

Effect of a sharp change of the incidence function on the dynamics of a simple disease

Julien Arino^{a*} and C. Connell McCluskey^b

^aDepartment of Mathematics, University of Manitoba, Winnipeg, MB, Canada; ^bDepartment of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada

(Received 16 June 2009; final version received 16 March 2010)

We investigate two cases of a sharp change of incidence functions on the dynamics of a susceptible-infective-susceptible epidemic model. In the first case, low population levels have mass action incidence, while high population levels have proportional incidence, the switch occurring when the total population reaches a certain threshold. Using a modified Dulac theorem, we prove that this system has a single equilibrium which attracts all solutions for which the disease is present and the population remains bounded. In the second case, an increase of the number of infectives leads to a mass action term being added to a standard incidence term. We show that this allows a Hopf bifurcation to occur, with periodic orbits being generated when a locally asymptotically stable equilibrium loses stability.

Keywords: epidemiology; incidence function

1. Introduction

One of the most critical aspects in modelling disease propagation is the choice of the incidence function, which describes the rate at which new infectives are produced. If there are S susceptible individuals and I infective individuals in a population, then the rate of creation of new infectives typically involves a product of terms involving S and I . The most common incidence functions are mass action SI , proportional incidence SI/N , where N is the total population, and general incidence $S^p I^q$, but there are many other functions that have been used. There has been a debate for quite some time about the nature of the functions to use; see, for example, McCallum *et al.* [9] and the references therein. Most studies conducted to this day assume that the form of the incidence function is fixed for a given model. In models where there are several groups, parameters of the incidence function can vary from group to group, but the general nature of the function is unique. In Fromont *et al.* [3], a model is introduced where different geographical locations have different types of incidence functions; in this case, this is justified by the fact that some of the locations are farms, others are villages, leading to different contact structures. In single populations [15], a number of SI and SIS models are formulated with various types of demographic components, with a smooth, density-dependent incidence function. Further mathematical results on some of those

*Corresponding author. Email: arinoj@cc.umanitoba.ca

models are provided in Zhou [14]. In Brauer [1], the incidence takes the form $\beta(N)SI$, with $\beta(N)$ differentiable. In Zegeling and Kooij [13], an SI model is studied for micro- and macro-parasitic diseases, with strong dependence of the incidence on the population size. In Liu *et al.* [8], an incidence function that takes into account the number of hospitalized infectives is used; in both Liu *et al.* [8] and Zegeling and Kooij [13], it is shown that such dependence leads to oscillations. See also, for instance, [4,10,11].

In the present paper, we consider an SIS model for the spread of an infectious disease in a single population. These models, with the population split between susceptible and infective individuals, constitute the most elementary description of the spread of a disease in a population, and have been extensively studied in the literature. However, we suppose that there is a population level across which the contact structure is fundamentally modified. Hence, contrary to models with varying contact levels [1], there is a sharp threshold where the nature of the incidence function itself changes. A similar situation was considered in Kribs-Zaleta [7], with logistic dynamics.

We consider here an SIS model in an exponential population dynamics framework, with no self-regulation effects due to intra-specific competition (crowding). Thus, if the birth rate is too low or too high, the population can go extinct or become unbounded. In Section 2, we consider this general model with unspecified incidence and establish some elementary properties.

We then specialize this framework to two different incidence functions with switching.

The first case we consider, in Section 3, has the same incidence form as in Kribs-Zaleta [7]. The switching occurs when the total population reaches a certain threshold, with mass action incidence for small populations and standard incidence for large populations. This can model a global (and abrupt) modification of the contact structure for high population levels. To study the stability of the endemic equilibrium, we use an extension of Dulac's theorem that is developed in Appendix A.

The second switching scenario addresses the situation where treatment facilities are overcome. At low incidence levels, standard incidence is used; when a certain threshold number of infectives is reached, a mass action type term is added to the incidence function. This describes the situation where, for example, a community has a limited number of hospital beds. When this capacity is overcome, infectives are treated at home rather than in hospital, where their contact pattern is more prone to the propagation of the disease. This change in the incidence is manifested in the incidence function; if the number of infectives is below a threshold value then the incidence is given by proportional mixing, whereas the additional incidence due to any amount of infectives above the threshold is of mass action form. We formulate and analyse this model in Section 4, studying in particular the occurrence of periodic solutions.

We now call attention to an important difference in the dynamics exhibited by the two models. For the second model, it is possible to have a Hopf bifurcation, whereas this cannot happen for the first model.

2. The general framework

We make the following assumptions. There is no vertical transmission of the disease, so all newborns are susceptibles; birth occurs with rate constant $b > 0$, proportional to the total population. The mortality coefficient for non-disease-related reasons is $d > 0$. Infective individuals are subject to disease-induced death, which occurs with rate constant $\delta \geq 0$. Infective individuals may recover after an average duration of infection of $1/\gamma$ time units. Finally, contacts between susceptible and infective individuals result in new infection with rate $F(S, I)$. We thus consider the following SIS epidemic model,

$$S' = bN - dS - F(S, I) + \gamma I, \quad (1a)$$

$$I' = F(S, I) - (d + \delta + \gamma)I. \quad (1b)$$

The precise nature of the incidence function F will be made explicit later. For now, we only assume that it is a continuous piecewise differentiable non-negative function such that $F(S, 0) = F(0, I) = 0$. It follows that the non-negative quadrant is positively invariant under Equation (1). Let $N = S + I$. Then,

$$N' = (b - d)N - \delta I. \quad (2)$$

Clearly, if N tends to zero, then so must S and I .

2.1. Equilibria

Disease-free equilibria are obtained by setting $I = 0$ in Equation (1). This yields the extinction equilibrium $(S, I) = (0, 0)$, which is present for all parameter values; whereas, if $b = d$, then $(S, I) = (S, 0)$ is an equilibrium for any $S > 0$. In this paper, we assume that $b \neq d$ and so we ignore this latter case.

From Equation (2), endemic equilibria (EEP), that is, equilibria (S^*, I^*) such that $I^* > 0$, can only exist if $d < b \leq d + \delta$, and must satisfy

$$I^* = \frac{b - d}{\delta} N^*,$$

where $N^* = S^* + I^*$ is determined by the exact form of the incidence function F that is used in Equation (1). Additionally, if $b = d + \delta$, then Equation (2) implies that an endemic equilibrium would satisfy $N^* = I^*$, and so S^* would be zero. Substituting this into Equation (1b) yields $I^* = 0$. Thus, endemic equilibria are only possible if $d < b < d + \delta$.

PROPOSITION 1 *A necessary condition for the SIS model (1) to have endemic equilibria is that*

$$d < b < d + \delta. \quad (3)$$

2.2. The system in proportions

In order to simplify the problem, we consider from now on the system in proportions. We let

$$s = \frac{S}{N}, \quad i = \frac{I}{N},$$

and, noting that $s + i = 1$, we transform Equation (1) into a system in terms of (i, N) . Differentiating i with respect to t gives

$$i' = \frac{I'}{N} - i \frac{N'}{N}.$$

Substituting Equation (1b) into this expression and using the notation

$$f(i, N) = \frac{F(S, I)}{N},$$

we obtain

$$i' = f(i, N) - (b + \gamma + \delta)i + \delta i^2.$$

This definition of f is only valid for $N > 0$, so we define

$$f(i, 0) = \lim_{N \rightarrow 0} f(i, N)$$

and require that this limit exist for $i \in [0, 1]$. This precludes, for example, the possibility that $F(\alpha S, \alpha I)$ is sublinear in α near 0.

The system under consideration is then

$$i' = f(i, N) - (b + \delta + \gamma)i + \delta i^2, \quad (4a)$$

$$N' = (b - d - \delta i)N. \quad (4b)$$

Under the dynamics of Equation (4), $i(t)$ remains in $[0, 1]$ for all positive times since f is zero if i is 0 or 1. If a solution starts with $i(0)$ less than or equal to one, then $i(t)$ is strictly less than one for all positive time. Also, if $N(0) \geq 0$, then $N(t)$ remains non-negative for all time, and satisfies $N(t) \leq N(0) \exp((b - d)t)$, giving the following result.

PROPOSITION 2 *Solutions to system (4) beginning in the positively invariant rectangular half-strip of the (i, N) -plane defined by*

$$\mathcal{D} = \{(i, N) : 0 \leq i \leq 1, N \geq 0\}$$

exist for all positive times.

Equation (4b) is separable, yielding

$$N(t) = N(0) \exp\left(\int_0^t b - d - \delta i(z) dz\right).$$

Considering this in the cases $b < d$ and $d + \delta < b$ gives a monotone function for N .

- If $b < d$, then $N(t) < N(0) \exp(-\varepsilon t)$ for some $\varepsilon > 0$.
- Since $i(t) \leq 1$ for all t , $b - d - \delta i(z) > b - d - \delta$, and it follows that if $b > d + \delta$, then $N(t) > N(0) \exp(\varepsilon t)$ for some $\varepsilon > 0$.

The following result has been proved.

PROPOSITION 3 *If $b < d$, then solutions of Equation (4) satisfy $\lim_{t \rightarrow \infty} N(t) = 0$. If $b > d + \delta$ and $N(0) > 0$, then solutions of Equation (4) satisfy $\lim_{t \rightarrow \infty} N(t) = \infty$.*

The Jacobian matrix associated with system (4) will often be used. At an arbitrary point (i, N) for which f is differentiable, the Jacobian is given by

$$J = \begin{pmatrix} \frac{\partial f(i, N)}{\partial i} - (b + \delta + \gamma) + 2\delta i & \frac{\partial f(i, N)}{\partial N} \\ -\delta N & b - d - \delta i \end{pmatrix}. \quad (5)$$

3. The model with switching on population density

The novelty of the model discussed here lies in the incidence function F . In this section, we assume that the incidence is a continuous function of the form

$$F(S, I) = \begin{cases} \beta SI & \text{if } N \leq 1, \\ \beta \frac{SI}{N} & \text{if } N > 1. \end{cases} \quad (6)$$

The situation described by Equation (6) is one where the type of contact structure varies sharply as the total population density crosses a threshold value. Here, the population has been scaled to

this threshold so that the transition occurs at $N = 1$. When the population is small enough, it is assumed that every infective individual can potentially meet every susceptible in the population, so that the incidence is of mass action type. As the population becomes larger, this hypothesis must be relaxed, and every infective is able to make contact with only a certain fraction of the susceptible population, leading to the use of proportional incidence. Note that $F(S, 0) = F(0, I) = 0$, as assumed in Section 2.

3.1. The system in proportions

The incidence function f for the system in proportions is as follows:

$$f(i, N) = \begin{cases} \beta(1-i)iN & \text{if } N \leq 1, \\ \beta(1-i)i & \text{if } N > 1. \end{cases} \quad (7)$$

This incidence function is substituted into Equation (4) to obtain the main equations for this section. For $N \leq 1$, the dynamics are described by

$$i' = (\beta N(1-i) - (b + \delta + \gamma) + \delta i)i, \quad (8a)$$

$$N' = (b - d - \delta i)N, \quad (8b)$$

whereas if $N > 1$, we have

$$i' = (\beta(1-i) - (b + \delta + \gamma) + \delta i)i, \quad (9a)$$

$$N' = (b - d - \delta i)N. \quad (9b)$$

We call Equation (8) the lower system and Equation (9) the higher system. We distinguish between the low strip \mathcal{D}_L , where $N \leq 1$, and the high strip \mathcal{D}_H , where $N \geq 1$.

3.2. Exact solution in \mathcal{D}_H

Let us first remark that the nature of those parts of a solution of Equation (4) that lie in \mathcal{D}_H are known. Indeed, we have the following theorem.

PROPOSITION 4 *Suppose that, at some instant $\tau \geq 0$, we have $(i_\tau, N_\tau) := (i(\tau), N(\tau)) \in \mathcal{D}_H$. Then there exists a potentially infinite interval \mathcal{I} with left endpoint τ such that, for all $t \in \mathcal{I}$, we have $(i(t), N(t)) \in \mathcal{D}_H$ with*

$$i(t) = \frac{K i_\tau}{\Psi(t - \tau)}, \quad (10)$$

$$N(t) = N_\tau e^{(b-d)(t-\tau)} \exp\left(-\delta K i_\tau \int_\tau^t \frac{du}{\Psi(u - \tau)}\right),$$

where $K = \beta - (b + \gamma + \delta)$ and $\Psi(u) = i_\tau(\beta - \delta)(1 - e^{-Ku}) + K e^{-Ku}$.

Proof Assume that $(i_\tau, N_\tau) \in \mathcal{D}_H$ for some $\tau \geq 0$. Then consider the initial value problem consisting of Equation (9) together with the initial condition $i(\tau) = i_\tau$ and $N(\tau) = N_\tau$. Integration of Equation (9b) gives, for $t \geq \tau$,

$$N(t) = N_\tau e^{(b-d)(t-\tau)} \exp\left(-\delta \int_\tau^t i(u) du\right). \quad (11)$$

We denote by \mathcal{I} the interval on which it remains true that $(i(t), N(t)) \in \mathcal{D}_H$. Now note that, denoting $K = \beta - (b + \delta + \gamma)$, Equation (9a) can be written as follows:

$$i' - Ki = -(\beta - \delta)i^2,$$

a Bernoulli equation (it is also separable). It follows that we have, for $t \in \mathcal{I}$,

$$i(t) = \frac{Ki_\tau}{\Psi(t - \tau)},$$

where

$$\Psi(u) = i_\tau(\beta - \delta)(1 - e^{-Ku}) + K e^{-Ku}.$$

Therefore, for $t \in \mathcal{I}$,

$$\int_\tau^t i(u) du = Ki_\tau \int_\tau^t \frac{du}{\Psi(u - \tau)}. \tag{12}$$

Substituting Equation (12) into Equation (11), we obtain the result for $t \in \mathcal{I}$. ■

There are two ways for a solution curve to intersect \mathcal{D}_H . It can start in \mathcal{D}_H , that is, correspond to an initial condition in \mathcal{D}_H , or it may correspond to a switching of incidence functions because $N = 1$ is crossed as N increases. To take this information into account, we introduce the sequence $\{t_k\}$ of *switching times*, which we define as follows. For $k = 1, 2, \dots$, the $t_k \in \bar{\mathbb{R}}$ are the times at which there is a change in Equation (6), that is, when N crosses the value 1, with $t_{k+1} = \infty$ if there is no such crossing for $t > t_k$. We take $t_0 = 0$ to be the initial time; so $N(t_0)$ is the initial condition. For subsequent t_k , $N(t_k) = 1$. So if the solution enters \mathcal{D}_H at t_k , then the integral (11) can be written as follows:

$$N(t) = e^{(b-d)(t-t_k)} \exp\left(-\delta \int_{t_k}^t i(u) du\right), \tag{13}$$

for $t_k \leq t < t_{k+1}$.

3.3. Local stability analysis

THEOREM 1 *For all values of the parameters, the disease-free equilibrium is given by*

$$e_0 := (i_0, N_0) = (0, 0).$$

If $b < d$, then there are no other equilibria and e_0 is globally asymptotically stable. If $d < b$, then e_0 is unstable.

If $d < b < d + \delta$, then the presence of endemic equilibria $e_ = (i_*, N_*)$, where $i_* = b - d/\delta$, is related to the quantity $N_\Delta = \delta(d + \delta + \gamma)/\beta(d + \delta - b)$.*

$N_\Delta < 1$		e_*
$N_\Delta = 1$		$e_* = (i_*, N_\Delta)$ is LAS
$N_\Delta > 1$		$e_* = (i_*, N)$ for any $N \geq 1$
		Does not exist

If $d + \delta \leq b$, then there are no endemic equilibria.

Proof For all parameter values, e_0 is an equilibrium of the lower system. The point e_* is an equilibrium of the lower system if $d < b < d + \delta$, guaranteeing that $i_* \in (0, 1)$, and $N_* < 1$.

A third solution to the equilibrium equations for the lower strip exists, but it is not biologically relevant since it occurs at $(i, N) = (1 + (b + \gamma/\delta), 0)$, which lies outside of \mathcal{D} .

The higher system (9) only yields equilibria for which $N_* > 1$ if $N_\Delta = 1$. In this case, there is a continuum of equilibria of the form $(i, N) = ((b - d)/\delta, N)$ for $N > 1$. However, the condition $N_\Delta = 1$ is satisfied with probability zero, and so these equilibria are not considered to be biologically significant.

Evaluating the Jacobian matrix (5) at e_0 , we have

$$J = \begin{pmatrix} -(b + \delta + \gamma) & 0 \\ 0 & b - d \end{pmatrix}.$$

It follows that the local stability of e_0 is ruled by the sign of $b - d$, with e_0 locally asymptotically stable if $b < d$ and unstable if $b > d$.

In the case $b < d$, Theorem 3 implies that $N(t) \rightarrow 0$ as $t \rightarrow \infty$. Now rewrite Equation (8a) as follows:

$$i' = (\beta N(1 - i) + \delta(i - 1) - (b + \gamma))i < (\beta N - (b + \gamma))i$$

since $i \leq 1$. As $N(t) \rightarrow 0$, there exists $\tau \geq 0$ such that $\beta N(t) < b$ for $t \geq \tau$. As a consequence, for sufficiently large t , $i' < -\gamma i$ and it follows that $i(t) \rightarrow 0$ as $t \rightarrow \infty$. Thus e_0 is globally asymptotically stable for $b < d$.

For $N_\Delta < 1$, the sign pattern of $J(e_*)$ is

$$J = \begin{pmatrix} - & + \\ - & 0 \end{pmatrix},$$

which is a sign stable pattern [6], and so e_* is locally asymptotically stable when it lies in the low strip \mathcal{D}_L . ■

3.4. Ruling out periodic solutions

Note that under the flow (4), the set $\{i = 0\}$ is positively invariant, as is the set $\{N = 0\}$. Also, $i' = -(b + \gamma)$ when $i = 1$. This implies that any periodic solution of Equation (4) is bounded away from $\{i = 0\}$, $\{N = 0\}$ and $\{i = 1\}$, that is, any periodic solution lies entirely in the interior of \mathcal{D} .

We apply Theorem 6 (Appendix 1), taking $D = \text{Int}(\mathcal{D})$, with $\Gamma = \{N = 1\}$, $D_1 = \mathcal{D}_L \cap D$ and $D_2 = \mathcal{D}_H \cap D$. Then hypotheses **(H1)** and **(H2)** in Appendix 1 are satisfied.

We use the Dulac function $\alpha(i, N) = 1/i(1 - i)N$, getting

$$\text{div}(\alpha g) = -\frac{b + \gamma}{N(1 - i)^2} \tag{14}$$

(which is strictly negative) on both D_1 and D_2 . Thus, by Theorem 6 (Appendix 1), there are no simple closed curves that are invariant under the flow (4). This rules out periodic solutions, homoclinic orbits and heteroclinic cycles. Since e_0 repels solutions from the interior of \mathcal{D} , we have the following theorem.

THEOREM 2 *Suppose $b \in (d, d + \delta)$. If $N_\Delta < 1$, then each bounded solution in the interior of \mathcal{D} limits to the endemic equilibrium e_* . If $N_\Delta > 1$, then there are no bounded solutions in the interior of \mathcal{D} .*

We note that if $i = 0$ is attracting for $N > 1$, then for $d < b$ there are solutions for which i tends to zero and N tends to infinity. Thus, the presence of e_* does not guarantee that it is globally stable.

This is distinct from the behaviour of the system studied in Kribs-Zaleta [7], where the endemic equilibrium can be shown (using the extended Dulac theorem in Appendix 1) to be globally stable whenever it is present. The difference in dynamics stems from the fact that the model in Kribs-Zaleta [7] includes logistic demographics which prevents solutions from being unbounded.

4. A model where infection overcomes treatment facilities

In this section, we consider the case where instead of switching for a given value of N , the switching occurs for a given value of I .

We want to model a situation, such as would occur in the event of overflow of treatment capacities. Suppose that a given population, say a city, has a certain number of hospital beds available, and that the disease under consideration requires hospitalization. While the number of infectives remains lower than the number of beds, infectives are cared for in the hospital, in what can be considered ideal conditions: their contacts with susceptibles from the general population are much less frequent. Then, as the epidemic progresses, there is a point where not enough beds are available, and some patients are sent home. In this situation, the contact pattern is much less favourable, with some infectives mixing more freely in the population (or, at least, in a less controlled setting). We suppose that the incidence function takes the form

$$F(S, I) = \begin{cases} \beta_1 \frac{SI}{N} & \text{if } I \leq \hat{I}, \\ \beta_1 \frac{SI}{N} + \beta_2 S(I - \hat{I}) & \text{if } I > \hat{I}, \end{cases} \quad (15)$$

where $\hat{I} > 0$ is given. The term $\beta_2 S(I - \hat{I})$ describes the additional infections caused by an elevated mixing rate for the non-hospitalized infectives.

Note that this is a toy model. It is generally assumed (as is done, for example, in Section 3) that proportional incidence is more appropriate to describe high population densities. Here, however, we want to focus on the effect of a sudden change in the contact pattern that results in the contacts becoming much more frequent. Hence the form chosen for Equation (15).

The basic analysis is very similar to the case of Section 3, and many results can be adapted.

4.1. The system in proportions

We consider the system in proportions, and so the incidence function is given by

$$f(i, N) = \begin{cases} \beta_1(1-i)i & \text{if } iN \leq \hat{I}, \\ \beta_1(1-i)i + \beta_2(1-i)(iN - \hat{I}) & \text{if } iN > \hat{I}. \end{cases} \quad (16)$$

Substituting into Equation (4), we obtain the main equations for Section 4. For $iN \leq \hat{I}$, the lower system is given by

$$i' = \beta_1(1-i)i - (b + \delta + \gamma)i + \delta i^2, \quad (17a)$$

$$N' = (b - d - \delta i)N, \quad (17b)$$

whereas if $iN > \hat{I}$, we consider the higher system

$$i' = \beta_1(1-i)i + \beta_2(1-i)(iN - \hat{I}) - (b + \delta + \gamma)i + \delta i^2, \quad (18a)$$

$$N' = (b - d - \delta i)N. \quad (18b)$$

One difference with the case of switching on the population density lies in the nature of the low and high strips, \mathcal{D}_L and \mathcal{D}_H . Here, the boundary Γ is given, for $i \in (0, 1]$, by the hyperbola $N = \hat{I}/i$. Once the strips are thus defined, some of the analysis of Section 3 carries through to the present model with little adaptation.

Note that Theorem 4 holds in \mathcal{D}_L , so that an explicit expression of the solution can be found in the lower strip.

In the lower strip \mathcal{D}_L , there are two nullclines. The i nullcline, resulting from Equation (17a), is given by

$$i = i_{LS} = 1 - \frac{b + \gamma}{\beta_1 - \delta}, \tag{19}$$

for all $N \in \mathcal{D}_L$. The second nullcline is the N nullcline, resulting from Equation (17b). It is given by

$$i = i_N = \frac{b - d}{\delta}, \tag{20}$$

for all $N \in \mathcal{D}$ (it exists both in the lower and upper strips). In the upper strip, apart from Equation (20), there is a nullcline resulting from Equation (18a) given by

$$N = N_{HS}(i) = \frac{\hat{I}}{i} + \frac{b + \gamma}{\beta_2(1 - i)} - \frac{\beta_1 - \delta}{\beta_2}. \tag{21}$$

Note that the nullclines i_{LS} and N_{HS} coincide with the hyperbola Γ , providing a strong connection between the lower and the upper strips.

Evidently, if $i_{LS} = i_N$, then there is a continuum of equilibria in the lower strip. However, the set of parameters for which this happens has measure zero in parameter space, and we choose to ignore this situation. So the equilibria are $e_0 = (0, 0)$, $\bar{e} = (i_{LS}, 0)$, $e_* = (i_N, N_{HS}(i_N))$, when $\bar{e}, e_* \in \mathcal{D}$.

THEOREM 3 *System (4) with incidence function (16) potentially has three equilibria, whose existences and stabilities are summarized in the following table*

	e_0	\bar{e}	e_*
$b < d$	GAS	Does not exist	Does not exist
$\begin{cases} \mathcal{E} < 1 \\ \mathcal{E} > 1 \end{cases}$	Unstable	GAS	Does not exist
$d < b < d + \delta$	Unstable	Theorem 4	Theorem 4
$d + \delta < b$	Unstable	Unstable	Does not exist

where

$$\mathcal{E} = \frac{\beta_1}{b + \delta + \gamma}.$$

Proof The equilibrium e_* does not exist if i_N does not belong to $(0, 1)$, hence the cases for e_* . (e_* might not exist even when $i_N \in (0, 1)$, which is treated in Theorem 4).

The extinction equilibrium $e_0 = (0, 0)$ always exists. At e_0 , the Jacobian matrix (5) takes the form

$$\begin{pmatrix} \beta_1 - (b + \delta + \gamma) & 0 \\ 0 & b - d \end{pmatrix}.$$

Thus, e_0 is locally asymptotically stable if, and only if, $b < d$ and $\mathcal{E} < 1$. Furthermore, if $b < d$ then N tends to 0. In this case, the asymptotic dynamics of i are governed by Equation (17a). If $\mathcal{E} < 1$, then i tends to 0, and if $\mathcal{E} > 1$, then i tends to i_{LS} . Thus, for $b < d$, if $\mathcal{E} < 1$ then e_0 is

globally asymptotically stable, and if $\mathcal{E} > 1$ then \bar{e} is globally asymptotically stable. (Note that $\mathcal{E} > 1$ requires $\beta_1 < \delta$, making $i_{LS} < 1$.)

For $d < b < d + \delta$, it follows from the Jacobian matrix that e_0 is unstable. (The other equilibria are treated in Theorem 4.)

Now suppose $b > d + \delta$. Then, e_* does not exist and both e_0 and \bar{e} are unstable (since in this case, $i_N > 1$, implying that both Jacobians are evaluated on the left of i_N). Also, note that from Theorem 3, $N(t)$ becomes unbounded. ■

Note that \mathcal{E} resembles the classical basic reproduction number \mathcal{R}_0 in its interpretation. Indeed, $1/(b + \delta + \gamma)$ is the mean time spent in the infective class, if natural death is replaced with birth as is the case when working in proportions. Multiplied by the rate β_1 at which individuals enter the infective class, this gives an estimation of the average number of new infections produced by an infective individual. However, for $b < d$, as \mathcal{E} passes through 1, the stability merely passes from one equilibrium with $N_* = 0$ to another. Since this role is very different from the classical role of \mathcal{R}_0 , we use a different notation.

From now on, we choose to consider only the cases when all three nullclines exist in the interior of \mathcal{D} , that is, $i_{LS}, i_N \in (0, 1)$. There are three cases to consider, which are shown in Figure 1. The following theorem summarizes these cases.

THEOREM 4 Suppose that $d < b < d + \delta$. Then, for system (4) with incidence function (16), e_0 is unstable, and the stability of \bar{e} and the existence and stability of e_* are summarized in the table

	\bar{e}	e_*
$i_{LS} < i_N$ $\left\{ \begin{array}{l} \mathcal{S}_N > 0 \\ \mathcal{S}_N < 0 \end{array} \right.$	Unstable	LAS
	Unstable	Unstable
$i_{LS} > i_N$	GAS	Does not exist

where

$$\mathcal{S}_N = \frac{(b + \gamma)(b - d)}{d + \delta - b} - \frac{\beta_2 \hat{I}(d + \delta - b)}{b - d}$$

is the slope of N_{HS} at e_* , when e_* exists.

Proof That e_0 is unstable is established in Theorem 3.

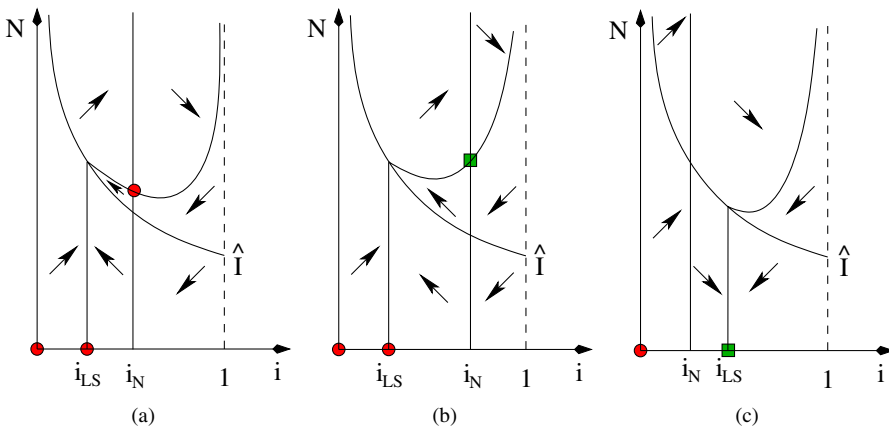


Figure 1. Phase plane situations when the incidence function takes the form (16), with the equilibria represented (circles are unstable, squares are stable). In each panel, the switching curve is the monotone decreasing curve, the nullcline for N' is given by $i = i_N$, and the nullcline for i' is the curve which has $i = 1$ as an asymptote (given by $N = N_{HS}(i)$).

When evaluated at $\bar{e} = (i_{LS}, 0)$ and to the left of i_N (i.e. when $i_{LS} < i_N$), the sign pattern of the Jacobian (5) is

$$\begin{pmatrix} * & 0 \\ 0 & + \end{pmatrix}$$

(where $*$ corresponds to a quantity that may be positive or negative), and thus \bar{e} is unstable when $i_{LS} < i_N$ because there is at least one positive eigenvalue. At \bar{e} and to the right of i_N (i.e. when $i_N < i_{LS}$), the sign pattern is

$$\begin{pmatrix} \text{sgn}(1 - \mathcal{E}) & 0 \\ 0 & - \end{pmatrix}$$

and since $\mathcal{E} > 1$ when \bar{e} exists, \bar{e} is locally asymptotically stable. That it is also globally asymptotically stable follows from a phase plane argument. First, because $i_{LS} > i_N$ (see Figure 1(c)), it is clear that the strip $\mathcal{D}_{i_N} = \{(i, N); i_N \leq i \leq 1, N \geq 0\}$ is positively invariant under the flow of Equation (4) with incidence function (16). Second, all points such that $i < i_N$ satisfy $i' > i'|_{i=i_N} > 0$. It follows that all solutions with $i < i_N$ enter \mathcal{D}_{i_N} in finite time. When a solution is in \mathcal{D}_{i_N} , N decreases so that eventually, the system is governed by the lower system. When this happens, Theorem 4 applies with $\beta = \beta_1$, and since $\mathcal{E} > 1$, $\beta_1 > b + \gamma + \delta$ and Equation (10) implies that $i(t) \rightarrow i_{LS}$ as $t \rightarrow \infty$.

If $i_N < i_{LS}$, then $N_{HS}(i_N)$ lies below the switching curve. Thus, if $i_N < i_{LS}$, then e_* is not present. For $i_N > i_{LS}$, the equilibrium e_* is present. We now determine the stability of e_* for $i_{LS} < i_N$.

Evaluating the derivative of $N_{HS}(i)$ at i_N gives the slope \mathcal{S}_N of the tangent to $N_{HS}(i)$ at e_* , when e_* exists. Then, at e_* , the Jacobian matrix (5) takes the form

$$\begin{pmatrix} -\mathcal{S}_N & \frac{\beta_2(b-d)(d+\delta-b)}{\delta^2} \\ \frac{\beta_1\delta}{\beta_2} - \frac{\delta^2\hat{I}}{b-d} - \frac{\delta^2(d+\delta+\gamma)}{\beta_2(d+\delta-b)} & 0 \end{pmatrix},$$

and thus has the sign pattern

$$J = \begin{pmatrix} -\text{sgn}(\mathcal{S}_N) & + \\ - & 0 \end{pmatrix}.$$

(The sign of J_{21} results from the fact that $J_{21} = -\delta N$). If $\mathcal{S}_N > 0$, this is a stable sign pattern [6], and so e_* is locally asymptotically stable. If $\mathcal{S}_N < 0$, then this is an unstable pattern and so e_* is unstable. ■

Interestingly, when it exists the stability of e_* is thus determined by the sign of the slope of the tangent to the nullcline N_{HS} at e_* . In fact, we show in the next section that when the sign of this slope changes, a Hopf bifurcation occurs at e_* .

4.2. Existence of periodic solutions

Theorem 4 states that the loss of stability of e_* occurs when \mathcal{S}_N changes sign. The following result relates this change in stability to a Hopf bifurcation. We treat \mathcal{S}_N as a bifurcation parameter, noting that by using β_2 , we may consider \mathcal{S}_N to be independent of b, d, δ, i_{LS} and i_N . In the theorem statement and proof, $\mathcal{B}_\epsilon(e_*)$ denotes the open ball centred at e_* with radius ϵ .

THEOREM 5 *Suppose that $d < b < d + \delta$ and $i_{LS} < i_N$. Then for any $\epsilon > 0$ and any $\mathcal{S}_N^0 > 0$, there exists $\bar{\mathcal{S}}_N \in (-\mathcal{S}_N^0, 0)$ such that system (4) with incidence function (16) has a non-trivial periodic orbit in $\mathcal{B}_\epsilon(e_*)$ for $\mathcal{S}_N = \bar{\mathcal{S}}_N$.*

Proof Using Lemma 1 in Appendix 2 with $\tau = S_N$, we see that the eigenvalues of the Jacobian matrix at e_* are purely imaginary conjugates when $S_N = 0$, and the real part of these eigenvalues decreases with S_N in a neighbourhood of $S_N = 0$. By using S_N as a bifurcation parameter (or β_2 as a proxy for S_N), the transversality as it passes through 0 is immediate. It follows that a Hopf bifurcation occurs at $S_N = 0$ [5]. In particular, e_* is attracting for $S_N > 0$ and is repelling for $S_N < 0$.

The Hopf bifurcation is occurring at e_* as S_N passes through 0, but always with the underlying assumption that $i_{LS} < i_N$. Since S_N is the derivative of $N_{HS}(i)$ at $i = i_N$ (as is the case at e_*), it follows that the bifurcation is happening in the interior of the high strip where the vector field is C^3 . Any points of non-differentiability on the periodic orbit are the result of further (minor) bifurcations as the expanding periodic orbit crosses the switching curve.

It remains to be shown that periodic orbits exist for negative S_N . As seen in Figure 2, any solution for which $N(0)$ is sufficiently large begins in one of the Regions I, II, III or IV (where the divisions between the regions are the isoclines and the switching curve). In any case, the solution will eventually advance to Region IV. In Region IV, however, N is bounded below by N_1 . Then, after leaving Region IV, the N coordinate of the positive semi-trajectory is bounded above by N_2 . Thus, solutions are bounded and do not approach e_0 or \bar{e} .

However, for S_N negative, e_* is repelling. Thus, by the Poincaré–Bendixson Theorem, the omega limit set of a solution starting near e_* , or in Regions I, II, III or IV, must be a periodic orbit. In fact, aside from e_* , any solution for which $i(0)$ and $N(0)$ are non-zero must limit to a periodic orbit. ■

In Figure 3, the parameters are $1/b = 20$ years, $1/d = 80$ years, $\delta = 0.0004$, $\gamma = 1/3$, $\beta_1 = 0.4$ and $\beta_2 = 0.001$. The initial population is $N(0) = 1000$, with one infective individual, and the switching occurs at $\hat{i} = 150$.

Note (Figure 3(a)) that the periodic orbit intersects both the lower and the upper strips. This is the situation after a so-called grazing bifurcation [2], where the periodic orbit whose existence is established in Theorem 5 hits a smooth switching boundary.

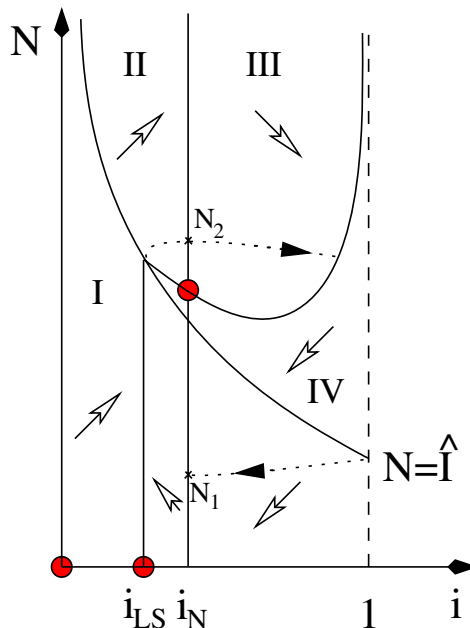


Figure 2. Phase plane diagram showing the nullclines, directions (empty headed arrows) as well as two solutions (filled head arrows), for the system with incidence function (16), in parameter regions where a periodic solution exists.

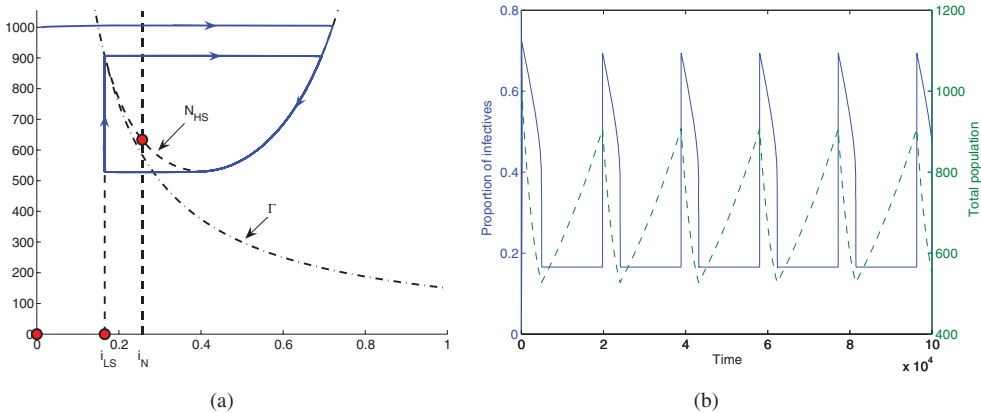


Figure 3. (a) Phase plane diagram showing the nullclines (dashed curves), switching boundary Γ (dash-dotted curve), as well as one solution along the hysteresis loop, for the system with incidence function (16). Parameters are indicated in the text. (b) A solution along the hysteresis loop in the case of Figure 3(a). The left vertical axis shows the proportion of infectives in the population, corresponding to the solid curve, while the right vertical axis shows the total population, corresponding to the dashed curve.

5. Discussion

The model we introduce here is, in essence, quite different from other models that were published on the subject. It does share some common properties with other models with exponential demography and disease induced death, in which two extreme cases occur when the birth rate is either too low or too high. For $b < d$, the population size limits to zero and extinction is inevitable, independent of the disease dynamics. For $b > d + \delta$, the population size grows without bounds. In this case, it is impossible for the disease to limit the population growth. In between these two extreme cases is a region where the disease persists.

For the model with switching on the population density of Section 3, if the population does not become extinct because of excessive disease-induced death, then the disease remains endemic. This result could be foreseen: we assume that the system switches between two well-studied vector fields, with the upper vector field having no equilibrium *per se*. To show that the endemic equilibrium attracts all bounded solutions, we use an extension of Dulac’s theorem for switching vector fields developed in Appendix 1.

A different situation arises in the model of Section 4, with switching on the number of infectives. An analysis of the phase plane, for the case when $d < b < d + \delta$, shows that there are three major configurations. In each of them, there are two equilibria where the population becomes extinct, one that is always unstable, and one whose local stability depends on the sign of an \mathcal{R}_0 -type quantity. In two of the configurations, there also exists an endemic equilibrium. We show that either this endemic equilibrium is locally asymptotically stable, or it is unstable and there is a periodic orbit. In Zhou and Hethcote [15], periodic orbits are not present. There are several major differences between the model presented here and that of Zhou and Hethcote [15], so pinpointing one single cause for these oscillations is hard. However, it is likely that the single most important difference between the models lies in the fact that in [15], the incidence function is a non-decreasing function of the *total* population, whereas here it is a function of the *infective* population. Note that the expression of S_N suggests that oscillations occur when β_2 and \hat{I} are rather large. This means that in order to observe oscillations, the prevalence of the disease in the population and the ‘cost’ of being treated outside the hospital must both be quite high. If, as in Zhou and Hethcote [15], incidence depended on the total population, then prevalence cannot reach such high levels as it does here.

It is clear that the period of the oscillations obtained in Section 4 are unrealistic (e.g. 120 years in the case of Figure 3). However, a more thorough exploration of the parameter space may reveal areas where such behaviours occur with more realistic frequencies. The models we develop here are toy models; SIS models are seldom used in real case studies, because of their simplicity. More realistic SEIRS models would most likely exhibit the same type of behaviour as our models here, if the same incidence functions were used, with perhaps more realistic oscillation periods. Furthermore, the oscillations we observe seem quite robust to the type of incidence functions used, provided the switching occurs as I passes through some threshold. For example, another way to model the saturation of treatment resources of Section 4 would be to use an incidence of the form

$$F(S, I) = \begin{cases} \beta_1 \frac{SI}{N} & \text{if } I \leq \hat{I}, \\ \beta_1 \frac{S\hat{I}}{N} + \beta_2 S(I - \hat{I}) & \text{if } I > \hat{I}, \end{cases} \quad (22)$$

with $\hat{I} > 0$ given. Such an incidence function produces the same type of oscillations as Equation (15).

Acknowledgements

The authors thank two anonymous referees for helpful suggestions. The work of both the authors was supported in part by NSERC. The work of J.A. is supported in part by MITACS.

References

- [1] F. Brauer, *Some simple epidemic models*, Math. Biosci. Eng. 3(1) (2006), pp. 1–15.
- [2] M. di Bernardo, C.J. Budd, A.R. Champneys, and P. Kowalczyk, *Piecewise-smooth Dynamical Systems*, Springer, Berlin, 2008.
- [3] E. Fromont, D. Pontier, and M. Langlais, *Disease propagation in connected host populations with density-dependent dynamics: The case of the Feline Leukemia Virus*, J. Theor. Biol. 223 (2003), pp. 465–475.
- [4] L.Q. Gao and H.W. Hethcote, *Disease transmission models with density-dependent demographics*, J. Math. Biol. 30(7) (1992), pp. 717–731.
- [5] J.K. Hale and H. Koçak, *Dynamics and Bifurcations*, Texts in Applied Mathematics Vol. 3, Springer-Verlag, New York, 1991.
- [6] C. Jeffries, V. Klee, and P. van den Driessche, *Qualitative stability of linear systems*, Linear Algebra Appl. 87 (1987), pp. 1–48.
- [7] C. Kribs-Zaleta, *To switch or taper off: The dynamics of saturation*, Math. Biosci. 192 (2004), pp. 137–152.
- [8] R. Liu, J. Wu, and H. Zhu, *Media/psychological impact on multiple outbreaks of emerging infectious diseases*, Comput. Math. Methods Med. 8(3) (2007), pp. 153–164.
- [9] H. McCallum, N. Barlow, and J. Hone, *How should pathogen transmission be modelled?*, Trends Ecol. Evol. 16(6) (2001), pp. 295–300.
- [10] J. Mena-Lorca and H.W. Hethcote, *Dynamic models of infectious diseases as regulators of population size*, J. Math. Biol. 30 (1992), pp. 693–716.
- [11] H.R. Thieme, *Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations*, Math. Biosci. 111(1) (1992), pp. 99–130.
- [12] W. Wang, *Backward bifurcation of an epidemic model with treatment*, Math. Biosci. 201(1–2) (2006), pp. 58–71.
- [13] A. Zegeling and R.E. Kooij, *Uniqueness of limit cycles in models for microparasitic and macroparasitic diseases*, J. Math. Biol. 36 (1998), pp. 407–417.
- [14] J. Zhou, *An epidemiological model with population size dependent incidence*, Rocky Mt. J. Math. 24(1) (1994), pp. 429–445.
- [15] J. Zhou and H.W. Hethcote, *Population size dependent incidence in models for diseases without immunity*, J. Math. Biol. 32 (1994), pp. 809–834.

Appendix 1. A Dulac-type theorem

We give here an extension of Dulac's Theorem that can be used to rule out periodic solutions when the vector field is C^1 everywhere except on a curve Γ on which it is non-differentiable. It was brought to our attention that this result is

quite similar to a result of Wang [12]. However, the result here covers the more general case of the non-differentiability occurring along a curve rather than along a straight line.

Let D be a simply connected open subset of \mathbb{R}^2 . Let $\Gamma \subset D$ be a simple curve such that $D \setminus \Gamma$ consists of two components D_1 and D_2 . We require the following:

(H1) For each $\epsilon > 0$, there exist simple C^1 curves $\Gamma_{j\epsilon}$ in D_j , $j = 1, 2$, such that the C^1 distance $d(\Gamma_{1\epsilon}, \Gamma_{2\epsilon})$ is less than ϵ .

A sufficient (but not necessary) condition for (H1) to be satisfied is that Γ be C^2 .

Consider the differential equation

$$\vec{x}' = g(\vec{x}), \tag{A1}$$

where $x \in D$, g is a C^0 vector field given by

$$g = \begin{cases} g_1 & \text{on } D_1, \\ g_2 & \text{on } D_2, \end{cases} \tag{A2}$$

and g_j is C^1 on D_j , $j = 1, 2$. (Note that we are assuming g to be C^1 everywhere except at Γ , where it is not differentiable.) We also assume the following:

(H2) Solutions to Equation (A1) are unique, and are transverse to Γ except possibly at a finite number of points.

THEOREM 6 (A Dulac-type Theorem) *Let α be a scalar function on D , bounded on compact subsets of D . If $\text{div}(\alpha g)$ has constant sign on $D_1 \cup D_2$, then there are no piecewise smooth, simple closed curves in D which are invariant under (A1).*

Proof Suppose $\gamma \subset D$ is a piecewise smooth, simple closed curve which is invariant under Equation (A1). The possibility of γ being contained entirely in D_1 or D_2 is precluded by Dulac's Theorem. Thus we may assume that γ intersects Γ . Let $\gamma_j = \gamma \cap D_j$ (see Figure A1).

By the Jordan Curve Theorem, there is a connected region U which is bounded by γ . Let $U_j = U \cap D_j$. (We note that U_1 and U_2 may not be connected.) Take $\epsilon > 0$ small enough so that $\Gamma_{j\epsilon}$ intersects each component of U_j .

Let $U_{j\epsilon}$ be the maximal subset of U_j which is bounded by γ_j and $\Gamma_{j\epsilon}$ and let $C_{j\epsilon}$ be its boundary. Finally, let $C_{j\epsilon}^y$ and $C_{j\epsilon}^r$ be the parts of $C_{j\epsilon}$ which coincide with γ_j and $\Gamma_{j\epsilon}$, respectively.

Then denoting $g = (u, v)$ and $\vec{x} = (x, y)$, and using Green's Theorem, we have

$$\begin{aligned} \iint_{U_{1\epsilon}} \text{div}(\alpha g) \, dx \, dy + \iint_{U_{2\epsilon}} \text{div}(\alpha g) \, dx \, dy &= \oint_{C_{1\epsilon}} (\alpha u \, dy - \alpha v \, dx) + \oint_{C_{2\epsilon}} (\alpha u \, dy - \alpha v \, dx) \\ &= \int_{C_{1\epsilon}^y} (\alpha u \, dy - \alpha v \, dx) + \int_{C_{1\epsilon}^r} (\alpha u \, dy - \alpha v \, dx) \\ &\quad + \int_{C_{2\epsilon}^y} (\alpha u \, dy - \alpha v \, dx) + \int_{C_{2\epsilon}^r} (\alpha u \, dy - \alpha v \, dx). \end{aligned}$$

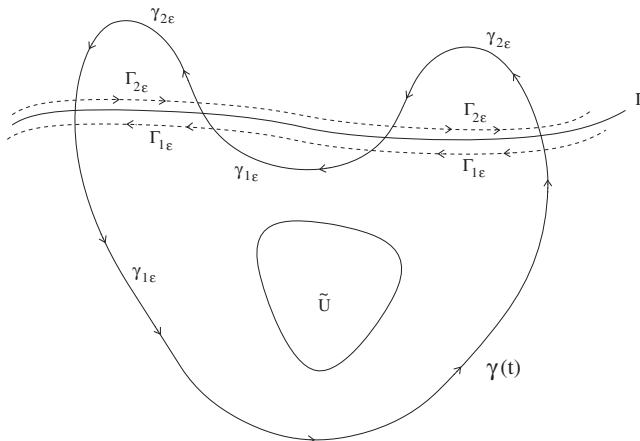


Figure A1. Example of a situation covered by Theorem 6.

Note that $dx/dt = u$ and $dy/dt = v$ on γ , it is clear that the integral over $C_{j\epsilon}^\gamma$ has an integrand which is zero leaving

$$\int \int_{U_{1\epsilon}} \operatorname{div}(\alpha g) \, dx \, dy + \int \int_{U_{2\epsilon}} \operatorname{div}(\alpha g) \, dx \, dy = \int_{C_{1\epsilon}^\Gamma} (\alpha u \, dy - \alpha v \, dx) + \int_{C_{2\epsilon}^\Gamma} (\alpha u \, dy - \alpha v \, dx). \tag{A3}$$

Note that the integrations along $C_{1\epsilon}^\Gamma$ and $C_{2\epsilon}^\Gamma$ are well-defined as ϵ tends to zero, and are taken in opposite directions in the sense that for $\epsilon = 0$, they would involve integrating along the same segment of Γ but in opposite directions. Since the C^1 distance $d(\Gamma_1, \Gamma_2) < \epsilon$ and α, u and v are bounded on the closure of U , it follows that as ϵ tends to zero, the right-hand side of Equation (A3) tends to zero.

On the other hand, the left-hand side of Equation (A3) is bounded away from zero. To see this, we note that there is an open set \tilde{U} which is contained in $U_{1\epsilon} \cup U_{2\epsilon}$ for sufficiently small ϵ . Combining this with the fact that $\operatorname{div}(\alpha g)$ has constant sign on $U_1 \cup U_2 \supseteq U_{1\epsilon} \cup U_{2\epsilon}$ implies that the left-hand side of Equation (A3) is bounded away from zero for sufficiently small ϵ .

This gives a contradiction and so we see that a piecewise smooth, simple closed curve in D cannot be invariant under Equation (A1). ■

Appendix 2. A small lemma on eigenvalues

LEMMA 1 *Let*

$$M = \begin{pmatrix} \alpha & a \\ -b & 0 \end{pmatrix}, \tag{A4}$$

where α, a, b are C^1 functions of a parameter $\tau \in \mathbb{R}$. Suppose $a(\bar{\tau}), b(\bar{\tau}) > 0$ and $\alpha(\bar{\tau}) = 0$. Then at $\tau = \bar{\tau}$, the matrix M has a pair of purely imaginary eigenvalues $\lambda_{1,2}$ that satisfy

$$\frac{d}{d\tau} \Re(\lambda_{1,2}) = \frac{1}{2} \frac{d\alpha}{d\tau}. \tag{A5}$$

Proof The characteristic equation is

$$\lambda^2 - \alpha\lambda + ab = 0.$$

It follows that eigenvalues of M are given by $\lambda_{1,2} = \frac{1}{2}(\alpha \pm \sqrt{\Delta})$, where $\Delta = \alpha^2 - 4ab$. When $\tau = \bar{\tau}$, we obtain $\lambda_{1,2} = \sqrt{-ab}$ which implies $\lambda_{1,2}$ form a pair of purely imaginary eigenvalues, since $ab > 0$.

Since Δ is negative at $\bar{\tau}$, it is negative on a neighbourhood of $\bar{\tau}$, and so $\Re(\lambda_{1,2}) = \alpha/2$ in this neighbourhood. On differentiating, we obtain Equation (A5). ■