Journal of Nonlinear Science (2019) 29:2447–2500
https://doi.org/10.1007/s00332-019-09550-7

Journal of Nonlinear Science

Complex Dynamics in a Unified SIR and HIV Disease Model: A Bifurcation Theory Approach

Pei Yu[1] · **Wenjing Zhang**[2]

Received: 7 May 2018 / Accepted: 24 April 2019 / Published online: 29 April 2019

Abstract

This paper is concerned with complex dynamical behaviors of a simple unified SIR and HIV disease model with a convex incidence and four real parameters. Due to the complex nature of the disease dynamics, our goal is to explore bifurcations including multistable states, limit cycles, and homoclinic loops in the whole parameter space. The first contribution is the proof of the existence of multiple limit cycles giving rise from Hopf bifurcation, which further induces bistable or tristable states because of the coexistence of stable equilibria and periodic motion. Next, we propose that the existence of Bogdanov–Takens (BT) bifurcation yields the bifurcation of homoclinic loops, which provides a new mechanism for generating disease recurrence, for example, the relapse–remission, viral blip cycles in HIV infection. Last, we present a novel method for the derivation of the normal forms of codimension two and three BT bifurcations. The method is based on the simplest normal form theory from Yu's previous works.

Keywords A unified SIR and HIV disease model · Recurrent infection · Stability · Hopf bifurcation · Bogdanov–Takens bifurcation · Limit cycle · Homoclnic orbit · The simplest normal form

Mathematics Subject Classification 34C07 · 34C23

Communicated by Paul Newton.

✉ Pei Yu
pyu@uwo.ca

[1] Department of Applied Mathematics, Western University, London, ON N6A 5B7, Canada

[2] Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA

1 Introduction

Susceptible-infected-recovered epidemiological models are well-known to predict possible disease outbreaks and design disease controls. But the complex disease dynamics are rarely explored in whole parameter space analytically due to the complexity of the model (including the model dimension and the number of parameters), model reduction techniques, and the computation capacity. Therefore, before carrying out the rigorous mathematical analysis, we first introduce the derivation of the simplest two-dimensional models. The basic SIR compartment frame is described as follows (Liu et al. 1986):

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= \lambda - G(I,\ S) - d\,S + \gamma\,R,\\ \frac{\mathrm{d}I}{\mathrm{d}t} &= G(I,\ S) - (d+v)\,I,\\ \frac{\mathrm{d}R}{\mathrm{d}t} &= v\,I - (d+\gamma)\,R.\end{aligned} \tag{1}$$

We assume the per capita death rate is d for three population compartments, see Earn et al. (2000). The newborns are all susceptibles and contribute to the susceptible population at the rate of λ. The infection duration is v^{-1}. The infection rate, $G(I,\ S)$, is a function in terms of both susceptible and infected populations. It follows from Griffin (1995) that the recovered population gets lifelong immunity (so that $\gamma = 0$), model (1) is reduced to a two-dimensional model, given by

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= \lambda - G(I,\ S) - d\,S,\\ \frac{\mathrm{d}I}{\mathrm{d}t} &= G(I,\ S) - (d+v)\,I.\end{aligned} \tag{2}$$

Considering the global incidence rate function in further studies, for example, see Korobeinikov and Maini (2005), Wang et al. (2010), Hethcote and van den Driessche (1991), Liu et al. (1986), the infection rate takes a special case:

$$G(I,\ S) = h(I)\,I\,S = \left(\beta_0 + \tfrac{k\,I}{1+m\,I}\right) I\,S, \tag{3}$$

where β_0, k and m are positive constants, and thus the model (2) is rewritten as

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= \lambda - \left(\beta_0 + \tfrac{k\,I}{1+m\,I}\right) I\,S - d\,S,\\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \left(\beta_0 + \tfrac{k\,I}{1+m\,I}\right) I\,S - (d+v)\,I.\end{aligned} \tag{4}$$

To simplify the analysis, we introduce the change of state variables, $S = \frac{\lambda}{d+v}\,X$, $I = \frac{\lambda}{d+v}\,Y$, and the time rescaling, $t = \frac{1}{d+v}\,\tau$, into (4) to obtain the following

dimensionless model,

$$\begin{aligned}\frac{\mathrm{d}X}{\mathrm{d}\tau} &= 1 - D\,X - \left(B + \tfrac{A\,Y}{Y+C}\right) X\,Y,\\ \frac{\mathrm{d}Y}{\mathrm{d}\tau} &= \left(B + \tfrac{A\,Y}{Y+C}\right) X\,Y - Y,\end{aligned} \tag{5}$$

where the new parameters are defined as

$$A = \frac{k\,\lambda}{m\,(d+v)^2},\quad B = \frac{\lambda\beta_0}{(d+v)^2},\quad C = \frac{d+v}{m\,\lambda},\quad D = \frac{d}{d+v}. \tag{6}$$

All these parameters take positive real values.

Next, we consider a four-dimensional in-host HIV model which was developed by van Gaalen and Wahl (2009) and used to study the viral blips phenomenon in Zhang et al. (2013, 2014a, b). The model is described by the following ordinary differential equations:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \lambda_x - d_x x - (1-\epsilon)\beta(r)xy,\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= (1-\epsilon)\beta(r)xy - d_y y,\\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \lambda_r + ky - mar - d_r r,\\ \frac{\mathrm{d}a}{\mathrm{d}t} &= \lambda_a + \alpha - par - d_a a,\end{aligned} \tag{7}$$

where x, y, r, and a represent the population densities of the uninfected $\mathrm{CD4}^+$ T cells, infected $\mathrm{CD4}^+$ T cells, reactive oxygen species (ROS), and antioxidants, respectively. The constant λ_x is the production rate of $\mathrm{CD4}^+$ T cells, and $d_x x$ denotes the death rate. The infectivity $\beta(r)$ which plays an important role in the infection process is modeled as a positive, increasing, and saturating function with respect to the concentration of ROS, r,

$$\beta(r) = b_0 + \frac{r(b_\mathrm{m} - b_0)}{r + r_{1/2}}, \tag{8}$$

where b_0 represents the infection rate in the ROS-absent case, while b_m denotes the maximum infection rate, and $r_{1/2}$ is the ROS concentration at half maximum. Then, $(1-\epsilon)\beta(r)xy$ represents the rate at which uninfected cells become infected, where $\epsilon \in (0,1)$ is the effectiveness of drug therapy, and d_y is the per capita death rate of infected $\mathrm{CD4}^+$ T cells. The ROS are produced naturally at rate λ_r, and by the infected cells at rate $k\,y$; but decay at rate $d_r\,r$ and are eliminated by interaction with antioxidants at rate mar. Antioxidants are introduced into the model via natural dietary intake at a constant rate λ_a and through antioxidant supplementation at rate α, and are eliminated from the system by natural decay at rate $d_a a$ and by reacting with the ROS

at rate par, where p is much smaller than m. All the parameters in (7) and (8) are positive, and their typical values have been chosen with careful reference to clinical studies, as given in van Gaalen and Wahl (2009), Zhang et al. (2014a, b).

We apply the quasi-steady state assumption (Flach and Schnell 2006; Korobeinikov et al. 2005; Boie et al. 2016) to reduce the four-dimensional system (7) to a two-dimensional system. To achieve this, we note from Zhang et al. (2013, 2014a, b) that parameters $\lambda_a, \alpha, \lambda_r$ and k typically have much bigger values than other parameters in the corresponding equations. Thus, a is obtained by solving the equation $\frac{da}{dt} = 0$ as

$$a = \frac{\lambda_a + \alpha}{d_a + p\,r} > 0, \quad (r \geq 0);$$

and analogously r from the equation $\frac{\mathrm{d}r}{\mathrm{d}t} = 0$ as

$$\begin{aligned} r &= \frac{-\big[d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r + ky)\big] + \sqrt{[d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r+ky)]^2 + 4pd_a d_r(\lambda_r + ky)}}{2\,p\,d_r} \\ &= \frac{2d_a\,(\lambda_r + ky)}{d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r+ky) + \sqrt{[d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r+ky)]^2 + 4pd_a d_r(\lambda_r+ky)}} > 0, \quad (y \geq 0). \end{aligned} \tag{9}$$

Then, the above solutions a and r are substituted into the first two equations in system (7) to obtain

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lambda_x - d_x x - \beta(y)xy, \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = \beta(y)xy - d_y y, \tag{10}$$

where $\beta(y)$ is derived from $(1-\varepsilon)\beta(r)$ with the use of (8) and (9), given by

$$\beta(y) = (1-\varepsilon)\Big[b_0 + \frac{2d_a(b_\mathrm{m} - b_0)(\lambda_r + ky)}{2d_a(\lambda_r + ky) + \widetilde{r}_{1/2}(y)}\Big], \tag{11}$$

in which $\widetilde{r}_{1/2}(y)$ is defined as

$$\begin{aligned} \widetilde{r}_{1/2}(y) = r_{1/2}\big\{&d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r + ky) \\ &+ \sqrt{[d_a d_r + m(\alpha+\lambda_a) - p(\lambda_r+ky)]^2 + 4pd_a d_r(\lambda_r+ky)}\big\} > 0, \quad (y \geq 0). \end{aligned}$$

To reveal the complex disease dynamics, we focus on bifurcation analysis (i.e., on the asymptotic behavior of the system when the system reaches its steady state). Noting that the value of y is much smaller than that of x, see Zhang et al. (2013, 2014a, b), we may further simplify the two-dimensional system (10) by taking the average value of $\widetilde{r}_{1/2}(y)$, given by

$$\begin{aligned} \bar{r}_{1/2} = \lim_{h\to\infty} \tfrac{1}{h} \int_0^h \widetilde{r}_{1/2}(y)\,dy = r_{1/2} \lim_{h\to\infty} \tfrac{1}{h}\big\{\big[d_a d_r - p\lambda_r + m(\alpha+\lambda_a)\big]h - \tfrac{1}{2}\,pkh^2 \\ + \frac{1}{2}pk\Big[\Big(h + \frac{d_a d_r + p\lambda_r - m(\alpha+\lambda_a)}{pk}\Big)^2 \end{aligned}$$

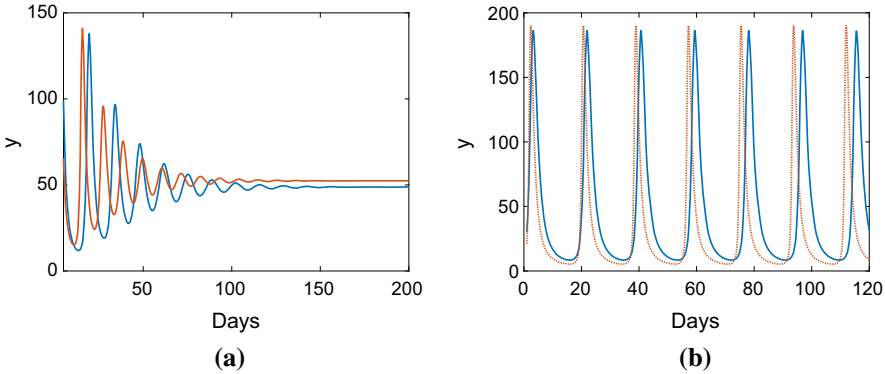

Fig. 1 Simulation of system (10) showing the comparison on y using the original $\widetilde{r}_{1/2}(y)$ (red color curve) and $\bar{r}_{1/2}$ (blue color curve), with the parameter values used in Zhang et al. (2013, 2014a, b), Yu et al. (2016) (Color figure online)

$$+\frac{4d_a d_r m(\alpha+\lambda_a)}{p^2k^2}\ln\left(2\left(h+\frac{d_a d_r+p\lambda_r-m(\alpha+\lambda_a)}{pk}\right)\right)\Big]\Big\}$$
$$=2d_a d_r r_{1/2},$$

and obtain the new infectivity function,

$$\beta(y)=b+\frac{ay}{y+c}, \tag{12}$$

where

$$a=\frac{(1-\epsilon)d_r r_{1/2}(b_{\mathrm{m}}-b_0)}{\lambda_r+d_r r_{1/2}},\quad b=\frac{(1-\epsilon)(\lambda_r b_{\mathrm{m}}+d_r r_{1/2}b_0)}{\lambda_r+d_r r_{1/2}},\quad c=\frac{\lambda_r+d_r r_{1/2}}{k}, \tag{13}$$

all of them are positive since $\epsilon\in(0,1)$ and $b_{\mathrm{m}}>b_0$.

The verification of the averaging step is demonstrated by using the original function $\widetilde{r}_{1/2}(y)$ and the averaged function $\bar{r}_{1/2}$ to simulate system (10) with typical values in Zhang et al. (2013, 2014a, b), Yu et al. (2016). The simulation comparison in Fig. 1a, b indeed shows a very good agreement, which implies that the average is reliable and yields very good approximation to the original system.

Finally, we introduce the scaling: $x=\frac{\lambda_x}{d_y}X$, $y=\frac{\lambda_x}{d_y}Y$, $t=\frac{1}{d_y}\tau$, and the new defined parameters:

$$A=\frac{a\lambda_x}{d_y^2},\quad B=\frac{\lambda_x b}{d_y^2},\quad C=\frac{cd_y}{\lambda_x},\quad D=\frac{d_x}{d_y}, \tag{14}$$

to transform the model (10) into the dimensionless system (5). Thus, with appropriate parameter values, the unified model (5) represents either an SIR epidemiological model (susceptible and infected individuals) with parameters given in (6) or an in-host

infection (susceptible and infected cells) model with parameters given in (14). It is noted that the key difference between system (5) and the class of models studied in Korobeinikov et al. (2005) is that the incidence function in system (5), $(\frac{B+AY}{Y+C})XY$, is convex, while that considered in Korobeinikov et al. (2005) is concave.

Recently, the model (5) was considered by Yu et al. (2016) to demonstrate various dynamics. However, a large part of the analysis, in particular on Hopf bifurcation and Bogdanov–Takens bifurcation is restricted to a two-dimensional parameter space with parameters B and D fixed. Moreover, some part of previous work in (Yu et al. 2016) are heavily dependent upon simulations. Therefore, in this contribution, we shall explore dynamical behaviors of the model (5) in the full four-dimensional parameter space, particularly for the analysis on Hopf and generalized Hopf bifurcations, and BT bifurcation. In this paper, we will first give an analysis on dynamics and bifurcation of system (5), which is different from that given in (Yu et al. 2016), and we then particularly focus on more generic dynamical study of the system in the four-dimensional (A, B, C, D) parameter space. We will focus on the properties of multistable states, limit cycles and homoclinic loops, showing complex behaviors in diseases. In particular, we use theory of Hopf and generalized Hopf bifurcations to prove the existence of multiple limit cycles, giving rise to a different type of bistable or tristable states, with coexistence of stable equilibria and periodic motion. Further, we will carry out the analysis on the BT bifurcation to show homoclinic loop bifurcation, which leads to a new mechanism of generating disease recurrence, that is, cycles of remission and relapse such as the viral blips observed in HIV infection.

The rest of the paper is organized as follows. In the next section, we consider basic solution properties of system (5), and study stability of and bifurcations from the equilibrium solutions of this system. Hopf and generalized Hopf bifurcations will be studied in detail in Sect. 3. In Sect. 4, BT bifurcation and homoclinic bifurcation are investigated, giving rise to a new scenario/mechanism for generating recurrence. Simulations are given in Sects. 3 and 4 for various cases to confirm analytical predictions. Finally, conclusions are drawn in Sect. 5.

2 Dynamics and Bifurcation of System (5)

In this section, we consider general solution properties of system (5), and study stability and bifurcations of its equilibrium solutions. For convenience, define the parameter space as

$$\gamma = (A, B, C, D) \in R_+^4, \tag{15}$$

where R_+^4 means that all the four parameters take positive real values. Further, define the trapping region,

$$\Omega = \left\{(X, Y) | X \geq 0,\ Y \geq 0,\ X + Y \leq \max\left\{1, \tfrac{1}{D}\right\} + \varepsilon\right\}, \tag{16}$$

where $0 < \varepsilon \ll 1$. Then it can be shown (Yu et al. 2016) that for any set of parameter values belonging to γ, all solutions of system (5) are nonnegative provided initial

conditions are taken nonnegative. Moreover, all solutions are attracted into Ω and so bounded.

System (5) has two equilibrium solutions:

$$\begin{aligned} &\text{disease-free equilibrium}: \mathrm{E}_0 = \left(\tfrac{1}{D},\ 0\right), \\ &\text{endemic equilibrium}: \quad \mathrm{E}_1 = \left(X_1, Y_1\right) = \left(X_1,\ 1 - DX_1\right), \end{aligned} \tag{17}$$

where X_1 is determined from the equation,

$$F_1(X) = D(A+B)X^2 - (A+B+D+BC)X + C + 1 = 0. \tag{18}$$

The solutions of $F_1(X) = 0$ can be written as

$$X_\pm = \tfrac{(A+B+D+BC) \pm \sqrt{\Delta}}{2D(A+B)}, \tag{19}$$

in which

$$\begin{aligned} \Delta &= (A+B+D+BC)^2 - 4D(1+C)(A+B) \\ &= (A+B-D-BC)^2 - 4C(A+B)(D-B). \end{aligned} \tag{20}$$

Further, define the reproduction number as

$$R_0 = \frac{B}{D}. \tag{21}$$

Then the dynamical behavior of system (5) can be conveniently analyzed according to $R_0 < 1$, $R_0 > 1$ and $R_0 = 1$. Note that E_0 is a boundary equilibrium (on the X-axis), while E_1 is a positive (an interior) equilibrium.

We choose D as the bifurcation parameter (which is fixed in the study Yu et al. (2016)) and treat other parameters A, B and C as control parameters, then consider the solution $X(D)$ determined by the graph $F_1(X) = 0$, as shown in the D-X plane (see Fig. 2). It will be seen later that the relation between the control parameters also plays an important role on bifurcation property and dynamical behavior of the system. We define the saddle-node bifurcation point and transcritical bifurcation point, respectively, as

$$\begin{aligned} &(D_s, X_s) = (D_-,\ X_{s-}) \quad \text{and} \quad (D_t, X_t) = \left(B,\ \tfrac{1}{B}\right), \\ &\text{where} \quad X_{s\pm} = \tfrac{A+B+BC+D_\pm}{2(A+B)D_\pm}, \quad D_\pm = A + B + BC + 2AC \\ &\qquad\qquad \pm 2\sqrt{AC(1+C)(A+B)}. \end{aligned} \tag{22}$$

Note that $\Delta = 0$ at $(D_\pm, X_{s\pm})$.

An elementary proof based on the function F_1 leads to the following result for the property of the solution X_1.

Lemma 2.1 *In the D-X plane, the function $X(D)$ has two horizontal asymptotes: $X = 0$ and $X = \frac{1}{A+B}$; and two vertices at $D = D_{\pm}$. The property of the function $X(D)$ is given in the following table.*

X	D	$X(D)$ monotonically
$\in (-\infty, 0)$	$\in (-\infty, 0)$	$\searrow$
$\in (0, X_{s+})$	$\in (D_+, +\infty)$	$\searrow$
$\in (X_{s+}, \frac{1}{A+B})$	$\in (D_+, +\infty)$	$\nearrow$
$\in (\frac{1}{A+B}, X_s)$	$\in (-\infty, D_s)$	$\nearrow$
$\in (X_s, +\infty)$	$\in (0, D_s)$	$\searrow$

Therefore, the biologically meaningful equilibria (nonnegative solutions) are defined as

$$\begin{aligned} &\mathrm{E}_0: \quad (X_0, Y_0) = \left(\tfrac{1}{D}, 0\right), && \text{for } D > 0, \\ &\mathrm{E}_{1\pm}: (X_1, Y_1) = \left(X_{\pm}, Y_{\pm}\right) = \left(X_{\pm}, 1 - DX_{\pm}\right), && \text{for } 0 \le X_{\pm} \le \tfrac{1}{D}, \end{aligned} \tag{23}$$

where the condition $0 \le X_{\pm} \le \frac{1}{D}$ comes from $Y_1 = 1 - DX_1 \ge 0$. This means that the biological reasonable equilibrium solutions in Fig. 2a–d can only locate in the first quadrant.

We have the following lemma.

Lemma 2.2 *The disease-free equilibrium solution $X_0 = \frac{1}{D}$ is a monotonic function of D. The solution curves X_0 and X_1 have a unique intersection point at $(D, X) = (D_t, X_t) = (B, \frac{1}{B})$. The biologically meaningful endemic equilibrium solution X_1 only exists for $\frac{1}{A+B} < X < \frac{1}{B}$ and below the curve $X = \frac{1}{D}$. Moreover, the equilibrium solution X_1 has a turning point at $(D, X) = (D_s, X_s)$, which is above the curve $X = \frac{1}{D}$ if $A < BC$, below the curve if $A > BC$, and on the curve if $A = BC$.*

Proof It is obvious that the function $X = \frac{1}{D}$ is monotonic with respect to D. It is easy to show that the two curves $X = \frac{1}{D}$ and $F_1 = 0$ have a unique intersection point by solving $F_1(\frac{1}{D}) = 0$, yielding $D = B$ and thus $X = \frac{1}{B}$, which results in a critical point at $(D, X) = (D_t, X_t) = (B, \frac{1}{B})$. The biologically meaningful solution curve X_1 must be below the curve $X = \frac{1}{D}$ since it requires $DX_{\pm} < 1$, i.e., $X_{\pm} < X_0$. It is easy to see from Fig. 2 that $X_1 > \frac{1}{A+B}$ for $D < D_s$. More precisely, a direct computation shows that

$$D < D_s \iff A + B + BC - D > 2\sqrt{4D(1+C)(A+B)} - 2AC > 0,$$

and then

$$\begin{aligned} X_- > \tfrac{1}{A+B} &\iff A + B + D + BC - \sqrt{\Delta} > 2D \\ &\iff A + B + BC - D > \sqrt{\Delta} \iff AC > 0. \end{aligned}$$

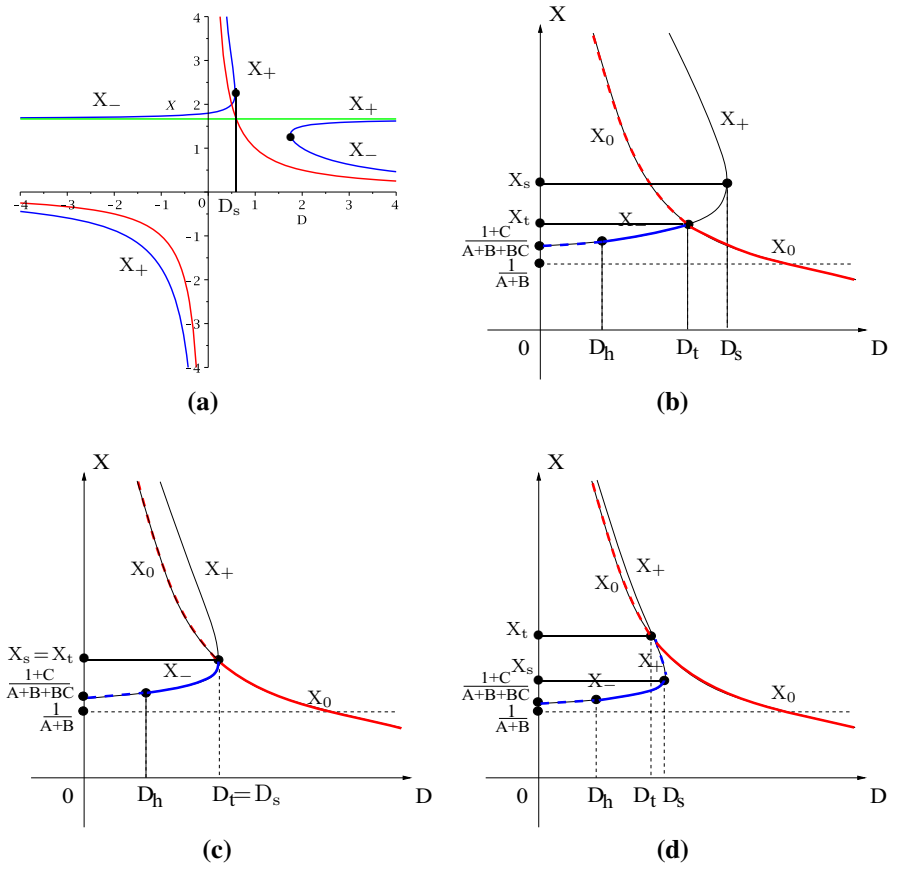

Fig. 2 Bifurcation diagrams for system (5): **a** in the whole D-X plane; **b** in the biologically meaningful part with $A < BC$; **c** in the biologically meaningful part with $A = BC$; and **d** in the biologically meaningful part with $A > BC$, with E_0 in red and E_1 in blue, and stable in solid and unstable in dotted curves, respectively. The green line in **a**, and black dotted lines in **b**–**d** denote the same asymptote $X = \frac{1}{A+B}$ to the biologically meaningful solution $X_\pm$, which satisfies $X_\pm \le X_0$ (Color figure online)

Now, we show that the part of solution X_1 for $D \ge D_+$ is above the curve $X = \frac{1}{D}$, see Fig. 2a. To achieve this, we only need to prove that $X_- > \frac{1}{D}$ for $0 < X < \frac{1}{A+B}$ and $D > D_+$, which is equivalent to

$$X_- > \tfrac{1}{D} \iff BC + D - A - B > \sqrt{(A+B+BC+D)^2 - 4D(1+C)(A+B)}$$
$$\iff (D-B)C > 0,$$

which is true since $D > D_+ > A + B$.

For the last conclusion in this lemma, since the two curves $X_0 = \frac{1}{D}$ and $X_1 = X(D)$ have a unique intersection point at $(D, X) = (D_t, X_t) = (B, \frac{1}{B})$, it suffices to show that $X_s > X_t$ (or $X_s < X_t$) if and only if $A < BC$ (or $A > BC$). A direct calculation yields that

$$X_s - X_t = \frac{A+B+BC+D_s}{2(A+B)D_s} - \frac{1}{B} = \frac{-(A-BC)\,A\,(A+B+BC)}{BD_s\left[A(1+2C)(A+B)+(2A+B)\sqrt{AC(1+C)(A+B)}\right]},$$

which clearly shows that the conclusion holds, and the proof is complete. □

The three different cases for $A < BC$, $A = BC$ and $A > BC$ are depicted in Fig. 2b–d, respectively. These figures are actually bifurcation diagrams for system (5), but only the red curve and the part of the blue curve below the red cure are biologically meaningful. The D_s, D_t and D_h denote the *saddle node, transcritical* and *Hopf* bifurcations, respectively, and $D_t = B$, implying that at the transcritical point $R_0 = 1$. The difference on the conditions of $A < BC$ and $A > BC$ causes a fundamental effect on the dynamics and stability of system (5). In particular, when $A > BC$, the system exhibits bistable phenomena if $R_0 < 1$ (i.e., $D > B$), which may involve two stable equilibrium solutions E_0 and E_{1-}, or stable equilibrium E_0 and stable limit cycles. This can not happen if $A \le BC$. The condition $A = BC$ distinguishes the system into two fundamental different types of bifurcations: *forward bifurcation* when $A \le BC$, and the other *backward bifurcation* when $A > BC$. Backward bifurcation usually exhibits more complex dynamical behaviors such as bistable phenomena (Zhang et al. 2016). Biologically, the threshold value of the contact rate at $A = BC$ means that the interaction between X and Y produces sufficient infection such that Y persists even there exists stable disease-free equilibrium E_0.

Stability analysis on the equilibria of a two-dimensional dynamical system is usually not difficulty. However, if the system contains multiple parameters, then it is not easy to find the explicit stability conditions expressed in terms of the system coefficients. For our purpose of studying stability and bifurcation of the equilibrium E_1, define

$$\begin{aligned}
&I_B = \left(0, \tfrac{1}{4}\right), \quad A_l = \tfrac{4B^2}{1-4B}, \quad A_u = \begin{cases} \frac{B\,(2B+\sqrt{B})}{1-4B}, & \text{if } A \le BC, \\ +\infty, & \text{if } A > BC, \end{cases} \\
&C_{l_2} = \frac{A-\sqrt{A^2(1-4B)-4AB^2}}{2B}, \qquad C_{u_1} = \frac{C_{l_2}}{B} - \left(\tfrac{A}{B}+1\right), \quad C_{u_2} = \min\left\{\tfrac{A}{B}, C_{u_1}\right\}, \\
&C_{l_1} = \frac{2B(A+B)C_{u_1}}{2B(1+A+B)+A-\sqrt{A^2(1-4B)-4AB^2}}, \\
&D_h = \frac{B\left[2(A+B)(A+B+BC)-C\left(A+\sqrt{A^2(1-4B)-4AB^2}\right)\right]}{(A+B)\left[A+2B-\sqrt{A^2(1-4B)-4AB^2}\right]}, \\
&v_0 = -\frac{1}{4(A+B)D^2/(\sqrt{\Delta}}\Big\{(A+B)\big[(A+B)(BC-D^2)+2BC^2(B-D)-BCD \\
&\qquad + D^2(D-BC-2AC)\big] + B^2C^2(D+BC) \\
&\qquad -\big[BC(A+B+BC)+(A+B)D^2\big]\sqrt{\Delta}\Big\} \\
&\mathrm{H}_1 = \left\{B \in I_B,\ A \in (A_l, A_u),\ C \in (C_{l_1}, C_{u_1})\right\}, \\
&\mathrm{H}_2 = \left\{B \in I_B,\ A \in (A_l, +\infty),\ C \in (C_{l_1}, C_{u_2})\right\}, \\
&\mathrm{H}_3 = \left\{B \in I_B,\ A \in (A_l, +\infty),\ C \in (C_{l_1}, C_{l_2})\right\},
\end{aligned} \tag{24}$$

where D_h defines a Hopf critical point when D is treated as a bifurcation parameter, and v_0 (which is borrowed from the notation used in normal forms) denotes its transversality condition with respect to D.

Then, we have the following result.

Theorem 2.3 *For system (5), the stability of equilibrium solutions and bifurcation properties are given in the following table.*

Equilibrium		$A \le BC$	$A > BC$
E_0		GAS for $D \ge D_t$	GAS for $D \ge D_s$, LAS for $D_t < D < D_s$
E_{1+}		Saddle	Saddle
	Hopf at $D = D_h$	$H_1 \bigcup D \in (D_h, D_t)$	$H_2 \bigcup D \in (D_h, D_s)$
E_{1-} LAS	Otherwise, No Hopf	$D \in (0, D_t)$	$D \in (0, D_s)$
Bistable* States		No	$H_3 \bigcup D_h \in (D_t, D_s)$ (E_0, E_1) for $D \in (D_h, D_s)$ (E_0, LC) for $D \in (D_t, D_h)$

where GAS, LAS and LC denote the Globally asymptotically stable, Locally asymptotically stable and Limit cycle, respectively. Moreover, no limit cycles can bifurcate from homoclinic orbits if $A \le BC$

Note in Theorem 2.3 that for the bistable states involving limit cycles, the results on the stability of limit cycles will be given in next section (see Theorems 3.2, 3.3, 3.4 and 3.5).

Proof The disease-free equilibrium, $E_0 = (\frac{1}{D}, 0)$, is a boundary equilibrium, located on the X-axis. Evaluating the Jacobian matrix of system (5) at the E_0 yields two eigenvalues, $\xi_1 = -D$ and $\xi_2 = R_0 - 1$, showing that E_0 is asymptotically stable (a stable node) if $R_0 < 1$ (i.e., $D > D_t$) and unstable (a saddle) if $R_0 > 1$ (i.e., $D < D_t$). To prove the global stability, first note that all trajectories are attracted into the trapping region Ω. Secondly, if $A \le BC$, then when $D \ge B$ (i.e., $R_0 \le 1$), there exists only one stable equilibrium E_0 on the boundary of Ω, thus all trajectories converge to the stable equilibrium E_0. If $A > BC$, then when $D > D_s$ (i.e., $R_0 < \frac{B}{D_s}$), there again exists only one stable equilibrium E_0 on the boundary of Ω, thus all trajectories converge to the stable equilibrium E_0.

Next, for stability of the E_1, we evaluate the Jacobian matrix of system (5) at the endemic equilibrium E_1 to obtain

$$J_1 = J(E_1) = \begin{bmatrix} -\frac{1}{X_1} & -1 - \frac{ACX_1Y_1}{(Y_1+C)^2} \\ \frac{1}{X_1} - D & \frac{ACX_1Y_1}{(Y_1+C)^2} \end{bmatrix}, \tag{25}$$

which, together with (19) and (20), in turn results in

$$\det(J_1) = \frac{-(1-DX_1)}{2D(A+B)(1-DX_1+C)X_1} \sqrt{\Delta}\big[\sqrt{\Delta} \pm (A + B + D + BC)\big]. \tag{26}$$

If we first solve C from the equation $F_1 = 0$ and then obtain the trace of J_1 as

$$\begin{aligned} C &= \tfrac{1-DX_1}{1-BX_1}\big[(A+B)X_1 - 1\big], \\ \mathrm{Tr}(J_1) &= -\tfrac{1}{AX_1}\big\{\big[(A+B)X_1 - 1\big](BX_1 - 1) + A\big\}. \end{aligned} \tag{27}$$

Note that at the critical point $R_0 = 1$ (i.e., $D = B$), $X_1 = X_0 = \frac{1}{B}$.

We first consider the equilibrium $E_{1+} : (X_+, Y_+)$. Since the term in the square bracket of $\det(J_1)$ (taking the positive sign) is positive, and the biologically meaningful solution requires $0 < DX_+ < 1$, it is obvious that $\det(J_1) < 0$, implying that E_{1+} is a saddle.

Now, for the equilibrium E_{1-}, we only need to consider the biologically meaningful solution $E_{1-} : (X_-, Y_-)$ in the first quadrant with $\frac{1}{A+B} < X_- < \frac{1}{D}$ (see Fig. 2b–d). At the critical point $D = D_t = B$, $X_- = X_0 = \frac{1}{B}$, with a zero eigenvalue at this point, indicating that $R_0 = 1$ is a transcritical bifurcation point, though E_{1-} does not biologically exist for $R_0 < 1$. For stability of the E_{1-}, first it is seen from (26) that $\det(J_1) > 0$ since $DX_- < 1$ and the term in the square bracket in (26) with the negative sign is negative, and thus the stability is determined by the sign of $\mathrm{Tr}(J_1)$. Since $X_- > \frac{1}{A+B}$, $DX_- < 1$, and $X_- < X_t = \frac{1}{B}$, yielding $BX_- < 1$, $C > 0$ is guaranteed. However, the term in the bracket of the trace $\mathrm{Tr}(J_1)$ given in (27) can be positive or negative for $D \in (0, D_t)$ if $A \le BC$ (see Fig. 2b, c) or for $D \in (0, D_s)$ if $A > BC$ (see Fig. 2d and note that when $A > BC$, E_{1+} is a saddle for $D \in (D_t, D_s)$). This indicates that the only possible bifurcation from the E_{1-} is Hopf bifurcation, arising from a critical point at which $\mathrm{Tr}(J_1) = 0$. In order to determine the Hopf critical point, we solve the two equations in (27) for D and X_- to obtain

$$D_h = \frac{(A+B+BC)X_- - (1+C)}{X_-[(A+B)X_- - 1]}, \tag{28}$$

where X_- is given by

$$X_- = \frac{A+2B-\sqrt{A^2(1-4B)-4AB^2}}{2(A+B)B}, \tag{29}$$

which is positive when $B \in I_B = (0, \frac{1}{4})$ and $A > A_l = \frac{4B^2}{1-4B}$. Substituting (29) into (28) results in the expression of D_h given in (24). To find the transversality condition v_0, we solve $F_1 = 0$ for X_1, which is substituted into $\mathrm{Tr}(J_1)$, and then take the derivative of the resulting trace to obtain

$$\frac{1}{2}\frac{\partial \mathrm{Tr}(J_1)}{\partial D}\bigg|_{E_{1-}} = \frac{1}{2}\frac{\partial}{\partial D}\left[-\frac{1}{X_1} + \frac{ACX_1(1-DX_1)}{(1+C-DX_1)^2}\right] = v_0,$$

where v_0 is given in (24). In general, $v_0 \neq 0$, implying that Hopf bifurcation occurs at the critical point $D = D_h$.

It is clear that no Hopf bifurcation occurs for $D_h \le 0$. To have $D_h > 0$, one more condition is required from $X_- > \frac{1+C}{A+B+BC}$, which in turn yields $0 < C < C_{u_1}$. So when the above conditions are satisfied, we have $D_h > 0$. However, note that only if $D_h < D_t$ (i.e., $X_- < X_t$) when $A \le BC$, or $D_h < D_s$ (i.e., $X_- < X_s$) when $A > BC$, then $D_h > 0$ defines a true Hopf critical point. Hence, we need to consider two cases: $A \le BC$ and $A > BC$.

(a) $A \le BC$. For this case, the disease-free equilibrium E_0 is globally asymptotically stable if $R_0 \le 1$ (for which E_{1-} does not biologically exist), and unstable if $R_0 > 1$ (i.e., $D < D_t$) for which E_{1-} emerges to exist. To have a Hopf bifurcation from

E_{1-}, it needs $X_- < X_t = \frac{1}{B}$, which is easy to be proved by using (27) to obtain $BX_- - 1 < 0$ because of $C > 0$. On the other hand, the condition $A \leq BC$, or $C \geq \frac{A}{B}$, together with $C < C_{u_1}$ yields $C \in [\frac{A}{B}, C_{u_1})$. Further, it requires that $C_{u_1} > \frac{A}{B}$, yielding $A < A_u = \frac{B\left(2B+\sqrt{B}\right)}{1-4B}$ (for $B \in I_B$). Hence, for this case, Hopf bifurcation appears from the equilibrium E_{1-} at the critical point $D = D_h$ if the parameters satisfy $B \in I_B$, $A \in (A_l, A_u)$, $C \in (0, C_{u_1})$ and $D_h \in (0, D_t)$.

(b) $A > BC$. For this case, the turning point $(D, X) = (D_s, X_s)$ is below the curve $X = \frac{1}{D}$ and so X_- must be below the curve $X = \frac{1}{D}$ as well. To have a Hopf bifurcation, the condition, $\frac{1+C}{A+B+BC} < X_- < X_s$, needs to be satisfied for $D_h > 0$. The condition for $\frac{1+C}{A+B+BC} < X_-$ is obtained in part (a) as $C < C_{u_1}$, while the condition for $X_- < X_s$ yields $C > C_{l_1}$. In addition, to determine if Hopf bifurcation can occur for $D_h \in (D_t, D_s)$, which may lead to bistable phenomenon involving the stable equilibrium E_0 and a stable limit cycle, we find another critical point at the second intersection point of X_- with the vertical line $D = D_t$, where $X_- = \frac{1+C}{A+B}$. Then, $\frac{1+C}{A+B} < X_s$ yields the critical point C_{l_2}. Finally, we need to ensure that $A > BC$ for this case, which is guaranteed by simply defining $C_{u_2} = \min\left\{\frac{A}{B}, C_{u_1}\right\}$. It is easy to show that $C_{l_2} < \frac{A}{B}$ and $C_{l_2} < C_{u_1}$, implying that $C_{l_2} < C_{u_2}$. Moreover, it can be shown that $C_{l_1} < C_{l_2}$, and thus we have $C_{l_1} < C_{l_2} < C_{u_2}$. Therefore, for this case, Hopf bifurcation occurs from E_{1-} at the critical point $D = D_h$ for the parameters satisfying $B \in I_B$, $A > A_l$ and $C \in (C_{l_1}, C_{u_2})$; and E_{1-} is asymptotically stable for $D \in (D_h, D_s)$. In particular, when $C \in (C_{l_1}, C_{l_2})$ and $D_h \in (D_t, D_s)$, bistable phenomena happen, with the coexistence of stable equilibria E_0 and E_{1-} for $D \in (D_h, D_s)$, and the coexistence of stable equilibrium E_0 and stable limit cycles for $D \in (D_t, D_h)$. While when $C \in (C_{l_2}, C_u)$ and $D_h \in (0, D_t)$, only stable equilibria E_0 and E_{1-} coexist for $D \in (D_t, D_s)$.

Finally, to prove that no limit cycles can bifurcate from homoclinic orbits if $A \leq BC$, note that E_0 becomes a saddle when $A \leq BC$. If a limit cycle bifurcates from a homoclinic orbit, there must exist a homoclinic loop which connects E_0, i.e., leaving the E_0 along the unstable manifold and return to it along the stable manifold. But we know that stable manifold of the E_0 is the X-axis which is an invariant manifold of the system. Thus, connecting the unstable manifold to the stable manifold must violate the uniqueness of the solutions of the system, and so it is not possible to have limit cycles bifurcating from homoclinic orbits when $A \leq BC$.

The proof for Theorem 2.3 is complete. □

3 Hopf and Generalized Hopf Bifurcations

In this section, we consider bifurcation of limit cycles due to Hopf and generalized Hopf bifurcations. In particular, we shall consider the existence of multiple limit cycles due to generalized Hopf bifurcation and prove that the maximal number of limit cycles is two for the whole four-dimensional γ parameter space. All Hopf bifurcations occur from the equilibrium E_{1-}, which can be classified into three categories according to

$R_0 = 1$, $R_0 < 1$ and $R_0 > 1$, namely $D = B$, $D > B$ and $D < B$, respectively. Since the treatment for the cases $R_0 < 1$ and $R_0 > 1$ are similar, we will combine the latter two cases as one: $R_0 \neq 1$. Also, the treatment is same for $A \leq BC$ and $A > BC$, we will not distinguish the two cases.

3.1 Case $R_0 = 1$ ($D = B$)

When $R_0 = 1$ (or $D = B$), system (5) is in a critical situation, i.e., the two equilibria E_0 and E_1 exchange their stability at the critical point and complex dynamics of the system happens on the center manifold which is characterized by a single zero eigenvalue (the other eigenvalue is $-B$).

Since in general we treat D as a bifurcation parameter and define $D_t = B$, to avoid confusing in this subsection we let $D = B$ and choose B as a bifurcation parameter. Then, when $D = B$ (i.e., at $R_0 = 1$), the equilibrium solutions become

$$\mathrm{E}_{1+} = \mathrm{E}_0 = \left(\tfrac{1}{B},\, 0\right) \quad \text{and} \quad \mathrm{E}_{1-} = \left(\tfrac{1+C}{A+B},\, \tfrac{A-BC}{A+B}\right), \quad \left(B \leq \tfrac{A}{C}\right), \tag{30}$$

where the existence condition $A \geq BC$ also guarantees $Y_{1-} \geq 0$ and $X_{1-} \leq \frac{1}{B}$. Note that when $A = BC$, the equilibrium $\mathrm{E}_{1-} \equiv \mathrm{E}_1$ also coincides with E_0, at which a bifurcation occurs. For this case, we define a new reproduction number as

$$B_t = \frac{A}{C}. \tag{31}$$

Then, E_1 emerges to exist for $B < B_t$, which requires $A > BC$. In order to study the stability of E_0, we need to find the center manifold at the critical point.

3.1.1 Center Manifold Reduction and Stability Analysis

We first use the center manifold theory to find the differential equation describing dynamics on the center manifold, and then discuss the stability of E_0. To achieve this, we first introduce an affine transformation, given by

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{1}{B} \\ 0 \end{pmatrix} + \begin{bmatrix} 1 & 1 \\ -B & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{32}$$

into (5) with $D = B$ to obtain a system,

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}\tau} &= Bu_1u_2 - \frac{1}{C}(A - BC)\,u_1^2 - \frac{AB}{C}\,u_1^2u_2 + \cdots, \\ \frac{\mathrm{d}u_2}{\mathrm{d}\tau} &= -Bu_2 + B(B-1)u_1u_2 + \frac{1}{C}(A - BC)(1-B)u_1^2 + \frac{AB}{C}(1-B)u_1^2u_2 + \cdots, \end{aligned} \tag{33}$$

whose linear part is in Jordan canonical form with eigenvalues 0 and $-B$. To find the center manifold, let $u_2 = h(u_1) = \eta u_1^2 + O(u_1^3)$ and then use (33) to find $\eta =$

$\frac{(1-B)(A-BC)}{BC}$. Therefore, the center manifold up to second order is given by

$$W^C = \{(u_1, u_2) \mid u_2 = \tfrac{(1-B)(A-BC)}{BC}\, u_1^2 + O(u_1^3)\},$$

and the differential equation describing the dynamics on the center manifold is given by

$$\frac{\mathrm{d}u_1}{\mathrm{d}\tau} = \tfrac{1}{C}(A - BC)\big[-u_1^2 + (1-B)\,u_1^3 + O(u_1^4)\big]. \tag{34}$$

Since when $D = B$, the system still has three free parameters A, B and C, we will give a brief analysis on the dynamics and bifurcation of the system when $R_0 = 1$. Now system has only two equilibria E_0 and E_1, given in (30). Note that for this critical case, $Y = 1 - BX$ still holds and the equilibrium solution X_1 for E_1, given by $X_1 = \frac{1+C}{A+B}$, is monotonically decreasing as B increases, indicating that no backward bifurcation can occur, and so bistable equilibria are not possible. Define

$$B_h = \tfrac{A(C-A)}{A+C^2}. \tag{35}$$

Then, we have the following theorem for this critical case.

Theorem 3.1 *For system* (5)*, when* $R_0 = 1$ *(i.e.,* $D = B$*), the disease-free equilibrium* E_0 *is globally asymptotically stable for* $B > B_t$ *and unstable for* $B < B_t$ *for which* E_1 *emerges to exist.* E_1 *is asymptotically stable for* $B < B_t$ *if* $A \geq C$*; and Hopf bifurcation occurs from* E_1 *at* $B = B_h$ *if* $A < C$ *for which* E_1 *is asymptotically stable for* $B_h < B < B_t$ *and unstable for* $B < B_h$*.*

Proof For the stability of the equilibrium E_0, linearization does not work since one eigenvalue is zero. However, we can apply Eq. (34) to study its stability because (u_1, u_2) represents small perturbation from this equilibrium. Since $Y = -Bu_1$, Y decreases from a positive initial point if u_1 is increasing, implying that Y is stable and converges to zero. It follows from (34) that u_1 is increasing if $A < BC$, or $B > B_t$. Recall that the X-axis is invariant and this trajectory converges to E_0 along the X-axis. Therefore, E_0 is a degenerate stable node when $B > B_t$, and a degenerate saddle when $B < B_t$. Note that E_1 does not exist for $B > B_t$, and there exists only one stable equilibrium E_0 on the boundary of the trapping region Ω, and thus all trajectories converge to E_0, implying that E_0 is globally asymptotically stable for $B > B_t$.

To find the stability of the E_1, evaluating the Jacobian of (5) when $D = B$ at E_1 yields the trace and determinant as

$$\mathrm{Tr}(J_1) = \tfrac{(C^2+A)(B_h-B)}{A(1+C)} \quad \text{and} \quad \det(J_1) = \tfrac{C^2(B-B_h)^2}{A(1+C)} > 0.$$

Hence, the equilibrium E_1 is asymptotically stable if $\max\{0, B_h\} < B < B_t$. Hopf bifurcation occurs from E_1 at $B = B_h$, if $B_h > 0$, and E_1 is unstable for $B \in (0, B_h)$. □

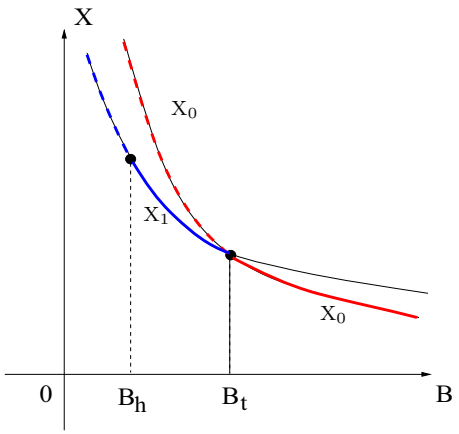

Fig. 3 Bifurcation diagram for system (5) when $R_0 = 1$ ($D = B$)

Further, transforming (33) with (34) back to the original coordinates (X, Y), we obtain the equations describing the dynamics on the center manifold, given by

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}\tau} &= 1 - BX - \tfrac{1}{B}\,Y + (1-B)\,XY - \tfrac{A-BC}{CB^2}\,Y^2 + \cdots, \\ \frac{\mathrm{d}Y}{\mathrm{d}\tau} &= \tfrac{A-BC}{CB^2}\big[B + (1-B)Y\big]\,Y^2 + \cdots, \end{aligned} \tag{36}$$

which clearly indicates that for small positive initial conditions, Y converges to zero if $A < BC$, i.e., $B > B_t$. Moreover, the non-trivial equilibrium solution of (36), given by $(X, Y) = \big(1 - \frac{A}{C(B-1)^2}, \frac{B}{B-1}\big)$, shows that the graph $(1-X)(B-1)^2 = \frac{A}{C}$ in the B-X plane does not have a turning point, again indicating that no backward bifurcation can occur. The bifurcation diagram for this special case is shown in Fig. 3.

3.1.2 Hopf Bifurcation

For convenience, define a subset in the γ-parameter space,

$$\gamma_1 = \big\{\gamma \mid C = (\rho+1)A,\ 0 < B < \tfrac{1}{\rho+1},\ \rho > 0\big\}, \tag{37}$$

and let

$$\begin{aligned} \bar{A} &= \frac{\sqrt{\rho^4+4\rho^2(\rho+1)^2+4(\rho+1)} - \rho^2}{2(\rho+1)\big[\rho^2(\rho+1)+1\big]} \\ &= \frac{2}{\sqrt{\rho^4+4\rho^2(\rho+1)^2+4(\rho+1)} + \rho^2} \in (0,1), \quad (\rho > 0). \end{aligned} \tag{38}$$

Then, we have the following result.

Theorem 3.2 *For system (5), when $R_0 = 1$, Hopf bifurcation occurs from the equilibrium E_1 at the critical point $B = B_h$, if the parameter values are taken from the set γ_1. The bifurcation is supercritical (or subcritical) if $A > \bar{A}$ (or $A < \bar{A}$), and a family*

of bifurcating limit cycles is stable (unstable), enclosing an unstable (a stable) focus at E_1.

Proof First, the necessary conditions for having a Hopf bifurcation from E_1 are obtained from the Hopf critical condition $B_h > 0$ and $B < B_t$ as $C > A > BC$, yielding $0 < B < 1$. Further, let $C = (\rho + 1)A$ ($\rho > 0$), under which the above two conditions become $0 < B < \frac{1}{\rho+1}$. In order to apply normal form theory to calculate the first-order focus value (or the first Lyapunov constant), we first multiply the equations in (5) by $Y + C > 0$ and then introduce the affine transformation, given by

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{1+C}{A+B} \\ \frac{A-BC}{A+B} \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ \frac{-CA}{C^2+(1+C)A} & \frac{-C\omega_c}{C^2+(1+C)A} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{39}$$

where $\omega_c = \frac{A}{C}\sqrt{(1+C)A}$, into the resulting equations to yield a system to be expanded around $(u_1, u_2) = (0, 0)$ up to third-order terms. Next, we apply the Maple program for computing the normal forms associated with Hopf and generalized Hopf bifurcations (Yu 1998) to the new system to obtain the normal form in polar coordinates up to third-order terms as follows:

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = r\left[v_0\,\mu + v_1\,r^2 + O(r^4)\right], \quad \frac{\mathrm{d}\theta}{\mathrm{d}\tau} = \omega_c + t_0\,\mu + t_1\,r^2 + O(r^4), \tag{40}$$

where $\mu = B_h - B$, is a perturbation parameter, and v_0 and v_1 are the zero-order and the first-order focus values. The first equation of (40) can be used to perform bifurcation analysis and the sign of v_1 determines whether the Hopf bifurcation is supercritical or subcritical. The values v_0 and t_0 can be found from a linear analysis, while v_1 and t_1 are obtained by applying the Maple program. The calculation shows that

$$v_0 = -\frac{(A+C^2)^2}{2AC(1+C)}, \qquad t_0 = -\frac{(A+C^2)^2}{C^2\sqrt{(1+C)A}}, \tag{41}$$

and the output from the Maple program gives v_1 and t_1 as

$$v_1 = \frac{-A^3C^2}{8(A+AC+C^2)(C^2+A)^2}\left[(\rho+1)(\rho^3+\rho^2+1)A^2 + \rho^2 A - 1\right] \tag{42}$$

and

$$\begin{aligned} t_1 = \frac{-A^3C\sqrt{(1+C)A}}{24(1+C)(A+AC+C^2)(C^2+A)^2}\big[&(\rho+1)^2(\rho^4+12\rho^3+20\rho^2+13\rho+5)A^3 \\ &+(\rho+1)(10\rho^3+44\rho^2+45\rho+13)A^2 + (25\rho^2+45\rho+21)A + 13\big]. \end{aligned}$$

It is easy to verify that when $A > \bar{A}$ (or $A < \bar{A}$), $v_1 < 0$ (or $v_1 > 0$), yielding a supercritical (subcritical) Hopf bifurcation and so the bifurcating limit cycles are stable (unstable). □

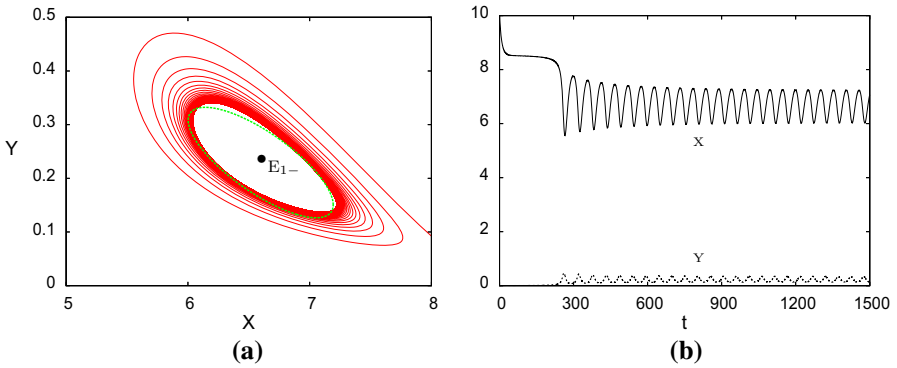

Fig. 4 A stable limit cycle of system (5) when $R_0 = 1$ for $A = \frac{6937}{65600}$, $C = \frac{96}{205}$, $B = D = \frac{73}{625}$: **a** phase portrait with the red and green color curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 0.5954$), respectively; and **b** the time history of the stable limit cycle (Color figure online)

To illustrate the theoretical result, we present a simulation using the following parameter values:

$$A = \frac{6937}{65600}, \quad C = \frac{96}{205}, \quad D = B = \frac{73}{625} = 0.1168, \tag{43}$$

yielding $B_h \approx 0.117947$ at which $\mathrm{Tr}(\mathrm{J}_1) = 0$ (or $v_0\,\mu = 0$), as well as $v_0 \approx -0.726539$ and $v_1 \approx -0.002351$, indicating that the Hopf bifurcation is supercritical and the bifurcating limit cycle is stable. To estimate the amplitude of the limit cycle, we note that $\mu = B_h - B = 0.001147$, which implies that B is decreasing to pass the Hopf critical point B_h as $v_0 < 0$ indicates. Thus, it is easy to use the truncated normal form $v_0\mu + v_1 r^2$ to obtain the estimate of the amplitude of the bifurcating limit cycle as $r \approx 0.5954$. The simulation of this stable limit cycle is shown in Fig. 4a as the red curve, and the green curve in the same figure denotes the first-order approximation of the analytical prediction, which is obtained by using the transformation (39), the normal form (40) together with (41) as well as the following additional transformation,

$$u_1 = r\cos(\omega_c t), \quad u_2 = -r\sin(\omega_c t).$$

The simulation shows an excellent agreement with the analytical prediction.

3.1.3 Multiple Limit Cycles Bifurcation

In the previous subsection, we have shown that system (5) always exhibits limit cycles due to Hopf bifurcation. But it is limited to a single limit cycle for a given set of parameter values. Now, we want to ask: for what feasible parameter values in the γ_1 set, we can obtain maximal number of limit cycles bifurcating from the Hopf critical point near the equilibrium E_1? Here, "feasible" means that the values of the chosen parameters must be positive and the equilibrium solution for the chosen parameter values must be also positive. Bifurcation of multiple limit cycles here is related to the

so-called generalized Hopf bifurcation. The condition for a generalized Hopf bifurcation is that at least the first-order focus value vanishes, i.e., $v_1 = 0$. When more focus values vanish, the generalized Hopf bifurcation is more degenerate and more limit cycles can bifurcate from the critical point. More precisely, we rewrite the first equation of (40) as

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = r\left[v_0\,\mu + v_1\,r^2 + \cdots + v_{k-1}\,r^{2k-2} + v_k\,r^{2k} + O(r^{2k+2})\right],$$

where all the focus values v_i's are expressed in terms of the system parameters. If we can find the conditions on k parameters, say, $\mu = (\mu_1,\ \mu_2,\ \ldots,\ \mu_k)$, such that $v_0 = v_1 = \cdots = v_{k-1} = 0$, but $v_k \neq 0$, at the critical point defined by $\mu_c = (\mu_{1c},\ \mu_{2c},\ \ldots,\ \mu_{kc})$ and

$$\mathrm{rank}\left[\frac{\partial(v_0,\ v_1,\ \ldots,\ v_{k-1})}{\partial(\mu_1,\ \mu_2,\ \ldots,\ \mu_k)}\right]_{\mu=\mu_c} = k,$$

then k limit cycles can bifurcate from the critical point near the equilibrium by using appropriate perturbations on μ. More details on the topic of bifurcation of limit cycles can be found, for example, in the book Han and Yu (2012).

For multiple limit cycles bifurcation when $R_0 = 1$, we have the following theorem.

Theorem 3.3 *For system (5), when $R_0 = 1$, there exist feasible parameter values in the set γ_1 such that maximal two small-amplitude limit cycles can bifurcate from the endemic equilibrium E_1 due to generalized Hopf bifurcation. The outer limit cycle is stable while the inner one is unstable, and both of them enclose the stable equilibrium E_1.*

Proof To prove this theorem, we need to compute the focus values up to v_3 since there are two free parameters A and C. However, we will show that v_3 is not needed since no feasible parameter values can be chosen such that $v_2 = 0$. Actually, taking $A = \bar{A}$ we have $v_1 = 0$, and then v_2 is obtained from the Maple program (Yu 1998) as

$$\begin{aligned} v_2 = -\frac{4\rho^2(\rho+1)^2C^4\bar{A}^7}{192(1+C)(C^2+\bar{A})^4(C^2+\bar{A}+\bar{A}C)^4(\rho^3+\rho^2+1)^6}\big[&7\rho^{14} + 54\rho^{13} \\ &+ 182\rho^{12} + 384\rho^{11} + 656\rho^{10} + 998\rho^9 \\ &+ 1269\rho^8 + 1344\rho^7 + 1303\rho^6 + 1110\rho^5 + 791\rho^4 \\ &+ 527\rho^3 + 276\rho^2 + 117\rho + 54 \\ &+ \bar{A}\,(\rho+1)(4\rho^{15} + 33\rho^{14} + 127\rho^{13} + 315\rho^{12} \\ &+ 565\rho^{11} + 821\rho^{10} + 1071\rho^9 + 1272\rho^8 \\ &+ 1262\rho^7 + 1153\rho^6 + 966\rho^5 + 656\rho^4 + 449\rho^3 \\ &+ 234\rho^2 + 90\rho + 54)\big] < 0, \quad \text{for } \rho > 0. \end{aligned}$$

This clearly shows that there are no feasible parameter values that can yield three limit cycles. There exist infinitely many solutions for the existence of two limit cycles since there are two free parameters, as long as the conditions satisfy $v_1 = 0$ with $v_2 < 0$. □

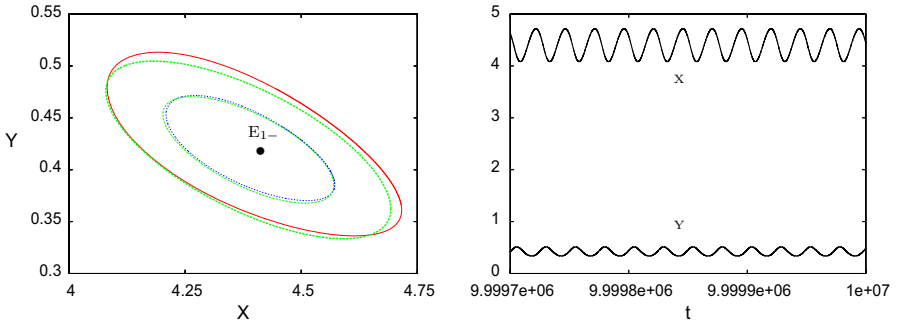

Fig. 5 Two limit cycles of system (5) when $R_0 = 1$ with $A = 0.2095365226$, $C = 0.5$ and $B = D = 0.1324446775$: **a** the phase portrait with the simulations in red and blue colors and the normal form predictions in green color (with amplitudes $r_1 \approx 0.1839$ and $r_2 \approx 0.3067$); and **b** the time history of the stable (outer) limit cycle (Color figure online)

To give an example to demonstrate the bifurcation of two limit cycles, we take $\rho = 1.397661\cdots$, which gives $A = \bar{A} = 0.208536\cdots$, $C = 0.5$, and $B = B_h = 0.132553\cdots$, resulting in $v_0 = v_1 = 0$ and $v_2 = -0.000222\cdots < 0$, as expected. Then, we perturb the parameters A and B such that $v_1 > 0$, $v_0 < 0$ and $|v_0| \ll v_1 \ll |v_2|$, and so two limit cycles can be obtained. More precisely, let $A = 0.208536 + \varepsilon_1$ and $B = B_h + \varepsilon_2$, where $\varepsilon_1 = 10^{-3}$ and $\varepsilon_2 = -7 \times 10^{-6}$, which yield $A = 0.209536$, $B = 0.132444$ and

$$v_0\mu = -0.731036 \times 10^{-6}, \quad v_1 = 0.293971 \times 10^{-4}, \quad v_2 = -0.229863 \times 10^{-3}. \tag{44}$$

Note that $\varepsilon_2 < 0$ again implies that B is decreasing to pass the Hopf critical point B_h since $v_0 \approx -0.672163 < 0$. Thus, the truncated normal form equation $v_0\mu + v_1 r^2 + v_2 r^4 = 0$ has two real roots: $r_1 \approx 0.1839$ and $r_2 \approx 0.3067$, which approximate the amplitudes of the two limit cycles. Since $v_2 < 0$, the outer limit cycle is stable while the inner one is unstable, and the equilibrium solution at this critical point is a stable focus.

The simulation, shown in Fig. 5, takes the exact parameter values:

$$A = 0.2095365226, \quad C = 0.5, \quad D = B = 0.1324446775.$$

The simulated phase portrait is shown in Fig. 5a where the stable (the larger one) and unstable (the smaller one) limit cycles are denoted by the red and blue curves, respectively. Analytic predictions based on the normal form are also shown in Fig. 5a as the green curves. It indeed indicates a good agreement between the simulations and the analytic predictions. Figure 5b depicts the time history of the stable (outer) limit cycle. Note that the simulation for the unstable limit cycle (or the unstable periodic motion) is obtained by using a negative time step in a fourth-order Runge–Kutta integration scheme.

3.2 Case $R_0 \neq 1$

When $R_0 \neq 1$ (i.e., $D \neq D_t = B$), the Hopf bifurcation conditions are given in Theorem 2.3, and the critical point is defined in (24) as D_h. However, unlike the case $R_0 = 1$ considered in the previous subsection, using the explicit solution of X_- will cause difficulty in computing the focus values (or the normal form), in particular for higher-order focus values. Thus, instead of using the explicit expression X_-, we use the parameter A to solve the equation $F_1 = 0$ and then use the parameter D to solve the trace of the Jacobian of the system to obtain

$$A = \frac{(1-BX_-)(1+C-DX_-)}{X_-(1-DX_-)} \quad \text{and} \quad D = D_h = \frac{1+C(1-X_-+BX_-^2)}{X_-}, \tag{45}$$

where $D = D_h$ denotes a Hopf critical point. Note that solving X_- from (45) yields the solution X_- given in (29). It follows from $A > 0$ that $1 - X_- + BX_-^2 < 0$.

3.2.1 Hopf Bifurcation

To study Hopf bifurcation which arises from the equilibrium E_{1-}, we use the Hopf bifurcation conditions given in Theorem 2.3 to define a subset γ_2 in the γ-parameter space as

$$\gamma_2 = \left\{\gamma \;\middle|\; B \in I_B,\ A \in (A_l, A_u),\ C \in (C_l, C_u),\ D \in \left(0, \tfrac{1}{2}\right)\right\}, \tag{46}$$

where

$$I_B = \left(0, \tfrac{1}{4}\right), \quad C_l = C_{l_1}, \quad C_u = \begin{cases} C_{u_1}, & \text{if } A \leq BC, \\ C_{u_1}, & \text{if } A > BC, \end{cases}$$

and A_l, A_u, C_{l_1}, C_{u_1} and C_{u_2} are defined in (24). Then for the parameter values taken from γ_2, Hopf bifurcation occurs at the critical point $D = D_h$.

We have the following theorem for Hopf bifurcation when $R_0 \neq 1$.

Theorem 3.4 *For system (5), when $R_0 \neq 1$ and the parameter values are taken from the γ_2 set, Hopf bifurcation occurs from the equilibrium E_{1-} at the critical point D_h. The bifurcation is supercritical (subcritical) if the first-order focus value v_1 is negative (positive), and a family of bifurcating limit cycles is stable (unstable), enclosing the unstable (stable) focus E_{1-}.*

Proof First, note that the equilibrium solution satisfies $Y_- = 1 - DX_- > 0$ (i.e., $0 < X_- < \frac{1}{D}$), and $1 - X_- + BX_-^2 < 0$ yields

$$\frac{1-\sqrt{1-4B}}{2B} < X_- < \frac{1+\sqrt{1-4B}}{2B}, \quad 0 < B < \frac{1}{4}.$$

Under these conditions, the frequency at the Hopf bifurcation point can be obtained as

$$\omega_c = \frac{(1-DX_-)(1-BX_-)}{-1+X_--BX_-^2}\sqrt{X_-(1-D-DX_-)}. \tag{47}$$

The condition $\omega_c > 0$ yields $0 < X_- < \frac{1}{D} - 1$ $(0 < D < 1)$, and so we have

$$\tfrac{1-\sqrt{1-4B}}{2B} < X_- < \min\left(\tfrac{1+\sqrt{1-4B}}{2B},\ \tfrac{1}{D} - 1\right), \quad 0 < B < \tfrac{1}{4}. \tag{48}$$

Further, note that

$$1 - BX_- > 1 - \tfrac{1+\sqrt{1-4B}}{2} = \tfrac{1-\sqrt{1-4B}}{2} = \tfrac{2B}{1+\sqrt{1-4B}} > B,$$
$$\text{and } X_- > \tfrac{1-\sqrt{1-4B}}{2B} = \tfrac{2}{1+\sqrt{1-4B}} \in (1,2) \implies \tfrac{1}{D} - 1 > 1 \implies D < \tfrac{1}{2}.$$

So the condition (48) becomes

$$\tfrac{1-\sqrt{1-4B}}{2B} < X_- < \min\left(\tfrac{1+\sqrt{1-4B}}{2B},\ \tfrac{1}{D} - 1\right), \quad 0 < B < \tfrac{1}{4},\ 0 < D < \tfrac{1}{2}. \tag{49}$$

Summarizing the above conditions defines the subset γ_2. Note that if the equilibrium solution $X_- > \frac{1}{D} - 1$, then it follows from (47) that the E_{1-} is a saddle point, rather than an elementary center.

Now, multiplying $Y + C$ to system (5) and applying the transformation,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X_- \\ 1 - DX_- \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ \frac{-1}{1+X_-} & -\frac{\omega_c}{C(1+X_-)(1-BX_-)} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

to the resulting system yields a system with its linear part in Jordan canonical form, and then using the Maple program (Yu 1998), we obtain the first focus value, given by

$$\begin{aligned} v_1 = {} & \frac{-1}{8X(1+X)(1-BX)(1-D-DX)(-1+X-BX^2)} \\ & \big\{(1+X)^2(1-BX)^2(1-X+BX^2)C^2 \\ & + \big[(X+3)(1-X+BX^2)^2 + (3BX+2X+B-4)(1-X+BX^2) \\ & + 3 - B - 4BX\big]C + 1\big\}, \end{aligned} \tag{50}$$

where $X = X_-$. Note that v_1 can be positive or negative for taking suitable parameter values, and its sign determines the stability of bifurcating limit cycles. In order to determine the critical condition such that $v_1 = 0$. We substitute X_- given in (29) into (50) to obtain an equation which is equivalent to $v_1 = 0$:

$$\begin{aligned} F_{v_1} = {} & A^2B^2\big[(1+B)^2 + (B+2)A\big]^2C^4 + A\big[2(B-1)(B^2+10B-2)A^4 \\ & + 2(B-1)(3B^3+29B^2+3B-2)A^3 \\ & + (6B^5+48B^4-42B^3-36B^2+5B+1)A^2 \\ & + 2B(B^5+6B^4-12B^3-18B^2-2B+1)A - 2B^3(B+3)(B+1)^2\big]C^3 \\ & + \big[(B^2+14B-3)A^6 + 2(26B^2+2B^3+4B-3)A^5 \\ & + (6B^4+68B^3+19B^2+14B-6)A^4 \end{aligned}$$

$$
\begin{aligned}
&+(4B^5+32B^4-20B^3+54B^2+B-2)A^3\\
&+B(B^5-2B^4-49B^3+48B^2+20B-3)A^2\\
&-2B^2(2B^4+10B^3-10B^2-11B+1)A+B^4(B+3)^2\big]C^2\\
&-(A+B)\big[2(B-1)A^5\\
&+2(B+1)(4B-1)A^4+2(6B^3+14B^2+4B-1)A^3\\
&+(8B^4+28B^3+16B^2+4B-1)A^2\\
&+2B^2(B^3+3B^2+3)A-2B^4(B+3)\big]C+(A+B)^6\\
=\;&0. \qquad (51)
\end{aligned}
$$

Therefore, for any values given in $B \in I_B$ and $A \in (A_l, A_u)$, we can use (51) to solve for C and then use (24) to get D_h. Further, we again use (24) to verify if these obtained parameter values satisfy $C \in (C_l, C_u)$; and if $D_h \in (0, D_t)$ when $A \le BC$, or $D_h \in (0, D_s)$ when $A > BC$. If all the conditions are satisfied, we have identified the critical point such that $v_1 = 0$. In next subsection, we will present a different method to investigate multiple limit cycles bifurcation. □

To this end, before considering multiple limit cycle bifurcation, we present several numerical examples to show supercritical and subcritical Hopf bifurcations, arising from the equilibrium E_{1-}.

(1) Taking $D = D_h = \frac{1}{10}$ and $B = \frac{7}{64}$ gives $R_0 > 1$ (i.e., $D < B$). Using (48) we obtain $\frac{8}{7} < X_- < 8$, and choose, for instance, $X_- = 7$, and use (45) to obtain $C = \frac{96}{205}$ and then $A = \frac{225}{2624}$. For these parameter values, we have $v_1 = -\frac{23}{17220}$ which indicates that the Hopf bifurcation is supercritical and the bifurcating limit cycle is stable. More precisely, it can be shown by using (45) and (50) that for $D = \frac{1}{10}$ and $B = \frac{7}{64}$, we have $v_1 < 0$ if $X_- \in (\frac{8}{7}, 2.419645) \cup (6.017355, 8)$; and $v_1 > 0$ if $X_- \in (2.419645, 6.017355)$.

(2) Choosing $D = D_h = \frac{1}{10}$ and $B = \frac{9}{100}$ yields $R_0 < 1$ (i.e., $D > B$), and $\frac{10}{9} < X_- < 9$. We take $X_- = 4$, and similarly use (45) and (50) to find $C = \frac{5}{13}$, $A = \frac{256}{975}$, and $v_1 = \frac{431}{133120}$. Thus, the Hopf bifurcation is subcritical and the bifurcating limit cycle is unstable. As a matter of fact, we can also use (45) and (50) to show that for $D = \frac{1}{10}$ and $B = \frac{9}{100}$, $v_1 < 0$ if $X_- \in (\frac{10}{9}, 2.116127)$; and $v_1 > 0$ if $X_- \in (2.116127, 9)$.

However, it is noted that all the parameter values of A, B and C in the above examples belong to the category $A > BC$. This suggests that it is pretty easy to find Hopf bifurcation when $A > BC$. Now we want to find Hopf bifurcation when $A \le BC$. We use the conditions on B and A defined in γ_2 and then apply (51) to determine the value of parameter C. We will give a couple of examples to show that for the case $A < BC$, v_1 can also be positive or negative. Let $B = 0.05 \in (0, 0.25)$. We take $A = 0.0156 \in (A_l, A_u) = (0.0125, 0.020225)$. Then we solve (51) for C to obtain two positive solutions: $C = 0.097104\cdots$ and $C = 0.461628\cdots$, both of them are located in the interval $(C_l, C_u) = (C_{l_1}, C_{u_1}) = (0.031912, 0.564006)$. However, only the second solution satisfies $v_1 = 0$. Thus, we choose two values for C: $C = 0.458 < 0.461628\cdots$ and $C = 0.476 > 0.461628\cdots$, both of them yield $A < BC$.

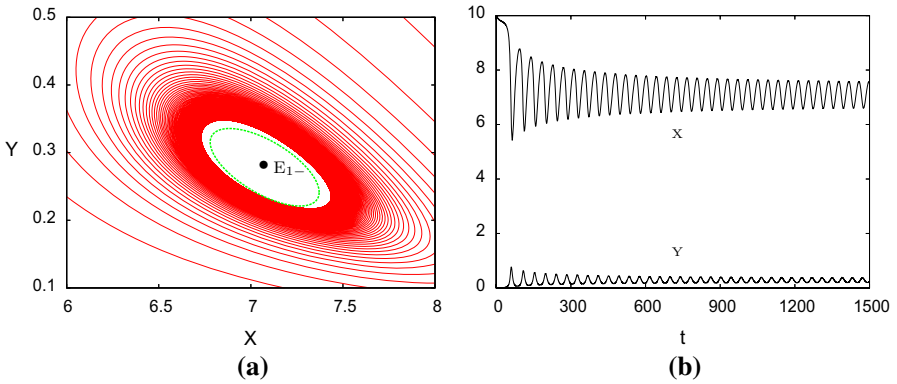

Fig. 6 A stable limit cycle of system (5) for $R_0 > 1$ ($A > BC$) with $A = \frac{225}{2624}$, $C = \frac{96}{205}$, $B = \frac{7}{64}$ and $D = \frac{51}{500}$: **a** phase portrait with the red and green curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 0.2958$), respectively; and **b** the time history of the stable limit cycle (Color figure online)

For the former, the Hopf critical point is given by $D_h = 0.011272 < 0.05 = D_t$, and we obtain $v_1 = 0.381609 \times 10^{-6} > 0$, showing a subcrtical Hopf bifurcation; while for the latter, the Hopf critical point is $D_h = 0.009358 < 0.05 = D_t$, and the first-order focus value becomes $v_1 = -0.149416 \times 10^{-5} < 0$, indicating a supercritical Hopf bifurcation.

The above examples indicate that Hopf bifurcations can be either supercritical or subcritical for both cases $A \leq BC$ and $A > BC$, implying that two limit cycles are possible to occur from the equilibrium E_{1-}, which will be considered in the next subsection by using a different method.

For simulation, we give two examples, one for $R_0 > 1$ and one for $R_0 < 1$.

(i) For $R_0 > 1$, we choose $A = \frac{225}{2624}$, $B = \frac{7}{64}$ and $C = \frac{96}{205}$. Then, $D_h = \frac{1}{10}$ under which $\mathrm{Tr(J)} = 2v_0\,\mu = 0$ (due to $\mu = 0$), where $v_0 = 0.058427 > 0$. As shown in above, $v_1 = -\frac{23}{17220}$, indicating that the Hopf bifurcation is supercritical and the bifurcating limit cycle is stable. To achieve a stable limit cycle, we need to perturb some parameter such that $\mu > 0$, yielding $v_0\mu > 0$. We add the perturbation $\mu = \frac{1}{500}$ to D so that $D = \frac{1}{10} + \mu = \frac{51}{500}$, for which v_1 is unchanged, and obtain $v_0\mu \approx 0.000116854$. Then the truncated normal form equation, $v_0\mu + v_1 r^2 = 0$ gives an estimation for the amplitude of the stable limit cycle as $r \approx 0.2958$. The comparison between the analytical prediction (the green curve) and the simulation (the red curve) is depicted in Fig. 6a, which again shows an excellent agreement and the time history of the stable limit cycle is given in Fig. 6b.

(ii) For $R_0 < 1$, we again choose $D_h = \frac{1}{10}$, but $B = \frac{3}{32} < D_h$. If we choose $A = \frac{121}{1440}$ and $C = \frac{16}{75}$, and get $v_1 = \frac{3}{3850} > 0$, indicating that this case is a subcritical Hopf bifurcation and the bifurcating limit cycle is unstable. Since $v_0 = 0.581928 > 0$, in order to obtain this unstable limit cycle, we take $\mu = -\frac{1}{2000}$ and let $D = \frac{1}{10} + \mu = \frac{199}{2000}$, yielding $v_0\mu \approx -0.0002910$, showing that the equilibrium E_1 is a stable focus and the limit cycle is unstable. The estimation of the amplitude of the unstable limit

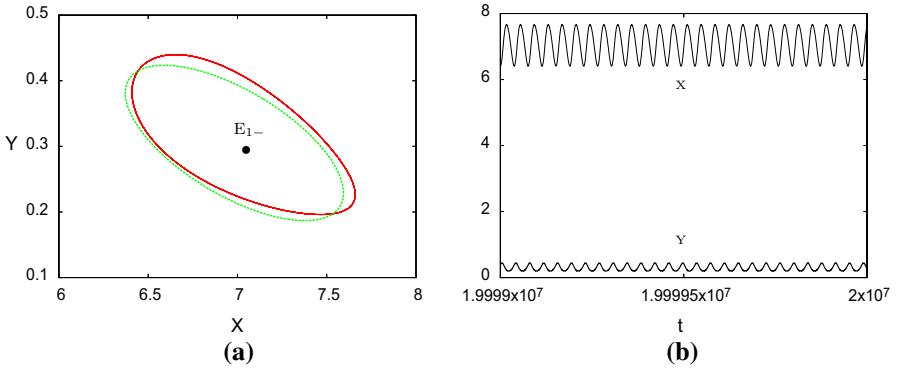

Fig. 7 An unstable limit cycle of system (5) for $R_0 < 1$ ($A > BC$) with $A = \frac{121}{1440}$, $C = \frac{16}{75}$, $B = \frac{3}{32}$ and $D = \frac{199}{2000}$: **a** phase portrait with the red and green curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 0.6111$), respectively; and **b** the time history of the unstable limit cycle (Color figure online)

cycle can be found from the normal form as $r \approx 0.6111$. The simulation is shown in Fig. 7a, where the unstable limit cycle (in red color) is obtained by using a negative time step in the numerical integration scheme. The green curve again denotes the normal form prediction, showing a good prediction even for a pretty large unstable limit cycle. Figure 7b shows the time history of the unstable limit cycle.

3.2.2 Multiple Limit Cycles Bifurcation

Since it has been shown in previous subsection that v_1 can be positive or negative by taking appropriate parameter values, it implies that v_1 can be zero and more limit cycles may bifurcate from the Hopf critical point. This is a more challenging task since the system contains four free parameters and in general it may exhibit at most four limit cycles. However, we will show that four and three limit cycles are not possible due to the physical restriction on the parameters, and there can still exist maximal two limit cycles even now it allows $B \neq D$. For this case, we obtain the focus values v_2, v_3 and v_4 in terms of the parameters B, D and X_-. We omit the lengthy expressions of these focus values for brevity. We have the following theorem for this case.

Theorem 3.5 *For system* (5), *when* $R_0 \neq 1$, *there exist feasible parameter values in the* γ_2 *set such that maximal two small-amplitude limit cycles can bifurcate from the endemic equilibrium* E_{1-} *due to generalized Hopf bifurcation. The outer limit cycle is stable while the inner one is unstable, and both of them enclose the stable equilibrium* E_{1-}.

Proof Here, we use the parameters A and C to solve the equation $F_1 = 0$ and the trace of the Jacobian of the system to get

$$A = \frac{(1-BX_-)^2}{-1+X_--BX_-^2} \quad \text{and} \quad C = \frac{1-DX_-}{-1+X_--BX_-^2}, \tag{52}$$

Note that $-1 + X_- - BX_-^2 > 0$ under the condition (49).

Since now there are three free variables B, D and X_- in the expressions of the focus values v_i ($i \geq 1$), the best expected result is to find the conditions such that $v_1 = v_2 = v_3 = 0$, but $v_4 \neq 0$, which implies possible existence of four limit cycles. We first consider whether there exist three limit cycles, i.e., to find the solutions such that $v_1 = v_2 = 0$, but $v_3 \neq 0$. v_1 is given in (50). Eliminating B from the two equations $v_1 = 0$ and $v_2 = 0$ we obtain $B = \frac{B_N}{B_D}$, where

$$\begin{aligned}
B_N = {} & 4D^3(4D^2 - 9D + 3)X^7 + 12D^2(2D^3 - 6D^2 + 9D - 3)X^6 \\
& - D(17D^4 - 51D^3 - 34D^2 \\
& + 102D - 36)X^5 - 2(20D^5 - 84D^4 + 145D^3 - 59D^2 \\
& - 12D + 6)X^4 - (13D^5 - 75D^4 \\
& + 240D^3 - 316D^2 + 132D - 6)X^3 \\
& + 2(D^5 - 7D^4 + 15D^3 + 28D^2 - 54D + 18)X^2 \\
& - (D - 1)(8D^3 - 43D^2 + 48D + 6)X - 6(2 - D)(1 - D)^2, \\
B_D = {} & 4D^3(4D^2 - 9D + 3)X^8 + 6D^2(5D^3 - 17D^2 + 20D - 6)X^7 \\
& - 2D(5D^4 - 6D^3 - 62D^2 \\
& + 69D - 18)X^6 - (45D^5 - 224D^4 + 289D^3 - 34D^2 \\
& - 60D + 12)X^5 - (20D^5 - 162D^4 \\
& + 472D^3 - 441D^2 + 114D + 6)X^4 \\
& + (D^5 + 2D^4 - 101D^3 + 308D^2 - 237D + 42)X^3 \\
& - (14D^4 - 88D^3 + 79D^2 + 48D - 42)X^2 - (1 - D)(25D^2 \\
& - 69D + 6)X - 12(1 - D)^2,
\end{aligned} \tag{53}$$

and a resultant:

$$R_{12} = DX(1 + X)(1 - DX)(1 - D - DX)\, R_{12a} R_{12b},$$

where

$$\begin{aligned}
R_{12a} = {} & D(1 - X)^2 - X + 2, \\
R_{12b} = {} & 12D^2(3D - 1)X^4 + 24D(2D^2 - 4D + 1)X^3 \\
& + (13D^3 - 55D^2 + 84D - 12)X^2 \\
& - (1 - D)(D^2 + 30D + 24)X + 15(1 - D)^2.
\end{aligned} \tag{54}$$

Note in the above expressions that $X = X_-$. It is seen from (49) that $1 < \frac{1-\sqrt{1-4B}}{2B} < X_- < \frac{1}{D} - 1$ for which the E_{1-} is an elementary center; and the E_{1-} is a saddle if $X_- > \frac{1}{D} - 1$.

To find the roots of R_{12}, first letting $R_{12a} = 0$ gives two real solutions:

$$X_-^{\pm} = \tfrac{1}{2D}\big(1 + 2D \pm \sqrt{1 - 4D}\big), \quad \text{for } 0 < D \leq \tfrac{1}{4}. \tag{55}$$

It is easy to show that $X_-^+ > \frac{1}{D} - 1$ for $0 < D < \frac{1}{4}$, and at $D = \frac{1}{4}$, $X_-^+ = X_-^- = 3$, leading to BT bifurcation. Thus, X_-^+ is not a feasible solution for bifurcation of multiple limit cycles from a Hopf critical point. Moreover, it can be verified that the solution X_-^- together with the solution $B = \frac{B_N}{B_D}$ results in $v_1 = v_2 = v_3 = v_4 = \cdots = 0$, implying that this is a center condition. As a matter of fact, the above solution X_-^- yields

$$A = \frac{1}{1+D}, \quad B = \frac{D}{1+D}, \quad C = 1 + D, \quad D \in \left(0, \frac{1}{4}\right], \tag{56}$$

under which (5) is an integrable system having the first integral,

$$F(X, Y) = X + Y + \ln\left(\frac{Y^{1+D}}{|1 - DX - XY|}\right). \tag{57}$$

Note that this integral system belongs to the case $A > BC$. We will not discuss this integral system further in this study.

Next, we consider the factor R_{12b} and will show that $R_{12b} = 0$ does not have feasible solutions to yield three limit cycles. First, for the solution $B = \frac{B_N}{B_D}$ and the equation $R_{12b} = 0$, a more precise upper bound for D can be obtained by considering the restriction $B \in (0, \frac{1}{4})$ on the solution $B = \frac{B_N}{B_D}$. To achieve this, we may use the two equations $R_{12b} = 0$ and $B = \frac{B_N}{B_D} = \frac{1}{4}$ to solve X and D, yielding the maximal value $D_{\max} = 0.252605\cdots$. Thus, we only need to consider the values of D in the interval $D \in (0, D_{\max})$ for the case of three limit cycles. Further, we want to prove that $R_{12b} = 0$ does not have real solutions for $X \in (1, \frac{1}{D} - 1)$. It is easy to verify that for $D \in (0, D_{\max})$,

$$\begin{aligned}
&R_{12b}|_{X\to-\infty} = -\infty,\\
&R_{12b}|_{X=0} = 15(1-D)^2 > 0,\\
&R_{12b}|_{X=1} = -21 + 72D - 119D^2 + 98D^3 < 0,\\
&R_{12b}|_{X=\frac{1}{D}-1} = -10(1-D)^2 < 0,\\
&R_{12b}|_{X\to\infty} = -\infty,
\end{aligned}$$

which shows that the polynomial R_{12b} has at least one real root for $X < 0$ and at least one real root for $0 < X < 1$. Then, if the R_{12b} has four real roots, then at most two real roots for $X > 1$. In fact, by using Sturm's theorem, we can show that the polynomial R_{12b} always has four real solutions for $X \in (-\infty, \infty)$ with $D \in (0, D_{\max})$. An example is shown in Fig. 8. However, since $R_{12b} < 0$ at $X = 1$, $X = \frac{1}{D} - 1$ and $X = +\infty$, the number of the roots of R_{12b} which can occur for $X > 1$ must be even, i.e., either two or zero, one is not possible.

More precisely, we can show that exactly one real root of R_{12b} is in the interval $X \in (-\infty, 0)$ and one real root in the interval $X \in (0, 1)$. Thus, two real roots must appear in the interval $X \in (1, \infty)$, and they must both appear either in the interval $X \in (1, \frac{1}{D} - 1)$ or in the interval $X \in (\frac{1}{D} - 1, \infty)$ due to $R_{12b} < 0$ at $X = 1$, $X = \frac{1}{D} - 1$ and $X = \infty$. In the following, we will show that the two real roots must locate in the interval $X \in (\frac{1}{D} - 1, \infty)$. To achieve this, consider the derivative,

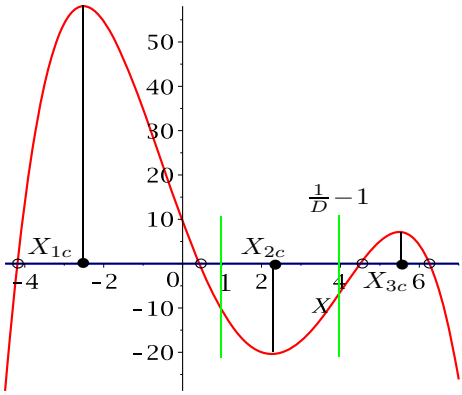

Fig. 8 Graph $R_{12b}(X)$ for $D = 0.2$ with two vertical lines (in green color) at $X = 1$ and $X = \frac{1}{D} - 1$, respectively (Color figure online)

$$\begin{aligned}\frac{\mathrm{d}R_{12b}}{\mathrm{d}X} &= -48D^2(1-3D)X^3 + 72D(1-4D+2D^2)X^2 \\ &\quad + 2(13D^3 - 55D^2 + 84D - 12)X - (1-D)(24+30D+D^2),\end{aligned}$$

whose discriminant is

$$\begin{aligned}\mathrm{Disc} = &-\frac{1}{746496D^6(1-3D)^4}\big(432 - 2592D + 12744D^2 - 43200D^3 + 56979D^4 \\ &+ 80784D^5 - 421418D^6 + 397326D^7 + 5556D^8 - 171736D^9 + 1425D^{10}\big).\end{aligned}$$

It can be shown that $\mathrm{Disc} < 0$ for $D \in (0, 0.34)$, and thus $\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}$ has (maximal) three distinct real solutions in $X \in (-\infty, \infty)$ for any value of $D \in (0, D_{\max})$. Further, for $D \in (0, D_{\max})$ we have

$$\begin{aligned}&\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X\to-\infty} = +\infty, \\ &\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=0} = -(1-D)(24+30D+D^2) < 0, \\ &\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1} = -48 + 234D - 417D^2 + 315D^3 < 0, \\ &\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=\frac{1}{D}-1} = (1-D)(24-44D+25D^2) > 0, \\ &\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X\to\infty} = -\infty,\end{aligned}$$

where $\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1} < 0$ is due to $\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1,D=0} = -48 < 0$, $\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1,D=D_{\max}} = -10.421481\cdots < 0$, and $\frac{\mathrm{d}}{\mathrm{d}D}\left(\left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1}\right) = 234 - 834D + 945D^2 > 0$ for $D \in (0, D_{\max})$. Therefore, the three real critical roots, X_{ic}, of $\frac{\mathrm{d}R_{12b}}{dX}$ are located in the intervals, $X_{1c} \in (-\infty, 0)$, $X_{2c} \in (1, \frac{1}{D} - 1)$ and $X_{3c} \in (\frac{1}{D} - 1, \infty)$, respectively, yielding local maximum of R_{12b} at $X = X_{ic}$, $i = 1, 3$, and local minimum at $X = X_{2c}$ (see black circles in Fig. 8). Moreover, by noticing

$$R_{12b}|_{X=1} < 0, \quad R_{12b}|_{X=\frac{1}{D}-1} < 0, \quad \left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=1} < 0, \quad \left.\frac{\mathrm{d}R_{12b}}{\mathrm{d}X}\right|_{X=\frac{1}{D}-1} > 0,$$

we know that $R_{12b} < 0$ for $X \in (1, \frac{1}{D} - 1)$. Now, it is easy to see the distribution of the four real roots of R_{12b}: one in the interval $X \in (-\infty, 0)$, one in the interval $X \in (0, 1)$, and two in the interval $X \in (1, \infty)$, see the circles in Fig. 8.

Summarizing the above results indicates that there are no feasible parameter values in the γ_2 set, which can yield three limit cycles. Hence, the best result is two limit cycles. The existence of two limit cycles for $R_0 \neq 1$ is obvious since we have already shown that there are two limit cycles for the case $R_0 = 1$ which has restriction $D = B$.

The remaining question is about the sign of v_2. Because we have shown that the maximal number of limit cycles is two for feasible parameter values, v_2 must keep the sign unchanged for feasible parameter values; otherwise, we have found solutions such that $v_2 = 0$. To determine the sign of v_2, we only need to use a special solution under which $v_1 = 0$ and then find the sign of v_2. We take $B = \frac{1}{8}$ which yields $A_l = \frac{1}{8}$, and then choose $A = \frac{1}{4} > A_l$ and use (29) to obtain $X_- = 4$. Next, using (51) to solve C we obtain $C = \frac{9+\sqrt{481}}{50}$, which satisfies $A > BC$. Further, we use (24) or (28) to get $D_h = \frac{41-\sqrt{481}}{200}$, and finally substitute these parameter values into v_2 and ω_c to obtain $v_2 = \frac{\sqrt{481}-191}{1228800} < 0$ and $\omega_c^2 = \frac{253+17\sqrt{481}}{3125} > 0$. This indicates that v_2 is always negative for feasible parameter values when $v_1 = 0$. Therefore, like the case $R_0 = 1$, if there exist two limit cycles bifurcating from a Hopf critical point for the case $R_0 \neq 1$, the outer one must be stable. □

Now, we present several simulations of multiple limit cycles bifurcation from a Hopf critical point, for $R_0 > 1$ and $R_0 < 1$ with either $A > BC$ or $A < BC$.

(1) First, consider $R_0 > 1$ with $A > BC$. We choose $D = \frac{1}{10}$, $B = \frac{7}{64}$ and then use (45) and (50) to obtain $X_- = 6.017354\cdots$. Further, it follows from (52) that $A = 0.110556\cdots$ and $C = 0.376772\cdots$ for which $v_0\mu = v_1 = 0$ and $v_2 = -0.000100\cdots < 0$, indicating that the larger one of the two limit cycles is stable, as expected. In order to have an unstable smaller limit cycles and stable equilibrium E_1, we need perturbations such that $v_1 > 0$ and $v_0\mu < 0$ satisfying $-v_0\mu \ll v_1 \ll -v_2$. To achieve this, we choose perturbations such that $A = 0.110934$ and $C = 0.376594$, which results in $v_0\mu = -0.889 \times 10^{-7}$, $v_1 \approx 0.0000101$, and $v_2 \approx -0.000103$, yielding the approximate amplitudes of the two limit cycles: $r_1 \approx 0.098849$ and $r_2 \approx 0.297144$. As depicted in Fig. 9a, the simulations given in red color (for the stable limit cycle) and blue color (for the unstable limit cycle) agree very well with the analytic predictions in green color. The time history of the stable (outer) limit cycle is shown in Fig. 9b. This example clearly indicates a bistable phenomenon involving a stable equilibrium and a stable (outer) limit cycle with an unstable (inner) limit cycle as their separatory.

(2) Next, consider $R_0 < 1$ with $A > BC$. Similarly we choose $B = \frac{3}{32} < D = \frac{1}{10}$ and then get $X_- = 8.129787\cdots$, yielding $A = 0.060591\cdots$ and $C = 0.200338\cdots$ for which $v_0\mu = v_1 = 0$ and $v_2 = -0.000182\cdots < 0$, implying that the larger one of the two limit cycles is stable. Then, we add perturbations to the parameters A and C to obtain $A = 0.06085981$ and $C = 0.20023941$, which yields $v_0\mu = -0.17779 \times 10^{-6}$, $v_1 \approx 0.0000196$, and $v_2 \approx -0.0001863$, from which we obtain the amplitudes of the two limit cycles as $r_1 \approx 0.1003$ and $r_2 \approx 0.3081$. These two limit cycles (see Fig. 10a) are similar to those shown in Fig. 9a for

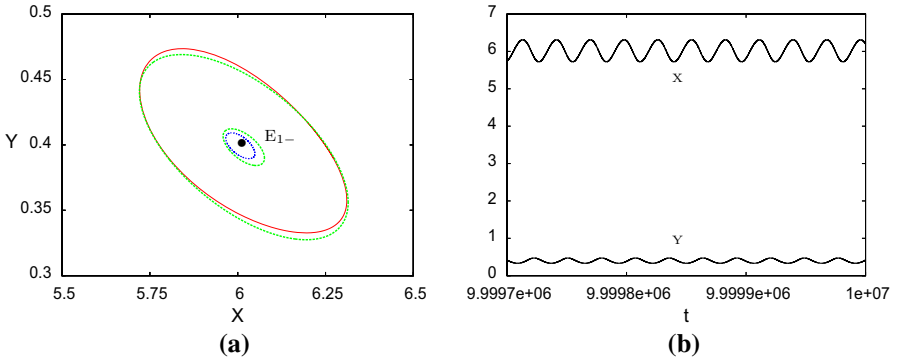

Fig. 9 Two limit cycles of system (5) for $R_0 > 1$ ($A > BC$) with $A = 0.110934$, $B = 0.109375$, $C = 0.376594$ and $D = 0.1$, showing a bistable phenomenon: **a** phase portrait with the red/blue and green color curves denoting the simulations and the estimations from normal form (with amplitudes $r_1 \approx 0.0989$ and $r_2 \approx 0.2971$), respectively; and **b** the time history of the stable (outer) limit cycle (Color figure online)

$R_0 > 1$, but for this case $R_0 < 1$, the phase portrait is different since the system in addition has a stable node at E_0 besides a saddle point at E_1^+ and a stable focus at E_1^-. Thus, this example indeed shows a tristable phenomenon with two stable equilibria (one node and one focus) and one stable limit cycle (another limit cycle is unstable, as a separatory between the stable focus and the stable limit cycle). The simulations (in red/blue color) and analytical predictions (in green color) are shown in Fig. 10a, again indicating a good agreement. The time history of the stable (outer) limit cycle is depicted in Fig. 9b.

(3) Note that the above two examples belong to the category $A > BC$. The easiness of finding feasible parameter values for two limit cycles when $A > BC$ indeed shows that the system is more likely to exhibit bistable or even tristable complex dynamics. For an illustration, we give an example to satisfy $A < BC$, yielding bistable phenomenon. As discussed above, we may take $B = 0.05$ and $A = 0.0156$, and then use (51) to find $C = 0.461627\cdots$ and (24) to obtain $D = D_h = 0.010886\cdots$ such that $v_0 = v_1 = 0$ and $v_2 = -0.771730\cdots \times 10^{-5} < 0$, as expected. Further, taking perturbations on C and D as $C = 0.4616276809 - 0.005 = 0.4566276809$ and $D = 0.011418264 + 0.00000005 = 0.011418314$, for which we have $v_0\mu = -0.513018 \times 10^{-7}$, $v_1 \approx 0.190025 \times 10^{-5}$ and $v_2 \approx -0.797500 \times 10^{-5}$. Thus the truncated normal form equation yields the approximations for the amplitudes of the two limit cycles: $r_1 \approx 0.1065$ and $r_2 \approx 0.4764$. The simulations and analytic predictions as depicted in Fig. 11 show a good agreement. For this case, bistable phenomenon appears to involve a stable equilibrium and a stable periodic motion.

The results obtained in this section, in particular for Theorems 3.3 and 3.5, indicate that regardless whether $R_0 = 1$ or $R_0 \neq 1$, system (5) can always exhibit complex dynamics including different types of bistability or even tristablity, due to multiple limit cycles arising from Hopf bifurcation. This suggests that the real situation could be very complex, showing the coexistence of a stable disease-free equilibrium, stable endemic

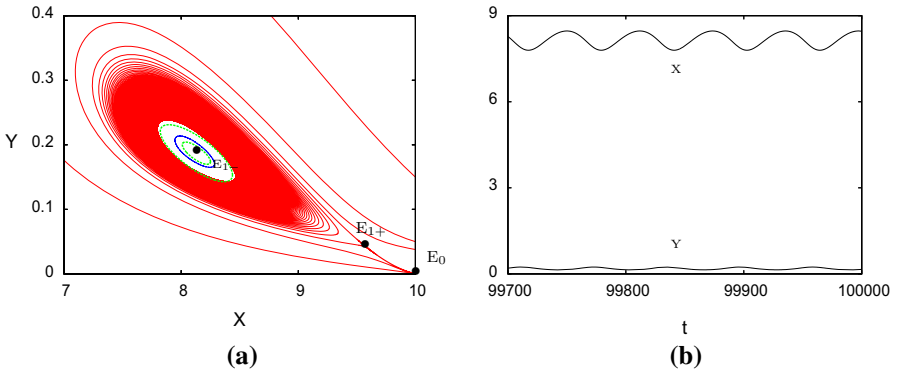

Fig. 10 Two limit cycles of system (5) for $R_0 < 1$ ($A > BC$) with $A = 0.06085981$, $B = 0.09375$, $C = 0.20023941$ and $D = 0.1$, showing a tristable phenomenon: **a** phase portrait with the red/blue and green color curves denoting the simulations and estimates from the normal form (with amplitude $r_1 \approx 0.1003$ and $r_2 \approx 0.3081$), respectively; and **b** time history of the stable (outer) limit cycle (Color figure online)

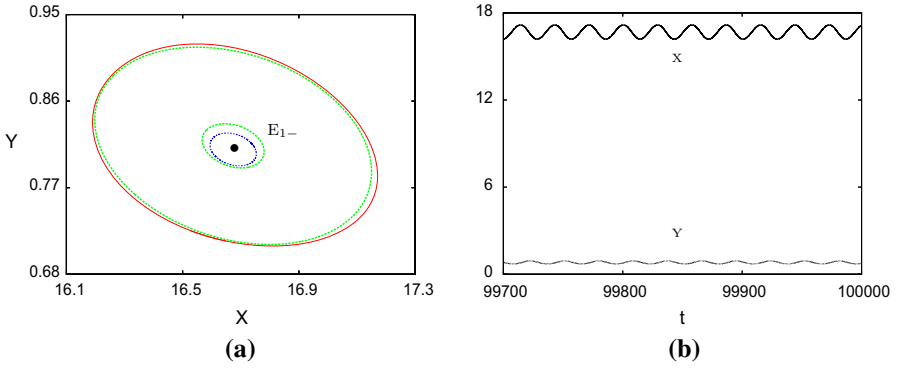

Fig. 11 Two limit cycles of system (5) for $R_0 > 1$ ($A < BC$) with $A = 0.0156$, $B = 0.05$, $C = 0.4566276809$ and $D = 0.011418214$: **a** phase portrait with the red/blue and green color curves denoting the simulations and estimates from normal form (with amplitude $r_1 \approx 0.1065$ and $r_2 \approx 0.4764$), respectively; and **b** time history of the stable (outer) limit cycle (Color figure online)

equilibria, and even stable oscillating motion, all of which are possible depending upon the initial conditions.

3.3 Recurrence Phenomenon (Viral Blips)

It has been shown in Zhang et al. (2013, 2014a, b),Yu et al. (2016) that the recurrence phenomenon can often appear in many disease models. It is characterized by short episodes of high viral reproduction, separated by long periods of relative quiescence. This recurrent pattern is observed in many persistent infections, including the "viral blips" observed during chronic infection with the human immunodeficiency virus (HIV). In fact, the model (5) considered in this paper indeed shows recurrence behav-

ior for a very large parameter region, see Zhang et al. (2013, 2014a, b) in which how this phenomenon occurs is discussed using dynamical system theory. Mathematically speaking, such "slow-fast" motion is a special type of limit cycles. However, it has been shown that such "slow-fast" motion cannot be analyzed by the well-known geometric singular perturbation theory (GSPT). It has been proposed that if the following conditions hold:

P_1: there exists at least one equilibrium solution;
P_2: there exists a saddle-node or transcritical bifurcation;
P_3: there is a Hopf bifurcation; and
P_4: there is a "window" between the Hopf bifurcation point and the saddle-node/transcritical bifurcation point in which oscillation continuously exists,

then the system exhibits relaxtion-type slow-fast motions. Note that Hopf bifurcation is necessary since it is the source of oscillations. To verify these conditions for higher-dimensional dynamical systems, identifying Hopf bifurcation (condition P_3) becomes crucial.

However, although such slow-fast oscillations are closely related to Hopf bifurcation, Hopf bifurcation cannot be used to predict or estimate such motions because normal form theory is no longer applicable for such a large perturbation, that is, such special bifurcating limit cycles are far away from the equilibrium. To illustrate this fact, we give simulations for three different cases: $R_0 = 1$, $R_0 > 1$ and $R_0 < 1$.

First we consider $R_0 = 1$. For this case, we use the same values of A and C given in (43), yielding the same B_h, v_0 and v_1 as that example shown in Fig. 4, but change B and D to $D = B = \frac{1}{10}$. For a comparison, we still use the normal form to estimate the amplitude of the oscillation. Here, $\mu = B_h - B = 0.017947$, and so the truncated normal form gives the approximation of the amplitude as $r \approx 2.3551$, which is very large, implying that the Hopf bifurcation theory and associated normal form are no longer applicable. The simulation is depicted in Fig. 12a as the red curve, where the green curve again denotes the normal form estimate, which indeed shows a very large deviation from the simulation, even with a negative part, implying that normal form theory is no longer applicable. The simulated time history given in Fig. 12b clearly shows the recurrence infection (viral blips) phenomenon, and this recurrence for the case $R_0 = 1$ was not considered in Zhang et al. (2013, 2014a, b), Yu et al. (2016).

For the case $R_0 > 1$, we take the perturbation on A as $\mu = \frac{1}{50}$ so that $A = \frac{225}{2624} + \mu = \frac{6937}{65600}$, and again take $C = \frac{96}{205}$ and $D = \frac{1}{10}$. Note that this set of parameter values (except B) is exactly the same as that for the case $R_0 = 1$, see Fig. 12. Now we apply the truncated normal form to obtain $r \approx 2.4778$, which is again very large and normal form theory is not applicable. The simulated phase portrait is given in Fig. 13a (see the red colored curve) and the green curve in this figure shows the normal form prediction, indicating a large discrepancy. The simulated time history given in Fig. 13b again shows the recurrence infection.

Finally for the case $R_0 < 1$, we again choose $D = \frac{1}{10}$, but $B = \frac{3}{32} < D$. If we choose $A = \frac{121}{1440}$ and $C = \frac{16}{75}$, then we have $v_1 = \frac{3}{3850} > 0$. To estimate the amplitude of this unstable limit cycle, we take $\mu = \frac{1}{50}$ and let $A = \frac{121}{1440} + \mu = \frac{749}{7200}$, yielding $r \approx 1.9346$. The simulation is shown in Fig. 14, where the unstable limit cycle is obtained by using a negative time step in a numerical integration scheme.

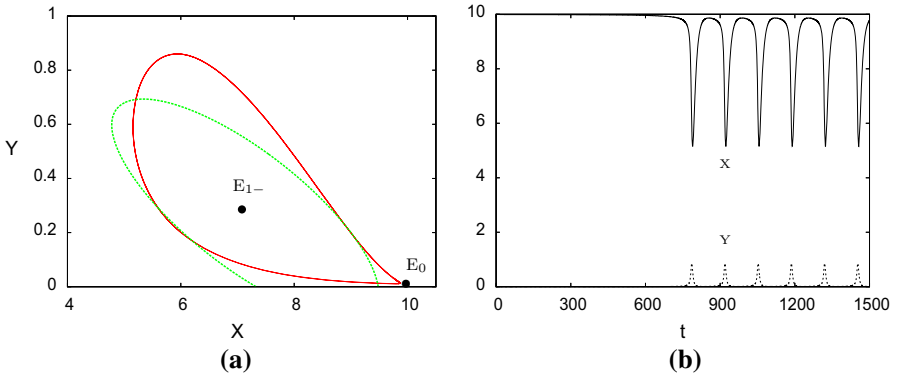

Fig. 12 Simulated recurrence oscillation of system (5) for $R_0 = 1$ with $A = \frac{6937}{65600}, C = \frac{96}{205}, B = D = \frac{1}{10}$: **a** phase portrait with the red and green color curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 2.3551$), respectively; and **b** the time history of the stable oscillation (Color figure online)

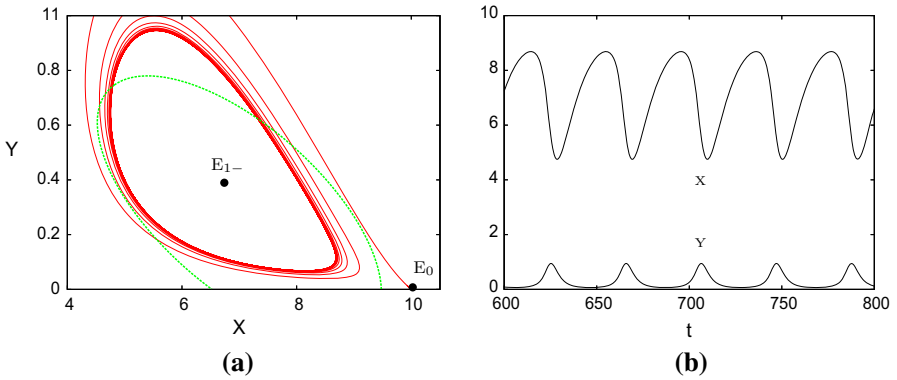

Fig. 13 Simulated recurrence oscillation of system (5) for $R_0 > 1$ ($A > BC$) with $A = \frac{6937}{65600}, C = \frac{96}{205}$, $B = \frac{7}{64}$ and $D = \frac{1}{10}$: **a** phase portrait with the red and green color curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 2.4778$), respectively; and **b** the time history of the stable oscillation (Color figure online)

Again, the green curve denotes the normal form prediction, showing a large difference from simulation. This recurrence for $R_0 < 1$ was also not studied in Zhang et al. (2013, 2014a, b), Yu et al. (2016).

The above three examples for the three different cases: $R_0 > 0$, $R_0 = 0$ and $R_0 < 0$ exhibit the recurrence phenomenon, which cannot be predicted by normal form theory or analyzed by using geometric singular perturbation theory. However, we have shown that the slow-fast motions can be induced from Hopf bifurcation if the four conditions P_1–P_4 are satisfied. This provides a mechanism to generate the recurrence behavior in disease models and has been discussed in Zhang et al. (2013, 2014a, b). In next section, we will provide another new mechanism of generating recurrence from homoclinic loops arising from BT bifurcation.

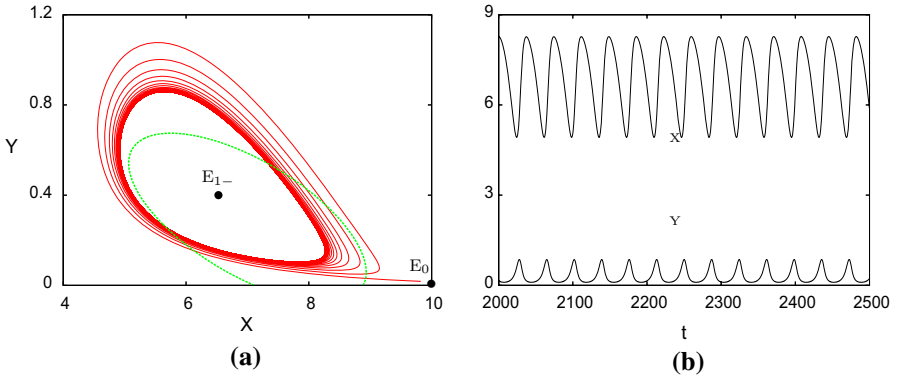

Fig. 14 Simulated recurrence oscillation of system (5) for $R_0 < 1$ ($A > BC$) with $A = \frac{749}{7200}$, $C = \frac{16}{75}$, $B = \frac{3}{32}$ and $D = \frac{1}{10}$: **a** phase portrait with the red and green color curves denoting the simulation and the estimation from normal form (with amplitude $r \approx 1.9346$), respectively; and **b** the time history of the unstable oscillation (Color figure online)

4 Bogdanov–Takens (BT) Bifurcations

In this section, we study Bogdanov–Takens (BT) bifurcation in system (5), which is characterized by a critical point associated with double-zero eigenvalues. First, BT bifurcation cannot occur (1) at the equilibrium $\mathrm{E}_0 = (\frac{1}{D}, 0)$ with eigenvalues $-D$ and $R_0 - 1$, because only one eigenvalue can be zero when $R_0 = 1$; (2) at E_{1+} because it is a saddle for all parameter values according to the result from Theorem 2.3; and (3) at E_{1-} when $R_0 \geq 1$, since the determinant of its Jacobian is positive if the trace equals zero. Therefore, BT bifurcation can only occur at E_{1-} when $R_0 < 1$, at which the Hopf critical point coincides with the saddle-node point, i.e., $D_h = D_s$. For example, in Fig. 2d, D_h approaches and collides with D_s.

First, we derive the parameter conditions for the occurrence of double-zero eigenvalues, that is, $\mathrm{Tr}(J_1) = \det(J_1) = 0$; or the collision of the Hopf and saddle-node bifurcations, namely $(D_h, X_h) = (D_s, X_s)$. Noting from (26) that $\mathrm{Tr}(J_1) = 0$ yields $\Delta = 0$, we solve $\Delta = 0$ and $\mathrm{Tr}(J_1)|_{\mathrm{E}_{1-}} = 0$ for A and B to obtain

$$A = A_0 \equiv \frac{D(C+D)^2}{C(1-D)^2}, \quad B = B_0 \equiv \frac{D[C(1-2D)-D^2]}{C(1-D)^2}. \tag{58}$$

The positiveness of B_0 requires

$$C > \frac{D^2}{1-2D} \quad \text{and} \quad 0 < D < \frac{1}{2}. \tag{59}$$

We then define a two-dimensional parameter hypersurface,

$$\mathcal{S} : \left\{\gamma \mid A = A_0,\ B = B_0,\ C > \frac{D^2}{1-2D},\ 0 < D < \frac{1}{2}\right\}. \tag{60}$$

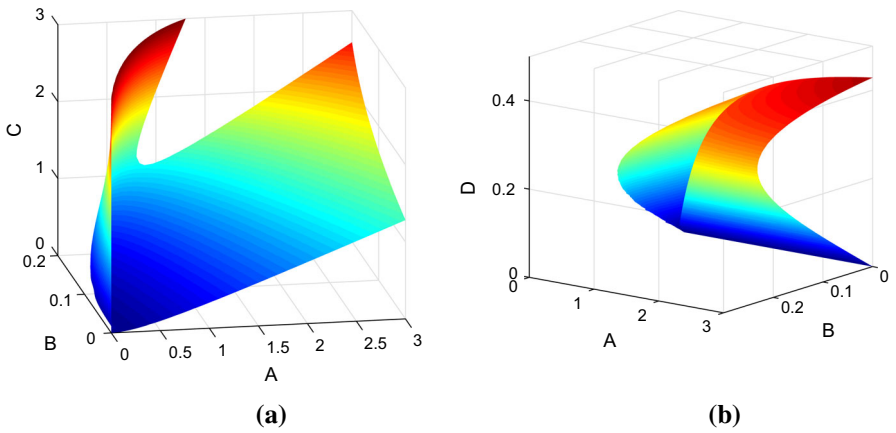

Fig. 15 Projection of the hypersurface $\mathcal{S}$ into **a** the A–B–C space; and **b** the A–B–D space (Color figure online)

System (5) undergoes a BT bifurcation if its paramerers $(A, B, C, D) \in \mathcal{S}$. The projections of $\mathcal{S}$ into the A–B–C space and A–B–D space are shown in Fig. 15a, b, respectively.

To analyze the BT bifurcation, the first step is usually to find the normal form of the dynamical system and then use the normal form to determine the codimension and unfolding of the BT bifurcation. The general idea of the derivation of normal forms is introducing nonlinear transforms on state variables together with necessary time rescaling. The derivation is standard for codimension-2 BT bifurcation at the BT bifurcation point without unfolding/bifurcation parameter. However, even for codimension-2 BT bifurcation, the normal form computation becomes much more complicated when bifurcation parameters are involved. For codimension-3 BT bifurcation, the computation burden is even heavier. The 6-step transformations approach developed by Dumortier et al. (1987) becomes a standard method and applied by researchers to find the parametric normal form for codimension-3 BT bifurcation. Some cases of codimension-4 BT bifurcation have been discussed in the literature (e.g., see Li and Rousseau 1989) by using a similar multiple-step transformation method to find the normal form. Multiple-step transformation method is tedious and yet hard to verify. Besides, the transformation between the original variables and the new variables in the last step of transformations is difficult to achieve. We are not aware of any work on developing one-step transformation and obtain explicit expressions for the transformation between the original variables and new variables. We have used the simplest normal theory (e.g., see Yu (1999); Yu and Leung (2003); Gazor and Yu (2010, 2012); Gazor and Moazeni (2015)) to develop a one-step transformation approach to find the parametric simplest normal form and associated transformations. Our new method not only gives the direct relation between original variables (including both state and parameter variables) but also yields results which can be easily verified. Programs based on a computer algebra system, Maple, and the new method and algorithm will be available in one of our forthcoming papers.

In order to determine the codimension of the BT bifurcation, we first find the equilibrium E_{1-}, under the conditions given in (58), as

$$\mathrm{E}_{1-} = \left(\tfrac{1}{D} - 1,\ D\right), \quad \text{for } D \in \left(0, \tfrac{1}{2}\right).$$

and then introducing the following change of state variables,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{1}{D} - 1 \\ D \end{pmatrix} + \begin{bmatrix} \frac{1}{D-1} & 0 \\ \frac{D}{1-D} & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{61}$$

into (5) we obtain

$$\frac{du_1}{d\tau} = u_2 + f(u_1, u_2), \qquad \frac{du_2}{d\tau} = f(u_1, u_2), \tag{62}$$

where

$$\begin{aligned} &f(u_1, u_2) \\ &= \frac{-[Du_1 + (1-D)u_2]\{D(1+C)[D(1-D)u_1 + Du_1^2 + (1-D)u_1u_2] - C(1-D)^3u_2\}}{(1-D)^3[(1-D)(C+D) + Du_1 + (1-D)u_2]} \end{aligned} \tag{63}$$

Now, expanding $f(u_1, u_2)$ in Taylor series around $(u_1, u_2) = (0, 0)$ and applying the simplest normal form theory, we introduce the following seventh-order near-identity nonlinear transformation,

$$u_1 = x_1 + \sum_{i+j=2}^{7} a_{ij} x_1^i x_2^j, \qquad u_2 = x_2 + \sum_{i+j=2}^{7} b_{ij} x_1^i x_2^j, \tag{64}$$

where a_{ij} and b_{ij} are constant coefficients, expressed in terms of C and D, and the time rescaling

$$\tau = \left[1 + \tfrac{C}{2(1-D)(C+D)}\, x_1 + x_1^3\right] \tau_1, \tag{65}$$

into (62) to obtain the following simplest normal form up to seventh-order terms:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}\tau_1} &= x_2 + \mathcal{O}(|(x_1, x_2)|^8), \\ \frac{\mathrm{d}x_2}{\mathrm{d}\tau_1} &= a_1 x_1^2 + a_2 x_1 x_2 + a_3 x_1^3 x_2 + a_4 x_1^4 x_2 + a_5 x_1^6 x_2 + \mathcal{O}(|(x_1, x_2)|^8). \end{aligned} \tag{66}$$

Comparing the above simplest normal form with the conventional normal form (e.g., see Guckenheimer and Holmes (1993),

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}\tau_1} &= x_2 + \mathcal{O}(|(x_1, x_2)|^{n+1}),\\ \frac{\mathrm{d}x_2}{\mathrm{d}\tau_1} &= \sum_{k=2}^{n} a_k\, x_1^k + b_k\, x_1^{k-1} x_2 + \mathcal{O}(|(x_1, x_2)|^{n+1}),\end{aligned} \tag{67}$$

it is seen that the simplest normal form (66) has less half of the terms in the conventional normal form (67).

Further, introducing the transformation,

$$x_1 \to x_1, \quad x_2 + \mathcal{O}(|(x_1, x_2)|^8) \to x_2,$$

into (66), we obtain

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}\tau_1} &= x_2,\\ \frac{\mathrm{d}x_2}{\mathrm{d}\tau_1} &= a_1\, x_1^2 + a_2\, x_1 x_2 + a_3\, x_1^3 x_2 + a_4\, x_1^4 x_2 + a_5\, x_1^6 x_2 + \mathcal{O}(|(x_1, x_2)|^8),\end{aligned} \tag{68}$$

where the coefficients a_i's are given in terms of C and D. In particular,

$$a_1 = -\tfrac{(1+C)D^3}{(C+D)(1-D)^3} \quad \text{and} \quad a_2 = \tfrac{D[C(1-3D)-D(1+D)]}{(C+D)(1-D)^3}. \tag{69}$$

It is obvious that $a_1 < 0$ due to $0 < D < \frac{1}{2}$, and it is easy to show that

$$\begin{aligned}&a_2 < 0 \quad \text{for } D \in \left[\tfrac{1}{3}, \tfrac{1}{2}\right), \text{ or if } C \in \left(\tfrac{D^2}{1-2D}, \tfrac{D(1+D)}{1-3D}\right) \text{ for } D \in \left(0, \tfrac{1}{3}\right),\\ &a_2 > 0 \quad \text{if } C > \tfrac{D(1+D)}{1-3D} \text{ for } D \in \left(0, \tfrac{1}{3}\right),\\ &a_2 = 0 \quad \text{if } C = \tfrac{D(1+D)}{1-3D} \text{ for } D \in \left(0, \tfrac{1}{3}\right).\end{aligned} \tag{70}$$

Therefore, if $a_2 \neq 0$, i.e., when $\frac{1}{3} \leq D < \frac{1}{2}$ or $0 < D < \frac{1}{3}$ with $C \neq \frac{D(1+D)}{1-3D}$, the BT bifurcation is codimension 2.

When $C = C_0 = \frac{D(1+D)}{1-3D}$, $(0 < D < \frac{1}{3})$, $a_2 = 0$, and a_1, a_3, a_4 and a_5 become

$$\begin{aligned}&a_1 = -\tfrac{D^2}{2(1-D)^2}, \quad a_3 = \tfrac{(1-4D)D(1+D)}{8(1-D)^5}, \quad a_4 = -\tfrac{(1-4D)D(1+D)(1-11D)}{64(1-D)^7},\\ &a_5 = \tfrac{(1-4D)D(68480D^5+3206517D^4+2472332D^3-837530D^2-107396D+60077)}{7741440(1-D)^{11}}.\end{aligned} \tag{71}$$

It can be seen that $a_1 < 0$, and $a_3 \neq 0$ for $D \neq \frac{1}{4}$. More precisely, we have

$$a_3 = \tfrac{(1-4D)D(1+D)}{8(1-D)^5} \begin{cases} > 0 & \text{if } 0 < D < \frac{1}{4},\\ = 0 & \text{if } D = \frac{1}{4},\\ < 0 & \text{if } \frac{1}{4} < D < \frac{1}{3}.\end{cases} \tag{72}$$

Thus, when $D \in (0, \frac{1}{4}) \bigcup (\frac{1}{4}, \frac{1}{3})$, the BT bifurcation is codimension 3.

When $D = \frac{1}{4}$, $a_3 = a_4 = a_5 = 0$. In fact, at this critical value, we obtain $C_0 = \frac{5}{4}$, yielding $A_0 = \frac{4}{5}$ and $B_0 = \frac{5}{4}$, which satisfy the center condition (56). That is, under these parameter values, system (5) is an integral system with the first integral (57), having a nilpotent point at $(X, Y) = (3, \frac{1}{4})$. We shall not discuss this special case in this paper. Finally, note that a necessary condition for the existence of the equilibrium E_{1-} is $\Delta > 0$, which is equivalent to

$$\Delta_1 = A + B - D - BC - 2\sqrt{C(A+B)(D-B)} > 0. \tag{73}$$

Summarizing the above results we have the following theorem.

Theorem 4.1 *For system (5), when $R_0 < 1$ (i.e., $D > B$) and $\Delta_1 > 0$ (so $A > BC$), BT bifurcation occurs from the endemic equilibrium $\mathrm{E}_{1-} : (\frac{1}{D} - 1, D)$ at the critical point $(A, B) = (A_0, B_0) = \big(\frac{D(C+D)^2}{C(1-D)^2}, \frac{D[C(1-2D)-D^2]}{C(1-D)^2}\big)$, with $C > \frac{D^2}{1-2D}$, $D \in (0, \frac{1}{2})$. Moreover, the BT bifurcation is*

(i) *codimension 2 if $D \in [\frac{1}{3}, \frac{1}{2})$ or $D \in (0, \frac{1}{3})$ with $C \neq \frac{D(1+D)}{1-3D}$; and*
(ii) *codimension 3 if $C = \frac{D(1+D)}{1-3D}$ with $D \in (0, \frac{1}{4}) \bigcup (\frac{1}{4}, \frac{1}{3})$.*

In the following two subsections, we will consider the two BT bifurcations with condimensions 2 and 3, respectively.

4.1 Codimension-2 BT Bifurcation

Suppose the condition (i) in Theorem 4.1 holds, under which $a_2 \neq 0$. To obtain the normal form with unfolding, we introduce the parameter transformation,

$$A = A_0 + \mu_1, \qquad B = B_0 + \mu_2, \tag{74}$$

together with the change of state variables (61), into (5) and expanding the resulting system around the critical point $(u_1, u_2, \mu_1, \mu_2) = (0, 0, 0, 0)$ yields

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}\tau} &= u_2 + f_2(u_1, u_2, \mu) + \mathcal{O}(|(u_1, u_2, \mu)|^3), \\ \frac{\mathrm{d}u_2}{\mathrm{d}\tau} &= f_2(u_1, u_2, \mu) + \mathcal{O}(|(u_1, u_2, \mu)|^3), \end{aligned} \tag{75}$$

where $\mu = (\mu_1, \mu_2)$ and f_2 as follows:

$$\begin{aligned} f_2 = {} & (1-D)^2 \left[\tfrac{D}{C+D}\,\mu_1 + \mu_2\right] \\ & + \left[\tfrac{D[CD+(2C+D)(1-2D)]}{(C+D)^2}\,\mu_1 + (1-2D)\,\mu_2\right] u_1 \\ & + (1-D)^2 \left[\tfrac{2C+D}{(C+D)^2}\,\mu_1 + \tfrac{1}{D}\,\mu_2\right] u_2 - \tfrac{(1+C)\,D^3}{(C+D)(1-D)^3}\,u_1^2 \\ & + \tfrac{D[C(1-2D)-D]}{(1-D)^2(C+D)}\,u_1 u_2 + \tfrac{C}{C+D}\,u_2^2. \end{aligned}$$

It should be pointed out that even the recurrence infection (viral blips) phenomenon has been studied in Zhang et al. (2013, 2014a, b), no detailed dynamical analysis is given for the two-dimensional model (5). In our later work (Yu et al. 2016), BT bifurcation is discussed with two pairs of parameter values on B and D, that is, $(B, D) = (\frac{27}{500}, \frac{57}{1000})$ and $(\frac{27}{500}, \frac{87}{1000})$. Therefore, the BT critical point is completely determined by the parameters $A = A(B, D)$ and $C = C(B, D)$. In fact, only two BT critical points on $\mathcal{S}$ are considered in (Yu et al. 2016):

$$(A, B, C, D) = \left(\tfrac{3078507}{206879500}, \tfrac{27}{500}, \tfrac{61731}{827518}, \tfrac{57}{1000}\right), \left(\tfrac{118428267}{2237439500}, \tfrac{27}{500}, \tfrac{219501}{8949758}, \tfrac{57}{1000}\right).$$

The dynamical behaviors from Yu et al. (2016) is limited. To get a global dynamical behavior in the whole parameter space, this paper studies the whole feasible parameter region near $\mathcal{S}$. Instead of fixing A, C and D in Zhang et al. (2013, 2014a, b) or B and D in Yu et al. (2016), this paper studies all parameter set $(A, B, C, D) \in \mathcal{S}$ in (60). In addition, bifurcation diagrams are given in terms of parameters $A = A_0(C, D)$ and $B = B_0(C, D)$ to show the global influence of control parameters C and D on the system's dynamical behavior.

Next, we apply the parametric simplest normal form theory (Yu and Leung 2003; Gazor and Yu 2010, 2012; Gazor and Moazeni 2015) to get the parametric normal form (up to second order) via the transformations of state variables and parameters. Note that unlike the series of transformations applied in many articles in the literature, here we only need to use one transformation to obtain the normal form. In general, the transformation is not unique, but in order to obtain the simplest parameterization, we use the following change of state variables,

$$\begin{aligned}
u_1 = \tfrac{1}{a_1}\Big[&-y_1 - \tfrac{(1-D)^2[D+C(1-D)]}{2(1+C)D^3}\,\xi_1 + \tfrac{D^2(1+C)}{(1-D)(C+D)}\,\xi_2 - \tfrac{C(2-3D)+D(1-2D)}{D(C+D)}\,\xi_2 y_1 \\
&- \tfrac{(1-D)[D(C+D)-C(1-D^2)]}{2D^3(1+C)}\,y_1^2 + \tfrac{(1-D)^3[C^2(4D^2+D+1)+CD(3D+4)+D^2]}{6D^5(1+C)^2}\,y_1 y_2 \\
&- \tfrac{(1-D)^3[C^2(3D^3-3D^2+7D-3)+2CD(D^2+4D-2)+D^2(3D-1)]}{8D^6(1+C)^2}\,y_2^2\Big], \\
u_2 = \tfrac{1}{a_1}\Big[&-y_2 + \xi_1 - \tfrac{(1-D)^3[C^2(D^3+D^2-9D+3)+4CD(D^2-4D+1)+D^2(1-5D)]}{4D^6(1+C)^2}\,\xi_1(\xi_1 - y_2) \\
&+ \tfrac{D(1-3D)+2C(1-2D)}{D(C+D)}\,\xi_2(\xi_1 - y_2) + y_1^2 + \tfrac{(1-D)^3C}{D^3(1+C)}\,y_1 y_2 \\
&- \tfrac{(1-D)^3[C^2(2D^2-D-1)+CD(3D-4)-D^2]}{6D^5(1+C)^2}\,(\xi y_1 + y_2^2)\Big],
\end{aligned} \tag{76}$$

and then obtain the following simplest parameterization,

$$\begin{aligned}
\mu_1 &= -\tfrac{(1-D)(C+D)^2}{D^4(1+C)}\,\xi_1 + \tfrac{2D(1+C)(C+D)}{C(1-D)^2}\,\xi_2, \\
\mu_2 &= -\tfrac{2(1+C)D^2}{C(1-D)^2}\,\xi_2 - \tfrac{D(1+C)}{(1-D)(C+D)}\,\xi_2^2.
\end{aligned} \tag{77}$$

Implicit function theorem implies that local perturbations on ξ_1 and ξ_2 near the critical poin $(\xi_1, \xi_2) = (0, 0)$ is topologically equivalent to local perturbations on μ_1 and μ_2

near $(\mu_1, \mu_2) = (0, 0)$ since

$$\det\left[\frac{\partial(\mu_1,\mu_2)}{\partial(\xi_1,\xi_2)}\right] = \frac{2(C+D)^2}{CD^2(1-D)} + \mathcal{O}(\xi) \neq 0 \ \text{ for small } \ \xi = (\xi_1, \xi_2).$$

Under the above transformations (76) and (77), (75) becomes

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}\tau} &= y_2 + \mathcal{O}(|(y_1, y_2, \xi)|^3),\\ \frac{\mathrm{d}y_2}{\mathrm{d}\tau} &= \xi_1 + \xi_2\, y_2 + y_1^2 - \tfrac{a_2}{a_1}\, y_1 y_2 + \mathcal{O}(|(y_1, y_2, \xi)|^3),\end{aligned}$$

where a_1 and a_2 are given in (69). To remove the higher-order terms $O(|(y_1, y_2, \xi)|^3)$ in the first equation of the above equations, we introduce an additional transformation:

$$y_1 = x_1, \quad y_2 + \mathcal{O}(|(y_1, y_2, \xi)|^3) = x_2,$$

into the above equations to obtain the normal form with unfolding up to second-order terms:

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}\tau} &= x_2,\\ \frac{\mathrm{d}x_2}{\mathrm{d}\tau} &= \xi_1 + \xi_2\, x_2 + x_1^2 - \tfrac{a_2}{a_1}\, x_1 x_2 + \mathcal{O}(|(x_1, x_2, \xi)|^3),\end{aligned} \tag{78}$$

where the coefficient $-\frac{a_2}{a_1}$ not being normalized to ± 1 is for us to easily see the effect of the original parameters C and D. Note that $-\frac{a_2}{a_1}$ has the same sign of a_2 due to $a_1 < 0$.

Now, we use the normal form (78) to analyze the codimension-2 BT bifurcation. Note that the normal form (78) is in the standard form given in Guckenheimer and Holmes (1993). Thus, we follow the approach described in Guckenheimer and Holmes (1993) to obtain the following theorem.

Theorem 4.2 *For system* (5)*, with the conditions given in Theorem* 4.1*, codimension-*2 *BT bifurcation occurs for* $D \in [\frac{1}{3}, \frac{1}{2})$*, or for* $D \in (0, \frac{1}{3})$ *with* $C \neq \frac{D(1+D)}{1-3D}$*. Moreover, three local bifurcations, with the representations of the bifurcation curves, are given below.*

(1) *Saddle-node bifurcation occurs from the bifurcation curve:*

$$\mathrm{SN} = \big\{(\xi_1, \xi_2) \mid \xi_1 = 0\big\}.$$

(2) *Hopf bifurcation occurs from the bifurcation curve:*

$$\mathrm{H} = \Big\{(\xi_1, \xi_2) \mid \xi_1 = -\big(\tfrac{a_1}{a_2}\big)^2 \xi_2^2\Big\} \begin{cases} \xi_2 > 0\ (a_2 < 0), \text{ subcritical},\\ \xi_2 < 0\ (a_2 > 0), \text{ supercritical}.\end{cases}$$

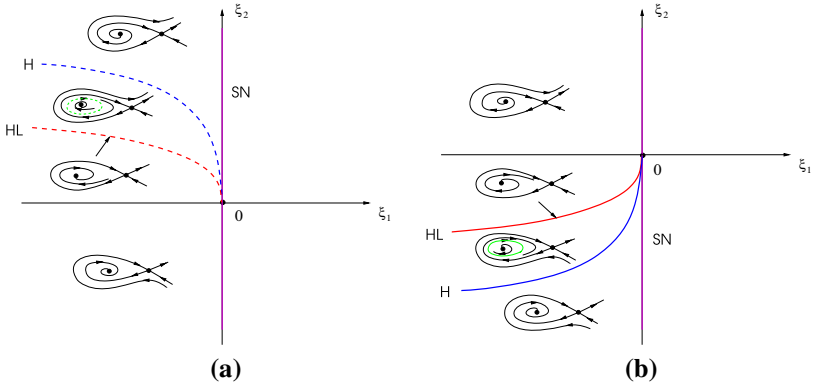

Fig. 16 Bifurcation diagrams for the codimension-2 BT bifurcation based on the normal form (78): **a** $a_2 < 0$; and **b** $a_2 > 0$

(3) *Homoclinic orbit occurs from the bifurcation curve:*

$$\mathrm{HL} = \left\{(\xi_1, \xi_2) \mid \xi_1 = -\tfrac{49}{25}\left(\tfrac{a_1}{a_2}\right)^2 \xi_2^2\right\} \begin{cases} \xi_2 > 0 \ (a_2 < 0), \text{ unstable}, \\ \xi_2 < 0 \ (a_2 > 0), \text{ stable}. \end{cases}$$

The bifurcation diagrams for the codimsion-2 BT bifurcation are shown in Fig. 16. Note that the above formulas for bifurcation curves given in terms of (ξ_1, ξ_2) can be expressed in terms of μ_1 and μ_2 for the original system (5) by using (77).

The homoclinic loop in the BT bifurcation studied above implies a new mechanism to generate the recurrence phenomenon, which is completely different from what discussed in the previous section, based on the four conditions P_1–P_4 in Sect. 3.3. Here, it is noted that the trajectory starting from a point on the homoclinic loop will approach the saddle point either as $\tau \to +\infty$ or $\tau \to -\infty$. Consider the case that a limit cycle near the homoclinic bifurcation curve is stable (see Fig. 16b), we can see that the limit cycle moves extremely slowly near the saddle point but moves fast when it is away. This yields a slow-fast motion which generates recurrence behavior. The first three conditions for this mechanism are same as P_1-P_3, but the last condition should be modified as

P_4^*: there exists Bogdanov–Takens bifurcation, leading to homoclinic loops near a Hopf bifurcation.

We notice that the new recurrence-generating mechanism offers a small amplitude of oscillation, or extremely slow convergence to the disease-free equilibrium. The big difference between the mechanism discussed in the previous section and the new mechanism is that the former yields very large oscillations in both amplitude and frequency, while the latter only has significant change in frequency, but very little variation in the amplitude. Patients with chronic infections may undergo a short-period of worse symptom.

To end this subsection, we present simulations for the codimension-2 BT bifurcation. It has been shown that the BT bifurcation can occur only for $R_0 < 1$ (i.e., for

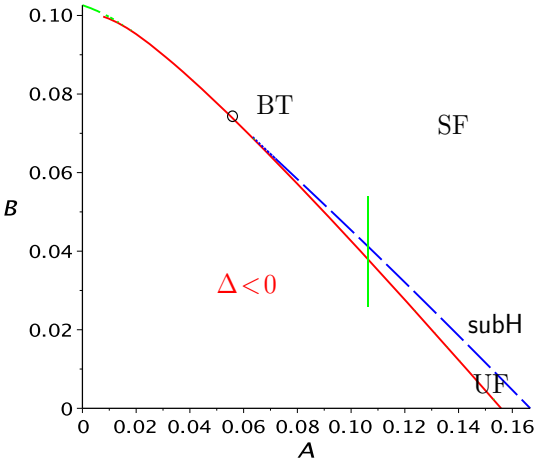

Fig. 17 Codimension-2 BT bifurcation diagram around the critical point $(A, B, C, D) = (\frac{1}{18}, \frac{2}{27}, \frac{1}{20}, \frac{1}{10})$

$D < B$). It should be also noted that in the neighborhood of the codimension-2 BT critical point, there exists only one limit cycle from the Hopf bifurcation. To give a more direct impression, here we shall use the original equation (5), rather than the normal form (78), to perform the simulation. Since the perturbation parameters A and B are given in terms of C and D, we choose $(C, D) = (\frac{1}{20}, \frac{1}{10})$, then obtain the BT critical point as $(A, B) = (\frac{1}{18}, \frac{2}{27})$, then further have $a_1 = -\frac{7}{729}$ and $a_2 = -\frac{50}{729}$, and $\frac{a_2}{a_1} = \frac{50}{7}$. The bifurcation diagram near the above critical point is shown in Fig. 17.

We can take perturbations on A and B such that A, B pass through the blue curve from above, see Fig. 17. To investigate the effect of B on the BT bifurcation, we let $A = \frac{1}{18} + \mu_1$, $B = \frac{2}{27} + \mu_2$, and fix $\mu_1 = \frac{1}{20}$.

Taking

$$\mu_2 = -0.028, \quad -0.028301496, \quad -0.03, \quad -0.034, \tag{79}$$

the four corresponding points $(A, B, C, D) \in \mathcal{S}$ locate from top to the bottom on the blue vertical line segment in Fig. 17. Four corresponding simulations are shown in Fig. 18a–d, respectively. We can use (77) to obtain ξ_1 and ξ_2 expressed in terms of μ_2 as follows:

$$\xi_1 = -\frac{7}{27000} + \frac{7}{1500}\,\mu_2 - \frac{98}{675}\,\mu_2^2, \quad \xi_2 = -\frac{9}{7}\,\mu_2 - 40\,\mu_2^2.$$

The simulations given in Fig. 17a–d correspond to the four phase portraits shown in Fig. 18a. However, note that stability is reversed since the transformation given in (61) yields

$$\det\begin{bmatrix} \frac{1}{D-1} & 0 \\ \frac{D}{1-D} & 1 \end{bmatrix} = \frac{1}{D-1} < 0, \quad \text{for } D \in \left(0, \tfrac{1}{2}\right),$$

and thus the Hopf and homoclinic bifurcations exchange their positions. Moreover, for the parameter values $(A, B, C, D) = (\frac{1}{18}, \frac{2}{27}, \frac{1}{20}, \frac{1}{10})$, we can use (61), (76) and

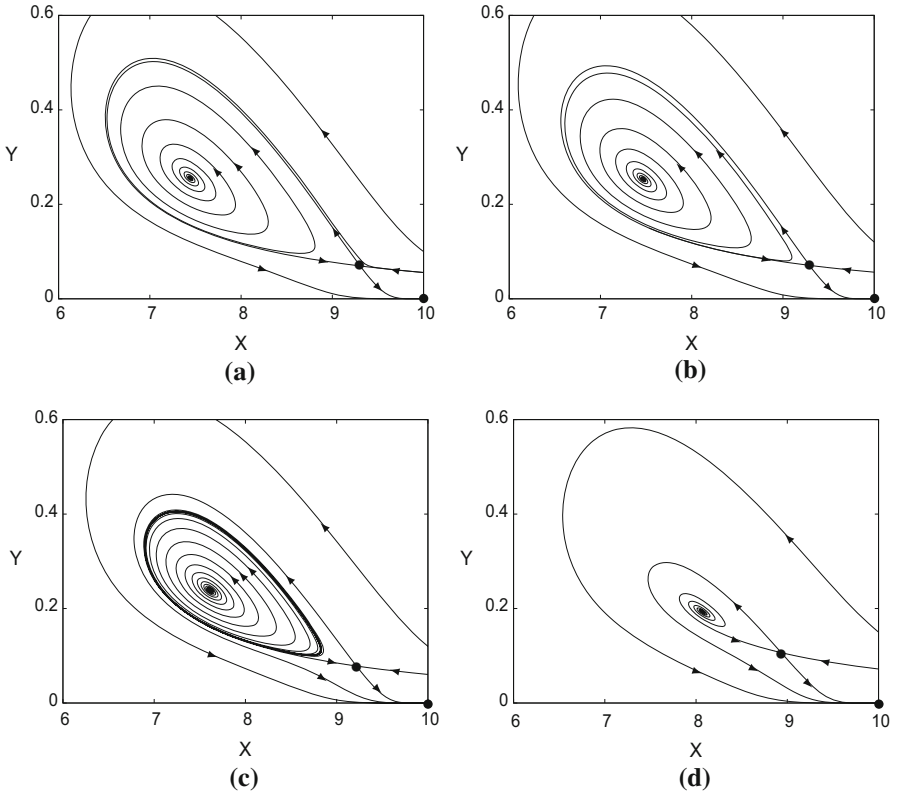

Fig. 18 Simulations of system (5) when $A = \frac{19}{180}$, $C = \frac{2}{27}$ and $D = \frac{1}{10}$ for **a** $B = 0.0460740741$, showing convergence of the half-unstable manifold of the saddle point E_{1+} to the stable focus E_{1-}; **b** $B = 0.0457725781$, showing an unstable homoclinic loop enclosing a stable focus; **c** $B = 0.0440740741$, showing stable focus E_{1-} enclosed by an unstable limit cycle; and **d** $B = 0.0400740741$, showing an unstable focus E_{1-} with a trajectory connecting to half-stable manifold of the saddle point E_{1+}

(77) to find the coordinates of the focus E_{1-} and the saddle E_{1+} in the original system (5), corresponding to the focus $(x_1, x_2) = (-\sqrt{-\xi_1}, 0)$ and the saddle $(x_1, x_2) = (\sqrt{-\xi_1}, 0)$, respectively, as

$$\begin{aligned}(X_-, Y_-) \approx \Big(&\tfrac{963}{140} - \tfrac{9\sqrt{210}}{70} + \big(\tfrac{324}{7} - \tfrac{783\sqrt{210}}{980}\big)\mu_2,\ \tfrac{8011729}{28000000}\\ &+ \tfrac{14193\sqrt{210}}{612500} - \big(\tfrac{717741}{200000} + \tfrac{94149\sqrt{210}}{245000}\big)\mu_2\Big),\\ (X_+, Y_+) \approx \Big(&\tfrac{963}{140} + \tfrac{9\sqrt{210}}{70} + \big(\tfrac{324}{7} - \tfrac{783\sqrt{210}}{980}\big)\mu_2,\ \tfrac{8011729}{28000000}\\ &+ \tfrac{14193\sqrt{210}}{612500} - \big(\tfrac{717741}{200000} - \tfrac{94149\sqrt{210}}{245000}\big)\mu_2\Big),\end{aligned}$$

which shows that

$$X_+ - X_- = \tfrac{9\sqrt{210}}{490}\big(14 + 87\mu_2\big) > 0, \quad Y_- - Y_+ = \tfrac{9\sqrt{210}}{612500}\big(3154 - 52305\mu_2\big) > 0,$$

for the values of μ_2 given in (79), as expected.

4.2 Codimension-3 BT Bifurcation

In this section, we consider the codimension-3 BT bifurcation under condition (ii) in Theorem 4.1. To find the parametric normal form of BT bifurcation, a widely used method was developed by Dumortier et al. (1987) which involves transformation with six steps. In this paper, we present a one-step transformation method to reach the parametric normal form. This novel method involves transformations on state variables and parameters, as well as a time rescaling. To achieve this, let

$$A = \tfrac{4D^2}{(1+D)(1-3D)} + \mu_1, \quad B = \tfrac{D}{1+D} + \mu_2, \quad C = \tfrac{D(1+D)}{1-3D} + \mu_3, \tag{80}$$

where $D \in (0, \frac{1}{4}) \bigcup (\frac{1}{4}, \frac{1}{3})$. We denote $\mu = (\mu_1, \mu_2, \mu_3)$ and then apply (80) together with (61) into (5) and expand the resulting system around the point $(u_1, u_2, \mu) = (0, 0, 0)$ to obtain the following system:

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}\tau} &= u_2 + f_4(u_1, u_2, \mu) + \mathcal{O}(|(u_1, u_2, \mu)|^5), \\ \frac{\mathrm{d}u_2}{\mathrm{d}\tau} &= f_4(u_1, u_2, \mu) + \mathcal{O}(|(u_1, u_2, \mu)|^5), \end{aligned} \tag{81}$$

where

$$\begin{aligned} f_4 = {} & \mathcal{F}_{400}(\mu) + \mathcal{F}_{310}(\mu)\, u_1 + \mathcal{F}_{301}(\mu)\, u_2 + \mathcal{F}_{220}(\mu)\, u_1^2 + \mathcal{F}_{311}(\mu)\, u_1 u_2 \\ & + \mathcal{F}_{302}(\mu)\, u_2^2 + \mathcal{F}_{130}(\mu)\, u_1^3 + \mathcal{F}_{121}(\mu)\, u_1^2 u_2 + \mathcal{F}_{112}(\mu)\, u_1 u_2^2 + \mathcal{F}_{103}(\mu)\, u_2^3 \\ & + \tfrac{(1-3D)(1+D)}{8(1-D)^3} \Big[\tfrac{D^2}{(1-D)^3}\, u_1^4 + \tfrac{2(2-3D)}{(1-D)^3}\, u_1^3 u_2 + \tfrac{6(1-2D)}{(1-D)^2}\, u_1^2 u_2^2 \\ & + \tfrac{2(2-5D)}{D(1-D)}\, u_1 u_2^3 + \tfrac{(1-3D)}{D^2}\, u_2^4 \Big], \end{aligned}$$

in which $\mathcal{F}_{kij}$ are kth-degree polynomials in μ. Then, we employ the parametric simplest normal form theory (Yu and Leung 2003; Gazor and Yu 2010, 2012; Gazor and Moazeni 2015) to obtain the parametric normal form. More precisely, we apply the following state variables transformation:

$$\begin{aligned} u_1 = {} & -\tfrac{2(1-D)^2}{D^2}\, y_1 + \tfrac{(1-D)^2(2D^2+3D-3)}{2D^4}\, \xi_1 + \tfrac{(1-D)^2}{D}\, \xi_2 \\ & + \sum_{i+j+k+l+m=2,3,4} \mathcal{U}_{1ijklm}\, y_1^i y_2^j \xi_1^k \xi_2^l \xi_3^m, \\ u_2 = {} & -\tfrac{2(1-D)^2}{D^2}\, y_2 + \tfrac{2(1-D)^2}{D^2}\, \xi_1 + \sum_{i+j+k+l+m=2,3,4} \mathcal{U}_{2ijklm}\, y_1^i y_2^j \xi_1^k \xi_2^l \xi_3^m, \end{aligned} \tag{82}$$

and a time rescaling:

$$\tau \ \to \ \left[1 - \tfrac{1+D}{2D^2}\, y_1 + \tfrac{1}{6D}\, \xi_2 + \tfrac{D}{3(1+D)}\, \xi_3\right]\tau, \tag{83}$$

then obtain the parametric normal form,

$$\begin{aligned}
\frac{\mathrm{d}y_1}{\mathrm{d}\tau} &= y_2 + \mathcal{O}(|y_1, y_2, \xi|^5),\\
\frac{\mathrm{d}y_2}{\mathrm{d}\tau} &= \xi_1 + \xi_2\, y_2 + \xi_3\, y_1 y_2 + y_1^2 - b_1\, y_1^3 y_2 + \mathcal{O}(|(y_1, y_2, \xi)|^5),
\end{aligned} \tag{84}$$

where $b_1 = \frac{(1-4D)(1-D^2)}{D^5}$ for $D \in (0, \frac{1}{4}) \bigcup (\frac{1}{4}, \frac{1}{3})$, and the parametrization is given by

$$\begin{aligned}
\mu_1 &= \tfrac{4(1-D)^2(8D^2-3D-1)}{D^2(1+D)(1-3D)^2}\, \xi_1 + \tfrac{4D(1-D)}{(1+D)(1-3D)^2}\, \xi_2 - \tfrac{8D^4(1-D)}{(1+D)^2(1-3D)^2}\, \xi_3,\\
&\quad + \sum_{i+j+k=2,\,3} \mathcal{M}_{1ijk} \xi_1^i \xi_2^j \xi_3^k,\\
\mu_2 &= -\tfrac{(1-D)(1-4D)}{D^2(1+D)}\, \xi_1 - \tfrac{D}{2(1+D)}\, \xi_2 - \tfrac{D^3}{(1+D)^2}\, \xi_3 + \sum_{i+j+k=2,3} \mathcal{M}_{2ijk} \xi_1^i \xi_2^j \xi_3^k,\\
\mu_3 &= -\tfrac{(1+D)(1-D)^2(1-D+4D^2)}{D^3(1-3D)^2}\, \xi_1 + \tfrac{3(1+D)(1-D)^2}{2(1-3D)^2}\, \xi_2 - \tfrac{D^2(1-D)^2}{(1-3D)^2}\, \xi_3\\
&\quad + \sum_{i+j+k=2,3} \mathcal{M}_{3ijk} \xi_1^i \xi_2^j \xi_3^k - \tfrac{2D^5(72D^5-36D^4+33D^3-3D^2+15D-1)(1-D)^2}{9(1+D)^3/(1-3D)^4}\, \xi_3^4.
\end{aligned} \tag{85}$$

Finally, introducing the transformation:

$$y_1 = x_1, \quad y_2 + \mathcal{O}(|(y_1, y_2, \xi)|^5) = x_2,$$

into (84) yields the normal form with the unfolding up to fourth-order terms:

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}\tau} &= x_2,\\
\frac{\mathrm{d}x_2}{\mathrm{d}\tau} &= \xi_1 + \xi_2\, x_2 + \xi_3\, x_1 x_2 + x_1^2 - b_1\, x_1^3 x_2 + \mathcal{O}(|(x_1, x_2, \xi)|^5).
\end{aligned} \tag{86}$$

Note that the coefficient of y_1 in the time rescaling (83), $-\frac{1+D}{2D^2}$, is actually equivalent to the coefficient of x_1 in the time rescaling (65), $\frac{C}{2(1-D)(C+D)} = \frac{1+D}{4(1-D)^2}$, because there is a scaling $-\frac{2(1-D)^2}{D^2}$ for u_1 and u_2 in (82). In fact, $\frac{1+D}{4(1-D)^2} \times \left(-\frac{2(1-D)^2}{D^2}\right) = -\frac{1+D}{2D^2}$. For the same reason, we see that $b_1 = \frac{(1-4D)(1-D)^2}{D^5} = -\frac{a_3}{a_1} \times \left(-\frac{2(1-D)^2}{D^2}\right)^2$, where a_1 and a_3 are given in (71). Thus, due to $a_1 < 0$, b_1 has the same sign of a_3, which is positive for $D \in (0, \frac{1}{4})$ and negative for $D \in (\frac{1}{4}, \frac{1}{3})$. To examine the influence of D on the system's dynamics, we use $b_1 = \frac{(1-4D)(1-D)^2}{D^5}$ without normalization.

Further, it is easy to use (85) to verify that

$$\det\left[\frac{\partial(\mu_1,\mu_2,\mu_3)}{\partial(\xi_1,\xi_2,\xi_3)}\right] = -\frac{8D(1-D)^3}{(1+D)(1-3D)^3} + \mathcal{O}(\xi) \neq 0 \ \text{ in } D \in (0, \tfrac{1}{3}) \ \text{ for small } \xi,$$

which shows that near the critical point $(A, B, C) = (A_0, B_0, C_0)$, system (5) has the same bifurcation set with respect to μ as system (86) has with respect to ξ, up to a homeomorphism in the parameter space.

Now, based on the normal form (86), we can follow a similar procedure used for analyzing the codimension-2 BT bifurcation to obtain the Hopf bifurcation surface and homoclinic bifurcation surface in the ξ_1–ξ_2–ξ_3 parameter space. To achieve this, we first find the two equilibrium solutions $\mathrm{E}_\pm$ as

$$\mathrm{E}_\pm = (x_{1\pm}, 0), \quad \text{where} \quad x_{1\pm} = \pm\sqrt{-\xi_1} \quad \text{for} \quad \xi_1 < 0. \tag{87}$$

The Jacobian of (86) evaluated at $\mathrm{E}_\pm$ is given by

$$J_\pm = \begin{bmatrix} 0 & 1 \\ 2x_{1\pm} & \xi_2 + \xi_3 x_{1\pm} - b_1 x_{1\pm}^3 \end{bmatrix}, \tag{88}$$

which indicates E_{1+} is a saddle, and E_{1-} is focus or node. It is easy to see that the plane

$$\mathrm{SN} = \big\{(\xi_1, \xi_2, \xi_3) \mid \xi_1 = 0\big\}, \tag{89}$$

excluding the origin in the parameter space is the saddle-node bifurcation surface.

Next, we consider Hopf bifurcation and generalized Hopf bifurcation from which multiple limit cycles can occur. We have the following result.

Theorem 4.3 *For system* (86), *Hopf bifurcation occurs from the equilibrium solution* E_- *at any point on the critical surface, defined by*

$$\mathrm{H} = \Big\{(\xi_1, \xi_2, \xi_3)\Big| \, \xi_2 - (\xi_3 + b_1\xi_1)\sqrt{-\xi_1} = 0\Big\}. \tag{90}$$

The generalized Hopf bifurcation occurs from the equilibrium solution E_- *at any point on the critical line which is the intersection of the critical surface* H *and the generalized critical surface* GH, *defined by*

$$\mathrm{GH} = \Big\{(\xi_1, \xi_2, \xi_3)\Big| \, \xi_3 + 3b_1\,\xi_1 = 0\Big\}, \tag{91}$$

yielding maximal two small-amplitude limit cycles, and outer one is stable (*unstable*) *if* $b_1 > 0$ ($b_1 < 0$).

Proof It is easy to see from the Jacobian J_- given in (88) that Hopf bifurcation can occur from the equilibrium E_{1-} at any point on the critical surface, defined by $\xi_2 + \xi_3 x_{1-} - b_1 x_{1-}^3 = 0$, which can be written in the form of (90). Note that ignoring the term

$b_1\xi_1$ and letting $\xi_3 = -\frac{a_2}{a_1}$ in (90) yields the Hopf bifurcation curve, $\xi_1 = -(\frac{a_1}{a_2})^2\xi_2^2$, given in Theorem 4.2, as expected.

To consider the generalized Hopf bifurcation and possible multiple limit cycles bifurcating from the Hopf critical surface (90), we need to find the focus values. Introducing the transformation,

$$x_1 = -\sqrt{-\xi_1} + w_1, \quad x_2 = \omega_c\, w_2, \quad \text{where} \quad \omega_c = \sqrt{2\sqrt{-\xi_1}}, \quad (\xi_1 < 0),$$

into (86) yields the following system:

$$\begin{aligned} \frac{\mathrm{d}w_1}{\mathrm{d}\tau} &= \omega_c\, w_2, \\ \frac{\mathrm{d}w_2}{\mathrm{d}\tau} &= -\omega_c\, w_1 + \tfrac{1}{\omega_c}\, w_1^2 + \big(\xi_3 - \tfrac{3b_1}{4}\,\omega_c^4\big)\, w_1 w_2 + \tfrac{3b_1}{2}\omega_c^2 w_1^2 w_2 - b_1\, w^3 w_2. \end{aligned} \tag{92}$$

Then applying the Maple program (Yu 1998) we obtain the focus values:

$$v_1 = \tfrac{1}{32\omega_c^2}(4\xi_3 + 3b_1\omega_c^4) = \tfrac{1}{8\omega_c^2}(\xi_3 - 3b_1\xi_1), \quad v_2|_{v_1=0} = -\tfrac{5b_1}{48\omega_c^2} \neq 0.$$

It is easy to find that $v_1 = 0$ leads to

$$\xi_3 = -\,3b_1\,\xi_1, \quad (\xi_1 < 0),$$

which defines the generalized Hopf critical surface (91). Since when $v_1 = 0$, $v_2 \neq 0$, at most two limit cycles can bifurcate from the E_{-1}. Moreover, simple perturbations on ξ_3 for v_1 and ξ_2 for v_0 yield $|v_0| \ll |v_1| \ll |v_2|$, implying that two limit cycles do occur, with the outer one stable (unstable) if $b_1 > 0$ ($b_1 < 0$). □

It should be noted that the above result on the number of small-amplitude limit cycles bifurcating from the Hopf critical surface (90), near the codimension-3 BT bifurcation point, agrees with that for the general original system (5), given in Theorems 3.3 (for $R_0 = 1$) and 3.5 (for $R_0 \neq 1$), but the proof here is much simpler.

Now, we turn to consider the homoclinic bifurcation surface as well as the degenerate homoclinic bifurcation points on the surface by following the method described in Han and Yu (2012). For this part, we have the following theorem.

Theorem 4.4 *For system* (86)*, homoclinic bifurcation can occur from the critical surface, defined by*

$$\mathrm{HL} = \Big\{(\xi_1, \xi_2, \xi_3)\Big|\, \xi_2 - \tfrac{5}{7}\,\big(\xi_3 + \tfrac{103}{55}\, b_1\,\xi_1\big)\sqrt{-\xi_1} = 0\Big\}, \tag{93}$$

and the degenerate homoclinic bifurcation occurs from any point on the critical line which is the intersection of the critical surface HL *and the degenerate critical surface* DHL*, defined by*

$$\mathrm{DHL} = \Big\{(\xi_1, \xi_2, \xi_3)\Big|\, \xi_2 + \big(\xi_3 + b_1\,\xi_1\big)\sqrt{-\xi_1} = 0\Big\}, \tag{94}$$

leading to bifurcation of two limit cycles.

Proof We use the method given in Han and Yu (2012) to prove the theorem. We first introduce the following scaling:

$$x_1 = \varepsilon^{\frac{2}{5}} w_1, \quad x_2 = \varepsilon^{\frac{3}{5}} w_2, \quad \xi_1 = \varepsilon^{\frac{4}{5}} \nu_1, \quad \xi_2 = \varepsilon^{\frac{6}{5}} \nu_2,$$
$$\xi_1 = \varepsilon^{\frac{4}{5}} \nu_3, \quad \tau_1 = \varepsilon^{\frac{1}{5}} \tau, \quad (0 < \varepsilon \ll 1), \tag{95}$$

into (86) we obtain

$$\begin{aligned} \frac{\mathrm{d}w_1}{\mathrm{d}\tau_1} &= w_2, \\ \frac{\mathrm{d}w_2}{\mathrm{d}\tau_1} &= \nu_1 + w_1^2 + \varepsilon\,(\nu_2 w_2 + \nu_3 w_1 w_2 - b_1 w_1^3 w_2) \equiv \nu_1 + w_1^2 + \varepsilon\, q(w_1, w_2, \nu), \end{aligned} \tag{96}$$

where $\nu = (\nu_1, \nu_2, \nu_3)$. The system $(96)\big|_{\varepsilon=0}$ is a Hamiltonian system. In order to have the Hamiltonian function given in the form of that in (Han and Yu 2012), we introduce a further transformation,

$$w_1 = \bar{\nu}_1 + z_1, \quad w_2 = \sqrt{2\bar{\nu}_1}\, z_2, \quad \tau_2 = \sqrt{2\bar{\nu}_1}\, \tau_1, \quad \nu_1 = -\bar{\nu}_1^2 \ \ (\bar{\nu}_1 > 0), \tag{97}$$

into (96) to yield

$$\begin{aligned} \frac{\mathrm{d}z_1}{\mathrm{d}\tau_2} &= z_2, \\ \frac{\mathrm{d}z_2}{\mathrm{d}\tau_2} &= z_1 + \tfrac{1}{2\bar{\nu}_1} z_1^2 + \tfrac{\varepsilon}{\sqrt{2\bar{\nu}_1}} \big[(\nu_2 + \bar{\nu}_1\, \nu_3 - b_1 \bar{\nu}_1^3) z_2 + (\nu_3 - 3b_1 \bar{\nu}_1^2) z_1 z_2 \\ &\quad - 3b_1 \bar{\nu}_1 z_1^2 z_2 - b_1 z_1^3 z_2)\big] \\ &\equiv z_1 + \tfrac{1}{2\bar{\nu}_1} z_1^2 + \varepsilon\, q(z_1, z_2, \bar{\nu}), \end{aligned} \tag{98}$$

where $\bar{\nu} = (\bar{\nu}_1, \nu_2, \nu_3)$.

System $(98)\big|_{\varepsilon=0}$ is a Hamiltonian system with two equilibrium solutions:

$$\widetilde{\mathrm{E}}_- = (-2\bar{\nu}_1, 0) \quad \text{and} \quad \widetilde{\mathrm{E}}_0 = (0, 0), \tag{99}$$

with $\widetilde{\mathrm{E}}_-$ and $\widetilde{\mathrm{E}}_0$ being center and saddle, respectively. These two equilibria correspond to the $\mathrm{E}_\pm$ defined in (87). The Hamiltonian function is

$$H(z_1, z_2) = \tfrac{1}{2}\,(z_2^2 - z_1^2) - \tfrac{1}{6\bar{\nu}_1}\, z_1^3, \tag{100}$$

and the homoclinic orbit connecting E_0 is given by

$$\Gamma_0: \quad H(z_1, z_2) = \tfrac{1}{2}\,(z_2^2 - z_1^2) - \tfrac{1}{6\bar{\nu}_1}\, z_1^3 = H(0, 0) = 0, \tag{101}$$

and $H(-2\bar{\nu}_1, 0) = -\frac{2}{3}\,\bar{\nu}_1^2$. Thus, any closed orbit of the Hamiltonian system (98)$\big|_{\varepsilon=0}$ inside Γ_0 can be described by

$$\Gamma_{\rm h}: \quad H(z_1, z_2, h) = \tfrac{1}{2}\,(z_2^2 - z_1^2) - \tfrac{1}{6\bar{\nu}_1}\,z_1^3 - h = 0, \quad h \in \big(-\tfrac{2}{3}\,\bar{\nu}_1^2, 0\big). \tag{102}$$

Now the Abelian integral or the (first-order) Melnikov function for the perturbed system (98) can be written as

$$\begin{aligned} M(h,\nu) &= \oint_{\Gamma_{\rm h}} q(z_1, z_2, \nu)\, dz_1 - p(z_1, z_2, \nu)\, dz_2 \;|_{\varepsilon=0} \qquad (p=0) \\ &= \oint_{\Gamma_{\rm h}} q(z_1, z_2, \nu) \;|_{\varepsilon=0}\; dz_1 = \oint_{\Gamma_{\rm h}} H_{z_2} q(z_1, z_2, \nu) \;|_{\varepsilon=0}\; dt \\ &= \frac{1}{\sqrt{2\bar{\nu}_1}} \oint_{\Gamma_{\rm h}} z_2^2 \big[\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3 + (\nu_3 - 3b_1\bar{\nu}_1^2)z_1 - 3b_1\bar{\nu}_1 z_1^2 - b_1 z_1^3\big]\, dt. \end{aligned} \tag{103}$$

If we consider bifurcation of limit cycles around the center $\rm E_-$, we may expand the $M(h, \nu)$ in h for $-\frac{2}{3}\,\bar{\nu}_1^2 < h \ll 1$, which is equivalent to using the method of focus values or normal forms (Tian and Yu 2018), and the same result as that given in Theorem 4.3 will be obtained.

In the following, we consider the limit cycles bifurcating from the homoclinic orbit Γ_0. We use the method and formulas given in Han and Yu (2012) to expand the Melnikov function $M(h, \nu)$ for $0 < -h \ll 1$ to obtain

$$\begin{aligned} M(h,\nu) = {} & C_0(\nu) + C_1(\nu)\, h \ln|h| + C_2(\nu)\, h \\ & + C_3(h)\, h^2 \ln|h| + \cdots, \quad (0 < -h \ll 1), \end{aligned} \tag{104}$$

where

$$\begin{aligned} C_0(\nu) &= \frac{1}{\sqrt{2\bar{\nu}_1}} \oint_{\Gamma_0} z_2^2\big[\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3 + (\nu_3 - 3b_1\bar{\nu}_1^2)z_1 - 3b_1\bar{\nu}_1 z_1^2 - b_1 z_1^3\big]\, dt, \\ C_1(\nu) &= a_{10} + b_{01}, \end{aligned} \tag{105}$$

in which a_{10} and b_{01} are the coefficients in the functions $p(z_1, z_2, \nu)$ and $q(z_1, z_2, \nu)$, given by

$$a_{10} = 0, \qquad b_{01} = \frac{1}{\sqrt{2\bar{\nu}_1}}(\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3). \tag{106}$$

To compute $C_0(\nu)$, introducing the parametric transformation:

$$z_1(t) = -3\,\mathrm{sech}^2(t), \quad z_2(t) = 3\,\bar{\nu}_1^2 \mathrm{sech}^2(t)\tanh(t), \tag{107}$$

into $C_0(\nu)$ we obtain

$$\begin{aligned} C_0(\nu) &= \frac{1}{\sqrt{2\bar{\nu}_1}} \int_{-\infty}^{\infty} \big[\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3 + (\nu_3 - 3b_1\bar{\nu}_1^2)z_1 - 3b_1\bar{\nu}_1 z_1^2 - b_1 z_1^3\big] z_2^2\, dt, \\ &= \frac{1}{\sqrt{2\bar{\nu}_1}} \int_{-\infty}^{\infty} \big[\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3 + (\nu_3 - 3b_1\bar{\nu}_1^2)z_1 - b_1 z_1^3\big] z_2^2\, dt, \end{aligned}$$

$$= \frac{1}{\sqrt{2\bar{\nu}_1}} \int_{-\infty}^{\infty} \left[\nu_2 + \bar{\nu}_1\,\nu_3 - b_1\bar{\nu}_1^3 - 3\,(\nu_3 - 3b_1\bar{\nu}_1^2)\,\mathrm{sech}^2(t) + 27\,b_1\,\mathrm{sech}^6(t)\right] \times 9\,\bar{\nu}_1^4 \mathrm{sech}^4(t)\tanh^2(t)\,dt$$
$$= \frac{1}{\sqrt{2\bar{\nu}_1}} \times \frac{12}{5}\bar{\nu}_1^2\left[\nu_2 - \frac{5}{7}\bar{\nu}_1\,\nu_3 - \frac{103}{77}\,b_1\bar{\nu}_1^3\right] = \frac{6\bar{\nu}_1\sqrt{2\bar{\nu}_1}}{5}\left[\nu_2 - \frac{5}{7}\bar{\nu}_1\,\nu_3 - \frac{103}{77}\,b_1\bar{\nu}_1^3\right]. \tag{108}$$

Finally, we express $C_0(\nu)$ and $C_1(\nu)$ in terms of the original perturbation parameters ξ_j by using

$$\bar{\nu}_1 = \sqrt{-\nu_1} = \sqrt{-\varepsilon^{-4/5}\,\xi_1} = \varepsilon^{-2/5}\sqrt{-\xi_1}, \quad \nu_2 = \varepsilon^{-6/5}\xi_2, \quad \nu_3 = \varepsilon^{-4/5}\xi_3,$$

as

$$\begin{aligned} C_0(\xi) &= \frac{6\bar{\nu}_1\sqrt{2\bar{\nu}_1}}{5}\varepsilon^{-6/5}\left[\xi_2 - \frac{5}{7}(\xi_3 + \frac{103}{55}\,b_1\xi_1)\sqrt{-\xi_1}\right], \\ C_1(\xi) &= \frac{1}{\sqrt{2\bar{\nu}_1}}\left[\xi_2 + (\xi_3 + b_1\xi_1)\sqrt{-\xi_1}\right]. \end{aligned} \tag{109}$$

Hence, the homoclinic bifurcation surface is defined by $C_0(\xi) = 0$, leading to the equation HL given in (93). Again note that ignoring the term $b_1\xi_1$ and letting $\xi_3 = -\frac{a_2}{a_1}$ in $C_0(\xi) = 0$ yields the homoclinic curve, $\xi_1 = -\frac{49}{25}\left(\frac{a_1}{a_2}\right)^2\xi_2^2$, which is exactly the same as that given in Theorem 4.2 for the codimension-2 BT bifurcation, as expected. Further, degenerate homoclinic bifurcation happens at $C_1(\xi) = 0$, which defines another critical surface DHL given in (94). □

It is obvious that system (86) has no equilibria for $\xi_1 > 0$. So the bifurcation surfaces are in the half space $\xi_1 \le 0$, implying that the bifurcation diagram of system (86) is a cone, which can be represented by its intersection with the 2-sphere,

$$S_\sigma = \left\{(\xi_1, \xi_2, \xi_3) \mid \xi_1^2 + \xi_2^2 + \xi_3^2 = \sigma^2,\ 0 < \sigma \ll 1\right\}. \tag{110}$$

The intersection curve of the Hopf bifurcation surface (90) and the 2-sphere (110) can be described as

$$\mathrm{H}_{\mathrm{curve}} = \left\{(\xi_2, \xi_3)\,\middle|\, \xi_2 = \left[\xi_3 - b_1\left(\sigma^2 - \xi_2^2 - \xi_3^2\right)^{\frac{1}{2}}\right]\left(\sigma^2 - \xi_2^2 - \xi_3^2\right)^{\frac{1}{4}},\ \ \xi_2^2 + \xi_3^2 \le \sigma^2\right\}, \tag{111}$$

which is shown in Fig. 19 as the blue curve. Similarly, we can find the intersection curve of the homoclinic bifurcation surface (93) and the 2-sphere (110) as

$$\mathrm{HL}_{\mathrm{curve}} = \left\{(\xi_2, \xi_3)\,\middle|\, \xi_2 = \frac{5}{7}\left[\xi_3 - \frac{103}{55}b_1\left(\sigma^2 - \xi_2^2 - \xi_3^2\right)^{\frac{1}{2}}\right]\left(\sigma^2 - \xi_2^2 - \xi_3^2\right)^{\frac{1}{4}},\ \ \xi_2^2 + \xi_3^2 \le \sigma^2\right\}, \tag{112}$$

which is also shown in Fig. 19 as the green curve.

The intersection point of the Hopf and homoclinic bifurcation curves, denoted by C (see Fig. 19), can be obtained from the following equations:

$$\mathrm{C}:\ \xi_2 = \tfrac{24}{11}\, b_1\xi_1\sqrt{-\xi_1},\ \ \xi_3 = \tfrac{13}{11}\, b_1\xi_1,\ \ f(\xi_1) = \xi_1 + \sqrt{\sigma^2 - \xi_2^2 - \xi_3^2} = 0 \Longrightarrow \xi_1. \tag{113}$$

Similarly, the GH denotes the intersection point of the Hopf bifurcation (in blue) and the generalized Hopf bifurcation (in pink) curves, and DHL denotes the intersection point of the homoclinic bifurcation (in green) and the degenerate homoclinic bifurcation (in brown) curves. These two points are shown in Fig. 19, determined from the following equations:

$$\mathrm{GH}:\ \begin{cases} \xi_2 = -\,2b_1\,\xi_1\sqrt{-\xi_1}, \quad \xi_3 = -\,3\,b_1\xi_1, \\ f(\xi_1) = \xi_1 + \sqrt{\sigma^2 - \xi_2^2 - \xi_3^2} = 0 \implies \xi_1, \end{cases} \tag{114}$$

and

$$\begin{aligned} \mathrm{DHL}:\quad \xi_2 &= \tfrac{4}{11}\, b_1\xi_1\sqrt{-\xi_1}, \quad \xi_3 = -\,\tfrac{15}{11}\, b_1\xi_1, \\ & f(\xi_1) = \xi_1 + \sqrt{\sigma^2 - \xi_2^2 - \xi_3^2} = 0 \implies \xi_1. \end{aligned} \tag{115}$$

As an example, taking $D = 0.2499$, which yields $b_1 = 0.230923$, and $\sigma = 0.1$, we obtain

$$\begin{aligned} \mathrm{C} = (\xi_2, \xi_3)_{\mathrm{C}} &= (-\,0.014847, -\,0.026036), \quad \text{for } \xi_1 = -\,0.095403, \\ \mathrm{GH} = (\xi_2, \xi_3)_{\mathrm{GH}} &= (0.010790, 0.056614), \quad \text{for } \xi_1 = -\,0.081721, \\ \mathrm{DHL} = (\xi_2, \xi_3)_{\mathrm{DHL}} &= (-\,0.002473, -\,0.030026), \quad \text{for } \xi_1 = -\,0.095354, \end{aligned}$$

as shown in Fig. 19a.

The two points GH and DHL correspond to the Melnikov function $M(h, \nu)$ given in (103) satisfying $M(h, \nu) = M_h(h, \nu) = 0$ at $h = -\frac{2}{3}\,\bar{\nu}_1^2$ and $h = 0$, respectively. In fact, for each value of $h \in (-\frac{2}{3}\,\bar{\nu}_1^2, 0)$, one can find a point corresponding to $M(h, \nu) = M_h(h, \nu) = 0$. These points form a curve connecting the two points GH and DHL and tangent to the Hopf bifurcation curve at the point GH and to the homoclinic bifurcation curve at the point DHL. This curve is usually called double limit cycle bifurcation curve, denoted as DLC. The existence of this curve is proved by Dumortier et al. (1987), but no explicit formula or computational method given to determine the curve. Figure 19a is an exact bifurcation diagram for $D = 0.2499$ and $\sigma = 0.1$, where the double limit cycle bifurcation curve DLC is not shown since it is too close to the Hopf and homoclinic bifurcation curves. In order to clearly see the bifurcations we show a schematic general bifurcation diagram in Fig. 19b with associated phase portraits for $b_1 > 0$ and $0 < \sigma \ll 1$. The case $b_1 < 0$ can be easily obtained from Fig. 19b as a reflection of this figure with stability changed.

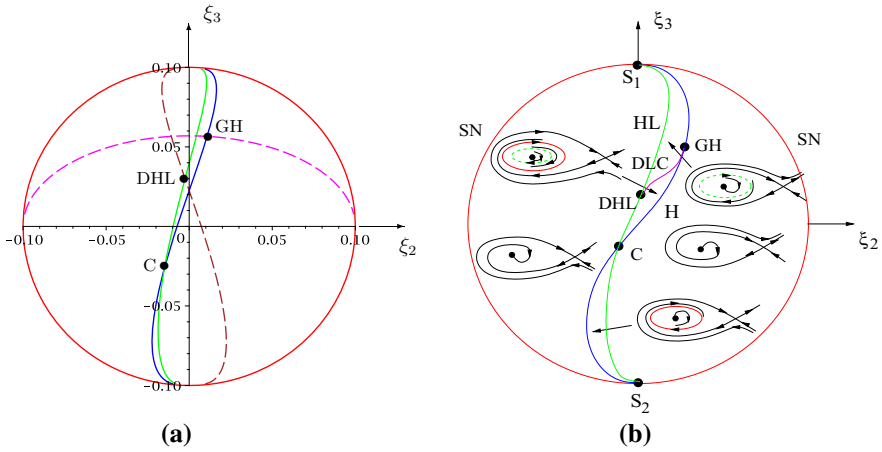

Fig. 19 Bifurcation curves for the codimension-3 BT bifurcation based on the normal form (86), displayed in the intersection of the cone and the 2-sphere $\xi_1^2+\xi_2^2+\xi_3^2=\sigma^2$, with red color for saddle-node, blue for Hopf and green for homoclinic, respectively: **a** with $D=0.2499$ and $\sigma=0.1$, where the intersection point of the pink and blue curves is the degenerate Hopf bifurcation, and the intersection point of the brown and green curves denotes the degenerate homoclinic loop bifurcation; and **b** a schematic bifurcation diagram for $b_1>0$ (i.e., $a_3>0$) (Color figure online)

We summarize the results for the codimension-3 BT bifurcation as follows (see Fig. 19b).

(a) Saddle-node bifurcation occurs along the circle, $\xi_2^2+\xi_3^2=\sigma^2$, excluding the two points S_1 and S_2, while the two points S_1 and S_2 correspond to the BT bifurcation of condimension two.
(b) Hopf bifurcation appears along the H curve excepting the point GH, while the point GH represent the generalized Hopf bifurcation.
(c) Homoclinic bifurcation happens along the HL curve excluding the point DHL, while the DHL point denotes the degenerate homoclinic bifurcation.
(d) Double limit cycle bifurcation occurs along the curve DLC, which connects the two points GH and DHL and is tangent to the curves H and HL at these two points, respectively.

5 Conclusions

In this contribution, we have provided a fairly complete dynamical analysis on a two-dimensional disease model, which can be used for either in-host disease modeling or epidemiologic modeling. In particular, we explored the dynamical behaviors of the system in a full four-dimensional parameter space. In particular, we have shown that when the reproduction number R_0 is varied near $\mathrm{R}_0=1$, the system exhibits rich dynamical behaviors, including equilibrium solutions which exchange their stability at the transcritical point $R_0=1$. Both Hopf and generalized Hopf bifurcations can occur regardless of whether $R_0<1$ or $R_0\geq 1$, and yield bistability or even tristability,

showing that a simple model can still catch certain types of complexity in a disease model. This new type of bistability or tristability reveals a more complex but more realistic situation: The predicted state may not necessarily be an equilibrium (either the disease-free equilibrium or the endemic equilibrium), but may also involve disease periodic oscillation. This implies that the infective individuals and removed individuals are not necessarily fixed, but in a more realistically, mutually stable periodic motion.

Our study in this paper has also indicated that when $R_0 < 1$, the system can have BT bifurcation leading to more complex dynamical behavior such as homoclinic orbit bifurcation. The analysis is based on the parametric normal form for BT bifurcation. We developed a novel computational method to derive the normal form for BT bifurcation of codimension two and three via one-step general nonlinear transformation. The homoclinic loop bifurcation, arising from BT bifurcation provides a new scenario/mechanism for generating recurrence, which is characterized by a special type of oscillation with slow and fast motions on the trajectories. This new mechanism is different with the one proposed in Zhang et al. (2013, 2014a, b).

Acknowledgements This work was supported by the National Sciences and Engineering Research Council of Canada (No. R2686A02). The comments and suggestions, received from two anonymous reviewers, for improving this manuscript are greatly appreciated.

References

Boie, S., Kirk, V., Sneyd, J., Wechselberger, M.: Effects of quasi-steady-state reduction on biophysical models with oscillations. J. Theor. Biol. **393**, 16–31 (2016)

Dumortier, F., Roussarie, R., Sotomayor, J.: Generic 3-parameter family of vector feilds on the plane, unfolding a singularity with nilpotent linear part. The cusp case of codimension 3. Ergod. Theory Dyn. Syst. **7**(3), 375–413 (1987)

Earn, D.J.D., Rohani, P., Bolker, B.M., Grenfell, B.T.: A simple model for complex dynamical transitions in epidemics. Science **287**(5453), 667–670 (2000)

Flach, E.H., Schnell, S.: Use and abuse of the quasi-steady-state approximation. IEEE Proc. Syst. Biol. **153**(4), 187–191 (2006)

Gazor, M., Yu, P.: Formal decomposition method and parametric normal form. Int. J. Bifurc. Chaos **20**, 3415–3487 (2010)

Gazor, M., Yu, P.: Spectral sequences and parametric normal forms. J. Differ. Equ. **252**, 1003–1031 (2012)

Gazor, M., Moazeni, M.: Parametric normal forms for Bogdanov–Takens singularity; the generalized saddle-node case. Discret. Contin. Dyn. Syst. **35**, 205–224 (2015)

Griffin D. E.: Immune Responses during measles virus infection. In: Measles Virus (Eds ter Meulen, Volker and Billeter, Martin A.) pp. 117–134, Springer, Berlin (1995)

Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 4th edn. Springer, New York (1993)

Han, M., Yu, P.: Normal Forms, Melnikov Functions and Bifurcations of Limit Cycles. Sringer, New York (2012)

Hethcote, H.W., van den Driessche, P.: Some epidemiological models with nonlinear incidence. J. Math. Biol. **29**(3), 271–287 (1991)

Korobeinikov, A., Maini, P.K.: Non-linear incidence and stability of infectious disease models. Math. Med. Biol. **22**, 113–128 (2005)

Korobeinikov, A., Shchepakina, E., Sobolev, V.: Paradox of enrichment and system order reduction: bacteriophages dynamics as case study. Math. Med. Biol. **33**(3), 359–369 (2005)

Li, C., Rousseau, C.: A system with three limit cycles appearing in a Hopf bifurcation and dying in a homoclinic bifurcation: the cusp of order 4. J. Differ. Equ. **79**, 132–167 (1989)

Liu, W., Levin, S., Iwasa, Y.: Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. J. Math. Biol. **23**, 187–204 (1986)
Tian, Y., Yu, P.: Bifurcation of small limit cycles in cubic integrable systems using higher-order analysis. J. Differ. Equ. **264**(9), 5950–5976 (2018)
van Gaalen, R.D., Wahl, L.M.: Reconciling conflicting clinical studies of antioxidant supplementation as HIV therapy: a mathematical approach. BMC Public Health **9**, 1–18 (2009)
Wang, J.J., Zhang, J.Z., Jin, Z.: Analysis of an SIR model with bilinear incidence rate. Nonlinear Anal. Real World Appl. **11**(4), 2390–2402 (2010)
Yu, P.: Computation of normal forms via a perturbation technique. J. Sound Vib. **211**, 19–38 (1998)
Yu, P.: Simplest normal forms of Hopf and generalized Hopf bifurcations. Int. J. Bifurc. Chaos **9**, 1917–1939 (1999)
Yu, P., Leung, A.Y.L.: The simplest normal form of Hopf bifurcation. Nonlinearity **16**, 277–300 (2003)
Yu, P., Zhang, W., Wahl, L.M.: Dynamical analysis and simulation of a 2-dimensional disease model with convex incidence. Commun. Nonlinear Sci. Numer. Simul. **37**, 163–192 (2016)
Zhang, W., Wahl, L.M., Yu, P.: Conditions for transient viremia in deterministic in-host models: viral blips need no exogenous trigger. SIAM J. Appl. Math. **73**, 853–881 (2013)
Zhang, W., Wahl, L.M., Yu, P.: Viral blips may not need a trigger: how transient viremia can arise in deterministic in-host models. SIAM Rev. **56**, 127–155 (2014)
Zhang, W., Wahl, L.M., Yu, P.: Modelling and analysis of recurrent autoimmune disease. SIAM J. Appl. Math. **74**, 1998–2025 (2014)
Zhang, W., Wahl, L.M., Yu, P.: Backward bifurcation underlies rich dynamics in simple disease models. J. Math. Biol. **73**, 947–976 (2016)