i-1}}{x_i - x_{i-1}} \right]. \quad (35)$$

An example is shown in figure 8. The synthetic data is a single period of a sine wave, on an interval from $x = 0$ to 3, sampled at 200 random positions. To simulate noisy data we add normally distributed random numbers with a standard deviation of $\sigma = 0.1$ to the ground-truth data. For the normal Whittaker–Henderson smoother we find an optimal smoothing parameter $\alpha \approx 5.10 \times 10^{-5}$ and an MAE of 0.027. For the scaled Whittaker–Henderson smoother we find an optimal smoothing parameter $\alpha \approx 8.47 \times 10^{-5}$ and a slightly smaller MAE of 0.023. Both smoothers use a second finite difference as roughness measure. The result from the scaled smoother is equally smooth over the entire x -axis, while the result from the normal smoother shows many sudden steps. These sudden steps are due to data points being very close together, causing a local underestimation in the roughness metric for normal Whittaker–Henderson smoothing. This simple fact makes it impossible for the normal smoother to make a smooth curve from randomly sampled data. Care has to be taken when two data points are extremely close together compared to other data points, as this can lead to numerical instabilities.

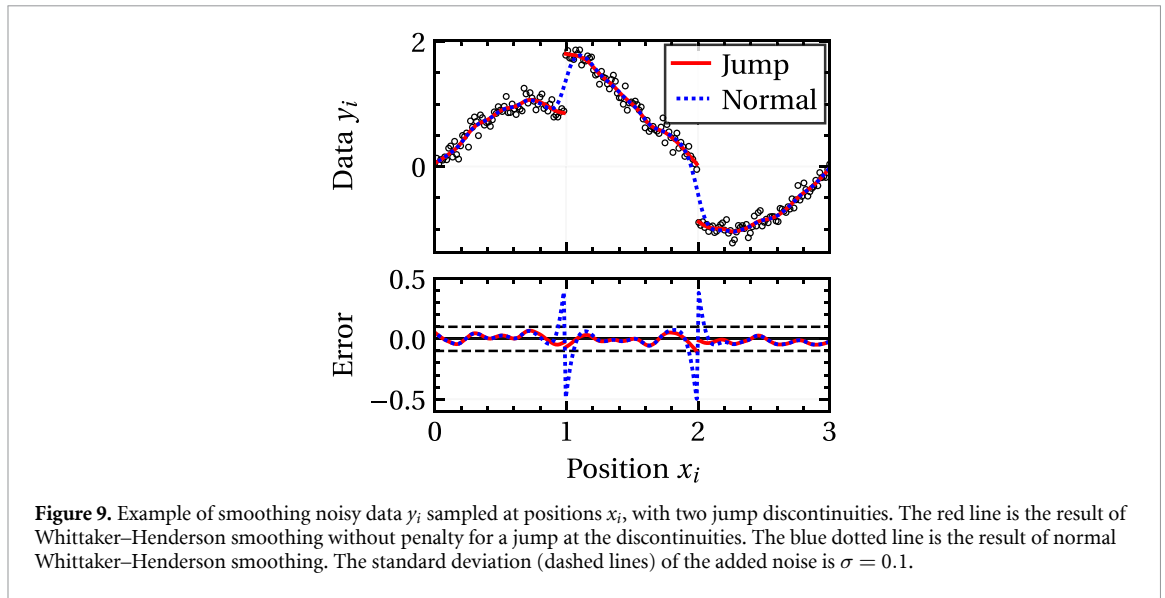
7. Smoothing discontinuous data

A major challenge in data analysis occurs when discontinuities exist, either in the original data $\mathbf{y}$ or in its derivatives. Discontinuities are associated with strong high-frequency components, which makes them exceedingly difficult to properly smooth. Simply splitting the data in parts at a discontinuity does not take advantage of information on both sides of the discontinuity and leads to large errors. In this section we distinguish two important types of discontinuities, but many more kinds can be implemented using our smoother.

7.1. Continuous first derivative: Jumps

A first type of discontinuity is when there is a jump in the data at some point x_d , whereas the derivative is continuous, *i.e.*, when

$$y(x_d^-) \neq y(x_d^+) \quad (36)$$



and

$$y'(x_d^-) = y'(x_d^+). \quad (37)$$

The information on the existence of a jump at x_d can be included in the roughness matrix $\mathbf{D}$. An example of a second finite difference matrix with jump discontinuity is

$$\mathbf{D} = \frac{1}{\Delta^2} \left[\begin{array}{cccc|cccc} 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{yellow } -\frac{1}{2} & \text{yellow } \frac{1}{2} & \frac{1}{2} & \text{yellow } -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \text{yellow } -\frac{1}{2} & \text{yellow } \frac{1}{2} & \frac{1}{2} & \text{yellow } -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{array} \right]. \quad (38)$$

At the yellow highlighted cells, the second finite difference is replaced by a different kind of roughness measure. This roughness measure has a penalty when the slopes y' at both sides of the discontinuity differ, yet there is no penalty when the values y differ at both sides of the discontinuity.

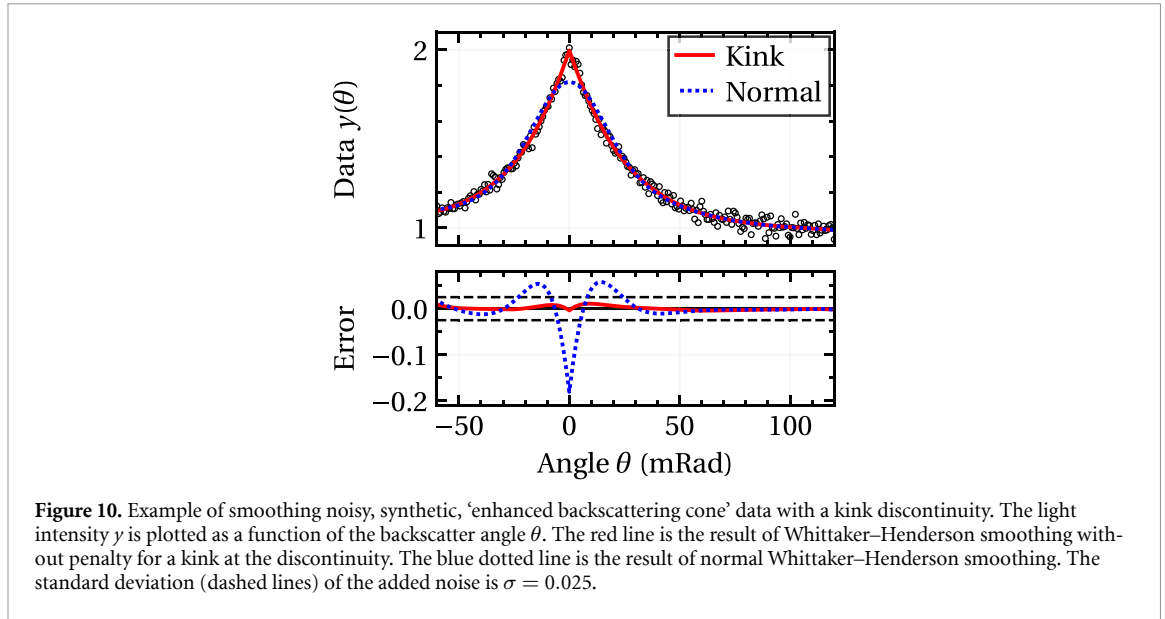
A comparison of two equivalent Whittaker–Henderson smoothers, one with jump discontinuities and one without, is shown in figure 9. The synthetic data is a single period of a sine wave, on an interval from $x = 0$ to 3, sampled on a uniform grid of 200 data points. At the index $x = 1$ the data jumps up by one, at the index $x = 2$ the data jumps back down by one. To simulate noisy data, we add normally distributed random numbers with a standard deviation of $\sigma = 0.1$ to the ground-truth data. For the Whittaker–Henderson smoother with jump discontinuity we find an optimal smoothing parameter $\alpha \approx 3.86 \times 10^{-6}$. For the normal Whittaker–Henderson smoother we use the same smoothing parameter, to compare the effects of adding the jump discontinuity. Both smoothers use a second finite difference as roughness measure. The smoother with discontinuities faithfully reproduces the true jumps in the curve. The error stays within the standard deviation of the added noise and is independent of the size of the jump. The smoother without discontinuity removes all high frequencies from the jumps, which turns the jumps into smooth slopes, and thus loses the essential features, namely the discontinuities. The error is much larger than the standard deviation of the added noise, and scales with the size of the jump.

7.2. Discontinuous first derivative: Enhanced backscattering cone

A second type of discontinuity is when the first derivative is discontinuous at some point x_d , whereas the curve itself is continuous, *i.e.*, when

$$y(x_d^-) = y(x_d^+) \quad (39)$$

and



$$y'(x_d^-) \neq y'(x_d^+). \quad (40)$$

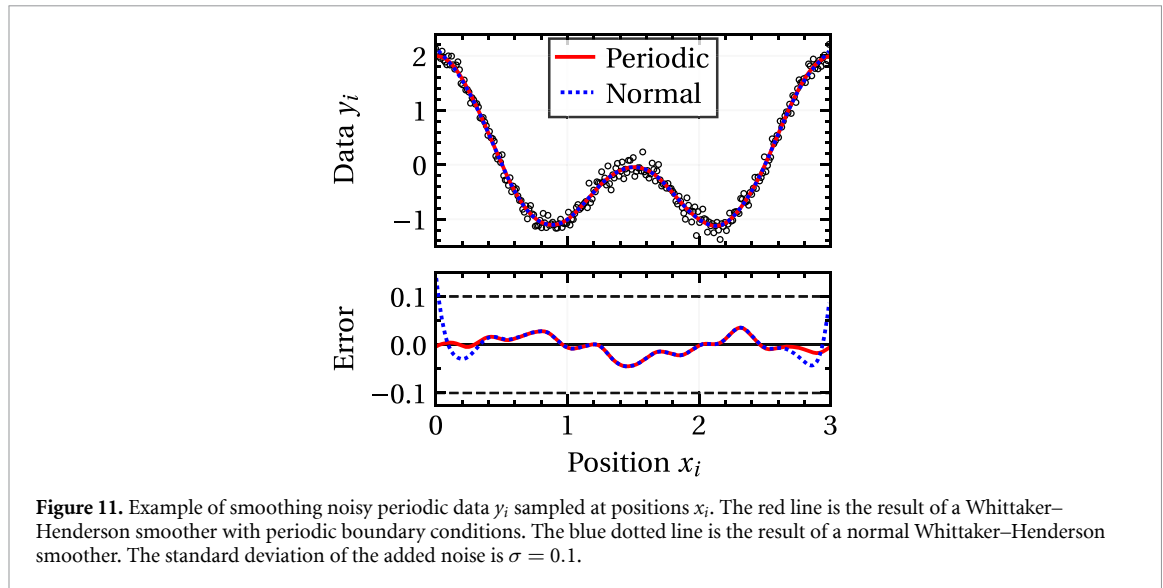
This situation is also known as a kink and a practical situation where this occurs is at the center of an enhanced backscattering cone [62]. Enhanced backscattering is a remarkable interference phenomenon that occurs in multiple scattering of light and other waves, where counterpropagating paths constructively interfere, leading to an intensity enhancement into the exact backscattering direction up to a factor two, compared to the diffuse background intensity. Moving away from the exact backscattering direction, the phases between the counterpropagating paths decohere, leading to a gradual decrease in enhancement, hence why it is named a cone. In the absence of absorption, the backscattering cone decays approximately exponentially as a function of the backscattering angle. The cone is then not analytic at the exact backscattering angle but is a cusp.

The information of the existence of a kink at x_d can be included in the roughness matrix $\mathbf{D}$. An example of a second finite difference matrix with kink discontinuity is

$$\mathbf{D} = \frac{1}{\Delta^2} \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}. \quad (41)$$

The yellow highlighted cells are set to zero, instead of to the second finite difference coefficients. This means that there is no penalty for smoothness at the position of the discontinuity, and a kink is free to appear in $\mathbf{z}$. Since the curve surrounding the kink has to be smooth, a jump in $\mathbf{z}$ is not allowed to appear without a penalty.

A comparison of two equivalent Whittaker–Henderson smoothers, one with a kink discontinuity and one without kink, are shown in figure 10. The synthetic data represents ‘enhanced backscattering cone’ data [63, 64] and is a symmetrically decaying exponential with background. It is sampled on an interval from $\theta = -60$ mrad to 120 mrad on a uniform grid of 226 data points. At the angle $\theta = 0$ mrad a cusp is present. To simulate noisy data, we add normally distributed random numbers with a standard deviation of $\sigma = 0.025$ mrad to the ground-truth data. For the Whittaker–Henderson smoother with kink discontinuity we find an optimal smoothing parameter $\alpha \approx 1.54 \times 10^4$. For the normal Whittaker–Henderson smoother we use the same smoothing parameter, to compare the effects of adding the kink discontinuity. Both smoothers use a third finite difference as roughness measure. The smoother with kink discontinuity faithfully reproduces the noiseless true data, including the cusp. The normal Whittaker–Henderson smoother simply removes all high-frequency components, regardless of the cusp, and is thus unable to reproduce the noiseless true data.



8. Smoothing with boundary conditions

An omnipresent yet often neglected challenge is the smoothing of a data set at the edges of the domain. The abrupt absence of data outside the domain often leads to large artifacts. Boundary conditions can be naturally included in Whittaker–Henderson smoothers, leading to significantly reduced errors.

8.1. Periodic boundary conditions

One of the simplest kinds of boundary conditions is a periodic boundary condition. This poses no issues for normal convolution-based filters, as the data is simply repeated. For Whittaker–Henderson smoothers it is only a matter of defining a roughness penalty that connects one side of the data to the other side. An example of a second finite difference matrix with a periodic boundary condition is

$$\mathbf{D} = \frac{1}{\Delta^2} \begin{bmatrix} -2 & 1 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 1 & -2 \end{bmatrix}, \quad (42)$$

where Δ is the finite difference step size. The yellow highlighted cells show where the finite difference coefficients wrap around from one side of the $\mathbf{D}$ matrix to the other, which enforces the periodicity.

An example of smoothing with periodic boundary conditions is shown in figure 11. We compare a Whittaker–Henderson smoother without boundary conditions to an otherwise equivalent Whittaker–Henderson smoother with periodic boundary conditions. Both smoothers use a second finite difference, similar to equation (42), as roughness measure. For the Whittaker–Henderson smoother with periodic boundary conditions we find an optimal smoothing parameter $\alpha \approx 3.34 \times 10^{-5}$. For the normal Whittaker–Henderson smoother we use the same smoothing parameter, to compare the effects of adding the boundary condition. The errors of the smoother with periodic boundary conditions are about the same order of magnitude everywhere, and it does not suffer from boundary artifacts. The smoother without boundary conditions, on the other hand, has significantly larger errors, and is also not continuous at the boundaries.

8.2. Dirichlet boundary conditions

When a precise value at a boundary is known, we enforce this condition using a Dirichlet boundary condition. Usual convolution filters apply this boundary condition by padding the edges of the data with chunks of the known value. This is not necessary for (weighted) Whittaker–Henderson smoothers. As an example, we consider the case of a weighted Whittaker–Henderson smoother where it is known that $z_1 = a$. We then have to solve a linear system defined as

$$\begin{bmatrix} 1 & 0 & 0 & \dots \\ c_{21} & c_{22} & c_{23} & \dots \\ c_{31} & c_{32} & c_{33} & \dots \\ c_{41} & c_{42} & c_{43} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} a \\ w_2 y_2 \\ w_3 y_3 \\ w_4 y_4 \\ \vdots \end{bmatrix}, \quad (43)$$

where the c_{ij} elements are given by the $\mathbf{C}$ matrix, defined as

$$\mathbf{C} = \mathbf{W} + \alpha \mathbf{D}^T \mathbf{D}. \quad (44)$$

This means that this boundary condition only requires us to change the first row of the matrix $\mathbf{C}$ and vector $\mathbf{W}\mathbf{y}$ and is thus just as efficient to solve as the weighted Whittaker–Henderson smoother without boundary conditions. Solving this system is equivalent to finding the smooth curve $\mathbf{z}$ minimizing the functional $Q[\mathbf{z}]$, while enforcing that $z_1 = a$.

8.3. Neumann boundary conditions

When the slope of the data at a boundary is known, we enforce this condition using a Neumann boundary condition. This boundary condition is straightforward to implement for Whittaker–Henderson smoothers. As an example, we consider the case of a weighted Whittaker–Henderson smoother where it is known that $z'_1 = b$. The derivative at the left boundary is approximated using a second-order forward finite difference. We then have to solve a linear system defined as

$$\begin{bmatrix} \frac{-3}{2\Delta} & \frac{2}{\Delta} & \frac{-1}{2\Delta} & 0 & \dots \\ c_{21} & c_{22} & c_{23} & c_{24} & \dots \\ c_{31} & c_{32} & c_{33} & c_{34} & \dots \\ c_{41} & c_{42} & c_{43} & c_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} a \\ w_2 y_2 \\ w_3 y_3 \\ w_4 y_4 \\ \vdots \end{bmatrix}, \quad (45)$$

where the c_{ij} elements are given by the $\mathbf{C}$ matrix as defined in equation (44) and Δ is the finite difference step size. This has the disadvantage that we lose the data point y_1 , yet to solve this we simply add temporary data points y_0 and z_0 which are removed again after solving the linear system.

8.4. Cauchy boundary conditions

When both the precise value and the slope of the data are known at a boundary, we enforce this condition using a Cauchy boundary condition. Even if these values are not known beforehand, we can still find an approximation by fitting a polynomial through a number of data points at the boundary. This hybrid approach makes smoothing at the edges of data sets much more robust and reduces artifacts. As an example, we consider the case of a weighted Whittaker–Henderson smoother where it is known that $z_1 = a$ and $z'_1 = b$. The derivative at the left boundary is approximated using a first-order forward finite difference. We then have to solve a linear system defined as

$$\begin{bmatrix} 1 & 0 & 0 & \dots \\ \frac{1}{\Delta} & \frac{-1}{\Delta} & 0 & \dots \\ c_{31} & c_{32} & c_{33} & \dots \\ c_{41} & c_{42} & c_{43} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} a \\ b \\ w_3 y_3 \\ w_4 y_4 \\ \vdots \end{bmatrix}, \quad (46)$$

where the c_{ij} elements are given by the $\mathbf{C}$ matrix as defined in equation (44) and Δ is the finite difference step size.

An example of smoothing with Cauchy boundary conditions is shown in figure 12. Here we compare a Whittaker–Henderson smoother with Cauchy boundary conditions to an otherwise equivalent Whittaker–Henderson smoother without boundary conditions. The synthetic data is a chirped sine wave, on an interval from $x = 0$ to 2.05, sampled on a uniform grid of 500 data points. At the right edge of the data the downward trend abruptly changes in an upward trend, which usually causes issues with smoothing. To simulate noisy data, we add normally distributed random numbers with a standard deviation of $\sigma = 0.1$ to the ground-truth data. For the Whittaker–Henderson smoother with Cauchy boundary conditions we find an optimal smoothing parameter $\alpha \approx 3.86 \times 10^{-6}$. For the normal Whittaker–Henderson smoother we use the same smoothing parameter, to compare the effects of adding the boundary condition. Both smoothers use a second finite difference as roughness measure.

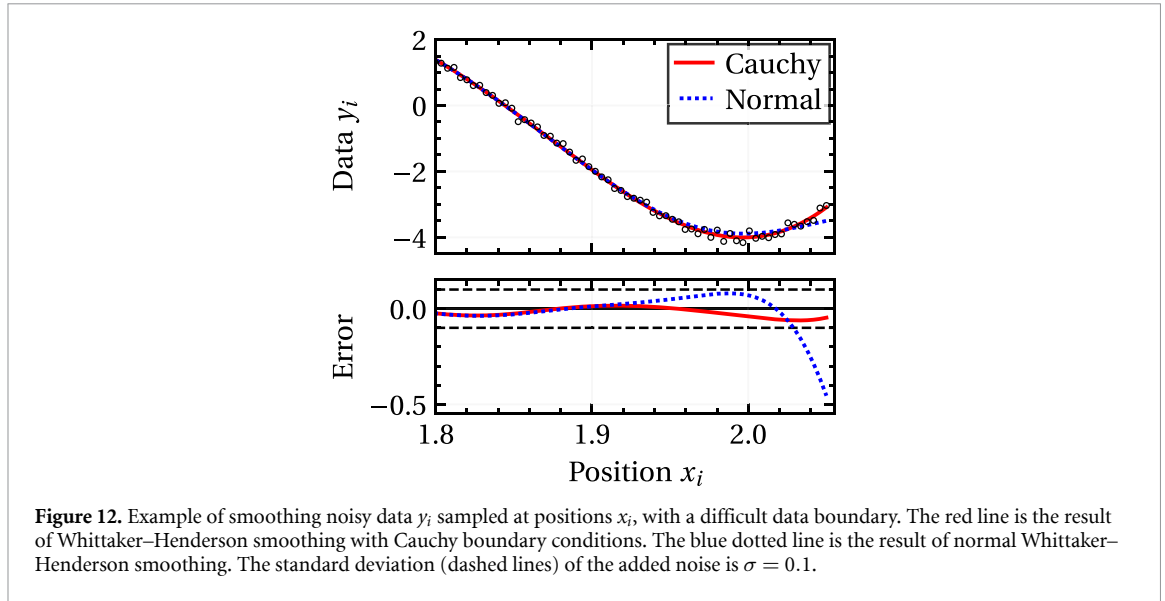


Figure 12. Example of smoothing noisy data y_i sampled at positions x_i , with a difficult data boundary. The red line is the result of Whittaker–Henderson smoothing with Cauchy boundary conditions. The blue dotted line is the result of normal Whittaker–Henderson smoothing. The standard deviation (dashed lines) of the added noise is $\sigma = 0.1$.

The Whittaker–Henderson smoother without boundary conditions suffers from boundary artifacts and is not able to accurately follow the data. For the Whittaker–Henderson smoother with Cauchy boundary conditions we find the right boundary value and slope by fitting a second-order polynomial through the last 25 data points. This forces the smoothed curve to follow the data more closely at the boundaries. The error at the boundaries is then given by the error of the fit, as opposed to the error of the smoother. This leads to significantly smaller errors, as shown by the error in figure 12.

9. Multidimensional smoothing

Whittaker–Henderson smoothing is readily extended to multidimensional data [65], such as for instance multidimensional spectral data collected with an array detector (CCD, CMOS sensor). We have to flatten the multidimensional data into a 1D array and calculate a roughness matrix $\mathbf{D}$ that encodes the multidimensional roughness measure. For simplicity we consider two-dimensional (2D) data $Y_{i,j}$ sampled on a rectangular grid of N_y by N_x points respectively. As a roughness measure, we take the Laplace operator, which is approximated using finite differences as

$$(\nabla^2 Y)_{i,j} \approx \frac{Y_{i,j+1} - 2Y_{i,j} + Y_{i,j-1}}{\Delta_x^2} + \frac{Y_{i+1,j} - 2Y_{i,j} + Y_{i-1,j}}{\Delta_y^2}, \quad (47)$$

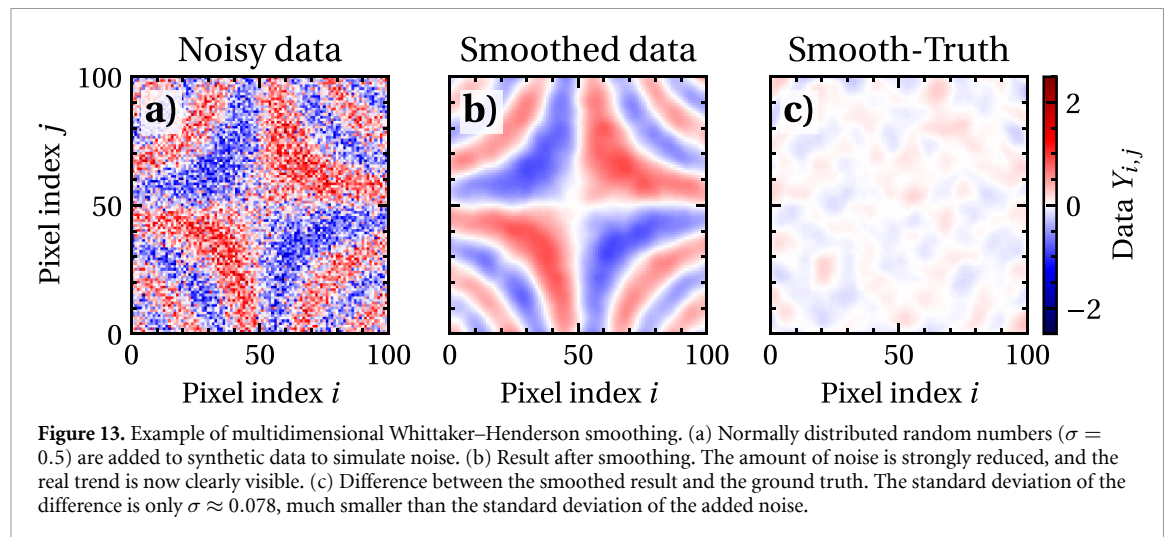
where Δ_x and Δ_y are the finite difference step sizes in the x and y direction respectively. The data is flattened into a one-dimensional (1D) array according to

$$y_{(iN_x+j)} \equiv Y_{i,j}. \quad (48)$$

Next, we calculate the elements of the roughness matrix $\mathbf{D}$ such that $(\mathbf{D}\mathbf{y})_{(iN_x+j)}$ is equivalent to the Laplace operator on $Y_{i,j}$ as defined in equation (47). By transforming the 2D problem into a 1D problem, we simply have to solve a linear system of the exact same form as the normal Whittaker–Henderson smoother defined in equation (2). Since this involves sparse matrices, this is still relatively efficient to solve. Weighted or scaled Whittaker–Henderson smoothing can also be used, depending on the problem at hand, see the previous sections. Finally, we transform the flattened 1D data back to the 2D data, according to equation (48).

An example of 2D smoothing is shown in figure 13. We have a grid of a 101 by 101 data points with synthetic data to which we added normally distributed noise with a standard deviation $\sigma = 0.5$. The function describing the synthetic data is

$$Y_{i,j} = \frac{\sin\left(15\left(\frac{i}{50} - \frac{1}{2}\right)\left(\frac{j}{50} - \frac{1}{2}\right)\right)}{1 + \left(\frac{i}{50} - \frac{1}{2}\right)^2 + \left(\frac{j}{50} - \frac{1}{2}\right)^2}. \quad (49)$$



The noisy data in figure 13(a) is then smoothed by a normal Whittaker–Henderson smoother with the Laplacian operator as roughness measure, as defined in equation (47). Using generalized cross-validation we find an optimal smoothing parameter $\alpha \approx 18.4$ that best balances the features against the noise. Finally, we calculate the difference between the smoothed data and the true data without noise, which is shown in figure 13(c). The standard deviation of the difference is $\sigma \approx 0.078$, which is more than six times smaller than the standard deviation of the added noise, therefore we accurately resolve the underlying true data from the noise. Multidimensional Whittaker–Henderson smoothing can also be applied to unequally spaced meshes. For this problem the hardest part is to calculate the elements of the roughness matrix $\mathbf{D}$ for the given mesh, but this is beyond the scope of this paper.

10. Discussion

In this paper we have introduced a weighted Whittaker–Henderson smoother with cross-validation. The continually variable weights allow to optimally balance the spectral features against the noise. Furthermore, Whittaker–Henderson smoothing does not use a moving window, or knots, to do smoothing. This leads to less problems at the edges of the data and leads to smaller artifacts in higher derivatives of smoothed data. Of course, our method has several limitations. Firstly, our method is computationally relatively expensive, which deters its use for large data volumes. The solution may lie in better minimization algorithms, faster linear solvers, or methods for describing the weights w_i that do not rely on interpolating splines. Secondly, there is a need for accurate noise variance to do error analysis. Thirdly, there is always a risk of under- or over-adaptation of the weights. This in turn leads to under- or oversmoothing respectively, and gives wrong credibility intervals. Therefore, it is vital to always verify results and to never blindly trust the results of any smoother.

For our synthetic data, we assumed independent, homoscedastic, normally distributed and stationary noise, which is the optimal case for linear (penalized) least-squares smoothers. On the contrary, real data can have outliers, require calibration, or have correlated, non-normal noise. Outliers are a common problem and causes large errors if they are not taken care off. If a way of detecting outliers is available, then we can, for example, use weighted Whittaker–Henderson smoothing to completely ignore outliers [66]. Whenever noise is correlated, two considerations must be taken. Firstly, the bandwidth of the smoothed result is limited by the noise bandwidth. This means that features past, or close to, the noise bandwidth will no longer be resolvable. Secondly, the statistical tests, such as cross-validation, used to verify whether the smooth is correct have to take the noise correlations into account.

Whenever the noise is heteroscedastic or non-stationary, weighted Whittaker–Henderson smoothing should be used, as discussed in section 4. Whittaker–Henderson smoothing, like other linear (penalized) least-squares smoothers, works optimally for smoothing normally distributed noise. Somewhat non-normally distributed errors are generally not a problem, as long as the errors have mean zero, but will lead to less-than-optimal smoothness. In case that noise is highly non-normally distributed, such as for binomial data [67], it is necessary to use a different smoothing formalism. Therefore, future work on a more robust smoother that takes these real features into account is highly beneficial.

11. Conclusions

In this article we study a frequently occurring challenge in experimental and numerical observations, namely, how to resolve characteristic features, such as spectral peaks, and derivatives from measured data with unavoidable noise. Therefore, we have developed a modified Whittaker–Henderson smoothing procedure that balances the spectral features and the noise. In this surprisingly little-known procedure, our contribution is to introduce adjustable weights that are optimized using cross-validation. Using the measurement errors, a straightforward error analysis of the smoothed results is feasible.

As an example, we calculate the optical group delay dispersion of a Bragg reflector from synthetic phase data with noise to illustrate the effectiveness of the smoothing algorithm. The calculated credibility interval σ_z of the smoothed dispersion is on average almost 41 times smaller than the standard deviation of the noise without smoothing. We find that 68.9% of all smoothed dispersion data points agree with the ground-truth dispersion within the $1\sigma_z$ credibility interval. Within the $2\sigma_z$ credibility interval we find no less than 95.8% agreement. Thus, the smoother faithfully reconstructs the group delay dispersion, allowing to observe details that otherwise remain buried in noise.

We recommend our weighted Whittaker–Henderson smoother with cross-validation whenever a single smoothing parameter does not suffice due to a large range of feature widths. To further illustrate the power of our smoother, we demonstrate our smoother with several commonly occurring difficulties in data and data analysis. We recommend our general framework of Whittaker–Henderson smoothing to properly smoothen unequally sampled data, and to obtain discontinuities, including discontinuous derivatives or kinks, and to properly smooth data in the vicinity of boundaries to the measurement domains.

Acknowledgments

We thank Lyuba Amitonova for stimulating discussions. This work is supported by the Dutch Research Council NWO-TTW Perspectief program P21-20 ‘Optical coherence; optimal delivery and positioning’ (OPTIC) in collaboration with TUE, TUD, and ARCNL, and with industrial partners Anteryon, ASML, Demcon, JMO, Signify, and TNO.

Data availability statement

The data that support the findings of this study (a collection of seven Python Notebooks) are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.17131152> [68].

Conflict of interest

The authors have no conflicts to disclose.

Author contributions

Bert Mulder  0009-0003-3576-5767

Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Project administration (equal), Software (equal), Visualization (equal), Writing – original draft (equal)

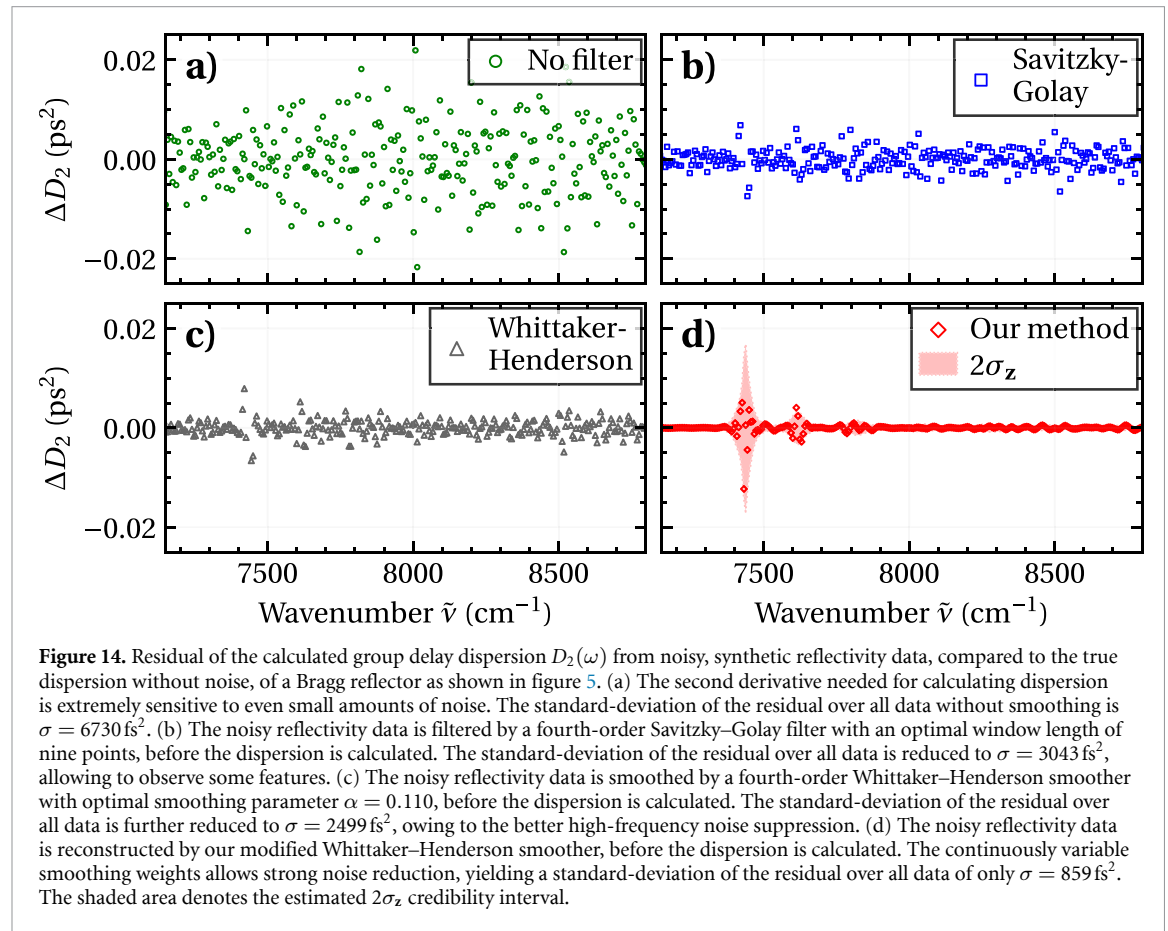
Ad Lagendijk  0000-0002-9165-8128

Supervision (supporting), Validation (supporting)

Willem L Vos  0000-0003-3066-859X

Funding acquisition (equal), Resources (equal), Supervision (lead), Validation (lead), Writing – review & editing (equal)

Appendix



References

- [1] Sussman R 2016 *Observational Methods: The First Step in Science* (Wiley) ch 2, p 9–27
- [2] Shapere D 1982 The concept of observation in science and philosophy *Phil. Sci.* **49** 485
- [3] Horowitz P and Hill W 1983 *The Art of Electronics* 1st edn (Cambridge University Press)
- [4] Goodman J W 2000 *Statistical Optics* (Wiley)
- [5] Orfanidis S J 2007 *Optimum Signal Processing: An Introduction* 2nd edn (Rutgers University)
- [6] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 *Numerical Recipes: The Art of Scientific Computing* 3rd edn (Cambridge University Press)
- [7] Altman G D, Machin D, Bryant T N and Gardner M J 2013 *Statistics With Confidence: Confidence Intervals and Statistical Guidelines* 2nd edn (Wiley)
- [8] Wiener N 1949 *Extrapolation, Interpolation and Smoothing of Stationary Time Series: With Engineering Applications* (The MIT Press) (<https://doi.org/10.7551/mitpress/2946.001.0001>)
- [9] Einicke G A 2012 *Smoothing, Filtering and Prediction: Estimating The Past, Present and Future* 1st edn (INTECH d.o.o.) (<https://doi.org/10.5772/2706>)
- [10] Buja A, Hastie T and Tibshirani R 1989 Linear smoothers and additive models *Ann. Stat.* **17** 453
- [11] Yee T W 2015 *Vector Generalized Linear and Additive Models: With an Implementation in R* 1st edn (Springer) (<https://doi.org/10.1007/978-1-4939-2818-7>)
- [12] Eilers P H C and Marx B D 1996 Flexible smoothing with b-splines and penalties *Stat. Sci.* **11** 89
- [13] de Boor C 2001 *A Practical Guide to Splines vol 27 of Applied Mathematical Sciences* rev. edn (Springer)
- [14] Cleveland W S and Devlin S J 1988 Locally weighted regression: An approach to regression analysis by local fitting *J. Am. Stat. Assoc.* **83** 596
- [15] Cleveland W S 1979 Robust locally weighted regression and smoothing scatterplots *J. Am. Stat. Assoc.* **74** 829
- [16] Savitzky A and Golay M J E 1964 Smoothing and differentiation of data by simplified least squares procedures *Anal. Chem.* **36** 1627
- [17] Whittaker E T 1922 On a new method of graduation *Proc. Edinburgh Math. Soc.* **41** 63
- [18] Henderson R 1924 A new method of graduation *Trans. Actuarial Soc. Am.* **25** 29
- [19] Eilers P H C 2003 A perfect smoother *Anal. Chem.* **75** 3631
- [20] Biessy G 2025 Whittaker–Henderson smoothing revisited: A modern statistical framework for practical use *ASTIN Bull.* **56** 1
- [21] Demtröder W 2002 *Laser Spectroscopy: Concepts and Instrumentation* 3rd edn (Springer)

- [22] Novotny L and Hecht B 2012 *Principles of Nano-Optics* 2nd edn (Cambridge University Press)
- [23] Imhof A, Vos W L, Sprik R and Lagendijk A 1999 Large dispersive effects near the band edges of photonic crystals *Phys. Rev. Lett.* **83** 2942
- [24] Gersen H, Karle T J, Engelen R J P, Bogaerts W, Korterik J P, van Hulst N F, Krauss T F and Kuipers L 2005 Real-space observation of ultraslow light in photonic crystal waveguides *Phys. Rev. Lett.* **94** 073903
- [25] Krauss T F 2008 Why do we need slow light? *Nat. Photon.* **2** 448
- [26] Baba T 2008 Slow light in photonic crystals *Nat. Photon.* **2** 465
- [27] Polak A G 2001 Indirect measurements: combining parameter selection with ridge regression *Meas. Sci. Technol.* **12** 278
- [28] Engl H W, Hanke M and Neubauer A 2000 *Regularization of Inverse Problems vol 375 of Mathematics and Its Applications* 1st edn (Kluwer Academic)
- [29] Candès E J, Romberg J K and Tao T 2006 Stable signal recovery from incomplete and inaccurate measurements *Commun. Pure Appl. Math.* **59** 1207
- [30] Caravaca-Aguirre A M et al 2023 Single-pixel photoacoustic microscopy with speckle illumination *Intell. Comput.* **2** 0011
- [31] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc. B* **58** 267
- [32] Simonoff J S 1996 *Smoothing Methods in Statistics Springer Series in Statistics* (Springer New York) (<https://doi.org/10.1007/978-1-4612-4026-6>)
- [33] Bohlmann G 1899 Ein ausgleichungsproblem Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse 260 (available at: <http://eudml.org/doc/58455>)
- [34] Hodrick R J and Prescott E C 1997 Postwar U.S. business cycles: An empirical investigation *J. Money Credit Bank.* **29** 1
- [35] Wood S N 2006 *Generalized Additive Models, An Introduction With R* 1st edn (CRC Press)
- [36] Marra G and Wood S N 2012 Coverage properties of confidence intervals for generalized additive model components *Scand. J. Stat.* **39** 53
- [37] Smith III J O 2011 *Spectral Audio Signal Processing* (W3K Publishing) (available at: <https://ccrma.stanford.edu/jos/sasp/>)
- [38] Silverman B W 1985 Some aspects of the spline smoothing approach to non-parametric regression curve fitting *J. R. Stat. Soc. B* **47** 1
- [39] Weinert H L 2007 Efficient computation for Whittaker-Henderson smoothing *Comput. Stat. Data Anal.* **52** 959
- [40] Golub G H, Heath M and Wahba G 1979 Generalized cross-validation as a method for choosing a good ridge parameter *Technometrics* **21** 215
- [41] Lukas M A 2006 Robust generalized cross-validation for choosing the regularization parameter *Inverse Probl.* **22** 1883
- [42] Nocon A S and Scott W F 2012 An extension of the Whittaker-Henderson method of graduation *Scand. Actuarial J.* **2012** 70
- [43] Nowak R 1997 Optimal signal estimation using cross-validation *IEEE Signal Process. Lett.* **4** 23
- [44] Urbas A A and Choquette S J 2011 Automated spectral smoothing with spatially adaptive penalized least squares *Appl. Spectrosc.* **65** 665
- [45] Li K, Zhang Y, Li Y and Liu B 2021 Local penalized least squares combined with the segment similarity for ecg denoising *Circuits Syst. Signal Process.* **41** 532
- [46] Gosteva A, Haiml M, Paschotta R and Keller U 2005 Noise-related resolution limit of dispersion measurements with white-light interferometers *J. Opt. Soc. Am. B* **22** 1868
- [47] Fowles G R 1989 *Introduction to Modern Optics Dover Books on Physics* 2nd edn (Dover)
- [48] Euser T G, Harding P J and Vos W L 2009 Broadband sensitive pump-probe setup for ultrafast optical switching of photonic nanostructures and semiconductors *Rev. Sci. Instrum.* **80** 073104:1
- [49] Yüce E, Ctistis G, Claudon J, Gérard J M and Vos W L 2016 Optimal all-optical switching of a microcavity resonance in the telecom range using the electronic Kerr effect *Opt. Express* **24** 239
- [50] Granata M et al 2020 Amorphous optical coatings of present gravitational-wave interferometers *Class. Quantum Grav.* **37** 095004
- [51] Singhapong W, Bowen C, Wang H, Sawhney K and Lunt A J G 2024 Multilayer optics for synchrotron applications *Adv. Mater. Technol.* **9** 2302187
- [52] Born M and Wolf E 1986 *Principles of Optics* 6th edn (Pergamon Ltd.)
- [53] Siegman A E 1986 *Lasers* 1st edn (University Science Books)
- [54] Jackson J D 1998 *Classical Electrodynamics* 3rd edn (Wiley)
- [55] Boyd R W 2008 *Nonlinear Optics* 3rd edn (Elsevier)
- [56] Diels J C and Rudolph W 2006 *Ultrashort Laser Pulse Phenomena* 2nd edn (Academic) (<https://doi.org/10.1016/b978-0-12-215493-5.x5000-9>)
- [57] Pickering J D 2021 *Dispersion* (IOP Publishing) ch 3, p 3–1 to 3–15
- [58] Diddams S and Diels J C 1996 Dispersion measurements with white-light interferometry *J. Opt. Soc. Am. B* **13** 1120
- [59] Naganuma K, Mogi K and Yamada H 1990 Group-delay measurement using the Fourier transform of an interferometric cross correlation generated by white light *Opt. Lett.* **15** 393
- [60] Schmid M, Rath D and Diebold U 2022 Why and how Savitzky-Golay filters should be replaced *ACS Meas. Sci. Au* **2** 185
- [61] Yeh P, Yariv A and Hong C S 1977 Electromagnetic propagation in periodic stratified media. I. General theory *J. Opt. Soc. Am.* **67** 423
- [62] Wiersma D S, van Albada M P and Lagendijk A 1995 An accurate technique to record the angular distribution of backscattered light *Rev. Sci. Instrum.* **66** 5473
- [63] van Albada M P and Lagendijk A 1985 Observation of weak localization of light in a random medium *Phys. Rev. Lett.* **55** 2692
- [64] Wolf P E and Maret G 1985 Weak localization and coherent backscattering of photons in disordered media *Phys. Rev. Lett.* **55** 2696
- [65] Moulart R and Rotinat R 2014 On the use of a penalized least squares method to process kinematic full-field measurements *Meas. Sci. Technol.* **25** 075001
- [66] Garcia D 2010 Robust smoothing of gridded data in one and higher dimensions with missing values *Comput. Stat. Data Anal.* **54** 1167
- [67] Henderson R 1928 Some points in the general theory of graduation in JC Fields, ed *Proc. Of The Int. Mathematical Congress* vol 2 pp 815–820
- [68] Mulder B, Lagendijk A and Vos W L 2026 Resolving features and derivatives in noisy data using weighted Whittaker-Henderson smoothing *Zenodo* (<https://doi.org/10.5281/zenodo.17131152>)