An Algebraic Solution for the Kermack-McKendrick Model

Alexsandro M. Carvalho
Escola de Gestão e Negócios, Programa de Pós-Graduação em Economia,
Universidade do Vale do Rio dos Sinos, Caixa Postal 275, 93022-000 São Leopoldo RS, Brazil

Sebastián González
Instituto de Física, Universidade Federal do Rio Grande do Sul,
Caixa Postal 15051, 90501-970 Porto Alegre RS, Brazil

We present an algebraic solution for the Susceptible-Infected-Removed (SIR) model originally presented by Kermack-McKendrick in 1927. Starting from the differential equation for the removed subjects presented by them in the original paper, we re-write it in a slightly different form in order to derive a formal solution, unless one integration. Then, using algebraic techniques and some well justified numerical assumptions we obtain an analytic solution for the integral. We compare the numerical solution of the differential equations of the SIR model with the analytic solution here proposed, showing an excellent agreement. Finally, the present scheme allow us to represent analytically two key quantities: time of the infection peak and fraction of immunized to stop the epidemic.

Keywords: SIR model, epidemic models, exact solution

I. INTRODUCTION

The Kermack-McKendrick model [1] or commonly called the SIR model is a cornerstone of the theoretical mathematical models applied to the dynamic of disease spreading, or simply epidemic models. Presented in 1927 in the Proceeding of Royal Society A, soon became evident that it represents an excellent frame for the understanding of the behavior of epidemics.

The basic reproductive number $R_0 = \beta \tau$, the infection curve (that is the asymptotic number of removed as function of $R_0$) are concepts that found in the SIR model the basement that provide the perfect sense, origin, and explanation. It is true, however, that real epidemics are hardly described strictly in terms of the SIR model (or one of the variants like SIS, SIRS, etc). Nevertheless, it is almost impossible to “speak” the language of the mathematical description of epidemics at any level of complexity, without having the SIR model in mind. If a physics metaphor could be applied, we would say that the SIR model is for epidemics dynamics as the harmonic oscillator is for physics. But the metaphor can not be sustained at all levels because different from the given physical example, the SIR equations lack an analytic solution. To obtain the time evolution of the $S$, $I$, and $R$ quantities we have to resort on numerical integration of the finite differences representation of the SIR equations. Fortunately, from the 1980, the increasing power of affordable computers made the finding of numerical solutions a trivial task, avoiding all the pain of obtaining them by hand or with the help of pocket calculators. But the present facilities do not diminish the beauty and intellectual satisfaction of an analytic solution. In this contribution we present a close solution for the classical SIR model as was formulated by Kermack-McKendrick.

Except for the SIS model, the simplest of all epidemic models, whose differential equations, because of the $S + I = 1$ constrain, end up to be a single and solvable one, the Riccati equation, all the others, remarkably the SIR one, which was the originally presented by Kermack-McKendrick, lack a close analytic solution for their differential equations [2].

While analytic tools like stability analysis, asymptotic analysis, or phase diagram analysis can be used to get some insight into the behavior of these models, by close analytic solution we mean the temporal evolution of the number of infectives or susceptibles, and this is what we want to pursue in this contribution. In doing this task we first present a solution which is exact and closed except for one integral. To cope with the integral, regarding the denominator as a Taylor series or polynomial, it is then represented as a partial fraction expansion. From that expansion, after making an approximation which is then numerically well justified, we arrive at the final solution of the SIR model.

II. SIR MODEL AND THE INTEGRAL FORM

The differential equation of the Kermack-McKendrick model are:

\[
\begin{align*}
\frac{dS}{d\tau} &= -R_0SI \\
\frac{dI}{d\tau} &= R_0SI - I \quad (1) \\
\frac{dR}{d\tau} &= I,
\end{align*}
\]

where $S$, $I$, and $R$ represent respectively the fraction of susceptible, infected, and removed subjects inside a pop-
ulation of fixed size. This is the no vital dynamics version of the model where the condition $S + I + R = 1$ holds. $R_0 = \beta/\gamma$ is the basic reproduction number and $\tau = \gamma t$. To obtain the time evolution of the state of the epidemic it is necessary to integrate the equations (1), which cannot be performed in this case. Kermack and McKendrick, from the first and third differential equations (Eq. 1), arrive to the following expression for $S(R)$:

$$S(\tau) = S(0)e^{-R_0 T(\tau)}$$

(2)

in which it is assumed that at the beginning of the infection there is an initial fraction of infected but no removed subjects, i.e. $I(0) = i_0$, $R(0) = 0$, so $S(0) = 1 - i_0$. From that expression they obtain an independent differential equation for the removed subjects:

$$\frac{dR}{d\tau} = I = 1 - R - S(0)e^{-R_0 T}$$

(3)

At this point, in order to go a step further, they use a strong approximation valid for $R_0 \ll 1$ (only valid for initial times or very weak epidemics). Here, we continue without approximations by now, making a change of variables to $w = R_0(1 - R)$, in terms of which, the last equation transforms into:

$$\frac{dw}{d\tau} = R_0 S(0)e^{w - R_0} - w.$$  

(4)

Thus, we can express the three quantities $(S, I, R)$ in terms of $w$, i.e.,

$$S(\tau) = S(0)e^{w - R_0}$$

$$I(\tau) = \frac{w}{R_0} - S(0)e^{w - R_0}$$

$$R(\tau) = 1 - \frac{w}{R_0}$$

(5)

Therefore, the problem of solving the original $S, I, R$ differential equations of the Kermack Mc-Kendrick model, Eq. (1), was translated into solving the equation for $w$ whose formal solution can be expressed in term of an integral:

$$\int \frac{dw}{R_0 S(0)e^{w - R_0} - w} = \tau + k,$$

(6)

where $k$ is a constant of integration. We will see in the next section how we can compute this integral.

**III. ALGEBRAIC SOLUTION**

In the previous section we have translated the SIR problem of solving the set of differential equations into the quest of a primitive for the integral (6). However, as far as we know, there is no such a primitive; instead we use a functional form that represent the integrand, in the hope that the functional form can be integrated somehow.

We denote the denominator of the integrand of Eq. (6) as $P(w)$,

$$P(w) = R_0 S(0)e^{w - R_0} - w$$

(7)

which can be considered, if the exponential is represented by its Taylor expansion, as a polynomial in complex domain. Thus, its inverse can be expanded as a partial fraction:

$$\frac{1}{P(w)} = \sum_j A(z_j) + F(w),$$

(8)

where $z_j$ are the real roots of $P(w)$. The coefficients $A(z_j)$ are given by

$$A(z_j) = \left(\frac{dP(w)}{dw}\right)^{-1}_{w=z_j} = \frac{1}{R_0 S(0)e^{z_j - R_0} - 1}$$

(9)

and $F(w)$ represents the residual part, i.e., the partial
we were looking for:

$$z_1 \approx \frac{R_0S(0)(R_0 - 1)}{R_0 S(0) - 1}$$
$$z_2 = w(\infty) = R_0(1 - R(\infty)) = R_0 S(\infty)$$

both of them in terms of $S(0)$ and $R_0$, and $S(\infty)$ which in turn can be implicitly expressed using Eq. (2) in terms of the first two quantities

$$S(\infty) = S(0)e^{-R_0(1-S(\infty))}$$

Therefore, we can write the Eq. (8) as

$$\frac{1}{P(w)} \approx \frac{A(z_1)}{w - z_1} + \frac{A(z_2)}{w - z_2} + F(w)$$

In order to verify the contribution of $F(w)$, we evaluate numerically the module ratio between $F(w)$ and $1/P(w)$, $|F(w)/P(w)|$. It can be seen in Fig. 1(b) that that ratio is small, then from this point we neglect $F(w)$, arriving to the following expression for $w$:

$$\tau + k = \int \frac{dw}{P(w)}$$
$$\simeq A(z_1) \ln |w - z_1| + A(z_2) \ln |w - z_2|$$

We can give explicit expression of the coefficients $A(z_1)$ and $A(z_2)$ from Eq.(9) and Eq.(11)

$$A(z_1) = \frac{1}{R_0 S(0)e^{-R_0(1-S(0))} - 1} \equiv A_1$$
$$A(z_2) = \frac{1}{R_0 S(0)e^{-R_0(1-S(\infty))} - 1} \equiv A_2$$

And the constant $k$ is obtained form the initial condition $w(0) = R_0$. So finally, we can approximately determine the integral (14) as

$$\tau \simeq A_1 \ln \left| \frac{w - z_1}{R_0 - z_1} \right| + A_2 \ln \left| \frac{w - z_2}{R_0 - z_2} \right|$$

This way, based on Eq. (15) and (16), the SIR dynamics can be algebraically obtained given the initial conditions: $I(0) = i_0$, $S(0) = 1 - i_0$, and the basic reproductive number $R_0$. Note that the function $\omega(\tau)$ has no inverse (which would be the ideal situation), however that does not keep us of getting an explicit representation of the dynamics, i.e to obtain explicit functions for the three quantities $S$, $I$, and $R$. Fig. 2 shows a comparison between the numerical integration of the differential equations (Eq. (1)) and the (numerically computed) proposed algebraic solution (Eq. (4) and (10)).
order to quantify the error of the algebraic solution, we compute the relative differences between removed \( R(t) \) obtained from the numerical solution of the differential equations (Eq. 5) and from the present solution (Eq. 10), averaged up to the asymptotic state, as follows:

\[
\text{Error} = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{R_N(\tau_k) - R_A(\tau_k)}{R_N(\tau_k)} \right| \times 100
\]

where \( N = \text{int}(\tau_F/d\tau) \), \( \text{int}(x) = \max\{m \in \mathbb{Z} | m \leq x\} \), and \( \tau_k = (k-1)d\tau \). In order to evaluate such error we take \( d\tau = 10^{-3} \) (time step of the numerical integration) and \( R(\tau_F) = 0.99R(\infty) \) (asymptotic or final time). The labels \( N \) and \( A \) refer to numerical and algebraic solutions, respectively. In Fig. 3(a) we present the percentage error as defined above versus the basic reproductive number \( R_0 \) and initial fraction of infected subjects \( i_0 \). It can be seen that the error is sensitive to both \( R_0 \) and the initial fraction of infected \( i_0 \). In particular, if \( i_0 < 10^{-3} \), the Error is proportional to \( \log R_0 \) (see Fig. 3(b)). Furthermore, the Error is practically constant in relation to \( i_0 \) (see Fig. 3(c)). Using linear regression method we estimate that \( \text{Error}(R_0, i_0) \approx 9\log R_0 \).

V. CONSEQUENCES OF THE ALGEBRAIC SOLUTION

From the previous results, we can arrive at an analytic expression for the time of the infection peak, \( \tau_p \). For that purpose we plug the expressions \( w(\tau_p) = R_0 - \ln \left| R_0 S(0) \right| \) into the equation Eq. (10), we obtain

\[
\tau_p \simeq \frac{\ln \left[ 1 - \frac{\ln \left| R_0 S(0) \right|}{R_0 (1 - S(\infty))} \right]}{R_0 S(0) S(\infty) - 1} + \frac{\ln \left[ 1 + \frac{\ln \left| R_0 S(0) \right| (R_0 S(0) - 1)}{R_0 (1 - S(0))} \right]}{R_0 S(0) \exp \left( \frac{R_0 (1 - S(0))}{R_0 S(0) - 1} \right) - 1}
\]

In the Fig. 4(a) we present the \( \tau_p \) as function of \( R_0 \). Note that previous equation is according with the numerical solution (numerical integration Eq. 11). This last result, together with the expression for the basic reproductive number, \( R_0 \), the asymptotic fraction of removed subjects \( R(\infty) \), and the maximum fraction of infected people, \( I_{\text{max}} \), i.e,

\[
R_0 = \frac{\beta}{\gamma}
\]

\[
I_{\text{max}} = 1 - \frac{1}{R_0} (1 + \ln \left| R_0 S(0) \right|)
\]

\[
R(\infty) = 1 - S(0) e^{-R_0 R(\infty)}
\]

IV. INSPECTION OF THE ALGEBRAIC SOLUTION

More detailed inspection of that figure as it displayed in Fig. 2(b) indicates that the agreement is not perfect as a consequence of the approximation done in Eq. 5. In

![Error Plot](image)

FIG. 3. (a) Percentage relative error on removed \( R(t) \) as function of \( R_0 \) and \( i_0 \). Two cross-section cuts or views of the previous curve: (b) cut at \( i_0 = 10^{-3} \) and (c) cut at \( R_0 = 5 \).
It is worth noting that the above results allow us to obtain close expression for other important quantities related to extensions of the Kermack-McKendrick formulation, such as the case of the immunization necessary to stop an epidemic. According to Anderson and May [2], the fraction of the population \( f \) that have to be vaccinated in order to stop the spread of the disease in the population, is \( f = 1 - 1/(R_0 S(0)) \). However, in many cases, we can not immunize the population before the start of the infection process. Thus, we can generalize the expression of the fraction of immunized with time dependence, \( f(\tau) = 1 - 1/(R_0 S(\tau)) \). Note that with the assistance of the Eq. (5) and Eq. (16), we have algebraic expression for the fraction of individuals that should be immunized in time \( \tau \) for the disease to be extinct. To do so, just make the substitution \( w(\tau) = R_0 - \ln[R_0 S(0)(1 - f(\tau))] \) in the Eq. (10), results

\[
\tau \approx \ln \left[ 1 - \frac{\ln[R_0 S(0)(1 - f(\tau))]}{R_0 (1 - S(\infty))} \right] \frac{R_0 S(0) S(\infty) - 1}{R_0 S(0)} + \ln \left[ 1 + \frac{\ln[R_0 S(0)(1 - f(\tau))][R_0 S(0) - 1]}{R_0 (1 - S(0))} \right] \frac{R_0 S(0) \exp \left[ \frac{R_0 (1 - S(0))}{R_0 S(0) - 1} \right] - 1}{R_0 S(0)} \] (22)

Fig. 4(b) shows the dynamics of the fraction of immunized. \( \tau \) that have to be vaccinated in order to stop the spread of the disease in the population is \( \tau = 1 - 1/(R_0 S(0)) \). However, in many cases, we can not immunize the population before the start of the infection process. Thus, we can generalize the expression of the fraction of immunized with time dependence, \( f(\tau) = 1 - 1/(R_0 S(\tau)) \). Note that with the assistance of the Eq. (5) and Eq. (16), we have algebraic expression for

\[
\tau \approx \ln \left[ 1 - \frac{\ln[R_0 S(0)(1 - f(\tau))]}{R_0 (1 - S(\infty))} \right] \frac{R_0 S(0) S(\infty) - 1}{R_0 S(0)} + \ln \left[ 1 + \frac{\ln[R_0 S(0)(1 - f(\tau))][R_0 S(0) - 1]}{R_0 (1 - S(0))} \right] \frac{R_0 S(0) \exp \left[ \frac{R_0 (1 - S(0))}{R_0 S(0) - 1} \right] - 1}{R_0 S(0)} \] (22)

We have presented an algebraic solution for an important and long standing problem of the mathematical biology, which is the solution of the differential equations of the SIR model, as first presented by Kermack and McKendrick. In this solution, the dynamics of the fraction susceptible, infected, and removed are given explicitly in terms of a time-dependent expression \( w(\tau) \) (Eq. (5)). Using well justified approximations, we finally arrive to a transcendental expression for \( w(\tau) \) (Eq. (16)), in terms of which all the dynamic quantities can be expressed. We verified that the present general solution is in excellent agreement with numerical solutions of the same equations over the entire dynamic for different threshold \( R_0 \) values. We showed that the difference between these solutions (algebraic and numeric) is proportional to the logarithm of the epidemic threshold \( R_0 \). In this sense, the error in the proposed solution is small and grows slowly as the value of \( R_0 \). Lastly, this study enables us to represent analytic expressions for the time of the infection peak (Eq. (18)) and fraction of immunized (Eq. (22)).

**VI. FINAL REMARKS**

We acknowledge support from the Brazilian agencies CNPq and CAPES and partial support from CNPq project #551974/2011-7.

[1] Kermack, W. and McKendrick, A., Proceeding of Royal Society A 115 (1927) 700.

[2] Anderson, R. M. and May, R. M., Infectious Diseases in Humans (Oxford University Press, Oxford, 1992).