Near-exact explicit asymptotic solution of the SIR model well above the epidemic threshold

Gregory Kozyreff
Physics Department, Université libre de Bruxelles (U.L.B.), CP 231, Campus de la Plaine, 1050 Bruxelles, Belgium.

March 24, 2021

Abstract

A simple and explicit expression of the solution of the SIR epidemiological model of Kermack and McKendrick is constructed in the asymptotic limit of large basic reproduction numbers $R_0$. The proposed formula yields good qualitative agreement already when $R_0 \geq 3$ and rapidly becomes quantitatively accurate as larger values of $R_0$ are assumed. The derivation is based on the method of matched asymptotic expansions, which exploits the fact that the exponential growing phase and the eventual recession of the outbreak occur on distinct time scales. From the newly derived solution, an analytical estimate of the time separating the first inflexion point of the epidemic curve from the peak of infections is given.

1 Introduction

The COVID-19 pandemic has impacted all aspects of our daily lives. It has thus triggered an extraordinary world-wide response across all fields of science, from virology and pharmacology to sociology. Within the scientific crowd, applied mathematicians and theoretical physicists also brought their contribution, notably by using their mathematical expertise in epidemiological modelling [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Deterministic compartment models allow one in principle to develop an accurate global view of the contagion dynamics within large populations. Compartments can be used to divide the population into age categories as well as according to the evolution of the illness, for those who have been infected [13, 8, 10]. These multiple levels of description are useful and necessary, and can be made to fit closely to the data. However, they contain a plethora of fitting parameters and can therefore be complex to interpret. Fortunately, it turns out that the simplest of all compartment models, SIR, faithfully reproduces the global dynamics at the level of a country or a large city with COVID-19 [9, 14]. As a result, the SIR model remains a useful
tool at such global level of description. We write it as follows
\[
\frac{dS}{dt} = -\beta SI/N, \\
\frac{dI}{dt} = \beta SI/N - \gamma I, \\
\frac{dR}{dt} = \gamma I,
\]
where \(S\), \(I\), and \(R\) respectively denote the number of susceptible, infected and removed individuals, with constant sum \(N = S + I + R\). The population \(R\) includes both those who have recovered from the illness and those who have died. Above, \(\beta\) is the contact rate, \(\gamma\) is the recovery rate, and their ratio
\[R_0 = \frac{\beta}{\gamma}\]
is the basic reproduction number. Despite its long history [15], there continues to be many efforts to try and solve it analytically [16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. In their seminal paper, Kermack and McKendrick offered an approximate solution in the limit of a “small epidemic”, i.e. when the reproduction number \(R_0\) is just above 1 [15]. So far, this has remained the only explicit approximate analytical solution available, in the sense that it can be justified as an asymptotic limit of the true solution. Well above the epidemic threshold, an analytical expression of the solution can be written in parametric form [26]. Unfortunately, it is only given implicitly, such that time is expressed in terms of one of the dynamical variables through an integral. Hence, this solution is not directly interpretable and its analysis can be quite involved [23]. Alternatives have been proposed in the form of converging series [17, 20], but these can involve expansions with as many as 15 up to 60 terms and, hence, are again impractical for analysis. More down-to-earth is the approach by which the solution is fitted by well-chosen ansatz with a few parameters [27, 24]. However, these ansatz are not derived from the model and their parameters must be fitted to the data, rather than deduced from the model.

1.1 Statement of results

The aim of this paper is to show that
\[
\frac{R(t)}{N} \equiv n_r(t) \sim \begin{cases} 
\frac{1}{R_0 - 1} \ln \left[ ae^{(\beta - \gamma)t} + 1 - a \right], & t < t_*, \\
\frac{1}{R_0 - 1} \ln \left[ 1 + \frac{1-a}{a} e^{-(\beta - \gamma)t} \right] + n_{r,\infty} - e^{-\gamma(t-t_*)}, & t > t_*,
\end{cases}
\]
where \(a\) is the infected fraction of the population at \(t = 0\), while \(n_{r,\infty}\) and \(t_*\) are given by
\[
n_{r,\infty} \sim 1 - e^{-R_0(1-e^{-\tau_0})}, \quad t_* \sim \frac{n_{r,\infty} - 1}{\gamma} - \frac{\ln a}{\beta - \gamma}.
\]
Given \(n_r(t)\), one may infer the infected fraction of the population exactly as
\[
\frac{I(t)}{N} \equiv n_i(t) = 1 - n_r(t) - (1 - a) e^{-R_0 n_r(t)}.
\]
Figure 1: Comparison between the approximate formulas (1) and (3) (full lines) and the numerical solution of the SIR equation (dotted lines) for several values of the reproduction number $R_0$. $n_i(0) = 0.01$. 

Eq. (1) is remarkably simple and compact. The “$\sim$” sign in it indicates that it holds asymptotically in the limit $R_0 \gg 1$. However, a good qualitative agreement with the numerical solution is already found for $R_0 = 3$ and the approximate formula rapidly becomes as good as exact with increasing $R_0$, see Fig. 1. Hence, Eq. (1) complements the classical “small epidemic” formula derived by Kermack and McKendrick [15], which is valid in the limit $0 < R_0 - 1 \ll 1$. When, in 1956, D. G. Kendall presented the implicit analytical solution of the SIR model, he noted:

"It is curious that the Kermack and McKendrick approximation should have been accepted without comment for nearly thirty years; the exact solution is easily obtained and the difference between the two can be of practical significance."

Similarly, the reader will find that the derivation of the present asymptotic formula is quite simple and could have been established much earlier with well-established asymptotic techniques.

As an illustration of the usefulness of the present analytical expression, let us compute the time elapsed between two important moments of the epidemic, namely the time $t_1$ of first inflexion of the curve $n_i(t)$ and the time $t_2$ of the peak of the epidemic. In appendix, we show that the former happens when $n_r$ is equal to $n_1$, solution of

$$R_0 \frac{R_0 e^{-R_0 n_1}}{1 - R_0 e^{-R_0 n_1}} = \frac{1 - R_0 e^{-R_0 n_1}}{1 - n_1 - e^{-R_0 n_1}}, \quad (4)$$
while the latter happens when
\[ n_r = n_2 = \ln \frac{R_0}{R_0 - 1}, \tag{5} \]
in the limit \( a \to 0 \). It turns out that \( t_1 < t^* \) and that \( t_2 > t^* \). Hence, using Eq. (1), we have
\[ \gamma t_1 \sim \frac{\ln(1/a)}{R_0 - 1} + \frac{1}{R_0 - 1} \ln \left[ e^{(R_0 - 1)n_1} - 1 \right]. \tag{6} \]
Meanwhile, for \( t_2 \), we have
\[ n_2 = \frac{1}{R_0 - 1} \ln \left[ 1 + \frac{1 - a}{a} e^{-(\beta - \gamma)t_2} \right] + n_{r,\infty} - e^{-\gamma(t_2-t^*)}. \tag{7} \]
Above, we may assume in first approximation that \( t_2 \) is sufficiently large that the logarithmic term is negligible. Thus, we obtain the more manageable expression
\[ n_2 \sim n_{r,\infty} - e^{-\gamma(t_2-t^*)}, \quad \Rightarrow \quad \gamma t_2 \sim \gamma t^* + \ln (n_{r,\infty} - n_2), \tag{8} \]
and obtain
\[ \gamma (t_2 - t_1) \sim \ln \left( \frac{R_0}{R_0 - 1} \right) - 1 + \ln (n_{r,\infty} - n_2) - \frac{1}{R_0 - 1} \ln \left[ e^{(R_0 - 1)n_1} - 1 \right]. \tag{9} \]
A comparison between this approximate formula and the numerics is given in Fig. 2. Collecting data in real-time allows one to detect when the slope of the epidemic curve has reached a maximum; \( t_2 - t_1 \) is then the time left before the epidemic starts to recede.

### 2 Preliminaries

Before embarking into the asymptotic analysis of the SIR model, we first recall a few well-known facts. Eliminating \( S \) thanks to the relation \( S(t) + I(t) + R(t) = \ldots \)
\( N \), the evolution equations for the fractions \( n_i = I/N \) and \( n_r = R/N \) are

\[
\begin{align*}
\frac{dn_i}{dt} &= (\beta - \gamma) n_i - \beta (n_i + n_r) n_i, \\
\frac{dn_r}{dt} &= \gamma n_i.
\end{align*}
\]

They should be solved with the initial condition

\[
n_i(0) = a \ll 1, \quad n_r(0) = 0.
\]

At the start of the outbreak, \( n_i, n_r \ll 1 \) and the linearised evolution equations immediately yield

\[
n_i \approx a e^{(\beta - \gamma)t}, \quad n_r \approx \frac{a}{R_0 - 1} e^{(\beta - \gamma)t}.
\]

On the other hand, dividing Eq. (10) by Eq. (11), one obtains

\[
\frac{d}{dn_r} (n_i + n_r - 1) = -R_0 (n_i + n_r - 1),
\]

which immediately yields Eq. (3). Thanks to the explicit dependence \( n_i[n_r(t)] \), Eq. (11) eventually becomes a first-order nonlinear ODE:

\[
\frac{dn_r}{dt} = \gamma \left[ 1 - n_r(t) - (1 - a) e^{-R_0 n_r(t)} \right].
\]

As \( t \to \infty \), \( n_r \) tends to the value \( n_{r,\infty} \) that makes the right hand side vanish:

\[
n_{r,\infty} = 1 - (1 - a) e^{-R_0 n_{r,\infty}} \approx 1 - e^{-R_0 n_{r,\infty}}.
\]

In the large-\( R_0 \) limit, the second term in the right hand side above is small, so that a recursive resolution can be set up. An excellent approximation is given by

\[
n_{r,\infty} \approx 1 - e^{-R_0 (1 - e^{-R_0})},
\]

with an absolute error of less than 0.002 for \( R_0 > 3 \). Alternatively, one may express \( n_{r,\infty} \) in terms of the tabulated Lambert function [23].

It is immediate to see that Eq. (15) can formally be integrated as

\[
\gamma t = \int_0^{n_r} \frac{dn}{1 - n - (1 - a) e^{-R_0 n}}.
\]

This expression has been known and studied for a long time [26]. Note that if both \( \gamma \) and \( \beta \) are functions of time, such that their ratio \( R_0 = \beta/\gamma \) remains constant, then Eq. (18) remains valid, with \( \gamma t \) replaced by the effective time \( \tau = \int \gamma(t) dt \) [22]. Recently, a thorough analysis of the integrand of Eq. (18) was carried out, yielding upper and lower bounds for the function \( t(n_r) \) [23]. In addition, series representations of the integral based either on Taylor expansion of the integrand near the origin or based on its poles in the complex-\( n \) plane were given in that work. The first of these approaches effectively amounts to assume that \( n \) is small and thus generalizes Kermack and McKendrick approximate solution. However, the formulas in [23] are rather involved, making the inversion of the relation \( t(n_r) \) laborious.
3 Matched asymptotic expansions

The method consists in deriving separate approximations at the early stage and at the later stage of the outbreak. One the one hand, the outer solution has a characteristic time \(\gamma^{-1}\) and describes the decaying phase of the epidemic. On the other hand, the inner solution applies to the growing phase and evolves on a shorter time scale. In the language of asymptotics, the \(R_0 \gg 1\) limit produces a boundary layer problem [28]. The two approximations are constrained by matching conditions at intermediate stages of the outbreak. Thanks to the fact that they have an overlapping range of validity, one can then construct a composite solution, Eq. (1), that is uniformly valid in time.

3.1 Outer solution

When \(n_r = O(1)\), \(e^{-R_0 n_r}\) is exponentially small and can at first be neglected in Eq. (15). We thus have

\[
\frac{dn_r}{dt} \sim \gamma (1 - n_r), \quad \rightarrow n_r \sim n_r^{(o)} = 1 - e^{-\gamma(t - t_o)},
\]

where the superscript \((o)\) refers to “outer” approximation and \(t_o\) is an arbitrary constant. However, knowing the actual limiting value of \(n_r\) as \(t \to \infty\), a better approximation is

\[
n_r^{(o)} = n_r, \quad \rightarrow e^{-\gamma(t - t_o)}.
\]

Since \(n_r, \infty\) and 1 only differ by an exponentially small quantity, they are asymptotically equivalent as \(R_0 \to \infty\). However, the correction brought about by using \(n_r, \infty\) in Eq. (20) considerably improves the approximation for moderately large \(R_0\) and thus expands the range of usability of the final result, Eq. (1).

Note that \(n_r^{(o)}\) is only defined for \(t > t_o\), since \(n_r\) must be a positive number.

3.2 Inner solution

At earlier stages of the epidemics, \(n_r\) is \(O(R_0^{-1})\). More precisely, Eq. (13) suggests to write

\[
n \sim n_r^{(i)}(t) = \frac{\nu(t)}{R_0 - 1},
\]

where \((i)\) refers to “inner” approximation. Then Eq. (15) becomes

\[
\frac{d\nu}{dt} = (R_0 - 1) \gamma \left[ 1 - e^{-\frac{R_0}{R_0 - 1}} \nu (1 - a) - \frac{1}{R_0 - 1} \nu \right],
\]

\[
\sim (\beta - \gamma) \left[ 1 - e^{-\nu} (1 - a) \right].
\]

Unlike Eq. (15), this last equation can be integrated explicitly:

\[
\frac{e^\nu}{e^\nu - 1 + a} \frac{d\nu}{dt} \sim \beta - \gamma, \quad \rightarrow \ln \left( \frac{e^\nu - 1 + a}{a} \right) \sim (\beta - \gamma) t,
\]

where we used the fact that \(\nu(0) = 0\). Hence,

\[
\nu \sim \ln \left[ a e^{(\beta - \gamma)t} - 1 + a \right],
\]

and

\[
n_r^{(i)} \sim \frac{1}{R_0 - 1} \ln \left[ a e^{(\beta - \gamma)t} - 1 + a \right].
\]
3.3 Matching

So far, we have derived two approximations that efficiently describe the solution at different moments of the epidemic. However, the constant $t_0$ in Eq. (20) is still unknown. To determine it, we must impose that (20) and (26) match in an intermediate region. Loosely speaking, one wants that

$$\lim_{t \to \infty} n_r^{(i)}(t) \sim \lim_{t \to 0} n_r^{(o)}(t).$$

To formalise this statement more precisely, let us define the matching region as

$$t = t_* + \eta \Delta t,$$

where $\eta$ is such that $\gamma \eta \Delta t \ll 1$, while $(\beta - \gamma) \eta \Delta t \gg 1$. This double asymptotic constraint is possible since $(\beta - \gamma) / \gamma = R_0 - 1 \gg 1$ by assumption. On the one hand, we have

$$n_r^{(o)}(t_* + \eta \Delta t) = n_r,\infty - e^{-\gamma(t_* - t_o)} - \gamma \eta \Delta t,$$

$$\sim n_r,\infty - e^{-\gamma(t_* - t_o)} + e^{-\gamma(t_* - t_o)} \gamma \eta \Delta t.$$ (30)

On the other hand, the inner solution yields

$$n_r^{(i)} \sim \frac{1}{R_0 - 1} \ln \left[ a e^{(\beta - \gamma) t} - 1 + a \right],$$

$$= \frac{1}{R_0 - 1} \left( \ln \left[ a e^{(\beta - \gamma) t} \right] + \ln \left[ 1 - \frac{1 - a}{a} e^{-(\beta - \gamma) t} \right] \right),$$

$$= \frac{1}{R_0 - 1} \left( (\beta - \gamma) t + \ln a + \ln \left[ 1 - \frac{1 - a}{a} e^{-(\beta - \gamma) t} \right] \right),$$

$$= \gamma t_* + \ln a + \frac{1}{R_0 - 1} \ln \left[ 1 - \frac{1 - a}{a} e^{-(\beta - \gamma) t} \right],$$

$$\sim \gamma t_* + \gamma \eta \Delta t + \frac{\ln a}{R_0 - 1}.$$ (35)

Equating expressions (30) and (35), we obtain the two conditions

$$n_r,\infty - e^{-\gamma(t_* - t_o)} = \gamma t_* + \frac{\ln a}{R_0 - 1}, \quad \text{and} \quad e^{-\gamma(t_* - t_o)} = 1.$$ (36)

Hence, we have $t_o = t_*$ and

$$\gamma t_* = n_r,\infty - 1 - \frac{\ln a}{R_0 - 1}.$$ (37)

3.4 Composite solution

As indicated above, $n_r^{(o)}$ is only defined for $t > t_o = t_*$. For $t < t_*$, only the inner solution holds:

$$n_r(t < t_*) \sim n_r^{(i)}. $$ (38)

On the other hand, for $t > t_*$, we may construct a uniformly valid approximation as

$$n_r(t > t_*) \sim n_r^{(i)} + n_r^{(o)} - \text{(common part)},$$ (39)
where the common part is the time-dependance that \( n_i^{(o)} \) and \( n_r^{(o)} \) have in common in the matching region, and which should be removed to avoid double counting. One has, from (35),

\[
\text{(common part)} = \gamma t + \frac{1}{R_0} \ln a - \gamma t + \frac{1}{R_0} \ln a.
\]  

(40)

Hence

\[
n_r(t > t_*) \sim \frac{1}{R_0} - 1 \ln \left[ \frac{e^{(\beta-\gamma) t} - a}{a} \right] + n_r^{(\infty)} - e^{-(\gamma t) - \gamma t} - \gamma t - \frac{1}{R_0} - 1 \ln a.
\]  

(41)

Eqs. (38) and (41) yield Eq. (1), which completes the derivation.

4 Conclusion

We have shown that an explicit approximate solution, Eq. (1), can be derived in the limit of large \( R_0 \). The interest of this solution is that it is already effective when the reproduction is only moderately large. Given that SIR is in itself a crude approximation of reality, it makes little sense from an applied perspective to look for an approximation that is exact to within 1%. Taking the time between inflexion and maximum of the epidemic curve as a measure of accuracy, the present theory, Eq. (9) yields an answer with less than 10% error for \( R_0 \) just above 4. With influenza, a reproduction number exceeding the value of 3 appears rare but possible [29], as was the case with the 1918 influenza pandemic [30]. In the case of the COVID-19 pandemic estimates yield \( R_0 \) in the range 1.5 to 6.49 [31] and particularly around 3 in France [32] and between 3.5 and 4 in Korea and in the Hubei province (China) during the first wave [33]. For measles, \( R_0 \) can be on the order of 10 or even larger [34]. Hence, the large-\( R_0 \) limit considered here is relevant, even though, of course, the assumption of a constant ratio \( \beta/\gamma \) does not apply to the COVID-19 pandemic. While the formulas derived in this paper appear accurate enough, it is to be noted that they can be improved. Only leading-order expressions have been used for the inner and outer solutions that apply to the ascending and receding phases of the outbreak, respectively.

The inner solution can be improved by keeping \( O(1/(R_0 - 1)) \) corrections in Eq. (22) and treating them as perturbations. The same can be done with \( O(e^{-R_0 n_r}) \) terms in the equation for the outer solution. Even within the present theory, a correction to \( t_2 \) could be computed by perturbation, thus improving Eq. (9).

Aknowledgement

G.K. is a Research Associate with the Belgian Fonds de la Recherche Scientifique (FNRS).

A Derivation of formulas (4) and (5)

From Eq. (10), \( \frac{dn_i}{dt} \) vanishes when \( n_i = 1 - \frac{1}{R_0} - n_r \). By substitution into (3), one immediately obtains (5).
Next, let us enquire when the second time derivative of $n_i$ vanishes. Writing $n_i = n_i \left[ n_r(t) \right]$, let us denote

$$n'_i = \frac{dn_i}{dn_r}.$$  \hspace{1cm} (42)

Then,

$$\frac{dn_i}{dt} = n'_i \frac{dn_r}{dt}, \hspace{1cm} (43)$$

and $d^2n_i/dt^2 = n''_i \left( \frac{dn_r}{dt} \right)^2 + n'_i \frac{d^2n_r}{dt^2} = n'_i \left( \frac{dn_r}{dt} \right)^2 \left[ \frac{n''_i}{n'_i} + \left( \frac{dn_r}{dt} \right)^{-2} \frac{d^2n_r}{dt^2} \right], \hspace{1cm} (44)$

and one thus seek for the condition

$$\frac{n''_i}{n'_i} = - \left( \frac{dn_r}{dt} \right)^{-2} \frac{d^2n_r}{dt^2}. \hspace{1cm} (45)$$

Dividing (10) by (11), one has

$$n'_i = R_0 - 1 - R_0 \left( n_i + n_r \right). \hspace{1cm} (46)$$

Hence, $n''_i = -R_0 \left( n'_i + 1 \right)$ and

$$\frac{n''_i}{n'_i} = -R_0 \left( \frac{n'_i + 1}{n'_i} \right) = -R_0 \left( \frac{R_0 - R_0 \left( n_i + n_r \right)}{R_0 - 1 - R_0 \left( n_i + n_r \right)} \right) \hspace{1cm} (47)$$

as $a \to 0$. On the other hand, neglecting $a$ compared to 1, (15) yields

$$\frac{d^2n_r}{dt^2} = -\gamma \frac{dn_r}{dt} \left( 1 - R_0 e^{-R_0 n_r} \right). \hspace{1cm} (48)$$

Hence,

$$\frac{dn_r}{dt} \left( 1 - R_0 e^{-R_0 n_r} \right) = \frac{1 - R_0 e^{-R_0 n_r}}{1 - n_r - e^{-R_0 n_r}}. \hspace{1cm} (49)$$

Equating (47) with (49) yields (4).

References


