their food resources. Moreover, the food resources themselves have powers of regeneration, and these powers are limited by density-dependent factors. These facts create the conditions necessary for the expression of limit cycles.

The saturating predator functional response used in the present analysis is that proposed by Holling (1965) for "nonlearning" predators. The label nonlearning is a bit misleading because even predators capable of learning should exhibit this type of response when given only 1 type of prey for which to search. The derivation of this response is as follows: Holling (1965) observed that there were 2 basic time-consuming elements of the "attack cycle" for a predator: (1) search time and (2) handling time (including digestion). If we let T equal the total attack cycle time, T_s equal the search time, h equal the handling time per prey item and N_a equal the number of prey items caught during the attack cycle, then

$$T = T_s + h \cdot N_a. \tag{1.1}$$

Holling (1965) then showed that the number of prey items caught is proportional to the density of prey (S) and to the search time, T_s:

$$N_a = c \cdot S \cdot T_s \tag{1.2}$$

where c is the encounter rate per unit prey density. In the Nicholson-Bailey model, the slope of the feeding rate curve with prey density is constant with slope c. However, this is not the case here: as prey density increases, the slope of the functional response (feeding rate) decreases. Eliminating T_s between Eqs. 1.1 and 1.2 and rearranging gives an expression for the number attacked.

$$N_a = \frac{c \cdot S_2 \cdot T}{1 + c \cdot h \cdot S}.$$

Then the number caught per unit time, F, is the feeding rate, and is given by:

$$F = \frac{N_a}{T} = \frac{c \cdot S}{1 + c \cdot h \cdot S}. \tag{1.3}$$

This is the feeding rate per individual predator. The total consumption by the predator population as a whole is given by $F \cdot X$, where X is the number of predators in the population. The maximum feeding rate is given by the inverse of handling time, $1/h$; this rate is achieved as S, the density of prey, becomes large.

Saturating functional responses of this kind are nearly universal in biological rate processes. At the level of enzyme-mediated reactions, Eq. 1.3 is the well-known Michaelis-Menten equation, which is usually written:

$$v = \frac{V_m \cdot S}{K_m + S}, \tag{1.4}$$

where v is the specific rate of product formation or nutrient uptake, V_m is the maximum specific rate of product formation or nutrient uptake, S is resource density or substrate concentration, and K_m (the half-saturation constant) is the substrate concentration at which the rate of product formation or nutrient uptake is half maximal. Equation 1.4 has 2 constants which are easily converted from the constants in Holling's (1965) functional response, Eq. 1.3: $K_m = 1/ch$ and $V_m = 1/h$ The relationship between the Holling (1965) functional response and Michaelis-Menten kinetics has been explored in more detail by Real (1977).

In summary, the most important biological features of Eqs. 1.3 and 1.4 are: (a) at low resource density, the rate of uptake is limited by, and proportional to, resource concentration, whereas (b) at high resource density, the rate is limited by processing time (enzyme-substrate complex turnover time in enzyme-mediated reactions; handling time in predator–prey interactions), and is independent of resource concentration. In this paper, we have chosen to use the Michaelis-Menten formulation because the representation of the function is a little simpler in Eq. 1.4 than in 1.3.

Statement of the Model

This paper concerns the behavior of a predator–prey system consisting of 2 predator species, x_1 and x_2, and a single prey species, S. We specifically assume that the predator species compete purely exploitatively, with no interference between rivals. Both species have access to prey and compete only by lowering the population of shared prey. Death rates are assumed to be such that the number dying is proportional to the number currently alive. We also assume that there are no significant time lags in the system, that growth rates are logistic in the prey species in the absence of predation, and the functional responses of the predators obey the Holling (1965) "disc" (nonlearning) curve. With these assumptions, the model is given by:

$$\begin{aligned}
\frac{dS(t)}{dt} &= \gamma S(t)[1 - S(t)/K] - \left(\frac{m_1}{y_1}\right)\left(\frac{x_1(t)S(t)}{a_1 + S(t)}\right) \\
&\quad - \left(\frac{m_2}{y_2}\right)\left(\frac{x_2(t)S(t)}{a_2 + S(t)}\right), \\
\frac{dx_1(t)}{dt} &= \frac{m_1 x_1(t)S(t)}{a_1 + S(t)} - d_1 x_1(t), \\
\frac{dx_2(t)}{dt} &= \frac{m_2 x_2(t)S(t)}{a_2 + S(t)} - d_2 x_2(t),
\end{aligned} \tag{1}$$

where S, x_1, and x_2 are all positive at $t = 0$. The symbols are as follows: $x_i(t)$ is the number of the ith predator at time t, $S(t)$ is the number of the prey at time t, m_i is the maximum growth (birth) rate of the ith predator, d_i is the death rate for the ith predator, y_i is the yield conversion factor for the ith predator feeding on the prey, a_i is the half-saturation constant for the ith predator, which is the prey density at which

At $E_1 = (K, 0)$,

$$A(K,0) = \begin{bmatrix} -\gamma & -\frac{mK}{a+K} \\ 0 & \frac{mK}{a+K} - d \end{bmatrix}.$$

E_1 is a saddle point since $x^* < K$.

At $E^* = (x^*, y^*)$,

$$A(x^*,y^*) = \begin{bmatrix} \gamma\left(1 - \frac{2x^*}{K}\right) - \frac{ma}{(a+x^*)^2}y^* & -\frac{mx^*}{a+x^*} \\ \frac{ma}{(a+x^*)^2}y^* & 0 \end{bmatrix}.$$

The characteristic equation of $A(x^*, y^*)$ is

$$\lambda^2 - \lambda\left[\gamma\left(1 - \frac{2x^*}{K}\right) - \frac{ma}{(a+x^*)^2}y^*\right] + \frac{mx^*}{a+x^*} \cdot \frac{ma}{(a+x^*)^2}y^* = 0.$$

E^* is asymptotically stable iff

$$\gamma\left(1 - \frac{2x^*}{K}\right) - \frac{ma}{(a+x^*)^2}y^* < 0,$$

or

$$\gamma\left(1 - \frac{2x^*}{K}\right) - \frac{a}{a+x^*}\gamma\left(1 - \frac{x^*}{K}\right) < 0,$$

or

$$\frac{K-a}{2} < x^*.$$

If $\frac{K-a}{2} > x^*$, then E^* is an unstable spiral.

If $\frac{K-a}{2} = x^*$, then the eigenvalues of $A(x^*, y^*)$ are $\pm\omega i$ for some $\omega \neq 0$. This is "Hopf Bifurcation" which will be discussed in section 6.3.

For $\frac{K-a}{2} < x^* < K$, the interior equilibrium E^* is a stable spiral, E_0, E_1 are saddles. From isocline analysis, we predict $(x(t), y(t)) \to E^*$ as $t \to \infty$ (see Fig. 4.1(a)). For $x^* < \frac{K-a}{2}$, E^* is an unstable spiral, E_0 and E_1 are saddles. It can be shown that the trajectory $(x(t), y(t))$ approaches a unique limit cycle [C] (see Fig. 4.2(b)).

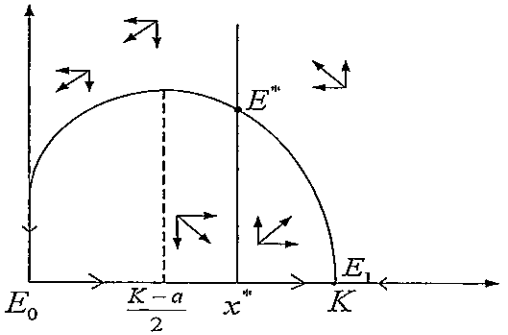

Fig. 4.2(a) $\frac{K-a}{2} < x^* < K$, E^* is a stable spiral.

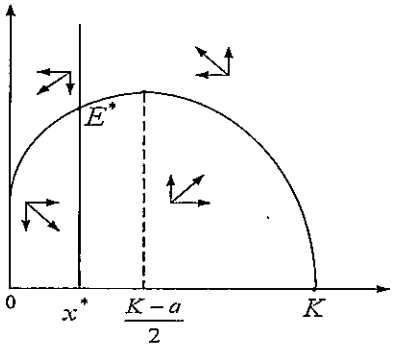

Fig. 4.2(b) $0 < x^* < \frac{K-a}{2}$, E^* is an unstable spiral.

Example 4.2.4 Lotka-Volterra two species competition model [W]:

$$\begin{cases} \frac{dx_1}{dt} = \gamma_1 x_1\left(1 - \frac{x_1}{K_1}\right) - \alpha_1 x_1 x_2, \\ \frac{dx_2}{dt} = \gamma_2 x_2\left(1 - \frac{x_2}{K_2}\right) - \alpha_2 x_1 x_2, \\ x_1(0) > 0,\ x_2(0) > 0, \end{cases}$$

where $\gamma_1, \gamma_2, K_1, K_2, \alpha_1, \alpha_2 > 0$. Three equilibria, $E_0 = (0,0)$, $E_1 = (K_1, 0)$ and $E_2 = (0, K_2)$ always exist. The interior equilibrium $E^* = (x^*, y^*)$ exists in the following cases (iii) and (iv). The varational matrix

From (4.4), given ε, $0 < \varepsilon < \sigma$, there exists $\delta > 0$ such that

$$|g(y)| < \frac{\varepsilon}{M}|y| \text{ for } |y| < \delta. \tag{4.7}$$

Let $0 < \gamma < \min\left\{\frac{\delta}{2M}, \frac{\delta}{2}\right\}$. We claim that if $|y(t_0)| < \gamma$ then $|y(t)| < \delta$ and $|y(t)| \le \frac{\delta}{2}e^{-(\sigma-\epsilon)(t-t_0)}$ for all $t \ge t_0$. If $|y(t)| < \frac{\delta}{2}$ for $t_0 \le t < t^*$ and $|y(t^*)| = \frac{\delta}{2}$, then from (4.7), for $t_0 \le t < t^*$ we have

$$|y(t)| \le M|y(t_0)|e^{-\sigma(t-t_0)} + \epsilon\int_{t_0}^{t} e^{-\sigma(t-s)}|y(s)|ds.$$

Applying Gronwall's inequality to $e^{\sigma t}|y(t)|$ yields

$$|y(t)| \le M|y(t_0)|e^{-(\sigma-\epsilon)(t-t_0)} < \frac{\delta}{2}e^{-(\sigma-\epsilon)(t-t_0)}.$$

we have

This is a desired contradiction to $|y(t^*)| = \frac{\delta}{2}$. Hence for $t \ge t_0$ $|y(t)| < \frac{\delta}{2}$ and $|y(t)| \le \frac{\delta}{2}e^{-(\sigma-\epsilon)(t-t_0)}$. Hence the equilibrium $y = 0$ is asymptotically stable. In fact, it is exponentially stable. □

Definition 4.2.1 *An equilibrium x^* of the autonomous system (4.2) is hyperbolic if the variational matrix $D_x f(x^*)$ has no eigenvalues with zero real parts.*

Definition 4.2.2 *We say two flows φ^t and ψ^t on $U \subseteq R^n$ are topologically conjugate if there is a homeomorphism $h : U \to h(U) \subseteq R^n$ such that $h \circ \varphi^t(x) = \psi^t \circ h(x)$ for all $x \in U$* (see Fig. 4.1).

$\subseteq U$

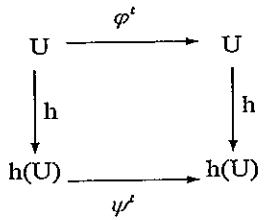

Fig. 4.1

If two flows φ^t and ψ^t are topologically conjugate, then obviously the homeomorphism h takes the trajectories of φ^t into the trajectories of ψ^t.

In the following we state without proof the Hartman-Grobman Theorem which says that if an equilibrium x^* is hyperbolic, then linearized stability of x^* implies nonlinear stability of x^* in the sense of topological conjugacy.

Hartman-Grobman Theorem [R]

Let x^* be a hyperbolic equilibrium of the system $\frac{dx}{dt} = f(x)$. Then the flow φ^t of f is topologically conjugate in a neighborhood U of x^* to the affine flow $x^* + e^{At}(x - x^*)$ where $A = D_x f(x^*)$. More precisely, there is a neighborhood U of x^* and a homeomorphism $h : U \to U$ such that $\varphi^t(h(x)) = h\left(x^* + e^{At}(x - x^*)\right)$ as long as $x^* + e^{At}(x - x^*) \in U$.

Example 4.2.3 Predator-Prey system with Holling's type II functional responses ([HHW]). Let $x(t)$, $y(t)$ be population densities of prey and predator species respectively. Assume the prey grows logistically in the absence of predation with intrinsic growth rate γ and carrying capacity K. Assume the per capital growth rate of predator, $\frac{y'}{y}$ is $\frac{mx}{a+x}$ where a is the half saturation constant or Michaelis-Menten constant and m is the maximum birth rate of predator. Let c be the conversion rate of predator, i.e., c is the ratio of the reproduction of predators to the consumption of prey and d is the death rate of predator. Then the equations take the form

$$\begin{aligned}
\frac{dx}{dt} &= \gamma x\left(1 - \frac{x}{K}\right) - c\frac{mx}{a+x}y,\\
\frac{dy}{dt} &= \left(\frac{mx}{a+x} - d\right)y,\\
x(0) &> 0, \quad y(0) > 0.
\end{aligned}$$

We may assume $c = 1$ if we do the scaling $y \to cy$.
Let $m > d$. Then we have three equilibria: $(0,0), (K,0)$ and (x^*, y^*) where $x^* = \frac{a}{\left(\frac{m}{d}\right)-1} > 0$, $y^* > 0$. The interior equilibrium (x^*, y^*) exists if $x^* < K$. Assume $0 < x^* < K$. Then the variational matrix is

$$A(x,y) = \begin{bmatrix} \gamma\left(1 - \frac{2x}{K}\right) - \frac{ma}{(a+x)^2}y & -\frac{mx}{a+x} \\ \frac{ma}{(a+x)^2}y & \left(\frac{mx}{a+x} - d\right) \end{bmatrix}.$$

At $E_0 = (0,0)$

$$A(0,0) = \begin{bmatrix} \gamma & 0 \\ 0 & -d \end{bmatrix}.$$

E_0 is a saddle point for $\gamma > 0, -d < 0$.

19. Epidemic Models and the Dynamics of Infectious Diseases

The study of epidemics has a long history with a vast variety of models and explanations for the spread and cause of epidemic outbreaks. Even today they are often attributed to evil spirits or displeased gods. AIDS (acquired immunodeficiency syndrome), *the* epidemic of the 1980's and probably of the 20th century, has been ascribed by many as a punishment sent by God. Hippocrates (459–377 BC), in his essay on 'Airs, Waters and Localities' wrote that one's temperament, personal habits and environment were important factors – not unreasonable even today. Somewhat less relevant, but not without its moments of humour, is Alexander Howe's (1865) book in which he sets out his 'Laws of Pestilence' in 31 propositions of which the following, proposition 2, is typical: 'The length of the interval between successive periodic visitations corresponds with the period of a single revolution of the lunar node, and a double revolution of the lunar apse time'.

In this chapter and the following, we shall describe some models for the population dynamics of disease agents and the spatio-temporal spread of infections. We can then try to exploit them in the control, or ideally the eradication, of the disease or infection we are considering. The practical use of such models must rely heavily on the realism put into the models. As usual, this does not mean the inclusion of all possible effects, but rather the incorporation in the model mechanisms, in as simple a way as possible, what appear to be the major components. Like most models they generally go through several versions before qualitative phenomena can be explained or predicted with any degree of confidence. Great care must be exercised before practical use is made of any epidemic models. However, even simple models should, and frequently do, pose important questions with regard to the underlying process and possible means of control of the disease or epidemic. One such case study is the model proposed by Capasso and Paveri-Fontana (1979) for the 1973 cholera epidemic in Bari in southern Italy.

An interesting early mathematical model, involving a nonlinear ordinary differential equation, by Bernoulli (1760), considered the effect of cow-pox innoculation on the spread of smallpox. The article has some interesting data on child mortality at the time. It is probably the first time that a mathematical model was used to assess the practical advantages of a vaccination control programme.

Models can be extremely useful in giving reasoned estimates for the level of vaccination for the control of directly transmitted infectious diseases: see, for example, Anderson and May (1982a, 1985, 1986). The theoretical papers on

epidemic models by Kermack and McKendrick (1927, 1932, 1933) have had a major influence in the development of mathematical models: we describe one of these in Section 19.1. The modelling literature is now extensive. A good introduction and survey of the variety of problems and models for the spread and control of infectious diseases are given, for example, by the books by Bailey (1975) and Hoppensteadt (1975) on mathematical models, the survey by Wickwire (1977) and the collection of articles on the population dynamics of infectious diseases edited by Anderson (1982).

In this chapter we discuss several models which incorporate some general aspects of epidemiological modelling of disease transmission and the time development of epidemics. In the following chapter we consider the geographic spread of infectious diseases and describe in detail a practical model for the spatial spread of rabies and a possible means of its control.

There are basically two broad types of model. In one the total population is taken to be approximately constant with, for example, the population divided into susceptible, infected and immune groups: other groupings are also possible, depending on the disease. We discuss models in this category in Section 19.1. In the other, the population size is affected by the disease via the birth rate, mortality and so on. Host-parasite interacting populations often come into this category: we discuss one such model later in the chapter.

19.1 Simple Epidemic Models and Practical Applications

In the models we consider here the total population is taken to be constant. If a small group of infected individuals are introduced into a large population, a basic problem is to describe the spread of the infection within the population as a function of time. Of course this depends on a variety of circumstances, including the actual disease involved, but as a first attempt at modelling directly transmitted diseases we make some not unreasonable general assumptions.

Consider a disease which, after recovery, confers immunity (which includes deaths: dead individuals are still counted). The population can then be divided into three distinct classes; the susceptibles, S, who can catch the disease; the infectives, I, who have the disease and can transmit it; and the removed class, R, namely those who have either had the disease, or are recovered, immune or isolated until recovered. The progress of individuals is schematically described by

$$S \longrightarrow I \longrightarrow R\,.$$

Such models are often called *SIR* models.

The assumptions made about the transmission of the infection and incubation period are crucial in any model. With $S(t)$, $I(t)$ and $R(t)$ as the number of individuals in each class we assume here that: (i) The gain in the infective class is at a rate proportional to the number of infectives and susceptibles, that is rSI, where $r > 0$ is a constant. The susceptibles are lost at the same rate. (ii) The

rate of removal of infectives to the removed class is proportional to the number of infectives, that is aI where $a > 0$ is a constant. (iii) The incubation period is short enough to be negligible; that is a susceptible who contracts the disease is infective right away.

We now consider the various classes as uniformly mixed: that is every pair of individuals has equal probability of coming into contact with one another. The model mechanism is then

$$\frac{dS}{dt} = -rSI\,, \tag{19.1}$$

$$\frac{dI}{dt} = rSI - aI\,, \tag{19.2}$$

$$\frac{dR}{dt} = aI\,. \tag{19.3}$$

where $r > 0$ is the infection rate and $a > 0$ the removal rate of infectives. This is the classic Kermack-McKendrick (1927) model. We are, of course, only interested in non-negative solutions for S, I and R. This is a primitive model but, even so, we can make some highly relevant general comments about epidemics and, in fact, adequately describe some specific epidemics with such a model.

The constant population size is built into the system (19.1)–(19.3) since, on adding the equations,

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad \Rightarrow \quad S(t) + I(t) + R(t) = N\,, \tag{19.4}$$

where N is the total size of the population. Thus, S, I and R are all bounded above by N. The mathematical formulation of the epidemic problem is completed given initial conditions such as

$$S(0) = S_0 > 0, \quad I(0) = I_0 > 0, \quad R(0) = 0\,. \tag{19.5}$$

A key question in any epidemic situation is, given r, a, S_0 and the initial number of infectives I_0, whether the infection will spread or not, and if it does how it develops with time, and of course when it will start to decline. From (19.2)

$$\left[\frac{dI}{dt}\right]_{t=0} = I_0(rS_0 - a) \begin{matrix} > \\ < \end{matrix} 0 \quad \text{if} \quad S_0 \begin{matrix} > \\ < \end{matrix} \frac{a}{r} = \rho\,. \tag{19.6}$$

Since, from (19.1), $dS/dt \leq 0$, $S \leq S_0$ we have, if $S_0 < a/r$,

$$\frac{dI}{dt} = I(rS - a) \leq 0 \quad \text{for all} \quad t \geq 0\,, \tag{19.7}$$

in which case $I_0 > I(t) \to 0$ as $t \to \infty$ and so the infection dies out: that is, no epidemic can occur. On the other hand if $S_0 > a/r$ then $I(t)$ initially increases and we have an epidemic. The term 'epidemic' means that $I(t) > I_0$

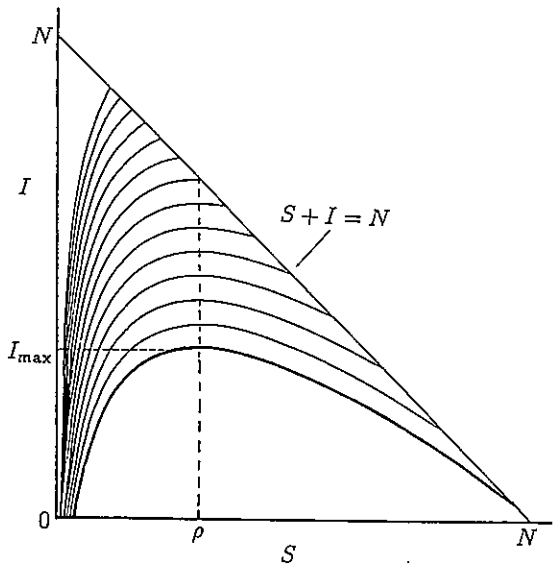

Fig. 19.1. Phase trajectories in the susceptibles (S)-infectives (I) phase plane for the SIR model epidemic system (19.1)–(19.3). The curves are determined by the initial conditions $I(0) = I_0$ and $S(0) = S_0$. With $R(0) = 0$, all trajectories start on the line $S + I = N$ and remain within the triangle since $0 < S + I \leq N$ for all time. An epidemic situation formally exists if $I(t) > I_0$ for any time $t > 0$: this always occurs if $S_0 > \rho(= a/r)$ and $I_0 > 0$.

for some $t > 0$: see Fig. 19.1 above. We thus have a *threshold phenomenon.* If $S_0 > S_c = a/r$ there is an epidemic while if $S_0 < S_c$ there is not. The critical parameter $\rho = a/r$ is sometimes called the *relative removal rate* and its reciprocal $\sigma(= r/a)$ the infection's *contact rate.*

We write

$$R_0 = \frac{rS_0}{a},$$

where R_0 is the basic *reproduction rate* of the infection, that is, the number of secondary infections produced by one primary infection in a wholly susceptible population. Here $1/a$ is the average infectious period. If more than one secondary infection is produced from one primary infection, that is $R_0 > 1$, clearly an epidemic ensues. The whole question of thresholds in epidemics is obviously important. A mathematical introduction to the subject is given by Waltman (1974).

We can derive some other useful analytical results from this simple model. From (19.1) and (19.2)

$$\frac{dI}{dS} = -\frac{(rS - a)I}{rSI} = -1 + \frac{\rho}{S}, \quad \rho = \frac{a}{r}, \quad (I \neq 0).$$

The singularities all lie on the $I = 0$ axis. Integrating the last equation gives the (I, S) phase plane trajectories as

$$I + S - \rho \ln S = \text{constant} = I_0 + S_0 - \rho \ln S_0 \,, \tag{19.8}$$

where we have used the initial conditions (19.5). The phase trajectories are sketched in Fig. 19.1. Note that with (19.5), all initial values S_0 and I_0 satisfy $I_0 + S_0 = N$ since $R(0) = 0$ and so for $t > 0$, $0 \leq S + I < N$.

If an epidemic exists we should like to know how severe it will be. From (19.7) the maximum I, $I_{\max}$, occurs at $S = \rho$ where $dI/dt = 0$. From (19.8), with $S = \rho$,

$$\begin{aligned} I_{\max} &= \rho \ln \rho - \rho + I_0 + S_0 - \rho \ln S_0 \\ &= I_0 + (S_0 - \rho) + \rho \ln \left(\frac{\rho}{S_0} \right) \\ &= N - \rho + \rho \ln \left(\frac{\rho}{S_0} \right) \,. \end{aligned} \tag{19.9}$$

For any initial values I_0 and $S_0 > \rho$, the phase trajectory starts with $S > \rho$ and we see that I increases from I_0 and hence an epidemic ensues. It may not necessarily be a severe epidemic as is the case if I_0 is close to $I_{\max}$. It is also clear from Fig. 19.1 that if $S_0 < \rho$ then I decreases from I_0 and no epidemic occurs.

Since the axis $I = 0$ is a line of singularities, on all trajectories $I \to 0$ as $t \to \infty$. From (19.1), S decreases since $dS/dt < 0$ for $S \neq 0, I \neq 0$. From (19.1) and (19.3)

$$\begin{aligned} &\frac{dS}{dR} = -\frac{S}{\rho} \\ &\Rightarrow \quad S = S_0 \exp\left[-R/\rho\right] \geq S_0 \exp\left[-N/\rho\right] > 0 \\ &\Rightarrow \quad 0 < S(\infty) \leq N \,. \end{aligned} \tag{19.10}$$

In fact from Fig. 19.1, $0 < S(\infty) < \rho$. Since $I(\infty) = 0$, (19.4) implies that $R(\infty) = N - S(\infty)$. Thus, from (19.10)

$$S(\infty) = S_0 \exp\left[-\frac{R(\infty)}{\rho}\right] = S_0 \exp\left[-\frac{N - S(\infty)}{\rho}\right]$$

and so $S(\infty)$ is the positive root $0 < z < \rho$ of the transcendental equation

$$S_0 \exp\left[-\frac{N - z}{\rho}\right] = z \,. \tag{19.11}$$

We then get the total number of susceptibles who catch the disease in the course of the epidemic as

$$I_{\text{total}} = I_0 + S_0 - S(\infty) \,, \tag{19.12}$$

where $S(\infty)$ is the positive solution z of (19.11). An important implication of this analysis, namely that $I(t) \to 0$ and $S(t) \to S(\infty) > 0$, is that the disease dies out from a lack of *infectives* and *not* from a lack of susceptibles.

The threshold result for an epidemic is directly related to the relative removal rate ρ – if $S_0 > \rho$ an epidemic ensues whereas it does not if $S_0 < \rho$. For a given disease, the relative removal rate varies with the community and hence determines whether an epidemic may occur in one community and not in another. The number of susceptibles S_0 also plays a role, of course. For example, if the density of susceptibles is high and the removal rate, a, of infectives is low (through ignorance, lack of medical care, inadequate isolation and so on) then an epidemic is likely to occur. Expression (19.9) gives the maximum number of infectives while (19.12) gives the total number who get the infection in terms of $\rho(= a/r)$, I_0, S_0 and N.

In most epidemics it is diffult to determine how many new infectives there are each day since only those that are removed, for medical aid or whatever, can be counted. Public Health records generally give the number of infectives per day, week or month. So, to apply the model to actual epidemic situations in general, we need to know the number removed per unit time, namely dR/dt as a function of time.

From (19.10), (19.4) and (19.3) we get an equation for R alone, namely

$$\frac{dR}{dt} = aI = a(N - R - S) = a\left(N - R - S_0 \exp\left[-\frac{R}{\rho}\right]\right), \quad R(0) = 0\ , \tag{19.13}$$

which can only be solved analytically in a parametric way: the solution in this form however is not very convenient. Of course, if we know a, r, S_0 and N it is a simple matter to compute the solution numerically. Usually we do not know all the parameters and so we have to carry out a best fit procedure assuming, of course, the epidemic is reasonably described by such a model. In practice, however, it is often the case that if the epidemic is not large, R/ρ is small; certainly $R/\rho < 1$. Following Kermack and McKendrick (1927) we can then approximate (19.13) by

$$\frac{dR}{dt} = a\left[N - S_0 + \left(\frac{S_0}{\rho} - 1\right)R - \frac{S_0 R^2}{2\rho^2}\right]\ .$$

Factoring the right hand side quadratic in R, we can integrate this equation to get, after some elementary but tedious algebra, the solution

$$R(t) = \frac{\rho^2}{S_0}\left[\left(\frac{S_0}{\rho} - 1\right) + \alpha \tanh\left(\frac{\alpha a t}{2} - \phi\right)\right] \tag{19.14}$$

$$\alpha = \left[\left(\frac{S_0}{\rho} - 1\right)^2 + \frac{2S_0(N - S_0)}{\rho^2}\right]^{1/2}, \quad \phi = \frac{\tanh^{-1}\left(\frac{S_0}{\rho} - 1\right)}{\alpha}\ .$$

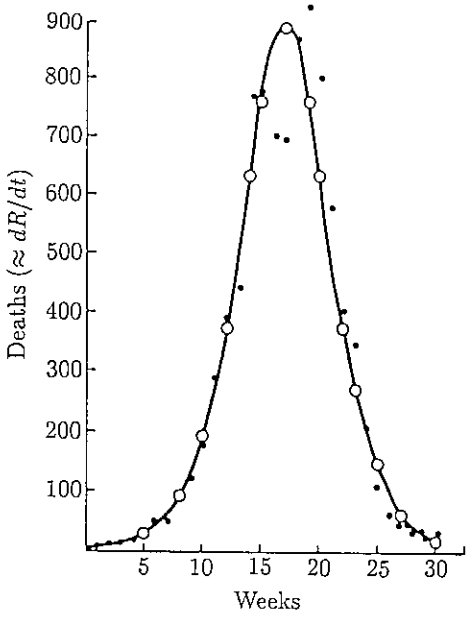

Fig. 19.2. Bombay plague epidemic of 1905–6. Comparison between the data (•) and theory (o) from the (small) epidemic model and where the number of deaths is approximately dR/dt given by (19.16). (After Kermack and McKendrick 1927)

The removal rate is then given by

$$\frac{dR}{dt} = \frac{a\alpha^2\rho^2}{2S_0}\,\mathrm{sech}^2\left(\frac{\alpha a t}{2} - \phi\right) , \tag{19.15}$$

which involves only 3 parameters, namely $a\alpha^2\rho^2/(2S_0)$, αa and ϕ. With epidemics which are not large, it is this function of time which we should fit to the Public Health records. On the other hand, if the disease is such that we know the actual number of the removed class then it is $R(t)$ in (19.14) we should use. If R/ρ is not small, however, we must use the differential equation (19.13) to determine $R(t)$.

We now apply the model to two very different epidemic situations.

Bombay Plague Epidemic 1905–6

This plague epidemic lasted for almost a year. Since most of the victims who got the disease died, the number removed per week, that is dR/dt, is approximately equal to the number of deaths per week. On the basis that the epidemic was not severe (relative to the population size), Kermack and McKendrick (1927) compared the actual data with (19.15), determined the best fit for the three parameters and got

$$\frac{dR}{dt} = 890\,\mathrm{sech}^2(0.2t - 3.4) . \tag{19.16}$$

This is illustrated in Fig. 19.2 together with the actual epidemic data.

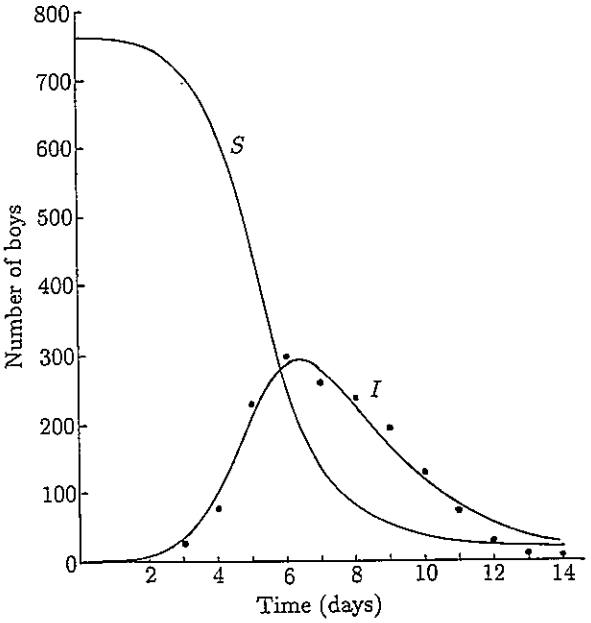

Fig. 19.3. Influenza epidemic data (•) for a boys boarding school as reported in British Medical Journal, 4th March 1978. The continuous curves for the infectives (I) and susceptibles (S) were obtained from a best fit numerical solution of the SIR system (19.1)–(19.3): parameter values $N = 763$, $S_0 = 762$, $I_0 = 1$, $\rho = 202$, $r = 2.18 \times 10^{-3}$/day. The conditions for an epidemic to occur, namely $S_0 > \rho$ is clearly stisfied and the epidemic is severe since R/ρ is not small.

Influenza Epidemic in an English Boarding School 1978

In the 4th March 1978 issue of the British Medical Journal there is a report with detailed statistics of a flu epidemic in a boys boarding school with a total of 763 boys. Of these 512 were confined to bed during the epidemic, which lasted from 22nd January to 4th February 1978. It seems that one infected boy initiated the epidemic. This situation has many of the requirements assumed in the model derivation. Here however, the epidemic is severe and the full system has to be used. Here, when a boy was infected he was put to bed and so we have $I(t)$ directly from the data. Since in this case we have no analytical solution for comparison with the data, a best fit numerical technique was used directly on the equations (19.1)–(19.3) for comparison of the data. Fig. 19.3 illustrates the resulting time evolution for the infectives, $I(t)$, together with the epidemic statistics. The R-equation (19.3) is uncoupled: the solution for $R(t)$ is simply proportional to the area under the $I(t)$ curve.

Raggett (1982) applied the SIR model (19.1)–(19.3) to the outbreak of plague in the village of Eyam in England from 1665–66. In this remarkable altruistic incident, the village sealed itself off when plague was discovered, so as to prevent it spreading to the neighbouring villages, and it was successful. By the end of the epidemic only 83 of the original population of 350 survived. Thus here,

$S(\infty) = 83$ out of an inital $S_0 = 350$. This is another example, like the school flu epidemic, where the epidemic was severe. Raggett (1982) shows how to determine the parameters from the available data and knowledge of the etiology of the disease. He reiterates the view that although the initial form was proabaly bubonic plague, the pneumonic form (see the following chapter, Section 20.2) most likely became prevalent; the latter form can be transmitted from the cough of a victim. The comparison between the solutions from the determinsitc model and the Eyam data is very good. The comparison is much better than that obtained from the corresponding stochastic model, which Raggett (1982) also considered. We discuss a model for the spatial spread of plague in Section 20.2 in the following chapter.

If a disease is *not* of short duration then (19.1), the equation for the susceptibles, should include birth and death terms. Mortality due to natural causes should also be included in equation (19.2) for the infectives and in (19.3) for the removed class. The resulting models can be analysed in a similar way to that used here and in Chapter 3 on interacting populations: they are all just systems of ordinary differential equations. It is not surprising, therefore, that oscillatory behaviour in disease epidemics is common: these are often referred to as epidemic waves. Here they are *temporal* waves. *Spatial* epidemic waves appear as an epidemic spreads geographically. The latter are also common and we consider them in the next chapter.

Many diseases have a latent or incubation period when a susceptible has become infected but is not yet infectious. Measles, for example, has an 8–13 day latent period. The incubation time for AIDS, on the other hand, is anything from a few months to years after the patient has been shown to have antibodies to the human immunodeficiency virus (HIV). We can, for example, incorporate this as a delay effect, or by introducing a new class, $E(t)$ say, in which the susceptible remains for a given length of time before moving into the infective class. Such models give rise to integral equation formulations and they can exhibit oscillatory behaviour as might be expected from the inclusion of delays. Some of these are described by Hoppensteadt (1975). Nonlinear oscillations in such models have been studied by Hethcote, Stech and van den Driessche (1981). An alternative approach recently used in modelling AIDS will be discussed below in Section 19.4. Finally age, a, is often a crucial factor in disease susceptibility and infectiousness. The models then become partial differential equations with independent variables (t, a): we consider one such model in Section 19.6.

There are many modifications and extensions which can and often must be incorporated in epidemic models; these depend critically on the disease. In the following sections we discuss a few more general models to illustrate different but important points. The books and references already cited describe numerous models and go into them in considerable detail.

CHAPTER 8

Equilibria and Stability Analyses—Nonlinear Models with Multiple Variables

Chapter Goals:

- To construct dynamical models involving nonlinear combinations of multiple variables
- To find equilibria of these models
- To analyze the stability of these equilibria

Chapter Concepts:

- Multivariable equilibria
- Linearization near equilibria
- Stability analyses with multiple equilibria
- Linkage disequilibrium
- Approximate stability analyses

8.1 Introduction

Chapter 7 discussed methods for determining equilibria and their stability properties for models involving multiple variables, but it only considered cases where the dynamical equations were linear functions of the variables. Most interesting biological models are not linear, because any interaction among individuals requires a nonlinear model. For example, Phillips' (1996) model of within-human HIV dynamics involved interactions: virus particles contact and infect CD4+ cells (note the nonlinear $R\ V$ term in (2.4.1)). Similarly, the model of Blower et al. (2000) describing the spread of HIV involved interactions: unprotected sexual contact between infected and uninfected males (note the nonlinear λX terms in (2.5.1)). In this chapter, we describe how the methods of Chapter 7 can be extended to nonlinear models.

In section 8.2 we present nonlinear models with multiple variables in continuous time. Section 8.3 then presents the discrete-time counterpart. Again, we develop the techniques in the context of two variable models for simplicity, and then present general recipes for handling models with an arbitrary number of variables. Finally, section 8.4 illustrates how perturbation techniques can be used in nonlinear models with multiple variables to obtain useful approximations when conducting stability analyses.

8.2 Nonlinear Multiple-Variable Models

We begin with an example to illustrate the process of finding equilibria in nonlinear models with multiple variables. Consider a two-variable model for the spread of a disease that tracks the dynamics of the number of susceptible and infected individuals in a population (denoted by S and I):

$$\begin{aligned} \frac{dS}{dt} &= \theta - dS - \beta SI + \gamma I, \\ \frac{dI}{dt} &= \beta SI - (d + \nu + \gamma)I. \end{aligned} \tag{8.1}$$

For simplicity, we have assumed that the host population is replenished by immigration at total rate, θ, and that recovered individuals immediately become susceptible again (i.e., there is no immunity). In equation (8.1), β denotes the transmission rate of the disease (i.e., $\beta = c\,a$ where c is the rate of contact between susceptible and infected hosts and a is the probability of transmission

given that a contact occurs; Chapter 3), d denotes the per capita background mortality rate of the host, ν denotes the additional mortality that is caused by infection, and γ denotes the rate of clearance of disease through host defense mechanisms. Given these definitions, all parameters in the model are positive.

8.2.1 Finding Equilibria

To identify the equilibria $\hat{S}, \hat{I}$ of the model, we set $dS/dt = 0$ and $dI/dt = 0$. This gives two equilibrium conditions:

$$0 = \theta - d\hat{S} - \beta\hat{S}\,\hat{I} + \gamma\hat{I}, \tag{8.2a}$$

$$0 = \beta\hat{S}\,\hat{I} - (d + \nu + \gamma)\hat{I}. \tag{8.2b}$$

There can be several equilibria in a nonlinear model, unlike a linear model, and we would like to obtain explicit expressions for all of them.

In general, finding all equilibria can be a difficult task because the equations that these equilibria must satisfy might be complicated functions of the dynamic variables. One good place to start is to factor these equations. If we can do this, then we can look for values of the dynamic variables that make any of the factors zero (because the entire expression will then be zero). For the present model we can see that equation (8.2a) cannot be factored, but equation (8.2b) can, giving

$$0 = \hat{I}\{\beta\hat{S} - (d + \nu + \gamma)\}. \tag{8.2c}$$

This simplifies our task because it is much clearer from (8.2c) that dI/dt equals zero only if $\hat{I} = 0$ or $0 = \beta\hat{S} - (d + \nu + \gamma)$. To find an equilibrium of the model as a whole, however, we also require that the dynamic variable S is unchanging (i.e., that equation (8.2a) holds). Therefore, for each of the different ways in which the variable I can be unchanging, we must determine the conditions required for S to be unchanging as well. Specifically, when $\hat{I} = 0$, we must determine if there are conditions under which S is also constant. Similarly, when $0 = \beta\hat{S} - (d + \nu + \gamma)$, we must determine if there are conditions under which S is again constant.

We begin with the case where $\hat{I} = 0$. Substituting this into equation (8.2a), we see that the equation, $0 = \theta - d\hat{S}$ must hold for S to remain constant. This implies that $\hat{S} = \theta/d$. As a result, one equilibrium of this model is

$$\hat{S} = \frac{\theta}{d}, \quad \hat{I} = 0 \tag{8.3a}$$

Similarly, we can substitute $\hat{S} = (d + \nu + \gamma)/\beta$ into equation (8.2a) to obtain another equilibrium (try this)

$$\hat{S} = \frac{d + \nu + \gamma}{\beta}, \quad \hat{I} = \frac{\theta - \dfrac{d}{\beta}(d + \nu + \gamma)}{d + \nu} \tag{8.3b}$$

Equilibrium (8.3a) corresponds to the case where the disease is absent, and equilibrium (8.3b) corresponds to the case where the disease is present, which is often referred to as the *endemic* equilibrium.

Before proceeding, we should step back for a moment and consider how the above procedure works generally. Model (8.1) is based on a specific set of assumptions about how the population of susceptible hosts gets replenished (i.e., by immigration) as well as how susceptible hosts become infected. A general model involving two variables can be written as

$$\begin{aligned} \frac{dS}{dt} &= f(S, I), \\ \frac{dI}{dt} &= g(S, I), \end{aligned} \tag{8.4}$$

where $f(S,I)$ and $g(S,I)$ can be any functions of the variables S and I. To find the equilibria of this model, we again need to identify values of the variables $\hat{S}$ and $\hat{I}$ that result in no change in either variable. That is, we set $dS/dt = 0$ and $dI/dt = 0$ to get two equilibrium conditions, which we must solve for the two unknowns $\hat{S}$ and $\hat{I}$:

$$0 = f(\hat{S}, \hat{I}), \tag{8.5a}$$

$$0 = g(\hat{S}, \hat{I}). \tag{8.5b}$$

The equilibrium conditions (8.5) describe the null clines of the model (Figure 8.1). If (8.5a) holds, then S remains constant. If (8.5b) holds, then I remains constant. But unless both equilibrium conditions hold, changes to one variable will typically lead to changes in the other. Therefore, for both variables to remain constant, an equilibrium must simultaneously satisfy both equilibrium conditions. Graphically, this means that any equilibrium must lie on null clines for every variable in a model (see filled circles in Figure 8.1).

At a *multivariable equilibrium*, all variables must remain unchanged. There can be multiple equilibria in nonlinear models.

Depending on the functions f and g, identifying the possible equilibria of a model can be straightforward, difficult, or even impossible. A good strategy is to factor each of the equilibrium conditions, identify which one is easiest to solve for its null clines, and then plug in these null clines, one by one, into the other equilibrium condition to see if it can be solved as well.

These procedures can be generalized for models involving any number of variables:

Definition 8.1: General Nonlinear Models in Continuous Time

A general, nonlinear, continuous-time model with n dynamic variables $x_1, \ldots, x_n$ can be written as

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \ldots, x_n), \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \ldots, x_n), \\ &\vdots \\ \frac{dx_n}{dt} &= f_n(x_1, x_2, \ldots, x_n), \end{aligned}$$

where $f_1, f_2, \ldots, f_n$ denote different functions specifying the rate of change of each variable.

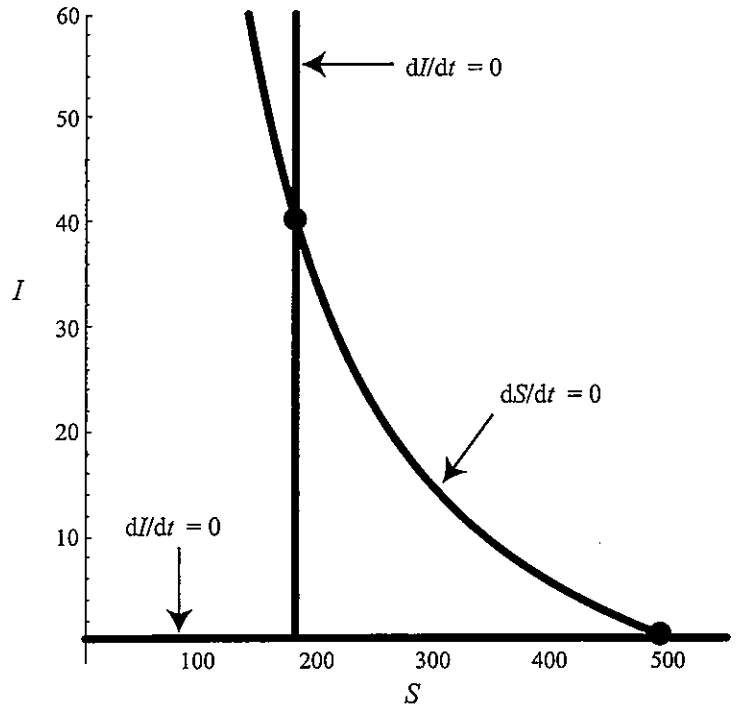

Figure 8.1: Null clines for the disease transmission model. For model (8.1), the equilibrium condition (8.2a) defines the null cline along which the number of susceptible individuals does not change ($dS/dt = 0$). The equilibrium condition (8.2b) defines two null clines along which the number of infected individuals does not change ($dI/dt = 0$). An equilibrium must lie on the null clines for both variables, which occurs at $\hat{S} = 180$, $\hat{I} = 40$ and at $\hat{S} = 500$, $\hat{I} = 0$ (filled circles). Parameter values are $\nu = 0.7$, $d = 0.1$, $\gamma = 0.1$, $\beta = 0.005$, $\theta = 50$.

Recipe 8.1 then describes how to find equilibria of models of the form in Definition 8.1:

Recipe 8.1: Equilibria of Nonlinear Multivariable Models in Continuous Time

Equilibria are found by determining the values of the variables that cause all of the variables to remain constant: $dx_1/dt = 0$, $dx_2/dt = 0, \ldots, dx_n/dt = 0$. This results in n equations in n unknowns: $f_1(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$, $f_2(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0, \ldots, f_n(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$. Any point that satisfies *all* of these conditions simultaneously is an equilibrium. To identify the equilibria:

Step 1: Factor each equation, $f_i(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$.

Step 2: Identify all possible solutions to one of these equations (start with the simplest one).

Step 3: Plug each possible solution into the remaining equations, and repeat the above steps until equilibrium values for all variables are identified.

For nonlinear models, there can be more than one equilibrium. Depending on the complexity of the model, it may or may not be possible to identify all of the equilibria explicitly.

8.2.2 Determining the Stability of the Equilibria

Once the equilibria have been identified, our next task is to determine the behavior of the model if we move the system slightly away from one of these points. In other words, we want to know when each equilibrium is *locally stable* (Definition 5.2). Stability conditions can be used to address very fundamental biological questions. For example, in our example model, can the disease spread and become established within a population? We can answer this question by asking if the equilibrium without the disease, $\hat{S} = \theta/d, \hat{I} = 0$, is unstable following the introduction of a small number of infected individuals.

To address this question, we imagine starting the model very close to an equilibrium. We define ε_S and ε_I to be the *displacements* of the starting values of the variables from the equilibrium $\hat{S}, \hat{I}$. That is, $\varepsilon_s = S - \hat{S}$ and $\varepsilon_I = I - \hat{I}$ where S and I are the numbers of susceptible and infected individuals, respectively. The terms ε_S and ε_I are dynamic variables that change over time as S and I change. These displacements will get larger or smaller depending on whether the system moves away from the equilibrium (an unstable equilibrium) or toward the equilibrium (a stable equilibrium). Thus, to determine whether the displacements grow or decay we need equations describing how the displacements change through time.

Differential equations for the displacements can be obtained by differentiating ε_S and ε_I with respect to time:

$$\begin{aligned}\frac{d\varepsilon_S}{dt} &= \frac{d}{dt}(S - \hat{S}) \\ &= \frac{dS}{dt} \\ &= \theta - dS - \beta S I + \gamma I\end{aligned} \tag{8.6a}$$

and

$$\begin{aligned}\frac{d\varepsilon_I}{dt} &= \frac{d}{dt}(I - \hat{I}) \\ &= \frac{dI}{dt} \\ &= \beta SI - (d + \nu + \gamma) I.\end{aligned} \tag{8.6b}$$

These results follow from the fact that $\hat{S}$ and $\hat{I}$ are constants and hence their derivatives with respect to time are zero. Consequently, the dynamics of the displacements are governed by the same equations that govern the dynamics of the variables S and I. We have not quite finished our derivation, however, because the right-hand side of each equation is written in terms of the original variables S and I. To have a self-contained model for the displacements ε_S and ε_I, we need to rewrite equations (8.6) in terms of ε_S and ε_I alone. This can be accomplished by using the fact that $S = \hat{S} + \varepsilon_S$ and $I = \hat{I} + \varepsilon_I$, giving

$$\begin{aligned}\frac{d\varepsilon_S}{dt} &= \theta - d(\hat{S} + \varepsilon_S) - \beta(\hat{S} + \varepsilon_S)(\hat{I} + \varepsilon_I) + \gamma(\hat{I} + \varepsilon_I) \\ &= \theta - d\hat{S} - d\varepsilon_S - \beta\hat{S}\hat{I} - \beta\hat{I}\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\varepsilon_S\varepsilon_I + \gamma\hat{I} + \gamma\varepsilon_I,\end{aligned} \tag{8.7a}$$

$$\frac{d\varepsilon_I}{dt} = \beta(\hat{S} + \varepsilon_S)(\hat{I} + \varepsilon_I) - (d + \nu + \gamma)(\hat{I} + \varepsilon_I)$$
$$= \beta\hat{S}\hat{I} + \beta\hat{I}\varepsilon_S + \beta\hat{S}\varepsilon_I + \beta\varepsilon_S\varepsilon_I - d\hat{I} - \nu\hat{I} \qquad (8.7b)$$
$$- \gamma\hat{I} - d\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I.$$

We are almost done, but we can now simplify (8.7) by using the fact that the displacements ε_S and ε_I are small. As a consequence, higher powers in these terms (e.g., ε_S^2, ε_I^2, and $\varepsilon_S\varepsilon_I$) will be extremely small and thus we can ignore them. Doing so, and grouping terms together that involve the displacements ε_S and ε_I, we get

$$\frac{d\varepsilon_S}{dt} = (\theta - d\hat{S} - \beta\hat{S}\hat{I} + \gamma\hat{I}) + (-d\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\hat{I}\varepsilon_S + \gamma\varepsilon_I), \qquad (8.8a)$$

$$\frac{d\varepsilon_I}{dt} = (\beta\hat{S}\hat{I} - d\hat{I} - \nu\hat{I} - \gamma\hat{I}) + (-d\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I + \beta\hat{S}\varepsilon_I + \beta\hat{I}\varepsilon_S). \qquad (8.8b)$$

The beauty of writing the equations in this way is that the first parenthetical term in each equation is identical to one of the equilibrium conditions, and therefore it must be zero. The equilibrium condition (8.2a) tells us that $(\theta - d\hat{S} - \beta\hat{S}\hat{I} + \gamma\hat{I})$ is zero, while the equilibrium condition (8.2b) tells us that $(\beta\hat{S}\hat{I} - d\hat{I} - \nu\hat{I} - \gamma\hat{I})$ is zero. As a result, the dynamics of the small displacements ε_S and ε_I near any equilibrium point of the original model are governed by the equations

$$\frac{d\varepsilon_S}{dt} = -d\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\hat{I}\varepsilon_S + \gamma\varepsilon_1, \qquad (8.9a)$$

$$\frac{d\varepsilon_I}{dt} = -d\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I + \beta\hat{S}\varepsilon_I + \beta\hat{I}\varepsilon_S. \qquad (8.9b)$$

The most important point to take away from the above calculations is that we started with a *nonlinear* model and we have arrived at a *linear* system of equations that governs the dynamics of the displacements from an equilibrium. We have been able to do this because we allow only small displacements, and therefore we can approximate the nonlinear model with a linear one near the equilibrium by ignoring higher-powered terms in the ε's. This process is called a *local stability analysis* or *linearization* near an equilibrium. Given that we have a linear system of equations, we can then use the techniques from Chapter 7 to determine whether the equilibrium is stable or unstable. If the displacements get smaller with time then the equilibrium is locally stable.

The stability of an equilibrium is determined by a local stability analysis, which is a *linearization* of the nonlinear model near the equilibrium of interest.

As in Chapter 7, we can write equations (8.9) in matrix form:

$$\begin{pmatrix} \frac{d\varepsilon_S}{dt} \\ \frac{d\varepsilon_I}{dt} \end{pmatrix} = \begin{pmatrix} -\beta\hat{I} - d & -\beta\hat{S} + \gamma \\ \beta\hat{I} & \beta\hat{S} - (d + \nu + \gamma) \end{pmatrix} \begin{pmatrix} \varepsilon_S \\ \varepsilon_I \end{pmatrix}. \qquad (8.10)$$

Matrix equation (8.10) describes the dynamics of small displacements from any equilibrium of the original model. According to Rule 7.2 for continuous-time linear models, the equilibrium is stable provided that all eigenvalues have

negative real parts. This ensures that the system moves toward the equilibrium along all eigenvectors.

To use the above results for determining local stability properties, we must first specify which equilibrium we are near. Let us begin with the equilibrium where the disease is absent. The stability of this equilibrium is determined by the eigenvalues of the matrix in (8.10). For the disease-absent equilibrium, $\hat{S} = \theta/d, \hat{I} = 0$, and this matrix therefore simplifies to

$$\begin{pmatrix} -d & -\beta\frac{\theta}{d} + \gamma \\ 0 & \beta\frac{\theta}{d} - (d + \nu + \gamma) \end{pmatrix}. \tag{8.11}$$

Matrix (8.11) is upper triangular, and therefore its eigenvalues can be read directly from the diagonal (Rule P2.26):

$$r_1 = \beta\frac{\theta}{d} - (d + \nu + \gamma) \quad \text{and} \quad r_2 = -d. \tag{8.12}$$

Because d is a positive constant, the stability of the disease-absent equilibrium is completely determined by the sign of the first eigenvalue $\beta(\theta/d) - (d + \nu + \gamma)$. If this is negative, then the equilibrium will be stable and the disease will not spread. If it is positive, then the disease will spread into the population.

The requirement for a disease to spread, $\beta(\theta/d) - (d + \nu + \gamma) > 0$, can be rewritten in terms of the number of susceptible individuals $\hat{S} = \theta/d$. Making this substitution and rearranging, we get the condition $\beta\hat{S}/(d + \nu + \gamma) > 1$ for a disease to spread. The quantity on the left-hand side is sometimes called R_0, the reproductive number, and represents the expected number of new infections produced per infected host when a disease is introduced into a susceptible population. For R_0 to be greater than one, the population size in the absence of the disease, $\hat{S}$, must be large enough that an infected individual encounters enough susceptible individuals to ensure infection of at least one other member of the population before the infected individual dies or recovers.

When there are *multiple equilibria*, the *stability* of each must be evaluated separately.

We have just completed a local stability analysis of the $\hat{S} = \theta/d, \hat{I} = 0$ equilibrium, but this model has a second equilibrium: the endemic equilibrium $\hat{S} = (d + \nu + \gamma)/\beta, \hat{I} = (\theta - (d/\beta)(d + \nu + \gamma))/(d + \nu)$. Each equilibrium of a model has its own stability properties, and we must identify these properties by performing a local stability analysis for each equilibrium separately.

Before we begin, we should determine the conditions under which the second equilibrium is biologically feasible (i.e., biologically valid) by asking when $\hat{S}$ and $\hat{I}$ are both positive. Given that the parameters are assumed to be positive, $\hat{S}$ will always be positive, but $\hat{I}$ need not be. For $\hat{I}$ to be positive, we require that $\theta > (d/\beta)(d + \nu + \gamma)$, which can be reorganized as $\beta(\theta/d) - (d + \nu + \gamma) > 0$. This is exactly the same condition required for the disease-absent equilibrium to be unstable. Therefore, the endemic equilibrium is biologically feasible only when the disease can spread when rare.

Once again, we use the linear version of the model (8.10) that describes the dynamics near an equilibrium for our stability analysis, but we now plug in the second equilibrium. After factoring, equation (8.10) becomes

$$\begin{pmatrix} \frac{d\varepsilon_S}{dt} \\ \frac{d\varepsilon_I}{dt} \end{pmatrix} = \begin{pmatrix} \frac{-\beta\theta + \gamma d}{d + \nu} & -\nu - d \\ \frac{\beta\theta - d(d + \nu + \gamma)}{d + \nu} & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_S \\ \varepsilon_I \end{pmatrix}. \tag{8.13}$$

Again, according to Rule 7.2, the equilibrium of (8.13) will be stable only if all of the eigenvalues of the matrix have negative real parts. We could calculate the eigenvalues of (8.13) directly and then examine their real parts, but it is easier to use the trace and determinant conditions in Rule P2.25 of Primer 2. According to Rule P2.25, for the real part of both eigenvalues to be negative, the determinant of a 2×2 matrix must be positive and its trace must be negative.

The determinant of the matrix in (8.13) is $\beta\theta - d(d + \nu + \gamma)$. The sign of this expression might not be obvious at first, but recall that we require that $\beta(\theta/d) - (d + \nu + \gamma) > 0$ for $\hat{I}$ to be positive. This implies that the determinant is positive whenever the endemic equilibrium is biologically feasible. The trace of (8.13) is $(-\beta\theta + \gamma d)/(d + \nu)$, and because all the parameters are positive, the sign of the trace depends on the sign of $(-\beta\theta + \gamma d)$. The fact that $\hat{I}$ must be positive implies that $\beta\theta$ must be greater than $d^2 + d\nu + d\gamma$ (from our condition for biological feasibility), which in turn implies that $\beta\theta$ is greater than $d\gamma$. Therefore, $(-\beta\theta + \gamma d)$ must be negative, meaning that the trace must be negative. Thus, according to Rule P2.25, the real parts of both eigenvalues are negative, and we conclude that the endemic equilibrium is locally stable whenever it is biologically feasible.

The above example illustrates how we can use the techniques of Chapter 7 for linear models to conduct a local stability analysis of our nonlinear epidemiological model. The approach rests on the assumption that our linear approximation adequately captures the dynamics of the nonlinear model near equilibria. Figure 8.2 shows that, indeed, this linearization provides a remarkably good approximation to the nonlinear model (8.1) near both equilibria. In fact, mathematicians have proven that such approximations will typically work for local stability analyses of general nonlinear models as well. These general results are presented next.

8.2.3 The General Approach for Determining the Local Stability of Equilibria

In the above derivation, we approximated the dynamics near an equilibrium with linear equations by first writing out nonlinear equations (8.7) and then dropping higher-order terms in the displacement, like ε_S^2, ε_I^2, and $\varepsilon_S\varepsilon_I$. In other models, however, it might not be so obvious which terms should be kept and which can be dropped. For example, what should we do with a term like $(1 + \varepsilon_I)/(1 - \varepsilon_S)$? Fortunately, there is a much simpler route to reaching matrix (8.10b), both conceptually and computationally, which automatically drops the correct terms.

We describe the procedure using the more general two-variable model (8.4). Having found the equilibria of a model, we can again define the displacements ε_S and ε_I from one of the equilibria and derive equations for their dynamics.