

Asymptotic analysis of the SIR model. Applications to COVID-19 modelling

Dimitar Prodanov (✉ dimitar.prodanov@imec.be)

Interuniversitair Micro-Elektronica Centrum <https://orcid.org/0000-0001-8694-0535>

Research Article

Keywords: SIR model, Lambert W function, asymptotic analysis, incomplete gamma function

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-255095/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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, Bulgarian Academy of Sciences, Sofia, 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 ongoing coronavirus 2019 (COVID-19) pandemics became a global emergency in 2020. Since the pandemic is still ongoing new epidemiological tools, which can accurately predict the time course of different outbreaks, can be of considerable public utility. While sophisticated mathematical models may be desirable for simulating different control strategies, it is difficult to accurately define all necessary inputs. Therefore, relatively simple epidemiological models can turn out also to be of comparable merits. 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 for a number of countries [1, 2, 3, 4, 5, 6].

The SIR model was introduced by Kermack and McKendrick in 1927 but it still remains a cornerstone of mathematical epidemiology [7]. The model can be extended in two directions – either by adding a final state, e.g. "dead" individuals – D ; or by adding one or more intermediate non-observable populations – e. g. "exposed" E individuals. Very recently, it was also demonstrated [9] that COVID-19 can be approximated by a SIRD model. This result is supported by the findings in [10] where the deceased population could also be fitted by the SIR model.

The analytical solution of the SIR model was derived only relatively recently [11] as an equation of state between the phase space variables. As an alternative

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

approach, Barlow and Weinstein derived a series for the S-variable and introduced rational convergents having better regions of convergence [5]. A different approximation scheme was introduced in [12]. Very recently, an inverse parametric solution for the I-variable has been also obtained [6, 10, 13].

The exact analytical solution of the SIR model was obtained by numerical inversion of a non-elementary integral equation for the I-variable by a Newton iteration (NI) scheme [10]. However, the numerical stability of the iteration scheme could not be thoroughly established, since the asymptotic analysis of the model is non-trivial.

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 NI scheme. Moreover, the presented results demonstrate that the approximation scheme also provides asymptotic solutions for the incidence variable, which are 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 [8] or [14]). The SIR model is formulated in terms of 3 populations of individuals [8]. 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 population cannot become infected and the individuals cannot transmit the disease to others. The model comprises a set of three ordinary differential equations ODEs:

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

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

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

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

$$\dot{s} = -si \quad (4)$$

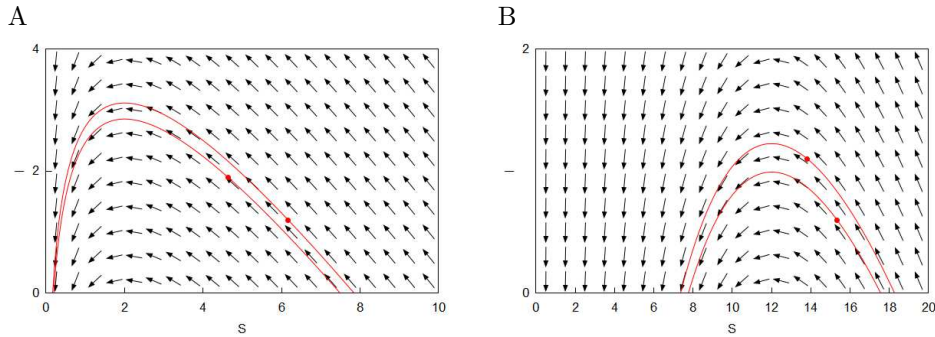
$$\dot{i} = si - gi, \quad g = \frac{\gamma}{\beta} = \frac{1}{R_0} \quad (5)$$

$$\dot{r} = gi, \quad (6)$$

subject to normalization $s + i + r = 1$ and time rescaling as $\tau = \beta t$.

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

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

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 $\alpha = 0.1, \beta = 1.2$.

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:

$$\frac{di}{ds} = -1 + \frac{g}{s} \quad (7)$$

$$\frac{di}{dr} = \frac{s}{g} - 1 \quad (8)$$

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

$$i = -s + g \log s + c \quad (9)$$

where the constant c is a first integral that can be determined, for example, 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 [15]):

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

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

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

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

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

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

$$\dot{s} = -s(-s + g \log s + c) \quad (13)$$

which can be solved implicitly as

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

This is, in fact, the solution obtained in [11].

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,

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

So that

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

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

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

Therefore,

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

4. THE PARAMETRIC SOLUTION

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

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

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

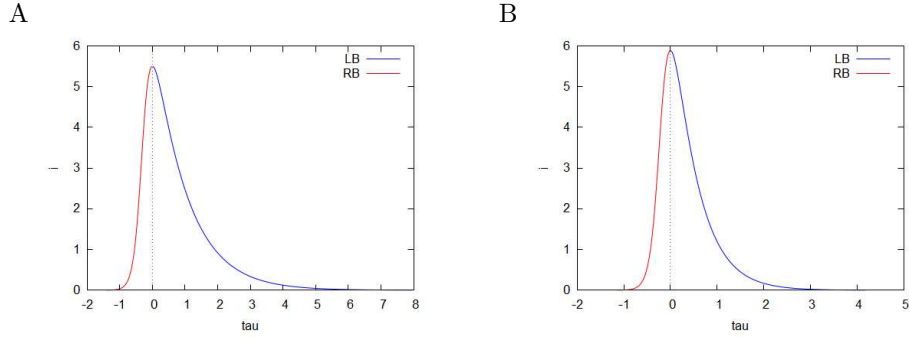


FIGURE 2. Parametric solution $i(\tau)$

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

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

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

4.3. The r -variable. To solve for $\tau(r)$ 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 $c_1 = -gW_-(-e^{-c/g}/g) - c$ by eq. 16. Therefore,

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

for the c -parametrization. The domain of r is

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

Remarkably, all of the involved integrals are non-elementary as demonstrated in [10]. 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 as

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

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:

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

where the negative branch is taken for $\tau < \tau_m$ and the positive branch is taken in the alternative case.

It is easy to re-parametrize 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

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

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 value problem in Fig. 4.

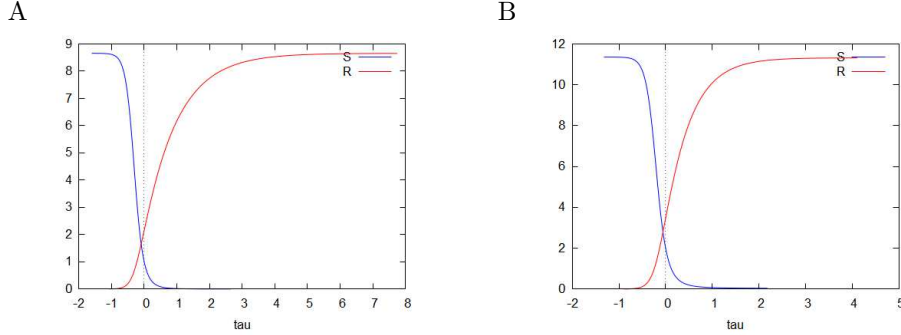
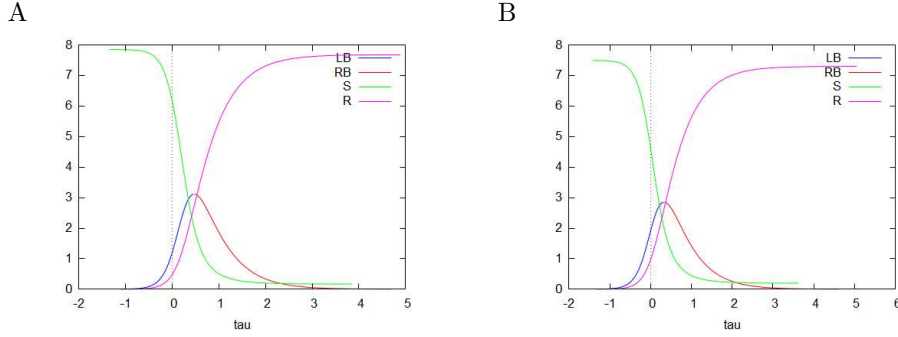


FIGURE 3. Parametric solutions $s(\tau)$, $r(\tau)$

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.

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

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

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.

4.6. 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 [17]. 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 (DJM method) for solving the equivalent integral equations [18]. 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

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

Therefore, formally,

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

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

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

However, this does not guarantee convergence of the iteration. To establish convergence we observe that $s = g$ is a fixed point of eq. 7 since $di/ds = 0$ for this point and, therefore, Banach theorem can be applied. Therefore, we must take

$s_0 = g$ as an initial condition. This corresponds to the peak-value parametrization so $i_0 = i_m$, $i'(0) = 0$ and

$$i(\tau) = i_m \exp\left(g \int_0^\tau e^{-\int_0^z i(y) dy} dz - g\tau\right) \quad (24)$$

can be formulated as a functional integral equation. From this, the 1st order approximation for the i -variable becomes

$$i^{(1)} = i_m e^{\frac{g}{i_m}(1-e^{-i_m t})-gt} \quad (25)$$

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

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

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 limit at $t \rightarrow \infty$ results in the equation

$$r^{(1)} = g \left(W_+ \left(-e^{-\frac{i_m}{g}-1} \right) - W_- \left(-e^{-\frac{i_m}{g}-1} \right) \right) \frac{\Gamma\left(\frac{g}{i_m}, \frac{g}{i_m} e^{-i_m t}\right)}{\Gamma\left(\frac{g}{i_m}\right)} \quad (26)$$

which can also be used 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,

$$s^{(2)} = g \exp\left(\left(\Gamma(a, a) - \Gamma(a, a e^{-i_m \tau})\right) \left(\frac{e}{a}\right)^a\right), \quad a = \frac{g}{i_m} \quad (27)$$

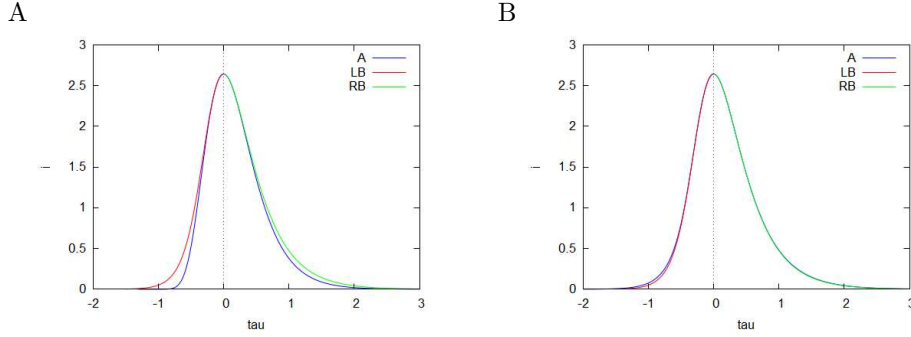
Finally, following the same procedure for the i -variable we obtain

$$i^{(2)} = i_m \exp\left(g \int_0^\tau e^{q(\Gamma(a, a) - \Gamma(a, a e^{-i_m z}))} dz - g\tau\right), \quad q = \left(\frac{e}{a}\right)^a \quad (28)$$

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

Furthermore, let

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

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. 25; B – the asymptotic is computed from eq. 28. Plots were produced using the *quad.qags* Maxima numerical integration command.

Therefore, for $t \rightarrow \infty$

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

since $W_+(-e^{-1}) = -1$ 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. The same argument holds also for $t \rightarrow -\infty$, however in this case

$$q = \lim_{t \rightarrow -\infty} \frac{W_- \left(-e^{\frac{i(t)-i_m}{g}} - 1 \right) + 1}{1 - e^{-i_m t}} = 0$$

since the nominator is bounded while the denominator grows exponentially. The sign of q is again positive since $W_-(z)$ decreases away from -1 for $z \in [-1/e, 0)$ while the sign of the denominator is negative.

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 [10], The parametric fitting was conducted using least-squares constrained optimization algorithm.

The least-squares constrained optimization was performed using the `fminsearchbnd` routine [20]. The fitting equation is given by

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

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 [10].

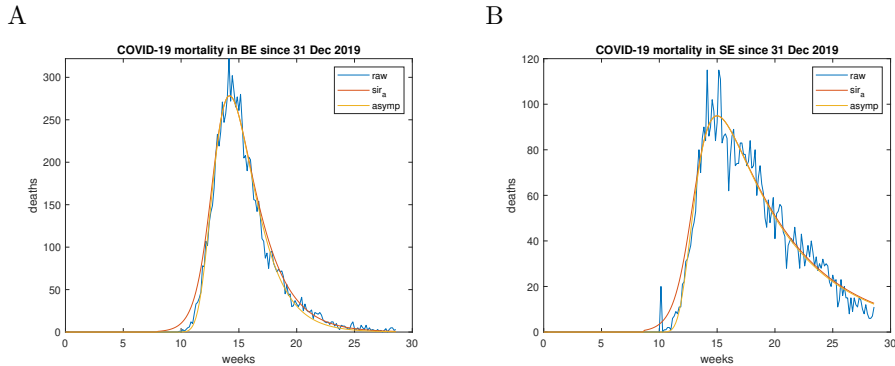


FIGURE 6. Case fatality parametric fitting for Belgium and Sweden
 Parametric fitting of the case fatality data of Belgium (A) and Sweden (B);
 'asypm' refers to parametric fit using the asymptotic formula eq. 25, 'sir_a' refers
 to the i -variable computed by numerical inversion using the parameters estimated
 by eq. 25

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 [10]. This incurs a relatively high computational cost associated with computation of integrals, followed by Newtonian iteration. Alternatively, the parameters can be fitted using the asymptotic eq. 25 only (see Fig. 6). Moreover, the asymptotic parameter estimation demonstrated faithful representation of the exact model (see Fig. 6). The numerical inversion approach is illustrated in Fig. 7 and compared with the asymptotic fitting in the subsequent tables. The raw series for the incidence demonstrated pronounced weekly variation. However, this did not preclude successful parameter estimation.

The optimal fitting scheme involved three steps

- (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.

Case fatality parameters for the first and second waves are presented in Tables 1 and 2. Incidence parameters are presented in Tables 3 and 4. From the data it is apparent that the new fitting procedure did not change parameter estimates for the countries with pronounced outbreaks. The estimates obtained using the method in [10] are marked with asterisks. In contrast, variation in the fitting could be observed for the first wave in Sweden. There the method did not converge well for the incidence and the data are not presented.

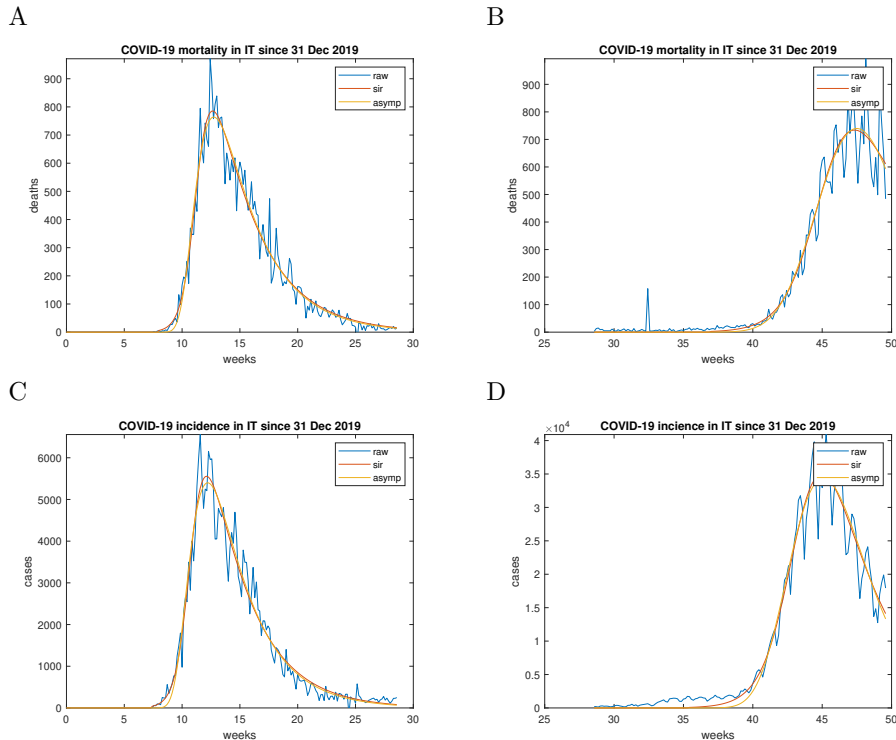


FIGURE 7. Parametric fitting of the first and second waves in Italy
 A, B – case fatality fitting; C, D – incidence fitting; 'asympt' refers to parametric fit using the asymptotic formula eq. 25, 'sir' refers to fitting the i -variable computed by numerical inversion using the parameters estimated by eq. 25. Note the pronounced weekly variation of the reported numbers.

The data demonstrate very good numerical (Tables 1, 2, 3, 4) 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.

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. 25 'sir' refers to fitting using numerical inversion of eq. 25.

Country	g	R_0	T[weeks]	i_m	fitting type
Belgium	0.4913	2.06	44.8	201.88	sir*
Italy	0.3039	3.29	47.61	731.09	sir*
Belgium	0.5780	1.73	56.85	198.51	asymp
Belgium	0.5051	1.98	56.77	202.29	sir
Italy	0.3942	2.54	12.74	764.05	asymp
Italy	0.3479	2.87	59.58	734.07	sir

TABLE 2. Case fatality parameters, second wave

T is given in weeks starting from 1st Jan 2020. 'asymp' refers to fitting using asymptotic formula eq. 25 'sir' refers to fitting using numerical inversion of eq. 25.

Country	g	R_0	T[weeks]	i_m	fitting type
Belgium	0.5500	1.82	14.00	1499.59	sir*
Italy	0.4006	2.50	12.30	5560.30	sir*
Belgium	0.5949	1.68	13.51	1495.37	asymp
Belgium	0.5991	1.67	13.51	1527.71	sir
Italy	0.4190	2.39	12.22	5401.13	asymp
Italy	0.3976	2.51	12.12	5557.86	sir

TABLE 3. Incidence parameters, first wave

* – fitting method as used in [10]. T is given in weeks and refers to the time passed since 1st Jan 2020.

Country	g	R_0	T[weeks]	i_m	fitting type
Belgium	1.0000	1.00	42.43	14970.01	sir*
Italy	0.5548	1.80	45.25	34949.74	sir*
Belgium	1.3772	0.73	54.46	14642.57	asymp
Belgium	2.0921	0.48	54.53	15084.92	sir
Italy	0.7835	1.28	57.32	34412.50	asymp
Italy	0.5715	1.75	57.22	34958.10	sir

TABLE 4. Incidence parameters, second wave

* – fitting method as used in [10]. T is given in weeks and refers to the time passed since 1st Jan 2020.

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

The previous publication using Newtonian iteration [10], 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 [21]. 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 [22, 1, 2, 3, 23, 4, 24]. In contrast to the approach in [24], here the asymptotics are derived from first principles, that is the non-linear algorithm of [18].

The new asymptotic approach simplifies and stabilizes the data fitting procedure compared to [10]. 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

$$W(z)e^{W(z)} = z, \quad z \in \mathbb{C} \quad (29)$$

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 [15]. Useful identities

$$e^{-W(z)} = \frac{W(z)}{z} \quad (30)$$

$$e^{nW(z)} = \left(\frac{z}{W(z)} \right)^n \quad (31)$$

$$\log W(z) = \log z - W(z) \quad (32)$$

APPENDIX B. USEFUL INTEGRALS

The following results have been derived in [10, 13]. Since the source is open access the proofs will not be repeated here.

Proposition 1.

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

Proposition 2.

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

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] M. Martcheva, *An Introduction to Mathematical Epidemiology*, Springer US, 2015. doi:10.1007/978-1-4899-7612-3.
- [9] T. Carletti, D. Fanelli, F. Piazza, COVID-19: The unreasonable effectiveness of simple models, *Chaos, Solitons & Fractals: X* 5 (2020) 100034. doi:10.1016/j.csf.2020.100034.
- [10] D. Prodanov, Analytical parameter estimation of the sir epidemic model. applications to the covid-19 pandemic., *Entropy (Basel, Switzerland)* 23. doi:10.3390/e23010059.
- [11] 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.
- [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] D. Prodanov, Comments on some analytical and numerical aspects of the SIR model, *Applied Mathematical Modelling* 95 (2021) 236–243. doi:10.1016/j.apm.2021.02.004.
- [14] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Review* 42 (4) (2000) 599–653. doi:10.1137/s0036144500371907.
- [15] 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.
- [16] D. Prodanov, Numerics for the SIR model, *Zenodo* (2020). doi:10.5281/ZENODO.4399944.
- [17] R. Piessens, E. Doncker-Kapenga, C. W. Überhuber, D. K. Kahaner, *Quadpack*, Springer Berlin Heidelberg, 1983. doi:10.1007/978-3-642-61786-7.
- [18] 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.
- [19] S. Bhalekar, V. Daftardar-Gejji, Convergence of the new iterative method, *International Journal of Differential Equations* 2011 (2011) 1–10. doi:10.1155/2011/989065.
- [20] D. Errico, Fminsearchbnd, <https://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon>, MATLAB Central Exchange (2012).
- [21] 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.
- [22] 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.
- [23] 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.
- [24] R. Giubilei, Closed form solution of the SIR model for the COVID-19 outbreak in italy- doi:10.1101/2020.06.06.20124313.

Figures

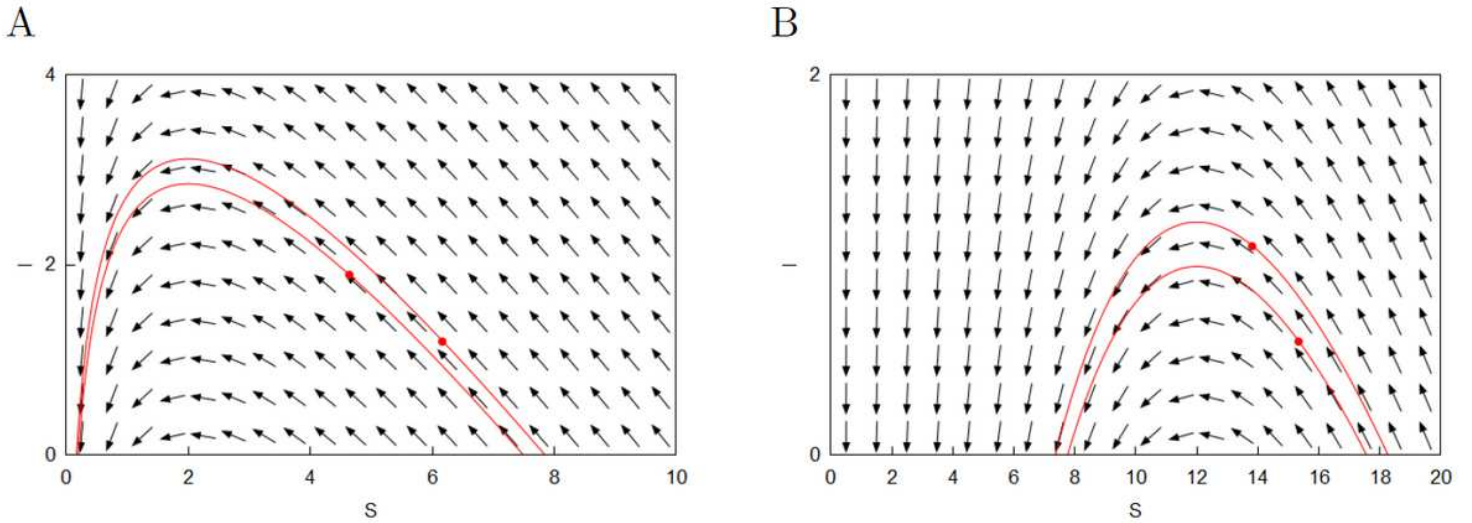


Figure 1

Phase plot of the (l, 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 $\alpha = 0.1, \beta = 1.2$.

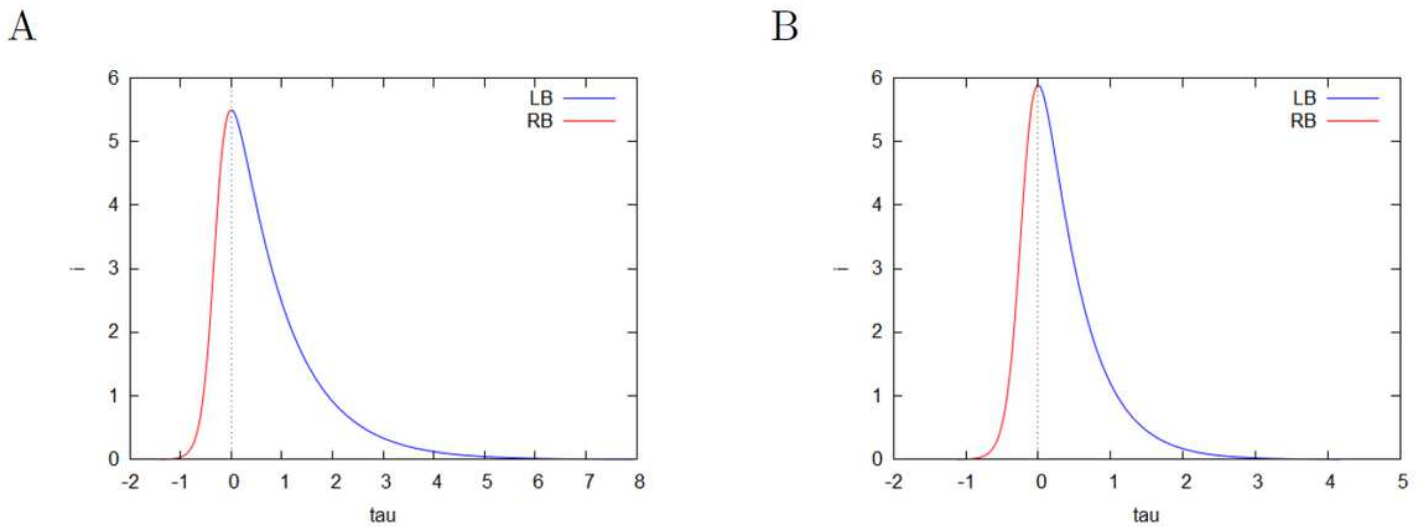


Figure 2

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.

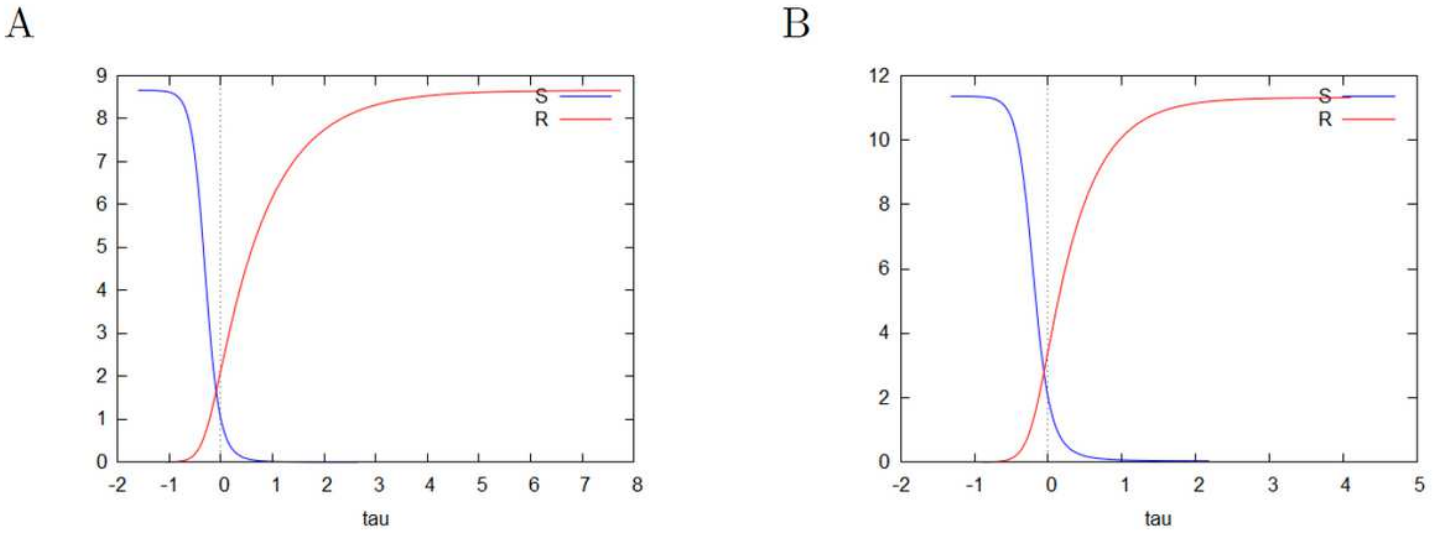


Figure 3

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.

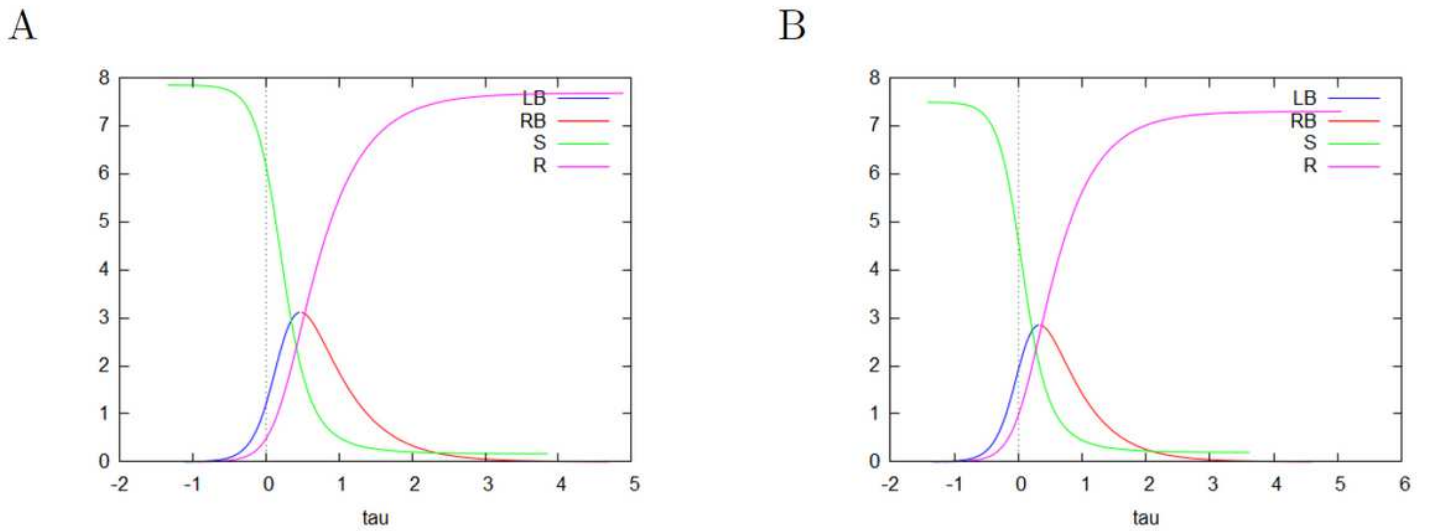


Figure 4

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.

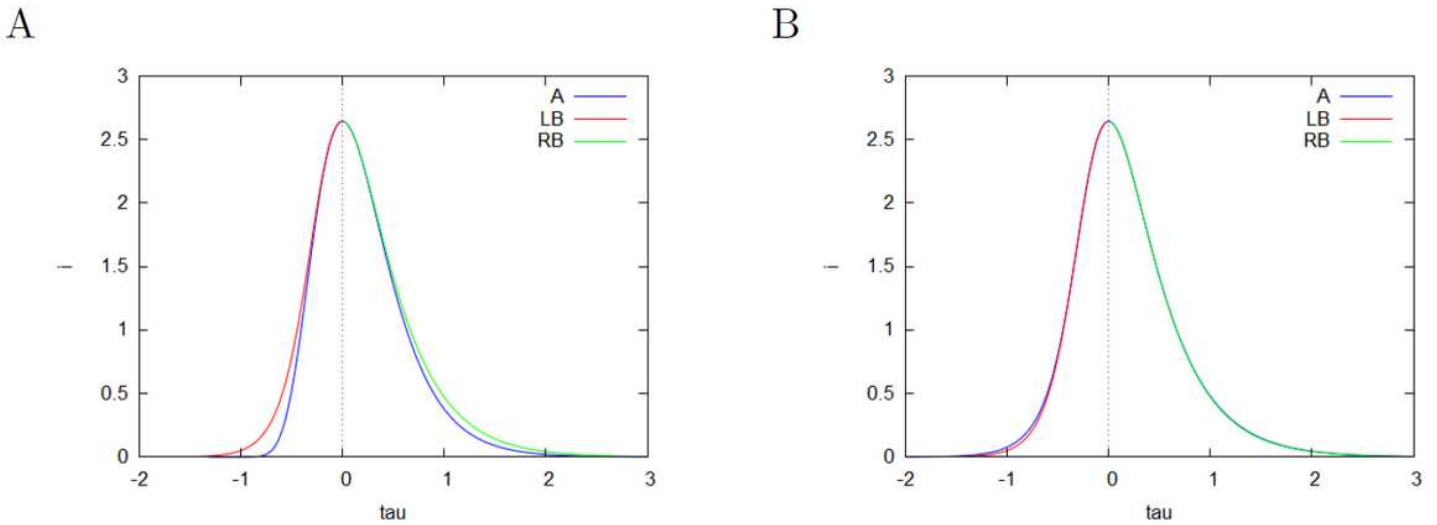


Figure 5

Asymptotic solutions $i(\tau)$ Asymptotic solutions compared to parametric plots of $(\tau(i), i)$ parametrized by $im = 3.7283$ 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. 25; B – the asymptotic is computed from eq. 28. Plots were produced using the quad qags Maxima numerical integration command.

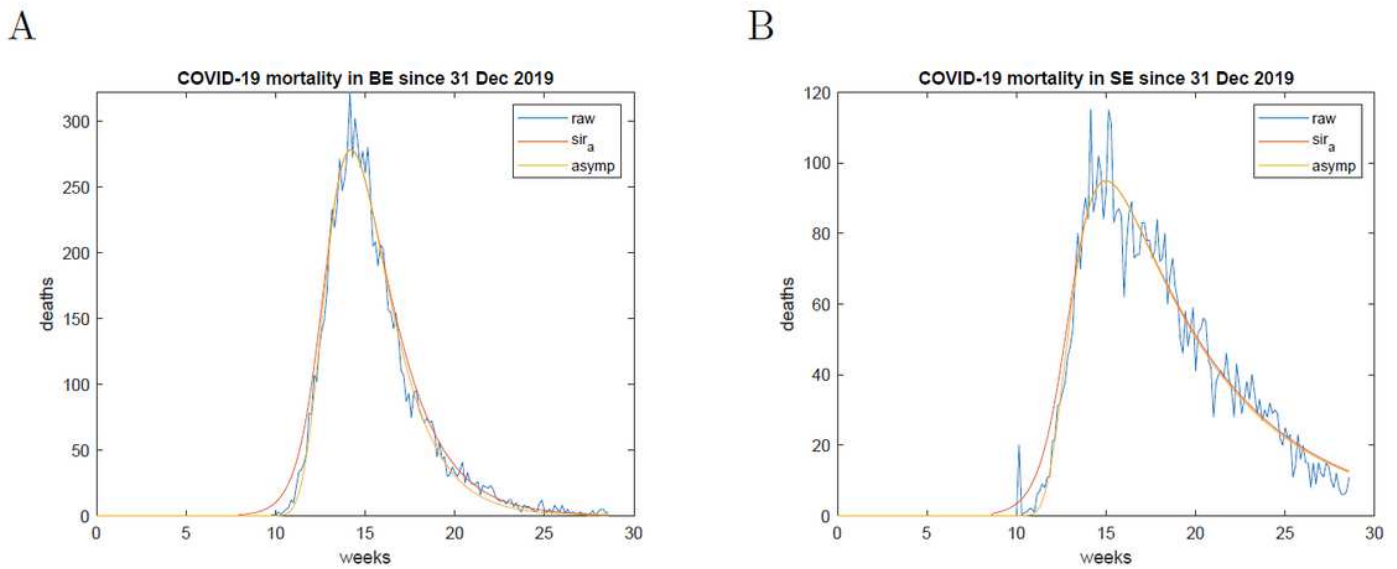


Figure 6

Case fatality parametric fitting for Belgium and Sweden Parametric fitting of the case fatality data of Belgium (A) and Sweden (B); 'asyp' refers to parametric fit using the asymptotic formula eq. 25, 'sira' refers to the i -variable computed by numerical inversion using the parameters estimated by eq. 25

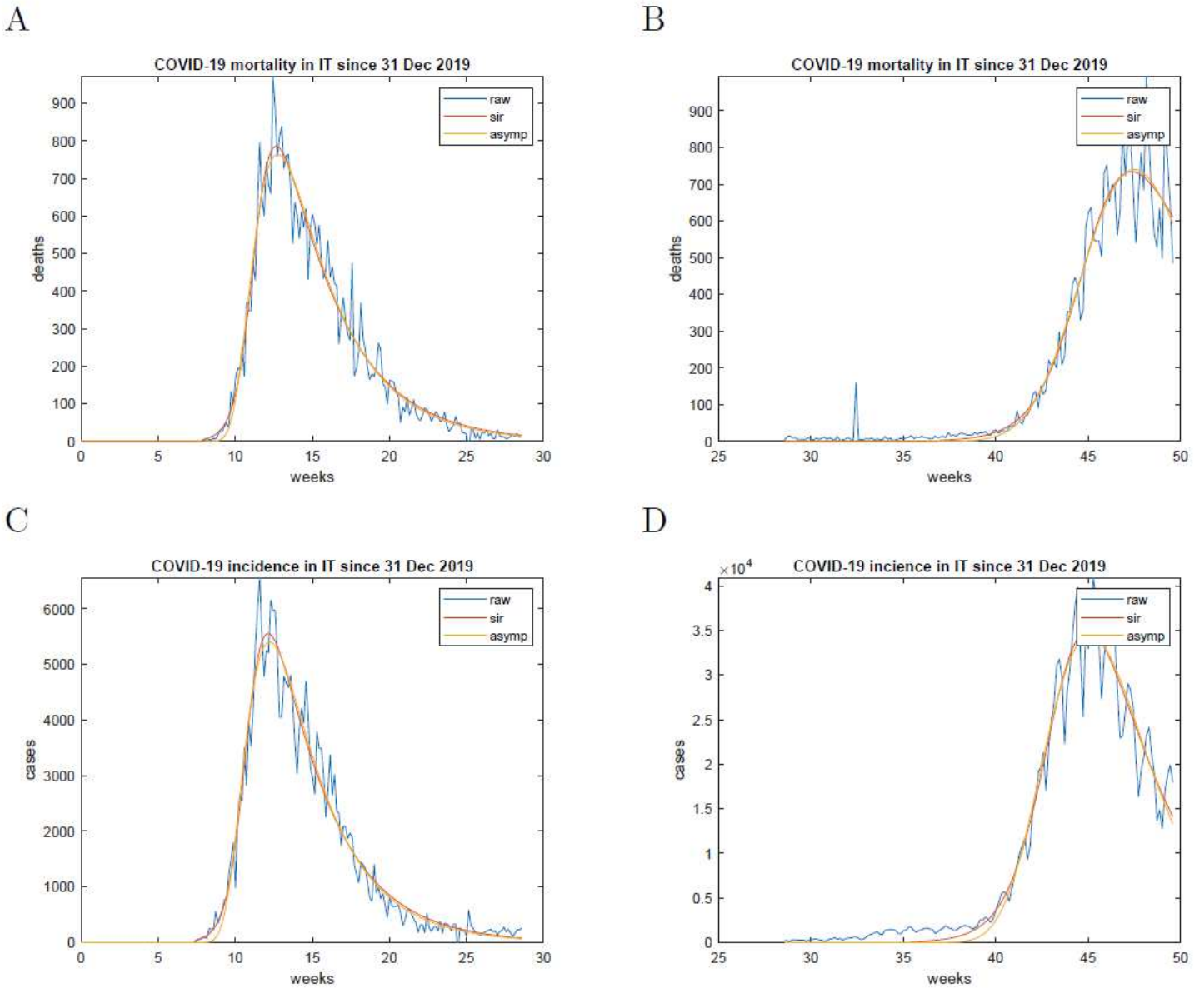


Figure 7

Parametric fitting of the first and second waves in Italy A, B – case fatality fitting; C, D – incidence fitting; 'asyp' refers to parametric fit using the asymptotic formula eq. 25, 'sir' refers to fitting the i -variable computed by numerical inversion using the parameters estimated by eq. 25. Note the pronounced weekly variation of the reported numbers.