ANALYTICAL PARAMETER ESTIMATION OF THE SIR EPIDEMIC MODEL. APPLICATIONS TO THE COVID-19 PANDEMIC.

DIMITER PRODANOV

1) Environment, Health and Safety, IMEC, Belgium
2) MMSIP, IICT, BAS, Bulgaria

Abstract. The dramatic outbreak of the coronavirus disease 2019 (COVID-19) pandemics and its ongoing progression boosted the scientific community's interest in epidemic modeling and forecasting. The SIR (Susceptible-Infected-Removed) model is a simple mathematical model of epidemic outbreaks, yet for decades it evaded the efforts of the community to derive an explicit solution. The present work demonstrates that this is a non-trivial task. Notably, it is proven that the explicit solution of the model requires the introduction of a new transcendental special function, related to the Wright's Omega function. The present manuscript reports new analytical results and numerical routines suitable for parametric estimation of the SIR model. The manuscript introduces iterative algorithms approximating the incidence variable, which allows for estimation of the model parameters from the numbers of observed cases. The numerical approach is exemplified with data from the European Centre for Disease Prevention and Control (ECDC) for several European countries in the period Jan 2020 – Jun 2020.

Keywords: SIR model; special functions; Lambert W function; Wright Omega function MSC: 92D30; 92C60; 26A36; 33F05; 65L09

1. Introduction

The coronavirus 2019 (COVID-19) disease was reported to appear for the first time in Wuhan, China, and later it spread to Europe, which is the subject of the presented case studies, and eventually worldwide. While there are individual clinical reports for COVID-19 re-infections, the present stage of the pandemic still allows for the application of a relatively simple epidemic model, which is the subject of the present report. The motivation behind the presented research was the intention to accurately model the short-term dynamics of the outbreaks of COVID-19 pandemics, which by the time of writing (Sept 2020), has infected more than 27 million individuals worldwide, while by time of the latest submission – more than 43 million cases. The efforts to contain the spread of the pandemic induce sustained social and economic damage. Therefore, the ability to accurately forecast short to medium-term epidemic outbreak's dynamics is of substantial public interest and brings certain sense of urgency to the present work.

The present manuscript gives a comprehensive analytical and numerical treatment of the SIR (Susceptible-Infected-Removed) epidemiological model. The SIR model was introduced in 1927 by Kermack and McKendrick in 1927 to study the plague and cholera epidemics in London and Bombay [9]. To date the SIR model remains as a cornerstone of mathematical epidemiology. It is a deterministic model formulated in terms of ordinary differential equations (ODEs). The model has been extensively used to study the spread of various infectious diseases (see the monograph of Martcheva [11]).

The objective of the present paper is to demonstrate numerical routines for curve-fitting allowing for estimation of the parameters of the SIR model from empirical data. To this end, the paper exhibits an algorithm, which can be used to compute the population variables as functions of time. In contrast to previous approaches, I do not consider the SIR model as an initial value problem but as a problem in the theory of special functions. This change of perspective allows for handling noisy data, e.g. time series having fluctuations caused by delays and accumulation of case reporting. The numerical approach is applied to the COVID-19 incidence and case fatality data in different European countries, having different population densities and dynamics of the epidemic outbreaks.

2. The SIR model

The SIR model is formulated in terms of 3 populations of individuals. The S population consists of all individuals susceptible to the infection of concern. The I population comprises the infected individuals. These persons have the disease and can transmit it to the susceptible individuals. The R poulation cannot become infected and the individuals cannot transmit the disease to others. The model comprises a set of three ODEs:

$$\dot{S}(t) = -\frac{\beta}{N} S(t) I(t) \tag{1}$$

$$\dot{I}(t) = \frac{\beta}{N} S(t) I(t) - \gamma I(t) \tag{2}$$

$$\dot{R}(t) = \gamma I(t) \tag{3}$$

The model assumes a constant overall population $N = S + I + R$. An disease carrier infects on average β individuals per day, for an average time of $1/\gamma$ days. The β parameter is called *disease transmission rate*, while γ – *recovery rate*. The average number of infections arising from an infected individual is then modelled by the number $R_0 = \frac{\beta}{\gamma}$, the *basic reproduction number*. Typical initial conditions are $S(0) = S_0, I(0) = I_0, R(0) = 0$ [9].

The model can be re-parametrized using normalized variables as

$$\dot{s} = -si \tag{4}$$

$$\dot{i} = si - gi, \quad g = \frac{\gamma}{\beta} = \frac{1}{R_0} \tag{5}$$

$$\dot{r} = gi, \tag{6}$$

subject to normalization $s + i + r = 1$ and time rescaling $\tau = \beta t$. Therefore, since $i(\tau)$ is integrable on $[0, \infty)$ then $i(\infty) = 0$.

3. The analytical solution

The analytical solution will be formulated first in an implicit form. Since there is a first integral by construction the system can be reduced to two equations:

$$\frac{di}{ds} = -1 + \frac{g}{s} \tag{7}$$

$$\frac{di}{dr} = \frac{s}{g} - 1 \tag{8}$$

Remark 1. *From this formulation*

$$R_e = N\frac{s_0}{g} = \frac{S_0\beta}{\gamma} \geq 1$$

must hold for the infection to propagate. R_e is called the effective reproductive number, while the basic reproduction number is $R_0 = R_e N$ [21].

In order to solve the model we will consider the two equations separately. Direct integration of the equation 7 gives

$$i = -s + g\log s + c$$

where the constant c can be determined from the initial conditions. In the present treatment, the constant c will be left indeterminate to be assigned by the different re-parametrization procedures. The s variable can be expressed explicitly in terms of the Lambert W function [4]:

$$s = -gW_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) \tag{9}$$

where the signs denote the two different real-valued branches of the function. Note, that both branches are of interest since the argument of the Lambert W function is negative. Therefore, the ODE 5 can be reduced to the first-order autonomous system

$$\dot{i} = -ig\left(W_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) + 1\right) \tag{10}$$

valid for two disjoined domains on the real line. The ODEs can be solved for the time τ as

$$-\int \frac{di}{i\left(W_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)+1\right)} = g\tau \tag{11}$$

Remark 2. *There is another equivalent form of the system using the Wright Ω function* [5] *since*

$$W\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) = \Omega\left(\frac{i-c}{g} - \log\left(-g\right)\right)$$

so that

$$\dot{i} = -gi\left(\Omega\left(\frac{i-c}{g} - \log g \mp i\pi\right) + 1\right)$$

Remarkably, Bronstein [1] *gave a proof that the Lambert W and Wright Ω are not Liouvillian.*

The s variable can be determined by substitution in equation 4, resulting in the autonomous system

$$\dot{s} = -s\left(-s + g\log s + c\right) \tag{12}$$

which can be solved implicitly as

$$\int \frac{ds}{s\left(s - g\log s - c\right)} = \tau \tag{13}$$

Finally, the r variable can also be conveniently expressed in terms of i. For this purpose we solve the differential equation

$$\frac{dr}{di} = \frac{g}{s-g} = \frac{-1}{1 + W_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)}$$

Therefore,

$$r = c_1 - g\log\left(-gW_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)\right) = c_1 - g\log s$$

by Prop. 4. On the other hand,

$$g\log\left(-gW\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)\right) = g\left(\log\left(\frac{e^{\frac{i-c}{g}}}{g}\right) - W\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)\right) =$$
$$-gW\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) + i - g\log g - c = s + i - g\log g - c$$

So that

$$r = gW\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) - i + c_1$$

For the purposes of curve fitting we assume that $i(-\infty) = r(-\infty) = 0$. Therefore,

$$c_1 = -gW_{-}\left(-\frac{e^{-\frac{c}{g}}}{g}\right)$$

Furthermore,

$$\frac{ds}{dr} = -\frac{s}{g}$$

whence $\log s = -\frac{r}{g} + \frac{c_1}{g}$ in accordance with the previous result.

3.1. Peak value parametrization.

The upper terminal of integration can be determined by the requirement for the real-valuedness of i. This value of i is denoted as i_m; that is

$$W_{\pm}\left(-\frac{e^{\frac{i_m-c}{g}}}{g}\right) = -1$$

Therefore,

$$i_m = c - g\log g - g \tag{14}$$

The peak-value parametrization is supported by the following result.

Proposition 1. *$i(t)$ attains a global maximum $i = i_m = c - g\log g - g$.*

Proof. We use a parametrization for which $i(0) = i_m$. Then

$$\dot{i}(\tau) = -gi\left(W_\pm\left(-e^{\frac{i-i_m}{g}-1}\right)+1\right)$$

It follows that

$$\dot{i}(0) = 0, \quad i(0) = i_m$$

so i_m is an extremum. In the most elementary way since W(z) should be real-valued then

$$-e^{\frac{i-i_m}{g}-1} \geq -1/e \Longrightarrow \frac{i-i_m}{g} \leq 0$$

Hence, $i \leq i_m$. □

If we consider formally the phase space $\left(z \times y = -gz\left(W_\pm\left(-e^{\frac{z-i_m}{g}-1}\right)+1\right)\right)$ the following argument allows for the correct branch identification. For $i \to -\infty$ $W_-\left(-e^{\frac{z-i_m}{g}-1}\right) \to -\infty$ so $y < 0$; while $W_+\left(-e^{\frac{z-i_m}{g}-1}\right) \to 0^+$ so $y > 0$. Therefore, if we move the origin as $t(0) = i_m$ then conveniently

$$-\int_{i_m}^{i} \frac{dz}{z\left(W_+\left(-e^{\frac{z-i_m}{g}-1}\right)+1\right)} = g\tau, \quad \tau > 0 \tag{15}$$

$$-\int_{i_m}^{i} \frac{dz}{z\left(W_-\left(-e^{\frac{z-i_m}{g}-1}\right)+1\right)} = g\tau, \quad \tau \leq 0 \tag{16}$$

Furthermore, the recovered population under this parametrization is

$$r = gW_\pm\left(-e^{\frac{i-i_m}{g}-1}\right) - gW_-\left(-e^{\frac{-i_m}{g}-1}\right) - i \tag{17}$$

under the same choice of origin.

3.2. **Initial value parametrization.** As customarily accepted, the SIR model can be recast as an initial value problem. The indeterminate constant c can be eliminated using the initial condition

$$i_0 = -s_0 + g\log s_0 + c$$

Therefore,

$$i = i_0 + s_0 - s + g\log s/s_0 = 1 - s + g\log\frac{s}{1-i_0} \tag{18}$$

For this case, the following autonomous differential equation can be formulated:

$$\dot{i} = -gi\left(W_\pm\left(-\frac{1-i_0}{g}e^{\frac{i-1}{g}}\right)+1\right) \tag{19}$$

This can be solved implicitly by separation of variables as

$$-\int_{i_0}^{i} \frac{dz}{z\left(W_-\left(-\frac{1-i_0}{g}e^{\frac{z-1}{g}}\right)+1\right)} = g\tau, \quad \tau \leq t_m \tag{20}$$

$$-\int_{i_0}^{i} \frac{dz}{z\left(W_+\left(-\frac{1-i_0}{g}e^{\frac{z-1}{g}}\right)+1\right)} = g\tau, \quad \tau > t_m \tag{21}$$

It is noteworthy that the time to the peak of infections t_m can be calculated as

$$t_m = \int_0^{\log g/s_0} \frac{du}{s_0 e^u - gu - (s_0 + i_0)}$$

The result follows by considering the autonomous system 7 and fixing the upper terminal of integration $s = g$. However, by Prop. 3 this definite integral can be evaluated only numerically.

4. Is the incidence function "new"?

The incidence i-function of the SIR model appears to be an interesting object of study on its won. One may pose the question about the representation of this function by other, possibly, elementary functions. The answer to this question is in the negative as will be demonstrated below. In precise terms, this function is non-Liouvillian. However, this does not mean that the function can not be well approximated. Fortunately, this is the case, as the incidence function can be approximated for a sufficiently wide domain of parameter values by Newtonian iteration.

Definition 1. *An elementary function is defined as a function built from a finite number of combinations and compositions of algebraic, exponential and logarithm functions under algebraic operations (+,-,., /)*

Allowing for the underlying field to be complex numbers – $\mathbb{C}$, trigonometric functions become elementary as well.

Definition 2 (Liouvillian function)**.** *We say that $f(x)$ is a Liouvillian function if it lies in some Liouvillian extension of $(C(x),')$ for some constant field C.*

As a first point we establish the non-elementary character of the integral in eq. 11. The necessary introduction to the theory of differential fields is given in the Appendix B. From the work of Liouville it is known that if a function $F(x) = f(x)e^{q(x)}$, where f, q are elementary functions, has an elementary anti-derivative of the form [20]

$$\int F(x)dx = \int f e^q dx = h e^q$$

for some elementary function $h(x)$ [2]. Therefore, differentiating we obtain

$$f e^q = h' e^q + h q' e^q$$

so that if $e^q \neq 0$

$$h' + hq' = f$$

holds. The claim can be strengthened to demand that h be algebraic for algebraic f and q (see Th. 1).

Theorem 1. *The integrals*

$$I_\pm(\xi) = \int \frac{d\xi}{\xi \left(W_\pm \left(-\frac{e^{\frac{\xi - c}{g}}}{g} \right) + 1 \right)}$$

are not Liouvillian.

Proof. We use i_m parametrization. Let $c = i_m + g - g \log g$. The proof proceeds by change of variables – first $\xi = g \log y - yg + g + i_m$; followed by $z = \log\left((g \log y + i_m + g)/g\right)$.

$$I = \int \frac{d\xi}{\xi \left(W_{\pm} \left(-e^{\frac{\xi - i_m}{g} - 1} \right) + 1 \right)} = \int \frac{y-1}{y \left(g \log(y) - gy + i_m + g \right) \left(W\left(-ye^{-y}\right) + 1 \right)} dy$$

$$= - \int \frac{dy}{y(g\log(y) - gy + i_m + g)} = \frac{1}{g} \int \frac{e^{z + \frac{i_m}{g} + 1}}{e^{e^z} - e^{z + \frac{i_m}{g} + 1}} dz$$

since $W_{\pm}\left(-ye^{-y}\right) = -y$. The last integral has the form

$$\int \frac{Ae^z}{e^{e^z} - Ae^z} dz$$

which allows for the application of the Liouville theorem in the form of Corr. 1. We can identify

$$\int fe^z dx = he^z, \quad f(z) = \frac{A}{e^{e^z} - Ae^z}, \quad A = e^{\frac{i_m}{g} + 1}$$

so that

$$\frac{A}{e^{e^z} - Ae^z} = h'(z) + h(z)$$

for some unknown algebraic $h(z)$. Since the left-hand side of the equation is transcendental in z so is the right-hand side. Therefore, the integrand has no elementary antiderivative. □

The proof establishes also the validity of two additional propositions:

Proposition 2. *The integral*

$$I = \int \frac{dy}{y\left(g \log y - gy + c\right)}$$

is not Liouvillian.

Proposition 3. *The integral*

$$I = \int \frac{dy}{y + c - e^y}$$

is not Liouvillian.

Proof. By change of variables $y = e^x$

$$I = \int \frac{dy}{y\left(\log y - y + c\right)} = - \int \frac{dx}{e^x - x - c}$$

□

Theorem 2. *The incidence function $i(t)$, defined by the differential eq. 10, is not Liouvillian.*

Proof. For the present case, let us assume that $i(0) = i_m$ so that i attains the maximum by Prop 1. Therefore,

$$i' = -gi \left(W \left(-e^{\frac{i - i_m}{g} - 1} \right) + 1 \right)$$

Without loss of generality let $g = 1$, which amounts to scaling of the solution by the factor of $1/g$.

Suppose further that

$$i - i_m - 1 = \log u - u$$

for some algebraic function u (log-extension case). Then

$$-\frac{i'}{i} = W\left(-e^{\log u - u}\right) + 1 = -u + 1$$

On the other hand,

$$-\frac{i'}{i} = \log\left(\log u - u + i_m + 1\right)' = -\frac{1-u}{u\left(\log u - u + i_m + 1\right)} u'$$

so that

$$-\frac{1-u}{u\left(\log u - u + i_m + 1\right)} = \frac{1-u}{u'}$$

However, if u is algebraic so is u' by Th. 3. Therefore, we have a contradiction, since the left-hand-side is transcendental. Hence, i is not part of a logarithmic elementary extension.

Suppose that u is exponential, i. e. $u = e^f$ for some algebraic function f. In this case,

$$\frac{\left(e^f - 1\right) f'}{-e^f + f + i_m + 1} = 1 - e^f \to -f' = -e^f + f + i_m + 1$$

However,

$$f' + f + i_m + 1 = e^f$$

Therefore, the right-hand-side is exponential and can not be algebraic as it is demanded by Corr 1. This is a contradiction, hence f can not be algebraic, hence u is not part of an exponential elementary extension.

Finally, suppose that $i(t)$ is algebraic. Since the Wright function $\Omega(z)$ is transcendental [5] it follows that $W\left(-e^{i-i_m-1}\right) = \Omega\left(i - i_m - 1 + i\pi\right)$ can not be algebraic in i. Therefore, i'/i and hence i must be transcendental. Hence, the case of an algebraic $i(t)$ can not hold either.

In summary all three cases are rejected, therefore, i is not Liouvillian. □

4.1. **A functional equation for the incidence function.** The result of the last paragraph leaves the question about the form of $i(\tau)$ somehow wanting. Here we demonstrate a functional equation for i exhibiting its non-Liouvillian character. Ritt established that if an elementary function has an elementary inverse it is a sequential composition of algebraic and transcendental functions [19]. Prelle and Singer proved in Corollary 3 that if the autonomous system $y' = f(y)$ has an elementary first integral then

$$g(y) = \int \frac{dy}{f(y)}$$

is also elementary [16]. This presents a direct way of proving that the incidence function is non elementary by virtue of Th. 1 but leaves open the question about its non-Liouvillian character. In order to clarify the form of $i(\tau)$ make use of the integral $I(\xi)$. On the first place, the following identity holds for its kernel:

$$\frac{1}{\xi\left(W_\pm\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right) + 1\right)} = \frac{1}{\xi} - \frac{W_\pm\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)}{\xi\left(W_\pm\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right) + 1\right)}$$

Therefore,

$$I_{\pm}(\xi) = \log \xi - \int \frac{W_{\pm}\left(-\frac{e^{\frac{\xi - c}{g}}}{g}\right) d\xi}{\xi\left(W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)+1\right)} =$$

$$= \log \xi + \int \frac{dz\,\left(\frac{g}{z}-1\right) W\left(-\frac{e^{\frac{g\log z - z}{g}}}{g}\right)}{(g\log z - z + c)\left(W\left(-\frac{e^{\frac{g\log z - z}{g}}}{g}\right)+1\right)}\Bigg|_{z=-W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)} =$$

$$\log \xi + \int \frac{dz}{g\log z - z + c}\Bigg|_{z=-W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)} = \log \xi + L\left(-W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)\right)$$

where we have used the identity

$$W\left(-\frac{z\,e^{-\frac{z}{g}}}{g}\right) = -\frac{z}{g}$$

and defined

$$L(z) := \int \frac{dz}{g\log z - z + c} \qquad (22)$$

which is a Liouvillian extension non-elementary integral by Prop. 6. Exponentiating and using the notation $L(z)$ it can be seen that

$$\exp\left(I_{\pm}(\xi)\right) = \xi + e^{L\left(-W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)\right)} = e^{-g\tau}$$

However, $\xi = i(\tau)$; therefore

$$i(\tau) = e^{-g\tau - L\left(-W_{\pm}\left(-\frac{e^{\frac{i(\tau)-c}{g}}}{g}\right)\right)} \qquad (23)$$

from which we can infer the non-Liouvillian character of $i(\tau)$ from the form of the equation.

5. Series solutions

We will give two series for the i-function in view of the different re-parametrizations of the SIR model.

5.1. Series for the i_m parametrization. The natural parametrization is fixing the peak at the origin. The Taylor development can be computed as follows:

$$i(t) = i_m - \frac{{i_m}^2 g}{2}t^2 + \frac{{i_m}^3 g}{6}t^3 + \frac{\left(4{i_m}^3 g^2 - {i_m}^4 g\right)}{24}t^4 - \frac{\left(15{i_m}^4 g^2 - {i_m}^5 g\right)}{120}t^5 + \ldots \qquad (24)$$

and for the logarithm

$$\log i(t) = \log i_m - \frac{i_m g}{2}t^2 + \frac{{i_m}^2 g}{6}t^3 + \frac{\left({i_m}^2 g^2 - {i_m}^3 g\right)}{24}t^4 - \frac{\left(5{i_m}^3 g^2 - {i_m}^4 g\right)}{120}t^5 + \ldots \qquad (25)$$

5.2. **Series for the i_0-parametrization.** The Taylor series starting from an initial value i_0 is

$$i(t) = i_0 - i_0\,(i_0+g-1)\,t + \frac{g\,i_0\left(4{i_0}^2+5g\,i_0-7i_0-3g+3\right)}{2\,(i_0-1)}t^2 - \frac{g\,i_0\left(5{i_0}^3+21g\,{i_0}^2-9{i_0}^2+18g^2\,i_0-26g\,i_0+4i_0-10g^2+7g\right)}{6\,(i_0-1)}t^3+\ldots \quad (26)$$

The Taylor series for the logarithm starting from an initial value i_0 is

$$\log i(t) = \log(i_0) - (i_0+g-1)\,t + \frac{g\,i_0\,(i_0+2g-1)}{2\,(i_0-1)}t^2 - \frac{g\,i_0\,(i_0+2g-1)\,(2i_0+g-1)}{6\,(i_0-1)}t^3+\ldots \quad (27)$$

The series follow directly from successive differentiation of the differential equation 19.

6. Numerical approximation

The i-function can be efficiently approximated by the Newton's method. The Newton iteration scheme is given as follows for the c-parametrization:

$$i_{n+1} = i_n - i_n\left(W_{\pm}\left(-\frac{e^{\frac{i_n-c}{g}}}{g}\right)+1\right)\left(\int_{g\log g-g+c}^{i_n}\frac{d\xi}{\xi\,\left(W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)+1\right)}+gt\right)$$

This is a conceptually simple representation. However, it has the disadvantage of using the Lambert function for the quadrature routine. Another equivalent representation is

$$i_{n+1} = i_n + g i_n\left(W_{\pm}\left(-\frac{e^{\frac{i_n-c}{g}}}{g}\right)+1\right)\left(t-\int_{g}^{-gW_{\pm}\left(-\frac{e^{\frac{i_n-c}{g}}}{g}\right)}\frac{dy}{y\,\left(g\log y-y+c\right)}\right)$$

(see Prop. 5). This form has the advantage of requiring only 1 Lambert function evaluation per iteration.

A point of attention here is the choice of the initial value for the iteration scheme. Despite my best efforts, a rigorous analytical asymptotic valid on the entire real line and for all parameter values could not be found. Numerical experiments gave acceptable results using the formula

$$f(x) = b\,e^{1-xc-e^{-xc}}$$

$g>0,\ c=\sqrt{2bg}$ or $c=2\sqrt{bg/e}$, and additionally $c \leftarrow c/\sqrt{e}$, $x>0$ for the initial value of $i_0 = f(t)$.

6.1. **Plots.** Plots of the branches of the integral $I(x)$ (Fig. 1) were obtained by direct numerical integration using the QUADPACK [14] routines in the Computer Algebra System Maxima. Plots of the SIR model (Fig. 2) were obtained using a Java routine [17] implementing the double exponential integration method [12, 13]. Both methods turn out to be suitable for the numerical integration problem.

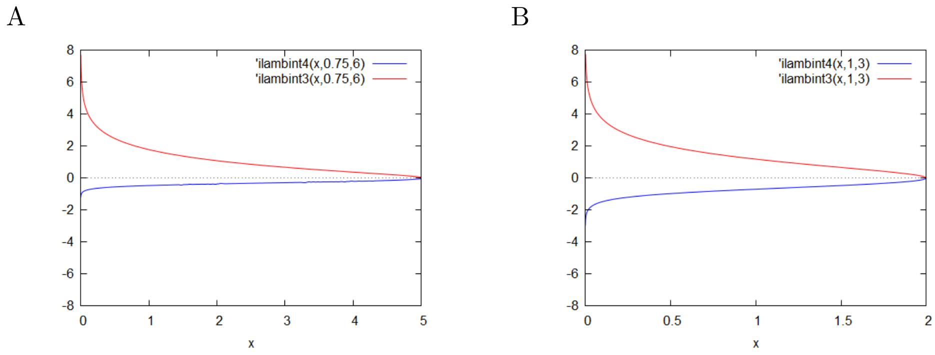

FIGURE 1. Plots of the integrals $I_{\pm}(x)$
A – c-parametrization with parameters g= 0.75, c = 6.0; B – c-parametrization with parameters g=1.0, c= 3.0. The negative branch is below 0.

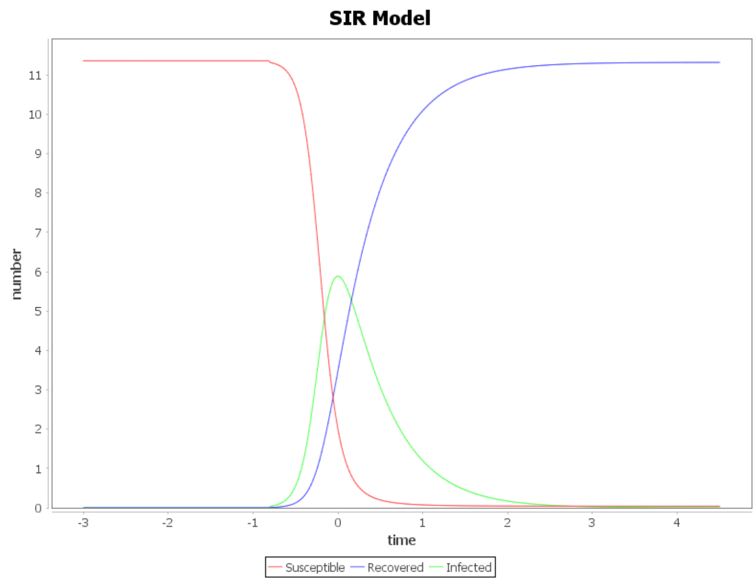

FIGURE 2. The SIR model variables as functions of time
The instance is parametrized by $i_m = 6.5$, $g = 2.0$.

7. Datasets

The COVID datasets were downloaded from the European Centre for Disease Prevention and Control (ECDC) website: `https://opendata.ecdc.europa.eu/covid19/casedistribution/csv`. The downloadable data file is updated daily and contains the latest available public data on COVID-19. Each row/entry contains the number of new cases reported per day and per country. The data collection policy is available from `https://www.ecdc.europa.eu/en/covid-19/data-collection`.

8. Data analysis

8.1. **Time series processing.** The downloaded data were imported in a SQLite `https://www.sqlite.org` database and further filtered by country. Subsequently, the country-specific time series were transferred to MATLAB for parametric fitting using native routines implementing the approach described above. Fitted parameters were stored in the same database. This allowed for easier cross-country comparisons.

8.2. **Parametric fitting.** The parametric fitting was conducted using the fminsearchbnd routine, which allows for constrained optimization. To reduce the impact of the fluctuation in the weekly reporting of data the parametric fitting procedure was applied first to 7 day moving average of the time series. The fitting algorithm is exemplified with datasets from Belgium, the Netwherlands, Italy, Germany and Bulgaria for the period Jan 2020 – Jun 2020. The fitting equation is given by

$$I_t \sim NI\left(t/10.0 - T|g, i_m\right)$$

where I_t is the observed incidence or mortality, respectively, while N, T, g and i_m are estimated from the data. For numerical stability reasons the time during the fitting procedure is rescaled by a factor of 10.

9. Results

The observed case fatality represented a parameter, which could be established with more confidence in the beginning of the pandemic due to the lack of testing and the non-specificity of the clinical signs of COVID-19. Hence, it was the primary target of the parametric fitting.

9.1. **Analysis of case fatality data.** The data fitting procedure is illustrated with the case fatality data of Belgium and are presented in Fig. 3. The T, N and i_m are estimated from the observed data. For the g parameter an initial estimate of 0.75 is used (i.e. $R_0 = 1.33$). The intermediate parameters are initially estimated on the 3-day moving average data. The final fit was performed on the raw data, using the intermediate parameters as initial values. The cumulative mortality data were estimated from eq. 17. As can be appreciated from Fig. 3 fluctuations in reporting did not have a detrimental effect on the estimation procedure. The results for Germany, Italy, Belgium and the Netherlands are presented in Table 1.

Country	g	R_0	T[weeks]	i_m
Belgium	0.7380	1.3549	14.8	313.97
Netherlands	0.5009	1.9962	14.1	156.72
Germany	0.6109	1.6370	15.1	226.51
Italy	0.3734	2.6760	12.8	785.39

Table 1. Case fatality parameters

T is given in weeks and refers to the time passed since 1^{st} Jan 2020.

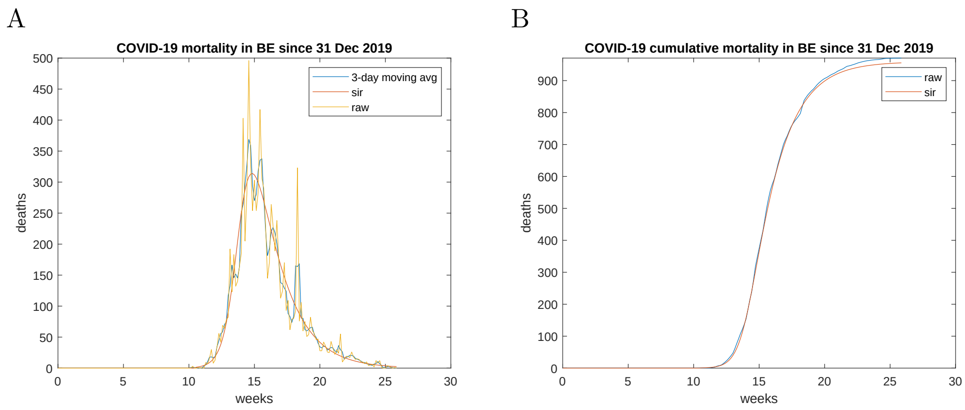

FIGURE 3. Case fatality model for Belgium
A – The parametric fit of the case fatality data; B – Cumulative deaths compared to the estimate from the r-variable.

9.2. **Analysis of incidence data.** The raw data demonstrate weekly fluctuations most probably caused by the reporting irregularity due to holidays and fluctuations in the testing demand. It is especially pronounced for the morbidity data of Germany for the presented period. The incidence data were analysed in the same way using the fitted parameters for the mortality as initial values. The fitting results demonstrate different lags of the peaks of incidence vs case fatality in the studied countries. For example for Germany it was 2.2 weeks, while for the Netherlands it was 0.3 weeks.

Country	g	R_0	T[weeks]	i_m
Belgium	0.5500	1.8183	14.0	1499.59
Netherlands	0.5149	1.9420	13.8	1143.25
Germany	0.5483	1.8237	12.9	5383.26
Italy	0.4006	2.4961	12.3	5560.30

TABLE 2. Incidence parameters

T is given in weeks and refers to the time passed since 1^{st} Jan 2020; $R_0 = 1/g$; i_m corresponds to the peak of the case fatality.

9.3. **Tracking of multiple outbreaks.** Changes of containment policies are meant to result in changes in the epidemic outbreak dynamics. This can be followed by the SIR model as demonstrated in the Bulgarian incidence dataset, where resuming of public sports events in the end of June correlates with the 3^{rd} and ongoing increase of incidence (Fig. 5). The process is difficult to automate because of the fluctuations in the data. Nevertheless, it was possible to accurately track the past outbreaks.

10. DISCUSSION

The SIR model was formulated to model epidemic outbreaks [9].The model has only two independent variables and two parameters, which allow for their estimation from data. There is a renewed interest in this model in view of the coronavirus

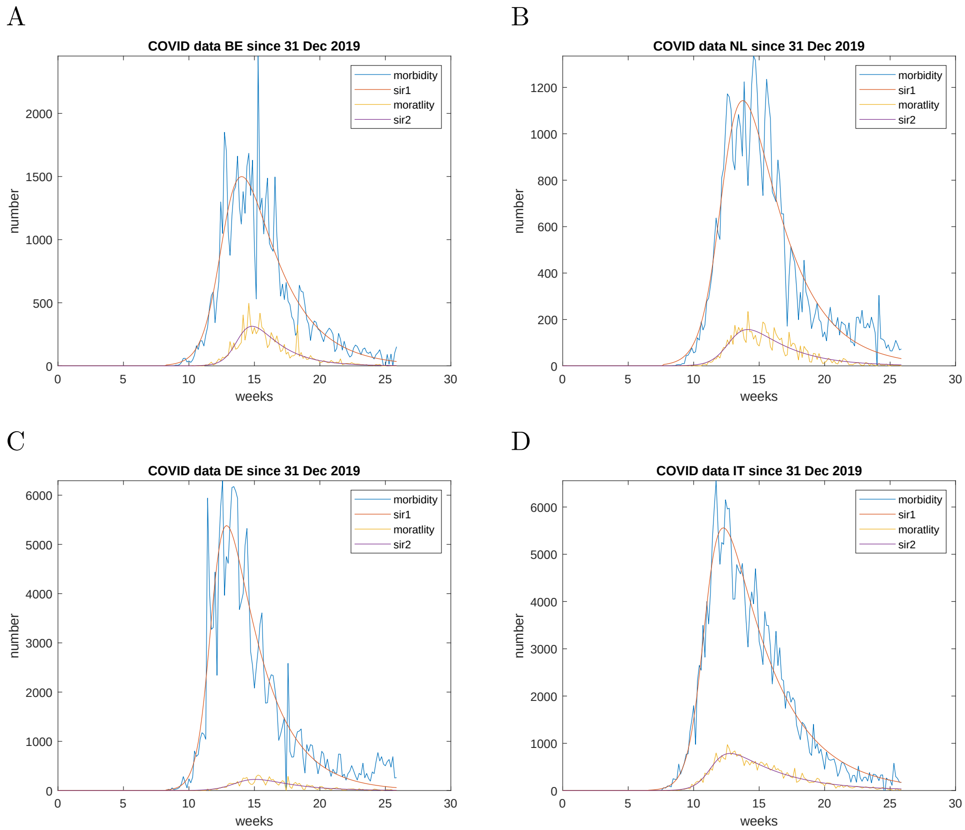

FIGURE 4. Incidence and case fatality model fits for Germany and Italy
A – combined data for Belgium by 29 Jun 2020; B – combined data for The Netherlands by 29 Jun 2020; C – combined data for Germany by 29 Jun 2020; D – combined data for Italy by 29 Jun 2020; The raw data are smoothed with a 3-day moving average filter; mortality represents the case fatality, morbidity represents the incidence.

disease 2019 (COVID-19) pandemics [6, 15, 3, 18]. For decades the model evaded the efforts of the community to derive explicit solution. The formal solution of the SIR model was obtained in 2014 [8] but it was not used for numerical approximation since the model was treated as an initial value problem, where the Runge-Kutta methods work well [11]. The integral of the time parameter in [8, eq. 26] corresponds to the eq. 13. Recently, Kudryashov et al. established some analytical features of the SIR model [10]. The integral form of the implicit solution presented in Kudryashov et al. (eq. 15) is completely different from the integral 11. It is not trivial to prove that this form is equivalent to eq. 11. The Kudryashov et al. [10] fit the data in the phase space. On the other hand, the approach presented here can be used for predictive purposes in time as demonstrated. Yet, another perspective can be also of merit – the model can be treated as a manifold problem, which can be parametrized by any point on the flow. This allows for efficient curve-fitting approach as demonstrated by the presented results. The approach is exemplified

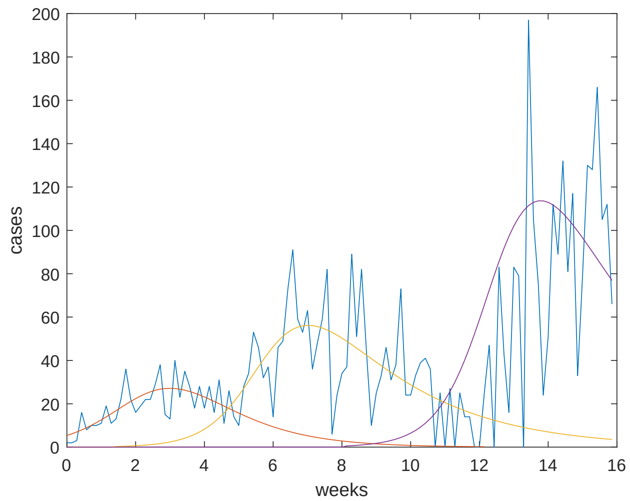

FIGURE 5. Modelling of consecutive outbreaks in Bulgaria by 29 Jun 2020
Raw data are compared to the fitted outbreaks. The origin corresponds to 8 March 2020 when the first COVID19 case was reported.

with data from the ECDC for several European countries for the period Jan 2020 – Jun 2020. The presented results can be discussed along three main directions.

10.1. **Analytical aspects.** On the first place, the present manuscript establishes original analytical results about the SIR model. A novel result of the present work is the form of the series solutions for the i-function 24 and 25. The non-elementarity proofs for the most of the presented integrals are interesting results of their own merits. An original contribution is the proof of the non-Liouvillian character of the incidence i-function. An alternative proof, based on the work of Prelle and Singer [16], is also discussed. This has the advantage of exhibiting a novel functional equation for the incidence i-function and in principle providing a pathway to establishing the asymptotics for the Newton approximation scheme used for the data fitting.

10.2. **Numerical aspects.** Presented results demonstrate the robustness of the fitting procedure with regard to the fluctuations in the raw data. On the other hand, more efforts are necessary in establishing a robust asymptotic of the incidence i-function. This is a clear direction for future research.

On the second place, the analytical formula involving first exponentiation and then computation of the Lambert W function has disadvantages for large arguments due to float under or overflows [5]. Therefore, another special function can be used in principle, notably the Wright Ω function. On the other hand, optimized routines for its calculation are not readily available in MATLAB.

10.3. **Epidemiological implications of the results.** Recorded data necessarily suffer from diverse biases. For example, the marked difference in the availability of tests in the early stages of the pandemics in different countries. Another bias is the unsteady reporting resulting in fluctuations of the numbers. Finally, the spread of a pathogen is a discrete process, which is only continuously approximated by a dynamical model, so bifurcation and chaos phenomena could arise. Altogether, these biases severely limit the usefulness of a model formulated as an initial value problem, which is the standard mathematical practice. This can result in a drastic overestimation of the infection peak (see for example the predictions for UK in [7]).

Presented results indicate that there is a universality in the time evolution of COVID-19 and the same epidemic model, notably SIR, can be applied to countries having large differences in populations sizes and densities. More interestingly, the model seems to fit well also the mortality data, which can be interpreted in the sense that the vulnerable population forms a distinct subpopulation from all susceptible individuals (e.g. elderly people). Presented data lend support to a simple modification of the SIR model: notably – the SIRD model with an independent population of dead (D) persons. This corresponds to the recent findings of other authors [6, 3].

A key finding of the present report is that simple models can be very useful in studying the epidemic outbreaks. This can be eventually extended to predicting the effects of different containment measures or the lack thereof [18].

Appendix A. Special functions

A.1. **The Lambert W function and related integrals.** The Lambert W function can be defined implicitly by the equation

$$W(z)e^{W(z)} = z, \quad z \in \mathbb{C}$$

We observe that by Lemma 1 $W(z)$ is transcendental. Furthermore, the Lambert function obeys the differential equation for $x \neq -1$

$$W(x)' = \frac{e^{-W(x)}}{1 + W(x)}$$

The W function is non-elementary and in particular it is non-Liouvillian [1]. Its indefinite integral is:

$$\int W(x)dx = xW(x) + \frac{x}{W(x)} - x$$

The Lambert W is a multivalued function. Properties of the W function are given in [4]. Useful identities

$$e^{-W(z)} = \frac{W(z)}{z} \tag{28}$$

$$e^{nW(z)} = \left(\frac{z}{W(z)}\right)^n \tag{29}$$

$$\log W(z) = \log z - W(z) \tag{30}$$

$$W\left(\frac{nz^n}{W(z)^{n-1}}\right) = nW(z), \quad n > 0, z > 0 \tag{31}$$

Proposition 4.

$$\int \frac{dy}{1 + W\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)} = g \log\left(-gW\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)\right) + C$$

Proof. We differentiate

$$g\left(\log W\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)\right)' = -\frac{e^{\frac{y-c}{g}-W\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)}}{gW\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)\left(1+W\left(-\frac{e^{\frac{y-c}{g}}}{g}\right)\right)} = \frac{1}{1+W\left(-ge^{\frac{i-c}{g}}\right)}$$

□

Proposition 5.

$$\int_{g\log g-g+c}^{i}\frac{d\xi}{\xi\left(W_{\pm}\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)+1\right)} = \int_{g}^{-gW_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)}\frac{dy}{y\left(g\log y-y+c\right)}$$

Proof. We use the change of variables $\xi-c=g\log y-y$ and then simplification by the defining identity of the Lambert W function.

$$\int_{g\log g-g+c}^{i}\frac{d\xi}{\xi\left(W\left(-\frac{e^{\frac{\xi-c}{g}}}{g}\right)+1\right)} = \int_{A}^{B}\frac{\frac{g}{y}-1}{\left(g\log y-y+c\right)\left(W\left(-\frac{e^{\frac{g\log y-y+c}{g}-\frac{c}{g}}}{g}\right)+1\right)}dy =$$

$$\int_{A}^{B}\frac{g-y}{-\frac{y\left(gy\log y-y^2+cy\right)}{g}+gy\log y-y^2+cy}dy = g\int_{A}^{B}\frac{dy}{y\left(g\log y-y+c\right)}$$

where

$$g\log A-A+c=g\log g-g+c,\quad g\log B-B+c=i$$

Therefore, $A=g$ and $B=-gW\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)$. □

Proposition 6. *$L(z)$ is a non-elementary integral.*

Proof. This can be established by an argument using change of variables as

$$L(z)=\int\frac{dz}{g\log z-z+c}=\int\frac{e^{u}du}{-e^{u}+gu+c}\bigg|_{z=e^{u}}$$

further applying the same reasoning as in the proof of Th. 1. □

A.2. **The Wright Ω function.** The Wright Ω function is related to the Lambert W function [4]

$$\Omega(z)=W_{K(z)}\left(e^{z}\right),\quad z\in\mathbb{C}$$

where $K(z)=\lceil\left(Im(z)-\pi\right)/2\pi\rceil$ is the unwinding number of z. Moreover,

$$\Omega(z)+\log\Omega(z)=z,\quad z\neq t\pm i\pi, t\leq -1,$$

for the principal branch of the logarithm. It is a transcendental function. It obeys the differential equation

$$\Omega(x)'=\frac{\Omega(x)}{1+\Omega(x)}$$

The Ω function is non-Liovillian [1]. Its indefinite integral is:

$$\int\Omega(x)dx=\frac{\Omega(x)^2}{2}+\Omega(x)+C$$

Appendix B. Differential fields

Definition 3. *Denote by* $\mathbb{C}\,(x, c_i, \theta_i)$ *the complex-valued ring, generated by the finite set of rational functions* $\{\theta_i\}_i^n$ *and constants* $\{c_i\}_i^n$.

Definition 4. *An element* θ *is called* algebraic *if* $P(x, \theta) = 0$ *for some polynomial*

$$P(x, t) = t^m + a_{m-1}t^{m-1} + \ldots + a_0,$$

where a_i *can be also rational functions of* x*, or else it is called* transcendental.

Lemma 1 (Composition lemma)**.** *Denote by* a *and* t *the algebraic or transcendental elementary functions, respectively. The following compositions hold*

$$a \circ a = a, \quad t \circ a = t, \quad a \circ t = t$$

Proof. The case $a \circ a$ when $a(x)$ is a polynomial is trivial. Suppose that a and b are both algebraic:

$$P(x, a) = 0, \quad Q(x, b) = 0$$

Without loss of generality suppose that a_i are polynomial. Formally, $b = \bar{f}_k(x)$ for any branch k with $\bar{f}$ algebraic since it is a root of a polynomial, where the bar denotes the inverse function in order to avoid confusion with exponentiation. Therefore,

$$a \circ b = b^m + a_{m-1}b^{m-1} + \ldots + a_0 = \bar{f}_k^m(x) + a_{m-1}\bar{f}_k^{m-1}(x) + \ldots + a_0$$

is algebraic since it is computed by a finite sequence of algebraic operations.

Suppose that $t = exp(x)$. Then $exp(a)$ is not algebraic, hence it is transcendental.

$$P(x, e^x) = e^{xm} + a_{m-1}e^{xm-x} + \ldots + a_0$$

is exponential.

Suppose that $t = log(x)$. Then $log(a)$ is not algebraic, hence it is transcendental.

$$P(x, \log(x)) = \log^m x + a_{m-1}\log^{m-1} x + \ldots + a_0$$

is not algebraic, hence it is transcendental. □

In what follows is assumed that the differential field is of characteristic zero and has an algebraically closed field of constants. An element y of a differential field is said to be an *exponential* of an element A if $y' = Ay$, an exponential of an integral of an element A if $y' = Ay$; logarithm of an element A if $y' = A'/A$, and an integral of an element A if $y' = A$.

The next definition is due to [1].

Definition 5. *Let* $(k,' \equiv d/dx)$ *be a differential field of characteristic 0. A differential extension* $(K,' \equiv d/dx)$ *of* k *is called Liouvillian over* k *if there are* $\theta_1, \ldots, \theta_n \in K$*, such that* $K = C(x, \theta_1, \ldots, \theta_n)$ *and for all i, at least one of the following*

1. θ_i *is algebraic over* $k(\theta_1, \ldots, \theta_{n-1})$
2. $\theta_i' \in k(\theta_1, \ldots, \theta_{n-1})$
3. $\theta_i'/\theta_i \in k(\theta_1, \ldots, \theta_{n-1})$

holds. The constant subfield $C(K)$ *of* K *is defined to be the set of* c *in* K*, such that* $c' = 0$.

The next theorem is due to [2].

Theorem 3. *If* K *is an elementary field, then it is closed under differentiation.*

An elementary integrability theorem due to Conrad [2].

Theorem 4 (Rational Liouville criterion)**.** *For $f, g \in \mathbb{C}(x)$ with f and g non-constant the function $f(x)e^g(x)$ can be integrated in elementary terms if and only if there exists a rational function $h \in \mathbb{C}(x)$ such that $h' + g'h = f$.*

The last result can be extended to algebraic functions as follows.

Corollary 1 (Algebraic Liouville criterion)**.** *For $f(x), g(x)$ algebraic and non-constant, the function $f(x)e^{g(x)}$ can be integrated in elementary terms if and only if there exists an algebraic function $h(x)$, for which $h' + g'h = f$.*

Proof. Suppose that f and g are arbitrary elementary algebraic functions. Denote the primitive of f as $f \div F$. The integral can be integrated by parts

$$I = \int f(x)e^{g(x)}dx = \int e^g(x)dF = F(x)e^{g(x)} - \int F(x)\left(e^{g(x)}\right)' dx$$

Therefore,

$$\int \left(f(x) + F(x)g'(x)\right) e^{g(x)}dx = F(x)e^{g(x)}$$

We observe that $g'(x)$ is elementary by Th. 3. The L.H.S has the form fe^g and since $f(x) + F(x)g'(x)$ is elementary we can identify

$$h \equiv F, \quad f_1 \equiv f + Fg' = h' + hg'$$

so that $(h' + hg')\, e^g = (he^g)'$ and the claim follows. □

References

[1] M. Bronstein, R. M. Corless, J. H. Davenport, and D. J. Jeffrey. Algebraic properties of the Lambert W function from a result of Rosenlicht and of Liouville. *Integral Transforms and Special Functions*, 19(10):709–712, oct 2008.

[2] B. Conard. Impossibility theorems for elementary integration. Technical report, Clay Mathematics Institute, 2005. 2005 Academy Colloquium Series.

[3] I. Cooper, A. Mondal, and C. G. Antonopoulos. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, oct 2020.

[4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, dec 1996.

[5] Robert M. Corless and D. J. Jeffrey. The wright ω function. In *Lecture Notes in Computer Science*, pages 76–89. Springer Berlin Heidelberg, 2002.

[6] D. Fanelli and F. Piazza. Analysis and forecast of COVID-19 spreading in china, italy and france. *Chaos, Solitons & Fractals*, 134:109761, may 2020.

[7] N Ferguson, D Laydon, G Nedjati Gilani, N Imai, K Ainslie, M Baguelin, S Bhatia, A Boonyasiri, ZULMA Cucunuba Perez, G Cuomo-Dannenburg, A Dighe, I Dorigatti, H Fu, K Gaythorpe, W Green, A Hamlet, W Hinsley, L Okell, S Van Elsland, H Thompson, R Verity, E Volz, H Wang, Y Wang, P Walker, P Winskill, C Whittaker, C Donnelly, S Riley, and A Ghani. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.

[8] T. Harko, F. S. N. Lobo, and M. K. Mak. Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, jun 2014.

[9] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, aug 1927.

[10] Nikolay A. Kudryashov, Mikhail A. Chmykhov, and Michael Vigdorowitsch. Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling*, 90:466–473, feb 2021.

[11] M. Martcheva. *An Introduction to Mathematical Epidemiology*. Springer US, 2015.

[12] M. Mori. Quadrature formulas obtained by variable transformation and the DE-rule. *Journal of Computational and Applied Mathematics*, 12-13:119–130, 1985.

[13] M. Mori and M. Sugihara. The double-exponential transformation in numerical analysis. *Journal of Computational and Applied Mathematics*, 127(1-2):287–296, 2001.

[14] R. Piessens, E.de Doncker-Kapenga, C. W. Überhuber, and D. K. Kahaner. *Quadpack*. Springer Berlin Heidelberg, 1983.

[15] E. B. Postnikov. Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals*, 135:109841, jun 2020.

[16] M. J. Prelle and M. F. Singer. Elementary first integrals of differential equations. *Transactions of the American Mathematical Society*, 279(1):215–215, jan 1983.

[17] D. Prodanov. DSP Quadrature library for Java. Technical report, 2020.

[18] N. R. Record and A. Pershing. A note on the effects of epidemic forecasts on epidemic dynamics. *PeerJ*, 8:e9649, aug 2020.

[19] J. F. Ritt. Elementary functions and their inverses. *Transactions of the American Mathematical Society*, 27(1):68–68, jan 1925.

[20] M. Rosenlicht. On the explicit solvability of certain transcendental equations. *Publications Mathématiques de l'IHÉS*, 36:15–22, 1969.

[21] H. Weiss. The SIR model and the foundations of public health. *Mat. Mat.*, 3:1–17, 2013.