# Chapter 3

# EPIDEMICS

In this chapter mathematical models will be studied that describe the spread of epidemics in a population. Although great mathematicians such as Euler and D. Bernoulli have already tried to describe the dynamics of epidemics by mathematical methods, the modern mathematical theory probably received impetus when papers by Kermack and McKendrick (1927, 1932, 1933) were published. Such models, even the most primitive ones, may help to find those points where one may most successfully fight an epidemic or forecast how it will pass. We deal first with the classical and simplest, the so-called SIR (Susceptibles, Infectives, Removed) model, then we consider the case of sexually transmitted diseases and so-called SIS (Susceptibles, Infectives, Susceptibles) models. For sexually transmitted diseases we treat the problem of pair formation in the human population. Finally, the spread of epidemics through space will be studied. Important references in the field are Murray (1989) and Capasso (1993).

## 3.1 The Spread of Diseases and Susceptibles/Infectives/Removed Models

These models are simple but still may yield some insight into the dynamics of a contagious illness in a densely populated city, an army barrack, or a student dormitory. Basic assumptions for such a model are:

(i) the total population is constant; the epidemic does not have a recognizable influence on population numbers;

(ii) the population is "well stirred," meaning that every individual has an equal chance to meet any other member of the population; and

(iii) any person in the population who caught the disease either obtained immunity or died (from the point of view of the individual this difference is far from being irrelevant but for the model, *horribile dictu*, the effect is the same).

We denote the number of *susceptibles, infectives* and those who obtained immunity (or died) called *removed* at time $t$ by $S(t)$, $I(t)$ and $R(t)$, respectively. We assume that susceptibles move into the group of infectives through

infection and a decrease in their number in unit time is proportional to the number of encounters of a susceptible and an infective individual and this, in turn, is proportional to their respective numbers. Denoting the *infection rate* by $r > 0$, the differential equation governing the variations in the number of susceptibles is then $\dot{S} = -rSI$. The class of infectives is recruited from the susceptibles by incorporating those who leave the latter class, and it is decreased by recovery from the disease (or death). Denoting the *recovery rate* by $a > 0$, the differential equation for the infectives is $\dot{I} = rSI - aI$; finally, the recovered follow the equation $\dot{R} = aI$. Thus, we have arrived at a 3D system of differential equations

$$\dot{S} = -rSI, \qquad \dot{I} = rSI - aI, \qquad \dot{R} = aI. \tag{3.1.1}$$

This system satisfies requirement (i), where the total population has constant size because by adding the three equations we get $(S + I + R)^{\cdot} = 0$, that is, the sum of the sizes of the three classes $N := S + I + R$ is constant. It is to be noted that in this model there is no latent period for the illness; a susceptible person who has contracted the disease becomes infective immediately. If incubation is short this abstraction may be accepted. The initial conditions attached to the system are $S_0 = S(0) > 0, I_0 = I(0) > 0, 0 = R(0)$ with $S_0 + I_0 = N$. Usually, one may assume that the disease starts with a small number of infectives, that is, $I_0$ is small with respect to $S_0$. The basic problem is whether or not the number of infectives will increase at the beginning. We say that we have an *epidemic* on hand if at the outset the number of infectives is increasing, which means that people get infected faster than they recover. It is easy to give a condition for that because from the second equation

$$\dot{I}(0) = I_0 (rS_0 - a) > \text{ respectively } < 0,$$

according to

$$S_0 > \text{ respectively } < a/r.$$

From the first of Eqs. (3.1.1) it is clear that $\dot{S}$ is always negative, so that $S(t) < S_0$ for $t > 0$. Thus, if $S_0 < a/r$ then $\dot{I}(t) = I(t)(rS(t) - a) < I(t)(rS_0 - a) < 0$ for $t \geq 0$; this means that if at the very beginning the number of infectives was decreasing this will remain. On the other hand, if the condition for an epidemic holds at the beginning then the number of infectives will increase for at least some time. The threshold parameter $a/r$ is called the *relative recovery rate*, which is the percentage of those recovered in unit time divided by the percentage of those infected by a single infective in unit time. We shall give here an intuitive interpretation of the condition of the outbreak of an epidemic. First it is to be noted that the reciprocal $1/a$ of the recovery rate can be interpreted as the average infectious period of an infective or the average time needed for recovery. This can be seen from the third equation–if we have just one infective at time $t$, that is, $I(t) = 1$ then during the time interval $1/a$ the number of recovered will change by $R(t + 1/a) - R(t) \approx \dot{R}(t)\, 1/a = I(t) = 1$, meaning that the single infective recovers. From the first equation during the time interval $1/a$ the number of those getting the infection is $S(t + 1/a) - S(t) \approx \dot{S}(t)\, 1/a = -(r/a) S(t)$. If this number is $< -1$, one infective passes on the disease during his infectious period to more than one person, thus if $(r/a) S(t) > 1$, implying that

$$S_0 r/a > 1, \tag{3.1.2}$$

then more persons get ill in unit time than recover; thus an epidemic breaks out. The ratio $r/a$, called the *contact rate*, is the infection rate multiplied by the average infectious period. The contact rate multiplied by the number of susceptibles gives the number of those infected by one infective during the infectious period of the latter. We determine now the projections of the trajectories of system (3.1.1) on the $S, I$ plane. For this we divide the second equation by the first one to obtain the differential equation of the trajectories, $dI/dS = -1 + a/(rS)$ . It is easy to integrate yielding, as the equation of the trajectories,

$$I = \ln(S^{a/r}) - S + c \tag{3.1.3}$$

where $c = I_0 + S_0 - (a/r) \ln S_0$. It is easy to see that for all these trajectories $\max I(S) = I(a/r)$, provided that $a/r$ is $< S_0$ (see Fig. 3.1.1 produced by MAPLE).

Thus, we see that if the number of susceptibles is greater than this threshold value at the start then the number of infectives will rise at the beginning until the number of susceptibles decreases enough to reach this value. If the number of susceptibles is less than the threshold value at the beginning, then the number of infectives is going down right away. By dividing the first equation by the third one, it is easy to obtain $S$ as a function of $R$; expressing $I$ in the third equation by $S$ and $R$ and substituting $S(R) = S_0 \exp(-Ra/r)$ obtains a first-order scalar differential equation for $R$. Solving the latter by MAPLE, we obtain the number of removed as a function of time (Fig. 3.1.2, thick line). Finally, the number of removed per unit time (the mortality in case of a lethal disease), that is, the derivative of $R$ with respect to time is shown in Fig. 3.1.2 (dotted line).

As the figures show even this simple model is able to help us in forecasting the number of all those who will catch the disease during the epidemic, the time when the number of infectives will be maximal, the time when the mortality will be maximal, and the time when the epidemic may be considered to be over etc., provided that we know the infection and recovery rates and the initial data. Naturally, refinements can be made in several directions–to take into consideration the incubation period, count the "carriers" who are not ill themselves but spread the disease, and consider the age structure etc.

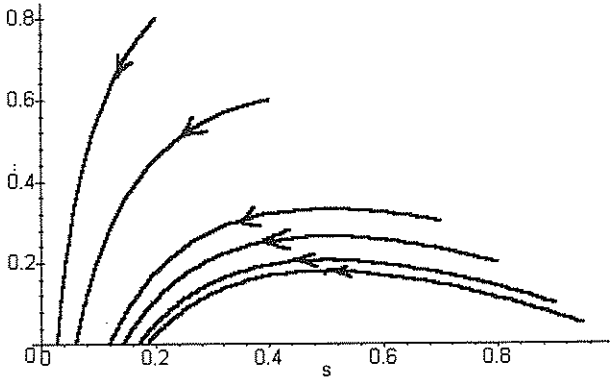


Figure 3.1.1: The spread of disease using the SIR model (3.1.1) with total population 1, infection rate $r = 2$, recovery rate $a = 1$, and relative recovery rate $a/r = 0.5$; the number of infectives versus the number of susceptibles in the first 10 time units of the outbreak (MAPLE).

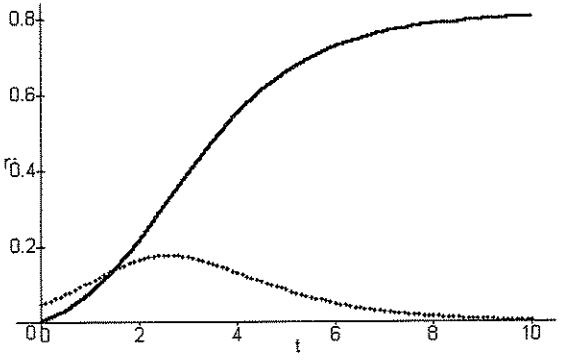


Figure 3.1.2: Thick line is the number of removed in the first 10 units of time and the dotted line is the mortality (in case of a lethal disease) in the first 10 units of time $I(0) = 0.05$ (MAPLE).



## 3.2 Sexually Transmitted Diseases

Venereal diseases differ from other epidemics prevalent in the human population in that the population is divided into two groups, males and females, and the disease is normally transmitted only from a member of one of the groups to a member of the *other* group, that is, males transmit the disease to females and vice versa. Further, because no immunity is conferred by going through the disease, if a person passes from the susceptible group to the infective one then following recovery from the disease he/she becomes susceptible again. Models without immunity are called SIS models. Naturally, if one considers AIDS lethal then it does not fall into this category. The model we handle in this Section considers a disease such as gonorrhea. Many mathematical models of this particular illness were presented by Hethcote and Yorke (1984). We present the simplest one.

We assume that the total population of sexually active males and females is constant. The number of susceptible males and females at time $t$ is denoted by $S_1(t)$ and $S_2(t)$, respectively, and the number of infective males and females by $I_1(t)$ and $I_2(t)$. By assumption $S_1(t) + I_1(t) = N_1$, $S_2(t) + I_2(t) = N_2$, where $N_1$ and $N_2$ are constants. The number of susceptible males decreases in unit time by those who get infected by infective females and vice versa, the number of males getting the infection in unit time is supposed to be proportional to the number of susceptibles and to the number of infective females. The number of susceptible males/females increases by those who recover from the disease–the number of those recovered in unit time is proportional to the number of infective males/females. The number of infectives increases in unit time by those who get the infection and decreases by those who recover. This way we arrive at the following four-dimensional (4D) system of differential equations:

$$\begin{aligned} \dot{S}_1 &= -r_1 S_1 I_2 + a_1 I_1 , \qquad \dot{S}_2 = -r_2 S_2 I_1 + a_2 I_2 \\ \dot{I}_1 &= r_1 S_1 I_2 - a_1 I_1 , \qquad \dot{I}_2 = r_2 S_2 I_1 - a_2 I_2 , \end{aligned} \tag{3.2.1}$$

where $r_1, a_1$ and $r_2, a_2$ are the infection and recovery rates of males, and females, respectively (cf. system (3.1.1)). Taking into account that the sum of susceptible and infective males/females is constant, this system can be reduced to a 2D one:

$$\dot{I}_1 = r_1(N_1 - I_1)I_2 - a_1 I_1 , \qquad \dot{I}_2 = r_2(N_2 - I_2)I_1 - a_2 I_2 . \tag{3.2.2}$$

This system is such that if $I_1 = 0$ then $\dot{I}_1$ is positive and, similarly, if $I_2 = 0$ then $\dot{I}_2$ is positive. This means that the positive quadrant of the $I_1, I_2$ plane is positively invariant–no trajectory may leave the positive quadrant. The system has two equilibria: $(0, 0)$ and

$$\begin{aligned} \left(\bar{I}_1, \bar{I}_2\right) = &\left(\left(N_1 N_2 - a_1 a_2 / (r_1 r_2)\right) / \left(N_2 + a_1/r_1\right) , \right. \\ &\left. \left(N_1 N_2 - a_1 a_2 / (r_1 r_2)\right) / \left(N_1 + a_2/r_2\right)\right) , \end{aligned}$$

with the last one in the positive quadrant iff $N_1 N_2 - a_1 a_2 / (r_1 r_2) > 0$, or

$$\left(N_1 r_1 / a_2\right)\left(N_2 r_2 / a_1\right) > 1 . \tag{3.2.3}$$

The last inequality has a clear intuitive meaning. The first factor on the left-hand side is the number of males multiplied by the contact rate with respect to males, that is the infection rate of males multiplied by the infectious period of an infective *female*. Thus, it gives the number of males infected by an infective female during her infective period. (To be sure, here $N_1$ ought to be replaced by the number of *susceptible* males but if the disease is not too widely spread the difference is slight.) The second factor has an analogous meaning for females. Hence, we see that the condition for the existence of an *endemic equilibrium* $(\bar{I}_1, \bar{I}_2)$ in the positive quadrant is that on the average one infective shall infect more than one person during his/her infective period. Naturally, the main question is whether the epidemic free state $(0,0)$ or the endemic state $(\bar{I}_1, \bar{I}_2)$ is stable. In order to be able to tell this we linearize the system at the two equilibria and apply the Routh-Hurwitz criterion (see Theorem A1.1.2). At $(0,0)$ we leave this to the reader as an exercise. It turns out by an easy calculation that if (3.2.3) is reversed then the epidemic free state is asymptotically stable, if (3.2.3) holds then it is unstable (as a matter of fact, a saddle point). The calculation for the endemic state is more tiresome. From now on (3.2.3) is assumed. The characteristic polynomial is

$$\begin{vmatrix} -a_1 - r_1 \frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_1 + a_2/r_2} - \lambda & r_1\left(N_1 - \frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_2 + a_1/r_1}\right) \\ r_2\left(N_2 - \frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_1 + a_2/r_2}\right) & -a_2 - r_2 \frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_2 + a_1/r_1} - \lambda \end{vmatrix}$$

$$= \lambda^2 + \lambda\left(a_1 + a_2 + r_1\frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_1 + a_2/r_2} + r_2\frac{N_1N_2 - a_1a_2/(r_1r_2)}{N_2 + a_1/r_1}\right)$$
$$+ \frac{(N_1N_2 - a_1a_2/(r_1r_2))(a_1r_2N_1 + a_2r_1N_2 + r_1r_2N_1N_2 + a_1a_2)}{(N_2 + a_1/r_1)(N_1 + a_2/r_2)}.$$

By (3.2.3) both the coefficient of $\lambda$ and the "constant term" are positive, hence with Theorem A1.1.2 the endemic equilibrium (if it exists in the interior of the positive quadrant) is asymptotically stable.

Condition (3.2.3) of an asymptotically stable endemic equilibrium shows clearly how the decrease of the average infectious period $1/a_1$ and $1/a_2$, respectively, and/or the infection rate $r_1$ and $r_2$, respectively, may destabilize the endemic state and lead to a disease-free state. If, for instance, we assume that the average infectious period of a male is 1.5 months and of a female is 3.5 months, respectively, the respective infection rates are $r_1 = 1.4 \cdot 10^{-8}$ per month and $r_2 = 3.8 \cdot 10^{-8}$ per month, and both the total sexually active male and female population are equal to $20 \cdot 10^6$, then the left-hand side of (3.2.3) is $1.127 > 1$ and the stable endemic equilibrium is $(\bar{I}_1, \bar{I}_2) = (0.515 \cdot 10^6, 1.27 \cdot 10^6)$. If the average infectious period of women could be brought down to 2.5 months then the left-hand side of (3.2.3) would become $0.805 < 1$, the endemic equilibrium would disappear, and the disease-free state would become stable.



## 3.3 A Model of Pair Formation

Early models on the dynamics of sexually transmitted diseases including the one treated in the previous Section have the disadvantage of being based on the assumption of "well stirredness" of the population, that is, it is assumed that each member of the population has an equal chance to meet any other member and mating is completely random. This assumption is clearly false for the human population, although there might be small highly promiscuous subgroups in which this assumption may be a not completely incorrect approximation of prevailing behavior. To better approximate reality one has to study how pairs are formed, how a sexual partnership emerges and how it ceases to exist. If a female and a male form a pair and they do not carry the disease then they can be considered immune until one of them does not have contact with another partner. The duration of these partnerships and the time interval between two partnerships of a given person are to be taken into account. If the problem of *pair formation* has been settled then one may build a model based on it to describe the spread of a sexually transmitted disease. Several models of pair formation can be found in the literature, with some of them dividing up the population to several less or highly promiscuous subgroups in which pair formation is governed by different laws. We present here a fairly simple one due to Dietz and Hadeler (1988) (see also the references therein) which is highly instructive, and then we shall describe how an epidemic model can be built upon it.

Denote the density of single females and males by $x$ and $y$, respectively, suppose that their densities are increasing by constant rates $\kappa_x$ and $\kappa_y$, respectively, due to the aging of younger generations into a sexually active population (we do not consider here dependence of the birth rate on the density of the population) and that these densities decrease due to deaths proportional to the numbers with mortalities $\mu_x$ and $\mu_y$, respectively. Denote the density of pairs formed by a female and a male by $p$. At this point we must decide what we consider a pair–how we define a pair. If we want to use the model for describing the spread of sexually transmitted diseases then the social or religious aspects of pair formation have to be disregarded. We say that a *pair* is formed when a female and a male have sexual contact with each other the first time and this pair ceases to exist the first time one of the members has sexual contact outside the pair. Denote the *divorce rate* (considered to be a positive constant) by $\sigma$. The number of single females and males increases by one when a pair separates, and the number of single males and females, respectively, increases by one when the female or male member of a pair dies. Pairs are recruited from the single population. The rate of pair formation is a function of $x$ and $y$; it is sometimes called the *marriage function* and will be denoted by $\varphi(x,y)$. It has to satisfy certain natural conditions: (i) it is defined for nonnegative values of $x$ and $y$ and must be zero when either the females or the males are absent, $\varphi(0,y) = \varphi(x,0) = 0$; (ii) it is increasing (or, rather, not decreasing) if the number of single females or males is increasing, $\varphi_x, \varphi_y \geq 0$ ; (iii) if the densities of both single females and males are increasing $\alpha > 0$ times the rate of pair formation also increases $\alpha$ times; this means that the function $\varphi$ is homogeneous of

degree one: $\varphi(\alpha x, \alpha y) = \alpha\varphi(x,y)$; and (iv) $\varphi$ is symmetric, $\varphi(x,y) = \varphi(y,x)$. With these assumptions and notations we arrive at the following system of differential equations:

$$\begin{aligned} \dot{x} &= \kappa_x - \mu_x x + (\mu_y + \sigma)\, p - \varphi(x,y) \\ \dot{y} &= \kappa_y - \mu_y y + (\mu_x + \sigma)\, p - \varphi(x,y) \\ \dot{p} &= -(\mu_x + \mu_y + \sigma)\, p + \varphi(x,y)\ . \end{aligned} \tag{3.3.1}$$

Here we have also tacitly assumed that the mortality of singles is equal to the mortality of those in pairs, even though sociological data suggest that people in permanent partnership live longer than singles. There are several possible ways to choose a marriage function that satisfies conditions (i)-(iv); here we suppose that *the rate of pair formation depends linearly on the density of that sex that is in minority.* The choice representing this assumption is

$$\varphi(x,y) := \rho \min(x,y) = \begin{cases} \rho x, & \text{if} \quad (x,y) \in K_y \\ \rho y, & \text{if} \quad (x,y) \in K_x\ , \end{cases} \tag{3.3.2}$$

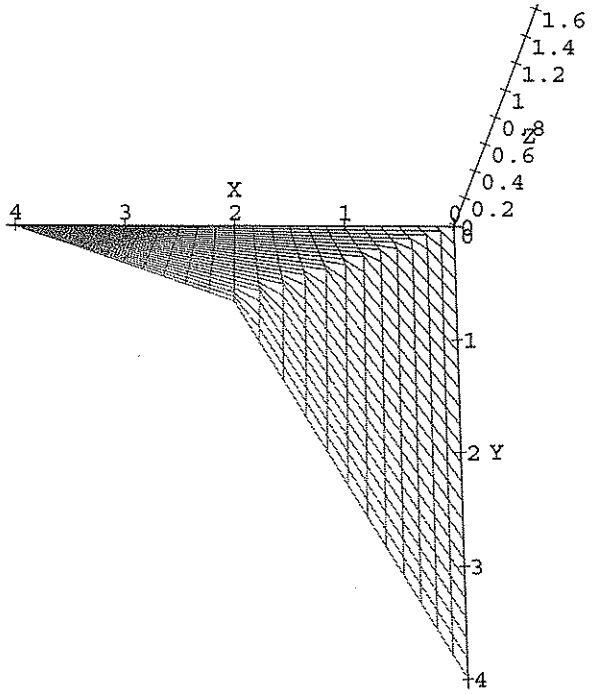


Figure 3.3.1: The marriage function "minimum" (MAPLE).

where $K_y = \{(x,y) : 0 \le x \le y\}$, $K_x = \{(x,y) : 0 \le y \le x\}$, and $\rho$ is a positive constant. The graph of this function, which looks like the part of a roof,



can be seen in Fig. 3.3.1 (produced by MAPLE). It is made up by two planes intersecting over the line $y = x$ of the plane $x, y$.

In most societies the number of females is increasing faster than the number of males and the mortality of women is lower than that of the men. Therefore we assume in the sequel that

$$\kappa_x \ge \kappa_y \text{ and } \mu_x \le \mu_y\ . \tag{3.3.3}$$

This way we arrive at the so-called *male dominance model.* Under condition (3.3.3) the set $K_x$ representing female majority is positively invariant, that is, no trajectory may leave $K_x$ because on the boundary $x = y$ we have

$$\begin{aligned} (x-y)^{\cdot} &= \kappa_x - \kappa_y - \mu_x x + \mu_y y + (\mu_y - \mu_x)p \\ &\ge \kappa_x - \kappa_y - \mu_x (x-y) + (\mu_y - \mu_x)p \\ &= \kappa_x - \kappa_y + (\mu_y - \mu_x)p \ge 0, \end{aligned}$$

and thus, if the number of females was greater than the number of males, this stays so, and on the boundary $y = 0$ we have from the second equation of system (3.3.1): $\dot{y} = \kappa_y + (\mu_x + \sigma)\, p > 0$ . In this case we may replace the function $\varphi$ by $\rho y$ in the system over $K_x$ and we have to deal with the linear system

$$\begin{aligned} \dot{x} &= \kappa_x - \mu_x x - \rho y + (\mu_y + \sigma)\, p \\ \dot{y} &= \kappa_y - (\mu_y + \rho) y + (\mu_x + \sigma)\, p \\ \dot{p} &= \rho y - (\mu_x + \mu_y + \sigma)\, p\,, \end{aligned} \tag{3.3.4}$$

$(x,y) \in K_x, \quad p \ge 0$ . This model can be explicitly calculated. It has a unique equilibrium:

$$\begin{aligned} &(\bar{x}, \bar{y}, \bar{p}) \\ &= \left( \frac{\kappa_x}{\mu_x} - \frac{\kappa_y}{\mu_y} \frac{\rho}{\mu_x + \mu_y + \sigma + \rho}, \frac{\kappa_y}{\mu_y} \frac{\mu_x + \mu_y + \sigma}{\mu_x + \mu_y + \sigma + \rho}, \frac{\kappa_y}{\mu_y} \frac{\rho}{\mu_x + \mu_y + \sigma + \rho} \right). \end{aligned}$$

It is easy to see that because of (3.3.3) $(\bar{x}, \bar{y}) \in K_x$. We note that in equilibrium the number of females and the number of males is $\bar{f} = \bar{x} + \bar{p} = \kappa_x/\mu_x$ and $\bar{m} = \bar{y} + \bar{p} = \kappa_y/\mu_y < \bar{f}$ , respectively. A simple calculation yields the characteristic polynomial of the coefficient matrix of system (3.3.4), it is

$$\begin{aligned} &\lambda^3 + \lambda^2 (2\mu_x + 2\mu_y + \sigma + \rho) \\ &+\lambda (\mu_x (\mu_x + 2\mu_y + \sigma + \rho) + \mu_y (\mu_x + \mu_y + \sigma + \rho)) \\ &+\mu_x \mu_y (\mu_x + \mu_y + \sigma + \rho)\,. \end{aligned}$$

All the coefficients are positive and one may check easily that condition (A1.1.1) is also satisfied, so that the equilibrium $(\bar{x}, \bar{y}, \bar{p})$ is globally asymptotically stable in $K_x \times \mathbf{R}_+$.

From the point of view of the spread of disease the average length of a partnership or the mean number of partners during a lifetime is crucial. In order

to estimate this we determine first the mean lifetime of a female and a male. In complete analogy to how the average infectious period has been shown to be the reciprocal of the recovery rate (Section 3.1 preceding (3.1.2)) from system (3.3.1) we obtain that the mean active lifetime of a female and a male is $1/\mu_x$ and $1/\mu_y$, respectively. For instance, if there are no pairs, males are not recruited, and at time $t$ there is just 1 male present, then at time $t + 1/\mu_y$ the number of males present will be $y(t + 1/\mu_y) \approx y(t) + \dot{y}(t)(1/\mu_y) = 1 - \mu_y \cdot 1 \cdot (1/\mu_y) = 0$. Similarly, the average duration of a partnership is $1/(\mu_x + \mu_y + \sigma)$. If no pairs are present at time $t$, and there is just 1 male, then from the third equation of system (3.3.4) $p(t + 1/\rho) - p(t) = p(t + 1/\rho) \approx \dot{p}(t)(1/\rho) = \rho \cdot 1 \cdot (1/\rho) = 1$, and thus, the average time needed for the formation of a pair ( for finding a partner) is $1/\rho$. As a consequence, the time consumed by the search for a partner plus the time spent in this partnership is $1/\rho + 1/(\mu_x + \mu_y + \sigma) = (\mu_x + \mu_y + \sigma + \rho)/(\rho(\mu_x + \mu_y + \sigma))$. The product of this duration and the average number of partners of a male during lifetime $N_y$ must be equal to the mean active lifetime of a male:

$$N_y(\mu_x + \mu_y + \sigma + \rho)/(\rho(\mu_x + \mu_y + \sigma)) = 1/\mu_y .$$

Hence, the average number of partners of a male during lifetime is

$$N_y = \frac{\rho}{\mu_y} \frac{\mu_x + \mu_y + \sigma}{\mu_x + \mu_y + \sigma + \rho} .$$

It is reasonable to suppose that the ratio of the average number of partners of a female and that of a male is equal to the ratio of their respective lifetimes divided by the ratio of their numbers in equilibrium:

$$\frac{N_x}{N_y} = \frac{1/\mu_x}{1/\mu_y} \frac{\bar{m}}{\bar{f}} = \frac{\kappa_y}{\kappa_x} \leq 1$$

by (3.3.3).

Dietz and Hadeler (1988) set up the model of pair formation and then built a model for the spread of the disease. Due to lack of space we can not present this model in detail here but we shall describe its main features. The population is divided into 8 groups: noninfected and infected females and males and pairs in which both partners are noninfected, both are infected, only the male or only the female, respectively, is infected. This way a system of eight dimensions is constructed in a fairly straightforward way. If there is no infection in the population the system reduces to system (3.3.1) or (3.3.4). A condition can be given for the stability of the disease-free equilibrium of the latter system given in the preceeding from the point of view of the eight-dimensional (8D) system. If one assumes that the rates of increase, death, infection, recovery etc. do not depend on the sex one obtains a simpler five-dimensional (5D) system of differential equations. Even in this lower dimensional model one may get explicit results by analytical methods only if recovery is excluded (which is the case now if one tries to apply the results to HIV). Under this assumption



a threshold condition can be given that implies the existence of an endemic equilibrium. With the data assumed by the authors 4 years of average duration of partnership is the threshold below which the endemic equilibrium persists. Four years of duration corresponds to an average of 12 partners per 50 years of active lifetime.

## 3.4 The Spread of Epidemics in Space

In the classical models of epidemics the "well stirredness" assumption made at the beginning of Section 3.1 plays a crucial role. In the previous Section, as long as sexually transmitted diseases were involved, we got rid of the "random mating" part of it by taking into account the dynamics of pair formation in the human population. Up to this point, however, the population was considered to be concentrated in one point and it was not taken into account that, in fact, the population has a spatial distribution on a continent, in a country, or even in a large town. In previous centuries diseases such as plague, cholera, or influenza swept over continents like a wave spreading from one place to the neighboring one and so on. To be sure, in the twenty-first century, distances probably are not as important as they were because a few passengers on a plane from, for example, Hong Kong who are carrying influenza may bring it into an American or European capital before it even reaches Shanghai. However, it remains undeniable that some diseases spread from place to place and people who live far away from a disease source may have a better chance of avoiding epidemics than those who live near the nucleus. Therefore, in this Section a model will be treated in which the spatial distribution of the population will be taken into account. This model was used by Murray (1989) to describe the "Black Death," the bubonic plague that swept through Europe from 1347 to 1350 and killed about one-quarter of the population (see Langer, 1964). We treat the model somewhat differently here but recommend reading of the vivid description of the case in the literature quoted.

We denote the *areal density of susceptibles and infectives* at time $t$ and at place $x$ by $S(t,x)$ and $I(t,x)$, respectively, and the *infection rate* and the *mortality of infectives* by $r > 0$ and $a > 0$, respectively. A SIR model is to be built but the equation for the removed is not written out and it is assumed that susceptibles and infectives move around, following Fick's diffusion law (see Appendix 3.2) with a diffusion rate $D > 0$. The following system of partial differential equations describes the dynamics:

$$\frac{\partial S}{\partial t} = -rSI + D\Delta S , \qquad \frac{\partial I}{\partial t} = rSI - aI + D\Delta I , \tag{3.4.1}$$

where $\Delta$ is the Laplace delta-if the space is 2D (the surface of a part of the earth considered to be a plane) and $x$ and $y$ are Cartesian orthogonal coordinates $\partial^2/\partial x^2 + \partial^2/\partial y^2$, and if the space is 1D (straight line in a certain direction) just the second derivative with respect to the spatial variable $\partial^2/\partial x^2$. If there is no diffusion, $D = 0$ then we get back system (3.1.1) without the equation

for the removed. Although the problem could be treated in the realistic two spatial dimension setting, in order to simplify the mathematics we are to treat it by assuming a 1D space. This is surely an abstraction but it still may give an insight into the dynamics of the propagation of disease if we suppose that the epidemic propagates from the nucleus uniformly in every direction. Further, we suppose that the domain where the disease propagates is infinite without boundary. Although the problem could be treated also by assuming, say, no flux boundary conditions, the complications at the boundary would increase the mathematical difficulties without much gain. It is assumed that the diffusion rates of the susceptibles and the infectives are the same. One may say that this is not too realistic because the sick do not move around. However, during the incubation period there is no difference in behavior among those who carry the disease and those who do not. During large plague epidemics in Europe there was massive emigration from the large towns, which served to accelerate propagation (the most famous literary evidence of this is, perhaps, the setting of Boccaccio's *Decameron*, in which a company of gentlemen and ladies who fled the plague for a country house tell each other spicy stories). The plague was carried also by rats, and no one knows now what was the diffusion rate of healthy and sick rats. In model (3.4.1) the susceptible population is considered to be constant if infectives are not present and on a far away boundary there is no in- and outflow of people, that is, no birth and death process apart from the epidemics is taken into account. We suppose that those who get the disease die; in case of the plague, indeed, 80–90% of those who fell ill did not recover. Under all these conditions we are to treat system (3.4.1) in one spatial dimension,

$$\frac{\partial S}{\partial t} = -rSI + D\frac{\partial^2 S}{\partial x^2}\,, \qquad \frac{\partial I}{\partial t} = rSI - aI + D\frac{\partial^2 I}{\partial x^2}\,. \tag{3.4.2}$$

We may simplify the equations by changing the scales introducing the new variables

$$h = \frac{S}{S_0}, \qquad v = \frac{I}{S_0}, \qquad \tau = rS_0 t, \qquad y = \left(\frac{rS_0}{D}\right)^{1/2} x\,, \tag{3.4.3}$$

where $S_0$ is the initial value of susceptible density. A simple calculation yields the system in the new coordinates:

$$\frac{\partial h}{\partial \tau} = -hv + \frac{\partial^2 h}{\partial y^2}\,, \qquad \frac{\partial v}{\partial \tau} = hv - bv + \frac{\partial^2 v}{\partial y^2}\,, \tag{3.4.4}$$

where $b = a/(rS_0)$, its reciprocal $rS_0/a$ is the contact rate multiplied by the density of susceptibles, giving the density of those infected by a unit density of infectives (cf. Section 3.1 and especially the discussion on condition (3.1.2)). The equilibria of system (3.4.4) are $(h, v) = (h_0, 0)$ with arbitrary $h_0 > 0$; however, $(h, v) = (1, 0)$ corresponds to the initial value $(S, I) = (S_0, 0)$. Rather than treating system (3.4.4) in a general way, we try to find only those solutions that describe the spread of the disease in the form of a wave, that is *traveling wave* solutions. These are the solutions that depend on $\tau$ and $y$ only through



the expression $z = y - c\tau$ with some constant $c > 0$. In the "space-time" plane $y, \tau$ along the straight lines $y - c\tau = z$ with a constant $z$ the state of the system is the same and the values of $h, v$ are constant. If we pick two points on one of these straight lines $(y_1, \tau_1)$ and $(y_2, \tau_2)$, that is, $y_1 - c\tau_1 = z = y_2 - c\tau_2$ and $\tau_1 < \tau_2$, for example, then this means that if the system was in a certain state at moment $\tau_1$ at place $y_1$ then it will be in the same state at moment $\tau_2$ at place $y_2$. This obviously means that the state of the system is propagating along these parallel lines with *velocity* $c = (y_2 - y_1)/(\tau_2 - \tau_1)$. If we are interested in solutions of the form $h(z) = h(y - c\tau)$, $v(z) = v(y - c\tau)$ then the system reduces to the system of ordinary differential equations:

$$\frac{d^2h}{dz^2} + c\frac{dh}{dz} - hv = 0\,, \qquad \frac{d^2v}{dz^2} + c\frac{dv}{dz} + (h - b)\,v = 0\,. \tag{3.4.5}$$

The equilibria of this system are also $(h_0, 0)$; in particular, $(1, 0)$ interests us. The question is, what are the solutions doing as $z$ tends to infinity and to minus infinity? If place $y$ is fixed then *as time $\tau$ tends to infinity the variable $z$ tends to minus infinity and vice versa and as $\tau$ tends to minus infinity $z$ tends to plus infinity.* System (3.4.5) will be linearized at $(1, 0)$ and the eigenvalues will be determined. First we put the system into Cauchy normal form introducing the new phase variables,

$$x_1 = h, \qquad x_2 = \dot{h}, \qquad x_3 = v, \qquad x_4 = \dot{v}\,.$$

The derivative with respect to $z$ is denoted with an overdot and we obtain

$$\begin{aligned} \dot{x}_1 &= x_2\,, & \dot{x}_2 &= -cx_2 + x_1x_3\,, \\ \dot{x}_3 &= x_4\,, & \dot{x}_4 &= -cx_4 + (b - x_1)\,x_3\,. \end{aligned}$$

Linearizing at (1,0) the characteristic polynomial turns out to be

$$p(\lambda) = \lambda\,(c + \lambda)\,\left(\lambda^2 + c\lambda + 1 - b\right)\,.$$

The eigenvalues are $0$, $-c$, $(1/2)\left(-c \pm \sqrt{c^2 + 4\,(b-1)}\right)$. That one of the eigenvalues is zero is no wonder because the equilibria fill in the axis $h$, which is the center manifold of dimension one of each (see Appendix 2.3). The third and fourth eigenvalues have negative real parts iff a number $< c^2$ stands below the square root, that is, if $b < 1$. Assuming this we must also suppose that the expression below the square root is nonnegative, that is, $c \geq 2\sqrt{1-b}$. Otherwise the solutions of the linearized system and with them the solutions of the original nonlinear system would oscillate around (1,0) with $x_1$ and $x_3$ assuming negative values that have no meaning. Under these assumptions the solutions tend towards the equilibria on the axis $h$ as $z$ tends to infinity, that is, time tends to minus infinity. This means that if the system is perturbed out of the equilibrium (1,0) by a small positive initial value of infectives $v$ then the solutions tend away from the equilibrium and a traveling wave of epidemics starts to

propagate as *time is increasing*. Summing up what has been established already here, a traveling wave of epidemics sweeps through a region if $a/(rS_0) < 1$ or

$$\frac{r}{a}S_0 > 1\,, \tag{3.4.6}$$

that is, if the density of those infected by a unit density of infectives during their infectious lifetime is greater than one. The minimal velocity of the wave sweeping through the population in the region is $c = 2\sqrt{1-b}$ in the transformed coordinates, or as $x/t = (rS_0D)^{1/2}\,y/\tau$,

$$c_{\text{real}} = 2\sqrt{rS_0D - aD} \tag{3.4.7}$$

in the original spatial and time scale.

We try to fit this model to the "Black Death" epidemic of 1347-1350 that swept through Europe, starting from the port of Genoa (Genova earlier), on the western coast of Italy and reaching Russia, the Baltic, and Sweden in only 3 years. Most of the data used here are taken from Murray (1989). Eighty to 90% of those who contracted the plague died on average within 12 days, so that the infectious period including the incubation is $1/a = 12\ days = 0.033\ years$, meaning that $a = 30\ year^{-1}$. The infection rate is estimated at $r = 0.4\ mile^2/year$. The population density of Europe was at that time estimated to be $50/mile^2$; however, because not only men but also rats carry the fleas that carry the disease one has to increase this number considerably, so that we double this figure to $S_0 = 100/mile^2$. This yields $b = 0.75$. From these data we obtain for the minimal transformed velocity of the epidemic wave $c = 1$. By Langer (1964) the average real velocity was $c_{\text{real}} = 400\ miles/year$. From Eq. (3.4.7) we may calculate the diffusion rate as $D = 4000\ mile^2/year$.

We have solved system (3.4.5) by PHASER using these data with two sets of initial values near $(h, v) = (1, 0)$ and small initial derivative values. The result is shown in Fig. 3.4.1. In plane $h, v$ the projection of the trajectory shows how the density of susceptibles falls from 1.05 or 0.96, respectively, to 0.75 and 0.65, respectively, as the epidemic sweeps over the land. This corresponds to the estimate mentioned at the beginning of this Section that $\approx$ 25% of the population died in the epidemic. Then we show the graph of the susceptibles and infectives as functions of $z$. The horizontal axis $z$ directed to the right is at the same time the axis $y$ of the spatial coordinate at a fixed moment $t$; far to the right-hand side the population is not yet affected and is at its original level and far to the left-hand side the epidemic had already swept through and the population declined. The horizontal axis *directed to the left* corresponds to the time axis $\tau$ at a fixed place $y$; far to the right, there was no epidemic yet and so population density remained at the original level; over time, that is, toward the left, the effect of the epidemic began to be felt and the population began to decrease. One may look similarly at the graph of the infectives. At the height of the epidemic the maximal density of infectives was 0.034, and 0.022, respectively, meaning that $\approx$ 3% of the population was infected at the same time.



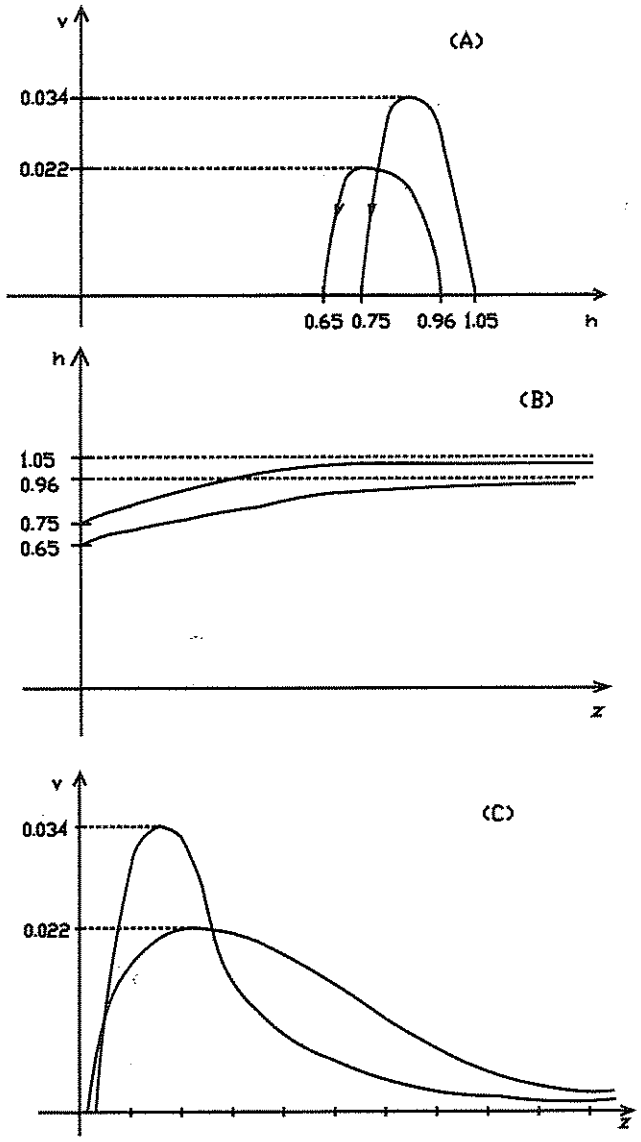


Figure 3.4.1: The traveling wave of an epidemic: system (3.4.5) $c = 1, b = 0.75$. (A) The trajectories in the plane $h, v$ of susceptibles and infectives starting near $(1.05, 0)$ or $(0.96, 0)$ and ending at $(0.75, 0)$ or $(0.65, 0)$, respectively. (B) the graphs of the functions $h(z)$ of the two solutions; and (C) the graphs of the functions $v(z)$ of the two solutions; $z \sim y \sim -\tau$ (PHASER).