Convex-Concave fitting to successively updated data and its application to covid-19 analysis

Demetrius E. Davos[1] · Ioannis C. Demetriou[1] 📷

Abstract

Let n measurements of a process be provided sequentially, where the process follows a sigmoid shape, but the data have lost sigmoidicity due to measuring errors. If we smooth the data by making least the sum of squares of errors subject to one sign change in the second divided differences, then we obtain a sigmoid approximation. It is known that the optimal fit of this calculation is composed of two separate sections, one best convex and one best concave. We propose a method that starts at the beginning of the data and proceeds systematically to construct the two sections of the fit for the current data, step by step as n is increased. Although the minimization calculation at each step may have many local minima, it can be solved in about $\mathcal{O}(n^2)$ operations, because of properties of the join between the convex and the concave section. We apply this method to data of daily Covid-19 cases and deaths of Greece, the United States of America and the United Kingdom. These data provide substantial differences in the final approximations. Thus, we evaluate the performance of the method in terms of its capabilities as both constructing a sigmoid-type approximant to the data and a trend detector. Our results clarify the optimization calculation both in a systematic manner and to a good extent. At the same time, they reveal some features of the method to be considered in scenaria that may involve predictions, and as a tool to support policy-making. The results also expose some limitations of the method that may be useful to future research on convex-concave data fitting.

Keywords Approximation · Convex-concave · Covid-19 pandemic data · Divided difference of order two · Inflection point · Least squares fit · Quadratic programming · Sigmoid · Substitution

✉ Ioannis C. Demetriou
    ioannis.demetriou@econ.uoa.gr

    Demetrius E. Davos
    demetrius.davos@econ.uoa.gr

[1] Department of Economics, National and Kapodistrian University of Athens, 1 Sofokleous and Aristidou Str, Athens 10559, Greece

Mathematics Subject Classification 41A29 · 65D10 · 90-11 · 90C27 · 90C90

1 Introduction

It has been about two years since the Covid-19 pandemic went into full force across the world, with its contagion seeing thousands of new cases and deaths every day. At the same time, policy makers are tasked with formulating policy to handle the pandemic as each wave of Covid-19 sweeps through society. Given that the pandemic is a phenomenon of growth, many analyses conducted in the years since the start of the pandemic have employed analytical models such as sigmoid curves in efforts to model and predict the evolution of the pandemic, to assist in the formulation of policy (Debecker and Modis 2021).

However, there has been a distinct lack of approaches to this problem that were not of analytical presumptions. In this paper, we propose a method for the first time, where the rates of change of new confirmed Covid-19 cases and deaths are estimated from the data through a convex-concave fitting process to data obtained from Our World in Data (Ritchie et al. 2020). We evaluate its performance as the Covid-19 pandemic develops and ascertain how well it can be used to assist in policy making.

Let $\{\phi_i : i = 1, 2, \ldots, n\}$ be given measurements of the real function values $\{f(x_i) : i = 1, 2, \ldots, n\}$, where the abscissae $\{x_i : i = 1, 2, \ldots, n\}$ satisfy the conditions $x_1 < x_2 < \ldots < x_n$, and the measurements contain random errors. When the data are collected from some process, such as a pandemic or a product substitution, the data may well take the form of the letter S when plotted, the underlying shape granting them the name 'sigmoid'.

Sigmoid functions have seen applications in a variety of fields, such as in the study of population growth since the beginning of the 19th century by Gompertz (1815) and Verhulst (1838), and throughout the last two centuries to date. When sigmoid functions are used for prediction, they are made to fit a dataset that exhibits a matching trend, in order to ascertain certain aspects of the data, as well as project on future data points; for example, when faced with a sigmoid curve of daily contagion growth data of a given population, it may be possible to tell if the contagion is on a rampant increase, which would, in turn, allow for proper decision-making in handling the emergency. As mentioned before, sigmoid functions are presently seeing great use for modeling purposes, tackling the problem of the growth of contagion in the pandemic of Covid-19 (Shen 2020). In this approach, the user relies on certain parameters and presumptions about an analytical model $f(x)$. If the assumption is corroborated by collected data, then the parameters of the algebraic form of $f(x)$ are evaluated so as to derive a useful approximation in accordance to some criterion, such as least squares.

In this paper we avoid the assumption that $f(x)$ has a form that depends on a few parameters. We take the view that some smoothing should be possible, if the data fail to possess a property that is usually obtained by the underlying function. We consider the problem of calculating numbers $\{y_i : i = 1, 2, \ldots, n\}$ from the measurements $\{\phi_i : i = 1, 2, \ldots, n\}$, that are smooth and that should be closer to the true function values $\{f(x_i) : i = 1, 2, \ldots, n\}$ compared to the measurements. The errors in the data tend to cause many sign alterations in the sequence of the second divided differences

$$\phi[x_{i-1}, x_i, x_{i+1}] = \frac{\phi_{i-1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + \frac{\phi_i}{(x_i - x_{i-1})(x_i - x_{i+1})}$$

$$+ \frac{\phi_{i+1}}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}, \quad i = 2, 3, \ldots, n-1. \quad (1)$$

If, however, the data are exact values of a function $f(x)$, $x_1 \leq x \leq x_n$ which has a continuous second derivative that changes sign at most once, then it can be proved that the number of sign changes in the sequence (1) is at most one (see, for example, Powell (1981)). Our approach is based on a method developed by Demetriou (2004a), which seeks numbers y_i, $i = 1, 2, \ldots, n$ that minimize the sum of squares

$$F(\underline{y}) = \sum_{i=1}^{n} (y_i - \phi_i)^2, \qquad \underline{y} \in \mathbb{R}^n, \quad (2)$$

subject to the constraints that the sequence

$$y[x_{i-1}, x_i, x_{i+1}], \quad i = 2, 3, \ldots, n-1 \quad (3)$$

changes sign at most once. Ideally, one sign change occurs in the second derivative of the underlying function $f(x)$, which would give a sigmoid form. Therefore, this method imposes the missing sigmoid property of $f(x)$ as a condition to the data smoothing calculation. Let $\underline{y}(n) \in \mathbb{R}^n$ be a solution to this problem, which we descriptively call an optimal convex-concave approximation to the data $\{\phi_i : i = 1, 2, \ldots, n\}$. Of note, is that a related problem is studied by Cullinan (2019), where the minimization of the objective function $\max\{|y_i - \phi_i| : i = 1, 2, \ldots, n\}$, subject to the same constraints on $\underline{y} \in \mathbb{R}^n$, is considered.

This sigmoid property plays a crucial part in our analysis of Covid-19 contagion. A key characteristic of the pandemic is that a given affected country usually has to face one or more 'waves' of Covid-19, which refer to a surge of Covid-19 cases in the population. A 'wave' is usually signified by a period where new cases slowly increase, followed by a period of rapid increase as each case begets others. This rapid increase is, in turn, normally followed by a period of slowdown,[1] until contagion slows to the point where new cases dwindle and total cases finally level off. Under this configuration, the data of new and total cases alike exhibit behaviour that, over time, lends itself to convex-concave approximation. The total cases evolve to exhibit a sigmoid trend, as is the result of the behaviour described previously; on the other hand, the new cases data can exhibit convexity-concavity, as they provide the rate of change of the total cases data, which are sigmoid in nature. Both the prevalence of and interaction between these sigmoid / convex-concave properties showcase both the need for this particular type of modeling, and why enforcing it as a condition for data fitting is an important part of the process.

With this in mind, we extend the mentioned smoothing calculation in a way that is applied to $\{\phi_i : i = 1, 2, \ldots, n\}$ for successive values of n. Sect. 2 gives some

---

[1] In our scenario, the conditions for this can be likely enforced artificially, such as with the imposition of policy-based preventative measures.

background results. First, it briefly describes the main property of the smoothing calculation. The property states that an optimal convex-concave fit consists of two separate sections, one best convex and one best concave that can be derived independently by two strictly convex quadratic programming calculations. Then, it states a B-spline representation of the best fit that is appropriate for presenting the application in Sect. 4. The main property is taken from Demetriou (2004a), and allows for the development of an efficient method that calculates the solution to this problem in about $\mathcal{O}(n^2)$ computer operations. The mentioned extension is considered in Sect. 3. The extended method starts at one end of the data and proceeds systematically as data enter the calculation. Specifically, the method produces a best convex-concave fit to the current data, and, in the long run, the fitting provides an approximation to the function underlying the data. The extension achieves substantial efficiencies in computation and savings in storage by taking advantage of the structure of the problem and the arrangement of the calculation.

In Sect. 4, the extended method is applied to real data regarding daily cases related to Covid-19 in the countries of Greece, the United States of America and the United Kingdom. In the timeframe covered by the data, spanning the period of June 1st 2021 to September 30th 2021 – a period of four months –, the three countries exhibit distinct data behaviours, facing waves of Covid-19 contagion that vary in scale and duration. The method is applied to the data on a monthly basis. It is first run on just the June 2021 data; after the output is extracted, the dataset is expanded to include the data of July and the method is run anew. This process continues until all data in the June-September dataset are used. The purpose that drives this process is twofold: first, it allows one to ascertain performance of the method as new data enter the optimization calculations carried out therein, regardless of the nature of the data used;[2] second, analysis of the output of the method may provide insights regarding the data that can support specific purposes, such as assisting in policy-making regarding Covid-19. For instance, a strong indicator for policy makers to consider is how the inflection point in the generated splines evolves, as new data enter the calculation. An additional analysis is also conducted, where data pertaining to new deaths in Greece are added on a weekly basis, to provide a more dynamic view of the performance of the method. Identifying the extent to which such insights can indeed prove useful, at least in the context of Covid-19, is also one of the purposes of this paper. Finally, our results are reviewed in Sect. 5.

Beyond its application to modeling the epidemic growth, the approach presented here may be applied to a variety of situations, when we know some properties of the underlying function, but do not have sufficient information to express $f(x)$ in a parametric form. For example, all industries are faced with the threat of substitution, where one product supplants another (Porter 1985). Our approach may provide an efficient tool for defending against a substitute or for promoting substitution, thus guiding in practice a competitive strategy. Technological substitution and forecasting include a wide range of problems where our method may find fruitful domains for applications (see, for example, sigmoid substitution curves from real data by Marchetti

---

[2] In this case, Covid-19 data are used, but any set of data that displays convexity-concavity may be used in general.

(1988), Modis (1993, 1999), Duncan (1999), and references therein). Other examples arise from machine maintenance, economic utility curves, and financial mathematics, to name a few.

2 Background of the convex-concave fit

This section consists of two parts. The first part states the main property of the optimal convex-concave approximation which was mentioned in Sect. 1. Specifically, the best approximation can be generated by solving two independent quadratic programming problems. The second part gives a brief description of the quadratic programming algorithm, and provides a spline representation of its solution that is instructive when employing the method for data analyses.

2.1 Some properties

A vector $\underline{y} \in \mathbb{R}^n$ is feasible if it satisfies the constraints

$$\left.\begin{array}{ll} y[x_{i-1}, x_i, x_{i+1}] \geq 0, & i = 2, 3, \ldots, j-1 \\ y[x_{i-1}, x_i, x_{i+1}] \leq 0, & i = j, j+1, \ldots, n-1, \end{array}\right\} \tag{4}$$

for some integer j in $[2, n]$ where we ignore the first line of (4) if $j = 2$ and the second line of (4) if $j = n$.

If $\underline{y} = \underline{y}(n) \in \mathbb{R}^n$ is optimal, then, for some integer ζ in $[2, n]$, the components y_i, $i = 1, 2, \ldots, \zeta - 1$ have the values y_i^{cx}, $i = 1, 2, \ldots, \zeta - 1$ that solve the quadratic programming problem (best convex approximation on $[x_1, x_{\zeta-1}]$)

$$\left.\begin{array}{l} \text{minimize } \sum_{i=1}^{\zeta-1} (y_i - \phi_i)^2, \\ \text{subject to } y[x_{i-1}, x_i, x_{i+1}] \geq 0, \ i = 2, 3, \ldots, \zeta-2, \end{array}\right\} \tag{5}$$

except that there are no constraints if $\zeta \leq 3$. Assuming that ζ is well inside the range $[2, n]$, which avoids trivialities of the presentation, the components y_i, $i = \zeta, \zeta + 1, \ldots, n$ have the values y_i^{cv}, $i = \zeta, \zeta + 1, \ldots, n$ that solve the quadratic programming problem (best concave approximation on $[x_\zeta, x_n]$)

$$\left.\begin{array}{l} \text{minimize } \sum_{i=\zeta}^{n} (y_i - \phi_i)^2, \\ \text{subject to } y[x_{i-1}, x_i, x_{i+1}] \leq 0, \ i = \zeta+1, \ldots, n-1, \end{array}\right\} \tag{6}$$

except that there are no constraints if $\zeta \geq n - 1$.

We define $\alpha(1, \zeta - 1; n)$ and $\beta(\zeta, n; n)$ to be the least values of the objective functions of the quadratic programming problems (5) and (6) respectively. Further, if we define $\gamma(\zeta; n) = \sum_{i=1}^{\zeta-1} (y_i^{cx} - \phi_i)^2 + \sum_{i=\zeta}^{n} (y_i^{cv} - \phi_i)^2$, which is the optimal value of the objective function (2), it follows that we can obtain the expression

$$\gamma(\zeta; n) = \alpha(1, \zeta - 1; n) + \beta(\zeta, n; n), \tag{7}$$

where we let $\alpha(1, 1; n) = \alpha(1, 2; n) = \beta(n - 1, n; n) = \beta(n, n; n) = 0$. Because ζ is not known in advance, one can calculate this sum for every ζ in $[2, n]$ in order to find one that gives the value $\gamma(\zeta; n)$. The assertion that the components of $\underline{y}(n)$ can be generated by solving separate quadratic programming problems on the convex and the concave section is proven by Demetriou (2004a).

In order to state a method that makes use of this idea, we define the quantities $\{\alpha(1, j; n) : j = 1, 2, \ldots, n\}$ by

$$\alpha(1, j; n) = \min_{y_1, \ldots, y_j} \left\{ \sum_{i=1}^{j} (y_i - \phi_i)^2 : y[x_{i-1}, x_i, x_{i+1}] \geq 0, \ 2 \leq i \leq j - 1 \right\}, \quad (8)$$

and, analogously, the quantities $\{\beta(j, n; n) : j = 1, 2, \ldots, n\}$. Algorithms for obtaining an optimal integer ζ are proposed by Demetriou and Powell (1997) and Demetriou (2004a). They seek an integer $j \in [2, n]$ that solves the problem

$$\left. \begin{array}{l} \text{minimize} \quad \gamma(j; n) = \alpha(1, j - 1; n) + \beta(j, n; n), \ 2 \leq j \leq n - 1 \\ \text{subject to} \quad y[x_{j-1}, x_j, x_{j+1}] < 0 \\ \text{or} \quad j = n, \end{array} \right\} \quad (9)$$

where the second difference $y[x_{j-1}, x_j, x_{j+1}]$ in formula (9) is evaluated on the vector $\underline{y}$ whose first $j - 1$ components occur at the definition of $\alpha(1, j - 1; n)$, and whose last $n - j + 1$ components occur at the definition of $\beta(j, n; n)$. We let $\zeta(n) = \zeta$ be an integer j that minimizes expression (9) when $\gamma(j; n)$ is calculated. Having found $\zeta(n)$, the components of the two sections of $\underline{y}(n)$ are calculated by solving problems (5) and (6). It is important to note that the optimality of ζ does not depend on the sign of the difference $y[x_{j-2}, x_{j-1}, x_j]$. Therefore, the constraint $y[x_{j-1}, x_j, x_{j+1}] < 0$ is a necessary and sufficient condition for the feasibility of $\underline{y}$.

2.2 The spline representation of the fit

We briefly describe the main ideas of the quadratic programming calculation with reference to problem (5) after we replace $\zeta - 1$ by n. The quadratic programming problem is solved by the method of Demetriou and Powell (1991). This method generates a finite sequence of subsets $\{\mathcal{A}^{(m)} : m = 1, 2, 3, \ldots\}$ of the constraint indices $\{2, 3, \ldots, n - 1\}$ with the property

$$y[x_{i-1}, x_i, x_{i+1}] = 0, \ i \in \mathcal{A}^{(m)}. \quad (10)$$

For each m, we let $\underline{y}^{(m)}$ be the vector that minimizes the objective function (2) subject to the equations (10). Unique Lagrange multipliers $\{\lambda_i^{(m)} : i \in \mathcal{A}^{(m)}\}$ are defined by the first order optimality condition

$$\underline{y}^{(m)} - \underline{\phi} = \tfrac{1}{2} \sum_{i \in \mathcal{A}^{(m)}} \lambda_i^{(m)} \underline{a}_i, \quad (11)$$

where $\underline{a}_i$ is the normal of the constraint function $y[x_{i-1}, x_i, x_{i+1}]$. If $\mathcal{A}^{(m)}$ is not the final set of the mentioned sequence, then the quadratic programming method makes adjustments to $\mathcal{A}^{(m)}$ until the solution is reached. The Karush-Kuhn-Tucker conditions provide necessary and sufficient conditions for optimality.

The equality constrained minimization problem that gives $\underline{y} = \underline{y}^{(m)}$ forms an important part of the calculation, because it is solved very efficiently by a reduction to an equivalent unconstrained one with fewer variables due to a linear B-spline representation. Specifically, if $s(x)$, $x_1 \leq x \leq x_n$ is the piecewise linear interpolant to the points $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, then $s(x)$ has its knots on the set $\{x_i : i \in \{1, 2, \ldots, n\} \setminus \mathcal{A}^{(m)}\}$ including x_1 and x_n. Indeed, the equation $y[x_{i-1}, x_i, x_{i+1}] = 0$, when $i \in \mathcal{A}^{(m)}$, implies the collinearity of the points (x_{i-1}, y_{i-1}), (x_i, y_i) and (x_{i+1}, y_{i+1}), but if $y[x_{i-1}, x_i, x_{i+1}] > 0$, then i is the index of a knot of $s(x)$. Thus, the knots of $s(x)$ are determined from the abscissae due to the constraints (10). Let $k_n = n - 1 - |\mathcal{A}^{(m)}|$, let $\{\xi_j : j = 0, 1, \ldots, k_n\}$ be the knots of $s(x)$ in ascending order, where $\xi_0 = x_1$ and $\xi_{k_n} = x_n$, and let $\{B_j : j = 0, 1, \ldots, k_n\}$ be a basis of normalized linear B-splines that are defined on the abscissae $\{x_i : i = 1, 2, \ldots, n\}$ and satisfy the equations $B_j(\xi_j) = 1$ and $B_j(\xi_i) = 0$, $j \neq i$. Then $s(x)$ may be written uniquely in the form

$$s(x) = \sum_{j=0}^{k_n} c_j B_j(x), \ x_1 \leq x \leq x_n, \tag{12}$$

where the spline coefficients $\{c_j : j = 0, 1, \ldots, k_n\}$ are the values of $s(x)$ at the knots and are calculated by solving the normal equations associated with the minimization of the objective function (2).

We assume that $\underline{\phi} \in \mathbb{R}^n$ is available throughout the calculation, and we accompany $\underline{y}(n)$ by the quintuple of the elements $(n, k_n, \underline{\xi}, \underline{c}, \zeta(n))$, where $\underline{\xi} \in \mathbb{R}^{k_n+1}$ is the vector whose components are the knots, $\underline{c} \in \mathbb{R}^{k_n+1}$ is the vector whose components are the spline coefficients, and $\zeta(n)$ is the optimal value of j obtained at the end of calculation (9). In this way we simplify the development of our approximation procedure in Sect. 3 by just referring to $\underline{y}(n)$ and $\zeta(n)$. In addition, the quintuple is a convenient way of representing the spline fittings in Sect. 4.

3 The approximation procedure

The user provides the data (x_i, ϕ_i), $i = 1, 2, \ldots, n$, and the procedure calculates a best convex-concave approximation $y_i(n)$, $i = 1, 2, \ldots, n$ and the associated integer variable $\zeta(n)$, as n is increased by one.

The procedure begins with $n = 3$. Then the components of the vectors $\underline{y}(1), \underline{y}(2)$ and $\underline{y}(3)$, and the associated integer variables $\zeta(1), \zeta(2)$ and $\zeta(3)$ are given the values

$$\begin{cases} y_1(1) = y_1(2) = y_1(3) = \phi_1, \ y_2(2) = y_2(3) = \phi_2, \text{ and } y_3(3) = \phi_3 \\ \zeta(1) = 1, \ \zeta(2) = 2; \text{ if } y[x_1, x_2, x_3] < 0 \text{ then } \zeta(3) = 2, \text{ else } \zeta(3) = 3. \end{cases} \tag{13}$$

It is worth noting that past approximations may provide for a starting point other than $n = 3$. After the procedure has started, $\underline{y}(n)$ is available and provides the starting point for the calculation of the best convex-concave approximation to the first $n + 1$ data. Now we increase n by one, and in order to obtain the best convex-concave approximation to the first n data, we add the new data point (x_n, ϕ_n) to the best approximation $\underline{y}(n - 1)$, which defines $\hat{y}(n) \in \mathbb{R}^n$ by

$$\hat{y}_i(n) = \begin{cases} y_i(n - 1), & i = 1, 2, \ldots, n - 1 \\ \phi_n, & i = n. \end{cases} \tag{14}$$

Then we define $\underline{y}(n) = \hat{\underline{y}}(n)$ and $\zeta(n) = \zeta(n - 1)$ unless

$$\hat{y}(n)[x_{n-2}, x_{n-1}, x_n] > 0, \tag{15}$$

which gives infeasibility. In this case, an equality constraint occurs at the concave section of the best fit, so we undertake the calculation of $\underline{y}(n)$ on the range $[x_1, x_n]$. The procedure is described below. The motivation for treating a convexity as an equality constraint is given in the following lemma.

Lemma 1 *We employ the notation of the first two paragraphs of this section. If $\hat{y}(n)[x_{n-2}, x_{n-1}, x_n] \leq 0$ then $\underline{y}(n) = \hat{\underline{y}}(n)$, but otherwise $\underline{y}(n)$ satisfies the equation $y(n)[x_{n-2}, x_{n-1}, x_n] = 0$.*

Proof By following Lemma 2 of Demetriou (2004a). □

The procedure depends on the important separation property of the convex and the concave section of a best approximation, as was stated in Sect. 2.1, and allows a constructive method for obtaining $\underline{y}(n)$. The method calculates an integer ζ such that the final fit has a convex section on $[x_1, x_{\zeta-1}]$ and a concave section on $[x_\zeta, x_n]$. It follows that the equality

$$\sum_{i=1}^{n} (y_i(n) - \phi)^2 = \alpha(1, \zeta - 1; n) + \beta(\zeta, n; n) \tag{16}$$

is derived. Precisely, the method seeks an integer $\zeta \in [2, n]$ that minimizes the right-hand side of expression (16), provided that the corresponding approximation $\underline{y}$ satisfies the constraint $y[x_{\zeta-1}, x_\zeta, x_{\zeta+1}] \leq 0$ or $\zeta = n$. The constraint allows for a convex section on $[x_1, x_{\zeta-1}]$ and a concave section on $[x_\zeta, x_n]$ as was written in the paragraph following the statement of problem (9).

In order to implement this technique for obtaining $\zeta(n)$, we let j be any trial value of ζ in the righthand side of the sum (16), and we obtain the quantity $\alpha(1, j; n)$ by solving problem (8). The calculation starts from the formula

$$\gamma(j; n) = \alpha(1, j; n), \quad j = 1, 2, \ldots, n,$$

and proceeds by employing the formula

$$\gamma(\zeta; n) = \min_{2 \le j \le n-1} \{\alpha(1, j-1; n) + \beta(j, n; n) : \quad y[x_{j-1}, x_j, x_{j+1}] < 0\}. \quad (17)$$

However, the constraint $y[x_{j-1}, x_j, x_{j+1}] < 0$ involves only three components of the trial vector whose components occur at the definitions of $\alpha(1, j-1; n)$ and $\beta(j, n; n)$. Therefore, we only need to pick the value y_{j-1} that occurs in the calculation of $\alpha(1, j-1; n)$ which we denote by $\psi^{(j-1)}(x_{j-1})$, and we take the components y_j and y_{j+1} from the fit that provides $\beta(j, n; n)$ which we denote by $\psi^{(j)}(x_j)$ and $\psi^{(j)}(x_{j+1})$, respectively. So far, this method, when n is fixed, is known and is briefly presented in Sect. 2.1.

Our procedure extends this method for successive values of n in a way that achieves substantial efficiencies in computation and savings in storage by taking advantage of the structure of the problem and the arrangement of the calculation. The gain in efficiency comes from the remark that all the numbers $\alpha(1, j; n)$ and $\beta(j, n; n)$, $j = 1, 2, \ldots, n$ are required (also including the limiting cases when $j = 1, n$) in order to implement formula (17) for the current n. Therefore, when n takes values in the set $\{4, 5, 6, \ldots\}$, these numbers have to be recomputed or stored. We avoid this task when n is increased, because the values $\alpha(1, j; n-1)$, $j = 1, 2, \ldots, n-1$ are kept in storage, and no storage is required for $\beta(j, n; n)$, $j = 1, 2, \ldots, n$. Indeed, j has already run through the set $\{1, 2, \ldots, n-1\}$, and the numbers $\alpha(1, j; n-1)$, $j = 1, 2, \ldots, n-1$ were calculated and placed in temporary storage together with the components specified in the paragraph after equation (17). Thus, they can be used again when n is increased. It follows that just the number $\alpha(1, n; n)$ need be calculated when n is increased, as we explain next.

The calculation of $\alpha(1, n; n)$ requires either $\mathcal{O}(1)$ or at most $\mathcal{O}(n)$ computer operations. We recall definition (14) and the $\mathcal{O}(1)$ complexity is obtained when the inequality

$$\hat{y}(n)[x_{n-2}, x_{n-1}, x_n] \ge 0, \quad (18)$$

is satisfied, because in this case $y(n) = \hat{y}(n)$ occurs, so $\alpha(1, n; n) = \alpha(1, n-1; n-1)$. If, instead, inequality $\hat{y}(n)[x_{n-2}, x_{n-1}, x_n] < 0$ occurs, then, by reference to Lemma 1, the best convex approximation to the first n data satisfies the constraint $y[x_{n-2}, x_{n-1}, x_n] \ge 0$ in equational form and has sum of squares of residuals equal to $\alpha(1, n; n)$. The $\mathcal{O}(n)$ complexity is achieved because the best convex approximation to the first $n-1$ data provides a very good starting point for the quadratic programming calculation that gives the best convex approximation to the first n data (Demetriou 2006). It is sufficient for future applications of formula (17) to retain the numbers y_{n-1}^{cx} and y_n^{cx}, because y_{n-1}^{cx} is used to test the feasibility condition (18), and y_n^{cx} is used to test the feasibility condition included in formula (17). Note that the storage requirements for all the values

$$\alpha(1, j; \cdot), \quad y_{j-1}^{cx}, \quad \text{and } y_j^{cx}, \quad j = 2, 3, \ldots, n, \quad (19)$$

are only $\mathcal{O}(n)$.

Further, the numbers $\beta(j, n; n)$, $j = 1, 2, \ldots, n$ are calculated on stream for the current value of n, and are then used in formula (17). We remember the remark made in the paragraph following inequality (18) about the $\mathcal{O}(n)$ requirements for having the best convex approximation to the first n data from the best convex approximation to the first $n - 1$ data, and, analogously, the calculation of the sequence $\beta(j, n; n)$, $j = 1, 2, \ldots, n$ is obtained by repeated applications of the mentioned quadratic programming algorithm in only $\mathcal{O}(n^2)$ computer operations. We arrange this calculation so that $\beta(j, n; n)$ need not be stored. Indeed, if j runs through the set $\{2, 3, \ldots, n-1\}$, then these numbers have to be recomputed or stored. We avoid these tasks by employing an outer loop that makes use of the formula

$$\gamma(j; n) = \alpha(1, j - 1; n) + \beta(j, n; n), \quad \text{if } y[x_{j-1}, x_i, x_{j+1}] < 0, \quad (20)$$

for $j = 2, 3, \ldots, n - 1$, before it obtains the least value of $\gamma(j; n)$ on those indices j in $[2, n]$, such that the feasibility condition $y[x_{j-1}, x_i, x_{j+1}] < 0$ is satisfied, or $j = n$. The numbers $\beta(j, n; n)$ are accompanied by the components

$$y_j^{cv} \text{ and } y_{j+1}^{cv}, \quad (21)$$

which are used to test the feasibility condition in formula (20), but no storage is required. The outer loop provides the values $\gamma(\zeta; n)$ and $\zeta(n)$. Furthermore, the concave section of the associated $\underline{y}(n)$ on $[x_\zeta, x_n]$ provides the components

$$y_{n-1}^{cv} \text{ and } y_n^{cv}, \quad (22)$$

which are needed to test the feasibility of $\hat{y}(n)$, when a new data point enters the calculation (as described in the paragraph after equation (14)). Then another cycle of the procedure begins. This procedure may well be applied to obtain a best concave-convex fit after a sign change in the components ϕ_i, $i = 1, 2, \ldots, n$.

4 The application to the covid-19 data

Davos (2021) created a Python Interface to the software of Demetriou (2006), which implements the algorithm described by Demetriou (2004a) through use of Fortran (Demetriou 2006). As of yet, this is a prototype of the approximation procedure in Sect. 3. Henceforth, we refer to it as the 'algorithm'.

Data can be input through the interface; in turn, the user is provided with the associated output of the algorithm. The output includes the convex-concave approximations to the data, the associated Lagrange multipliers, the knots that comprise the approximation spline, its inflection point (if any), as well as the slope of each linear segment of the spline, representing the rate of change of the underlying function in each spline segment. We represent $\underline{y}(n)$ by the quintuple $(n, k_n, \underline{\xi}, \underline{\sigma}, \zeta(n))$ as described in Sect. 2.2. An important note regarding the presentation: whenever inflection is detected, it is presented through two highlighted points on the approximation spline; the rightmost of which – or just the one, if only one exists – is what the algorithm recognizes as

the inflection point with index $\zeta(n)$. These highlighted points represent what can be called the inflection range of the data, within which the actual inflection point of the underlying function lies (Davos 2021).

Herein we shall present graphical and numerical output of the method and interface, through application to Covid-19 data of Greece (pop. 11 mil.), the United States of America (pop. 334 mil.) and the United Kingdom (pop. 68 mil.). The datasets used pertain to new Covid-19 cases on a daily basis, spanning the period of June 1st 2021 to September 30th 2021. For purposes of demonstrating certain aspects of the calculation, additional data on daily deaths for Greece, spanning the period of June 1st 2021 to October 31st 2021, were also used. The data have been gathered from Our World in Data (Ritchie et al. 2020), sourced from Johns Hopkins University. In this dataset, the abscissae used are the dates for each day; since the algorithm requires strictly numeric data as input, the dates have been converted into integers using the Microsoft Excel DATEVALUE() function. The data in question are too many to be presented in these pages.

An important factor to consider that Covid-19 data generally possess is a cycle-like diffusion stemming from weekly seasonality; the strength of this diffusion varies from dataset to dataset, which, in turn, can affect performance of the algorithm in a number of ways. In addition, different countries may be facing different periods of a Covid-19 wave over the same timeframe; where one might be approaching a peak in cases, another may be exiting a wave entirely. Differences in terms of the starting point can affect performance in convex-concave approximations. This follows naturally: if the dataset starts out as concave and then becomes convex, the concave section of the data will likely be misapproximated, as the convex-concave algorithm will attempt to construct a convex section first.

Material supplementary to this paper also contains results and analyses for the new deaths data corresponding to the new cases data studied herein. From this, among others, an insight regarding the relationship between the deaths data and the cases data of a given country was gleaned: the behaviour of the new deaths data tends to be quite similar to that of the new cases, with a delay of half a month, on average, as a result of Covid-19 pathology, in terms of cases that become mortalities.

With the above in mind, the aforementioned datasets were selected: the behaviour of the data in the case of Greece provides a more standard use scenario for the algorithm, highlighting its primary properties as a dynamically evolving process. On the other hand, the USA and UK data are such that peculiarities in the behaviour thereof are reflected in the output of the algorithm as time goes on; said peculiarities will be addressed accordingly, in their respective sections of this paper. The complexity of the data provides a good test of the power of the method. The results are analyzed so as to assist decision making, particularly regarding the evolution of the inflection point.

4.1 Greece—standard inflection behaviour

In this section, we apply the method to Covid-19 daily cases data of Greece, spanning the period of June 1st 2021 to September 30th 2021. In Fig. 1, we have the raw data presented as scattered dots and four convex-concave approximation splines as
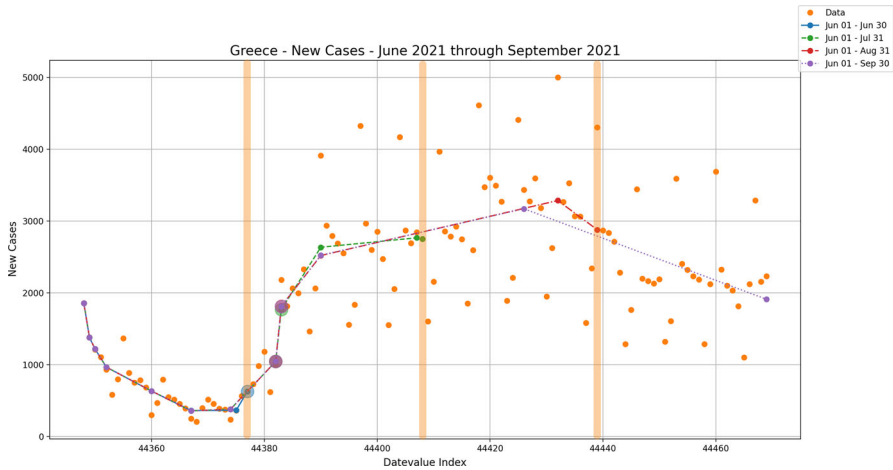
Fig. 1 New Cases in Greece from June 1st 2021, to September 30th 2021. (Large circles denote inflection point/range; the three highlighted columns section off each of the four months in the dataset, each on the last day of June, July and August, from left to right)

generated by the method. Each spline corresponds to an approximation run on a dataset expanded by one month, starting from a set of data pertaining to the period of June 1st to June 30th and ending with the full dataset of June 1st to September 31st.

As can be seen, up until the middle of July 2021, the cases data are generally non-diffuse, closely following a clear trend and inflecting, from convexity to concavity, in early July. This subset provides an example of a standard convex-concave approximation, where the splines generated by the algorithm are performing well, both as approximants and as trend detectors.

However, past the middle of July, the data grow significantly more diffuse, exhibiting a seasonality that is frequently found across Covid-19 datasets from other countries; the USA and the UK are no exception to this. Likely having their roots, at least in part, in administrative matters, the figures of the cases data recorded at the weekends are considerably smaller, while those in the middle of the week tend to be much higher; the rest of the weekdays tend to operate on similar levels. This property of the data prevents the approximation splines from performing well as approximants, unless measures are taken to counteract it, such as using moving 7-day averages (see Ritchie et al. (2020), for instance), or splitting the dataset into subsets isolating the problematic days and approximating them separately (Davos 2021).

Whatever the case may be, while the algorithm may be lacking as an approximant, its performance as a trend detector is still strong. As seen in Fig. 1, the approximation splines, after the middle of July, strongly assert that the end-of-June-to-present Covid wave in Greece inflects in early July and reaches its peak mid-to-end-of August. Post-inflection, both the ascent towards the peak and descent afterwards are characterized by non-steep slopes, leading to an extended period of high Covid-case readings. The knowledge of the current date in relation to the position of the inflection point can give indications to policy makers, regarding whether or not additional measures need be

taken to stifle the flow of viral transmission; for instance, seeing increasing slopes in successive, mostly convex splines (as is the case for the June and July splines in Fig. 1) would suggest that a significant influx in cases is still to be expected, which could, for example, lead to more restrictive measures in favour of reducing transmission.

This behaviour is also reflected in the numerical data provided by the interface - they are presented in Table 1, which summarizes the results of the run on the data of cases in Greece. It displays the knot and end point indices $j = 0, 1, \ldots, k_n(= 7)$, the dates at the knots and corresponding values ξ_j, the estimated values, i.e. the spline coefficients c_j, the first divided differences of the fit $s[\xi_{j-1}, \xi_j]$ (namely, the slopes of the line segments that join the two consecutive knots ξ_{j-1} and ξ_j) and the second divided differences of the fit $s[\xi_{j-1}, \xi_j, \xi_{j+1}]$ centered at knot ξ_j. Table 1 consists of four separate sets of rows, each of which provides the spline representation of a convex-concave approximation, starting with the approximation over the data of June and subsequently over the addition of the data of July, August and September. In this paper, we shall analyse the output regarding the period of June 1st to July 31st, as an example; the style of presentation is uniform across all tables and between cases and deaths data, so it can be easily extended to any output pertaining to either cases or deaths data, over different periods and/or different countries (or different data, in general).

The approximation spline that the algorithm calculates for these data is convex-concave in nature; this can be seen by studying the second divided differences column in relation to the inflection point of the spline. As seen in the table, the spline inflects in the range of the two knots $\{\xi_7, \xi_8\} = \{44382, 44383\}$, with associated spline coefficients (c_j) of 1043.465 and 1767.438, respectively. The change in convexity, from convex to concave, is marked by the change in sign in the second divided differences. From the first knot of the spline ($\xi_0 = 44348$, June 1st 2021), up to and including the left bound of the inflection range ($\xi_7 = 44382$, July 5th 2021), the second divided differences are positive in sign, which denotes convexity. On the other hand, starting from the right bound of the inflection range ($\xi_8 = 44383$, July 6th 2021) all the way to the last knot of the spline ($\xi_{11} = 44408$, July 31st 2021), second divided differences are negative, which denotes concavity. Whenever the approximation spline bears an inflection range (*comprised of two knots*), the leftmost bound of the inflection range will bear a positive second divided difference and the rightmost, negative. This is a clear display of the convex-concave property, as seen through the second divided differences.

It should be noted, that the first of the four splines generated, covering the June 1st - Jun 30th period, is entirely convex in nature (indeed, its 'inflection point' is actually the last knot of the spline, $\xi_7 = 44377$ - June 30th 2021). From this, the inflection point is observed to move rightwards, if only once, before becoming fixed on the June 5th - June 6th inflection range across all following splines (tantamount to a mostly secure detection of the actual inflection range in the data). This is a case of a forward-progressing and fixed inflection point as new data are included in the calculation, which is the standard scenario when the data are not overly irregular in their convexity-concavity.

In addition to the aforementioned, the output also contains information on the slopes of the individual linear segments of the spline, which denote the rates of change in the

Table 1 The approximation spline output from New Cases data in Greece. Each date is paired with its corresponding MS Excel DATEVALUE(), which was used in the calculation. Inflection Point/Range underlined

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
June 01 2021 - June 30 2021				
0	01/06/2021 (44348)	1857.000	–	–
1	02/06/2021 (44349)	1379.000	−478.000	157.780
2	03/06/2021 (44350)	1216.559	−162.441	12.255
3	05/06/2021 (44352)	965.205	−125.677	8.394
4	13/06/2021 (44360)	631.309	−41.737	0.199
5	20/06/2021 (44367)	360.088	−38.746	2.615
6	28/06/2021 (43375)	363.925	0.480	13.106
7	30/06/2021 (44377)	627.000	131.538	–
June 01 2021 - July 31 2021				
0	01/06/2021 (44348)	1857.000	–	–
1	02/06/2021 (44349)	1379.000	−478.000	157.819
2	03/06/2021 (44350)	1216.639	−162.361	12.148
3	05/06/2021 (44352)	964.806	−125.917	8.436
4	13/06/2021 (44360)	632.339	−41.558	0.139
5	20/06/2021 (44367)	355.998	−39.477	3.054
6	27/06/2021 (44374)	378.918	3.274	5.320
7	05/07/2021 (44382)	1043.465	83.068	71.212
8	06/07/2021 (44383)	1767.438	723.973	−75.006
9	13/07/2021 (44390)	2634.905	123.924	−4.840
10	30/07/2021 (44407)	2766.710	7.753	−1.248
11	31/07/2021 (44408)	2752.000	−14.710	–
June 01 2021 - August 31 2021				
0	01/06/2021 (44348)	1857.000	–	–
1	02/06/2021 (44349)	1379.000	−478.000	157.819
2	03/06/2021 (44350)	1216.639	−162.361	12.148
3	05/06/2021 (44352)	964.806	−125.917	8.436
4	13/06/2021 (44360)	632.339	−41.558	0.139
5	20/06/2021 (44367)	355.998	−39.477	3.054
6	27/06/2021 (44374)	378.918	3.274	5.320
7	05/07/2021 (44382)	1043.465	83.068	76.343
8	06/07/2021 (44383)	1813.624	770.159	−83.666
9	13/07/2021 (44390)	2519.440	100.831	−1.685
10	24/08/2021 (44432)	3285.817	18.247	−1.557
11	31/08/2021 (44439)	2879.423	−58.056	–
June 01 2021 - September 30 2021				
0	01/06/2021 (44348)	1857.000	–	–
1	02/06/2021 (44349)	1379.000	−478.000	157.819
2	03/06/2021 (44350)	1216.639	−162.361	12.148

Table 1 continued

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
3	05/06/2021 (44352)	964.806	−125.917	8.436
4	13/06/2021 (44360)	632.339	−41.558	0.139
5	20/06/2021 (44367)	355.998	−39.477	3.054
6	27/06/2021 (44374)	378.918	3.274	5.320
7	05/07/2021 (44382)	1043.465	83.068	76.138
8	06/07/2021 (44383)	1811.775	768.311	−83.319
9	13/07/2021 (44390)	2524.061	101.755	−1.948
10	18/08/2021 (44426)	3171.329	17.980	−0.599
11	30/09/2021 (44469)	1910.095	−29.331	–

studied measure (the daily cases, in this case) per segment; along with the information regarding the rates of change themselves, the behaviour of the sign changes in the slopes helps one readily identify the bottom and the peak of the spline. The first six knots of the June-July spline produce five linear segments with - gradually flattening - negative slopes, starting from a steep negative slope of –478 (which lasts for a day) and bottoming out with a less steep slope of –39.477, at the knot of $\xi_5 = 44367$, June 20th 2021, with a spline coefficient of 355.998 (approximately 356 daily cases at the lowest, in the period in question). Afterwards, the slopes turn positive, increasing as one heads through to the end of the inflection range (with a rampant increase of about 724 cases from the left bound of the inflection range to the rightmost bound). Since the inflection range signifies a change from convexity to concavity, as is also revealed by the second divided differences, the slopes of subsequent linear segments, expectedly, start to flatten and may even return to being negative; this behaviour, depending on the extent to which the slopes flatten over a given dataset, can also provide hints as to the position of the peak of the studied measure. For instance, in the June-July period, the slopes start to flatten following the July-6th knot (rightmost inflection bound); in doing so, they even change sign, with the last linear segment of the spline becoming a descending one, thus identifying a possible peak (of approximately 2767 cases) at the penultimate knot ($\xi_{10} = 44407$, July 30th 2021).

However, it is to be noted that the proximity of the last two knots ($\xi_{10} = 44407$, July 30th 2021 and $\xi_{11} = 44408$, July 31st 2021), with them being only one day apart (i.e. pertaining to successive observations in the original dataset), provides reasonable doubt as to whether or not the peak is actually there. That doubt is indeed confirmed, as a higher peak, of about 3200 cases, was identified in the period of August 18th (knot $\xi_{10} = 44426$ for 3171.329 cases - June-September spline) and August 24th (knot $\xi_{10} = 44432$ for 3285.817 cases - June-August spline).
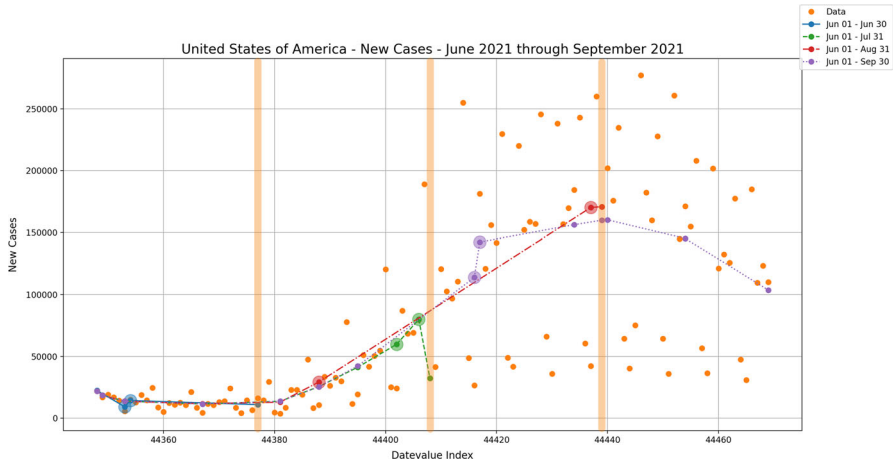
Fig. 2 New Cases in the USA from June 1st 2021, to September 30th 2021. Same presentation style as Fig. 1

4.2 United States of America—variant inflection bounds

In this section, we apply the method to Covid-19 cases data of the United States of America, as in Sect. 4.1. The studied period finds the USA entering and going through a new Covid-19 wave, reaching its peak in late July - early September, exhibiting a weekly seasonality standard to Covid-19 data, as seen in Fig. 2. In the same figure, another property of these data is made apparent; in their diffusion, the mid-range values appear to be rather proximal to the high-end values. This contrasts somewhat with the diffusion observed in the data of Greece in Sect. 4.1, where the mid-range values appeared relatively equidistant from either extreme. Besides this element of seasonality, however, the diffusion in the USA data in particular is also rooted in another important factor: the data are an aggregate of each individual State in the USA. Each State has its own population characteristics (size, demographics, et cetera) and its own state-wide policies. Additionally, any two States follow Covid-19 waves somewhat independently; for instance, one state may be entering a Covid-19 surge as another is leaving one (although, it is not beyond reason to assume that a wave in one state, particularly among the largest, will eventually cascade through the rest). Thus, while any one of the United States may have relatively standard Covid-19 readings for a given period, the aggregate for the USA will likely be particularly diffuse.

The inflection range in this dataset exhibits a particularly interesting property. In the case of Greece, the inflection range proceeded rightwards as new data were added, following the evolution of the Covid-19 wave, until the actual inflection point was reached. However, in the case of the USA, the inflection range actually changes its size and position in somewhat irregular ways.

While the shift from June through July is forward, as expected, the addition of the data of August cause the inflection range to expand dramatically—presenting nearly the entire ascent from growth to peak in the data as linear—so as to encompass the

inflection range of the previous period. This extensive linearity also serves to obfuscate the actual exact location of the inflection point in the data; bounds less proximal to the inflection range would give a larger room for error, in terms of pinpointing the exact inflection point. In addition, when the algorithm runs on the complete four-month dataset, the inflection range is once again compacted, with both its bounds within the inflection range of August. In essence, the inflection point in the data that as asserted by the spline does not necessarily proceed rightwards with the addition of new data; indeed, on the complete dataset, it proceeds leftwards compared to an earlier inflection point,[3] in reverse of what would be usually expected by adding new data. This is an important remark, as it rules out the possibility of reducing the number data that need be considered when finding the optimal inflection points as was observed by Demetriou (2004b).

This shift, among other things, can also be seen in the numerical output in Table 2; in the first two subperiods, the inflection range bounds proceed rightwards, from the 6th and 7th of June in the June data (knots ξ_1 and ξ_2, respectively), to the 25th and 29th in the June-July data (knots ξ_7 and ξ_8, respectively). The addition of the August data finds the inflection bounds distancing themselves from one another, with the left bound at July 11th (knot ξ_5) and the right bound at August 29th (knot ξ_6), a significant range of 49 days, as opposed to the much more common one-to-two-day span of most inflection ranges derived from the studied data. It is to be noted, however, that the inflection range in the last period (June - September), with its bounds situated on the 8th and 9th of August (knots ξ_7 and ξ_8 respectively), showcases a more standard rightward movement of the range in relation to the first two subperiods; in addition to its span being one day, as is common, this serves to identify it as a more likely location for the actual inflection range.

In terms of the general structure of the June-September wave, the approximation splines, as revealed through the slopes and second divided differences in Table 2, provide the following: the lowest point of the studied data (exit from the previous wave into the current one) is identified in the area of the 20th of June and 4th of July, seeing as all approximation splines, barring the June spline, see a sign change in their slopes at that point, from descending to ascending. It is noted that the slopes in this range are much flatter relative to slopes elsewhere on the splines, being practically horizontal in comparison, which lends to the extent of this 15-day range. On the other hand, each approximation spline, barring the June-September spline, identifies the peak of the data as being close to the end of their respective subperiod - the second divided differences change sign only near the end of the data subsets, which is to be expected, as the data, though diffuse, do display a mostly upward trend.

However, once the data of September are included, the resulting approximation spline provides clear insight on the structure of the studied wave: its lowest point is in the area of the 20th of June and the 4th of July, at around 11000 to 11500 cases, evidenced by the negative-to-positive sign change in the slopes; its inflection range spans the 8th and 9th of August, as evidenced by the sign change in the second divided differences; and its peak is identified around the 1st of September, with the spline coefficient asserting it at about 160247 cases, as the slopes change back into

---

[3] Reminding that the right bound of the inflection range is the inflection point returned by the algorithm

Table 2 The approximation spline output from New Cases data in the USA. Same presentation style as in Table 1

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
June 01 2021 - June 30 2021				
0	01/06/2021 (44348)	22450.238	–	–
1	06/06/2021 (44353)	9268.095	−2636.429	1284.340
2	07/06/2021 (44354)	14337.707	5069.611	−217.256
3	30/06/2021 (44377)	11013.627	−144.525	–
June 01 2021 - July 31 2021				
0	01/06/2021 (44348)	22119.000	–	–
1	02/06/2021 (44349)	18518.696	−3600.304	484.922
2	06/06/2021 (44353)	13815.913	−1175.696	56.233
3	20/06/2021 (44367)	11527.009	−163.493	11.375
4	04/07/2021 (44381)	13696.975	154.998	73.680
5	11/07/2021 (44388)	25612.946	1702.282	38.283
6	18/07/2021 (44395)	41280.621	2238.239	29.133
7	25/07/2021 (44402)	59803.366	2646.106	217.352
8	29/07/2021 (44406)	79951.278	5036.978	−4810.936
9	31/07/2021 (44408)	32294.000	−23828.639	–
June 01 2021 - August 31 2021				
0	01/06/2021 (44348)	22119.000	–	–
1	02/06/2021 (44349)	18542.854	−3576.146	475.259
2	06/06/2021 (44353)	13743.439	−1199.854	58.587
3	20/06/2021 (44367)	11709.495	−145.282	8.567
4	04/07/2021 (44381)	13033.890	94.600	106.315
5	11/07/2021 (44388)	29324.393	2327.215	9.794
6	29/08/2021 (44437)	170233.206	2875.690	−50.319
7	31/08/2021 (44439)	170852.000	309.397	–
June 01 2021 - September 30 2021				
0	01/06/2021 (44348)	22119.000	–	–
1	02/06/2021 (44349)	18517.135	−3601.865	485.546
2	06/06/2021 (44353)	13820.594	−1174.135	56.081
3	20/06/2021 (44367)	11515.221	−164.670	11.556
4	04/07/2021 (44381)	13739.806	158.899	71.572
5	11/07/2021 (44388)	25373.215	1661.916	53.266
6	18/07/2021 (44395)	42226.682	2407.638	35.798
7	08/08/2021 (44416)	113836.336	3409.983	1145.059
8	09/08/2021 (44417)	142437.622	28601.286	−1543.554
9	26/08/2021 (44434)	156331.848	817.307	−4.324
10	31/08/2021 (44439)	159942.692	722.169	−69.670
11	01/09/2021 (44440)	160246.841	304.150	−91.421
12	15/09/2021 (44454)	145306.585	−1067.161	−59.017
13	30/09/2021 (44469)	103626.651	−2778.662	–
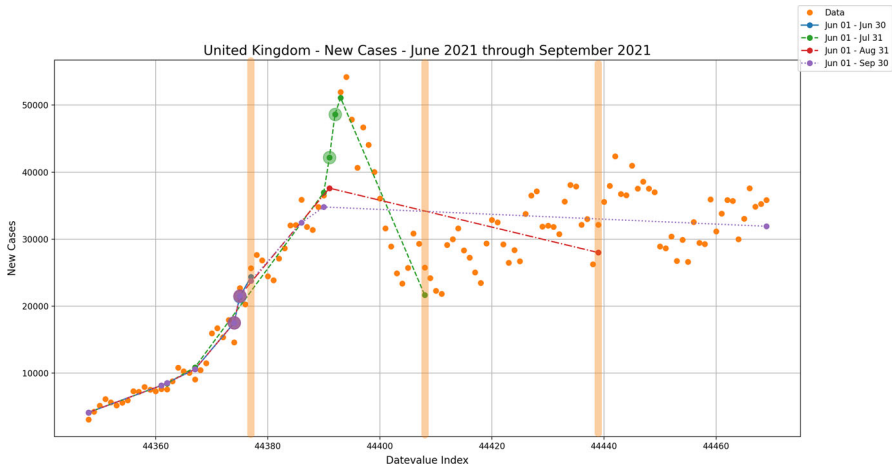
Fig. 3 New Cases in the UK from June 1st 2021, to September 30th 2021. Same presentation style as Fig. 1

descent after this point. It is to be noted that, while the value of the peak itself is misapproximated due to extensive diffusion in the data, its position at September 1st is more securely identified, which is a valuable piece of information in itself.

That said, the erratic behaviour of the inflection range in the period of July-August-September also highlights an important aspect of applying the method on overly diffuse data: though general trends can be tracked, the position of the inflection point may require several periods of data addition in order to be verifiably asserted. Compare this to the data of New Cases in Greece back in Fig. 1, where the inflection range persisted through all periods of data addition once it was reached.

4.3 United Kingdom—short-term data irregularities

In this section, we apply the method to Covid-19 cases data of the United Kingdom, as in Sect. 4.1. The data of cases in the UK, as far as the studied period is concerned, display a remarkably irregular behaviour compared to that of the other countries studied previously. As can be seen in Fig. 3, the period of June and July finds the UK facing what could be described as a rapid transition from the peak of one wave directly to the ascent and peak of another. The approximation splines in the data showcase this; the June spline detects the inflection point of its wave near the end of June (inflection range of $\{\xi_4, \xi_5\} = \{44374, 44375\}$, the 27th and 28th of June), while the addition of the July data shifts the inflection range rightwards, to the 14th and 15th of July. The latter is found remarkably close to the peak of the compounded wave, with nearly the entirety of the June wave becoming practically linear in comparison.

However, the June-July spline also descends quite rapidly; at the end of July, the behaviour of the data becomes much less wild in its variations. The weekly seasonality is still in effect, certainly; but the data are more closely gathered and follow a considerably subdued increase, compared to the rampant increases in June and July. The extent of this change in behaviour is such that the inflection range of the associated

approximation spline actually regresses entirely (and persists) back to that of June – verily, the June spline is a near-perfect subset of the post-July splines. In addition, the peaks of the splines are much lower and the splines themselves much smoother compared to the June-July spline; this can well be interpreted as the July Covid-19 surge having momentarily commandeered the wave of June, as if extruding a spike from the peak of a bell curve (with all deformities that would entail).

This change in spline structure is also in part because of the nature of the post-July data themselves; given that the algorithm employs least squares optimization, the least-squares measurement from maintaining the forward-moving July inflection range would be far larger compared to when using the June inflection range (one need only apply simple linear regression from the July inflection point onwards to verify this).

For their part, the post-July splines also serve in providing a hint regarding the location of the knot, as the data inflect from the concave descent of the July spike back into the slower convex ascent of the new data, in the form of the intersection of their last linear segments. While not providing decisive evidence so as to the value of the approximation coefficient at that knot (or even an exact value for the knot itself), it can nonetheless provide an indication that the data are in the process of shifting from concavity back into convexity (Davos 2021), leading to an eventual collapse of the currently detected inflection range and proceeding towards detecting the one that characterizes the new–now convex–splines.

As is the norm, the numerical output regarding the cases data is seen in Table 3. The slopes of all splines start positive and remain so at least until a peak is reached, which identifies the lowest point of all splines as the first data point (the splines all claim a measure of cases around 4140 for this, which is slightly higher from the first actual data point, on account of the optimization calculation); in this, this lowest point is not necessarily the actual low point of the wave (i.e. the point the wave actually started at) as it may well have appeared earlier. This is simply because the chosen period for our data happened to start after this point had already been reached.

Aside from the June-July spline, with the location of the inflection range pinpointed on the spike on July 14th and 15th, leading to its associated peak of 51091 cases on July 16th, all other splines have exactly the same structure up to and including the rightmost bound of the inflection range: six knots, starting from June 1st and inflecting in the range of June 27th and June 28th. The increasingly steepening slopes show a standard case of the growth phase of a wave, spanning the period from June 1st through July 28th; the abnormal extrusion of the mid-July spike does not, in the end, serve to destroy the form of the wave as described by the splines. On the other hand, the location of the peak showed minor changes as new data were added, and the image of the wave became clearer. The June spline ended with an ascending line segment, thus showing a peak on June 30th of 24358 cases, while also casting doubt as to whether this peak was the actual peak of the wave.

This doubt is confirmed by subsequent splines, suggesting that the peak of the wave appeared somewhere in the area of July 14th and July 16th; this includes the date of the spike in the June-July spline, which makes its constancy interesting when taking the last two splines into account.

Table 3 The approximation spline output from New Cases data in the UK. Same presentation style as in Table 1

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
June 01 2021 - June 30 2021				
0	01/06/2021 (44348)	4138.429	–	–
1	14/06/2021 (44361)	8171.429	310.231	4.402
2	15/06/2021 (44362)	8543.286	371.857	6.221
3	20/06/2021 (44367)	10589.214	409.186	48.407
4	27/06/2021 (44374)	17519.714	990.071	362.673
5	28/06/2021 (44375)	21411.167	3891.452	−805.984
6	30/06/2021 (44377)	24358.167	1473.500	–
June 01 2021 - July 31 2021				
0	01/06/2021 (44348)	4146.041	–	–
1	15/06/2021 (44362)	8464.678	308.474	9.105
2	20/06/2021 (44367)	10871.992	481.463	23.315
3	13/07/2021 (44390)	36960.280	1134.273	170.352
4	14/07/2021 (44391)	42183.000	5222.720	617.640
5	15/07/2021 (44392)	48641.000	6458.000	−2003.941
6	16/07/2021 (44393)	51091.118	2450.118	−275.702
7	31/07/2021 (44408)	21674.507	−1961.107	–
June 01 2021 - August 31 2021				
0	01/06/2021 (44348)	4138.429	–	–
1	14/06/2021 (44361)	8171.429	310.231	4.402
2	15/06/2021 (44362)	8543.286	371.857	6.221
3	20/06/2021 (44367)	10589.214	409.186	48.407
4	27/06/2021 (44374)	17519.714	990.071	380.386
5	28/06/2021 (44375)	21552.870	4033.156	−178.203
6	14/07/2021 (44391)	37612.204	1003.708	−18.813
7	31/08/2021 (44439)	27996.820	−200.320	–
June 01 2021 - September 30 2021				
0	01/06/2021 (44348)	4138.429	–	–
1	14/06/2021 (44361)	8171.429	310.231	4.402
2	15/06/2021 (44362)	8543.286	371.857	6.221
3	20/06/2021 (44367)	10589.214	409.186	48.407
4	27/06/2021 (44374)	17519.714	990.071	376.717
5	28/06/2021 (44375)	21523.520	4003.805	−955.788
6	30/06/2021 (44377)	23796.402	1136.441	−15.885
7	09/07/2021 (44386)	32451.747	961.705	−28.780
8	13/07/2021 (44390)	34801.999	587.563	−7.518
9	30/09/2021 (44469)	31921.455	−36.463	–

It is worth noting, that the final line segment of each of the June-August and June-September splines is remarkably long, spanning one-and-a-half and two-and-a-half months respectively. When considering the relative brevity of the July spike compared to the entire studied period (half a month of a rapid ascent and descent compared to nearly three months along an average of 30000 cases), it would appear that it is a momentary extrusion in a longer, sustained wave. In light of this, the June-August and June-September splines would indeed provide a clearer view of the actual form of the wave. Verily, this entire studied period is generally considered as being part of one single wave. However, should one require greater accuracy in approximation and not just in trend detection, it would be prudent to run the convex-concave algorithm with a date in the end of July as a starting point, as opposed to the 1st of June. That said, the intersection of the rightmost line segments in the June-August and June-September splines seems to be a good indicator as to where to place the first knot. Alternatively, the method of Demetriou and Powell (1997) might have been applied with 4 or 5 inflection points.
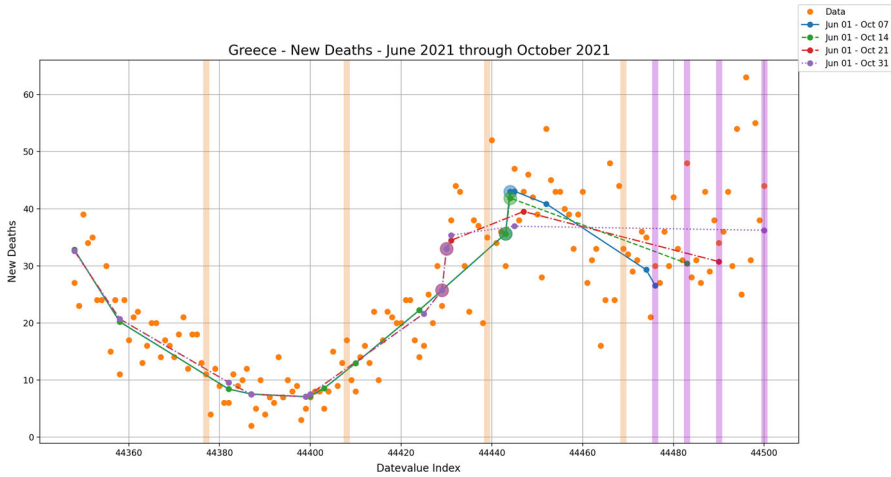
It should be mentioned that the United Kingdom, too, is an aggregate, comprised of four countries (England, Scotland, Wales and Northern Ireland). However, unlike in the case of the USA, the Covid-19 data from the UK are comparatively non-diffuse. One likely reason for this, is that the data mainly follow the same trends as the data for England which has by and large the greatest population among the four countries and the greatest proportion of the population in the United Kingdom overall (according to the latest census at the time of writing (July 2020 to July 2021),[4] England has about $55, 944, 000$ of the estimated $66, 329, 000$ residents of the UK, or about 84 percent of the population). In addition the four countries, partly due to being proximal, with England at the epicentre, generally follow waves in a synchronous fashion, barring differences in policy on the country level. Thus, any particular diffusions in the UK data that cannot be explained from weekly seasonality alone (at least when regarding the data of England), can usually be easily traced back to an irregularity in either of the four component countries.

4.4 An example of weekly analysis

The results on the daily data presented thus far have all been on a monthly basis (i.e. the inclusion of new data has always been in batches of a month's worth of data). It would be prudent for a presentation of weekly inclusions of data to also be provided, since analysis on a weekly basis can provide more immediate insights regarding the current state of affairs. Carried out on such a comparitively frequent basis, analysis of such data can provide information that may well assist in policy making, particularly since it can allow for insights regarding the near future (such as indications for the coming of a cases surge, for instance).

The data used for this purpose pertain to daily Covid-19 deaths in Greece, from June 1st 2021 through October 31st 2021. The process followed is largely the same as in the previous subsections, with the difference that each batch of data added contains

---

[4] https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/populationoftheunitedkingdombycountryofbirthandnationality.

Fig. 4 New Deaths in Greece from June 1st 2021, to October 31st 2021 on a weekly basis for October. (The first four highlighted columns denote end-of-month. The last four denote end-of-week throughout October. Large circles denote inflection point/range)

a week's worth of data, covering the entirety of October 2021. The splines generated here showcase a regressive inflection range, which stands to differentiate this particular subset from the one presented previously. In the material supplementary to this paper, a similar analysis is conducted for data of daily deaths in the UK, where the data batches added cover the month of July 2021. Said analysis provides a standard scenario of a forward-proceeding inflection range, which serves as a complement to the regressive inflection range presented herein.

Figure 4 displays the four approximation splines that resulted over the four weeks of October 2021 (with June 1st 2021 as the origin). As can be seen, the approximation splines for the first two weeks of October agree that the data inflect on the 4th and 5th of September[5] (knots $\xi_9 = 44443$ and $\xi_{10} = 44444$ in the associated subtables of Table 4). As revealed by the addition of data, this inflection range is remarkably close to the peak of deaths in the wave studied; the inflection range of the two splines, on September 4th and September 5th, either is adjacent to the peak (the June 1st–October 7th spline suggests a peak of about 43 deaths on September 6th, just one day after inflection, on knot ξ_{11}), or outright includes it (in the June 1st–October 14th spline, the right bound of the inflection range (knot ξ_{10}) also serves as the peak of the spline, suggesting about 42 deaths).

However, once the second half of October is introduced, in which the advent of a new Covid-19 wave in Greece is already in effect, the inflection range actually shifts backwards, transfixed as the range of August 21st and August 22nd (and, to be noted, both the knots and the coefficients in this inflection range are exactly the same for the latter half of October – knots $\xi_7 = 44429$ and $\xi_8 = 44430$, with coefficients at 25.725 (about 26 deaths) and 33 deaths, respectively –, whereas the inflection range in the first

---

[5] For reference, the same inflection range was suggested by the monthly analysis for the New Deaths in Greece, in the supplementary material

Table 4 The approximation spline output from the weekly-basis New Deaths data in Greece. Same presentation style as in Table 1

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
June 01 2021 - October 07 2021				
0	01/06/2021 (44348)	32.825	–	–
1	11/06/2021 (44358)	20.196	−1.263	0.023
2	05/07/2021 (44382)	8.415	−0.491	0.011
3	10/07/2021 (44387)	7.543	−0.174	0.008
4	22/07/2021 (44399)	7.075	−0.039	0.013
5	23/07/2021 (44400)	7.205	0.130	0.078
6	26/07/2021 (44403)	8.533	0.443	0.019
7	02/08/2021 (44410)	12.968	0.634	0.001
8	16/08/2021 (44424)	22.220	0.661	0.001
9	04/09/2021 (44443)	35.614	0.705	0.334
10	05/09/2021 (44444)	43.000	7.386	−3.662
11	06/09/2021 (44445)	43.063	0.063	−0.048
12	13/09/2021 (44452)	40.843	−0.317	−0.007
13	05/10/2021 (44474)	29.31	−0.524	−0.036
14	07/10/2021 (44476)	26.538	−1.386	–
June 01 2021 - October 14 2021				
0	01/06/2021 (44348)	32.825	–	–
1	11/06/2021 (44358)	20.196	−1.263	0.023
2	05/07/2021 (44382)	8.415	−0.491	0.011
3	10/07/2021 (44387)	7.543	−0.174	0.008
4	22/07/2021 (44399)	7.075	−0.039	0.013
5	23/07/2021 (44400)	7.205	0.130	0.078
6	26/07/2021 (44403)	8.533	0.443	0.019
7	02/08/2021 (44410)	12.968	0.634	0.001
8	16/08/2021 (44424)	22.220	0.661	0.001
9	04/09/2021 (44443)	35.614	0.705	0.278
10	05/09/2021 (44444)	41.879	6.265	−0.164
11	14/10/2021 (44483)	30.421	−0.294	–
June 01 2021 - October 21 2021				
0	01/06/2021 (44348)	32.610	–	–
1	11/06/2021 (44358)	20.699	−1.191	0.021
2	05/07/2021 (44382)	9.591	−0.463	0.001
3	10/07/2021 (44387)	7.473	−0.424	0.023
4	22/07/2021 (44399)	7.106	−0.031	0.036
5	23/07/2021 (44400)	7.547	0.441	0.005
6	17/08/2021 (44425)	21.626	0.563	0.016
7	21/08/2021 (44429)	25.725	1.025	1.250
8	22/08/2021 (44430)	33.000	7.275	−2.909

Table 4 continued

j	Date (ξ_j)	c_j	$s[\xi_{j-1}, \xi_j]$	$s[\xi_{j-1}, \xi_j, \xi_{j+1}]$
9	23/08/2021 (44431)	34.457	1.457	−0.067
10	08/09/2021 (44447)	39.513	0.316	−0.009
11	21/10/2021 (44490)	30.749	−0.204	–
June 01 2021 - October 31 2021				
0	01/06/2021 (44348)	32.610	–	–
1	11/06/2021 (44358)	20.699	−1.191	0.021
2	05/07/2021 (44382)	9.591	−0.463	0.001
3	10/07/2021 (44387)	7.473	−0.424	0.023
4	22/07/2021 (44399)	7.106	−0.031	0.036
5	23/07/2021 (44400)	7.547	0.441	0.005
6	17/08/2021 (44425)	21.626	0.563	0.016
7	21/08/2021 (44429)	25.725	1.025	1.250
8	22/08/2021 (44430)	33.000	7.275	−2.452
9	23/08/2021 (44431)	35.372	2.372	−0.151
10	06/09/2021 (44445)	36.939	0.112	−0.002
11	31/10/2021 (44500)	36.226	−0.013	–

half did have fixed knots, but a slightly variant right bound – the left bound was steady at 35.614 (about 36 deaths), but the latter ranged from 43 to 41.879 – in essence, about 42 or 43 deaths). This backstep of the inflection range actually places it at a more believable position for an inflection point of the wave (since, on average, one would not expect inflection directly next to a peak, barring unforseen events or phenomena); in addition, its continued stability, that even applies to spline coefficients, serves to suggest that this may be the true inflection range of the wave.

That said, all four splines, despite differences over their suggested inflection ranges, agree upon the general location of the peak of the wave, placing it in the range of September 5th to September 8th, though there are differences in coefficients between the forward-inflecting splines (ending on 7/10 and 14/10, suggesting about 43 deaths) and the backstep-inflecting splines (ending on 21/10 and 31/10, suggesting about 37 to 39 deaths). What is more remarkable still, however, is the fact that, when combined, they provide an additional important piece of information; as can be readily observed through the data themselves, a new Covid-19 death wave emerges after a period of decreasing deaths following the peak in the data. This emergence is, naturally, marked by an inflection in the data, from concavity descending from the peak, to a convexity leading into the growth of the new wave. The combined readings of the four splines suggest that this inflection point is around September 21st, which is the point where the dominant linear segments in the concave sections of each spline intersect. This is a significant piece of information to consider, as it is one way that the approximation splines can signal the emergence of a new wave, in addition to being applied on a more dynamically changing dataset. The accuracy of this indication may be tested by applying the algorithm of Demetriou and Powell (1997), searching for more inflection

points. This point of interest warrants further investigation over whether or not the intersection point defines an invariant of this combinatorial calculation locally.

5 Conclusion

We developed a method that gives a sigmoid-type approximation to noisy data, and applied it to daily Covid-19 pandemic data. Specifically, the method calculates the least squares convex-concave approximation to the first n data points for $n = 1, 2, 3, \ldots$, as the data enter the process. The statement of the convexity-concavity constraints in terms of second divided differences subject to one sign change gives rise to a combinatorial problem which is known and was solved some years ago. The solution to this problem allowed the development of the method we presented here. The important property of this calculation is that the piecewise linear interpolant to the optimal approximation for each n consists of two separate parts of components, one best convex and one best concave, that are calculated independently by two quadratic programming calculations. This serves for a reduction in the number of combinations.

Our method starts at the beginning of the data and proceeds systematically as data enter the calculation. Specifically, it produces a best convex-concave fit to the current data; in the long run, the fitting provides an approximation that reveals the structure of the function underlying the data. In the context of the Covid-19 data, this equates to revealing the general form of a major contagion wave. In light of this, the point where the convex part meets the concave one, ideally being the inflection point of an underlying function that has a continuous second derivative, is critical to management when projecting the future.

The method achieves substantial efficiencies in computation and savings in storage by taking advantage of the characteristics of the problem and the arrangement of the calculation. The efficiencies are due to a relation of the components of best approximations to consecutive data, a linear B-spline representation of the components and a suitable quadratic programming calculation. The quadratic programming part takes advantage of changes to the spline that occur on each n, and the use of splines reduces considerably the size of the internal matrix computations (numerical evidence for up to $n = 10000$ data shows it to be around $n/10$).

Moreover, the best convex-concave approximation to the first $n - 1$ data provides an excellent starting point for the best convex-concave approximation to the first n data. In effect, the latter approximation is obtained in about $\mathcal{O}(n^2)$ or just $\mathcal{O}(1)$ computer operations. However, the numerical results from the Covid-19 data confirm that much shorter computation times are achieved in practice.

The method was applied to data pertaining to cases related to Covid-19 in Greece, the USA and the UK. A common characteristic in all studied datasets was a diffusion of varying strengths, primarily due to weekly seasonality. On the other hand, there were differences between the datasets of the three countries, which helped in showcasing the performative capabilities of the method: Greece generally provided a more favourable dataset for purposes of both approximation and trend detection, the USA exhibited a more commonly diffuse dataset and the UK bore an irregular spike whose quick ascent and descent momentarily commandeered a comparatively extended, milder wave. As

a complement to this, the supplementary material contains a similar analysis of Covid-19 deaths data for each country in the same studied period, the contents thereof helped in understanding both the relationship between cases and deaths, as well as the deaths data themselves.

Several points of interest were identified in the output obtained through this process. It became clear that the level of diffusion in the data is one of the chief factors that affect the capabilities of the method, in terms of approximation. Generally, the approximation splines returned by the method serve to detect trends in the data, even if the data are misapproximated on account of the weekly seasonality that characterizes diffusion in the data. However, in some cases, such as the new cases data over the June-September period in the US, the diffusion can adversely affect the accuracy of the point of inflection suggested by the splines. On the other hand, the long-term analysis of the output proved to be unaffected by relatively short-term irregularities in the data. For instance, in the UK new cases data, while the July spike in the data served to severely alter spline structure in the June-July spline, subsequent splines revealed that it was indeed an irregularity, while showing a more accurate representation of what the wave would be without the spike in cases. Furthermore, a weekly-basis analysis was carried out and provided additional insights pertaining to both behaviours of forward and backward movement in the inflection range of the splines, on a more dynamic timescale compared to the monthly additions.

The former instances also served to highlight the fact that, when introducing new data to the optimization calculation, a rightward movement of the inflection point (i.e. towards the more recent values, in terms of time series) is not a foregone conclusion. Diffusion in the data may serve to obfuscate the true nature and/or location of the inflection point until a sufficient amount of data are introduced. That being said, when several successive splines report the same (or quite nearly so) inflection range, it is strongly suggested to be the actual one. Any erratic movement of the inflection point, be it forwards or backwards, was, insofar as the studied data revealed, a temporary effect, mitigated once convexity-concavity was made clearly present in the data.

The output provided by the method, be it through short-term or long-term application, can provide insight that may well support processes of decision making. In the case of Covid-19 time series, the method quite handily provides immediate indicators regarding the status of a given studied Covid-19 wave, in terms of both the convex-concave linear spline fit to the Covid-19 data and the corresponding inflection point, as shown in Tables 1–4 for the purposes of our presentation. Indeed, the fit reveals the rates of change of the underlying sigmoid function. Moreover, the spline representation of the data uses the minimal number of parameters, where the density of the number of spline knots to the number of data was kept between 1/7 and 1/14 as $n = 30$ and $n = 120$ respectively. The nature of the inflection point suggested by sequential splines gives indication regarding the growth phase of the studied wave: a steadily proceeding inflection point shows a wave in growth, which could call for policies to restrict transmission; a stationary inflection point most likely means that the wave starts to decelerate, which could have policy makers prepare for a peak and eventual reduction in the Covid-19 measures; while a drastically forward-shifting inflection point, coupled with a drastic change in spline structure, heralds a new Covid wave. An important aspect of this is that the inflection point was found to regress on occasion,

primarily in instances of overly diffuse or otherwise irregular data; depending on the scenario, this retrogression was either in favour of or against an accurate reading of where the inflection point is located. As such, one should always bear in mind the nature of the analyzed data itself; if the data are unfavourable towards the method, as in the UK cases spike or the overly diffuse USA cases data, their short-term results may be misleading by their lonesome. Thus, it is imperative that a decision-maker frequently run the method as datasets are updated. For one, insight from previous runs made on earlier data may well be used in analyzing the results from subsequent runs of the method (such as identifying concave-convex inflection through intersections of succesive splines, an example of which appears in Fig. 4). Furthermore, other properties of the data themselves, such as the seasonal diffusion present in Covid-19 data, can assist in analysis when identified, since they shape the lens through which one may view the results. These steps need be taken in order for the method to be used to its greatest effect.

Covid-19 data are not themselves considered exactly known. Therefore an immediate question is to what extent do these values influence the values of the solution components. The results of this analysis do confirm our convexity- concavity assumption for the evolution of the pandemic. That is, the linear spline fitting as the data entered the calculation is robust with respect to the uncertainty of the Covid-19 values. We remind of the variety in the datasets the method was applied to; each dataset may be characterized by a behaviour that separates it from the rest – be it diffusion, irregularities, both or lack thereof—but not once did the method fail to provide a convex-conave that adequately describes the data. Even in cases of misapproximation due to diffusion, the ability of the method to detect the general trends in the data held strong, thus providing pertinent information for Covid-19 policy. Apart from the real data considered here, this is also the conclusion of Demetriou (2015) when the quadratic programming algorithm of our technique, that gives the convex-concave linear spline fit, was applied to simulated data with infinitesimal changes.

The technique presented here may be valuable for applications to a variety of situations in a variety of ways, partly because it provides useful parameters for the phenomenon under investigation, and partly because it is very economical. For example, our technique may be employed at each node of the traffic-driven epidemic spreading model by Wu et al. (2021) when this model is applied to a real Covid-19 network, in order to observe the evolution of the pandemic and inform a decision-maker. Furthermore, certain features of our analysis may be combined with probabilistic suggestions of other techniques that take account of the dependence of data on the Covid spread (Lee et al. 2020; Overton et al. 2020).

All the work presented in Sects. 3 and 4 was done after the conference on Global Optimization in July 2021. The method is so new that there has not been enough time for more numerical experiments, to test its efficiency more extensively. However, we have managed to provide a Python interface to associated Fortran software that is very convenient, friendly and fast for interactive computation as the data enter the computation. A programmed version and numerical results and further applications will be published elsewhere by the authors. In addition, supplementary material, that

covers the deaths side of the Covid-19 data presented herein, can be found in the on-line version of this work.

Declarations

Conflict of interest Not applicable.

References

Cullinan MP (2019) Piecewise convex–concave approximation in the mini- max norm. In: Demetriou IC, Pardalos PM (eds) Approximation and Optimization: Algorithms, Complexity and Applications, vol 145. Springer Optimization and Its Applications, New York, pp 83–118 (**Springer Nature**)

Davos DE (2021) Efficient Convex/Concave Data Modelling by FORTRAN software L2CXCV: Analytics, Python Interface Design and Application to pandemic COVID-19 data from November 1st 2020 to February 28th 2021 (Master's Thesis). Department of Economics, National and Kapodistrian University of Athens. (155 pages)

Debecker A, Modis T (2021) Poorly known aspects of attening the curve of covid-19. Technol Forecasting and Social Change, 1–13

Demetriou IC (2004) Least squares convex-concave data smoothing. Comput Optim and Appl 29:197–217

Demetriou IC (2004) A theorem for piecewise convex-concave data approximation. J of Comput and Applied Mathematics 164–165:245–254

Demetriou IC (2006) L2CXCV: A Fortran 77 package for least squares convex/concave data smoothing. Computer Phys Commun 174(8):643–668

Demetriou IC (2015) On the sensitivity of least squares data fitting by nonnegative second divided differences. In: Migdalas A, Karakitsou A (eds.), Optimization, Control and Applications in the Information Age: In Honor of Panos M. Pardalos's 60th Birthday, pp 91–112. Switzerland: Springer Proceedings in Mathematics and Statistics, vol. 130, Springer International Publishing

Demetriou IC, Powell MJD (1991) The minimum sum of squares change to univariate data that gives convexity. IMA J of Numer Anal 11:433–448

Demetriou IC, Powell MJD (1997) Least squares fitting to univariate data subject to restrictions on the signs of the second divided differences. In: Buhmann MD, Iserles A (eds) Approx Theory and Optim, Tributes to M.J.D. powell. Cambridge University Press, Cambridge, England, pp 109–132

Duncan JF (1999) Chemistry of social interactions. Technol Forecasting and Social Change, 167–198

Gompertz B (1815) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S. & c. By Benjamin Gompertz, Esq. F. R. S Proc. R. Soc. Lond. 2:252–253

Lee SY, Lei B, Mallick B (2020) Estimation of covid-19 spread curves integrating global data and borrowing information. PLoS ONE 15(7):e0236860. https://doi.org/10.1371/journal.pone.0236860 (**17 pages**)

Marchetti C (1988) Kondratiev revisited | after one Kondratiev cycle. Int Conf on Regularities of Scientific–Tech Progress and Long-Term Tendencies in Econ Dev, 14 March 1988, 14 March 1988, Novosibirsk, USSR. (22 pages)

Modis T (1993) Technological substitutions in the computer industry. Technol Forecasting and Soc Change 43:157–167

Modis T (1999) Conquering Uncertainty. McGraw-Hill, New York

Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, ... Webb L (2020) Using statistics and mathematical modelling to understand infectious disease outbreaks: Covid-19 as an example. Infectious Disease Modelling, 5:409-441. Retrieved from https://www.sciencedirect.com/science/article/pii/S2468042720300245. https://doi.org/10.1016/j.idm.2020.06.008

Porter ME (1985) Competitive advantage-creating and sustaining superior performance. The Free Press, MacMillan Inc, New York

Powell MJD (1981) Approximation Theory and Methods. Cambridge University Press, Cambridge, England

Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, ... Roser M (2020) Coronavirus pandemic (Covid-19). Our World in Data. (https://ourworldindata.org/coronavirus. June 1st to October 30th, 2021)

Shen CY (2020) Logistic growth modelling of Covid-19 proliferation in China and its international implications. Int J of Infectious Diseases 96:582–589

Verhulst P-F (1838) Notice sur la loi que la population poursuit dans son accroissement. Corresp Math et Physique 10:113–121

Wu Y, Pu C, Zhang G, Pardalos P (2021) Traffic-driven epidemic spreading in networks: Considering the transition of infec- tion from being mild to severe. IEEE Trans on Cybernetics, 1–11. (https://ieeexplore.ieee.org/document/9652473. https://doi.org/10.1109/TCYB.2021.3132791)