ASYMPTOTIC ANALYSIS OF THE SIR MODEL. APPLICATIONS TO COVID-19 MODELLING

DIMITER PRODANOV[1,2]

[1] *Environment, Health and Safety, Leuven, IMEC, Leuven, Belgium*

[2] *MMSDP, IICT, BAS, Bulgaria*

Abstract. The SIR (Susceptible-Infected-Removed) model can be very useful in modelling epidemic outbreaks. The present paper derives the parametric solution of the model in terms of quadratures. The paper demonstrates a simple analytical asymptotic solution for the I-variable, which is valid on the entire real line. Moreover, the solution can be used successfully for parametric estimation either in stand-alone mode or as a preliminary step in the parametric estimation using numerical inversion of the parametric solution. The approach is applied to the ongoing coronavirus disease 2019 (COVID-19) pandemic in three European countries – Belgium, Italy and Sweden.

Keywords: SIR model; Lambert W function; asymptotic analysis; incomplete gamma function ORCID: 0000-0001-8694-0535

MSC: 92D30; 92C60; 65H05; 33-04; 33F05; 33B20

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 analysis, and eventually worldwide. COVID-19 became a global emergency in 2020 and draws a lot of effort in analysing and predicting the course of the pandemic. Very recently, some authors have convincingly demonstrated that the SIR (Susceptible-Infected-Removed) epidemiological model can successfully model the short-term dynamics of COVID-19 [1, 2, 3, 4, 5, 6].

The SIR epidemiological model was introduced by Kermack and McKendrick in 1927 to study the plague and cholera epidemics in London and Bombay [7]. The SIR model can be extended in two directions – either by adding a final state, e.g. a "dead" individuals – D; or by adding one or more intermediate not observable populations – e. g. "exposed" E population.

The analytical solution of the SIR model was obtained in 2014 [8] as an equation of state between the phase space of variables, and later on independently by [5] as a series for the S-variable. Very recently and again independently, an inverse parametric solution for the I-variable has been obtained by [6, 9]. While one may be tempted to question the immediate utility of such an analytical solution, it is worthwhile to discuss it as it borders the theory of special functions, integral equations and autonomous differential equations. Moreover, efficient numerical

E-mail address: dimiter.prodanov@imec.be; dimiterpp@gmail.com.

integration routines are available in all computer algebra systems, such as Maxima, and numerical environments, which enable potential future applications. One can draw an analogy with the advent of the Lambert W function itself, which became popular in applications only relatively recently [10, 11].

Numerical integration routines represent a very useful alternative to other approximation techniques, such as truncated series. The exact analytical solution of the SIR model was obtained by numerical inversion of a non-elementary integral equation by a Newton approximation scheme [9]. However, the numerical stability of the iteration scheme could not be thoroughly studied.

On the other hand, approximate and asymptotic solutions of the SIR model have been established only recently with the works of [5, 12] in terms of approximants.

The objective of the present work is to demonstrate an approximation scheme in terms of standard transcendental functions: exponents and gamma incomplete functions and to compare them with the previously introduced Newton iteration scheme. Moreover, the presented results demonstrate that the approximation scheme also results in asymptotic solutions for the incidence variable valid on the entire real line.

2. Preliminaries of the SIR model

The SIR model has been typically used to study the spread of various infectious diseases (see the monograph of Martcheva [13] or [14]). The SIR model is formulated in terms of 3 populations of individuals [13]. 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 ordinary differential equations ODEs:

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

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

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

By construction, the model assumes a constant overall population $N = S(t)+I(t)+R(t)$ [7]. The interpretation of the parameters is that a 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 $S(0) = S_0, I(0) = I_0, R(0) = 0$ [7]. A phase diagram of the model is plotted in Fig. 1.

For simplicity of the presentation, the SIR 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 as $\tau = \beta t$.

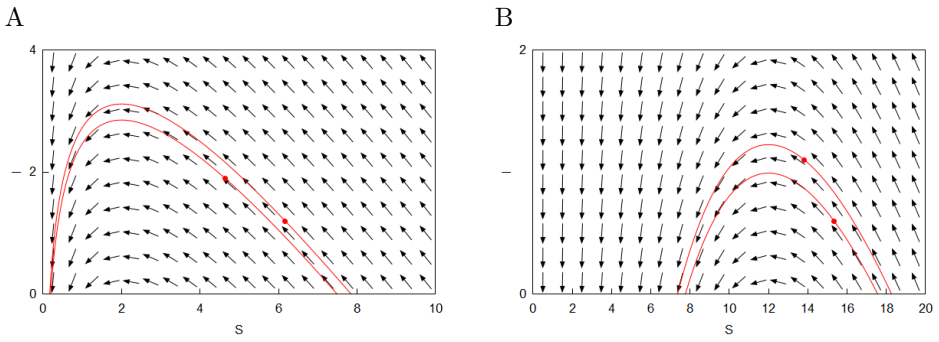

FIGURE 1. Phase plot of the (I, S) plane
A – Case g=2, the trajectories starting at (4.6316, 1.9) and (6.16(6), 1.2) are indicated with dots. This corresponds to the phase plot given in [6, Fig. 1A] for $\alpha = 0.1$, $\beta = 0.2$. B – Case g=12, the trajectories, starting at (13.81(81), 1.1) and (15.3(3), 0.6) are indicated with dots. This corresponds to the phase plot given in [6, Fig. 1B] for $\alpha = 0.1$, $\beta = 1.2$.

Eq. 5 has a fixed point at $s = g$. Moreover, it is a maximum since

$$\ddot{i} = \dot{s}i + s\dot{i} - g\dot{i} = \dot{s}i = -si^2\big|_{i=i(t_p)} < 0$$

and all variables are positive defined.

3. Solution procedure

Since there is a first integral by construction, the system can be reduced to two differential equations in the phase planes (i, s) and (i, r), respectively:

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

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

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

$$i = -s + g\log s + c \tag{9}$$

where the constant c can be determined from the initial conditions.

3.1. The i-variable.

The s variable can be represented explicitly in terms of the Lambert W function (see Appendix A and [11]):

$$s = -gW_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) \tag{10}$$

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, eq. 5 can be reduced to the first-order autonomous system

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

valid for two disjoint domains on the real line. The autonomous ODE can be solved for the rescaled time variable τ as

$$-\int \frac{di}{i\left(W_{\pm}\left(-\frac{e^{\frac{i-c}{g}}}{g}\right)+1\right)} = g\tau \tag{12}$$

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

$$\dot{s} = -s\left(-s + g\log s + c\right) \tag{13}$$

which can be solved implicitly as

$$\int \frac{ds}{s\left(s - g\log s - c\right)} = \tau \tag{14}$$

This is, in fact, the solution obtained in [8].

3.3. **The r-variable.** 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. 1. On the other hand,

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

So that

$$r = gW\left(-\frac{e^{\frac{i-c}{g}}}{g}\right) - i + c + c_1 \tag{15}$$

Since $(s_0, 0, 0)$ is a stable point it follows that

$$c + c_1 = -gW_{-}\left(-\frac{e^{\frac{-c}{g}}}{g}\right) \tag{16}$$

Therefore,

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

4. The parametric solution

4.1. **The i-variable.** The implicit solution eq. 12 can be computed as a definite integral, requiring the computation of the Lambert W function on every integration step. However, this does not seem to be efficient. Alternatively, the solution can be computed more efficiently by change of variables by Prop. 2:

$$\tau(i) = -\int_g^{-gW_{\pm}\left(-e^{\frac{i-c}{g}}/g\right)} \frac{dy}{y\,(g\log y - y + c)} \tag{18}$$

This equation involves quadrature of only elementary functions, which are present in any numerical package. The solution is plotted in Fig. 2.

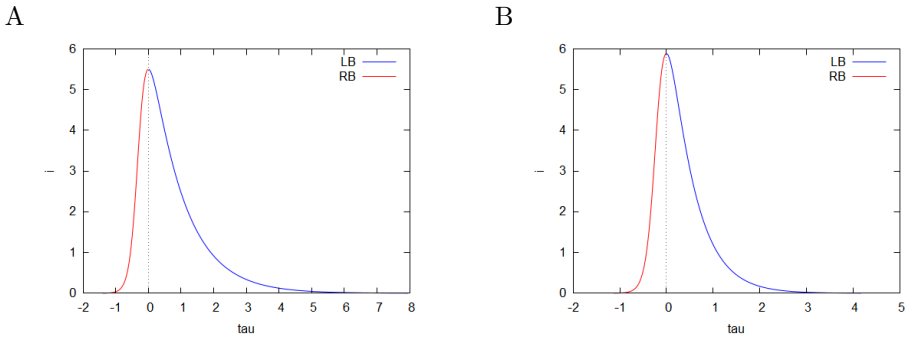

FIGURE 2. Parametric solution $i(\tau)$
Parametric plot of $(\tau(i), i)$ for parameter values A – $c = 6.5$, $g = 1.0$ and B – $c = 6.5$, $g = 2.0$; LB denotes the left branch involving $W_-(x)$, RB denotes the right branch involving $W_+(x)$, eq. 18. Plots were produced using the *quad_qags* Maxima numerical integration command.

4.2. **The s-variable.** The parametric solution can be computed from eq. 9. It follows that

$$\tau(s) = -\int_g^s \frac{dy}{y\,(g\log y - y + c)} \tag{19}$$

where the domain of s is $[-gW_+\left(-e^{-c/g}/g\right), -gW_-\left(-e^{-c/g}/g\right)]$.

4.3. **The r-variable.** The τ-variable can also be expressed as a function of r. We differentiate eq. 15 by s. From where

$$-gt = \int \frac{dr}{ds}dt = \int dr \frac{1}{s\,(s - g\log s - c)}\Big|_{s=e^{(c_1-r)/g}} = \int \frac{e^{\frac{r-c_1}{g}}dr}{(r-(c_1+c))\,e^{\frac{r-c_1}{g}}+1}$$

where $c_1 = -gW_-\left(-e^{-c/g}/g\right) - c$ by eq. 16. Therefore,

$$\tau(r) = \frac{-1}{g}\int_{q-c-g\log g}^{r} \frac{e^{\frac{y}{g}}\,dy}{(y-q)\,e^{\frac{y}{g}}+q}, \quad q = -gW_-\left(-e^{-c/g}/g\right) \tag{20}$$

for the c-parametrization. The domain of r is

$$r \in \left[0, gW_+\left(-e^{-c/g}/g\right) - gW_-\left(-e^{-c/g}/g\right)\right]$$

Remarkably, all of the involved integrals are non-elementary as demonstrated in [9]. Plots of the solutions are presented in Fig. 3.

4.4. **Initial value parametrization.** For parametrization, based on an initial value, the indeterminate constant c can be eliminated using the initial condition

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

Moreover, the peak i_m can be also determined from the initial conditions as

$$i_m = i_0 + s_0 - g - g \log s_0/g$$

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

$$\dot{i} = -gi \left(W_{\pm} \left(-\frac{s_0}{g} e^{\frac{i-(s_0+i_0)}{g}} \right) + 1 \right), \quad s_0 = \frac{\dot{i}_0}{i_0} + g \tag{21}$$

where the negative branch is taken for $i < i_m$ and the positive branch is taken in the alternative case.

It is easy to re-paramterize the solution for given initial values as this amounts to computing the value of c from the initial conditions and adding a time offset to the integrals, amounting to the time to the peak. The time to peak can be calculated from eq. 19 as

$$\tau_m = -\int_g^{s_0} \frac{dy}{y\left(g \log y - y + (s_0 + i_0)\right)} \tag{22}$$

considering that the upper pole of integration is the initial condition for s_0 and the lower pole is g, corresponding to the maximum of the i-variable. In this case, the parametric model can be formulated as

$$(\tau(s) + \tau_m, s)$$
$$(\tau(i) + \tau_m, i)$$
$$(\tau(r) + \tau_m, r)$$

This can be demonstrated for the initial valued problem in Fig. 4.

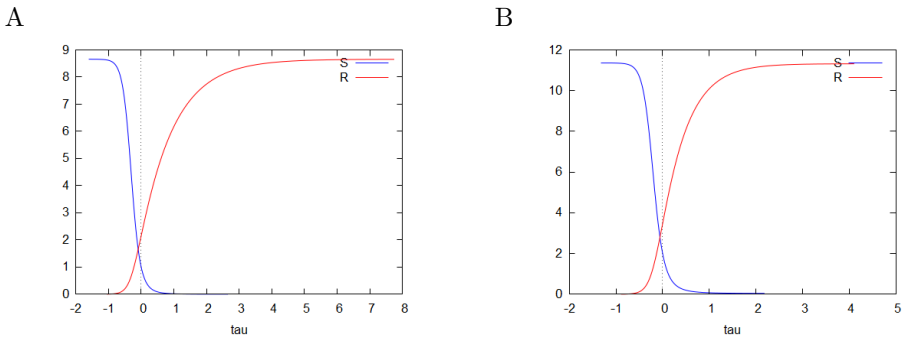

FIGURE 3. Parametric solutions $s(\tau)$, $r(\tau)$
Parametric plots of $(\tau(s), s)$ and $(\tau(r), r)$ for parameter values A – $c = 6.5$, $g = 1.0$ and B – $c = 6.5$, $g = 2.0$; S denotes $\tau(s)$, eq 19. R denotes $\tau(r)$, eq 20. Plots were produced using the *quad_qags* Maxima numerical integration command.

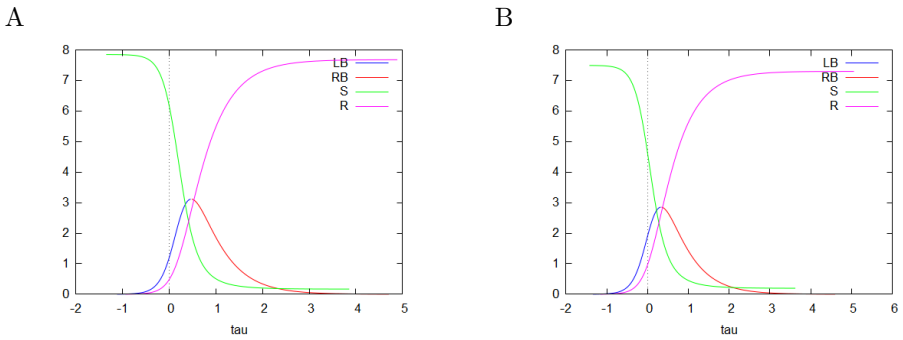

FIGURE 4. Parametric solutions $i(\tau)$, $s(\tau)$, $r(\tau)$
Parametric plots of $(\tau(i), i)$, $(\tau(s), s)$ and $(\tau(r), r)$ for initial conditions A – $i_0 = 1.2$, $i_0' = 5$, $g = 2.0$ and B – $i_0 = 1.9$, $i_0' = 5$, $g = 2.0$; LB denotes the left branch involving $W_-(x)$, RB denotes the right branch involving $W_+(x)$, eq. 18. S denotes $\tau(s)$, eq 19. R denotes $\tau(r)$, eq 20. Plots were produced using the *quad_qags* Maxima numerical integration command.

4.5. **Peak value parametrization.** Alternatively, the solution can be parametrized by the global maximum of the i-variable i_m. In such case, $s_0 = g$ so that

$$\dot{i} = -gi\left(W_{\pm}\left(-e^{\frac{i-i_m}{g}-1}\right)+1\right) \tag{23}$$

4.6. **Computational aspects of the parametric solution.** All three integrals can be efficiently computed by numerical quadratures. Reference implementation in the Computer Algebra System Maxima has been developed and the code is available through the Zenodo repository [16]. Maxima incorporates an efficient numerical integration routine ported from QUADPACK [15]. The integration functions compute a result to a user specified accuracy. Parametric plots of the parametric solution are exhibited in Fig. 2, 3 and 4. The numerical integrals were computed with relative precision 10^{-8}.

5. ASYMPTOTIC ANALYSIS OF THE SIR MODEL

Since the SIR solution is non-singular anywhere in $\mathbb{R}$ one can make use of the Banach Fixed-Point Theorem. Notably, one can use the non-linear approximation scheme of Daftardar-Gejji-Jafari for solving the equivalent integral equations [17]. If we treat the equations of the SIR model as independent we can formally solve eq. 4 as

$$s(\tau) = s_0 e^{-\int i(t)d\tau}$$

On the other hand, eq. 5 can be transformed formally as

$$\dot{i} + gi = si \iff e^{-g\tau} d\left(e^{g\tau} i\right) = si\, d\tau \iff \frac{d\left(e^{g\tau} t\right)}{e^{g\tau} i} = s\, d\tau$$

Therefore, formally,

$$i(\tau) = k_2 e^{-g\tau + \int s(\tau)d\tau}$$

Starting from the 0^{th} order approximation $i^{(0)} \approx i_0$, it follows that

$$s^{(0)} \approx s_0 e^{-i_0\tau}$$

This approximation is valid around the fixed point $s \approx g$, which we can take as an initial condition $s_0 = g$. However, this corresponds to the peak-value parametrization so $i_0 = i_m$, $i'(0) = 0$. From this, the 1st order approximation for the i-variable becomes

$$i^{(1)} = i_m \, e^{\frac{g}{i_m}\left(1-e^{-i_m t}\right)-gt} \tag{24}$$

A second iteration of the loop results in a non-elementary γ integral. Let

$$J := \int e^{\frac{g}{i_m}\left(1-e^{-i_m \tau}\right)-g\tau} d\tau$$

Then, by change of variables

$$J = -\int y^{\frac{g}{i_m}-1} \, e^{\frac{g}{i_m}-\frac{gy}{i_m}} \, dy\Big|_{y=e^{-i_m\tau}} = \frac{\Gamma\left(\frac{g}{i_m}, \frac{g}{i_m}e^{-i_m\tau}\right)}{\left(\frac{g}{i_m}\right)^{\frac{g}{i_m}}}$$

where Γ is the upper incomplete Euler's gamma function. Therefore,

$$r^{(1)} = gJ = k_3 \frac{\Gamma\left(\frac{g}{i_m}, \frac{g}{i_m}e^{-i_m\tau}\right)}{\left(\frac{g}{i_m}\right)^{\frac{g}{i_m}}}$$

Matching the limits results in the equation

$$r^{(1)} = g\left(W_+\left(-e^{-\frac{i_m}{g}-1}\right) - W_-\left(-e^{-\frac{i_m}{g}-1}\right)\right) \frac{\Gamma\left(\frac{g}{i_m}, \frac{g\,e^{-i_m t}}{i_m}\right)}{\Gamma\left(\frac{g}{i_m}\right)} \tag{25}$$

which can be used also for fitting purposes. Then

$$s^{(2)} = k_2 e^{-J} = k_2 \exp\left(-\frac{\Gamma\left(\frac{g}{i_m}, \frac{g}{i_m}e^{-i_m\tau}\right)}{\left(\frac{g}{i_m}\right)^{\frac{g}{i_m}}}\right)$$

Matching the initial condition gives the value of the constant

$$k_2 = g\, e^{\frac{-\Gamma\left(\frac{g}{i_m}, \frac{g}{i_m}\right)e^{\frac{g}{i_m}}}{\left(\frac{g}{i_m}\right)^{\frac{g}{i_m}}}}$$

Therefore,

$$s^{(2)} = g \exp\left(\left(\Gamma(a,a) - \Gamma\left(a, a\,e^{-i_m t}\right)\right)\left(\frac{e}{a}\right)^a\right), \quad a = \frac{g}{i_m} \tag{26}$$

Finally, following the same procedure for the i-variable we obtain

$$i^{(2)} = i_m \, \exp\left(g\, e^q \int_0^\tau e^{-\Gamma\left(a, ae^{-i_m z}\right)\left(\frac{e}{a}\right)^a} dz - g\tau\right), \quad q = \Gamma\left(a,a\right)\left(\frac{e}{a}\right)^a \tag{27}$$

where $a = \frac{g}{i_m}$. The asymptotics of the i-variable are shown in Fig. 5.

Furthermore, let

$$q := \frac{d\log i}{d\log i^{(1)}} = \frac{W_\pm\left(-e^{\frac{i-i_m}{g}-1}\right)+1}{1-e^{-i_m t}}$$

Therefore, for $t \to \infty$

$$q = W_+\left(-e^{\frac{-i_m}{g}-1}\right) + 1 \in [0,1)$$

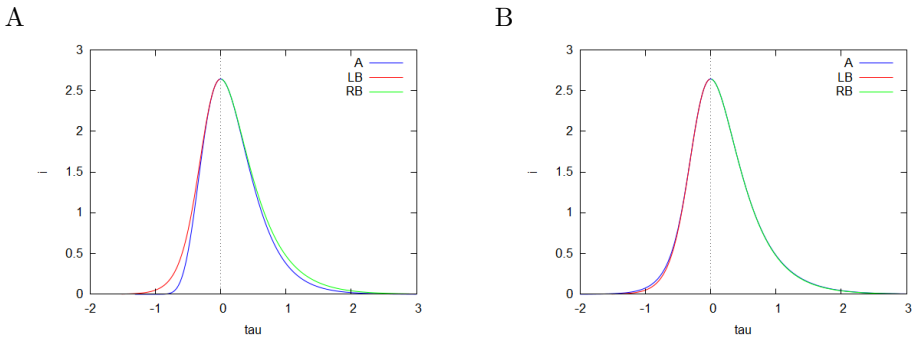

FIGURE 5. Asymptotic solutions $i(\tau)$
Asymptotic solutions compared to parametric plots of $(\tau(i), i)$ parametrized by $i_m = 3.7283$ and $g = 2.0$. Legends: LB denotes the left branch involving $W_-(x)$, RB denotes the right branch involving $W_+(x)$, eq. 18. A denotes the asymptotic solution. A – the asymptotic is computed from eq. 24; B – the asymptotic is computed from eq. 27. Plots were produced using the *quad_qags* Maxima numerical integration command.

since $W_+\left(-e^{-1}\right) = -1$ and $W_+(z)$ is monotone increasing in $z \in [-1/e, \infty]$. Therefore,

$$\frac{i}{i_m} \approx \left(\frac{i^{(1)}}{i_m}\right)^q$$

and we can conclude that $i(t)$ dominates $i^{(1)}(t)$ as t grows from 0.

6. 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 was updated daily until 14 Dec 2020 and contains the latest available public data on COVID-19 aggregated per country. The data collection policy is available from `https://www.ecdc.europa.eu/en/covid-19/data-collection`.

7. Data Processing

The data were imported in the SQLite `https://www.sqlite.org` database, filtered by country and transferred to MATLAB for parametric fitting using native routines. Quadratures were estimated by the default MATLAB integration algorithms. Estimated parameter values were stored in the same database. The processing is described in [9], The parametric fitting was conducted using least-squares constrained optimization algorithm.

The least-squares constrained optimization was preformed using the fminsearchbnd routine [18]. The fitting equation is given by

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

where I_t is the observed incidence or case fatality, respectively. For numerical stability reasons the time variable was rescaled by a factor of 10.

8. Case Studies

The asymptotic approach is exemplified with data from ECDC for several European countries in the period Jan 2020 – Dec 2020. The data analysis is limited to 14 Dec due to reporting reasons [9].

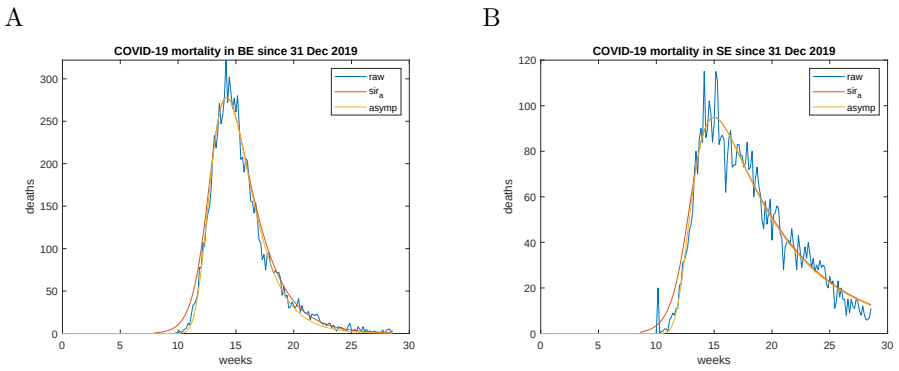

Figure 6. Case fatality parametric fitting for Belgium and Sweden
Parametric fitting of the case fatality data of Belgium (A) and Sweden (B); 'asymp' refers to parametric fit using the asymptotic formula eq. 24, '$\mathrm{sir_a}$' refers to the i-variable computed by numerical inversion using the parameters estimated by eq. 24

The most direct and "brute-force" approach is to directly fit the time series using the numerical inversion scheme. Such an approach has been followed in [9]. This incurs a relatively high computational cost associated with computation of integrals, followed by Newtonian iteration. Alternatively, an the parameters can be fitted using the asymptotic eq. 24 only (see Fig. 6). The numerical inversion approach is illustrated in Fig. 7 and compared with the asymptotic fitting. The raw series for the incidence demonstrated pronounced weekly variation. However, this did not preclude successful parameter estimation.

The optimal fitting scheme involves three iterations

(1) initial estimation from the raw data of the peak, the time to the peak and assuming $g_0 = 1$.
(2) asymptotic estimation of i_m, g, and T
(3) refined estimation of i_m, g, and T, involving numerical inversion of eq. 18.

The data demonstrate very good numerical (Table 1) and graphical agreement of both methods between themselves. Moreover, there is also an excellent agreement with the raw time series.

9. Discussion

The present results can be discussed in two directions.

9.1. **Analytical implications.** The asymptotic analysis of the SIR model is important in terms of the Newtonian approximation scheme since incorrect initial guess will converge the iteration to the wrong branch of the Lambert W function.

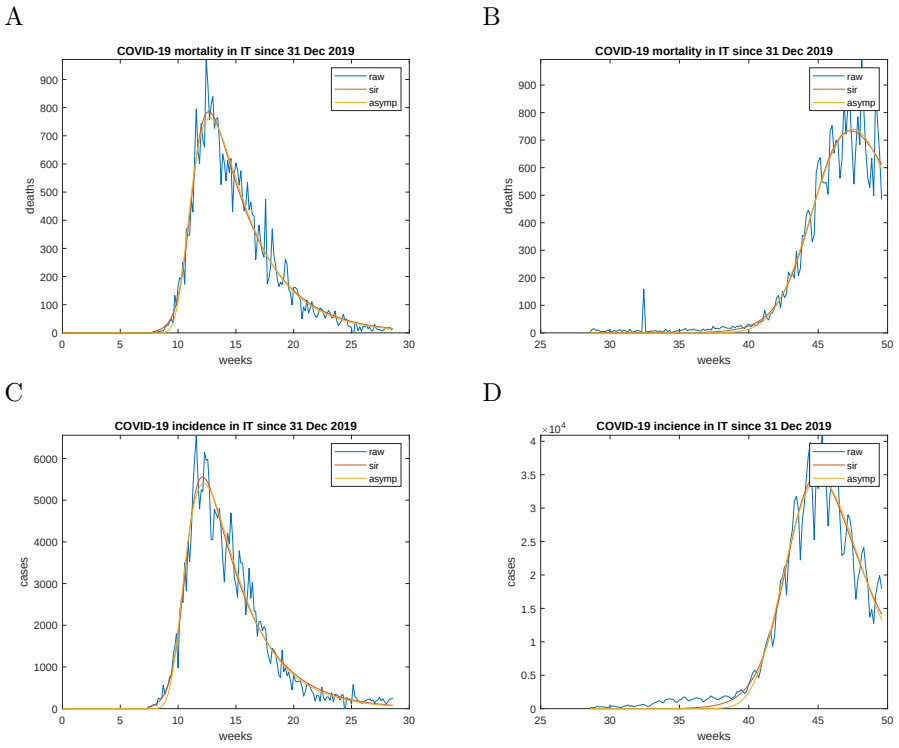

FIGURE 7. Parametric fitting of the first and second waves in Italy A, B – case fatality fitting; C, D – incidence fitting; 'asymp' refers to parametric fit using the asymptotic formula eq. 24, 'sir' refers to fitting the i-variable computed by numerical inversion using the parameters estimated by eq. 24

Country	g	R_0	T[weeks]	i_m	fitting type
Belgium	0.7518	1.33	14.85	303.80	asymp
Belgium	0.7328	1.36	14.32	285.85	sir
Sweden	0.2082	4.80	15.80	89.05	asymp
Sweden	0.1952	5.12	15.68	90.48	sir
Italy	0.3942	2.54	12.74	764.05	asymp
Italy	0.3767	2.65	12.64	785.75	sir

TABLE 1. Case fatality parameters, first wave

T is given in weeks starting from 1st Jan 2020. 'asymp' refers to fitting using asymptotic formula eq. 24 'sir' refers to fitting using numerical inversion of eq. 24.

Therefore, one needs an initial function, which is dominated by the i-variable on the entire real line.

The previous publication using Newtonian iteration [9], did not address the asymptotic analysis of the SIR model. Instead, the original asymptotic of Kermack

and McKendrick was used. However, this did not work for all combinations of parameters. Originally, Kermack and McKendrick proposed hyperbolic cosine, while Barlow and Weinstein[5] used Padé approximation scheme.

An interesting result related to the present asymptotic scheme is the emergence of two fundamental time scales – a longer one determined by g and a shorter one determined by i_m. Such phenomenon is not apparent from the form of the differential equations only, as these explicate only the longer time scale g. Therefore, the appearance of the shorter scale is a truly emergent phenomenon. It will be interesting to investigate such emergence also other epidemiological models derived from SIR.

9.2. **Epidemiological implications.** A key finding of the presented results is that simple models are very useful in studying epidemic outbreaks also in the quantitative sense. This comes in contrast to some shared beliefs that models should be made more complicated to tackle real world epidemics [19]. Presented results demonstrate the utility of the SIR model for estimating the basic reproduction number and the peaks of the infections, respectively case fatalities in the COVID-19 pandemic outbreaks. There is a renewed interest in the applications of the SIR model in view of the COVID-19 pandemics [20, 1, 2, 3, 21, 4, 22]. In contrast to the approach in [22], here the asymptotics are derived from first principles, that is the non-linear algorithm of [17].

The new asymptotic approach simplifies and stabilizes the data fitting procedure compared to [9]. The employed numerical procedures also demonstrate increased robustness to fluctuations.

10. Funding

No specific funding available.

11. Conflicts of interest

The author declares no conflict of interest.

12. Availability of data and material

The COVID datasets were downloaded from ECDC:`https://opendata.ecdc.europa.eu/covid19/casedistribution/csv`. The analysis pertains to the version from 14 Dec 2020.

13. Code availability

Reference implementation in the Computer Algebra System Maxima has been developed and the code is available through the Zenodo repository [16].

Appendix A. The Lambert W function

The Lambert W function can be defined implicitly by the equation

$$W(z)e^{W(z)} = z, \quad z \in \mathbb{C} \tag{28}$$

Furthermore, the Lambert function obeys the differential equation for $x \neq -1$

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

The Lambert W is a multivalued function and in particular it has 2 real-valued branches denoted by W_+ and W_-, respectively. Properties of the W function are surveyed in [11]. Useful identities

$$e^{-W(z)} = \frac{W(z)}{z} \tag{29}$$

$$e^{nW(z)} = \left(\frac{z}{W(z)}\right)^n \tag{30}$$

$$\log W(z) = \log z - W(z) \tag{31}$$

Appendix B. Useful integrals

The following results have been derived in [9].

Proposition 1.

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

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 2.

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

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.

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

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

$$\int_A^B \frac{g-y}{-\frac{y\,(gy\log y - y^2+cy)}{g} + gy\log y - y^2 + cy}dy = g\int_A^B \frac{dy}{y\ (g\log y - y + c)}$$

where

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

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

References

[1] S. Ahmetolan, A. H. Bilge, A. Demirci, A. Peker-Dobie, O. Ergonul, What can we estimate from fatality and infectious case data using the susceptible-infected-removed (SIR) model? a case study of covid-19 pandemic, Frontiers in Medicine 7. doi:10.3389/fmed.2020.556366.

[2] U. Nguemdjo, F. Meno, A. Dongfack, B. Ventelou, Simulating the progression of the COVID-19 disease in Cameroon using SIR models, PLOS One 15 (8) (2020) e0237832. doi:10.1371/journal.pone.0237832.

[3] 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 (2020) 109841. doi:10.1016/j.chaos.2020.109841.

[4] N. R. Record, A. Pershing, A note on the effects of epidemic forecasts on epidemic dynamics, PeerJ 8 (2020) e9649. doi:10.7717/peerj.9649.

[5] N. S. Barlow, S. J. Weinstein, Accurate closed-form solution of the SIR epidemic model, Physica D: Nonlinear Phenomena 408 (2020) 132540. doi:10.1016/j.physd.2020.132540.

[6] N. A. Kudryashov, M. A. Chmykhov, M. Vigdorowitsch, Analytical features of the SIR model and their applications to COVID-19, Applied Mathematical Modelling 90 (2021) 466–473. doi:10.1016/j.apm.2020.08.057.

[7] W. O. Kermack, 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) (1927) 700–721. doi:10.1098/rspa.1927.0118.

[8] T. Harko, F. S. N. Lobo, 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 (2014) 184–194. doi:10.1016/j.amc.2014.03.030.

[9] D. Prodanov, Analytical parameter estimation of the sir epidemic model. applications to the covid-19 pandemic., Entropy (Basel, Switzerland) 23. doi:10.3390/e23010059.

[10] M. Goličnik, On the lambert w function and its utility in biochemical kinetics, Biochemical Engineering Journal 63 (2012) 116–123. doi:10.1016/j.bej.2012.01.010.

[11] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth, On the Lambert W function, Advances in Computational Mathematics 5 (1) (1996) 329–359. doi:10.1007/bf02124750.

[12] M. Kröger, R. Schlickeiser, Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor., Journal of Physics A: Mathematical and Theoretical 53 (50) (2020) 505601. doi:10.1088/1751-8121/abc65d.
URL `https://doi.org/10.1088/1751-8121/abc65d`

[13] M. Martcheva, An Introduction to Mathematical Epidemiology, Springer US, 2015. doi:10.1007/978-1-4899-7612-3.

[14] H. W. Hethcote, The mathematics of infectious diseases, SIAM Review 42 (4) (2000) 599–653. doi:10.1137/s0036144500371907.

[15] R. Piessens, E. Doncker-Kapenga, C. W. Überhuber, D. K. Kahaner, Quadpack, Springer Berlin Heidelberg, 1983. doi:10.1007/978-3-642-61786-7.

[16] D. Prodanov, Numerics for the SIR model, Zenodo (2020). doi:10.5281/ZENODO.4399944.

[17] V. Daftardar-Gejji, H. Jafari, An iterative method for solving nonlinear functional equations, Journal of Mathematical Analysis and Applications 316 (2) (2006) 753–763. doi:10.1016/j.jmaa.2005.05.009.

[18] D. Errico, Fminsearchbnd, `https://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon`, mATLAB Central Exchange (2012).

[19] A. Huppert, G. Katriel, Mathematical modelling and prediction in infectious disease epidemiology, Clinical Microbiology and Infection 19 (11) (2013) 999–1005. doi:10.1111/1469-0691.12308.

[20] D. Fanelli, F. Piazza, Analysis and forecast of COVID-19 spreading in china, italy and france, Chaos, Solitons & Fractals 134 (2020) 109761. doi:10.1016/j.chaos.2020.109761.

[21] I. Cooper, A. Mondal, C. G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, Chaos, Solitons & Fractals 139 (2020) 110057. doi:10.1016/j.chaos.2020.110057.

[22] R. Giubilei, Closed form solution of the SIR model for the COVID-19 outbreak in italy-doi:10.1101/2020.06.06.20124313.