Proceedings of the World Congress on Engineering 2015 Vol II
WCE 2015, July 1 - 3, 2015, London, U.K.

An Example of Peak Finding in Univariate Data by Least Squares Approximation and Restrictions on the Signs of the First Differences

I. C. Demetriou

***Abstract*—We consider an application of the least squares piecewise monotonic data approximation method to the problem of locating significant extrema in univariate observations that are contaminated by random errors. The piecewise monotonic approximation method makes the smallest change to the data such that the first differences of the smoothed values change sign a prescribed number of times, but the positions of the sign changes are unknowns of the optimization process. We present a numerical example in order to show the efficiency of the method for peak finding. The example is an application to 31959 noisy observations of daily sunspots. Our results suggest some subjects for future research in automatic peak finding.**

***Index Terms*—data smoothing, divided differences, peak finding, piecewise monotonic approximation, sunspots**

I. Background

The piecewise monotonic data approximation method by Demetriou and Powell [8] provides useful applications in signal processing (see, for example, [2], [7], [12], [21] and references therein). In this paper an example is worked out as an illustration of the method for estimating turning points of a function from some measurements of its values that contain random errors.

Let $\{\phi_i : i = 1, 2, \ldots, n\}$ be a sequence of measured values of a function $f(x)$ at the abscissae $x_1 < x_2 < \cdots < x_n$, but the measurements include random errors (noise). We assume that if the function has turning points, then the number of measurements is substantially greater than the number of turning points. Therefore some algorithms have been developed by [8] and [5] that modify the measurements if their first differences $\{\phi_{i+1} - \phi_i : i = 1, 2, \ldots, n-1\}$ include more than $k-1$ sign changes, where k is a prescribed integer. This condition allows k monotonic sections to the smoothed data, alternately increasing and decreasing. Let $\{y_i : \ i = 1, 2, \ldots, n\}$ be the smoothed values, which we regard as components of a n-vector $\underline{y}$.

Specifically, the method calculates a vector $\underline{y}$ that minimizes the sum of squares of the errors

$$\Phi(\underline{y}) = \sum_{i=1}^{n} (y_i - \phi_i)^2 \tag{1}$$

Manuscript received April 16, 2015.

The author is very grateful to two referees for several useful comments on the first draft of this paper. He is also grateful to the Solar Influences Data Analysis Center of the Royal Observatory of Belgium for providing the data set with the sunspots. This work was supported in part by the University of Athens under Research Grant 11105.

I. C. Demetriou is with the Division of Mathematics and Informatics, Department of Economics, University of Athens, 1 Sofokleous and Aristidou street, Athens 10559, Greece (e-mail: demetri@econ.uoa.gr).

subject to the piecewise monotonicity constraints

$$\left.\begin{array}{ll} y_{t_{j-1}} \leq y_{t_{j-1}+1} \leq \cdots \leq y_{t_j}, & \text{if } j \text{ is odd} \\ y_{t_{j-1}} \geq y_{t_{j-1}+1} \geq \cdots \geq y_{t_j}, & \text{if } j \text{ is even} \end{array}\right\}, \tag{2}$$

where the integers $\{t_j : \ j = 1, 2, \ldots, k-1\}$, namely the positions of the *turning points* or *extrema* of the fit, satisfy the conditions

$$1 = t_0 \leq t_1 \leq \cdots \leq t_k = n. \tag{3}$$

The integers $\{t_j : j = 1, 2, \ldots, k-1\}$ are not known in advance and they are variables in the optimization calculation that gives a best fit. This raises the number of combinations of integer variables to the order $O(n^k)$, but fortunately the method of Section II allows an efficient and automatic calculation of an optimal fit $\underline{y}$ together with the associated integers $\{t_j : j = 1, 2, \ldots, k-1\}$ in only $O(kn^2)$ computer operations. This complexity reduces to $O(n)$ when $k = 1$ or $k = 2$. Except when $k = 1$, the problem need not have a unique solution.

The method is suitable for data smoothing when the data errors are so large that they can be detected by the first differences. By contrast, the method may not be suitable if the data contains significant random errors that are too small to cause many sign changes in its first differences. An initial advantage of the method is that it is a projection operator, because if the data satisfy the constraints, then they remain unchanged, so if P takes $\underline{\phi}$ to $\underline{y}$, then $P^2 = P$. Further, by considering piecewise linear functions, it is straightforward to see that there exists a piecewise monotonic function $\{y(x) : \ x_1 \leq x \leq x_n\}$ that interpolates the smoothed values, so $\{y(x_i) = y_i : \ i = 1, 2, \ldots, n\}$.

The piecewise monotonic approach to data smoothing avoids the assumption that f depends on a set of parameters and takes the view that some useful smoothing should be possible if the data fails to satisfy a property that is usually satisfied by the underlying function [4]. Indeed, it is different from methodologies, where the data are approximated by spline functions, wavelets or other functional forms (see, for example, [1], [3] and [11]) that have to be chosen a priori. The properties, however, of monotonicity (i.e. $k = 1$) or piecewise monotonicity property (i.e. $k > 1$) in the function being sought are readily diagnosed in the data. In addition, piecewise monotonic data smoothing can be used in several disciplines because the properties it gives occur in a wide range of underlying functions. It can also be used in estimating turning points of a function from noisy measurements of its values. It is reported by [2] that our method was used competitively for automatic peak position finding in P[31] spectral analysis. Peak finding is a major problem in signal processing, spectroscopy and chromatography that is supported by computer packages, like PeakFit^{TM} and Autosignal^{TM} by Systat Software Inc., computing environments like Matlab and many websites (see, for example, [14]). For presentations of medical uses of proton spectroscopy see [16]. The piecewise monotonic method provides optimal extreme values over the whole sequence of data as it is defined by the constraints (2). However, if the user is only interested in a fairly small region about an extremum, he may well combine the global results of our method with his local analyses.

In order to apply the piecewise monotonicity method to a sequence of data, only the parameter k must be set by the user. Then the method automatically and simultaneously provides the optimal turning points and the best fit. The author has developed the Fortran software package L2WPMA [6], which implements the method of [5], as a version of a method in [8]. It is suitable for processing very large numbers of data in real time. The software package has been tested on a variety of data sets showing a performance that provides in practice far shorter computation times than those indicated in theory.

A value for k may be selected by inspecting the plotted data, or by forming tables of the first differences of the data and checking for sign alterations, or by increasing k until the differences $\{y_i - \phi_i : \ i = 1, 2, \ldots, n\}$ seem to be due to the errors of the data. Prior knowledge about $f(x)$ or about the underlying process may provide estimates of k, but it is not inefficient to run the algorithm of [8] for a sequence of integers k if a suitable value is not known in advance.

The paper is organized as follows. In Section II we give a brief description of the method with emphasis on properties of the turning points of the smoothed data. In Section III we consider a numerical example in order to illustrate the method. Specifically, we apply the method on 31959 observations of daily sunspots during the time period August 1928 - February 2015, we present some results and we demonstrate the capability of the method in locating turning points. The Karush-Kuhn-Tucker statistical test provided an adequate value of k automatically. In Section IV we present some concluding remarks and discuss on the possibility of future directions of this research.

II. Some Features of the Piecewise Monotonic Approximation Method

This section gives some details of the method that are needed in the application of Section III. For proofs one may consult the references stated previously.

Turning points are important to this calculation, because they have properties that are used in practice and enhance the computation greatly. In order to state these properties, we define $t \in [1, n]$ to be the index of a local minimum of the data if, moving to the left or right from ϕ_t in the sequence $\{\phi_i : \ i = 1, 2, \ldots, n\}$, we find either $\phi_i > \phi_t$ or the end of the sequence before $\phi_i < \phi_t$ occurs, and analogously for a local maximum. We denote the sets of the indices of local minima and local maxima of the data by $\mathcal{L}$ and $\mathcal{U}$ respectively. We note that $\mathcal{L}$ and $\mathcal{U}$ can be formed in $O(n)$ operations, each of these sets has fewer than $n/2$ elements that are in strictly ascending order and their interior elements interlace.

If the number of extrema in the data is less than $k-1$, then $\underline{y} = \underline{\phi}$, because in this case $\underline{\phi}$ satisfies the piecewise monotonicity constraints. If, however, $\underline{\phi}$ does not satisfy the piecewise monotonicity constraints, as it occurs in practice, then it is proved that the turning point indices $\{t_j : j = 1, 2, \ldots, k-1\}$ of a best fit $\underline{y}$ are all different and at the turning points we have the interpolation conditions

$$y_{t_j} = \phi_{t_j}, j = 1, 2, \ldots, k-1. \tag{4}$$

The component y_{t_j}, for some $j \in [1, k-1]$, need not be where $\max\{\phi_i : \ t_{j-1} \le i \le t_{j+1}\}$ occurs. Indeed, an example in [6] considers the data $n = 5$, $x_i = i$, $i = 1, 2, \ldots, 5$ and, $\phi_1 = \phi_2 = 2$, $\phi_3 = -7$, $\phi_4 = 3$ and $\phi_5 = -5$, where $\phi_4 = \max\{\phi_i : 1 \le i \le 5\}$. Then the example finds that the components of a monotonic increasing / decreasing fit $\underline{z}$ subject to the condition that its maximum is at the fourth data point satisfy $z_1 = z_2 = z_3 = -1$, $z_4 = 3$ and $z_5 = -5$ and give $\Phi(\underline{z}) = 54$. Further, the optimal fit $\underline{y}^*$ with two monotonic sections in (2) has the components $y^*_1 = y^*_2 = 2$, $y^*_3 = y^*_4 = -2$ and $y^*_5 = -5$ and gives $\Phi(\underline{y}^*) = 50 < \Phi(\underline{z}) = 54$. However, the maximum is at the first or at the second data point, different from the fourth where $\max\{\phi_i : 1 \le i \le 5\}$ occurs.

In view of the inequalities

$$\begin{cases} y_{t_j-1} \le y_{t_j} \ge y_{t_j+1}, & \text{if } j \text{ is odd} \\ y_{t_j-1} \ge y_{t_j} \le y_{t_j+1}, & \text{if } j \text{ is even} \end{cases} \tag{5}$$

and (4), we see that the constraints $y_{t_j-1} \le y_{t_j}$ and $y_{t_j} \ge y_{t_j+1}$ if j is odd, and analogously if j is even, need not be considered in the calculation of an optimal fit. Thus, each monotonic section in a best piecewise monotonic fit is the optimal fit itself to the corresponding data. Hence it can be obtained by a separate calculation and the components $\{y_i : i = t_{j-1}, t_{j-1}+1, \ldots, t_j\}$ on $[x_{t_{j-1}}, x_{t_j}]$ minimize the sum of the squares

$$\sum_{i=t_{j-1}}^{t_j} (y_i - \phi_i)^2 \tag{6}$$

subject to the constraints

$$y_i \le y_{i+1}, i = t_{j-1}, \ldots, t_j - 1, \ \text{ if } j \text{ is odd} \tag{7}$$

or subject to the constraints

$$y_i \ge y_{i+1}, i = t_{j-1}, \ldots, t_j - 1, \ \text{ if } j \text{ is even.} \tag{8}$$

In the former case the sequence $\{y_i : i = t_{j-1}, t_{j-1} + 1, \ldots, t_j\}$ is the best monotonic increasing fit to $\{\phi_i : i = t_{j-1}, t_{j-1}+1, \ldots, t_j\}$ and in the latter case the best monotonic decreasing one. The statement suggests expressing (1) in the form

$$\Phi(\underline{y}) = \alpha(t_0, t_1) + \beta(t_1, t_2) + \alpha(t_2, t_3) + \cdots + \delta(t_{k-1}, t_k),$$

where, for positive integers p and q such that $1 \le p \le q \le n$, we define

$$\alpha(p, q) = \min_{y_p \le y_{p+1} \le \cdots \le y_q} \sum_{i=p}^{q} (y_i - \phi_i)^2, \tag{9}$$

ISBN: 978-988-14047-0-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

and

$$\beta(p,q) = \min_{y_p \geq y_{p+1} \geq \cdots \geq y_q} \sum_{i=p}^{q} (y_i - \phi_i)^2 \qquad (10)$$

and δ denotes α if k is odd and β if k is even. The computation of all $\{\alpha(p,i) : i = p, p+1, \ldots, q\}$ with the best fit that occurs in (9) is achieved in only $O(q-p)$ computer operations and similarly for the β's.

Further, the required piecewise monotonic fit is obtained by a dynamic programming formula that makes use of the separation property of its components. We define $Y(k,n)$ to be the set of n-vectors $\underline{y}$ that satisfy the constraints (2). Then the first t_{k-1} components of an optimal $\underline{y}$ in $Y(k,n)$ give an optimal fit from $Y(k-1, t_{k-1})$ to ϕ_i, $i = 1, 2, \ldots, t_{k-1}$ and the last $n - t_{k-1}$ components give the optimal fit to the remaining data subject to the constraints $y_{t_{k-1}} \leq y_{t_{k-1}+1} \leq \cdots \leq y_{t_k}$, if k is odd, or subject to the constraints $y_{t_{k-1}} \geq y_{t_{k-1}+1} \geq \cdots \geq y_{t_k}$, if k is even. For any integers $m \in [1,k]$ and $t \in [1,n]$, we define

$$\gamma(m,t) = \min_{\underline{z} \in Y(m,t)} \sum_{i=1}^{t} (z_i - \phi_i)^2,$$

where $Y(m,t)$ is the set of t-vectors with m monotonic sections analogously to $Y(k,n)$. Therefore t_{k-1} satisfies the equation

$$\left.\begin{array}{l} \gamma(k-1, t_{k-1}) + \alpha(t_{k-1}, n) = \\ \min\limits_{1 \leq s \leq n} \left[\gamma(k-1,s) + \alpha(s,n)\right], \text{ if } k \text{ is odd} \\ \\ \gamma(k-1, t_{k-1}) + \beta(t_{k-1}, n) = \\ \min\limits_{1 \leq s \leq n} \left[\gamma(k-1,s) + \beta(s,n)\right], \text{ if } k \text{ is even.} \end{array}\right\} \qquad (11)$$

It follows that the least value of the right hand side of (11) can be found in $O(n)$ computer operations provided that the sequences $\{\gamma(k-1,s) : s = 1, 2, \ldots, n\}$ and $\{\alpha(s,n) : s = 1, 2, \ldots, n\}$ or $\{\beta(s,n) : s = 1, 2, \ldots, n\}$ are available.

Therefore in order to calculate $\gamma(k,n)$, which is the least value of (1), we begin the calculation from $\gamma(1,t) = \alpha(1,t)$, for $t = 1, 2, \ldots, n$, and proceed by applying the dynamic programming formulae

$$\gamma(m,t) = \begin{cases} \min\limits_{1 \leq s \leq t} \left[\gamma(m-1,s) + \alpha(s,t)\right], m \text{ odd} \\ \min\limits_{1 \leq s \leq t} \left[\gamma(m-1,s) + \beta(s,t)\right], m \text{ even,} \end{cases} \qquad (12)$$

for $t = 1, 2, \ldots, n$, for every value of $m \in [2,k]$. We store also $\tau(m,t)$, namely the value of s that minimizes expression (12), for each value of m and t.

At the end of the process $m = k$ occurs and the value $\tau(k,n)$ is the integer t_{k-1} that is required in equation (11). Then, because $\tau(k-1, t_{k-1})$ is the optimal value of s in expression (12) when $m = k-1$ and $t = t_{k-1}$, it is the required value t_{k-2}. Hence, we set $t_k = n$ and we obtain the sequence of optimal values $\{t_j : j = 1, 2, \ldots, k-1\}$ by the backward formula

$$t_{m-1} = \tau(m, t_m), \text{ for } m = k, k-1, \ldots, 2. \qquad (13)$$

Accordingly, the components of an optimal fit are monotonic increasing on $[1, t_1]$ and on $[t_j, t_{j+1}]$ for even j in $[1, k-1]$ and monotonic decreasing on $[t_j, t_{j+1}]$ for odd j in $[1, k-1]$. This dynamic programming process requires $O(kn^2)$ computer operations.

Formulae (12) provide the basis for the calculation, but far more efficient formulae are employed in practice. Indeed, we assume in this calculation that $\underline{\phi} \notin Y(k,n)$ and let $\{t_1, t_2, \ldots, t_{k-1}\}$ be optimal integers. If $t = t_\ell$ for some $\ell \in \{2, 3, \ldots, k-1\}$ such that $\phi_{t+1} = \phi_t$, then $\{t_1, t_2, \ldots, t_{\ell-1}, t_\ell+1, t_{\ell+1} \ldots, t_{k-1}\}$ are also optimal. Hence, it is not necessary to compute both $\gamma(m,t)$ and $\gamma(m,t+1)$ and in fact the integer $t_{m-1} = \tau(m, t_m)$ is the index of a local maximum of the data if m is even, that is $t_{m-1} \in \mathcal{U}$, and it the index of a local minimum if m is odd, that is $t_{m-1} \in \mathcal{L}$. Further, if the righthand side of (12) is least for some $s < t$, then s satisfies similar conditions. Therefore we only need to compute $\gamma(m,t)$ when t is the index of a local extremum of the data and similarly for s. With these choices, the process requires $O(n|\mathcal{U}| + k|\mathcal{U}|^2)$ computer operations, which provides a considerable operations saving in practice. More improvements of the calculation are available in [8] and [5], which are summarized by our software package L2WPMA [6].

In Section III we apply a version of L2WPMA that calculates initially an optimal fit with $k = 2$ monotonic sections and then an iterative procedure is started. On each iteration an optimal fit with $k+2$ monotonic sections is calculated, say it is $\tilde{\underline{y}}$, and then it is tested if $\tilde{\underline{y}}$ is improved with respect to the fit with k monotonic sections, say it is $\underline{y}$, on each interval $[x_{t_{j-1}}, x_{t_j}]$. If the test is affirmative then k is increased by 2 and another iteration is commenced. Otherwise $\underline{y}$ is accepted as an adequate approximation to the data. The test for whether the fit need be improved is based on the value of the Karush-Kuhn-Tucker (Lagrange multiplier) statistic (see, for example, [15]) that employs the components of $\underline{y}$ and the components of $\tilde{\underline{y}}$. We accepted $\tilde{\underline{y}}$ for values of the Karush-Kuhn-Tucker test that were larger than the associated 0.1% value of the F-distribution.

III. Numerical Example in Peak Finding of Daily Sunspots During 1927 - 2015

In order to illustrate the efficacy of our method for identifying important extrema in noisy data, we present a numerical example which considers 31959 data points that span the period from August 1927 to February 2015 of the daily sunspot numbers. These numbers show dark spots that appear periodically on the solar surface and affect terrestrial magnetism and other terrestrial phenomena [17], [18]. The datafile dayssnv0-1.dat was downloaded from the website of the Solar Influences Data Analysis Center (SIDC) of the Royal Observatory of Belgium [19]. The datafile contains individual yearly data of the daily sunspot number in three columns: The first column keeps year, month and day, the second column keeps year and fraction of year (in Julian years of 365.25 days) and, the third column keeps the sunspot number. The second and third column provided the data pairs (x_i, ϕ_i), which we plot in Fig. 1. The total number of local minima of the data is $|\mathcal{L}| = 6064$ and the total number of local maxima is $|\mathcal{U}| = 6063$. We see that the data vary considerably and exhibit cycles and spikes.

Without any preliminary analysis or assumption we sought turning points by applying the method of Section II together

ISBN: 978-988-14047-0-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

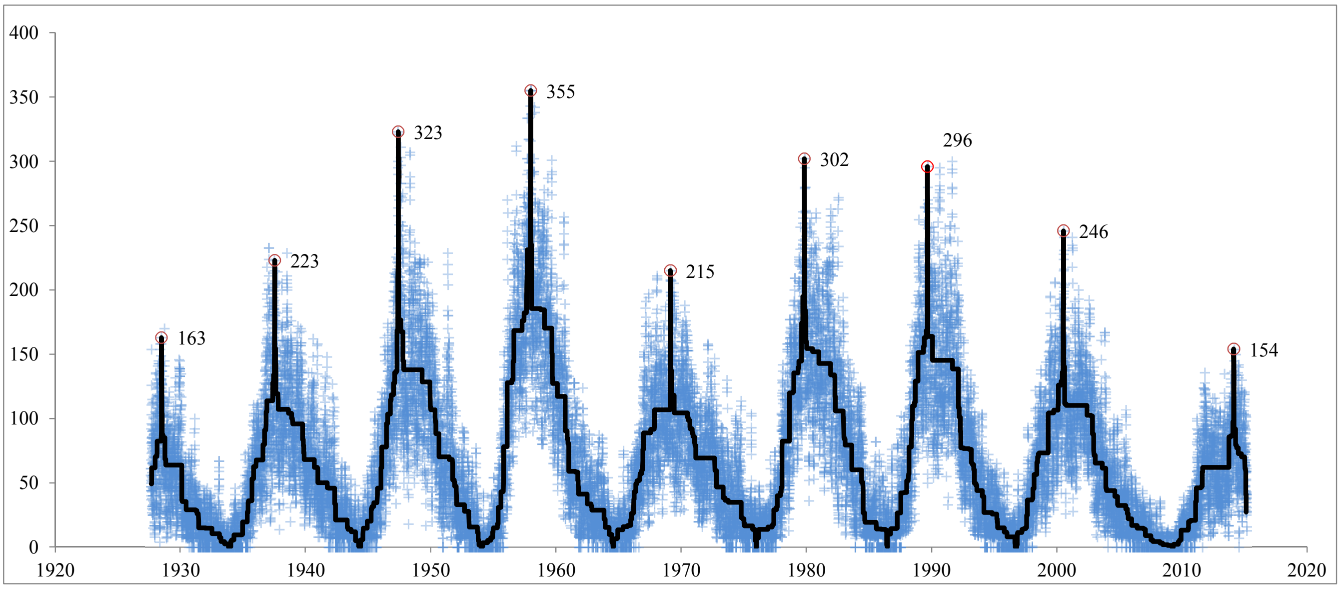

Fig. 1. Detected peaks (circles) by best piecewise monotonic fit with $k = 18$ to 31959 daily sunspots (plus signs) during 1927 - 2015. The solid line illustrates the best fit. The numbers give the sunspots at the corresponding dates.

TABLE I

LEFT FIVE COLUMNS: TURNING POINTS IN DAILY SUNSPOTS DURING 1927 - 2015 BY A BEST FIT WITH $k = 18$ MONOTONIC SECTIONS. RIGHT EIGHT COLUMNS: THE TURNING POINTS POSITION OF THE OPTIMAL FIT FOR $k = 2, 4, \ldots, 16$ ARE INDICATED BY THE TIMES SIGN

j	t_j	Date	Year (x_{t_j})	Sunspot number (ϕ_{t_j})	$k =$	2	4	6	8	10	12	14	16
0	1	19270831	1927.663	56		×	×	×	×	×	×	×	×
1	300	19280625	1928.482	163									
2	2259	19331105	1933.845	0									
3	3604	19370712	1937.528	223								×	×
4	6076	19440418	1944.296	0								×	×
5	7208	19470525	1947.395	323					×	×	×	×	×
6	9632	19540112	1954.031	0					×	×	×	×	×
7	11074	19571224	1957.979	355		×	×	×	×	×	×	×	×
8	13473	19640719	1964.548	0							×	×	×
9	15154	19690224	1969.150	215							×	×	×
10	17651	19751227	1975.986	0			×	×	×	×	×	×	×
11	19065	19791110	1979.858	302			×	×	×	×	×	×	×
12	21485	19860626	1986.483	0				×	×	×	×	×	×
13	22656	19890909	1989.689	296				×	×	×	×	×	×
14	25217	19960913	1996.701	0						×	×	×	×
15	26622	20000719	2000.548	246						×	×	×	×
16	29821	20090422	2009.306	0									×
17	31593	20140227	2014.157	154									×
18	31959	20150228	2015.159	41		×	×	×	×	×	×	×	×

with the Karush-Kuhn-Tucker (Lagrange) multiplier test for the automatic determination of k.

The data $\underline{\phi}$ were fed to the program and the best fit *with* $k = 18$ monotonic sections was calculated. Hence the method detected 17 turning points, which gives 9 peaks. In the left part of Table I we present the turning point indices $\{t_1, t_2, \ldots, t_{17}\}$ as well as $t_0 = 1$ and $t_{18} = 31959$, the associated sunspot date according to the format year, month and day, the year in decimal form and the sunspot number; the odd turning point indices indicate the peaks and the even ones indicate the troughs. Fig. 1 displays the fit and the detected peaks. Indeed, the fit to the data is much smoother than are the data values themselves, it has revealed turning points that seem to be adequately located and it has followed the in-between the turning points trends.

In the right part of Table I we indicate the positions of the turning points of each optimal fit for $k = 2, 4, \ldots, 16$ in correspondence with the column labeled "t_j" derived when $k = 18$. For example, when $k = 4$ the turning points occur at the positions 11074, 17651 and 19065 as indicated by the times signs in the column labeled "4". We remind that as k was increased by 2, all the turning points were found automatically by the method. The method iterated until the Karush-Kuhn-Tucker test indicated no need for further improvement of the fit. We see that the extra turning points of the optimal approximation with $k + 2$ monotonic sections occur between adjacent turning points of the optimal approximation with k monotonic sections. Although it is noticeable that the turning points of the optimal approximation with k monotonic sections are preserved by the optimal approximation with $k + 2$ monotonic sections, any algorithm based on local improvements of an optimal approximation with k monotonic sections cannot succeed in finding more than a local minimum of (1), which may not be a global minimum [4].

Because of the large number of data in this example, we restrict attention to a narrow window of data and discuss on some prominent features of the corresponding smoothed val-

ISBN: 978-988-14047-0-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE II
DAILY SUNSPOTS AND SMOOTHED DATA DURING 20131009-20140518

20131009	79	79.00	20131115	113	85.80	20131222	83	86.52	20140128	54	86.52	20140306	91	93.50	20140412	55	91.76
20131010	90	85.80	20131116	125	85.80	20131223	81	86.52	20140129	66	86.52	20140307	96	93.50	20140413	60	91.76
20131011	91	85.80	20131117	131	85.80	20131224	79	86.52	20140130	69	86.52	20140308	83	91.76	20140414	79	91.76
20131012	84	85.80	20131118	99	85.80	20131225	70	86.52	20140131	65	86.52	20140309	79	91.76	20140415	109	91.76
20131013	99	85.80	20131119	77	85.80	20131226	75	86.52	20140201	66	86.52	20140310	81	91.76	20140416	141	91.76
20131014	96	85.80	20131120	57	85.80	20131227	77	86.52	20140202	83	86.52	20140311	79	91.76	20140417	150	91.76
20131015	96	85.80	20131121	49	85.80	20131228	73	86.52	20140203	89	89.00	20140312	94	91.76	20140418	134	91.76
20131016	89	85.80	20131122	46	85.80	20131229	90	86.52	20140204	101	97.28	20140313	80	91.76	20140419	134	91.76
20131017	109	85.80	20131123	42	85.80	20131230	73	86.52	20140205	117	97.28	20140314	78	91.76	20140420	130	91.76
20131018	116	85.80	20131124	49	85.80	20131231	99	86.52	20140206	120	97.28	20140315	79	91.76	20140421	113	91.76
20131019	97	85.80	20131125	27	85.80	20140101	87	86.52	20140207	103	97.28	20140316	87	91.76	20140422	93	91.76
20131020	85	85.80	20131126	25	85.80	20140102	93	86.52	20140208	103	97.28	20140317	90	91.76	20140423	64	79.31
20131021	96	85.80	20131127	50	85.80	20140103	107	86.52	20140209	103	97.28	20140318	97	91.76	20140424	54	79.31
20131022	88	85.80	20131128	72	85.80	20140104	95	86.52	20140210	96	97.28	20140319	101	91.76	20140425	43	79.31
20131023	93	85.80	20131129	67	85.80	20140105	94	86.52	20140211	111	97.28	20140320	99	91.76	20140426	34	79.31
20131024	108	85.80	20131130	65	85.80	20140106	117	86.52	20140212	113	97.28	20140321	90	91.76	20140427	58	79.31
20131025	103	85.80	20131201	90	85.80	20140107	98	86.52	20140213	103	97.28	20140322	104	91.76	20140428	60	79.31
20131026	99	85.80	20131202	101	85.80	20140108	75	86.52	20140214	91	97.28	20140323	108	91.76	20140429	58	79.31
20131027	111	85.80	20131203	80	85.80	20140109	84	86.52	20140215	79	97.28	20140324	98	91.76	20140430	62	79.31
20131028	110	85.80	20131204	85	85.80	20140110	96	86.52	20140216	72	97.28	20140325	97	91.76	20140501	59	79.31
20131029	112	85.80	20131205	80	85.80	20140111	99	86.52	20140217	74	97.28	20140326	80	91.76	20140502	77	79.31
20131030	109	85.80	20131206	71	85.80	20140112	93	86.52	20140218	89	97.28	20140327	82	91.76	20140503	82	79.31
20131031	97	85.80	20131207	65	85.80	20140113	82	86.52	20140219	88	97.28	20140328	87	91.76	20140504	85	79.31
20131101	72	85.80	20131208	69	85.80	20140114	67	86.52	20140220	93	97.28	20140329	84	91.76	20140505	95	79.31
20131102	69	85.80	20131209	103	86.52	20140115	65	86.52	20140221	95	97.28	20140330	72	91.76	20140506	99	79.31
20131103	84	85.80	20131210	136	86.52	20140116	56	86.52	20140222	102	102.00	20140331	84	91.76	20140507	80	79.31
20131104	87	85.80	20131211	128	86.52	20140117	52	86.52	20140223	111	109.00	20140401	72	91.76	20140508	88	79.31
20131105	83	85.80	20131212	110	86.52	20140118	81	86.52	20140224	107	109.00	20140402	86	91.76	20140509	93	79.31
20131106	99	85.80	20131213	105	86.52	20140119	77	86.52	20140225	120	120.00	20140403	100	91.76	20140510	82	79.31
20131107	113	85.80	20131214	98	86.52	20140120	93	86.52	20140226	145	145.00	20140404	119	91.76	20140511	100	79.31
20131108	97	85.80	20131215	98	86.52	20140121	90	86.52	20140227	154	**154.00**	20140405	102	91.76	20140512	103	79.31
20131109	71	85.80	20131216	83	86.52	20140122	108	86.52	20140228	137	137.00	20140406	91	91.76	20140513	89	79.31
20131110	69	85.80	20131217	88	86.52	20140123	102	86.52	20140301	111	113.00	20140407	86	91.76	20140514	111	79.31
20131111	77	85.80	20131218	102	86.52	20140124	81	86.52	20140302	113	113.00	20140408	88	91.76	20140515	104	79.31
20131112	92	85.80	20131219	102	86.52	20140125	70	86.52	20140303	115	113.00	20140409	71	91.76	20140516	89	79.31
20131113	104	85.80	20131220	104	86.52	20140126	68	86.52	20140304	101	105.50	20140410	49	91.76	20140517	101	79.31
20131114	118	85.80	20131221	102	86.52	20140127	53	86.52	20140305	110	105.50	20140411	46	91.76	20140518	92	79.31

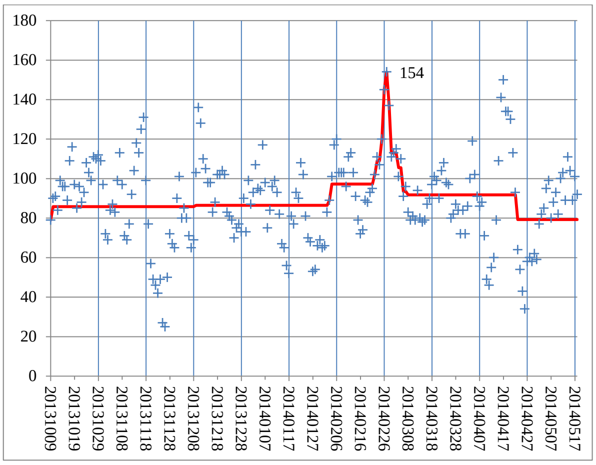

Fig. 2. Plot of the daily sunspots (plus signs) and the smoothed data (solid line) during 20131009-20140518, which are presented in Table II. Number 154 gives the sunspots on date 20140227.

ues. Specifically, in Table II we display 222 data and solution components around the rightmost turning point integer t_{17}, which span the time period 20131009-20140518. Index t_{17} occurs at the date 20140227 which is associated to the value $y_{t_{17}} = 154$ (it is typed with bold characters in Table II). Of course, at the peak we have the equation $y_{t_{17}} = \phi_{t_{17}} = 154$.

The data are presented in six triples of columns, where the first column of each triple keeps the dates, the second column keeps the sunspot numbers and the 3rd column keeps the smoothed values. Further, the smoothed values consist of the monotonic increasing components, which are part of the best monotonic increasing fit on $[x_{t_{16}}, x_{t_{17}}]$, and the best monotonic decreasing components on $[x_{t_{17}}, x_n]$. Fig. 2 displays the data of Table II. In order to simplify the discussion, let $\{s, s+1, \ldots, t_{17}\}$ be the data indices of the monotonic section on 20131009-20140227 and we note that the extracted components $\{y_s, y_{s+1}, \ldots, y_{t_{17}}\}$ from the best fit on $[x_1, x_n]$ give the best monotonic increasing fit to the data $\{\phi_s, \phi_{s+1}, \ldots, \phi_{t_{17}}\}$. Now, we see that each monotonic section consists of values of equal components and there are breakpoints $r \in [s, t_{17}]$ such that $y_r < y_{r+1}$. If r and ℓ are any integers in $[s, t_{17}]$ such that $y_r < y_{r+1} = y_{r+2} = \cdots = y_\ell < y_{\ell+1}$, then it is a consequence of the first order conditions (see, for example, [9]) of minimizing the function

$$\sum_{i=r+1}^{\ell} (y_i - \phi_i)^2$$

subject to the equality constraints

$$y_{r+1} = y_{r+2} = \cdots = y_\ell$$

that the solution satisfies the equations

$$y_{r+1} = \cdots = y_\ell = \frac{1}{\ell - r} \sum_{i=r+1}^{\ell} \phi_i.$$

Hence y_ℓ is the best least squares approximation by a constant to the data $\{\phi_{r+1}, \phi_{k+2}, \ldots, \phi_\ell\}$ and in general the values of the equal components in a best monotonic fit are averages of consecutive data. In Table II, we see that the monotonic increasing fit contains the components $\{79 \times 1, 85.8 \times 60, 86.52 \times 56, 89 \times 1, 97.28 \times 18, 102 \times 1, 109 \times 2, 120 \times 1, 145 \times 1, 154 \times 1\}$, where 79×1, 85.8×60 and so on, indicate that the components 79, 85.8 and so on are repeated once, 60 times and so on. Similarly, the best monotonic decreasing fit contains the components $\{154 \times 1, 137 \times 1, 113 \times 3, 105.5 \times 2, 93.5 \times 2, 91.76 \times 46, 79.31 \times 26\}$. Clearly, the best least squares monotonic fit consists of ranges

ISBN: 978-988-14047-0-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

of equal components between breakpoints. An important consequence of the breakpoints is that the fit has flexibility in following monotonic data trends.

IV. Concluding Remarks

We have presented an application that shows the effectiveness of the piecewise monotonic approximation method in identifying important extrema in discrete noisy data. Piecewise monotonic approximation as a data smoothing approach can have many applications, because piecewise monotonicity is a property that occurs in a wide range of underlying functions. Despite the large number of local minima that can occur in this optimization calculation, it obtains a global solution in quadratic complexity with respect to n, but in practice the complexity is far lower because the dynamic programming algorithm takes account of several properties of the turning points that reduce the numerical work. The accompanying Fortran software L2WPMA would be the most useful for real time processing applications, because it is user friendly, it does not need user intervention, it is able to process fast and effectively large data sets and provided that k is known it identifies the data extrema automatically. A question that deserves further study is the development of techniques of choosing automatically a value for k. The method of [20], which employs the trend test of [13], is a step towards this direction and worked successfully on a number of examples. Further, the Karush-Kuhn-Tucker test that we employed in Section II seems promising, but more work is needed to gain experience.

The example with the 31959 sunspots data is especially challenging, because the very many peaks that occur in the data raise the number of combinations for optimal extrema to about $6064^{17}/17!$. Nonetheless, the least value of (1) was reached automatically in negligible time on a common pc. This example drew attention to the effectiveness and accuracy of L2WPMA at finding the integers $\{t_1, t_2, \ldots, t_{k-1}\}$. Indeed, the number of peaks and their positions in the data sequence show that the results accord with human judgment, while no assumption on any structure of the sunspots data was made. However, some data, as for instance in NMR spectroscopy [10], have typical structures concerning the number of peaks. Therefore an interesting question for further study is whether the $O(n|\mathcal{U}| + k|\mathcal{U}|^2)$ complexity of our method is too high for problems involving such data sets. These studies may be helped by solving particular peak finding problems, in order to receive guidance from both underlying structures and numerical results.

References

[1] C. de Boor, *A Practical Guide to Splines, Revised Edition*, NY: Springer-Verlag, Applied Mathematical Sciences, vol. 27, 2001.

[2] A. P. Bruner, K. Scott, D. Wilson, B. Underhill, T. Lyles, C. Stopka, R. Ballinger and E. A. Geiser, "Automatic Peak Finding of Dynamic Batch Sets of Low SNR In-Vivo Phosphorus NMR Spectra", unpublished manuscript, Departments of Radiology, Physics, Nuclear and Radiological Sciences, Surgery, Mathematics, Medicine, and Exercise and Sport Sciences, University of Florida, and the Veterans Affairs Medical Center, Gainesville, Florida, U. S. Available online: cds.ismrm.org/ismrm-1998/.../P1857.PDF (accessed on 21 March 2015)

[3] W. Cheney and W. Light, *A Course in Approximation Theory*, NY: Brooks/Cole Publishing Company, An International Thompson Publishing Company, 2000.

[4] I. C. Demetriou, *Data Smoothing by Piecewise Monotonic Divided Differences*, Ph.D. Dissertation, Department of Applied Mathematics and Theoretical Physiscs, University of Cambridge, Cambridge, 1985.

[5] I. C. Demetriou, "Discrete piecewise monotonic approximation by a strictly convex distance function", *Mathematics of Computation*, vol. 64, 209, pp. 157-180, 1995.

[6] I. C. Demetriou, "Algorithm 863: L2WPMA, a Fortran 77 package for weighted least-squares piecewise monotonic data approximation", *ACM Trans. Math. Softw.*, vol. 33, 1, pp. 1-19, 2007.

[7] I. C. Demetriou and V. Koutoulidis, "On Signal Restoration by Piecewise Monotonic Data Approximation", in Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2013, Eds. S. I. Ao, L. Gelman, D. W. L. Hukins, A. Hunter, A. M. Korsunsky, U.K., 3-5 July, 2013, London, pp. 268-273.

[8] I. C. Demetriou and M. J. D. Powell, "Least squares smoothing of univariate data to achieve piecewise monotonicity", *IMA J. of Numerical Analysis*, vol. 11, pp. 411-432, 1991.

[9] R. Fletcher, *Practical Methods of Optimization*, Chichester, U. K.: J. Wiley and Sons, 2003.

[10] H. Gunther, *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*, 3rd ed., Chichester, U. K.: J. Wiley and Sons, 2013.

[11] W. Härdle, M. Müller, S. Sperlich and A. Werwatz, *Nonparametric and Semiparametric Models*, Heidelberg, Germany: Springer-Verlag, 2004.

[12] J. Lu, "Signal restoration with controlled piecewise monotonicity constraint", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 12 May 1998-15 May 1998, Seattle WA, vol. 3, pp. 1621 - 1624, 1998.

[13] M. J. D. Powell, "Curve fitting by splines in one variable", in *Numerical Approximation to Functions and Data*, ed. J. G. Hayes, The Institute of Mathematics and its Applications, London, U. K.: The Athlone Press, pp. 65 - 83, 1970.

[14] Peak Finding and Measurement. Available online: http://terpconnect.umd.edu/∼toh/spectrum/PeakFindingandMeasurement.htm (accessed on 21 March 2015).

[15] T. Robertson, F. T. Wright and R. L. Dykstra, *Order Restricted Statistical Inference*, New York: John Wiley and Sons, 1988.

[16] V. M. Runge, W. R. Nitz and S. H. Schmeets, *The Physics of Clinical MR Taught Through Images*, 2nd ed., New York: Thieme, 2009.

[17] F. Scholkmann, J. Boss and M. Wolf, "An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals", *Algorithms*, vol. 5, pp. 588-603, 2012.

[18] S. K. Solanki, "Sunspots: An overview", *Astronomy Astrophysics Review*, vol. 11, pp. 153-286, 2003.

[19] Solar Influences Data Analysis Center (SIDC) of the Royal Observatory of Belgium. Available online: http://sidc.oma.be/DATA, dayssnv0.dat (accessed on 21 March 2015).

[20] E. Vassiliou and I. C. Demetriou, "An adaptive algorithm for least squares piecewise monotonic data fitting", *Computational Statistics and Data Analysis*, vol. 49, pp. 591-609, 2005

[21] J. B. Weaver, "Applications of Monotonic Noise Reduction Algorithms in fMRI, Phase Estimation, and Contrast Enhancement", *International Journal on Innovation, Science and Technology*, vol. 10, pp. 177-185, 1999.

ISBN: 978-988-14047-0-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)