Mathematical Biology

Geometric singular perturbation theory analysis of an epidemic model with spontaneous human behavioral change

Stephen Schecter[1] (ORCID)

Abstract

We consider a model due to Piero Poletti and collaborators that adds spontaneous human behavioral change to the standard SIR epidemic model. In its simplest form, the Poletti model adds one differential equation, motivated by evolutionary game theory, to the SIR model. The new equation describes the evolution of a variable x that represents the fraction of the population following normal behavior. The remaining fraction $1 - x$ uses altered behavior such as staying home, social isolation, mask wearing, etc. Normal behavior offers a higher payoff when the number of infectives is low; altered behavior offers a higher payoff when the number is high. We show that the entry–exit function of geometric singular perturbation theory can be used to analyze the model in the limit in which behavior changes on a much faster time scale than that of the epidemic. In particular, behavior does not change as soon as a different behavior has a higher payoff; current behavior is sticky. The delay until behavior changes is predicted by the entry–exit function.

Keywords Epidemiological modeling · Entry–exit function · Geometric singular perturbation theory · Imitation dynamics · Evolutionary game theory

Mathematics Subject Classification 92D30 · 34E15 · 91A22

1 Introduction

An infectious disease epidemic in a human population, such as measles, influenza, or covid-19, spreads due to a combination of pathogen characteristics and human behavior. Pathogen characteristics determine the circumstances under which an infected

✉ Stephen Schecter
schecter@ncsu.edu

1 Department of Mathematics, North Carolina State University, Box 8205, Raleigh, NC 27695, USA

person can readily infect another. Human behavior determines how frequently those circumstances occur.

Baseline human behavior varies with the society. In a city, crowded conditions in housing, public transportation, schools and workplaces may lead to frequent close human interactions; in a rural area this may be less true. In East Asia mask-wearing in public is fairly common in normal circumstances (Teng 2020); in other parts of the world it is rare.

During an epidemic, human behavior may change due to government policies that close schools and businesses, require people to stay at home, or encourage social distancing and mask-wearing.

Spontaneous changes in human behavior also affect the course of an epidemic. People may react to an epidemic, or to information presented to them, by spontaneously reducing social contacts, staying home to the extent possible, adopting more stringent hygiene or social distancing, or wearing a mask. People may adopt these behaviors independent of government policies; and, to the extent that they feel motivated to adopt such behaviors, they are more likely to comply with government orders and encouragement to do so.

Similarly, when an epidemic wanes, or when people are presented with information that an epidemic is waning or that the infection is less dangerous than originally feared, people may spontaneously return to normal behavior. If restrictive government policies are still in place, compliance may decline.

In the simplest epidemic models, SIR models, the transmissibility of an infectious disease in captured in a single parameter, β, defined as the number of "adequate contacts" per unit time that an infected person has with other people (Hethcote 2009). If these other people are susceptible to the infection (not immune due to previous infection and not currently infected), an adequate contact results in a new infected individual. The basic reproduction number of the infectious disease, R_0, is β times the typical length of time that an infected person remains infective. If $R_0 > 1$, then initially, when the susceptible fraction of the population is close to 1, the number of infected individuals will grow.

Epidemic control measures aim to reduce β by enforcing or encouraging changes in behavior. To determine what measures to institute, governments rely on epidemic models that estimate β under normal circumstances (perhaps using information from similar pathogens or past epidemics) and under various restrictive policies.

Bauch et al. (2013) pointed out that a weakness of all epidemic models in then-current use is that they ignored spontaneous behavioral change. As far as I know, the situation has not changed. For example, the Imperial College covid-19 model (Ferguson et al. 2020), which influenced the United Kingdom and United States government to institute social distancing measures (Booth 2020), was based on a very detailed 2006 influenza epidemic model by the same group (Ferguson et al. 2006). According to the 2006 paper, "We do not assume any spontaneous change in the behaviour of uninfected individuals as the pandemic progresses, but note that behavioural changes that increased social distance together with some school and workplace closure occurred in past pandemics …and might be likely to occur in a future pandemic even if not part of official policy. …Such spontaneous changes in population behaviour might more easily reduce peak daily case incidence."

Epidemiologists appear to be well aware that spontaneous behavioral change should be incorporated in models. There is a fairly extensive literature on ways to do it; review articles include Bauch et al. (2013), Verelst et al. (2016), Wang (2016). There does not appear to be agreement on what modeling approach is best. This probably should not be regarded as a serious problem; a variety of different models are commonly used in epidemiology. A more serious issue is that there has been little work on how to determine the values of the parameters in the models (Verelst et al. 2016). Without approximate values for the parameters, models of spontaneous behavioral change can only yield qualitative predictions.

The goal of this paper is not to deal with the various issues of how best to account for behavioral change in epidemic models. Instead we want to direct attention to a particular approach to behavioral change (Poletti et al. 2009, 2012; Poletti 2010) in its simplest form. This model, which we call the Poletti model, adds one equation, motivated by evolutionary game theory (Nowak and Sigmund 2004; Hofbauer and Sigmund 2003), to the standard SIR model. It is thus a "coupled disease-behavior model" (Wang 2016). Our goal is to show how the entry–exit function (De Maesschalck 2008) of geometric singular perturbation theory (Jones 1995; Kuehn 2015) can be used to analyze this model. Given values for the parameters, the entry–exit function enables one to make precise predictions in the limit where behavioral change occurs on a much faster time scale than the epidemic itself.

To my knowledge, there have been two previous uses of the entry–exit function in epidemiological models, Li et al. (2016) and Jardón-Kojakhmetov et al. (2020).

Figure 1 shows a typical simulation of the Poletti model. There are three variables. Two, S and I, are the familiar susceptible and infective population fractions from the SIR model. The third variable, x, represents the fraction of the population following normal behavior. When $x = 1$, in this simulation, the model reduces to an SIR model with $R_0 = 3$. When $x = 0$, the entire population uses altered behavior; in this simulation, the model reduces to an SIR model with $R_0 = .6$. In the simulation, behavior changes on a time scale 200 times faster than that of the epidemic itself. Thus, if the time scale for the epidemic is days, 1000 time units in the simulation equals 5 days. The simulation shows 20,000 fast time units, or 100 days.

Altered behavior yields a negative payoff due to loss of income, loss of social interactions, and so on. However, altered behavior reduces the chance of getting the infectious disease. In this simulation, normal behavior yields a higher payoff to the individual when $I < .1$. When $I > .1$, altered behavior yields a higher payoff. There is therefore a tendency to adopt altered behavior, which moderates the epidemic, when I passes .1. When I falls below .1, there is a tendency to resume normal behavior. Resuming normal behavior can result in a "second wave" of infections, as seen in the simulation.

In the Poletti model, susceptibles are assumed to change their behavior from normal to altered, or vice-versa, due to imitation of other susceptibles they encounter who are using the opposite behavior and experiencing a higher payoff. The mathematical formulation of this notion is called imitation dynamics and comes from evolutionary game theory (Hofbauer and Sigmund 2003). It was introduced into mathematical epidemiology by Bauch (2005) in a study of vaccination behavior.
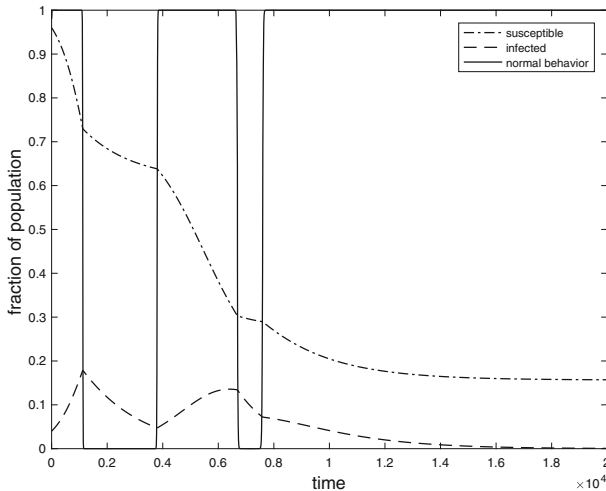
Fig. 1 A simulation of the Poletti model. At the start $(S, I, x) = (.96, .04, .98)$. Since $I < 1$, almost all the population quickly adopts normal behavior. After I rises to about .18 (showing behavior stickiness), the population switches to altered behavior. I falls to about .05 (again showing behavior stickiness); the population returns to normal behavior; and I rises to about .13 (second wave). After two more behavioral switches, the epidemic dies out

Since behavior changes because of encounters with other people, in the simulation, behavior does not change immediately when I passes .1; the current behavior is "sticky." The delay until behavior changes can be calculated in the limit from the entry–exit function.

d'Onofrio et al. (2011), in a discussion of imitation dynamics in a vaccination model, noted that imitation dynamics would inherently involve delay. In their model, however, the delay occurred near an equilibrium at a rate determined by an eigenvalue of the equilibrium. In our situation the delay has a different origin.

The Poletti model, in my view, plays a role similar to the SIR model: it gives the essence of the situation, stripped of complications, and can form the basis for more realistic models. I expect that geometric singular perturbation theory will also prove useful in analyzing more realistic extensions of the model.

In the next few sections of the paper we review the SIR model (Sect. 2), introduce the Poletti model (Sect. 3), and describe and exploit the model's slow–fast structure (Sect. 4). The main result of the paper, Theorem 1, is stated at the end of Sect. 4. We then provide examples (Sect. 5) and proofs (Sect. 6), and conclude with a brief discussion (Sect. 7).

2 The SIR model

In this section we review some standard results about the SIR model from Hethcote (2000) that will be used in the remainder of the paper. The SIR model for an epidemic is

$$\dot{S} = -\beta SI, \tag{1}$$

$$\dot{I} = \beta SI - \gamma I, \tag{2}$$

$$\dot{R} = \gamma I, \tag{3}$$

with $\dot{} = \frac{d}{dt}$. The variables S, I, and R are population fractions; they sum to 1. (Since $\dot{S} + \dot{I} + \dot{R} = 0$, the sum $S + I + R$ is constant.) S is the fraction of the population that is susceptible to acquiring the infectious disease; I is the fraction that is currently infected; R is the fraction that has recovered. (For simplicity we assume that the infection is serious but never fatal.) Thus the equation for R can be ignored; R can be recovered from $R = 1 - S - I$. The system reduces to

$$\dot{S} = -\beta SI, \tag{4}$$

$$\dot{I} = \beta SI - \gamma I \tag{5}$$

on the triangle $T = \{(S, I) : S \geq 0, \ I \geq 0, \ S + I \leq 1\}$.

Let $\hat{T} = \{(S, I) \in T : S > 0 \text{ and } I > 0\}$. In $\hat{T}$ the orbits of (4)–(5) satisfy the differential equation $\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}$, so they are curves

$$I + S - \frac{\gamma}{\beta} \ln S = C. \tag{6}$$

The parameter β was discussed in the introduction. The average length of time an individual is infected is $\frac{1}{\gamma}$. Thus the basic reproduction number of the infectious disease R_0 mentioned in the introduction is $\frac{\beta}{\gamma}$.

Phase portraits on T in the cases $0 < R_0 < 1$ and $R_0 > 1$ are shown in Fig. 2. In both cases the system has the line segment of equilibria $I = 0, 0 \leq S \leq 1$. Each solution approaches one of the equilibria $(S_f, 0)$. In other words, when the epidemic ends, no one is infected, and $R = 1 - S_f$ is the fraction of the population that contracted the infection in the course of the epidemic.

In $\hat{T}$, $\dot{S} < 0$, so the number of susceptibles steadily falls. If $0 < R_0 < 1$, $\dot{I} < 0$ in $\hat{T}$ as well, so the number of infectives also steadily falls. If $R_0 > 1$, $\dot{I} < 0$ (resp. $\dot{I} > 0$) for $0 < S < \frac{\gamma}{\beta}$ (resp. $\frac{\gamma}{\beta} < S < 1$). Thus if a solution starts with $\frac{\gamma}{\beta} < S < 1$, then I increases until S has fallen to $\frac{\beta}{\gamma}$; after that I decreases.

3 The Poletti model

In the Poletti model, susceptible individuals have two available behaviors, normal, for which $\beta = \beta_n$ with basic reproduction number $R_0 = \frac{\beta_n}{\gamma} > 1$, and altered, for which $\beta = \beta_a$ with basic reproduction number $R_0 = \frac{\beta_a}{\gamma} < 1$. Altered behavior may include staying home to the extent possible, practicing social distancing, mask wearing, etc.

Each behavior has a payoff to a susceptible who adopts it. The payoffs are

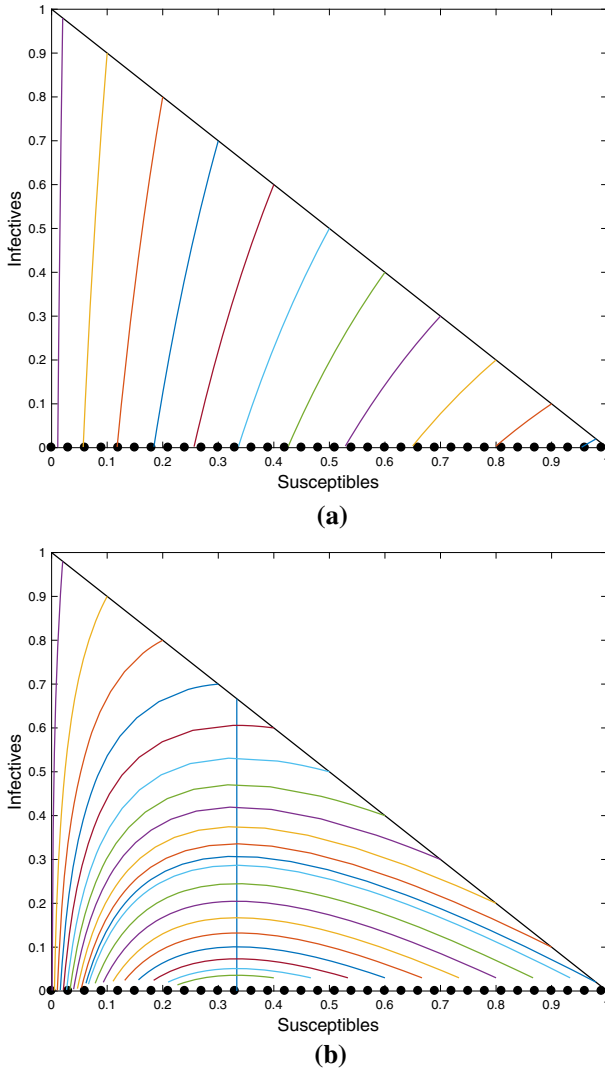$$p_n = -m_n I \text{ and } p_a = -k - m_a I$$

Fig. 2 Phase portraits of SIR models. **a** $\beta = 1/10$, $\gamma = 1/6$, so $R_0 = 6/10$. **b** $\beta = 1/2$, $\gamma = 1/6$, so $R_0 = 3$. In case (b), the vertical line $S = \gamma/\beta = 1/3$ at which solutions attain their maximum value of I is also shown. All solutions move to the left as time increases

with m_n, m_a and k positive and $m_n > m_a$. The negative payoff $-m_n I$ is due to the possibility that a susceptible with normal behavior will contract the infectious disease; it is proportional to I, the fraction of infectives in the population. The negative payoff $-m_a I$ is due to the possibility that a susceptible with altered behavior will contract the infection; it is also proportional to I, but the proportionality constant is less negative. In addition, altered behavior has a negative payoff $-k$ independent of I that represents loss of income, loss of valued social interactions, etc. The payoff from altered behavior is higher if and only if $I > \frac{k}{m_n - m_a}$, i.e., if and only if the fraction of infectives in the population is sufficiently high. We assume

$$\frac{k}{m_n - m_a} < 1.$$

This assumption allows altered behavior to sometimes have a higher payoff.

We recall from the introduction that susceptibles are assumed to change their behavior from normal to altered, or vice-versa, due to imitation of other susceptibles they encounter who are using the opposite behavior and experiencing a higher payoff. Let x denote the fraction of the susceptibles following normal behavior, so that $1 - x$ is the fraction using altered behavior. We continue to let S and I denote the susceptible and infected fractions of the population. Then the complete model is

$$\dot{S} = -\big(\beta_n x + \beta_a(1 - x)\big)SI, \tag{7}$$
$$\dot{I} = \big(\beta_n x + \beta_a(1 - x)\big)SI - \gamma I, \tag{8}$$
$$\dot{x} = x(1 - x)(\beta_a - \beta_n)I + \rho x(1 - x)\big(k - (m_n - m_a)I\big), \tag{9}$$

with $\dot{} = \frac{d}{dt}$. The state space is the prism

$$P = \{(S, I, x) : S \geq 0,\ I \geq 0,\ S + I \leq 1,\ 0 \leq x \leq 1\}.$$

There is also an equation for the recovered fraction of the population R, $\dot{R} = \gamma I$; we ignore it since R can be recovered from $S = 1 - R - I$.

For the derivation of the model, see (Poletti et al. 2009; Poletti 2010). It can be intuitively understood as follows.

The equations for $\dot{S}$ and $\dot{I}$ come from assuming that both susceptibles with normal behavior and susceptibles with altered behavior satisfy SIR models.

The first summand in the equation for $\dot{x}$ is negative; it expresses the fact that susceptibles with normal behavior acquire the infectious disease more easily than susceptibles with altered behavior, and hence more readily leave the susceptible group. Thus the fraction of susceptibles following normal behavior tends to decrease.

The second summand represents the the rate of change of x due to imitiation dynamics (Hofbauer and Sigmund 2003). The rate at which susceptibles using different behaviors encounter each other is proportional to $x(1 - x)$. The difference in payoffs of the two behaviors, given the current level of I, is

$$p_n - p_a = k - (m_n - m_a)I.$$

When this number is positive, normal behavior yields a larger payoff, so x increases at a rate proportional to the difference between the payoffs; when this number is negative, x decreases in the same manner.

The rate constant ρ is the product of two constants: the proportionality constant that determines how the rate at which suceptibles encounter each other depends on the product $x(1 - x)$, and the influence on behavior per encounter.

4 Slow–fast structure

In the remainder of the paper we will assume that the rate constant ρ in (9) is large, indicating that behavior can change on a much faster time scale than that of the epidemic itself. We therefore write $\rho = \frac{1}{\epsilon}$ with $\epsilon > 0$ small.

4.1 Slow–fast structure

We consider system (7)–(9) with $\rho = \frac{1}{\epsilon}$ and $\epsilon > 0$ small, and we recall that $\dot{} = \frac{d}{dt}$. This system is a slow–fast system (Jones 1995; Kuehn 2015) with two slow variables, S and I, and one fast variable, x; t is the slow time.

Such systems are more commonly written with the equation for the fast variable multiplied on both sides by ϵ:

$$\dot{S} = -\big(\beta_n x + \beta_a(1-x)\big)SI, \tag{10}$$

$$\dot{I} = \big(\beta_n x + \beta_a(1-x)\big)SI - \gamma I, \tag{11}$$

$$\epsilon \dot{x} = \epsilon x(1-x)(\beta_a - \beta_n)I + x(1-x)\big(k - (m_n - m_a)I\big). \tag{12}$$

The fast time τ satisfies $t = \epsilon \tau$. With $' = \frac{d}{d\tau}$, system (10)–(12) becomes

$$S' = -\epsilon\big(\beta_n x + \beta_a(1-x)\big)SI, \tag{13}$$

$$I' = \epsilon\big(\beta_n x + \beta_a(1-x)\big)SI - \epsilon\gamma I, \tag{14}$$

$$x' = \epsilon x(1-x)(\beta_a - \beta_n)I + x(1-x)\big(k - (m_n - m_a)I\big). \tag{15}$$

The slow system (10)–(12) and the fast system (13)–(15) have the same phase portraits for $\epsilon > 0$, but they have different limits at $\epsilon = 0$. For $\epsilon = 0$, the slow system (10)–(12) becomes the slow limit system

$$\dot{S} = -\big(\beta_n x + \beta_a(1-x)\big)SI, \tag{16}$$

$$\dot{I} = \big(\beta_n x + \beta_a(1-x)\big)SI - \gamma I, \tag{17}$$

$$0 = x(1-x)\big(k - (m_n - m_a)I\big), \tag{18}$$

and the fast system (13)–(15) becomes the fast limit system

$$S' = 0, \tag{19}$$

$$I' = 0, \tag{20}$$

$$x' = x(1-x)\big(k - (m_n - m_a)I\big). \tag{21}$$

A *singular orbit* is a sequence of orbits of the slow limit system (in our case (16)–(18)) and the fast limit system (in our case (19)–(21)), with each orbit after the first starting where the previous orbit ends. In many situations, orbits for small $\epsilon > 0$ are close to singular orbits.

4.2 Fast limit system

For the fast limit system (19)–(21), each line segment $(S, I) = (S_0, I_0)$ is invariant, and the triangles $x = 0$ and $x = 1$ consist of equilibria. (The plane $I = \frac{k}{m_n - m_a}$ also consists of equilibria, but we will not make direct use of them.) On line segments $(S, I) = (S_0, I_0)$ with $0 \leq I_0 < \frac{k}{m_n - m_a}$, $\dot{x} > 0$, so the solution $x(t)$ of (21) satisfies $\lim_{t \to -\infty} x(t) = 0$ and $\lim_{t \to \infty} x(t) = 1$. On line segments $(S, I) = (S_0, I_0)$ with $\frac{k}{m_n - m_a} < I_0 \leq 1$, $\dot{x} < 0$, so the solution $x(t)$ of (21) satisfies $\lim_{t \to -\infty} x(t) = 1$ and $\lim_{t \to \infty} x(t) = 0$. The fast dynamics just reflect the fact that normal behavior gives a higher payoff if $I < \frac{k}{m_n - m_a}$, and altered behavior gives a higher payoff if $I > \frac{k}{m_n - m_a}$.

If one linearizes the fast limit system (19)–(21) at an equilibrium, there must be at least two zero eigenvalues, since the first two rows of the matrix of the linearization are identically zero. The equilibrium is called *normally hyperbolic* if the other eigenvalue is nonzero. It is called *normally attracting* (resp. *normally repelling*) if the other eigenvalue is negative (resp. positive).

Thus equilibria of (19)–(21) are normally attracting (resp. normally repelling) if $\frac{\partial \dot{x}}{\partial x}$ is negative (resp. positive) at the equilibrium. One can check that equilibria with $x = 0$ are normally repelling for $I < \frac{k}{m_n - m_a}$ and normally attracting for $I > \frac{k}{m_n - m_a}$. Equilibria with $x = 1$ are the reverse.

4.3 Slow limit system

The slow limit system (16)–(18) makes sense on the triangles $x = 0$ and $x = 1$.

On the triangle $x = 0$, the slow limit system reduces to

$$\dot{S} = -\beta_a SI, \tag{22}$$

$$\dot{I} = \beta_a SI - \gamma I. \tag{23}$$

This is just an SIR model with $\beta = \beta_a$ and basic transmission number $R_0 < 1$.

Similarly, on the triangle $x = 1$, the slow limit system reduces to

$$\dot{S} = -\beta_n SI, \tag{24}$$

$$\dot{I} = \beta_n SI - \gamma I. \tag{25}$$

This is just an SIR model with $\beta = \beta_n$ and basic transmission number $R_0 > 1$.

For $\epsilon > 0$, the triangles $x = 0$ and $x = 1$ remain invariant. The slow system (10)–(12), restricted to $x = 0$, remains (22)–(23). Restricted to $x = 1$ it remains (24)–(25). Thus the line segments $\{(S, I, x) : 0 \leq S \leq 1, \ I = 0, x = 0\}$ and $\{(S, I, x) : 0 \leq S \leq 1, \ I = 0, x = 1\}$ remain equilibria.

We will use the following notation where convenient:

- $\phi^\epsilon\big((S_0, I_0, x_0), t\big) =$ solution of (13)–(15) with $\phi^\epsilon\big((S_0, I_0, x_0), 0\big) = (S_0, I_0, x_0)$.
- $\psi_0\big((S_0, I_0), t\big) =$ solution of (22)–(23) with $\psi_0\big((S_0, I_0), 0\big) = (S_0, I_0)$.
- $\psi_1\big((S_0, I_0), t\big) =$ solution of (24)–(25) with $\psi_1\big((S_0, I_0), 0\big) = (S_0, I_0)$.

4.4 Entry–exit function for the triangle $x = 0$

In the triangle $x = 0$, let $(S_0, I_0) \in \hat{T}$ with $I_0 > \frac{k}{m_n - m_a}$, so (S_0, I_0) lies in the attracting portion of the triangle. Let $(S(t), I(t)) = \psi_0\big((S_0, I_0), t\big)$, let $t_1 > 0$, and let $(S_1, I_1) = \big(S(t_1), I(t_1)\big)$. The solution $\big(S(t), I(t)\big)$ traces out a curve Γ, which from (6) has the equation

$$I + S - \frac{\gamma}{\beta_a} \ln S = v_0, \quad v_0 = I_0 + S_0 - \frac{\gamma}{\beta_a} \ln S_0 = I_1 + S_1 - \frac{\gamma}{\beta_a} \ln S_1. \quad (26)$$

We define the entry–exit integral

$$\begin{aligned}
\mathscr{I}_0\big((S_0, I_0), (S_1, I_1)\big) &= \int_0^{t_1} k - (m_n - m_a) I(t) \, dt \\
&= \int_{S_0}^{S_1} \frac{k - (m_n - m_a)(v_0 - S + \frac{\gamma}{\beta_a} \ln S)}{\beta_a S(v_0 - S + \frac{\gamma}{\beta_a} \ln S)} \, dS.
\end{aligned} \quad (27)$$

The second integral follows from the first by making the substitutions $S = S(t)$, $dS = -\beta_a S(t) I(t) \, dt$, and $I = v_0 - S + \frac{\gamma}{\beta_a} \ln S$, which follows from (26). It cannot be evaluated analytically, but is readily evaluated numerically.

The integrand of the first integral is negative when $I > \frac{k}{m_n - m_a}$ and positive when $I < \frac{k}{m_n - m_a}$. The integral represents accumulated attraction to (resp. repulsion from) the plane $x = 0$ where the integrand is negative (resp. positive).

Proposition 1 *For each point (S_0, I_0) in $\hat{T}$ with $I_0 > \frac{k}{m_n - m_a}$, there is exactly one $t_1 > 0$ such that $(S_1, I_1) = (S(t_1), I(t_1))$ satisfies $\mathscr{I}_0\big((S_0, I_0), (S_1, I_1)\big) = 0$.*
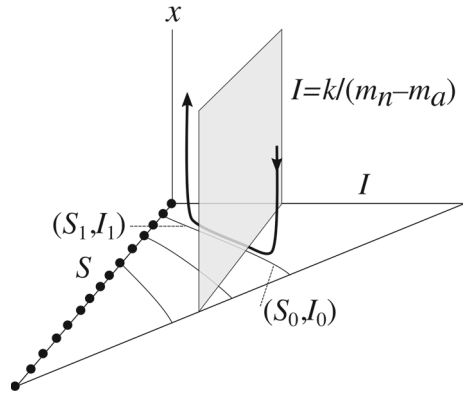
Of course, (S_1, I_1) lies in the region $I < \frac{k}{m_n - m_a}$. Intuitively, at (S_1, I_1) the accumulated repulsion from the plane $x = 0$ balances the accumulated attraction to the plane. We shall see that for small $\epsilon > 0$, a solution of (13)–(15) that enters a neighborhood of the plane $x = 0$ near (S_0, I_0) will track a solution of (22)–(23) near $(S(t), I(t))$ until it leaves the neighborhood near (S_1, I_1). See Fig. 3 and Sect. 6.2.

4.5 Entry–exit function for the triangle $x = 1$

In the triangle $x = 1$, let $(S_0, I_0) \in \hat{T}$ with $I_0 < \frac{k}{m_n - m_a}$, so (S_0, I_0) lies in the attracting portion of the triangle. Let $(S(t), I(t)) = \psi_1\big((S_0, I_0), t\big)$, let $t_1 > 0$, and let $(S_1, I_1) = \big(S(t_1), I(t_1)\big)$. The solution $\big(S(t), I(t)\big)$ traces out a curve Γ, which from (6) has the equation

$$I + S - \frac{\gamma}{\beta_n} \ln S = v_0, \quad v_0 = I_0 + S_0 - \frac{\gamma}{\beta_n} \ln S_0 = I_1 + S_1 - \frac{\gamma}{\beta_n} \ln S_1. \quad (28)$$

Fig. 3 A solution of (13)–(15) approaches the triangle $x = 0$ near a point (S_0, I_0), follows the solution of (22)–(23) through (S_0, I_0) until $\mathscr{I}_0\big((S_0, I_0), (S_1, I_1)\big) = 0$, then leaves the triangle

We define the entry–exit integral

$$\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = -\int_0^{t_1} k - (m_n - m_a)I(t) \, dt$$
$$= \int_{S_0}^{S_1} \frac{k - (m_n - m_a)(v_0 - S + \frac{\gamma}{\beta_n}\ln S)}{\beta_n S(v_0 - S + \frac{\gamma}{\beta_n}\ln S)} \, dS. \tag{29}$$

The second integral follows from the first as in the previous subsection.

The integrand of the first integral is negative when $I < \frac{k}{m_n - m_a}$ and positive when $I > \frac{k}{m_n - m_a}$. The integral represents accumulated attraction to (resp. repulsion from) the plane $x = 1$ where the integrand is negative (resp. positive).

The system (24)–(25) on T has a unique orbit that is tangent to the line $I = \frac{k}{m_n - m_a}$. The point of tangency is (S^*, I^*), $I^* = \frac{k}{m_n - m_a}$. Let Γ^* denote the part of this orbit with $S^* < S < 1$. Let Γ^* have the equation $S = S_*(I)$, $0 < I < \frac{k}{m_n - m_a}$. Let $V_- = \{(S, I) \in \hat{T} : 0 < S < S_*(I) \text{ and } I < \frac{k}{m_n - m_a}\}$, and let $V_+ = \{(S, I) \in \hat{T} : S_*(I) \leq S \text{ and } I < \frac{k}{m_n - m_a}\}$. See Fig. 4.

Let $(S_0, I_0) \in V_-$ and let $(S(t), I(t)) = \psi_1\big((S_0, I_0), t\big)$. Then $(S(t), I(t)) \in V_-$ for all $t \geq 0$. Thus $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big)$ is never 0. As $t \to \infty$, $(S(t), I(t))$ approaches an equilibrium $(S_f, 0)$ of (24)–(25). In this case, for small $\epsilon > 0$, a solution of (13)–(15) that enters a neighborhood of the plane $x = 1$ near (S_0, I_0) will track a solution of (24)–(25) near $(S(t), I(t))$ and approach an equilibrium $(S_f^\epsilon, 0, 1)$ of (13)–(15) with S_f^ϵ near S_f. See Sect. 6.4.

Let $(S_0, I_0) \in V_+$ and let $(S(t), I(t)) = \psi_1\big((S_0, I_0), t\big)$. Then $(S(t), I(t))$ enters the region $I \geq \frac{k}{m_n - m_a}$ at $t = t_{in} > 0$ and leaves that region at $t = t_{out} \geq t = t_{in}$.

If $-\int_0^{t_{out}} k - (m_n - m_a)I(t) \, dt < 0$, then there is no point (S_1, I_1) where $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = 0$. As in the previous paragraph, let $(S_f, 0) = \lim_{t \to \infty} (S(t), I(t))$. In this case also, for small $\epsilon > 0$, a solution of (13)–(15) that enters a neighborhood of the plane $x = 1$ near (S_0, I_0) will track a solution of (24)–(25) near $(S(t), I(t))$ and approach an equilibrium $(S_f^\epsilon, 0, 1)$ of (13)–(15) with S_f^ϵ near S_f.
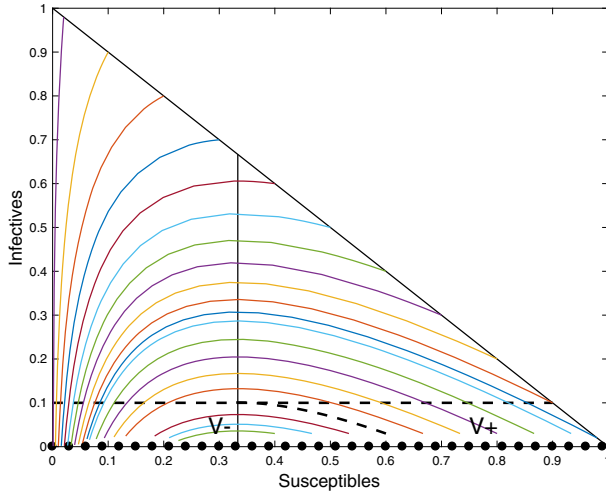
Fig. 4 Phase portrait of the fast limit system (24)–(25) in the triangle $x = 1$ with $\beta_n = 1/2$ and $\gamma = 1/6$. The vertical line $S = \gamma/\beta = 1/3$ and the horizontal line $I = \frac{k}{m_n - m_a} = \frac{1}{10}$ are shown, as are the sets V_- and V_+ bounded above by this line. Solutions that start in V_- approach equilibria without crossing the line $I = \frac{k}{m_n - m_a}$; solutions that start in V_+ cross the line

If $-\int_0^{t_{\text{out}}} k - (m_n - m_a) I(t) \, dt > 0$, then there is a unique point (S_1, I_1), with $I_1 > \frac{k}{m_n - m_a}$, where $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = 0$. For small $\epsilon > 0$, a solution of (13)–(15) that enters a neighborhood of the plane $x = 1$ near (S_0, I_0) will track a solution of (24)–(25) near $(S(t), I(t))$ until it leaves the neighborhood near (S_1, I_1). See Sect. 6.3.

4.6 Singular orbits

Motivated by the previous subsections, we construct singular orbits (which were defined at the end of Sect. 4.1) of the system (13)–(15) (or equivalently (10)–(12)).

First we consider a singular orbit $\mathscr{S}$ with starting point (S_0, I_0, x_0), $(S_0, I_0) \in \hat{T}$, $I_0 < \frac{k}{m_n - m_a}$, and $0 < x_0 < 1$. At this point, $\dot{x} > 0$.

1. The first orbit in $\mathscr{S}$ is a fast orbit: the portion of the line $(S, I) = (S_0, I_0)$ with $x_0 \le x < 1$.
2. To describe the next orbit in $\mathscr{S}$, there are three cases. Let $(S(t), I(t)) = \psi_1\big((S_0, I_0), t\big)$, and, given $t_1 > 0$, let $(S_1, I_1) = \big(S(t_1), I(t_1)\big)$.

   2a. Suppose there exists $t_1 > 0$ such $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = 0$ and $I_1 > \frac{k}{m_n - m_a}$. The next orbit of $\mathscr{S}$ is $\{(S(t), I(t)) : 0 \le t \le t_1\}$, a slow orbit.
   2b. Suppose there exists $t_1 > 0$ such that $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = 0$, and $I_1 = \frac{k}{m_n - m_a}$. In this case the construction of the singular orbit fails. (Notice $t_1 = t_{\text{out}}$ from the previous subsection. We discuss this case further after the proof of Theorem 2 in Sect. 6.1.)

2c. Suppose there is no $t_1 > 0$ such that $\mathscr{I}_1\big((S_0, I_0), (S_1, I_1)\big) = 0$. Then the next orbit of $\mathscr{S}$ is $\{(S(t), I(t)) : t \geq 0\}$, a slow orbit. This orbit approaches an equilibrium of (24)–(25). The construction of $\mathscr{S}$ terminates.

3. We continue the construction of $\mathscr{S}$ in case 2a. The next orbit in $\mathscr{S}$ is a fast orbit: the portion of the line $(S, I) = (S_1, I_1)$ with $0 < x < 1$.

4. Let $(S(t), I(t)) = \psi_0\big((S_1, I_1), t\big)$. By Proposition 1 there is exactly one $t_1 > 0$ such that $(S_2, I_2) = (S(t_1), I(t_1))$ satisfies $\mathscr{I}_0\big((S_1, I_1), (S_2, I_2)\big) = 0$. We have $I_1 < \frac{k}{m_n - m_a}$. The next orbit of $\mathscr{S}$ is $\{(S(t), I(t)) : 0 \leq t \leq t_1\}$, a slow orbit.

5. The next orbit in $\mathscr{S}$ is a fast orbit: the portion of the line $(S, I) = (S_2, I_2)$ with $0 < x < 1$.

We now continue the construction at step 2, setting (S_0, I_0) equal to (S_2, I_2).

Next we consider a singular orbit $\mathscr{S}$ with starting point (S_0, I_0, x_0), $(S_0, I_0) \in \hat{T}$, $I_0 > \frac{k}{m_n - m_a}$, and $0 < x_0 < 1$. In this case the first orbit in $\mathscr{S}$ is again a fast orbit: the portion of the line $(S, I) = (S_0, I_0)$ with $0 < x \leq x_0$. We continue the construction of $\mathscr{S}$ at step 4 above, setting (S_1, I_1) equal to (S_0, I_0).

In both cases the singular orbit $\mathscr{S}$ is an alternating sequence of fast and slow orbits, with the first orbit fast. The slow orbits alternate between orbits in $x = 0$ and orbits in $x = 1$. The last orbit is a slow orbit in $x = 1$ that approaches an equilibrium, for which $I = 0$.

Theorem 1 *Let (S_0, I_0, x_0) satisfy $(S_0, I_0) \in \hat{T}$, $I_0 \neq \frac{k}{m_n - m_a}$, and $0 < x_0 < 1$. Suppose the construction of the singular orbit $\mathscr{S}$ that starts at (S_0, I_0, x_0) never fails at step 2 and terminates after a finite number of steps at $(S_f, 0, 1)$. Let Γ^ϵ denote the orbit of (13)–(15) that starts at (S_0, I_0, x_0). Then as $\epsilon \to 0$, $\Gamma^\epsilon \to \mathscr{S}$. The terminal point $(S_f^\epsilon, 0, 1)$ of Γ^ϵ converges to $(S_f, 0, 1)$.*

Theorem 1 applies only to singular orbits with a finite number of segments. I suspect that, in the context of Theorem 1, all singular orbits have a finite number of segments, but have not been able to prove it.

Roughly speaking, the fast jumps between $x = 0$ and $x = 1$ occur because the predominant behavior among the susceptibles has become less rewarding than the alternative. When normal behavior predominates (x near 1), if the fraction of infectives becomes high, behavior may switch to the altered form (x near 0). When altered behavior predominates (x near 0), the fraction of infectives becomes low, and behavior swiches to the normal form (x near 1).

However, the switch does not occur immediately when the number of infectives crosses the threshhold value $I = \frac{k}{m_n - m_a}$. As was mentioned in the introduction, behavior changes because of encounters with other people whose behavior offers a higher payoff than one's own, so the current behavior is "sticky." The delay until behavior changes can be calculated in the limit from the entry–exit function.

5 Examples

We consider the system (10)–(12) with the parameter values

$$\beta_n = 1/2, \ \beta_a = 1/10, \ \gamma = 1/6, \ k = 3/10, \ m_n = 5, \ m_a = 2.$$

From the values of β_n, β_a, and γ, we see $R_0 = 3$ for normal behavior and .6 for altered behavior. The phase portrait of (22)–(23) in the triangle $x = 0$ (resp. (24)–(25) in the triangle $x = 1$) is given by Fig. 2a (resp. Fig. 2b). The plane $I = \frac{k}{m_n - m_a}$ is $I = 1/10$.

We shall consider singular orbits that start at $P_{\text{start}} = (S_0, I_0, .98)$ with $I_0 < 1/10$. Such a singular orbit starts with a fast solution from P_{start} to $(S_0, I_0, 1)$. One possibility is that the singular orbit immediately ends with an orbit of (24)–(25) from $(S_0, I_0, 1)$ to a point $P_{\text{end}} = (S_f, 0, 1)$; we would represent such a singular orbit by the sequence $(P_{\text{start}}, P_{\text{end}})$. Otherwise we represent the singular orbit by a sequence

$$(P_{\text{start}}, P_1, P_2, \ldots, P_{2k}, P_{\text{end}}),$$

where

- the first fast orbit goes from $P_{\text{start}} = (S_0, I_0, .98)$ to $(S_0, I_0, 1)$;
- the first slow orbit goes from $(S_0, I_0, 1)$ to $P_1 = (S_1, I_1, 1)$;
- the second fast orbit goes from $P_1 = (S_1, I_1, 1)$ to $(S_1, I_1, 0)$;
- the second slow orbit goes from $(S_1, I_1, 0)$ to $P_2 = (S_2, I_2, 0)$ (unless it's the last slow orbit, see below);
- the third fast orbit goes from $P_2 = (S_2, I_2, 0)$ to $(S_2, I_2, 1)$;

$$\vdots$$

- the last fast orbit goes from $P_{2k} = (S_{2k}, I_{2k}, 0)$ to $(S_{2k}, I_{2k}, 1)$;
- the last slow orbit goes from $(S_{2k}, I_{2k}, 1)$ to $P_{\text{end}} = (S_f, 0, 1)$.

In other words, P_1, ..., P_{2k} are the starting points of fast jumps; P_i with i odd is in $x = 1$, and P_i with i even is in $x = 0$.

Using the Matlab routines in the appendix, one can compute singular orbits for this system. We give three examples. Corresponding to each example we show the solution of the fast system (13)–(15) with the same starting point and $\epsilon = .005$, on the interval $0 \le t \le 20{,}000$, computed using the Matlab ODE solver ode23s with the options RelTol = 1e−10 and AbsTol = 1e−11. Because $\epsilon = .005$, 1000 units of fast time correspond to five units of slow time, i.e., 5 days. To compare with the singular orbits, we give the value of I at $x = 1/2$ along each jump, and the value of S at $t = 30{,}000$.

Example 1 A singular orbit with two jumps.

$$P_{\text{start}} = (.97, .03, .98)$$
$$P_1 = (.6713533014, .2059798507, 1)$$
$$P_2 = (.5714338970, .0373213930, 0)$$
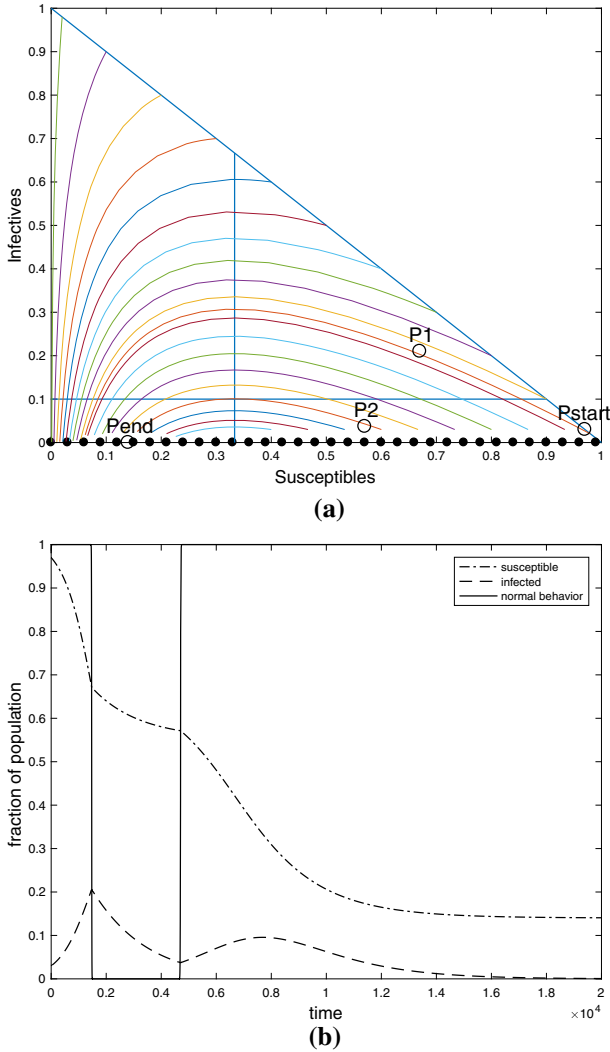
(a)

(b)

Fig. 5 Example 1, $P_{start} = (.97, .03, .98)$. **a** Phase portrait of the slow limit system in the triangle $x = 1$, with the vertical line $S = \frac{\beta_n}{\gamma} = \frac{1}{3}$ and the horizontal line $I = \frac{k}{m_n - m_a} = \frac{1}{10}$ shown. The slow orbits from P_{start} to P_1 and from P_2 to P_{end} are shown in this phase portrait. The slow orbit from P_1 to P_2 lies in the triangle $x = 0$; see Fig. 2a. **b** Solution of the fast system (13)–(15) with the same starting point and $\epsilon = .005$

$$P_{end} = (.1400580768, 0, 1)$$

See Fig. 5. For the computed solution, jumps in x occur successively at $I = .20535$ and $I = .03735$; $S = .14017$ at $t = 30,000$. Infections initially rise, then the epidemic is controlled by altered behavior for a while. When the population switches back to normal behavior, infections rise for a while, then fall to zero.
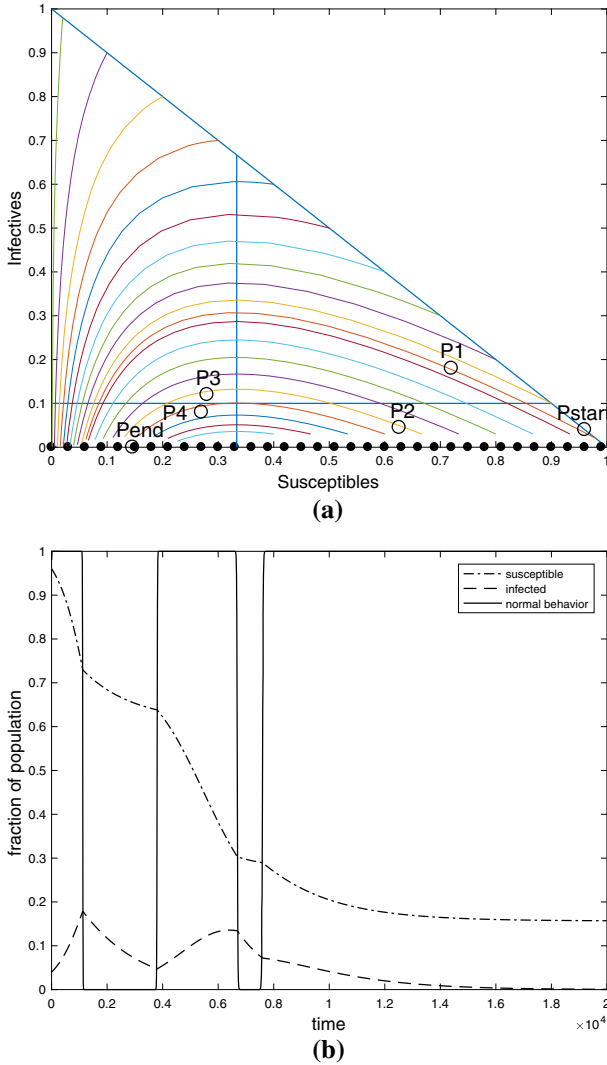
(a)

(b)

Fig. 6 Example 2, $P_{\text{start}} = (.96, .04, .98)$. **a** Phase portrait of the slow limit system in the triangle $x = 1$, with the vertical line $S = \frac{\beta_n}{\gamma} = \frac{1}{3}$ and the horizontal line $I = \frac{k}{m_n - m_a} = \frac{1}{10}$ shown. The slow orbits from P_{start} to P_1, from P_2 to P_3, and from P_4 to P_{end} are shown in this phase portrait. The slow orbits from P_1 to P_2 and from from P_3 to P_5 lie in the triangle $x = 0$; see Fig. 2a. **b** Solution of the fast system (13)–(15) with the same starting point and $\epsilon = .005$

Example 2 A singular orbit with four jumps. This example was shown in the introduction.

$$P_{\text{start}} = (.96, .04, .98)$$
$$P_1 = (.7197479246, .1842413292, 1)$$
$$P_2 = (.6258761345, .0451988482, 0)$$

$$P_3 = (.2763360357, .1222273345, 1)$$
$$P_4 = (.2682106618, .0806111708, 0)$$
$$P_{end} = (.1459222576, 0, 1)$$

See Fig. 6. For the computed solution, jumps in x occur successively at $I = .17815$, $I = .04756$, $I = .13366$, and $I = .07233$; $S = .15656$ at $t = 30,000$. In this example, when infections rise after the population switches back to normal behavior, the population again switches to altered behavior. Eventually it switches back to normal behavior; this time there is no rise in infections, and infections fall to zero.

Example 3 A singular orbit with six jumps.

$$P_{start} = (.93, .07, .98)$$
$$P_1 = (.8251362461, .1349850649, 1)$$
$$P_2 = (.7667769297, .0710900559, 0)$$
$$P_3 = (.6515152002, .1320533864, 1)$$
$$P_4 = (.6155232547, .0733319974, 0)$$
$$P_5 = (.4804269385, .1258291260, 1)$$
$$P_6 = (.4615380487, .0778669034, 0)$$
$$P_{end} = (.1387323862, 0, 1)$$

See Fig. 7. For the computed solution, jumps in x occur successively at $I = .13931$, $I = .07344$, $I = .12876$, $I = .07547$, $I = .12329$, and $I = .07959$; $S = .13760$ at $t = 30,000$. In this example, the population switches to altered behavior three times after a rise in infections with normal behavior. After the final episode of altered behavior, when the population switches back to normal behavior, infections rise and then fall to zero.

6 Proofs

6.1 Entry–exit function

Let U be an open subset of $\mathbb{R}^n$, $n \geq 1$, and consider the system

$$\dot{c} = p(c, z, \epsilon), \tag{30}$$
$$\epsilon \dot{z} = zq(c, z, \epsilon), \tag{31}$$

with $(c, z, \epsilon) \in U \times [0, z_0) \times [0, \epsilon_0)$ and $\dot{} = \frac{d}{dt}$. We assume p and q are of class C^r, $r \geq 2$. This is a slow–fast system with n-dimensional slow variable c and one-dimensional fast variable z; t is the slow time.

The fast time τ satisfies $t = \epsilon\tau$. With $' = \frac{d}{d\tau}$, system (30)–(31) becomes the fast system
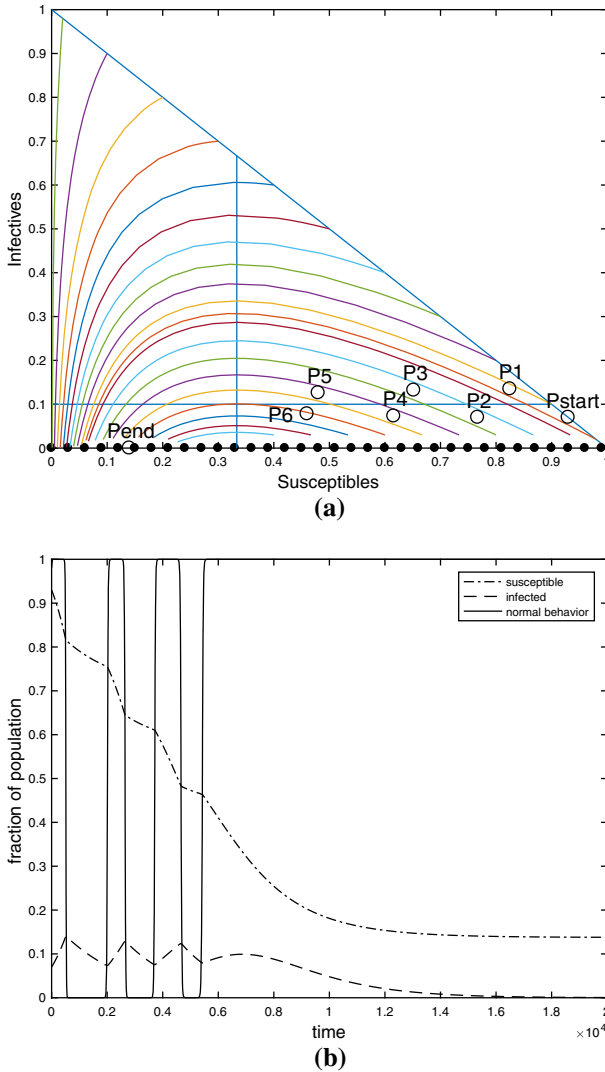
$$c' = \epsilon p(c, z, \epsilon), \tag{32}$$

(a)

(b)

Fig. 7 Example 3, $P_{start} = (.93, .07, .98)$. **a** Phase portrait of the slow limit system in the triangle $x = 1$, with the vertical line $S = \frac{\beta_n}{\gamma} = \frac{1}{3}$ and the horizontal line $I = \frac{k}{m_n - m_a} = \frac{1}{10}$ shown. The slow orbits from P_{start} to P_1, from P_2 to P_3, from P_4 to P_5, and from P_6 to P_{end} are shown in this phase portrait. The slow orbits from P_1 to P_2, from P_3 to P_4, and from P_5 to P_6 lie in the triangle $x = 0$; see Fig. 2a. **b** Solution of the fast system (13)–(15) with the same starting point and $\epsilon = .005$

$$z' = zq(c, z, \epsilon). \tag{33}$$

The slow and fast systems have the same phase portraits for $\epsilon > 0$ but have different limits at $\epsilon = 0$. For $\epsilon = 0$, the slow system (30)–(31) becomes the slow limit system

$$\dot{c} = p(c, z, 0), \tag{34}$$
$$0 = zq(c, z, 0), \tag{35}$$

and the fast system (32)–(33) becomes the fast limit system

$$c' = 0, \tag{36}$$

$$z' = zq(c, z, 0). \tag{37}$$

The fast limit system (36)–(37) has the n-dimensional plane of equilibria $z = 0$. The linearization of the fast limit system at one of these equilibria has at least n zero eigenvalues. The equilbrium is called normally hyperbolic if the $(n+1)$th eigenvalue is nonzero. More precisely, the equilibrium is called normally attracting (resp. normally repelling) if the $(n + 1)$th eigenvalue is negative (resp. positive).

Let $\tilde{p}(c) = p(c, 0, 0)$ and $\tilde{q}(c) = q(c, 0, 0)$. Then the equilibrium $(c, 0)$ of the fast limit system (36)–(37) is normally attracting (resp. normally repelling) if $\tilde{q}(c) < 0$ (resp. $\tilde{q}(c) > 0$). We assume

(E) if $\tilde{q}(c) = 0$, then $D\tilde{q}(c)\tilde{p}(c) > 0$.

Assumption (E) says in particular that if $\tilde{q}(c) = 0$, then $D\tilde{q}(c) \neq 0$. This implies that the equation $\tilde{q}(c) = 0$ defines a C^r codimension-one submanifold S of U. It separates points $(c, 0)$ where $\tilde{q}(c) < 0$, i.e., normally attracting equilibria of the fast limit system (34)–(35), from points $(c, 0)$ where $\tilde{q}(c) > 0$, i.e., normally repelling equilibria of the fast limit system.

For $c_0 \in U$ with $\tilde{q}(c_0) < 0$, let $\psi(c_0, t)$ denote the solution of (34) with $\psi(c_0, 0) = c_0$. Assumption (E) implies that $\psi(c_0, t)$ crosses the manifold S at most once, say at $t = \tilde{t} > 0$. Moreover, assumption (E) implies that for $t < \tilde{t}$, $\tilde{q}(\psi(c_0, t)) < 0$, so $(\psi(c_0, t), 0)$ is a normally attracting equilibrium of the slow limit system; and for $t > \tilde{t}$, $\tilde{q}(\psi(c_0, t)) > 0$, so $(\psi(c_0, t), 0)$ is a normally repelling equilibrium of the slow limit system.
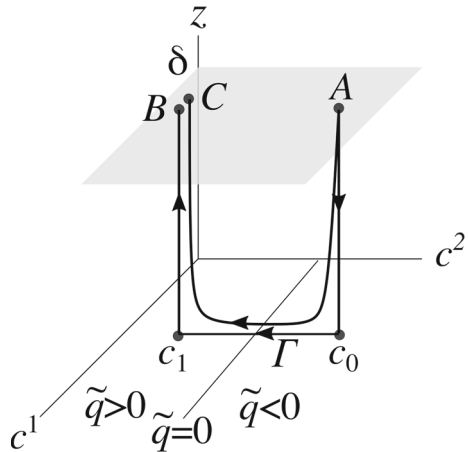
Given $t_1 > 0$, let $c_1 = \psi(c_0, t_1)$. Define the entry–exit integral

$$\mathscr{I}(c_0, c_1) = \int_0^{t_1} \tilde{q}(\psi(c_0, t)) \, dt. \tag{38}$$

Theorem 2 *For the system* (30)–(31) *with* p *and* q *of class* C^r, *assume* (E), $\tilde{q}(\bar{c}_0) < 0$, *and* $\mathscr{I}(\bar{c}_0, \bar{c}_1) = 0$, *so that we necessarily have* $\tilde{q}(\bar{c}_1) > 0$. *For a small neighborhood* $\tilde{U}$ *of* $\bar{c}_0$ *in* U, *define the entry–exit function* $P^0 : \tilde{U} \to U$ *by* $P^0(c_0) = c_1$, *where* c_1 *is defined implicitly by* $\mathscr{I}(c_0, c_1) = 0$. *Fix* $\delta > 0$ *sufficiently small. For a given* $\epsilon > 0$, *consider the solution of* (32)–(33) *that starts at* $(c, z) = (c_0, \delta)$, *with* $c_0 \in \tilde{U}$. *Then:*

1. *For* $\epsilon > 0$ *sufficiently small, the solution reintersects the section* $z = \delta$ *at a point* $(c, z) = (P^\epsilon(c_0), \delta)$.
2. P^ϵ *and* P^0 *are* C^r *functions, and* $P^\epsilon \to P^0$ *in the* C^r *sense as* $\epsilon \to 0$.
3. *Let* Γ^ϵ *denote the orbit of* (32)–(33) *from* (c_0, δ) *to* $(P^\epsilon(c_0), \delta)$. *As* $\epsilon \to 0$, Γ^ϵ *approaches the singular orbit of* (32)–(33) *consisting of*

   (a) *the line segment* $[(c_0, \delta), (c_0, 0))$, *an orbit of* (36)–(37);
   (b) $(\Gamma, 0)$, *where* Γ *is the orbit of* $\dot{c} = \tilde{p}(c)$ *from* c_0 *to* $c_1 = P^0(c_0)$; $(\Gamma, 0)$ *is an orbit of* (34)–(35).
   (c) *the line segment* $((c_1, 0), (c_1, \delta])$, *an orbit of* (36)–(37).

See Fig. 8.

Theorem 2 is proved in De Maesschalck (2008) under the assumption that for $\epsilon = 0$, the system (32)–(33) has been written in a standard form. The relation of the standard form to the form (32)–(33) is explained in Liu (2000).

In Theorem 2, we assumed $\mathscr{I}(\bar{c}_0, \bar{c}_1) = 0$, and we used the fact that for c_0 near $\bar{c}_0$ there exists c_1 near $\bar{c}_1$ such that $\mathscr{I}(c_0, c_1) = 0$. This can be shown using the Implicit Function Theorem; it requires that $\bar{c}_0$ lie in the region where $\tilde{q} > 0$. If $\bar{c}_0$ were in the boundary of this region, the construction would not work; in addition, we could not necessarily define $P^\epsilon(\bar{c}_0)$ near $P^0(\bar{c}_0)$. This is why, in Sect. 4.6, the construction of the singular orbit fails in case 2b.

6.2 Entry–exit function for the Poletti model in the triangle $x = 0$

The Poletti model (13)–(15) satisfies the hypotheses of Theorem 2 for $n = 1$ and any $r \geq 2$, with (S, I) corresponding to p and x corresponding to z. The set U is $\hat{T}$, and S is the line $I = \frac{k}{m_n - m_a}$. ($\hat{T}$ is not open, since it includes a segment of the line $S + T = 1$, but this does not cause any difficulty.)

Let $(S_0, I_0) \in \hat{T}$ with $I_0 > \frac{k}{m_n - m_a}$. Let $(S(t), I(t)) = \psi_0((S_0, I_0), t)$, let $t_1 > 0$, and let $(S_1, I_1) = (S(t_1), I(t_1))$. The formula (27) for the entry–exit integral $\mathscr{I}_0$ follows immediately from (38).

Proof of Proposition 1 We consider the solution $(S(t), I(t))$ of (22)–(23) defined above. Since $I(t)$ is decreasing, there is a unique $t_* > 0$ such that $I(t_*) = \frac{k}{m_n - m_a}$. The integral $\int_0^{t_1} k - (m_n - m_a)I(t) \, dt$ is negative and decreasing for $0 < t \leq t_*$ and is increasing for $t > t_*$. To prove Proposition 1 it suffices to show that $\int_0^\infty k - (m_n - m_a)I(t) \, dt = \infty$. Since $\int_0^\infty k \, dt = \infty$, it suffices to show that $\int_0^\infty (m_n - m_a)I(t) \, dt$ is finite. To see this, just note that $(S(t), I(t))$ approaches a normally attracting equilibrium $(S_f, 0)$, so $I(t) \to 0$ exponentially.

6.3 Entry–exit function for the Poletti model in the triangle $x = 1$

To treat the Poletti model (13)–(15) near $x = 1$, we first make the change of variables $y = 1 - x$. We obtain the system

$$S' = -\epsilon \left(\beta_n(1 - y) + \beta_a y \right) SI, \tag{39}$$

$$I' = \epsilon \left(\beta_n(1 - y) + \beta_a y \right) SI - \epsilon \gamma I, \tag{40}$$

$$y' = \epsilon(1 - y)y(\beta_n - \beta_a)I - (1 - y)y\left(k - (m_n - m_a)I \right). \tag{41}$$

Define the curve C to be the union of the line $I = \frac{k}{m_n - m_a}$, $0 < S \leq S^*$, and Γ^* defined in Sect. 4.5. Let U be the part of T above C.

The system (13)–(15) satisfies the hypotheses of Theorem 2 for $n = 1$ and any $r \geq 2$, with (S, I) corresponding to p and y corresponding to z. The set U is defined above, and S is the line segment $I = \frac{k}{m_n - m_a}$, $S^* < S < 1$. (Again the set U is not open because it includes a segment of the line $S + T = 1$, but this does not cause any difficulty.)

As in the previous subsection, the formula for the entry–exit integral $\mathcal{I}_1\big((S_0, I_0), (S_1, I_1)\big)$ follows immediately from (38).

6.4 Solutions that approach equilibria

Recall the sets V_- and V_+ defined in Sect. 4.5.

Proposition 2 *Let K be a compact subset of V_-. For each $(S_0, I_0) \in K$, let $(S_f, 0) = \lim_{t \to \infty} \psi_0\big((S_0, I_0), t\big)$. Define $Q^0 : K \to \mathbb{R}$ by $Q^0(S_0, I_0) = S_f$. Let $\delta > 0$ be small. Then:*

1. *For small $\epsilon > 0$ and for each $(S_0, I_0) \in K$, there exists $S_f^\epsilon \in \mathbb{R}$ such that $\lim_{t \to \infty} \phi^\epsilon\big((S_0, I_0, \delta), t\big) = (S_f^\epsilon, 0, 1)$.*
2. *Define $Q^\epsilon : K \to \mathbb{R}$ by $Q^\epsilon(S_0, I_0) = S_f^\epsilon$. Then Q^ϵ and Q^0 are C^{r-1} functions, and $Q^\epsilon \to Q^0$ in the C^{r-1} sense as $\epsilon \to 0$.*

Proof Let $\hat{K}$ be a compact subset of V_- that contains K in its interior. Let $\tilde{K}$ denote the union of $\hat{K}$, solutions of (24)–(25) that start in K, and the limits of these solutions.

For the system (13)–(15) with $\epsilon = 0$, $\tilde{K}$ is a union of equilbria that is compact, normally hyperbolic, and normally attracting. The point (S_0, I_0, δ) is in the stable fiber $(S_0, I_0, 1)$.

For small $\epsilon > 0$, the set $\tilde{K}$ remains normally hyperbolic and normally attracting. (S_0, I_0, δ) is in the stable fiber of a point $(S^\epsilon, I^\epsilon, 1)$ near $(S_0, I_0, 1)$. The slow system (10)–(12), restricted to $x = 1$, is still (24)–(25). The solution of (24)–(25) through (S^ϵ, I^ϵ) lies near the solution of (24)–(25) through (S_0, I_0).

Given these observations, the proposition follows from the theory of normally hyperbolic invariant manifolds. □

For each $(S_0, I_0) \in V_+$, there exists $t_{\text{in}}(S_0, I_0) > 0$ and $t_{\text{out}}(S_0, I_0) \geq t_{\text{in}}(S_0, I_0)$ such that $\psi_1\big((S_0, I_0), t\big)$ enters the region $I \geq \frac{k}{m_n - m_a}$ at $t = t_{\text{in}}(S_0, I_0)$ and leaves that region at $t = t_{\text{out}}(S_0, I_0)$.

Proposition 3 *Let K be a compact subset of V_+. Assume that for each $(S_0, I_0) \in K$,*
$-\int_0^{t_{out}(S_0, I_0)} k - (m_n - m_a)I(t) \ dt \ < \ 0$, where $I(t)$ is defined by $(S(t), I(t)) =$
$\psi_1\big((S_0, I_0), t\big)$. Choose $t_1 > \sup_K t_{out}(S_0, I_0)$. Define $Q^0 : K \to V_-$ by $Q^0(S_0, I_0) =$
$\psi_1\big((S_0, I_0), t_1\big)$. Let $\delta > 0$ be small. For small $\epsilon > 0$, define $Q^\epsilon(S_0, I_0)$ and $Z^\epsilon(S_0, I_0)$
by $(Q^\epsilon(S_0, I_0), Z^\epsilon(S_0, I_0)) = \phi^\epsilon\big((S_0, I_0, \delta), t_1\big)$. Then:

1. *Q^ϵ, Z^ϵ and Q^0 are C^r functions.*
2. *$Q^\epsilon \to Q^0$ in the C^r sense.*
3. *There exists $A > 0$ such that $Z^\epsilon \leq \delta e^{-At_1}$.*

Proof See Proposition 3 of De Maesschalck ([2008](#)) and the remark that follows it.
The key assumption needed is that for $(S_0, I_0) \in K$ and $0 < t_2 \leq t_1$, $\int_0^{t_2} k - (m_n - m_a)I(t) \ dt < 0$. □

Proposition 4 *Proposition [2](#) also holds for a compact subset K of V_+ that satisfies*
the assumption of Proposition [3](#).

Proof Roughly speaking, we want to apply Proposition [2](#) to the compact set $\phi_1(K, t_1)$
in V_-. However, corresponding to the point $(S_1, I_1) = \phi_1\big((S_0, I_0), t_1\big)$, we want to
look not at the solution of [(13)](#)–[(15)](#) that starts at (S_1, I_1, δ), but at the solution that
starts at $\phi^\epsilon\big((S_0, I_0, \delta), t_1\big)$. This requires minor changes to Proposition [2](#).

In the situation of Proposition [2](#) or [3](#), we can also describe the limiting position of
orbits.

Proposition 5 *Let K be a compact subset of V_- that satisfies the assumption of*
Proposition [2](#), or a compact subset of V_+ that satisfies the assumption of Proposi-
tion [3](#). Let $(S_0, I_0) \in K$. Then there is an equilibrium $(S_f, 0)$ of [(24)](#)–[(25)](#) such that
$\phi_1(S_0, I_0), t) \to (S_f, 0)$ as $t \to \infty$. Let Γ^ϵ denote the orbit of [(13)](#)–[(15)](#) that starts
at (S_0, I_0, δ). As $\epsilon \to 0$, Γ^ϵ approaches the singular orbit of [(13)](#)–[(15)](#) consisting of

1. *the line segment $[(S_0, I_0, \delta), (S_0, I_0, 1))$;*
2. *$\{\phi_1(S_0, I_0), t) : t \geq 0\}$.*

6.5 Proof of Theorem [1](#)

We will only consider one type of singular orbit; the proof for other types is similar.
Let (S_0, I_0, x_0) satisfy $(S_0, I_0) \in \hat{T}$, $I_0 > \frac{k}{m_n - m_a}$, and $0 < x_0 < 1$. We will consider
the following singular orbit $\mathscr{S}$:

1. Fast orbit from (S_0, I_0, x_0) to $(S_0, I_0, 0)$.
2. Slow orbit Γ_1 from $(S_0, I_0, 0)$ to $(S_1, I_1, 0)$ with $I_1 < \frac{k}{m_n - m_a}$ and
   $\mathscr{I}_0\big((S_0, I_0), (S_1, I_1)\big) = 0$.
3. Fast orbit from $(S_1, I_1, 0)$ to $(S_1, I_1, 1)$.
4. Slow orbit Γ_2 from $(S_1, I_1, 1)$ to $(S_f, 0, 1)$.

For a small $\delta > 0$, let $E_0 = \{(S, I, x) \in P : x = \delta\}$ and $E_1 = \{(S, I, x) \in P :$
$x = 1 - \delta\}$. For a small $\epsilon > 0$, let Γ^ϵ be the orbit of [(13)](#)–[(15)](#) that starts at (S_0, I_0, x_0).
We break Γ^ϵ into parts.

1. Γ_1^ϵ from (S_0, I_0, x_0) to $(\tilde{S}_0, \tilde{I}_0, \delta) \in E_0$.
2. Γ_2^ϵ from $(\tilde{S}_0, \tilde{I}_0, \delta)$ to the next intersection with E_0 at $(\tilde{S}_1, \tilde{I}_1, \delta)$.
3. Γ_3^ϵ from $(\tilde{S}_1, \tilde{I}_1, \delta)$ to $(\tilde{\tilde{S}}_1, \tilde{\tilde{I}}_1, 1-\delta) \in E_1$.
4. Γ_4^ϵ from $(\tilde{\tilde{S}}_1, \tilde{\tilde{I}}_1, 1-\delta) \in E_1$ to an equilibrium $(\tilde{S}_f, 0, 1)$.

Then, as $\epsilon \to 0$:

1. By considering the fast system (13)–(15) on $\delta \le x \le x_0$, we see that Γ_1^ϵ converges to the line segment $[S_0, I_0, x_0), (S_0, I_0, \delta)]$.
2. From Theorem 2, Γ_2^ϵ converges to the union of the line segment $[(S_0, I_0, \delta), (S_0, I_0, 0)]$, the curve Γ_1, and the line segment $[(S_1, I_1, 0), (S_1, I_1, \delta)]$.
3. By considering the fast system (13)–(15) on $\delta \le x \le 1-\delta$, we see $\Gamma_{\epsilon 3}$ converges to the line segment $[(S_1, I_1, \delta), (S_1, I_1, 1-\delta)]$.
4. From Proposition 5, $\Gamma_{\epsilon 4}$ converges to the union of the line segment $[(S_1, I_1, 1-\delta), (S_1, I_1, 0)]$ and the curve Γ_2. Moreover, the limiting equilibrium $(S_f^\epsilon, 0, 1)$ converges to $(S_f, 0, 1)$.

7 Discussion

One mathematical issue has been left hanging. We recall that Theorem 1 applies only to singular orbits with a finite number of segments. The question of whether, in the context of Theorem 1, singular orbits can have an infinite number of segments has been left open.

The model discussed in this paper could be generalized in several tantalizing directions.. One is to replace the susceptible group by several subgroups with different payoff functions. The groups could represent, for example, those with sufficient resources to survive staying home, or with the ability to work from home, and those who need to work outside the home. A second direction, suggested by the covid-19 pandemic, is to replace the infective group by subgroups. There could be a group that is infected, and infective, but so far asymptomatic, so unaware of being infective. Those in this group would continue to use the behavior they used when susceptible. Some in this group would later become symptomatic; they would presumably change their behavior at this point.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

A Matlab routines

The file findsingorbit.m is used to find a singular orbit. Parameter values are entered in the file epimconstants.m. The files entryexitint0.m and entryexitint1.m are used by findsingorbit.m to evaluate entry–exit integrals in $x = 0$ and $x = 1$ respectively.

epimconstants.m

```
% Constants used by other functions.

epsilon = 0.005;
betan = 0.5;
betaa = 0.1;
gam = 1/6;
k = 0.3;
mn = 5;
ma = 2;
```

entryexitint0.m

```
% Entry-exit function in the plane x=0.

function y = entryexitint0(S0,S1,v0)

% In the plane x=0, evaluates the integral from S0 to S1, with parameter v0,
% that is used to define the entry-exit function.  Given S0 we will want to
% find S1 such that the integral is 0.

syms S;
y = vpaintegral(top(S,v0)/bottom(S,v0),S0,S1);
return
end

function y1 = main(S,v0)
epimconstants;
syms S;
y1 = v0-S+(gam/betaa)*log(S);
return
end

function y2 = top(S,v0)
epimconstants;
syms S;
y2 = k-(mn-ma)*main(S,v0);
return
end

function y3 = bottom(S,v0);
epimconstants;
syms S;
y3 = -betaa*S*main(S,v0);
return
end
```

entryexitint1.m

```
function y = entryexitint1(S0,S1,v0)

% In the plane x=1, evaluates the integral from S0 to S1, with parameter v0,
% that is used to define the entry-exit function.  Given S0 we will want to
% find S1 such that the integral is 0.

syms S;
y = vpaintegral(top(S,v0)/bottom(S,v0),S0,S1);
return
end

function y1 = main(S,v0)
epimconstants;
syms S;
y1 = v0-S+(gam/betan)*log(S);
return
end

function y2 = top(S,v0)
epimconstants;
syms S;
y2 = k-(mn-ma)*main(S,v0);
return
end

function y3 = bottom(S,v0);
epimconstants;
syms S;
y3 = betan*S*main(S,v0);
return
end
```

findsingorbit.m

```
function singorbit = findsingorbit(S0,I0,x0)

epimconstants
syms u

singorbit = [S0 I0 x0];

while 0<S0 & 0<I0 & S0+I0<=1 & 0<x0 & x0<1 & I0 =k/(mn-ma)

    if I0>k/(mn-ma)
        disp( solution is attracted to x=0 )
        % In this case the solution is attracted to the plane x=0 and will
        % follow a solution in that plane.
        v0=I0+S0-(gam/betaa)*log(S0);

        % The solution will arrive at (S1,I1) with I1=0 where u=S1
        % satisfies:
        eqn1 = u-(gam/betaa)*log(u)==v0;
        S1=vpasolve(eqn1,u,[0,S0]);
```

```
            % The solution in SI-space will arrive at (S2,I2) with I2=k/(mn-ma)
            % where u=S2 satisfies:
            eqn2 = u-(gam/betaa)*log(u)==v0-k/(mn-ma);

            % Find where the solution in SI-space crosses the line I2=k/(mn-ma).
            disp( solution leaves the plane x=0 )
            S2=vpasolve(eqn2,u,[S1,S0]);
            syms uEnd;
            eqn3 = entryexitint0(S0,uEnd,v0)==0;
            u4=vpasolve(eqn3,uEnd,[S1,S2]);
            % Continue solution from following point.
            S0=u4;
            I0=v0-S0+(gam/betaa)*log(S0);
            x0=0.1;
            singorbit = [singorbit;[S0 I0 x0]];

    else
            disp( solution is attracted to the plane x=1 )
            % In this case the solution is attracted to the plane x=1 and will
            % follow a solution in that plane.
            v0=I0+S0-(gam/betan)*log(S0);

            % The solution will arrive at (S1,I1) with I1=0 where u=S1
            % satisfies:
            eqn1 = u-(gam/betan)*log(u)==v0;
            S1=vpasolve(eqn1,u,[0,S0]);

            % The following is the value of v0 for which the integral curve in
            % SI-space has its max on the line I=k/(mn-ma).
            v0tan = (gam/betan)-(gam/betan)*log(gam/betan)+k/(mn-ma);

            if v0 <= v0tan
                    disp( entire solution lies below I=k/(mn-ma), solution terminates )
                    % In this case the solution will not leave the plane x=1.
                    S0=S1;
                    I0=0;
                    x0=1;
                    singorbit = [singorbit;[S0 I0 x0]];

            else
                    disp( solution crosses the line I=k/(mn-ma) )
                    % In this case the solution in SI-space meets the line
                    % I=k/(mn-ma) in two points with S-values S3>S2.
                    eqn2 = u-(gam/betan)*log(u)==v0-k/(mn-ma);
                    S3=vpasolve(eqn2,u,[S1,gam/betan]);
                    int2 = entryexitint1(S0,S3,v0);

                    if int2<=0
                            disp( nevertheless solution terminates )
                            % In this case the solution will not leave the plane x=1.
                            S0=S1;
                            I0=0;
                            x0=1;
                            singorbit = [singorbit;[S0 I0 x0]];

                    else
                            disp( solution leaves the plane x=1 )
                            S2=vpasolve(eqn2,u,[gam/betan,S0]);
                            syms uEnd
                            eqn3 = entryexitint1(S0,uEnd,v0)==0;
                            u4=vpasolve(eqn3,uEnd,[S3,S2]);
                            % Continue solution from following point.
                            S0=u4;
                            I0=v0-S0+(gam/betan)*log(S0);
                            x0=0.9;
                            singorbit = [singorbit;[S0 I0 x0]];

                    end

            end

    end

end

return

end
```

References

Bauch C (2005) Imitation dynamics predict vaccinating behaviour. Proc R Soc B 272:1669–1675

Bauch C, d'Onofrio A, Manfredi P (2013) Behavioral epidemiology of infectious diseases: an overview. In: Manfredi P, d'Onofrio A (eds) Modeling the interplay between human behavior and the spread of infectious diseases. Springer, New York, NY, pp 1–19

Booth W (2020) A chilling scientific paper helped upend U.S. and U.K. coronavirus strategies, Washington Post, 17 March 2020. https://tinyurl.com/t8nfc34. Accessed 6 Oct 2020

De Maesschalck P (2008) Smoothness of transition maps in singular perturbation problems with one fast variable. J Differ Equ 244:1448–1466

d'Onofrio A, Manfredi P, Poletti P (2011) The impact of vaccine side effects on the natural history of immunization programmes: An imitation-game approach. J Theor Biol 273:63–71

Ferguson N, Cummings D, Fraser C, Cajka J, Cooley P, Burke D (2006) Strategies for mitigating an influenza pandemic. Nature 442:448–452

Ferguson N et al (2020) Imperial College COVID-19 response team report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, 16 March 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf. Accessed 6 Oct 2020

Hethcote H (2000) The mathematics of infectious diseases. SIAM Rev 42:599–653

Hethcote H (2009) The basic epidemiology models: models, expressions for r0, parameter estimation, and applications. In: Ma S (ed) Mathematical understanding of infectious disease dynamics. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore 16, World Science Publications, Hackensack, NJ, pp 1–61

Hofbauer J, Sigmund K (2003) Evolutionary game dynamics. Bull Am Math Soc (NS) 40:479–519

Jardón-Kojakhmetov H, Kuehn C, Pugliese A, Sensi M (2020) A geometric analysis of the SIR, SIRS and SIRWS epidemiological models. arXiv:2002.00354v1

Jones C (1995) Geometric singular perturbation theory. In: Dynamical systems (Montecatini Terme, 1994), lecture notes in mathematics, vol 1609. Springer, New York, NY, pp 44–118

Kuehn C (2015) Multiple time scale dynamics, applied mathematical sciences 191. Springer, New York, NY

Li M, Liu W, Shan C, Yi Y (2016) Turning points and relaxation oscillation cycles in simple epidemic models. SIAM J Appl Math 76:663–687

Liu W (2000) Exchange lemmas for singular perturbation problems with certain turning points. J Differ Equ 167:13–180

Nowak M, Sigmund K (2004) Evolutionary dynamics of biological games. Science 303:793–799

Poletti P (2010) Human behaviour in epidemic modelling. Ph.D. thesis, University of Trento, Trento, Italy. http://eprints-phd.biblio.unitn.it/422/1/tesi.pdf. Accessed 10 Oct 2020

Poletti P, Caprile B, Ajelli M, Pugliese A, Merler S (2009) Spontaneous behavioural changes in response to epidemics. J Theor Biol 260:31–40

Poletti P, Ajelli M, Merler S (2012) Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic. Math Biosci 238:80–89

Teng E (2020) 3 questions: historian Emma Teng on face masks. MIT News, 19 August 2020. https://news.mit.edu/2020/meanings-face-masks-emma-teng-0819. Accessed 12 Oct 2020

Verelst F, Willem L, Beutels P (2016) Behavioural change models for infectious disease transmission: a systematic review (2010–2015). J R Soc Interface 13:20160820

Wang Z et al (2016) Statistical physics of vaccination. Phys Rep 664:1–113